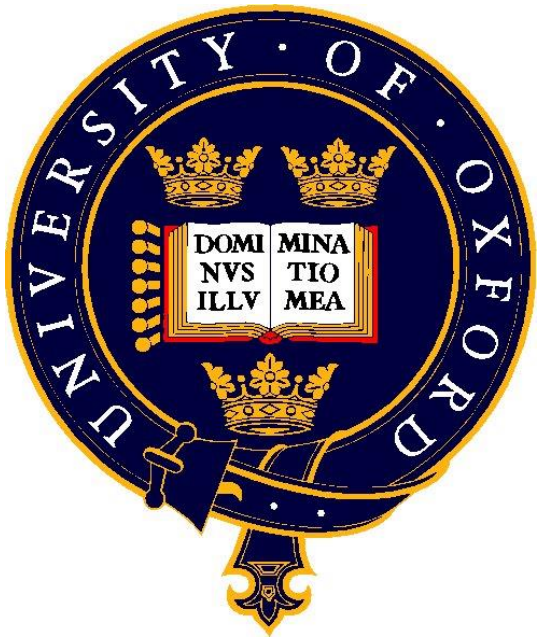
# Sesquiterpene Oxidation by Evolved CYP102A1 ($P450_{BM3}$) Monooxygenase




A Thesis submitted for the degree of
Doctor of Philosophy

Yang Cao

St Hugh's College

University of Oxford

Michaelmas Term 2018

# Abstract

Cytochrome P450 enzymes (CYPs) catalyse the oxidation of C–H bonds by inserting an oxygen atom from atmospheric dioxygen under mild conditions. The amino acid sequences of CYPs are highly diverse, so much so that the only conserved residue is the proximal cysteine that provides the thiolate ligand to the heme iron, but their structural folds are highly conserved. Hence, although one P450 enzyme or a family of enzymes, may be very substrate specific, the super family of P450 enzymes oxidise a very wide range of organic compounds due to the highly diversified sequences. CYP102A1, also known as $P450_{BM3}$ as the 3$^{rd}$ P450 enzyme identified from *Bacillus megaterium*, was the first discovered self-sufficient P450 which has been extensively studied as a biocatalyst for decades. Site-directed, site-saturation and random mutagenesis have been applied to create CYP102A1 variants to expand the substrate range and explore new reaction types. In synthetic chemistry, the functionalisation of naturally occurring compounds, such as steroids and terpenes, remains a huge obstacle for chemical synthesis methodologies because of the high energy reactants, harsh reaction conditions, and difficulty in controlling product selectivity. Engineered CYP102A1 variants with high selectivity and wide substrate range have great potential in synthesis. This thesis consists of six chapters and reports the results of research in sesquiterpene oxidation by CYP102A1variants and process development to scale up a laboratory scale reaction to the point where it is suitable for pilot scale development for industrial production. The mutation and experimental design for screening and engineering oxidation activity and selectivity for valencene, α-humulene and β-caryophyllene are described in Chapter 3.

Allylic oxidation and epoxidation dominated the reaction outcomes of these 3 substrates with the initial library of 24 variants. Having more allylic C–H bonds and C=C double bonds as target positions, 8 out of 11 potential sites in humulene were oxidised, including C3 oxidation to give the C3 alcohol the precursor to Zerumbone that has antitumor activity. β-Caryophyllene, with a fewer active positions, had 6 out of 12 potential targets oxidised by this 24-variant library. α-Cedrene has the fewest allylic C–H bonds and C=C double bonds, and only 3 out of 12 potential sites were oxidised. Therefore, α-cedrene was selected for further investigation and mutagenesis, generating expanded libraries of variants, leading to oxidation at unactivated aliphatic C–H bonds on the 5-membered ring. In addition, the C15 allylic methyl group was oxidised to the allylic alcohol, then the α,β-unsaturated aldehyde and carboxylic acid, potentially opening up routes to sesquiterpene lactones via hydroxylation then spontaneous cyclisation. Chapter 5 details the scaling up of valencene oxidation by one enzyme variant to 3 litre scale, with high valencene conversion at high product concentration, thus laying the foundation for an economically viable industrial process for the production of natural grade nootkatone. Chapter 6 provides a summary and suggestions for further research.

# Acknowledgement

I wish to express my sincere gratitude to the University of Oxford and St Hugh's College for accepting me as a graduate student and providing such a wonderful opportunity to carry on my research career in Oxford.

Great thanks to Dr Kouji Urata, Dr Alizé Pennec and Dr Mohamed Issouf being such wonderful colleagues in the group and Oxford Biotrans.

I would like to thank Wenyu Chen, Yushu Li, Matthew Fisher, Yuan Zhang and Alice Wong for all the joy memory in T4a. Wish you all a splendid future. Thanks to Georgia Lubbock and George Woodward for your great contribution to this project in your Part II year.

The beer tasting from different brewers with Dr Yang Xia was greatly enjoyable across so many pubs in Oxford. Thanks to Dr Yixing Wu, Dr Xiaochun Meng and Dr Yifan Cai for a marvellous year in 3 Great Mead. There are so many names that I should mention and all your encouragement, humorous, intelligence and spirit will forever stay with me.

I would like to make three special thanks at the last.
The first one is to my parents and my great family that I could not become who I am without your support and encouragement. The love and care from my family taught me that one should never give up.

A special thanks to my supervisor, Prof. Luet Wong, a wise and tolerant scholar who is always ready to advise and help me whenever I felt lost, depressed or helpless. I could not be more thankful for what you have done for me during my DPhil time in Oxford.

The last, to my love for life, Xue Min, we were so unlucky that missed the galaxy in the Sahara Desert, the aurora in Iceland and so many magnificent views in our trips. Finally, it is because we used up all our fortune to find each other in our lives.

# Table of contents

**Chapter 1 Introduction** ..... 1

1.1 Biocatalysis in synthesis ..... 2

1.2 Transition metal-catalysed C–H activation ..... 6

1.3 Cytochrome P450 enzymes: Direct C–H activation ..... 8

1.3.1 The P450 catalytic cycle ..... 9

1.3.2 Electron transfer chains in P450 systems ..... 11

1.4 $P450_{BM3}$ ..... 13

1.4.1 Crystal structures of wild-type $P450_{BM3}$ ..... 14

1.4.2 High-level $P450_{BM3}$ expression and large-scale biotransformation ..... 16

1.4.3 Rational design for unnatural substrate oxidation ..... 17

1.4.4 Site-saturation and random mutagenesis: Directed evolution ..... 18

1.4.5 Combinatorial and iterative saturation mutagenesis ..... 27

1.4.6 The mutant library approach ..... 29

1.5 Terpenoid compounds ..... 32

1.6 Thesis objectives ..... 36

**Chapter 2 Materials and Methods** ..... 38

2.1 Chemicals, Solutions and Cell Strains ..... 39

2.1.1 Chemicals, solvents and reagents ..... 39

2.1.2 Buffers, solution and growth media ..... 39

2.1.3 Equipment and analysis methods ..... 41

2.2 DNA and mutagenesis methods ..... 43

2.2.1 Plasmids and mutant generation ..... 43

2.2.2 Recombinant protein production and purification ..... 46

2.3 *In vitro* screening, preparative scale reaction and product characterization ..... 48

2.3.1 In vitro screening ..... 48

2.3.2 Preparative scale reactions and product purification ..... 49

2.4 Final conditions for valencene scale up study ..... 50

2.4.1 100 mL scale reaction conditions ..... 50

2.4.2 Reaction conditions at 3 litre scale in a bioreactor ..... 51

**Chapter 3 Engineering $P450_{BM3}$ for Sesquiterpene Oxidation** ..... 53

3.1 Sesquiterpene oxidation ..... 54

3.2 Mutants and general methods ..... 56

3.2.1 A $P450_{BM3}$ mutant library ..... 56

3.2.2 Activity screening and product analysis ..... 58

3.3 Humulene and caryophyllene ..... 58

3.3.1. Humulene screening result and product isolation ..... 58

3.3.2. Caryophyllene oxidation and product characterisation ..... 62

3.3.3. Product characterisation ..... 64

3.3.4 Discussion ..... 68

3.4 Valencene Screening Results ..... 71

**Chapter 4 Cedrene Oxidation: Generating Diversity** ..... 76

4.1. α-Cedrene derivatives and the woody note ..... 77

4.2. Initial Screening of α-Cedrene ..... 80

4.2.2 Summary ..... 87

4.3. Construction of new libraries of variants ..... 88

4.3.1 Variants with F87, A328 and A330 substitutions ..... 88

4.3.2 A sub-library of F87/A328 Variants ..... 93

4.3.2 A F87–I263 combinatorial library ..... 103

4.4 Discussion ..... 112

**Chapter 5 Valencene Oxidation: Scaling Up Study of $P450_{BM3}$ Biotransformation** ..... 115

5.1 The grapefruit flavour compound nootkatone ..... 116

5.1.1 Nootkatone synthesis ..... 117

5.2 Oxidation of valencene to nootkatone catalysed by $P450_{BM3}$ ..... 120

5.3 The effect of organic co-solvents ..... 124

5.4 Phase transfer catalysts ..... 129

5.4.2 The effect of cyclodextrins ..... 132

5.4.3 The effect of aeration and pH control ..... 135

5.5 Further optimisation ..... 139

**Chapter 6 Conclusions** ..... 139

**References** 148

**Appendix** 156

Appendix 1 Variants and mutations ..... 156

Appendix 2 NMR result of products ..... 160

# List of Figures

**Fig. 1.1 Examples of asymmetric synthesis using KREDs ..... 3**

**Fig. 1.2 Synthetic pathway of (*R*)-3-hydroxythiolane. ..... 4**

**Fig. 1.3 A biocatalytic asymmetric reduction route to montelukast sodium. ..... 4**

**Fig. 1.4 The first-generation process of sitagliptin synthesis ..... 5**

**Fig. 1.5 Transition metal-catalysed C–C bond formation ..... 6**

**Fig. 1.6 The generally accepted P450 catalytic cycle ..... 10**

**Fig. 1.7 P450 electron transfer chains ..... 12**

**Fig. 1.8 Dodecanoic acid, the natural substrate of $P450_{BM3}$ and oxidisation position. ..... 13**

**Fig. 1.9 The active site structure of the *N*-palmitoyl glycine (NPG)-bound form of the heme domain of wild type $P450_{BM3}$ ..... 14**

**Fig. 1.10 *p*-Nitrophenolate screening procedure for $P450_{BM3}$ directed evolution. ..... 20**

**Fig. 1.11 Examples of drug molecule oxidised by $P450_{BM3}$ variants library ..... 22**

**Fig. 1.12 Oxidation selectivity pattern of M01 variants as testosterone as substrate. ..... 23**

**Fig. 1.13 Overlay of the crystal structures of the substrate-free form of the WT (grey) and the A330P mutant (yellow) of $P450_{BM3}$ ..... 24**

**Fig. 1.14 Overlay of the structures of the substrate-free (SF) form of the WT (grey) and KT2 mutant (yellow) of $P450_{BM3}$. ..... 26**

**Fig. 1.15 Overlay of the crystal structures of the substrate-free forms of the WT (grey) and I401P (yellow).. ..... 27**

**Fig. 1.16 Examples of drug metabolism, fragment molecule oxidation, and new P450 reaction ..... 32**

**Fig. 1.17 Biosynthesis of terpenes and terpenoids. ..... 33**

**Fig. 1.18 Examples of terpenoid compounds with flavour, fragrance and medicinal properties. ..... 34**

**Fig. 1.19 Synthesis of artemisinin ..... 35**

**Fig. 3.1 Sesquiterpenes selected for P450-catalysed oxidation ..... 54**

**Fig. 3.2 Proposed reaction route to hydroxylate humulene and valencene to zerumbone and nootkatone ..... 55**

**Fig. 3.3 Two views of the active site structure of NPG bound wild-type $P450_{BM3}$. ..... 57**

**Fig. 3.4 GC profile of α-humulene initial screen. ..... 59**

**Fig. 3.5 GC analysis of preparative scale reactions of α-humulene ..... 61**

**Fig. 3.6 GC profile initial screen. ..... 62**

**Fig. 3.7 The GC profile of preparative scale reactions of β-caryophyllene ..... 63**

**Fig. 3.8 The $^1H$ NMR spectrum of product 2 in $CDCl_3$. ..... 64**

**Fig. 3.9 The $^1H$ NMR spectrum of product 3 in $CDCl_3$. ..... 65**

**Fig. 3.10 The $^1H$ NMR spectrum of product 10 in $CDCl_3$ ..... 66**

**Fig. 3.11 The $^1H$ NMR spectrum of product 13 in $CDCl_3$ ..... 67**

**Fig. 3.12 The $^1H$ NMR spectrum of product 15 in $CDCl_3$ ..... 68**

**Fig. 3.13 Selectivity comparison of variants with F87A and F87V mutations for the oxidation of humulene and Caryophyllene. ..... 70**

**Fig. 3.14 GC profile of selected variants showing selectivity of valencene oxidation ..... 72**

**Fig. 3.15 Valencene oxidation profile of the RKA variant. ..... 72**

**Fig. 3.16 Selected GC profile of valencene oxidation by $P450_{BM3}$ variants ..... 73**

**Fig. 3.17 Mutation study of C2 oxidation of valencene to form nootkatone. ..... 74**

**Fig. 4.1 Main cedranoid compounds in cedar oil and acetalyation product of α-cedrene ..... 77**

**Fig. 4.2 Chemical synthesis of oxygenated α-cedrene derivatives. ..... 78**

**Fig. 4.3 Functionisation of α-cedrene from microbial biotransformation. ..... 79**

**Fig. 4.4 Characterised cedrene oxidation products and the product profile. ..... 81**

**Fig. 4.5 WT $P450_{BM3}$ compared with 3 other variants selected for preparative reactions ..... 82**

**Fig. 4.6 The $^1H$ NMR spectrum in $CDCl_3$ of the allylic alcohol 15-hydroxy-cedrene ..... 84**

**Fig. 4.7 Product profile (GC) of variants containing F87 mutations ..... 85**

**Fig. 4.8 The $^1H$ NMR spectrum of 37 in $CDCl_3$ ..... 86**

**Fig. 4.9 Product profile (GC) of variants based on R19/F87A ..... 89**

**Fig. 4.10 The $^1H$ NMR spectrum of 31 in $CDCl_3$ ..... 90**

**Fig. 4.11 The $^1H$ NMR spectrum of 35 in $CDCl_3$ ..... 91**

**Fig. 4.12 The $^1H$ NMR spectrum of 34 in $CDCl_3$ and the possible mechanism of its formation ..... 92**

**Fig. 4.13 Product profile of variants containing A328F with different F87 mutations ..... 95**

**Fig. 4.14 GC product profile of variants based on R19/F87I and R19/F87L ..... 97**

**Fig. 4.15. GC results of R19/F87G/A328L and R19/F87G. ..... 98**

**Fig. 4.16 The construction cogitation of GVQ library ..... 100**

**Fig. 4.17 Major products from GVQ variants library and GC results of each variant. ..... 101**

**Fig. 4.18 The construction cogitation of the K19 library ..... 104**

**Fig. 4.19 GC result of selected variants screening reactions ..... 105**

**Fig. 4.20 GC profile of K19/FG/IV and K19/FA/IG. ..... 106**

**Fig. 4.21 The $^{1}$H NMR spectrum of 32 in $CDCl_3$ ..... 107**

**Fig. 4.22 The $^{1}$H NMR spectrum of 41 in $CDCl_3$ ..... 108**

**Fig. 4.23 The $^{1}$H NMR spectrum of 40 in $CDCl_3$ ..... 109**

**Fig. 4.24 The $^{1}$H NMR spectrum of 39 in $CDCl_3$ ..... 110**

**Fig. 4.25 Products from K19/FG/IV and K19/FA/IG ..... 111**

**Fig. 4.26 Mono-hydroxylated products and the corresponding variants with selectivity ..... 113**

**Fig. 5.1 Structure of (+)-valencene and (+)-nootkatone ..... 116**

**Fig. 5.2 Total synthesis of racemic nootkatone. ..... 117**

**Fig. 5.3 Total synthesis of (+)-nootkatone. ..... 117**

**Fig. 1.4 Standard curve of nootkatone concentration ..... 121**

**Fig. 5.5 Oxidation of valencene by the $P450_{BM3}$ mutant RKA. ..... 123**

**Fig. 5.6 GC analysis of valencene oxidation at 1 mL scale ..... 125**

**Fig. 5.7 The effect of organic co-solvents.. ..... 128**

**Fig. 5.8 GC analysis of the effect of PEG-400 concentration on valencene oxidation ..... 131**

**Fig. 5.9 The result of continuously feeding valencene from a PEG stock ..... 132**

**Fig. 5.10 Structure of α-, β- and γ-cyclodextrin and their sizes of cavity. ..... 133**

**Fig. 5.11 The effect of HPβCD on valencene conversion and product distribution. ..... 135**

**Fig. 5.12 Effect of aeration. ..... 136**

**Fig. 5.13 The GC plots and integration result of the reaction with aeration and pH adjustment. .. 138**

**Fig. 5.14 The formation of gluconic acid from NADPH regeneration system ..... 139**

# List of Tables

Table 1.1 **The conversion and selectivity of 16α- and β-hydroxylation of testosterone ..... 29**

Table 1.2 **P450BM3 variants which gave absolute selectivity for 7(R)- and 7(S)-hydroxy-artemisinin31**

Table 2.1 **Media and buffer components ..... 40**

Table 2.2 **Standard reaction setup for PCR reaction ..... 44**

Table 2.3 **Reaction components for sesquiterpene oxidation screening ..... 48**

Table 2.4 **Reaction condition of 100 mL scale ..... 51**

Table 2.5 **Reagents in 3 L valencene oxidation reaction ..... 52**

Table 3.1 **The selection of $P450_{BM3}$ variants for sesquiterpene oxidation. ..... 57**

Table 3.2 **Selectivity of mono-oxidation product of humulene and caryophyllene from selected variants. ..... 69**

Table 4.1 **Product distribution of WT, RP and KT2 variants ..... 83**

Table 4.2 **List of $P450_{BM3}$ variants with mutations at A328 and A330 added to R19/FA..... 88**

Table 4.3 **Conversion of variants containing different F87 mutations combined with n A328F mutation. ..... 94**

Table 4.4 **The activity and selectivity of R19/F87I and R19/F87L based variants. ..... 96**

Table 4.5 **Summarised results of R19/F87A and R19/F87G based variants. ..... 99**

Table 4.6 **Variants of the GVQ sub-library. ..... 100**

Table 4.7 **Variants of the K19 library. K19 = H171L/Q307H/N329Y ..... 104**

Table 5.1 **Percentage of valencene, nootkatol and nootkatone ..... 121**

Table 5.2 **Peak area of internal standard and nootkatone ..... 122**

Table 5.3 **The effect of co-solvents, surfactant and sonication on valencene oxidation. ..... 126**

Table 5.4 **Reagents in 100 mL valencene oxidation reaction. ..... 140**

Table 5.5 **Reagents in 3 L valencene oxidation reaction. ..... 141**

Table 5.6 **Comparison between the conditions developed in this work with the best published data.142**

# Abbreviations

| | |
|---|---|
| CYP | cytochrome P450 |
| WT | wild type |
| KRED | ketoreductases |
| GDH | glucose dehydrogenase |
| TON | turnover number |
| NAD(P)H | nicotine adenine dinucleotide (phosphate), reduced form |
| Fdx | ferredoxin |
| FdR | ferredoxin reductase |
| FAD | flavin adenine dinucleotide |
| FMN | flavin mononucleotide |
| PdR | putidaredoxin reductase |
| Pdx | putidaredoxin |
| SF and SB | substrate-free and substrate-bound |
| NPG | N-palmitoyl glycine |
| ep-PCR | error-prone polymerase chain reaction |
| CAST | combinatorial active-site saturation test |
| ISM | iterative saturation mutagenesis |
| GC | gas chromatography |
| FID | flame ionisation detector |
| HPLC | high performance liquid chromatography |
| SRS | substrate recognition sites |

| | |
|---|---|
| IPP | isopentenyl pyrophosphate |
| GPP | geranyl pyrophosphate |
| DMAPP | dimethylallylpyrophosphate |
| GGPP | geranylgeranyl pyrophosphates |
| SOB | super optimal broth |
| SOC | super optimal catabolite repression |
| IPTG | isopropyl-β-D-thiogalactopyanoside |
| TLC | thin-layer chromatography |
| dNTP | deoxynucleotide |
| HPβCD | hydroxylpropyl-β-cyclodextrin |
| COSY | correlation spectroscopy |
| HSQC | heteronuclear single quantum correlation spectroscopy |
| GMO | genetically modified organism |
| PTC | Phase transfer catalyst |
| PEG | Polyethylene glycols |
| DMSO | dimethyl sulfoxide |

# Chapter 1 Introduction

## 1.1 Biocatalysis in synthesis

Modern synthetic chemistry focuses on more complex molecules and the development of enantioselective methodologies. The chemical and pharmaceutical industries strive to develop new processes for fine chemicals, drugs and medicines. The thalidomide debacle in the late 1980's revealed to the public the importance of the enantiomeric purity of drug molecules for safe therapeutic efficacy.[1,2] The ingenuity of the synthetic chemist and chemical engineer has led to highly optimised chemical processes for the synthesis of even the most complex molecules. Nevertheless, classical organic synthesis methods are still limited by the frequent need for protection/deprotection strategies to control structural elaboration, and stereoselective synthesis remains a significant challenge.[3,4]

Biocatalysis, the way of Nature to transform molecules, was first utilised in the early 18$^{th}$ century by Louis Pasteur who used a bacterial culture to produce tartaric acid.[5] Enzymes as the functional biocatalyst have been studied for decades as isolated form or inside cells. There is increasing interest in biocatalytic applications due to the many advantages over chemical routes such as control of chemo-, regio- and sterero-selectivity, mild reaction conditions and minimal waste generation. From triose isomerase in the sugar industry to lipases and dehydrogenases, naturally occurring enzymes have been applied on an industrial scale as efficient and environmentally friendly catalysts.[4,6,7]

Despite the advantages of enzymatic biotransformation, the limitation of scaling reactions from laboratory research to industrial production remains. There is no

biological process in Nature which has the same scale in volume and product concentration as in industry. Enzymes are easily accessed by a substrate in a reaction *in vitro*. However, most enzymes are unstable to inactivation when the proportion of hydrophobic organic compounds in aqueous solution increases. Whole cells can protect enzymes from organics in the medium. However, cell viability must be maintained during the biotransformation phase and cellular engineering is often needed to modify the cell wall to increase the accessibility of hydrophobic substrates, e.g. by expression of porins into cell walls.[8]

Some of the most successful applications of biotransformation involve the use of engineered ketoreductases (KREDs) for asymmetric synthesis in the production of pharmaceuticals. KREDs are a large class of enzymes which could reduce ketones to alcohols. Combined with NAD(P)H cofactor regeneration by glucose/glucose dehydrogenase (GDH), KREDs developed by directed evolution have found important applications in enantioselective synthesis of chiral alcohols (Fig. 1.1).[7,9]

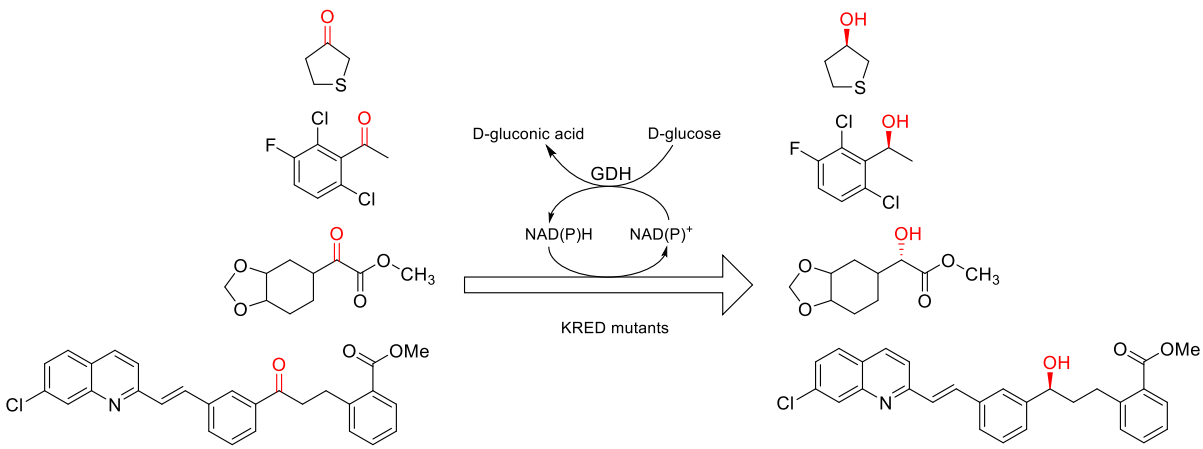



**Fig. 1.1** Examples of asymmetric synthesis using KREDs.

Screening of collections of KREDs followed by directed evolution led to a variant that reduced tetrahydrothiophene-3-one enantioselectively to (*R*)-3-hydroxy-thiolane (**A**, Fig. 1.2) with 99.3% *ee* and in 85% yield. (*R*)-3-Hydroxythiolane is a key intermediate in the synthesis of the antibiotic sulopenem and can be chemically synthesised in 5 steps starting from an acyclic precursor.[10] Clearly, the one-step asymmetric process using an engineered KRED is a superior process.

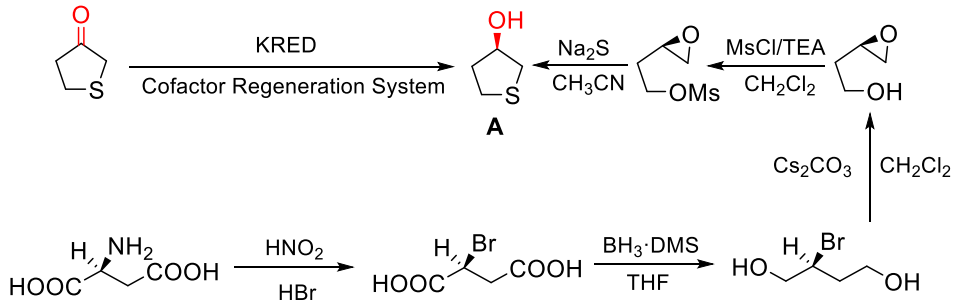



**Fig. 1.2** Synthetic pathway of (*R*)-3-hydroxythiolane.

An engineered KRED was applied to the synthesis of montelukast sodium, a leukotriene receptor antagonist for the treatment of asthma and allergies.[11] Asymmetric reduction of (*E*)-methyl-2-(3-(3-(2-(7-chloroquinolin-2-yl)vinyl)phenyl)-3-oxopropyl)-benzoate to the key (*S*)-alcohol **B** (Fig. 1.3) greatly simplified the process compared to the chemical route, with a lower catalyst loading and higher yield.[12]

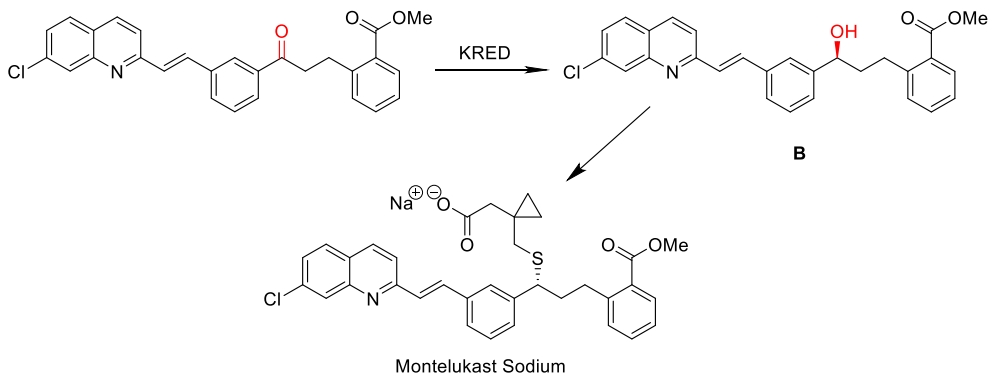



**Fig. 1.3** A biocatalytic asymmetric reduction route to montelukast sodium.

Transaminases have found increasing applications in asymmetric amine synthesis. Sitagliptin, a dipeptidyl peptidase IV (DPP-4) inhibitor in the treatment of type II diabetes mellitus, was chemically produced in 52% yield over 8 steps at a >100 kg scale.[13-15] The major limiting factor was the introduction of the chiral amine group (Fig. 1.4).[16] Alternative chemical routes still required high pressure hydrogenation and use of heavy metals. An engineered transaminase was introduced in sitagliptin synthesis to solve the problem of forming the chiral amine.[17] Directed evolution of the wild type enzyme from *Arthrobacter* sp. KNK168,[18] which has no activity on the key precursor prositagliptin ketone, led to a mutant that gave the desired product in 92% yield and >99.95% *ee*.

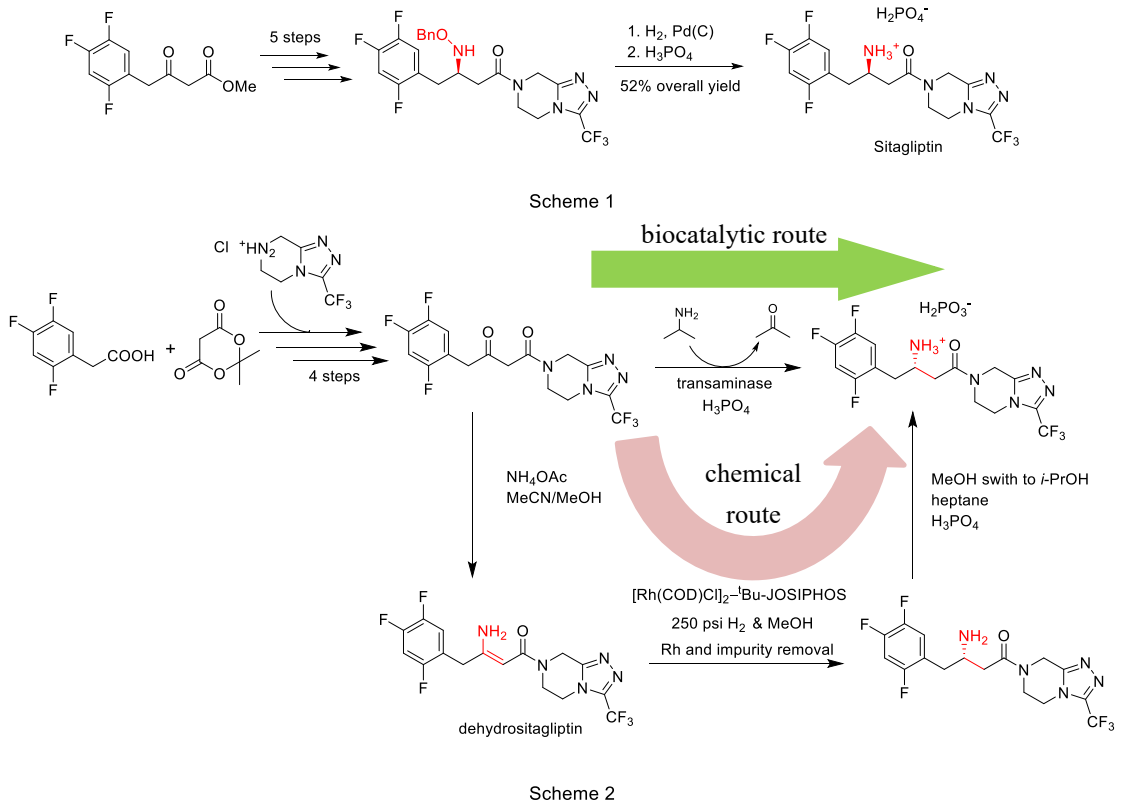



**Fig. 1.4** Scheme 1. The first-generation process of sitagliptin synthesis. Scheme 2. Multi-step chemical synthesis of sitagliptin via dehydrositagliptin compared with the biocatalytic route.

## 1.2 Transition metal-catalysed C–H activation

As the most basic building block of organic compounds, C−H bonds are chemically stable and unreactive due to their low bond polarity and high bond dissociation energy (~430 kJ $mol^{-1}$). Direct C–H bond functionalisation forming C–C, C–O and C–N bonds provides synthetic handles to elaborate structures to produce diverse properties.

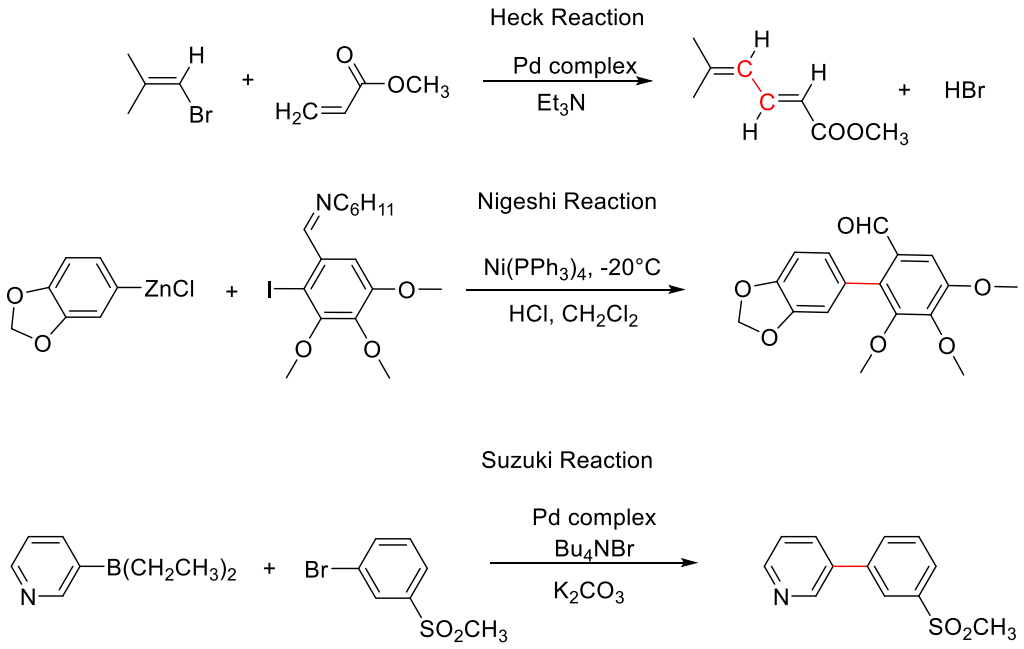



**Fig. 1.5** Transition metal-catalysed C–C bond formation via the Heck reaction, Negishi coupling, and Suzuki coupling.

Palladium (II) complexes (or other transition metal such as nickel in the Negishi reaction) can catalyse C−O and C−C bond formation with the assistance of an oxidant. Palladium complexes can be regenerated with Cu(II) and oxygen in a Pd(II)/Pd(0) catalytic cycle. No prior catalyst activation is needed. These reactions are relatively environmentally friendly and of a reasonable efficiency. As a result, Pd- and Ni-catalysed reactions are widely applied in many industrial processes.

Development of new approaches and variations of transition-metal catalysed C–C and C–N bond formation via C–H bond activation is one of the most active area of synthesis research, with numerous reports of refinements and broadening of the scope of these reactions. Some fundamental drawbacks remain: catalyst loading is relatively high (low turnover numbers, TON), lack of control of regio- and enantio-selectivity (although this is mitigated in some instances with directing groups), narrow substrate scope and functional group compatibility, and dependence of individual reactions on specific substrates or reaction conditions. On the other hand, once a set of conditions has been established for a particular selectivity pattern these reactions can be reliable and applicable to many, sometimes hundreds of related compounds.

A general pathway for C–H functionalisation is selective oxidation to an alcohol or carbonyl compound which can be converted readily to other functional groups, including C–C bond formation chemistry. Classical chemical methods are mostly stoichiometric and require high-energy reactants such as peroxides and high-valent metal species, often leading to low yield and poor selectivity. Therefore, selective catalytic C–H oxy-functionalisation under mild conditions without the need for high-energy compounds is highly desirable in the modern era of green and environmentally friendly processes.

## 1.3 Cytochrome P450 enzymes: Direct C–H activation

The cytochrome P450 (P450s) superfamily of heme monooxygenases catalyse the insertion of an oxygen atom from atmospheric oxygen into C–H bonds to form the alcohol.

$$R–H + 2H^+ + O_2 + 2e^- \rightarrow R–OH + H_2O$$

This reaction has no equivalent in classical synthetic methodologies. Small molecule catalysts tend to have low activity, low turnover numbers, lack selectivity, and where activity is observed it is often difficult to control the regio- and stereo-selectivity. The substrate range of P450 enzymes is extremely diverse, ranging from short chain alkanes and fatty acids to alicyclic, aromatic and heterocyclic compounds, cyclic and acyclic terpenes, peptides, steroids and macrolides.[19-24] P450 enzymes are found in almost every living organism in which they are involved in the biosynthesis of steroid hormones,[25] vitamin D homeostasis,[26] secondary metabolite production and degradation of xenobiotics.[27] At the time of writing there are >300,000 *CYP* genes in genome databases and the number will reach 1 million as more genome sequences of angiosperms are completed. *Escherichia coli* is devoid of P450 enzymes, which is a great advantage for P450 heterologous expression.

A scheme for systematic nomenclature of the superfamily of P450 enzymes based on sequence homology has been developed.[28] Enzymes with >40% amino acid sequence identity are grouped into families, e.g. CYP1, CYP101. Those that share >55% identities are put into sub-families, e.g. CYP2A, CYP2C. Each enzyme is then designated a number, e.g. CYP2C5, CYP2C19. This classification scheme is

invaluable in bioinformatic and phylogenetic analysis since, in general, enzymes within the same family have similar activity and function.

The C–H bond oxy-functionalisation activity of P450 enzymes has been extensively researched for applications in synthesis. Perhaps the most commonly known physiological role of P450 enzymes is in drug metabolism. Over 90% of clinical drugs are metabolised by human P450 enzymes, and drug metabolite synthesis is one of the most commonly tested activity of P450 enzymes from all sources. The engineering of $P450_{BM3}$ from *Bacillus megaterium* for potential biocatalytic applications is the subject of this thesis.

### 1.3.1 The P450 catalytic cycle

In the resting state the P450 heme iron is axially coordinated on the proximal side with a cysteine thiolate ligand. The distal heme axial position is occupied by a water ligand, making the heme iron hexa-coordinated and low spin (Fig. 1.6, **I**). Binding of a substrate close to the heme causes the dissociation of this coordinated water from the iron centre which becomes penta-coordinated and high spin (**II**). The heme redox potential increases, and electron transfer can occur to this species without the need to reorganise the iron ligand environment. Both factors promote electron transfer.

An electron supplied by NAD(P)H is shuttled to the heme to form the $Fe^{II}$ form (**III**). Dioxygen binds to give the oxy complex **IV**. A second electron is transferred, giving the $Fe^{III}$-peroxy complex (**V**) which is rapidly protonated to form the $Fe^{III}$-hydroperoxy complex (**VI**), or Compound 0. The distal oxygen is then protonated and the O–O bind is cleaved heterolytically to form the ferryl, Compound I species

$Fe^{IV}(=O)P^{\bullet+}$ (**VII**) which is the active intermediate for C–H bond activation.[29,30]

If a substrate binds at an appropriate position the hydrogen atom closest to the ferryl oxygen is abstracted to give a substrate radical and the $Fe^{IV}$–OH, Compound II species.[31] Fast recombination of the substrate radical and •OH forms the alcohol and returns the heme iron to the $Fe^{III}$ state; this is the oxygen or radical rebound mechanism of P450-catalysed C–H bond activation.[32] The hydroxylated product may be released and a water molecule re-coordinates to the heme iron to regenerate the resting state. Alternatively, another substrate molecule can displace the product from the active site and start another cycle of reaction without the enzyme returning to the resting state.

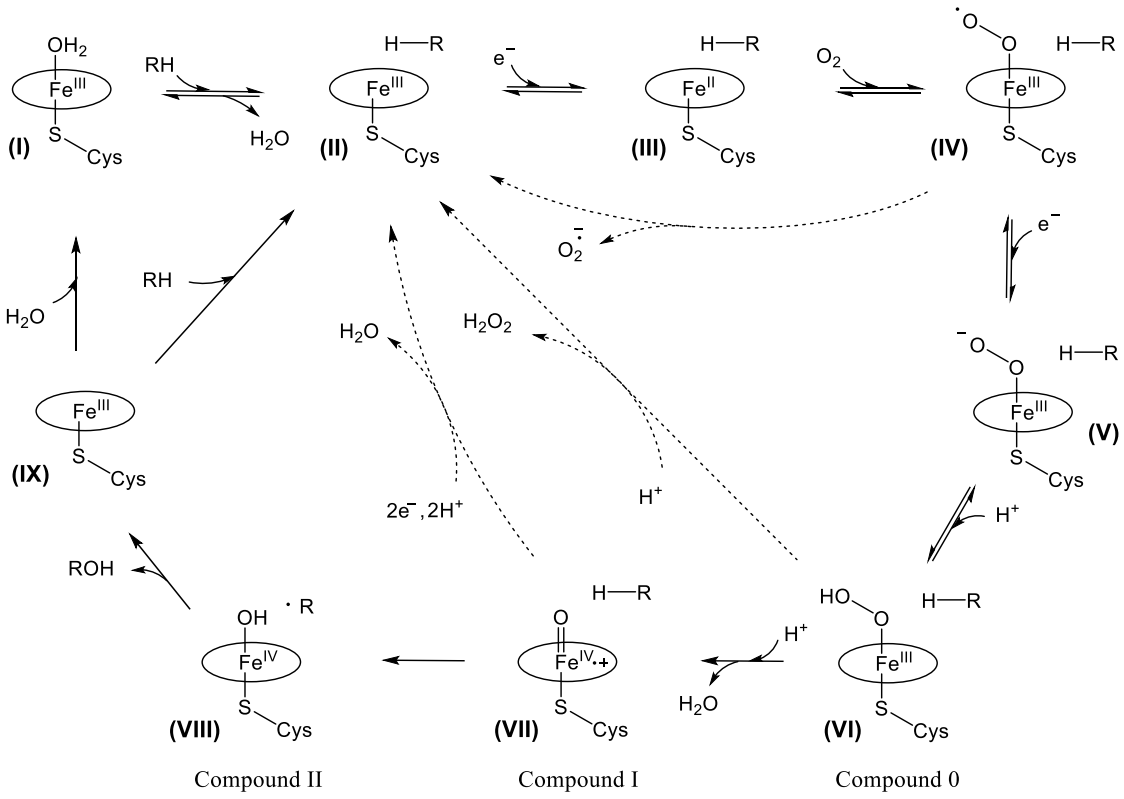



**Fig. 1.6** The generally accepted P450 catalytic cycle showing the sequential electron transfer steps, O–O bond cleavage and the radical rebound mechanism.

The P450 catalytic cycle is known to undergo uncoupling whereby the electrons supplied by NAD(P)H are not utilised for substrate oxidation. In the first pathway, the ferrous-oxy complex **IV** decomposes to form superoxide and the ferric heme, thus uncoupling one electron from the full cycle. This may occur if the enzyme does not adopt the correct conformation to stabilise oxygen binding or the substrate could sterically interfere with the Fe–$O_2$ interaction. In the second pathway, the hydroperoxide anion dissociates from the hydroperoxy intermediate **VI**, e.g. by protonation at the Fe-bound oxygen, and both electrons from NAD(P)H are lost. In this pathway the P450 enzyme displays peroxygenase activity. Finally, the ferryl species **VII** may be reduced by two electrons back to the ferric–$OH_2$ species, e.g. if the substrate is bound too far away from the ferryl oxygen or a chemically inert part of the substrate such as a C–F bond is closest to the ferryl. A total of four electrons are uncoupled from substrate oxidation and the P450 enzyme displays oxidase activity. These uncoupling pathways would be minimised if the enzyme–substrate interactions promote the adaptation of the necessary conformations for monooxygenase activity. For P450 enzymes that have been evolved for specific functions the reactions are tightly coupled, e.g. the $P450_{cam}$ (CYP101A1) enzyme shows 100% coupling for camphor oxidation, but xenobiotic detoxification enzymes tend to show low coupling.

### 1.3.2 Electron transfer chains in P450 systems

Electron transfer partner proteins are required for the sequential delivery of two electrons from NAD(P)H to the P450 monooxygenase for O–O bond activation. There are two major classes (class I and class II) of P450 electron transfer chains (Fig. 1.7).

Class I enzymes utilise a ferredoxin (Fdx) and a ferredoxin reductase (FdR). FdRs are FAD- or FMN-dependent enzymes that accept the hydride from NAD(P)H and pass the two electrons one-at-a-time to the Fdx which is typically a [2Fe-2S] protein.[33] The most well characterised FdR/Fdx combinations are the adrenoxin reductase (AdR) and adrenodoxin (Adx) pair that mediate steroid biosynthesis by mitochondrial P450s,[34] and putidaredoxin reductase (PdR) and putidaredoxin (Pdx) of the $P450_{cam}$ (CYP101A1) system from *Pseudomonas putida* that catalyses camphor oxidation to 5-*exo*-hydroxycamphor.[35] Class II P450s are coupled with cytochrome P450 reductase (CPR) which consists of separate FAD and FMN domains. For both class I and class II P450s, the P450 monooxygenase and electron transfer partners need to be expressed separately which increases the complexity of the reaction system.

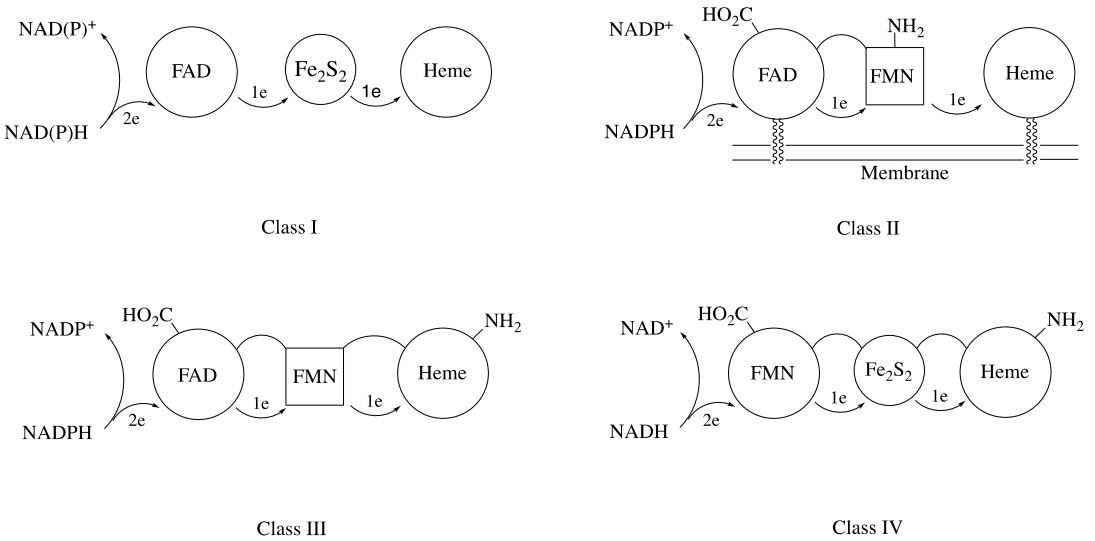



**Fig. 1.7** P450 electron transfer chains.

Fused systems, sometimes referred to as class III and exemplified by $P450_{BM3}$,[36] and the class IV enzyme $P450_{Rhf}$ (CYP116B2),[37] are also known. These enzymes are increasingly studied because the electron transfer partner is genetically fused to the

P450 monooxygenase, leading to simpler handling since only a single gene has to be expressed and the catalytic activity is generally higher than class I systems due to the intramolecular nature of the rate-limiting electron transfer step.

## 1.4 $P450_{BM3}$

In 1975, $P450_{BM3}$ was isolated from *Bacillus megaterium* as the first prokaryotic P450 enzyme naturally fused with CPR-like domain as the electron transfer partner in a single polypeptide chain.[38] The full-length enzyme has 1046 amino acid residues and a molecular weight of 120 kDa comprising of a ~55 kDa heme monooxygenase domain fused at its C-terminus to a ~65 kDa FAD-FMN reductase domain.[39] An unusual feature of $P450_{BM3}$ is that it is on obligate dimer.[40] At low ionic strengths the dimer dissociates to catalytically inactive monomers, suggesting that at least one of the electron transfer steps, i.e. FAD-to-FMN or FMN-to-heme, occurs across the dimer interface. In common buffers such as 50 mM phosphate, pH 7.4, the enzyme is in the dimer form.[41]

The intramolecular (intra-dimer) nature of the electron transfer step endows this catalytically self-sufficient enzyme with high activity (>3,000 turnovers per minute) for the oxidation of medium-chain ($C_{12}$–$C_{20}$) fatty acids, believed to be the natural substrates, at sub-terminal positions, giving approximately equal proportions of ω–1, ω–2 and ω–3 mono-hydroxy lauric acids, for example.

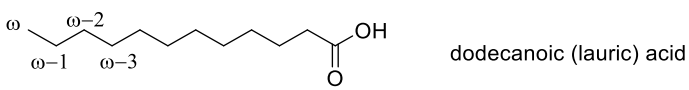



**Fig. 1.8** Dodecanoic acid, the natural substrate of $P450_{BM3}$ and oxidisation position.

### 1.4.1 Crystal structures of wild-type $P450_{BM3}$

The structure of the functional dimeric form of $P450_{BM3}$ has not been reported although there was a recent report of a cryo-electron microscopy image at low resolution.[42] The separately expressed heme domain of both the substrate-free (SF) and fatty acid-bound (SB) forms have been crystallised and their structures determined.[43] Both structures show an elongated, largely hydrophobic substrate access channel leading from the enzyme surface to the active site above the heme. The entrance of the channel is defined by two key residues, Arg47 and Tyr51, which interact with the carboxylate group of a bound fatty acid substrate such as palmitoleic acid,[44] palmitic acid,[45]and *N*-palmitoyl-glycine.[46] The aliphatic chain is in an extended conformation within the channel, reaching over and beyond the heme iron to contact the B′ helix residues V78 and A82 (Fig. 1.9).

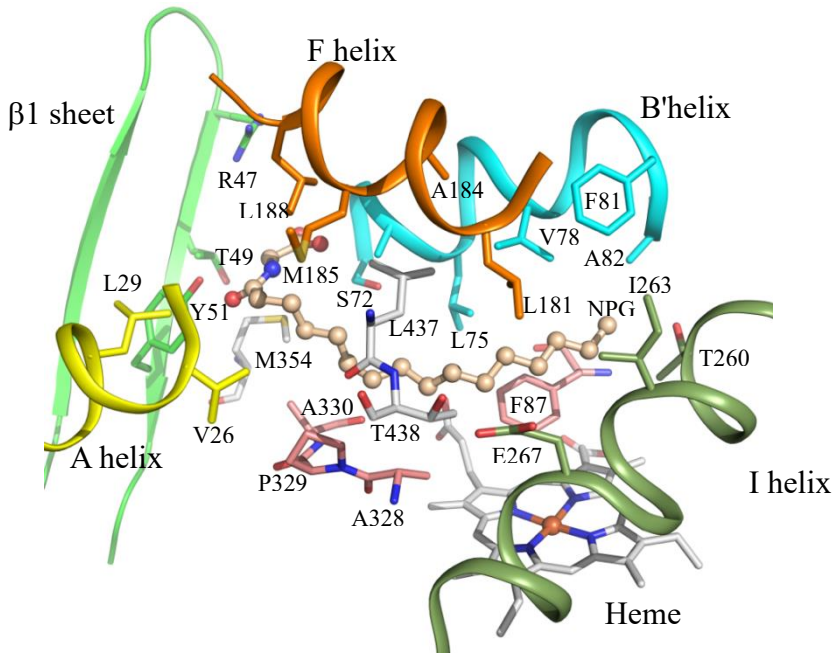



**Fig. 1.9** The active site structure of the *N*-palmitoyl glycine (NPG)-bound form of the heme domain of wild type $P450_{BM3}$ showing the residues in contact with the substrate which is in an extended conformation along the substrate channel. R47 and 51 at the entranc

Unexpectedly, none of the carbon atoms of the aliphatic chain, including the ω–1, ω–2 and ω–3 carbons that are oxidised by the enzyme, is within 7 Å of the heme iron. Therefore, the SB structure of the heme domain is not representative of the substrate binding conformation and orientation at the C–H bond activation step. Conformational changes must occur during the catalytic cycle to bring the carbons close to the heme for oxidation by the ferryl intermediate. Nevertheless, the structures revealed residues that contacted the fatty acid substrate and provided a framework for mutagenesis and protein engineering for structure-function studies and the oxidation of unnatural substrates.

There are other conformational changes between the SF and SB structures. The heme iron is hexa-coordinated in the SF structure, with a water bound on the distal side and hydrogen-bonded to the carbonyl of Ala264 of the I helix. This water ligand dissociates from the iron in the SB form while a new water molecule appears in the Ala264–Thr268 I-helix oxygen-binding groove. The I helix also moves away from the heme iron compared to the SF form.[47] The F87 side chain sits directly above the heme and undergoes a ~90° rotation. The loop containing Leu437 and Thr438 undergoes a conformation change to make space for the fatty acid aliphatic chain. Further from the heme, the F and G helices move relative to each other. Overall, the SB structure is much more compact or in a “closed” form compared to a more “open” conformation for the SF form, probably to allow residue side chains to contact the substrate. These multiple conformational changes have to be induced by enzyme-substrate interactions to move the enzyme from the SF structure with low turnover activity to the

"catalytically ready" structure of the SB form. Clearly, if the enzyme could be engineered so that the mutated variant has the SB-like, catalytically ready conformation in the absence of a bound substrate, the oxidation of unnatural substrates will be greatly promoted.[32]

### 1.4.2 High-level $P450_{BM3}$ expression and large-scale biotransformation

Once a P450 enzyme has been engineered, designed or evolved to possess the required selectivity of product formation and sufficient catalytic activity, production scale biotransformation requires large quantities of enzyme. In laboratory shake flasks *E. coli* typically reaches $OD_{600}$ of 3.5–4.0 and typically WT $P450_{BM3}$ is produced at 2 μmoles/L (~24 mg/L). At a turnover number (TON) of 10,000 the productivity would be 20 mM, or 4 g/L for a product with a molecular weight of 200, which is not economically viable. Moreover, many P450 enzymes do not show such high TON. Therefore, high enzyme production levels are required.

Pflug *et al*. were the first to report the fed-batch expression of $P450_{BM3}$ in high cell density *E. coli* cultures.[48] Glycerol was found to be the optimal carbon source and the cultures had to be induced with IPTG at OD ~20 for maximal expression. The conditions gave $P450_{BM3}$ expression of 14 μmoles/L (~1.7 g/L) at $OD_{600}$ ~100. Large scale biotransformations were not reported with the cultures or the expressed enzyme.

Kaluzna *et al*. reported fed-batch high cell-density *E. coli* fermentations to express WT $P450_{BM3}$ at up to 30 μM/L at 10,000 L scale using a glucose feed.[49] The acetate production issue found by Pflug *et al*. was not reported. The cell-free extracts were then used to carry out *in vitro* oxidation of isophorone to 4-hydroxy-isophorone

using glucose and glucose dehydrogenase to recycle the NADPH cofactor. The reaction mixtures were aerated with air mixed with pure oxygen to improve oxygen mass transport. Substrate concentration up to 10 g/L (TON 13,200) and kg-scale batch reactions were successfully conducted, with 55% product recovery. The system was further developed for the biotransformation of α-ionone, which is less soluble than isophorone in aqueous buffers, using non-ionic detergents such as Tween to solubilise the substrate, leading to a product concentration of 15 mM and TON of 4,000.[50] These reports of high-level expression, successful *in vitro* cofactor recycling and high product concentrations are important steps in developing the biotechnological applications of $P450_{BM3}$.

### 1.4.3 Rational design for unnatural substrate oxidation

The SF and SB structures spurred extensive efforts to engineer $P450_{BM3}$ for unnatural substrate oxidation. This area has been reviewed.[32,51] The most obvious target residue for mutagenesis is F87 as it is close to the heme iron. The F87A and F87V mutations were introduced to create space above the heme. The mutants were found to alter the regioselectivity of fatty acid oxidation and promote unnatural substrate oxidation.[52-54]

The R47/Y51 couplet of channel entrance residues were also early targets for mutagenesis. Reversing the charge of R47 led to a reduction in fatty acid oxidation activity and the R47E mutant accepted trialkylammonium compounds with long aliphatic chains as substrates for which the WT had no activity.[55] The double hydrophobic substitution mutant R47L/Y51F was proposed, and found to promote the entry, binding and oxidation of hydrophobic compounds, including polyaromatic

hydrocarbons (PAH) such as pyrene.[56] PAH oxidation activity was further increased when this couplet was combined with the F87A and A264G mutations that created space close to the heme. This rational design approach was extended to sesquiterpene oxidation and the first report of valencene oxidation to nootkatol and nootkatone by the R47L/Y51F/F87A mutant.[57] The R47L/Y51F, or closely related combinations, are some of the most used base mutations in $P450_{BM3}$ engineering.

Another interesting approach was the idea of "filling a hole" using bulky substitutions at A82.[58] In the fatty acid-bound structures of the WT the end of the aliphatic chain contacts A82. It was hypothesised that the side chain of Ala generated a hole to sequester the chain, and that a bulky substitution would fill in this hole, leading to a smaller substrate pocket and potentially enhancing unnatural substrate oxidation. This was found to be the case, with increased styrene oxidation activity for the A82F and A82W mutants.

### 1.4.4 Site-saturation and random mutagenesis: Directed evolution

Although rational, structure-based mutation design was successful in promoting unnatural substrate oxidation the approach is laborious and many mutants had to be prepared and studied. Residue and mutation selection and the linear nature of the engineering make this approach less efficient. Site-saturation and random mutagenesis cover more sequence space; residues and mutations that exert their effect without having to contact the substrate could be discovered by these approaches.

To engineer $P450_{BM3}$ from a long-chain fatty acid hydroxylase to a medium- to short-chain fatty acid hydroxylase, a method called rational evolution was applied.[50]

Eight ‘hotspot’ residues were selected from computational simulation and each was subjected to site saturation mutagenesis. Variants showing activity in initial screening were further engineered. Some variants were found to produce the blue pigment indigo from indole oxidation, indicating oxidation of an unnatural substrate. Further site-saturation and random mutagenesis followed by selection for indigo formation led to the identification of the triple mutant A74G/F87V/L188Q (GVQ) that was highly active for indigo formation. This mutant was also found to oxidise PAHs, thus validating the use of indole oxidation as an indicator of unnatural substrate oxidation. The GVQ mutant shows much higher activity than the F87V single-site mutant. The mutations at A74, which is within the substrate channel but far from the heme, and the non-active site residue L188 at the end of the F helix, promoted unnatural substrate oxidation by long-range effects.

Random mutagenesis methods such as error-prone PCR (ep-PCR) and DNA shuffling could generate large mutant libraries with mutations at random residues and random mutations at each residue. More of sequence space can be covered, the reactivity could be enhanced and the selectivity altered. However, the introduction of all 20 amino acid substitutions in a protein of 500 amino acids would generate $20^{500}$ mutants; it is impossible to test each mutant individually, let alone the analysis of reaction outcome. Simple, highly efficient screening procedures are necessary to probe the effect of mutations. Despite such problems, early ep-PCR and DNA shuffling studies gave promising results on improving enzyme stability, activity and selectivity.

Arnold and co-workers pioneered the application of ep-PCR and DNA shuffling

to $P450_{BM3}$ engineering. Using a screening assay similar to that for esterase directed evolution, a library of $P450_{BM3}$ mutants generated by random mutagenesis was screened against *p*-nitrophenyl octyl ether. C–H bond oxidation of this unnatural substrate followed by elimination of octanal released *p*-nitrophenolate which can be monitored by its absorption at 405 nm.[59]

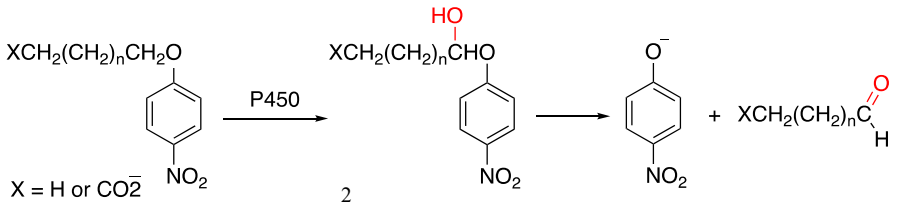



**Fig. 1.10** *p*-Nitrophenolate screening procedure for $P450_{BM3}$ directed evolution.

After multiple rounds of ep-PCR and DNA shuffling of positive mutants from the screens, new mutants were identified that showed high activity for the oxidation of short- to medium-chain ($C_3$–$C_8$) alkanes. Notably, many of the mutations were remote from the active site and yet high activities were observed, further indicating the potential value of random mutagenesis in revealing residues that would not be targeted in rational engineering approaches. One of the variants, 139-3, contained 11 mutations V78A/H138Y/T175I/V178I/A184V/H236Q/E252G/R255S/A290V/A295T/L353V and oxidised propane and butane to a single product. Only V78 is an active site residue. The mechanism by which remote mutations assert their effect was not clear, while mutations with the most dramatic effects on selectivity and substrate range remained those at active site residues.

Further random and site-directed mutagenesis developed mutants with high activity and selectivity for octane oxidation; the 9-10A-A328V mutant oxidised octane primarily at the 2-position to form (*S*)-2-octanol with 40% ee.[60] The A328 residue

contacts the heme and is on the opposite side of the substrate pocket to F87 (Fig. 1.9). Mutants capable of propane and ethane oxidation to 2-propanol and ethanol, respectively, were then engineered.[61,62] The most active "propane monooxygenase" has 25 mutations, some of which (G443D, V445M, T480M, T515M, P654Q, T664M, D698G, E1037G) are in the reductase domain. The early "hits" also served as starting points for directed evolution and screening to oxidise numerous drug molecules, in many cases to metabolites produced by human P450 enzymes.[63-66] The range of drug molecules oxidized was impressive, ranging from smaller molecules such as ibuprofen to larger ones including propranol[64] and Imatinib.[67]

Vermeulen and co-workers observed oxidation of drugs such as testosterone and amodiaquine by the R47L/F87V/L188Q mutant (Fig. 1.11).[68] This was then subjected to ep-PCR and mutants were screened for alkoxyresorufin *O*-dealkylation by detection of the fluorescence from the resorufin product. A series of mutants were identified that had >200-fold higher activity than the starting mutant for the oxidation of dextromethorphan and 3,4-methylenedioxymethylamphetamine (MDMA). The mutations with the greatest effect were found to be F81I, L86I and E267V.[69] These mutants were used for preparative scale synthesis of metabolites of clozapine and diclofenac for toxicology studies,[54] and screened for oxidation of 43 marketed drugs, showing diversity of product formation including human metabolites.[70]

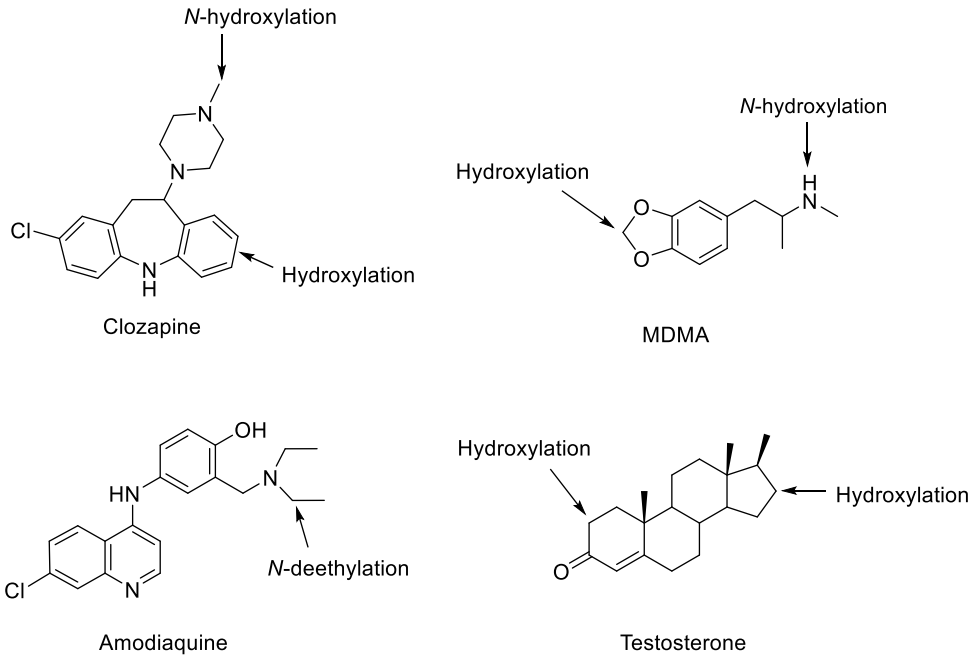



**Fig. 1.11** Examples of drug molecule oxidised by $P450_{BM3}$ variants library generated by ep-PCR base on variant bearing R47L/F87VL188Q mutations.

The F87 residue was again found to be critical, as were A82 and E267. One striking observation was the effect of the A82W mutation introduced in variant M01 (R47L/F87V/L188Q/E267V/G415S) that inverted the stereoselectivity of testosterone significantly towards 16β-hydroxylation,[71] while addition of the S72I mutation to the A82W mutant shifted the selectivity towards the 16α-hydroxy-testosterone (Fig. 1.12).[72] One of the most active mutants, M11, contained the mutations R47L/E64G/F81I/F87V/E143G/L188Q/E267V/G415S. Of these, R47L, F87V, L188Q and E267V were previously reported while the effect of E64G and F81I were newly found. Only F81, F87 and E267 are within the active site; mutations at these residues, as well as A82, have the greatest effect on product selectivity.

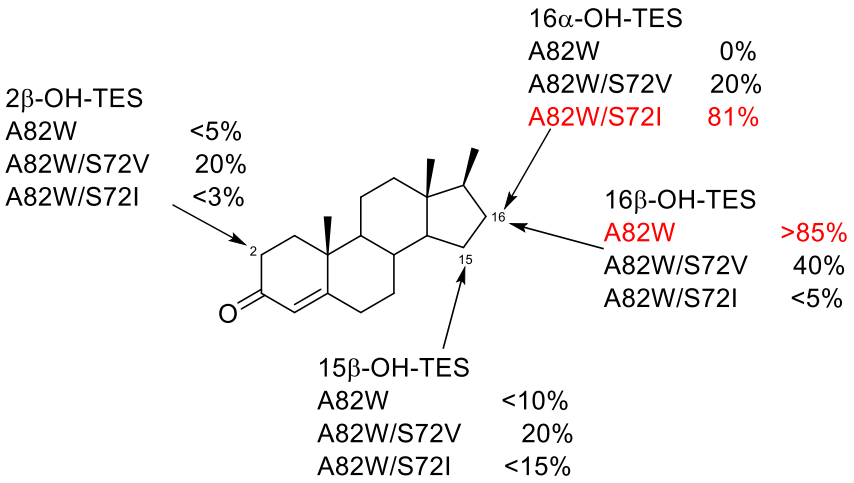



**Fig. 1.12** Oxidation Selectivity Pattern of single mutation and double mutations introduced in M01 variants as testosterone as substrate.

Most random mutagenesis/directed evolution of $P450_{BM3}$ utilised surrogate unnatural substrates to detect mutants with altered substrate specificity and increased activity, e.g. *p*-nitrophenolate formation and dealkylation of alkoxyresorufin. Wong and co-workers used indigo formation from indole oxidation within *E. coli* cells transformed with $P450_{BM3}$ mutants to screen random mutagenesis libraries of the WT and the F87A mutant.[73] After rounds of colony picking and re-plating, colonies showing the most intense blue coloration were selected and the mutants subjected to DNA shuffling. The new library was again screened for indigo formation. Of ~800 colonies, ~130 were picked and the mutants tested for whole-cell oxidation of two unnatural substrates, naphthalene and propylbenzene. GC analysis identified 20 mutants with enhanced activity and/or altered product profile. The four most active variants were F87A/H171L/Q307H/N319Y (KSK19), F87A/A330P/E377A/D425N (KT5), A330P and A191T/N239H/I259V/A276T/L353I (KT2). All these mutants showed enhanced activity by at least an order of magnitude over the WT for naphthalene and propylbenzene oxidation; the highest activity for propylbenzene was

2205 $min^{-1}$ compared to 3.1 $min^{-1}$ for the WT. One mutant produced 99% 1-phenyl-1-propanol while the others showed non-benzylic oxidation.

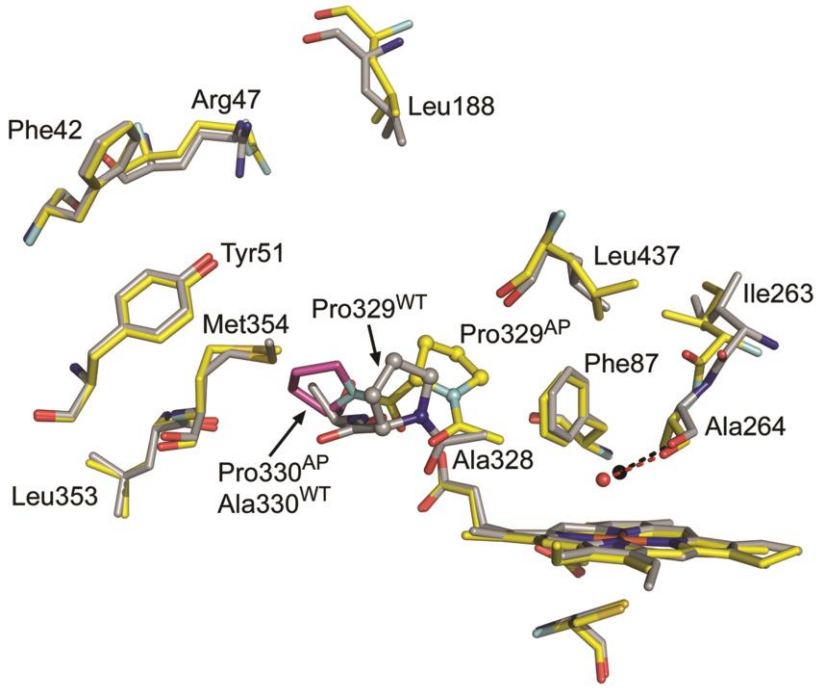



**Fig. 1.13** Overlay of the crystal structures of the substrate-free form of the WT (grey) and the A330P mutant (yellow) of $P450_{BM3}$ showing the relocation of the Pro329 residue in A330P into the substrate pocket such that the gap between it and F87 is only 6.0 Å compared to 9.1 Å in the WT.

Of the mutated residues in these variants only F87 was close to the heme. Ala330 is in a loop between two β-strands and recessed away from the substrate channel. His171 is in the E/F loop, Ala191 is in the F/G loop, Asn239 in the H/I loop, while Ile259 and Asn276 are in the I helix, and Leu353 is part of the substrate channel. The other residues (Gln307, Asn329, Glu377 and Asp425) are on the surface of the heme domain. The A330P variant was highly unusual in having a single mutation and yet it continued to be selected throughout the screening steps. Crystal structure determination revealed that this mutation resulted in relocation of the adjacent residue Pro329 into the substrate pocket and the closest approach to the F87 side chain was 6 Å (Fig. 1.13). Hence, the shape of the pocket above the heme was altered dramatically,

such that fatty acid oxidation activity was actually abolished but the activity towards unnatural substrates such as naphthalene and propylbenzene was increased.[46]

The crystal structure of the heme domain of the SF form of the KT2 mutant showed that a number of interhelix hydrogen bonds and salt bridges were broken, leading to movement of helices such that the I helix retreated from the heme to a virtually identical position as the fatty acid-bound form (Fig. 1.14). The hydrogen bond between the heme iron-bound water and the Ala264 carbonyl was broken.[74] The SF heme redox potential was unchanged compared to the WT but the rate of the first electron transfer step was comparable to that for the WT with palmitic acid bound. Therefore, the "gate" of the catalytic cycle for the WT, i.e. slow electron transfer without a bound substrate to induce a change in heme coordination, was no longer operational. The KT2 mutant was described as being in a "catalytically ready" conformation, hence the increased activity for unnatural substrates.

The unexpected effect of the A330P mutation on enzyme conformation led to the introduction of the I401P mutation, just to the C-terminus side of the heme proximal residue Cys400. The I401P mutant was found to have greatly enhanced activity for unnatural substrate oxidation.[75] The SF crystal structure showed that the heme group had moved away from the substrate pocket compared to the SF WT structure, and similar to that in SB WT (Fig. 1.15). Strikingly, the iron–water distance increased to 3.6 Å, and it is arguable whether this water is actually a ligand and in any case it would be very readily displaced by substrate binding compared to the WT. The heme redox potential and the rate constant for the first electron transfer step were almost the same as the palmitic acid-bound WT.[46]

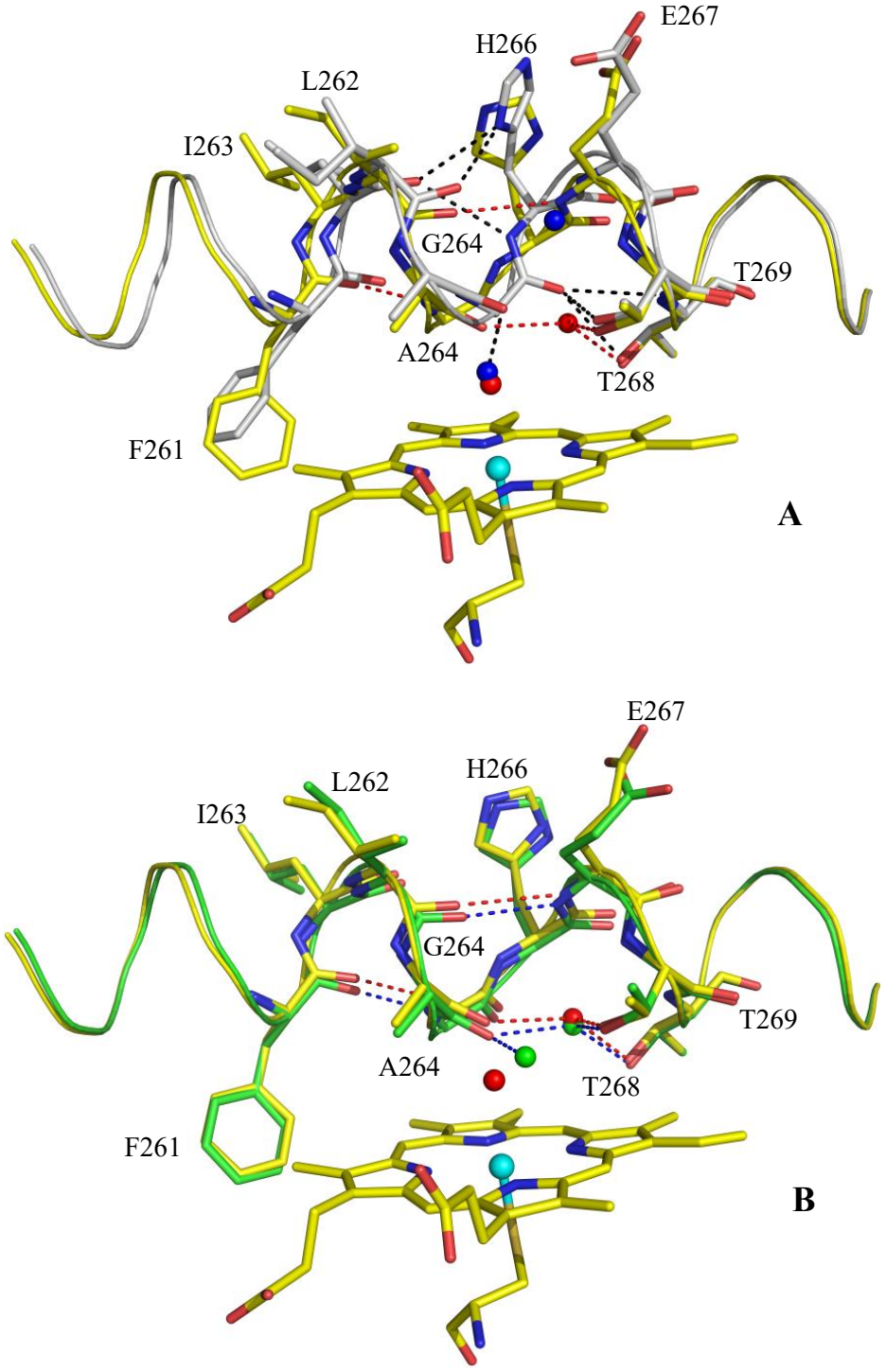



**Fig. 1.14** (**A**) Overlay of the structure of the substrate-free (SF) form of the WT (grey)[76] and KT2 mutant (yellow) of $P450_{BM3}$ showing the H-bond between the water ligand (blue) and A264 in the WT and its absence in KT2 and the appearance of an extra water in the A264–T268 oxygen-binding groove (waters in red). (**B**) Overlay of the substrate-bound WT (green)[47] with SF KT2 showing the virtually identical conformations of the I helices, the identical position of the extra water (green for SB WT) in the oxygen binding groove, and the location (WAT123) which the heme-bound water is proposed to move to when the substrate binds.

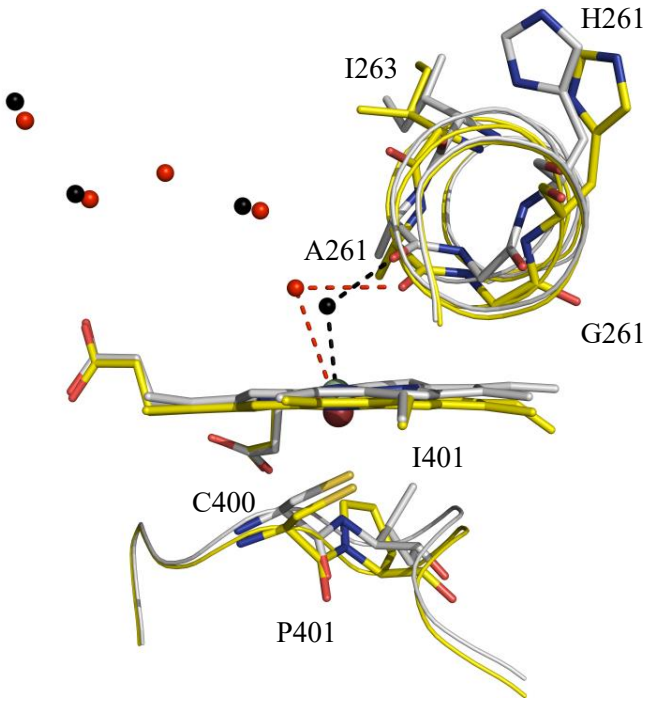



**Fig. 1.15** Overlay of the crystal structures of the substrate-free forms of the WT (grey) and I401P (yellow). Water molecules in the WT are in black and those in the mutant in red. In the mutant the I401P mutation caused the heme to drop away from the substrate pocket and the Fe–$OH_2$ distance increased to 3.6 Å.

Therefore, the I401P mutation induced conformational changes that also unlocked the gate to the catalytic cycle, and unnatural substrates are readily oxidised. It was noted, however, that the I helix remained in the SF conformation, unlike for the KT2 mutant. Overall, these four mutants revealed different mechanisms by which mutations increased the activity of $P450_{BM3}$ for the oxidation of unnatural substrates.

### 1.4.5 Combinatorial and iterative saturation mutagenesis

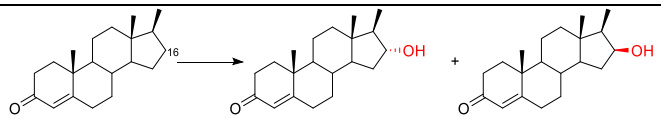
Although random mutagenesis by ep-PCR reveals residues and mutations that might not be considered by rational design, it also systematically generates superfluous mutations which are not necessary in enhancing activity/selectivity and have to be eliminated by effective screening procedures. Narrowing down the target list could

reduce the size of the mutant library to be screened. Since residues around the active site dramatically influence selectivity, Reetz developed two methods, Combinatorial Active-site Saturation Test (CAST) and Iterative Saturation Mutagenesis (ISM), to generate a smaller, active site-focused variant library guided by structural information on the enzyme.[77,78] Residues around the active site are divided into several spatially close groups. Site-saturation mutagenesis is applied to selected residues to identify these that affect activity/selectivity. Residues can be targeted in turn, i.e. the process is iterative (ISM), or combinatorially in pairs and threes (CAST). These two approaches can significantly reduce the time required to obtain hits, especially when a high throughput (e.g. colorimetric) screen is not available and medium throughput methods such as chromatographic analysis is required.

Using this strategy, the F87A mutant was selected as the template and a library targeting 23 residues was constructed to screen for testosterone oxidation.[79] These residues, most of which contact a fatty acid substrate, were divided into 9 sites A–I, each with 2–3 residues (A: R47, T49, Y51; B: V78, A82; C: M185, L188; D: S72, A74, L75; E: L181, A184; F: T260, I263; G: A328, A330; H: L437, T438; I: V26, L29). Residues S72, A328 and A330 were targeted for saturation mutagenesis while both V78 and A82 of site B were randomised, and residues in sites A and C were subjected to NDC mutagenesis. The NDC codon encompasses a reduced alphabet of 12, mostly hydrophobic amino acid substitutions. Up to a total of 10,000 transformants were screened by HPLC. With F87A, the 2β- and 15β-alcohols were formed in a 1:1 ratio. The final screening results showed that the F87A/A330W gave 95% 2β-hydroxy-

testosterone and the R47Y/T49F/V78L/A82M/F87A gave 96% of the 15β-alcohol. In further work using similar approaches but broader coverage of residues and more site-saturation mutagenesis, and taking into account steroid–enzyme binding interactions from docking and MD simulations, Reetz and co-workers developed $P450_{BM3}$ mutants capable of highly selective 16α- and 16β-hydroxylation of testosterone and related steroids. Interestingly, crystal structure determination showed that two testosterone molecules were bound within the active site, one close to and over the heme while the other was located near the side chains of residues 47 and 51.[80]

**Table 1.1** The conversion and selectivity of 16α- and β-hydroxylation of testosterone by $P450_{BM3}$ variants WIFI-WC and WWV-QRS




| Variant | **Mutation** | **Conversion** | **16α- hydroxylation** | **16β-hydroxylation** |
|---|---|---|---|---|
| **WIFI-WC** | R47W/Y51W/S72I/A82F/F87I/ L181C | 95% | 96% | 4% |
| **WWV-QRS** | R47W/A82W/F87V/ L181Q/T436R/M177S | 92% | 0% | 92% |

## 1.4.6 The mutant library approach

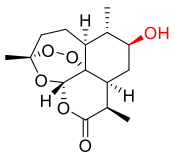
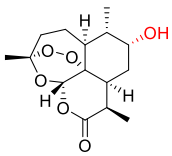
P450 enzymes can share as little as 15% sequence identity but the basic structural fold is conserved within the superfamily. Gotoh first noted that P450 enzyme–substrate interactions can be classed into 6 substrate recognition sites (SRS).[81] Computational docking and MD simulation were used to study interaction between $P450_{BM3}$ SRS and substrates in order to identify residues which may impact selectivity. Based on these analyses, a smaller but more focused library could be constructed for selectivity studies.

Seifert *et al*. constructed a $P450_{BM3}$ library containing mutations at F87 in SRS1 and A328 in SRS5, with 5 hydrophobic substitutions (alanine, valine, phenylalanine, leucine and isoleucine) at each.[82] In all, 24 variants were expressed into functional enzymes which were tested for oxidation of limonene, valencene, geranyl acetone, neryl acetone, octane and cyclic alkanes. Significant selectivity changes included the 97% selectivity for 9,10-epoxy-geranylacetone for the F87L mutant compared to 54% for the WT and allylic oxidation at C11 (37%) and C12 (80%) were observed while they were not preferred by the WT. Valencene was oxidised to more than 10 products at low conversion but two variants gave specific oxidation at C2 to produce nootkatol and nootkatone. High conversions were observed for mutants within the library for cyclooctane (>70%) and cyclodecane (>50%) compared to the WT. The variant F87V/A328F produced 92% 2-octanol from octane.[53] This approach appeared to be an efficient way to obtain good to excellent substrate conversion and selectivity changes.

Fasan and co-workers developed a larger library of ~500 mutants derived from the base mutants from Arnold's work and containing a variety of mutants selected from a number of random mutagenesis studies. The residues and mutations were not specifically selected or designed, and were at both active site and non-active-site residues. This library was called a "training set" that would be used to screen for residues that affected activity and selectivity.[83] Once hits were obtained, site-saturation mutagenesis could give higher selectivity, etc. This library was tested for artemisinin oxidation, and 20 out of the 500 showed activity. Subsequent mutagenesis at key residues then afforded mutants with 100% selectivity for the 7(*R*)- and 7(*S*)-alcohols of

artemisinin. Variant IV-H4 and II-H10 were identified to give 100% selectivity for the 7(*R*)- and 7(*S*)-alcohols, of which the difference are compared in Table 1.2. Although the library is slightly large, at 500 mutants, this approach is promising for identifying initial hits.

**Table 1.2** $P450_{BM3}$ variants which gave absolute selectivity for 7(*R*)- and 7(*S*)-hydroxy-artemisinin. Different combinations of mutations at the active site residues V78, F81, A82, F87 and A184 completely changed the stereoselectivity.

|  |  | **Variant** | **Mutations** |
|---|---|---|---|
| 83% | 10% | FL#62 | V78A/F81S/A82V/F87A/P142S/T175I/A180T/A184V/ A197V/F205C/S226R/H236Q/E252G/R255S/A290V/ L353V |
| **100%** | 0 | IV-H4 | V78S/F81V/P142S/T175I/A180T/A197V/F205C/ S226R/H236Q/E252G/R255S/A290V/L353V |
| 0 | **100%** | II-H10 | V78N/A82T/P142S/T175I/A180T/L181F/A184T/ A197V/F205C/S226R/H236Q/E252G/R255S/A290V/ L353V |

Wong and co-workers have reported a similar library approach using ~100 mutants based on four template enzymes A330P, I401P, KT2 and KSK19,[73] with additional, mainly hydrophobic substitutions at 2–4 active site residues. The aim was to generate varied active site topologies to identify residues and mutations that improved unnatural substrate oxidation activity and altered product selectivity. Further mutagenesis and directed evolution would then increase production of specific target compounds. The mutant library was shown to oxidise a range of drugs to human metabolites,[84] fragment compounds,[85] and led to the discovery of new P450 reactions (Fig. 1.16).[86] One significant property of the mutant library was the generation of

different products by different members so that the product scope accessible to $P450_{BM3}$ variants was revealed.

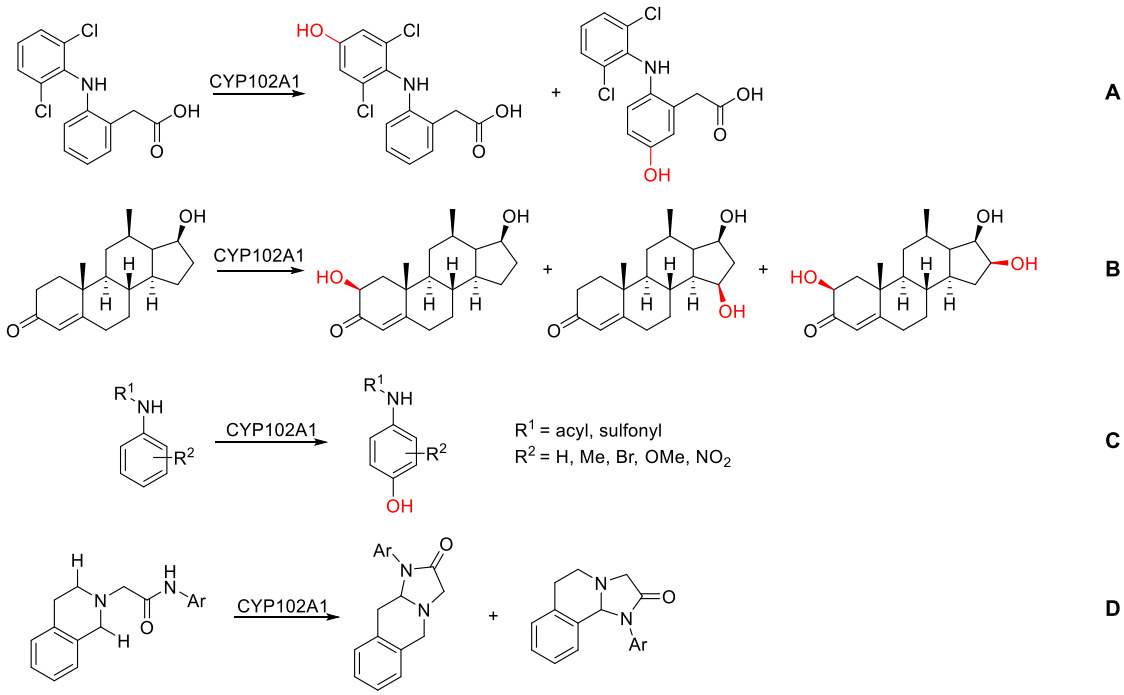



**Fig. 1.16** Examples of drug metabolism (A and B), fragment molecule oxidation (C), and new P450 reaction of C–H amination (D) from a library of $P450_{BM3}$ enzyme variants.

## 1.5 Terpenoid compounds

During metabolic processes in plants, animals and microorganisms, numerous organic compounds are produced as primary or secondary metabolites. Many of these possess biological activity and are valuable products in the flavour, fragrance and pharmaceutical industries. Compounds as diverse as alkaloids, alcohols, aldehydes, terpenoids, macrolides and sterols are drugs or have desirable biological activity.

Terpenes and their derivatives constitute one the largest classes of secondary metabolites in nature. All terpenes are biosynthesised from the 5-carbon building block isoprene and classified as hemiterpenes (C5), monoterpenes (C10), sesquiterpenes

(C15), diterpenes (C20), triterpenes (C30), and some mix-origin and polymeric terpenes.

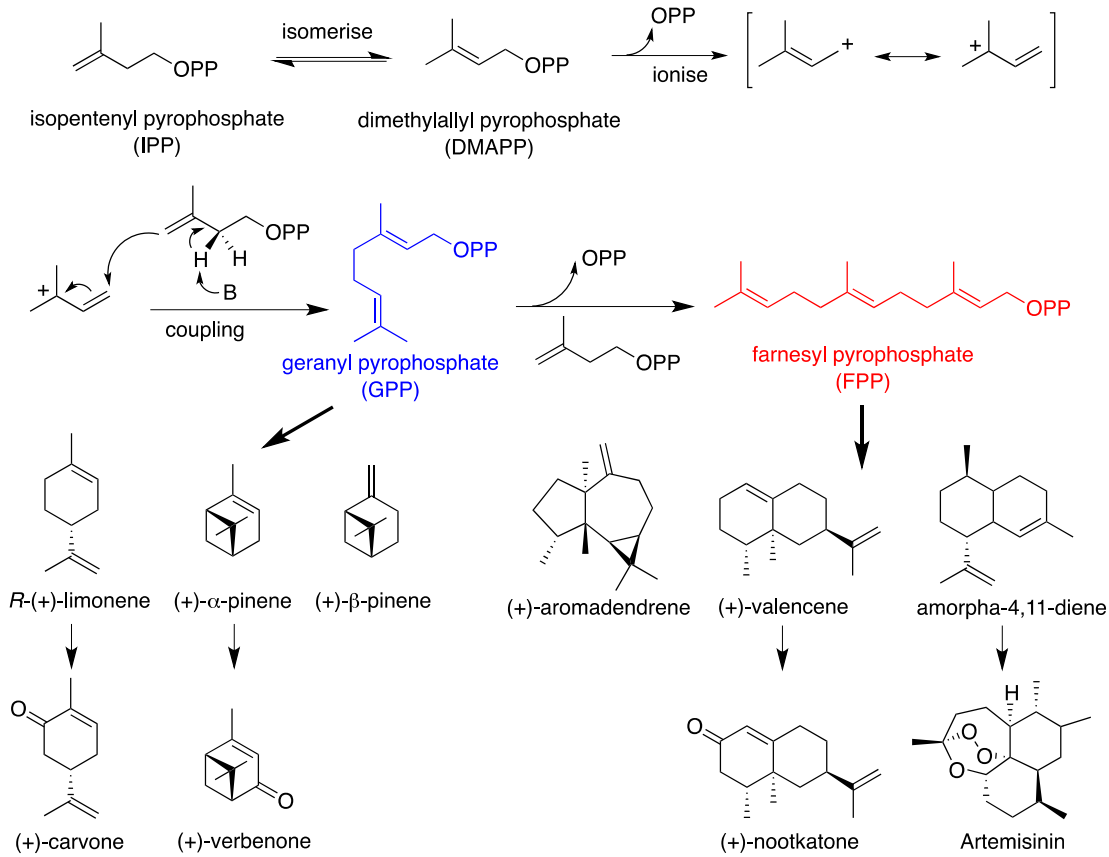



**Fig. 1.17** Biosynthesis of terpenes and terpenoids.

All terpenes in eukaryotes are biosynthesised via the mevalonate pathway which assembles mevalonic acid from acetyl-CoA, then forming isopentenyl pyrophosphate (IPP) which is the precursor to all terpenes (Fig. 1.17).[87] An isomerase converts IPP to dimethylallylpyrophosphate (DMAPP). Loss of pyrophosphate from DMAPP followed by trapping of the carbocation by IPP in a 'head-to-tail' coupling reaction generates geranyl pyrophosphate (GPP), the precursor to all monoterpenes (C10). Coupling of another IPP unit to GPP forms farnesyl pyrophosphate (FPP) which is the precursor to all sesquiterpenes. The sequence proceeds further to form geranylgeranyl

pyrophosphates (GGPP) from which diterpenes are derived. Coupling of two FPP units leads to squalene that forms steroid precursors.

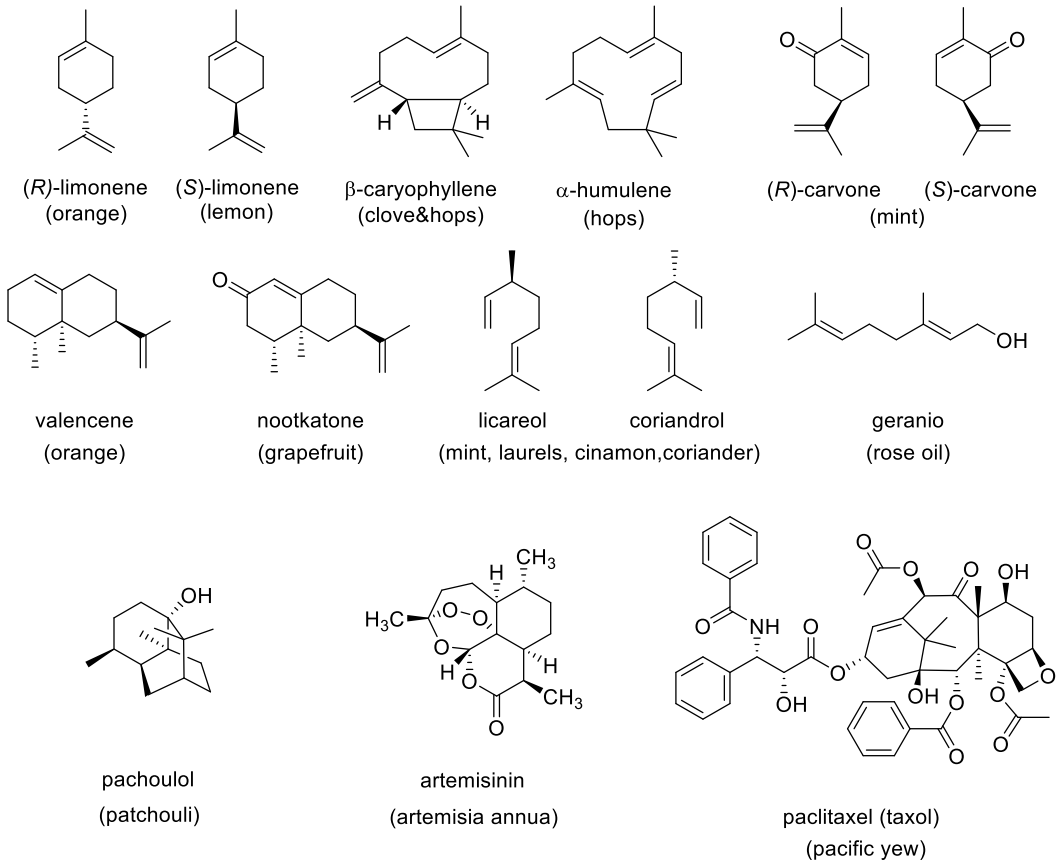



**Fig. 1.18** Examples of terpenoid compounds with flavour, fragrance and medicinal properties.

Terpenoids are the active compounds for the characteristic odour and flavour of many plants, flowers and fruits, such as (*R*)-limonene in oranges, (*S*)-limonene in lemon, α- and β-pinene from pines, β-caryophyllene and α-humulene in hops, valencene from oranges and its 2-keto derivative nootkatone in grapefruit. They are often the most important component in essential oils, e.g. carvone in spearmint and peppermint, linalool in lavender and coriander, geraniol in rose oil, patchouli alcohol in patchouli oil, and santalol in sandalwood. Biological activities are also common. α-Humulene has anti-inflammatory properties,[88] while its 3-keto derivative zerumbone

induces apoptosis and programmed cell death and is being investigated as an anti-cancer drug.[89-93] Nootkatone is not only the high-value flavour compound of grapefruit but also highly effective as an insecticide and insect repellent.[94] The anti-cancer drug taxol is a polyoxygenated diterpene derivative.

The low molecular weight of sesquiterpenes ($C_{15}H_{24,}$ RMM = 204) and yet the wide range of biological effects including anti-inflammatory, antibiotics and anti-cancer properties, make sesquiterpenoids potentially a rich source of molecules for drug discovery. A most important sesquiterpene drug is artemisinin, a sesquiterpene lactone active against *Plasmodium falciparum* malaria. It was isolated from *Artemisia annua* and employed for thousands of years in Chinese traditional medicine. Combination therapies of artemisinin with other drugs are recommended since it is well tolerated.

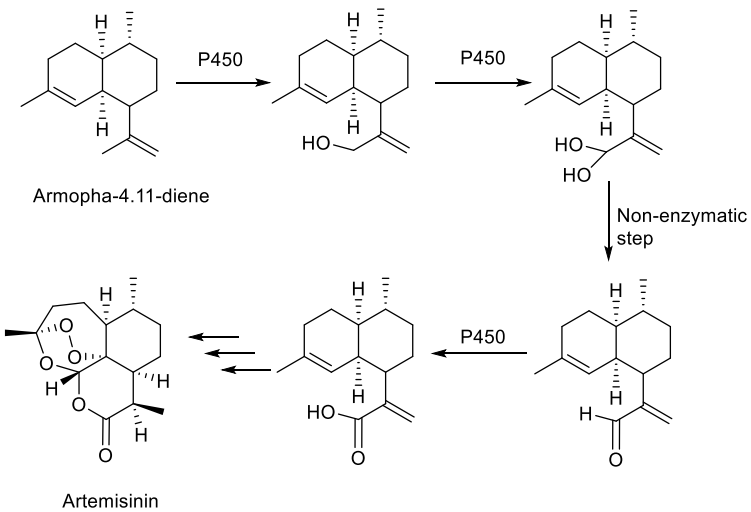



**Fig. 1.19** Synthesis of artemisinin, with the critical initial oxidation of amorpha-4,11-diene catalysed by a P450 enzyme.

In 2006, the total synthesis of artemisinin by assembling the entire terpene biosynthesis pathway in yeast to synthesise amorpha-4,11-diene from glucose, and

then oxidation by the P450 enzyme CYP71AV1, eventually to the precursor artemisinic acid, was reported (Fig. 1.19). [95] This seminal research is an excellent example of the application of synthetic biology in synthesis.

## 1.6 Thesis objectives

Terpenoids have huge potential in the fine chemical industry and in drug discovery, and P450 enzymes provide a powerful catalytic tool to functionalise terpenoids. Extensive research has focused on functionalising terpenoid substrates. Fine chemical synthesis via oxidation of valencene to nootkatone is a promising project for industrial applications, however, no process for nootkatone synthesis via valencene oxidation has been commercialised since the early reports of chemical routes or enzymatic nootkatone formation.[94] Comparing all experimental methods of sesquiterpene biotransformation reported to date, low solubility of substrate, poor stability of reaction systems and poor selectivity have impeded their application. Therefore, a sustainable, efficient and scaleable process using a suitable P450 enzyme is an attractive prospect.

There are two main aims of the research in this thesis. The first is to explore and understand the role of selected key residues on the activity and selectivity so as to design an appropriate mutagenesis strategy for $P450_{BM3}$ engineering for the oxidation of sesquiterpenes. Chapter 3 describes initial variant selection, screening for the oxidation of the sesquiterpenes caryophyllene, cedrene, humulene, and valencene, and characterisation of some of the oxidation products.

In Chapter 4, the more complex tricyclic compound cedrene was selected as the

target substrate to be screened against the expanded library designed following the strategy established. The results on manipulation and shifting the oxidation regioselectivity of cedrene from allylic alcohol and epoxide to a range of different oxidative products such as tertiary alcohol, aldehyde and carboxylic acid, alcohol at aliphatic position, etc. are reported.

The second aim is to use the screening data for valencene oxidation to select a mutant for scaling up nootkatone production from screening scale (1 mL) through benchtop reactions (20–200 mL) to reactions at up to 3 litre volume in a bioreactor. Chapter 5 describes experiments to optimise parameters such as the valencene-to-enzyme ratio, valencene addition methods and the role of surfactants, second organic phase and phase transfer agents to develop a biotransformation method to achieve high product yield.

# Chapter 2 Materials and Methods

## 2.1 Chemicals, solutions and cell strains

### 2.1.1 Chemicals, solvents and reagents

General reagents and chemicals were from Acros Organics, Alfa-Aesar, Fisher Scientific or Sigma-Aldrich and were used without further purification. The sesquiterpene substrates β-caryophyllene, α-cedrene, α-humulene and valencene were from Sigma-Aldrich, Fluka or MP Biomedicals.

HPLC-grade solvents were from Fisher Scientific or Sigma-Aldrich. Media components were from Melford Laboratories, Sigma-Aldrich or VWR International. Lysozyme from chicken egg white (≥40,000 U/mg) was from Sigma-Aldrich. Buffer components were from Fisher Scientific, Melford Laboratories or VWR International. Kanamycin (kan) was from Melford Laboratories. Isopropyl-β-D-thiogalactopyranoside (IPTG) was from Fisher Scientific. Restriction enzymes and the associated buffers were from NEB. Site-directed mutagenesis was carried out using the hot-start KOD polymerase from Merck. Oligonucleotides were supplied by Eurofins Genetic Services, UK. $NADP^+$ disodium salt was from Prozomix. GDH was from Codexis (California). Competent *Escherichia coli* cell strains DH5α for plasmid DNA propagation and BL21 (DE3) for recombinant enzyme production, were from NEB.

### 2.1.2 Buffers, solution and growth media

Type 1 water (ρ ≥12 $M\Omega.cm^{-1}$) provided by a Millipore Elix 3 system was used for the preparation of growth media and buffers for enzyme purification and biotransformation reactions. Type 2, "ultrapure" water (ρ ≥18 $M\Omega.cm^{-1}$) used for PCR, general DNA manipulations, SOB and SOC media, and competent cells preparation was from a

Millipore Milli-Q system. Phosphate buffers were prepared by mixing monobasic and dibasic solutions diluted from 1 M stocks, e.g. 50 mM monobasic solution was slowly mixed into 50 mM dibasic solution until the pH reached the required value to make the 50 mM phosphate buffer.

**Table 2.1** Media and buffer components

| Media and buffers | Components (per litre unless stated otherwise) |
|---|---|
| **$LB_{agar}$** | 10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g bacto-agar |
| **LB** | 10 g tryptone, 5 g yeast extract, 10 g NaCl |
| **$FB_{Glycerol}$** | 14.7 g $K_2HPO_4$, 3.6 g $NaH_2PO_4$, 2.5 g $(NH_4)_2SO_4$, 2 g $Na_2SO_4$, 1.2 g $Na_3$citrate.$2H_2O$, 0.5 g $NH_4Cl$, 5 mL glycerol<br>Post autoclaving – 2 mL 1 M $MgSO_4.7H_2O$, 3 mL trace elements solution, 1 mL 10% *w/v* thiamine |
| **Trace elements** | 20.1 g $Na_2EDTA$, 16.7 g $FeCl_3 \cdot 6H_2O$, 0.74 g $CaCl_2 \cdot H_2O$, 0.25 g $CoCl_2 \cdot 6H_2O$, 0.18 g $ZnSO_4 \cdot 7H_2O$, 0.132 g $MnSO_4 \cdot 4H_2O$, 0.1 g $CuSO_4 \cdot 5H_2O$ |
| **FB supplement** | 15 mg kanamycin, 250 mg $MgSO_4 \cdot 7H_2O$, 50 mg thiamine·HCl, add trace element solution to 1.5 mL |
| **Induction mixture** | IPTG 180 mg, δ-ALA·HCl 672 mg, $FeSO_4 \cdot 7H_2O$ 112 mg, Kanamycin 600 mg, add Milli-Q water to 40 mL |
| **SOC** | 20 g tryptone, 5 g yeast extract, 0.5 g NaCl; after autoclaving, add 10 mL 1 M $MgCl_2$, 10 mL 1 M $MgSO_4$, 2 mL 20% *w/v* glucose at ambient temperature. |
| **1 M $KH_2PO_4$** | 136.1 g |
| **1 M $K_2HPO_4$** | 174.2 g |
| **1 × TAE** | 4.84 g Tris, 1.14 mL glacial acetic acid, 2 mL 0.5 M $Na_2EDTA$ |
| **TFB 1** | 30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% glycerol (pH adjusted to 5.8 with 1 M acetic acid) |
| **TFB 2** | 100 mM MOPS or PIPES (pH 6.5), 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol (pH adjusted to 6.5 with 1 M KOH) |

Liquid media were sterilised by autoclaving at 121°C for 20 min. Supplement solution for $FB_{glycerol}$ and induction solution for inducing recombinant enzyme production were made using type 2 water and filter-sterilised using a 0.22 μM filter unit (Merck) before use. $LB_{kan}$ media was made with a kanamycin working concentration of 30 mg/L; this was added from a 30 mg/mL stock solution after the media had cooled to ambient temperature. $LB_{kan}$ agar plates contained 15 g/L agar and kanamycin was added to 30 mg/L after the agar melt had cooled to below ~45°C.

Phosphate buffer was used in enzyme preparation at 50 mM concentration. Tris-HCl (50 mM, pH 7.8) used for biotransformation reactions was made by diluting a 0.5 M stock. All buffer stock solutions were made in a small volume in order to minimise storage time. Buffer stocks over 1 month old were discarded.

### 2.1.3 Equipment and analysis methods

PCR sequences were performed using a Techne $^{3}$Prime thermocycler. Plasmid mini-preps were carried out using the GeneJet kit from Fisher Scientific. Liquid bacterial cultures were grown in sterile plastic centrifuge tubes (15 mL or 50 mL) or conical flasks at the appropriate temperature and speed using an Eppendorf Excella E24R or Innova 44R refrigerated shaker incubator. Beckman Coulter Allegra X-22R refrigerated bench top centrifuges were used for bacterial culture and enzyme manipulations using a combination of 4 × 250 mL swing-out and 6 × 50 mL, 8 × 40 mL and 48 × 2 mL fixed angle rotors. Smaller scale centrifugation at up to 21,000× *g* in microcentrifuge tubes were carried out on Thermo Scientific Heraeus Pico microcentrifuges. A Fisher Scientific FB505 sonicator equipped with a 1 cm diameter

sonicator probe was used for cell lysis.

UV/vis spectra were recorded on a Varian CARY40 spectrophotometer using 1-cm path length quartz cuvettes. Gas chromatography analysis were performed on a ThermoFisher Scientific TRACE1300 instrument equipped with a split-splitless injector operating in split mode, an AI1310 autosampler and a flame ionisation detector (FID) using helium as carrier gas (1.5 mL/min) and a DB-1 fused silica column (30 m × 0.25 μm diameter and 0.25 μm film thickness). Flash column chromatography was performed using Merck Geduran® silica gel (40–63 μm). Analytical thin-layer chromatography (TLC) was performed on aluminium-backed plates pre-coated with silica gel (0.2 mm, Merck 60 F254), which were developed using phosphomolybdic acid (PMA). Proton ($^{1}$H) and carbon ($^{13}$C) NMR spectra were recorded on Bruker AVII-500 (500/125 MHz), Bruker AVIII-400 (400/100 MHz), or Bruker AV-400 (400/100 MHz) spectrometers. Chemical shifts are quoted in parts per million (ppm), referenced to the appropriate residual solvent peak. Mass spectra (MS) were recorded by the staff of the Department of Chemistry, Oxford on a Bruker Daltonics MicroTOF spectrometer; mass to charge ratios (m/z) are reported in Daltons.

## 2.2 DNA and mutagenesis methods

### 2.2.1 Plasmids and mutant generation

The $P450_{BM3}$ gene had been cloned into the pET-28a(+) vector by Dr C.J.C. Whitehouse using the NcoI and BamH1 restriction sites.[96] The vector conferred kanamycin resistance and the T7 RNA polymerase promoter was inducible in BL21 (DE3) and related hosts.

Primers for mutagenesis (Appendix 1) were designed to replace the target amino acid code with the desired code. The overlapping primer method of Zheng *et al*. was adopted for mutagenesis in order to avoid using complementary primer pairs which suffered from non-specific binding and low PCR efficiency due to the sequence length limitation and high melting temperatures.[97] In general, the mutated amino acid codon was at least 10 bases away from the 5′-end and the staggered 3′-end sequence was around 10 bases. The primers were designed to have $T_m$ = 62°C by adjusting the length of the oligonucleotide.

PCR was carried out in 50 μL reaction volume. The template DNA used in PCR reaction was 10× diluted from the DNA stock from miniprep plasmid preparation experiment. Primers were 10× diluted and a primer mixture was made before adding into a PCR reaction. The thermal cycle was set as follows: initial denaturing at 95°C for 2 min, denaturing temperature 95°C for 20 s, annealing temperature 62°C for 10 s, extension temperature 70°C for 3 min 20 s; the cycle was repeated 30 times followed by a final extension at 70°C for 5 min. The reaction was then cooled to 4°C until further processing.

**Table 2.2** Standard reaction setup for PCR reaction

| Components | Volume / μL | Final Concentration |
|---|---|---|
| **10× buffer** | 5 | 1× |
| **25 mM $MgSO_4$** | 4 | 1.5 mM |
| **dNTPs (2 mM each)** | 5 | 0.2 mM (each dNTP) |
| **Forward & Reverse primer mixture (10 μM each)** | 3 | 0.3 μM (each primer) |
| **DNA (10× dilution)** | 1 | >10 ng each reaction |
| **KOD polymerase** | 1 | 0.02 U/μL |
| **$H_2O$** | 31 | To 50 μL |

A 10 μL aliquot of a PCR mixture was mixed with 2 μL 6× loading buffer to prepare the loading solution for agarose gel electrophoresis. The 1% *w/v* agarose suspension in TAE buffer was heated in a microwave (750 W, 4 × 30 s) to dissolve the agarose. The solution was then poured into a plastic mould and left to cool to room temperature to form the gel. Previous experiments had shown that the PCR mutagenesis product migrated with similar mobility to an 8.9 kbp linearised fragment in common DNA size ladders. Once an intense band with this mobility was observed in the gel, 5 μL of 10× restriction enzyme buffer and 1 μL DpnI were added to the remaining 40 μL of the PCR mixture to digest the template plasmid at 37°C for at least 2 h or at room temperature overnight.

The digested DNA was then transformed into *E. coli* DH5α competent cells for plasmid repair and propagation for purification for DNA sequencing. The digested reaction mixture (5 μL) was added to 20 μL of competent cells that had been thawed

out on ice from storage at –80°C. After gentle mixing the mixture was left on ice for 15 min, heat shocked at 42°C for 45 s and then left on ice for 15 min. SOC media (1 mL) that had been pre-warmed at 37°C was added and the transformed cells were gently resuspended to make the recovery culture. The recovery process was carried out by shaking the culture at 37°C and 180 rpm for at least 1 h. The culture (100 μL) was then spread on a $LB_{kan}$ agar plate and grown in an incubator oven at 37°C for 24 hours.

A single colony was picked from the plate and inoculated in to 5 mL $LB_{kan}$. After incubation at 37°C at 180 rpm for no more than 18 h, the culture was used for plasmid extraction. Cells were collected by centrifugation at 5,000× *g* for 5 min and the supernatant discarded. The pellet was resuspended in 250 μL resuspension buffer (RNAase added) and transferred into a 2 mL microcentrifuge tube. Lysis buffer (350 μL) was added and gently mixed. After standing for 1 min, 350 μL of neutralisation buffer was added into the mixture and gently mixed. Cell debris was pelleted by centrifugation at 21,000× *g* for 10 min and the supernatant was added into a DNA-binding filter column followed by centrifugation at 10,000× *g* for 1 min. The flow through was discarded and the column washed twice with 800 μL wash buffer, centrifuging at 10,000× *g* for 1 min after the first aliquot and for 2 min after the second in order to remove as much of the residual buffer as possible. The column was left to dry under ambient atmosphere for 30 min to allow any residual ethanol on the membrane to evaporate. Finally, elusion buffer (40 μL) was added directly onto membrane, left for 5 min before centrifuging at 10,000× *g* for 5 min to collect the DNA solution. A 2 μL aliquot was mixed with 2 μL of 6× gel loading buffer and 10 μL 1×

TAE buffer to prepare a loading solution for agarose electrophoresis. If the plasmid DNA was clean, 5 μL of the DNA solution was sent for DNA sequencing by Source BioScience to confirm the outcome of mutagenesis. The DNA solution was stored at –80°C.

### 2.2.2 Recombinant protein production and purification

Once the sequence of a target variant was confirmed, the plasmid was transformed into *E. coli* BL21 (DE3) competent cells for recombinant protein production. The transformation protocol was the same as it was for the DH5α strain. One single colony was picked from the plate to inoculate into 500 mL $FB_{glycrol}$ media with 30 mg/L kanamycin and the culture was grown at 37°C at 160 rpm for 24 h. A 10 mL aliquot of the culture was removed and mixed with an equal volume of autoclaved 60% *v/v* glycerol in type 1 water to make a cell bank in 30% *v/v* glycerol. The cell bank was stored in 1 mL aliquots and stored at –80°C for up to one year. In repeat growth each 500 mL of $FB_{glycerol}$ media was inoculated with 1 mL of the thawed cell bank.

The main culture was cooled to 20°C in the shaker and 1 mL each of filter-sterilised induction solution and autoclaved 0.5 g/mL tryptone were added to induce the production of the $P450_{BM3}$ variant. The culture was grown at 20°C and 120 rpm for at least 48 h. Further 1 mL aliquots of induction mixture and tryptone solution were added to each flask every 12 h in order to maintain the antibiotic selection pressure and provide sufficient supplement for protein production.

Cells were then harvested by centrifugation at 3,900× *g* for 15 min and the supernatant was discarded. The cell pellet was resuspended with 40 mL of 50 mM

phosphate buffer, pH 7.8. Lysozyme was then added to a 500 U/L working concentration and the cell suspension was incubated in a shaker at 25°C, 120 rpm for 2 h. Cells were lysed by sonication on ice using 10 cycles of 15 s sonication at 55% amplitude then 15 s of cooling with the power off, for a total of 5 min of sonication. Cell debris was removed by centrifugation at 9,600× *g* for 15 min at 4°C. The supernatant was collected and the $P450_{BM3}$ variant was partially purified by ammonium sulfate fractionation. The saturation gradient was set as 30% then 40% to remove impurity proteins, lipids, etc., and finally the $P450_{BM3}$ enzyme was precipitated at 60% saturation. Centrifugation at 9,600× *g* for 10 min was employed when each $(NH_4)_2SO_4$ gradient was achieved; retaining the supernatant at 30% and 40% but the enzyme pellet at 60% saturation was collected and re-dissolved in 5 mL of 50 mM phosphate, pH 7.8.

The enzyme concentration was determined by the reduced-CO difference assay. The UV-Vis spectrum of the enzyme solution was recorded between 250 nm and 700 nm. A few grains of solid sodium dithionite were added to reduce the enzyme and the spectrum of this reduced enzyme was set as the baseline. A few bubbles of CO was slowly passed through the solution and the spectrum was recorded.[98] This step was repeated to ensure that all the reduced enzyme had bound CO. The absorbance at 450 nm and 490 nm were noted and the P450 enzyme concentration was calculated using an extinction coefficient $\varepsilon_{450–490}$ = 91,000 $M^{-1}$ $cm^{-1}$.[99] The enzyme was then mixed with an equal volume of autoclaved 60% *v/v* glycerol and stored at –20°C.

## 2.3 *In vitro* screening, preparative scale reaction and product characterisation

### 2.3.1 *In vitro* screening

All sesquiterpene substrates were dissolved in ethanol to make 100 mM stock solutions. $NADP^+$, GDH and glucose were added as the regeneration system to supply NADPH required for $P450_{BM3}$ monooxygenase activity. 2-Hydroxylpropyl-β-cyclodextrin (HPβCD) was added to increase substrate solubility. Terpene oxidation reactions contained these components at the concentrations given in Table 2.3

**Table 2.3** Reaction components for sesquiterpene oxidation screening

| Reagent | Final Concentration |
|---|---|
| $NADP^+$ (2 mM stock in water) | 20 μM |
| GDH (1 U/μL stock in water) | 1 U/mL |
| Glucose (1 M stock in water) | 200 mM |
| Substrate (100 mM stock in EtOH) | 1 mM |
| Enzyme (20 μM) | 2 μM |
| HPβCD (250 mM in water) | 25 mM |
| Tris-HCl buffer, pH 7.8 | 50 mM |
| Total reaction volume | 1 mL |

Screening reactions were carried out in 24-well plates (Sarstedt). A small volume of the solution stock of each $P450_{BM3}$ variant was diluted to 20 μM and 100 μL of this was placed in one well. A master mixture was made that contained all the other components and a 900 μL aliquot was added into each well. Each plate was then

shaken at 120 rpm in an incubator shaker at 20°C for 16 to 24 h. A 500 μL aliquot was taken from each reaction and organics were extracted with 300 μL of ethyl acetate by vortexing the mixture for at least 1 min. The phases were separated by centrifuging at 14,000× *g* for 2 min. The organic phase was removed and placed in glass inserts held in glass vials sealed with a silicon rubber septum and analysed by gas chromatography. The autosampler injects 2 μL onto the column at a split ratio of 25. Both the injector and flame ionisation detector were held at 250°C. The oven temperature was initially at 120°C and held for 1 min and then increased at 15°C/min to 240°C and held for 4 min. Sesquiterpene substrates were eluted at ~5 min while the epoxides, with the shortest retention times of the products, eluted at ~6 min.

Substrate conversion and product distribution were determined from the GC output in order to identify the activity and selectivity of different enzyme variants. Substrate-variant combinations with high conversion (>70%), high selectivity or diverse product formation were selected for preparative scale reactions for product purification and characterisation by NMR and MS data.

### 2.3.2 Preparative scale reactions and product purification

Preparative scale reactions for product purification and characterisation were carried out under identical reaction conditions as those of screening reactions except that the volume was increased so that 100 mg of substrate was added. Reactions were monitored by GC analysis of ethyl acetate extracts. After no further conversion was observed the reaction mixture extracted by three times with an equal volume of ethyl acetate. The aqueous-organic mixture often became gel-like, necessitating

centrifugation at 2,000× *g* to separate the phases. Yields were found to be improved by shaking the centrifuge bottle to break up the material at the aqueous-organic interface and centrifuging again for each extraction step. The extracts were combined, washed with water and then brine, dried over anhydrous $MgSO_4$ and filtered. Solvent was then removed by rotary evaporation. Thin layer chromatography (TLC) was conducted to select the solvent mixture for product separation by silica gel column chromatography. The alcohol derivatives of a sesquiterpene differed so little in polarity that it was difficult to achieve good separation by a single chromatography step, and several rounds of chromatography using different solvent mixtures were necessary in order to obtain pure compounds.

## 2.4 Final conditions for valencene scale up study

### 2.4.1 100 mL scale reaction conditions

Reactions at 100 mL scale were carried out to optimise the conditions (Table 2.4). The reactions were carried out in a 3-neck 250 mL round bottom flask with pH titration using 1 M KOH and aeration using compressed air from a cylinder. It was not possible to control the aeration rate precisely. However, the flow rate was regulated at a suitable level to avoid foaming from the reaction mixture. Valencene was added from a PEG suspension of 320 μL valencene (MW = 204.35 g/mol, density = 0.92 g/mL at 25°C) in 10 mL PEG-400 and vortexed to suspend. An additional 0.5% *v/v* of PEG-400 was added into the reaction mixture to increase substrate solubility. The enzyme concentration was 3 μM.

**Table 2.4** Reaction condition of 100 mL scale

| Reagent | Final Concentration |
|---|---|
| $NADP^+$ | 10 μM |
| GDH | 5 U/mL |
| Glucose | 200 mM |
| Valencene | 10 mM |
| P450 enzyme | 3 μM |
| PEG-400 | 0.5% *v/v* |
| 2-Hydroxypropyl-β-cyclodextrin | 20 mM |
| Tris-HCl buffer, pH 7.8 | 50 mM |

### 2.4.2 Reaction conditions at 3 litre scale in a bioreactor

The reaction conditions for 3-litre scale reactions were optimised from those at 100 mL scale. Reactions were carried out using an Eppendorf Bioflo 3000 bioreactor system with a 5-litre working volume vessel with a water jacket. The reaction temperature was maintained at 25°C using a recirculating water bath, the impeller speed was between 200 and 400 rpm depending on the dissolved oxygen level, and the pH was maintained at 7.6; all these parameters were monitored and controlled in real-time by the central cascade software of the apparatus. A condenser maintained at 4°C was attached to the gas exhaust outlet to minimize loss of organics by evaporation. The mole ratio of substrate to enzyme was 3000 : 1 and that of HPβCD to substrate was 2 : 1. The valencene concentration was 30 mM and the concentration of GDH and glucose were

increased 2-fold compared to the 100 mL scale reaction. The cell-free enzyme extract was supplied by Oxford Biotrans Ltd. The enzyme concentration was between 50 and 100 μM, and the material was centrifuged at 9,600 g for 30 min to remove any precipitate before use.

**Table 2.5** Reagents in 3 L valencene oxidation reaction

| **Reagent** | **Final Concentration** |
|---|---|
| $NADP^+$ | 20 μM |
| GDH | 10 U/mL |
| Glucose | 400 mM |
| Valencene | 30 mM |
| P450 enzyme | 10 μM |
| 2-Hydroxypropyl-β-cyclodextrin | 60 mM |
| Tris-HCl buffer, pH 7.8 | 50 mM |

# Chapter 3

# Engineering $P450_{BM3}$ for Sesquiterpene Oxidation

## 3.1 Sesquiterpene oxidation

The commercially available sesquiterpenes (+)-valencene, α-humulene, β-caryophyllene and α-cedrene (Fig. 3.1), were selected as target compounds for P450 biotransformation.

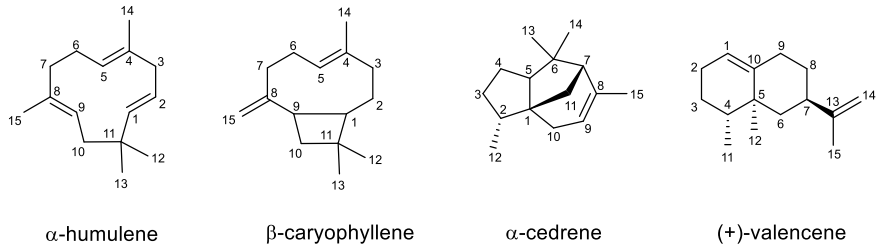



**Fig. 3.1** Sesquiterpenes selected for P450-catalysed oxidation.

As the two major components of hop essential oil (17 to 33% of α-humulene, and 6 to 10% of β-caryophyllene), the chemical oxidation of α-humulene and β-caryophyllene has been studied extensively.[100] Oxygenated derivatives of these two sesquiterpenes are suggested to contribute to the aroma of hops and beer.[101,102] Valencene is the main sesquiterpene in orange essential oil and could be obtained by fractional distillation. The annual production of orange oil is ~30,000 tonnes so valencene is in plentiful supply. α-Cedrene is one of the main sesquiterpenes (20 to 40%) in cedarwood oil; the other is thujopsene. Derivatives of α-cedrene provide woody scents; many of them have been produced and used in commercial woody fragrance products (see Chapter 4)

Recently, α-humulene and β-caryophyllene have been widely studied because of their anti-inflammatory properties and some of their derivatives have shown anti-cancer properties.[103,104] The C3 ketone of humulene, zerumbone (Fig. 3.2), was

extensively reported as a promising anti-tumour compound.[88,105,106] Chemical synthesis of zerumbone has been reported but there are limitations.[107,108] The C2 ketone of valencene, nootkatone, is the grapefruit aroma and flavour compound and has an attractive bitter taste.

A panel of $P450_{BM3}$ variants was screened for α-humulene and valencene oxidation in order to find variants and mutations that directed C3 oxidation of humulene and C2 oxidation of valencene for the synthesis of zerumbone and nootkatone, respectively. β-Caryophyllene was selected for direct comparison of selectivity traits with the closely related α-humulene.

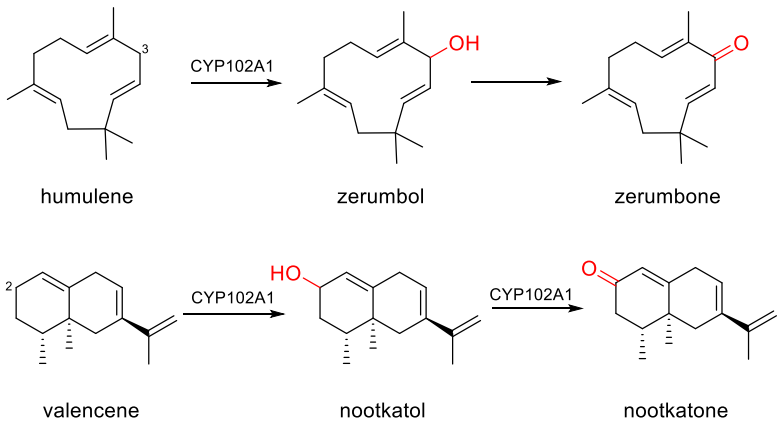



**Fig. 3.2** Proposed reaction route to hydroxylate at C3 position of humulene and C2 position of valencene to zerumbone and nootkatone.

The selectivity of α-cedrene oxidation could be compared with humulene and caryophyllene to explore the effect of the decrease in structural flexibility and the numbers of C=C double bonds. The bicyclic structure of β-caryophyllene shares a similar monocyclic skeleton with α-humulene but for the extra C–C bond between C1 and C9. There are three C=C double bonds in α-humulene and two in β-caryophyllene. α-Cedrene presents a rigid tricyclic structure with only one C=C double bond, between

C8 and C9. Based on the initial screening results, two-residue and three-residue orthogonal libraries were constructed to alter the regioselectivity of cedrene oxidation (see Chapter 4).

## 3.2 Mutants and general methods

### 3.2.1 A $P450_{BM3}$ mutant library

The $P450_{BM3}$ enzyme library was based on a set of variants identified from directed evolution to possess the desirable property of enhanced activity for the oxidation of unnatural substrates without affecting the product profile of the starting template enzyme.[73] The base mutants were KSK19 (F87A/H171L/Q307H/N319Y), KT2 (A191T/N239H/I259V/A276T/L353I), AP (A330P), and IP (I401P). Detailed biochemical and structural studies have shown that these variants have unexpected conformation changes that led to weakened interactions between the heme iron and the axial water molecule, resulting in lowered barriers to substrate binding and electron transfer.[75,109] Hence, the energetics for the binding of an unnatural substrate to displace the heme-bound water to initiate the catalytic cycle was more favourable, leading to catalytic turnover of a greatly expanded range of substrates for these variants.

Further active site mutations have been introduced over a period of time to alter the active site topology to accommodate a wide range of molecules including short and medium chain alkanes, steroids and drug molecules.[84] Target residues included Arg47 and Tyr51 at the entrance of the substrate access channel (Fig. 3.3). The R47L/Y51F (RLYF) combination had been shown to promote the binding and oxidation of hydrophobic molecules.[84] Combinations of two or more hydrophobic substitutions with

different side chain volumes have been introduced to the base mutants at Ser72, Ala74, Val78, Phe81, Ala82, Phe87, Ile263, Ala264, Glu267, Ala328, Ala330 and Leu437 to generate a library of >100 variants.[86] 24 variants from this collection were selected to fill in a 24-well plate for the screening experiments (Table 3.2). The principle of selection was to cover as many active site residues as possible and to include more types of combinations, in order to test the combined effect of mutations.

**Table 3.1** The selection of $P450_{BM3}$ variants for sesquiterpene oxidation.
**WT**: wild type; **RLYF**: R47L/Y51F; **RP**: R47L/Y51F/I401P; **GVQ**: A74G/F87V/L188Q; **K19**: H171L/Q307H/N319Y; **RT2**: R47L/Y51F/A191T/N239H/I259V/A276T/L353I;

| | |
|---|---|
| WT | GVQ |
| RP | GVQ/S72W |
| RP/F87V | GVQ/I263G |
| RP/E267V | K19/F87V |
| RP/F87V/E267V | K19/F87A/V78I |
| RP/I263A/A330W | K19/F87A/A328I/A184I |
| RLYF/Q403P/I263A/A82M | K19/F87A/A82M |
| RP/F87A | K19/F87A/I263A |
| RP/A330P | K19/F87A/A328I/I263A |
| KT2 | K19/F87A/I263A/A330W |
| RLYF/KT2 | RLYF/K19/F87V |
| RLYF/KT2/A330P | RLYF/K19/F87A/A328I(RKA) |

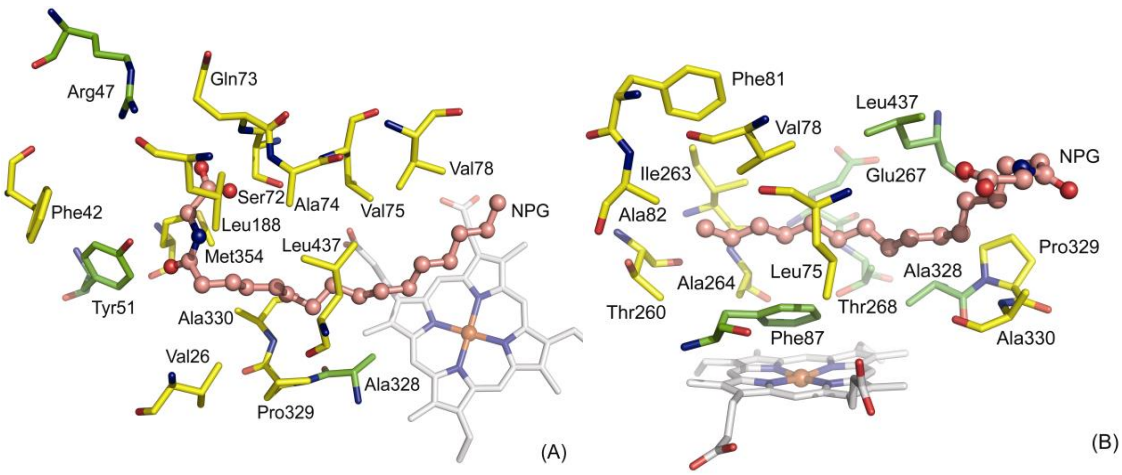



**Fig. 3.3** Two views of the active site structure of *N*-palmitoylglycine (NPG) bound wild-type $P450_{BM3}$ showing the key residues that contact the bound fatty acid substrate.

## 3.3 Humulene and caryophyllene

### 3.3.1. Humulene screening result and product isolation

WT $P450_{BM3}$ had barely detectable activity for humulene oxidation. The base variants without active site mutations, namely KT2 and RP, showed limited activity of 15% and <5% substrate conversion, respectively. Variants with the active site mutation F87A showed excellent conversion of over 90% and most of these showed similar regioselectivity (Fig. 3.4). The 8,9-epoxide **1** was the most common product, as well as the diepoxides **11** and **12** (see Appendix for NMR and MS data). The C7 alcohol **2** was a major mono-hydroxylation product in many reactions, in particular for variants containing active site mutations such as F87A, F87V and mutations at A328. Variants containing F87V mutations, such as GVQ/S72W, RP/FV and GVQ/IG showed C3 oxidation to give the allylic alcohol 3-hydroxy-humulene **3**, or zerumbol the precursor of zerumbone. The GVQ/IG variant displayed >95% conversion, with 12% of **3** and also showed preference for formation of the diepoxide **12** rather than **11.**

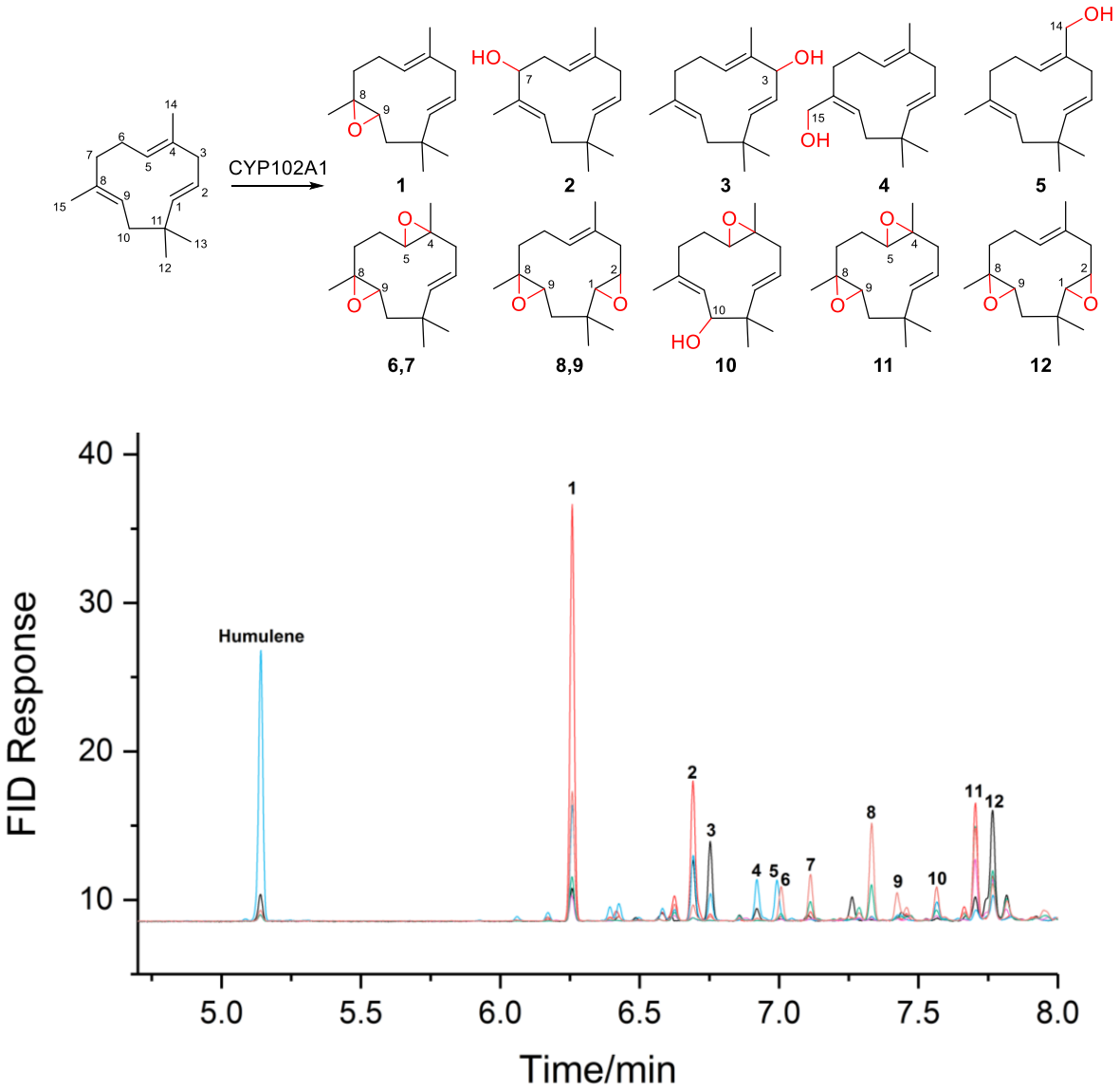



**Fig. 3.4** GC profile of α-humulene initial screen. Humulene and characterised products are labelled.

The RLYF/K19/F87A/A328I (RKA), GVQ/I263G (GVQ/IG), RP/F87A (RP/FA) and RP/F87V (RP/FV) variants were selected for preparative scale reactions (80 mg in 500 mL reaction volume) to isolate and characterise products **1** to **12** (Fig. 3.5). The reaction conditions are described in Chapter 2. The reaction mixtures were extracted with ethyl acetate and the products separated by flash silica gel column chromatography eluting with a gradient ratio of hexane : ethyl acetate mixture from 20 : 1 gradually increasing to 6 : 1, to achieve the isolation of different oxygenated

humulenes. Fractions were collected and analysed by TLC and GC before being pooled for analysis by NMR and MS.

The products **3** and **12** were isolated from a GVQ/I263G reaction in 25% and 42% yield, respectively. The C7 alcohol **2** showed very similar polarity as **3** during column chromatography, causing significant loss when separating both from the mixture. In the RKA reaction, **1**, **2** and **11** were major products and a reasonable yield of **2** was isolated for characterisation. Therefore, **3** could be characterised from a mixture of **2** and **3** from GVQ/IG reactions by comparing the NMR spectra with those of pure **2** isolated from the RKA reaction. The products **6**, **7**, **8**, **9**, and **10** were isolated from reactions using RP/FA, while **8** and **9** co-eluted regardless of the solvent mixture used. Fractions from reactions using GVQ/IG, RKA and RP/FA that contained **11** and **12** were combined and these two compounds were characterised as a mixture. Among these four preparative scale reactions, RP/FV gave the lowest conversion at 65% but with all products being limited within the early retention time range. Products **4** and **5** were also difficult to separate; however, **5** was purified from the mixture for characterisation. Therefore, in order to avoid further loss, **4** was characterised as the mixture with **5** by comparison with the spectra for pure **5**.

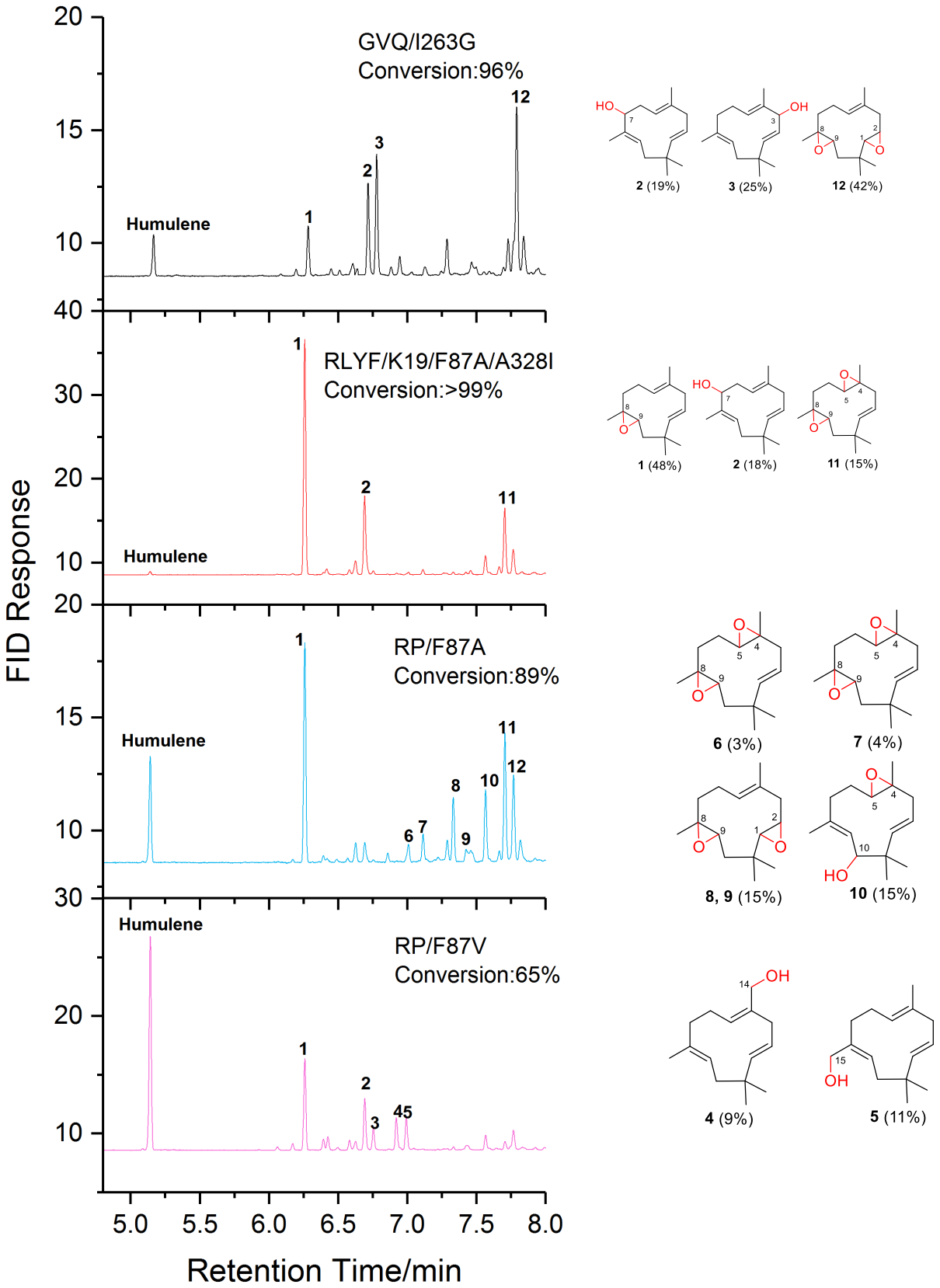



**Fig. 3.5** GC analysis of preparative scale reactions with variants selected from screening results. Products isolated and purified from the reactions are shown on the right.

### 3.3.2. Caryophyllene oxidation and product characterisation

As α-humulene is also known as α-caryophyllene, the ring-opened isomer of β-caryophyllene, the panel of variants was screened for β-caryophyllene oxidation.

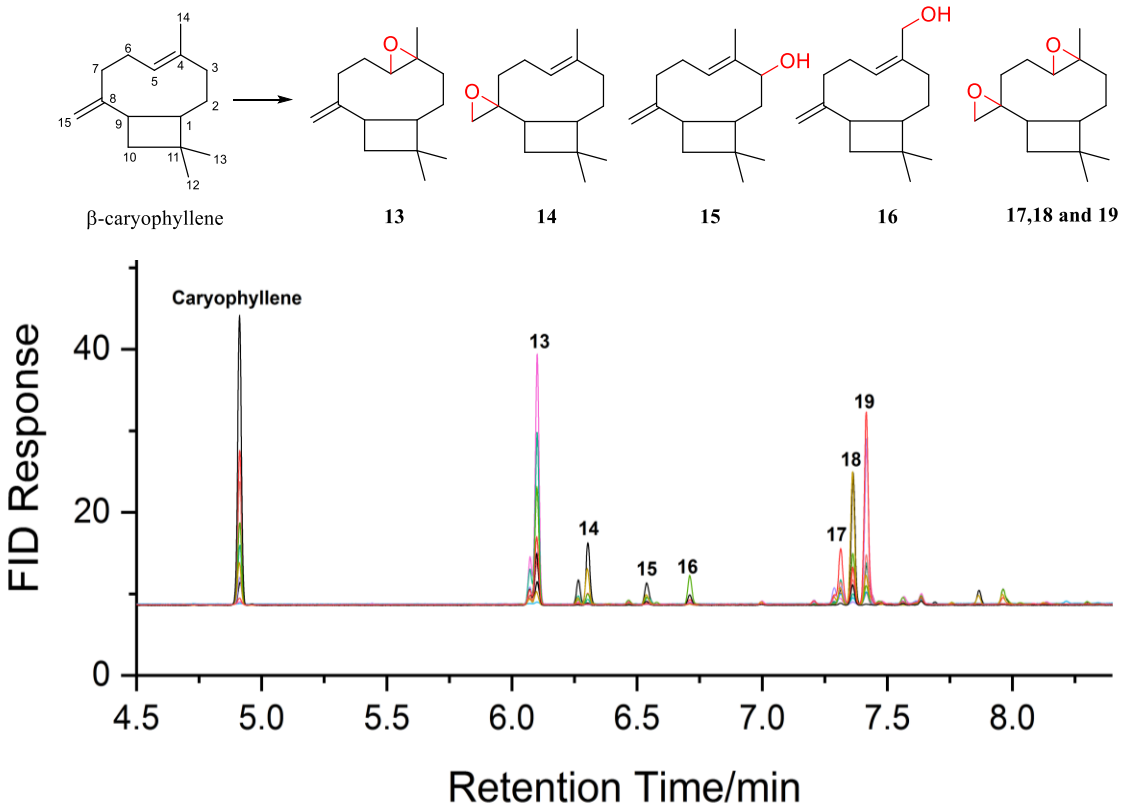



**Fig. 3.6** GC profile of β-caryophyllene initial screen. Characterised products are labelled.

Similar to humulene oxidation, WT and the base variants without active site mutations barely showed any caryophyllene conversion. The endocyclic 4,5-epoxide **13** and the di-epoxidation products **17**, **18** and **19** were major products across all screening reactions. This pattern minored that for humulene. Variants with an F87V mutation showed significant variations in selectivity (Fig. 3.7). RP/F87V was the only variant to give the C14 allylic alcohol **16** while GVQ/I263G increased the selectivity for the exocyclic 8,15-epoxide **14** and the C3 alcohol **15**.

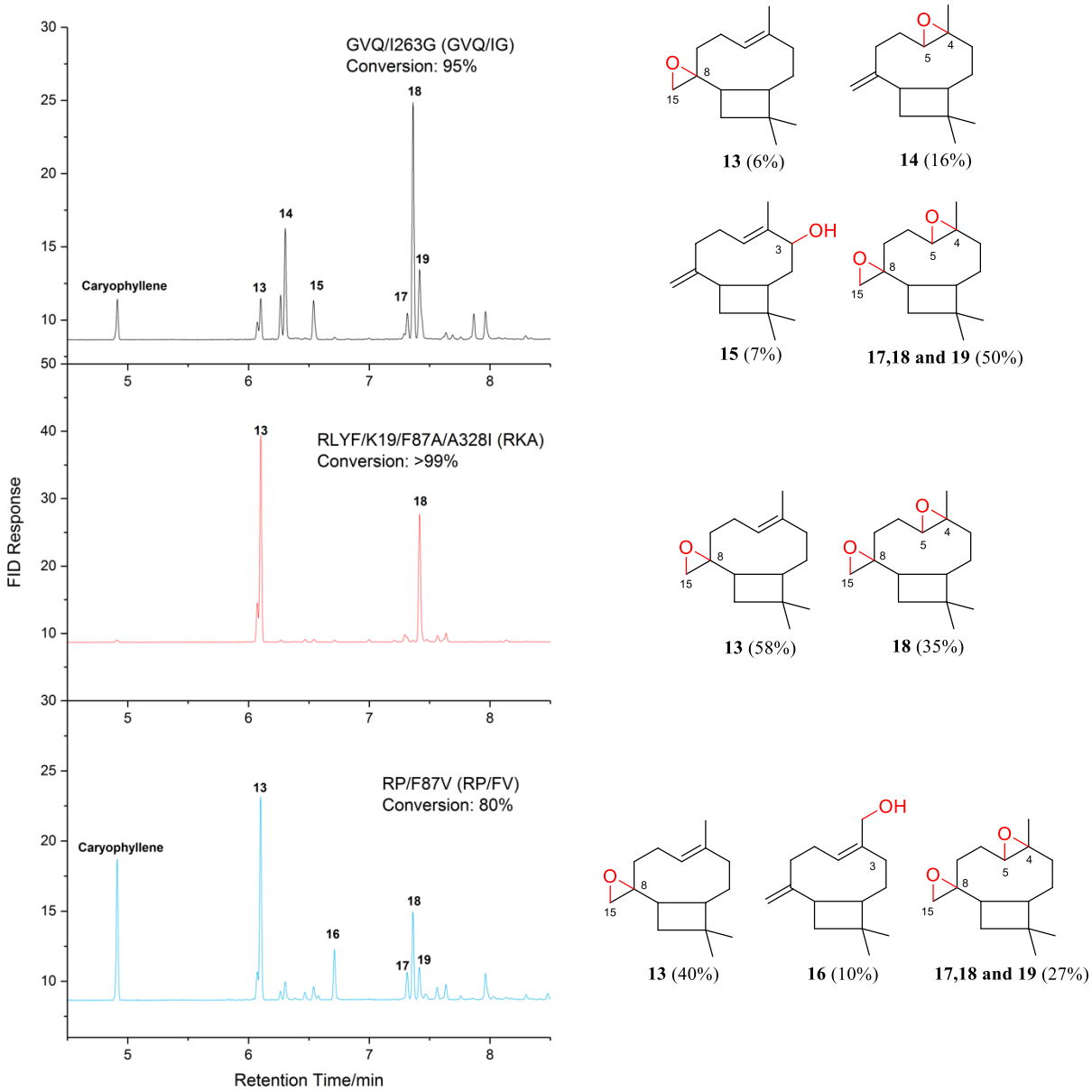



**Fig. 3.7** The GC profile of preparative scale reactions with selected variants from screening results. Products isolated and characterised from the relevant reactions are shown on the right.

Three variants were selected for preparative scale reactions according to the screening results (80 mg in 500 mL reaction volume). The reaction conditions, work up and purification procedures were similar to those for humulene. The products **13**, **14**, **15**, **17**, **18** and **19** were isolated from the preparative reaction with GVQ/IG, while **13** and **18** from the reaction with RKA were combined with those from GVQ/IG for a concentrated sample for characterisation. **16** was isolated from the reaction with RP/FV.

### 3.3.3. Product characterisation

All products were characterised by NMR and MS data. The spectral analysis and assignment of three products are presented here as examples to show the general procedure.

#### 3.3.3.1 Humulene Oxidation Products

Product **2** was characterised as 7-hydroxy-humulene from the NMR data (Fig. 3.8). The four alkene protons H1, H2, H5 and H9 were identified at low field suggesting the presence of all three double bonds in humulene. The low-field one-proton resonance at 4.08 ppm suggested an allylic CHOH proton which was confirmed as H7 via the coupling to two H6 protons, and coupling of these with H5 in the COSY spectrum. Resonances due to other protons were readily assigned from the COSY and HSQC data.

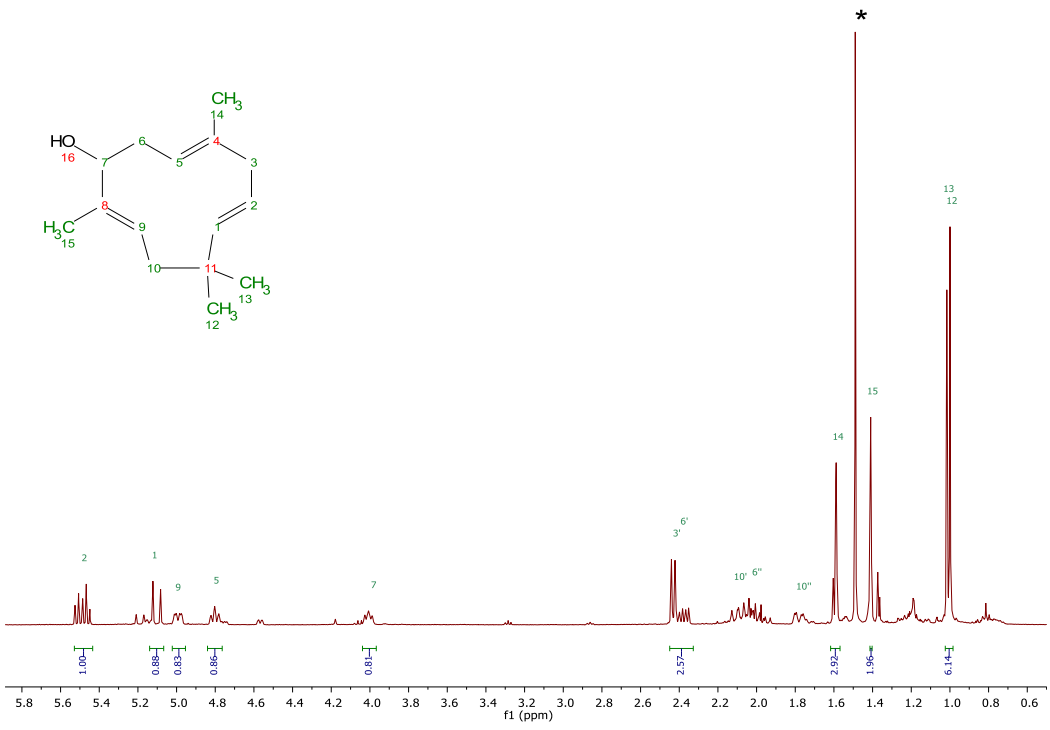



**Fig. 3.8** The $^{1}$H NMR spectrum of product **2** in $CDCl_3$. Asterisk denotes resonance due to water.

For product **3**, H1 and H2 were identified according to their chemical shift and coupling with each other (Fig. 3.9). H5 and H9 were identified via chemical shift and coupling with protons on adjacent carbon position of C6 and C10 respectively. The one-proton resonance at 4.56 ppm was coupled to H2. This is assigned to the C3HOH, with its diallylic position causing the significant shift to low field. The MS result of M+16 showed monooxygenated humulene, hence the product **3** was confirmed as 3-hydroxy-humulene.

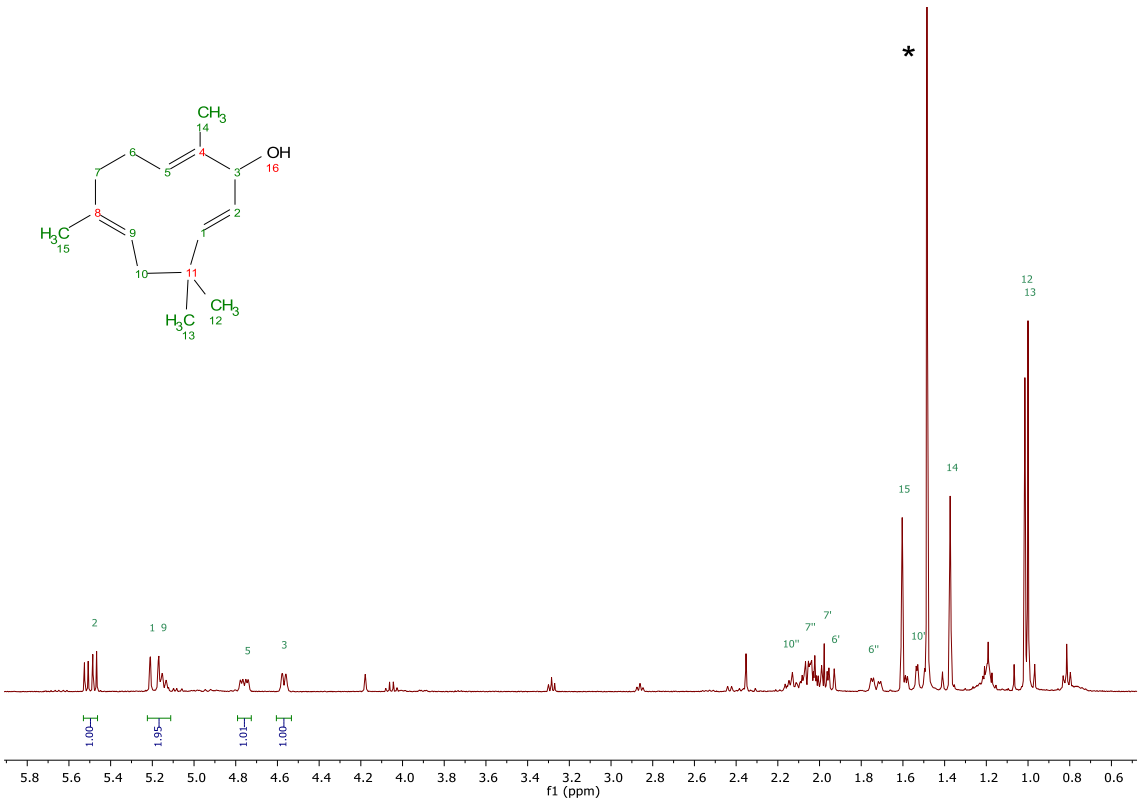



**Fig. 3.9** The $^{1}$H NMR spectrum of product **3** in $CDCl_3$. Asterisk denotes resonance due to water.

Product **10** was at a similar GC retention time to the diepoxides **11** and **12** but had very different polarity from diepoxides as it was eluted in earlier fractions than **11** and **12** with the same solvent combination. The alkene protons H1 and H2 were identified at 5.39 ppm and 5.07 ppm, therefore two H3 protons were identified via their coupling with H2. Another proton resonance was at very similar chemical shift to H1 showing strong mutual coupling with the proton doublet at 4.01 ppm. No further coupling was observed for these two resonances, which suggested an alkene proton next to a C<u>H</u>OH. As H5 was identified in the epoxide region via the coupling with H6 and H7, this alkene proton at 5.07 ppm was assigned as H9. Hence, the C<u>H</u>OH was confirmed as H10 and **10** was assigned as 10-hydroxy-humulene-4,5-oxide.

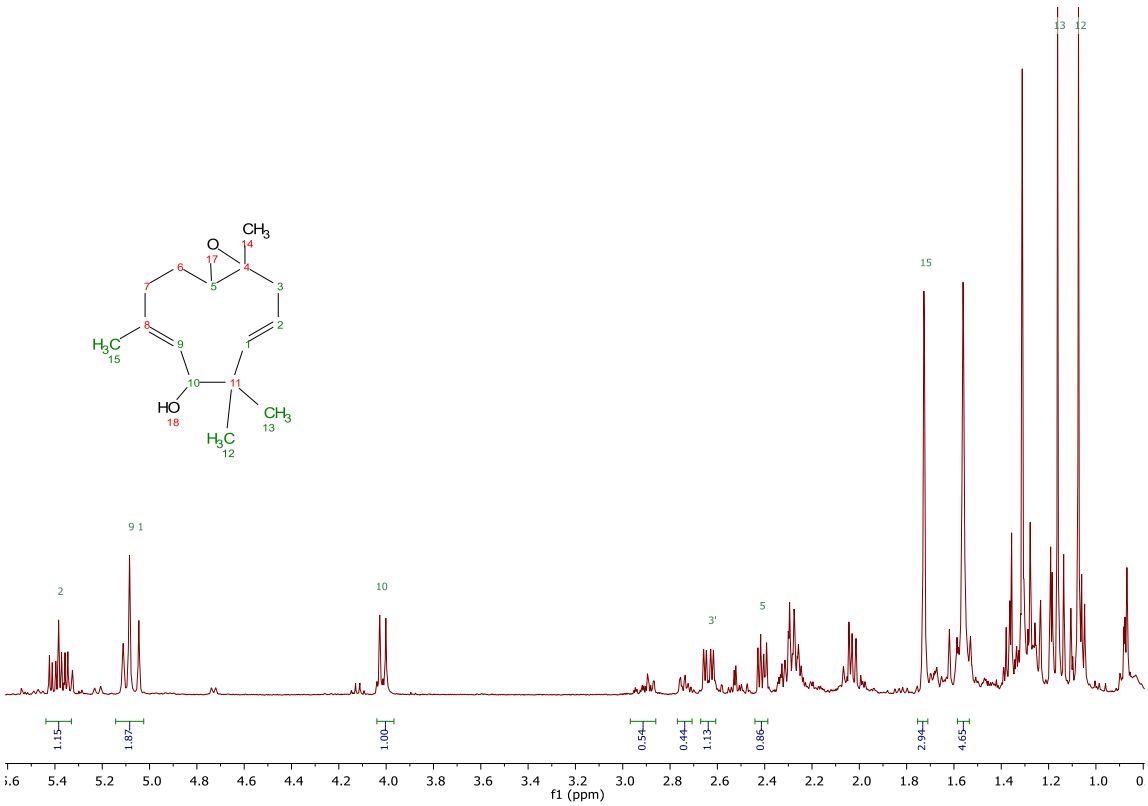



**Fig. 3.10** The $^{1}H$ NMR spectrum of product **10** in $CDCl_3$.

### 3.3.3.2 Caryophyllene Oxidation Products

Product **13**, the caryophyllene endo-epoxide**,** was one of the products found across reactions with many variants. The $^{1}$H NMR spectrum (Fig. 3.11) showed the presence of two H15 alkene protons at 4.97 ppm and 4.85 ppm but the H5 resonance had moved from the alkene region to the epoxide region (2.88 ppm). The other resonances were assigned by the COSY and HSQC data, and **13** was characterised as the C4=C5 double bond epoxidation product.

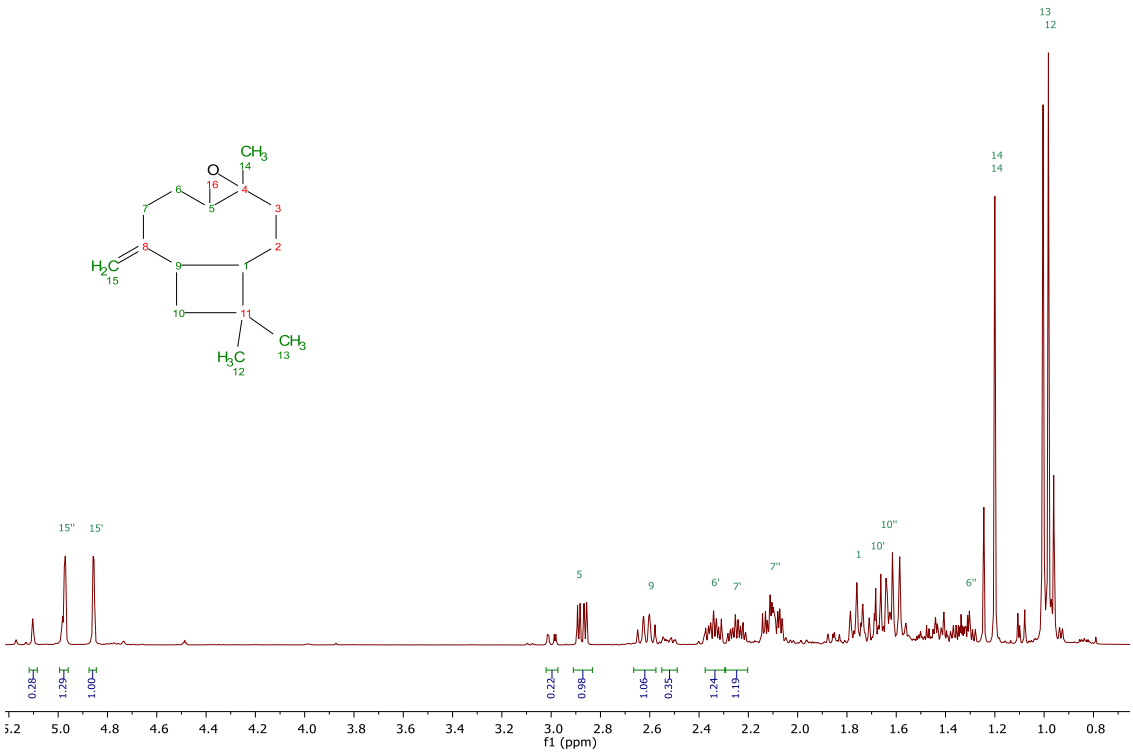



**Fig. 3.11** The $^{1}$H NMR spectrum of product **13** in $CDCl_3$.

Product **15** was characterised as the C3 allylic alcohol. The $^{1}$H NMR spectrum (Fig. 3.12) suggested the presence of the C8=C15 double bond and the H5 alkene proton. Two H6 protons and two H7 protons were then identified in the allylic region via the COSY spectrum. The C<u>H</u>OH resonance at 4.01 ppm was identified as H3 via

coupling with H2 and H1. Therefore, product **15** was characterised as 3-hydroxy-caryophyllene.

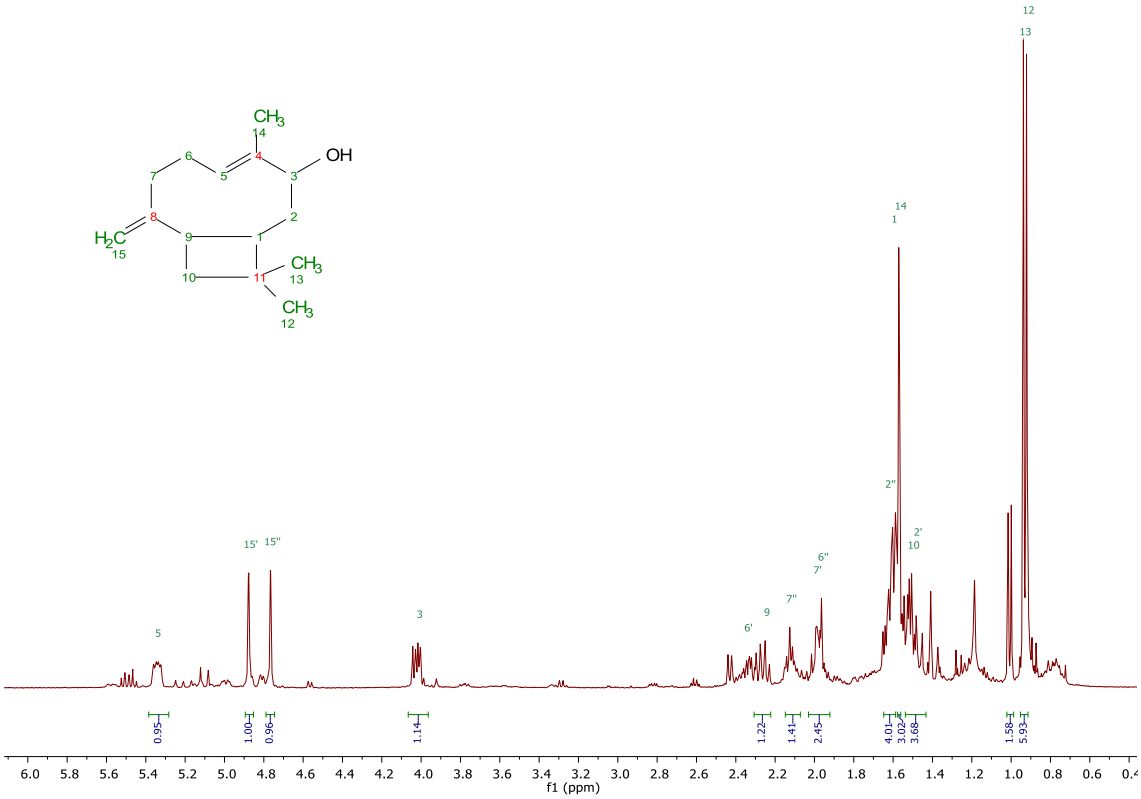


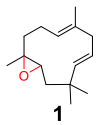
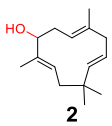
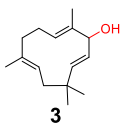
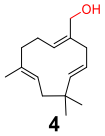
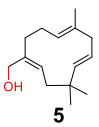
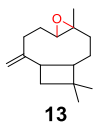
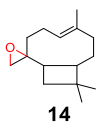
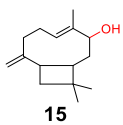
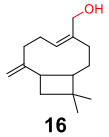

**Fig. 3.12** The $^{1}$H NMR spectrum of product **15** in $CDCl_3$.

## 3.3.4 Discussion

Of the three tested sesquiterpenes humulene has the highest number (three) of C=C double bonds. As a result, the secondary and primary C–H bonds are allylic, except for the aliphatic C12 and C13 methyl groups. With less ring strain from the monocyclic structure, the panel of $P450_{BM3}$ variants was able to oxidise all three double bonds and most of the C–H bonds via mono- and di-oxygenation except at C6, C12 and C13. With its fused cyclononenyl and cyclobutyl bicyclic structure leading to ring strain and the presence of fewer allylic C–H bonds, the caryophyllene oxidation activity was

decreased and selectivity was restricted to epoxidation being the major reactivity. Only two allylic hydroxylation products were characterised while there were five other epoxidation products.

**Table 3.2** Selectivity of mono-oxidation product of humulene and caryophyllene from selected variants.

| Variant | Humulene Product | | | | | Caryophyllene Product | | | |
|---|---|---|---|---|---|---|---|---|---|
| |  1 |  2 |  3 |  4 |  5 |  13 |  14 |  15 |  16 |
| RKA | 48% | 18% | 15% | – | – | 58% | – | – | – |
| GVQ/IG | 19% | 25% | 42% | 2% | – | 6% | 16% | 7% | – |
| RP/FV | 23% | 14% | 6% | 9% | 11% | 40% | 5% | 2% | 10% |
| RP/FA | 24% | 2% | – | – | – | 14% | – | – | – |

The humulene and caryophyllene oxidation product profiles of some selected variants are given in Table 3.2. The F87A mutation promoted epoxidation and further oxidation of both humulene and caryophyllene. The RP/FA and RKA variants gave the product **1** via C10 hydroxylation of humulene-4,5-oxide (Fig. 3.13). F87V had the effect of shifting the selectivity from epoxidation towards hydroxylation of allylic positions. RP/FV showed unique selectivity of allylic methyl oxidation in both substrates. F87V in GVQ/IG combined with I263G biased selectivity towards secondary allylic C–H bonds on the ring, especially enhancing the selectivity for C3 hydroxylation in both humulene and caryophyllene. The A328I mutation in RKA increased selectivity for C7 oxidation of humulene. However, there was no significant selectivity observed on caryophyllene on the similarly allylic C7 position, nor was the C8,C15 exocyclic double bond oxidised. This suggested different binding orientation

to humulene. The A328I mutation also increased the activity of RKA which gave >99% conversion of both humulene and caryophyllene in preparative scale reactions.

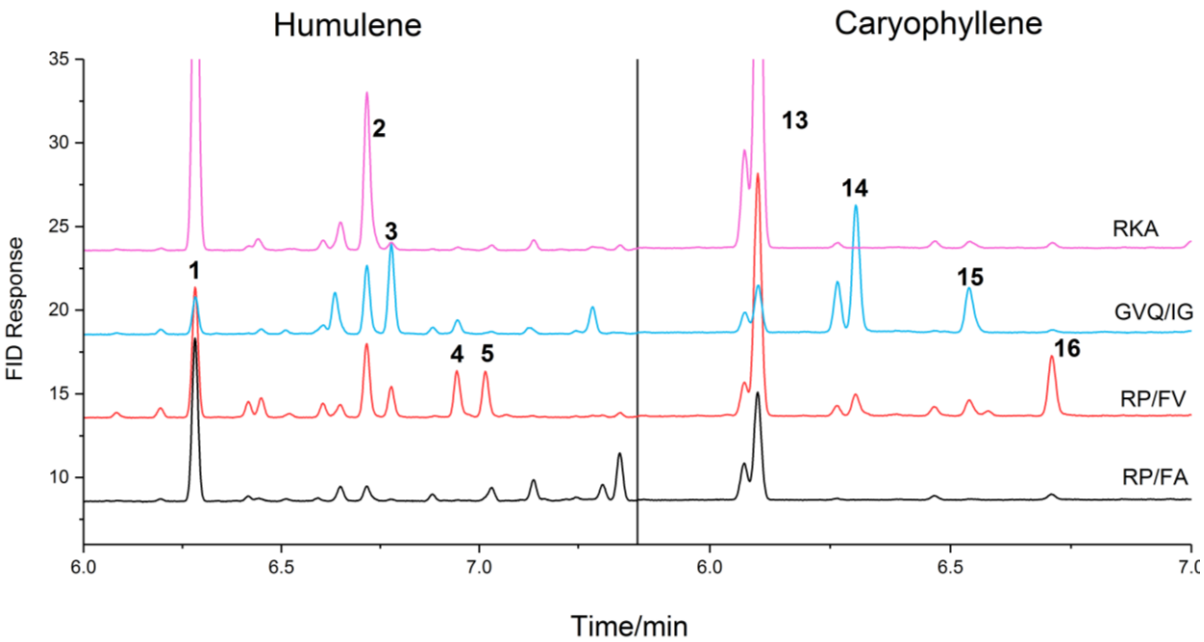



**Fig. 3.13** Selectivity comparison of variants with F87A and F87V mutations for the oxidation of (**A)** humulene and (**B)** Caryophyllene.

In summary, this set of screening results for just a small panel of 24 variants indicated that the residues F87, A328 and I263 have significant effects on product selectivity and should be considered as target residues for engineering and directed evolution. Further research on the oxidation of humulene and caryophyllene with an expanded variants library is being pursued by other members of the group.

### 3.2.2 Activity screening and product analysis

Details of experiments in this chapter were described in Chapter 2. Many variants had to be prepared; the procedure included mutagenesis via PCR, DNA plasmid transformation and cell culture incubation, enzyme production and purification, following by the $Fe^{II}(CO)$ assay to determine the enzyme concentration. The protocol for *in vitro* substrate oxidation screening in the 24-well plate was also described in Chapter 2, including screening reaction conditions and experimental set up, extraction of organics, GC analysis methods, scale-up reaction conditions and the work-up process. The characterisation of sesquiterpene oxidation products by NMR and MS data will be presented in this chapter.

## 3.4 Valencene Screening Results

Valencene was screened with the 24 selected $P450_{BM3}$ variants collection, and 10 products were detected from GC analysis (Fig. 3.14). Of these, **21** is β-nootkatol, **22** is α-nootkatol, **23** is nootkatone and **29** is the futher oxidation product, nootkatone-13,14-oxide (Fig. 3.15).[57] Mutations at F87 and A328 had been shown to have important effects on C2 oxidation of valencene.[82,110]

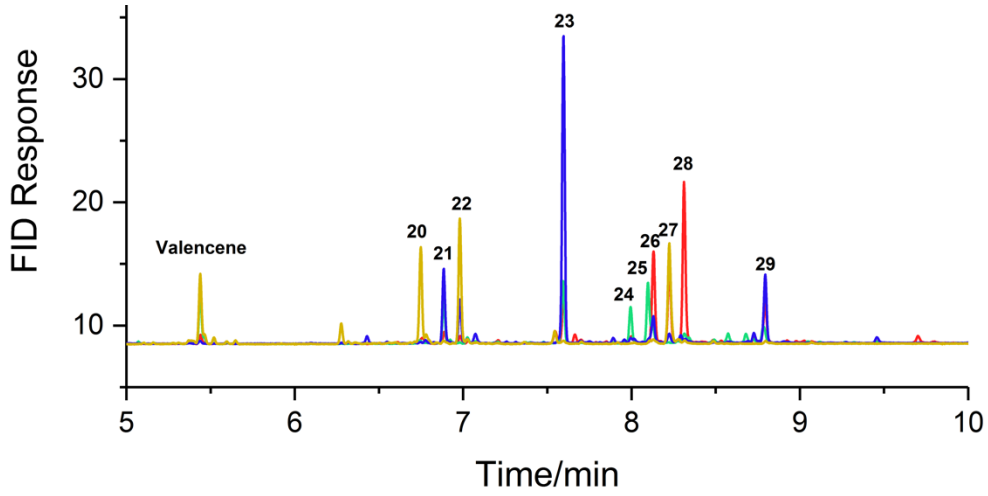


**Fig. 3.14** GC profile of selected variants showing selectivity of valencene oxidation.

Among variants which gave over 60% valencene conversion, RKA was the best candidate since it converted >99% of valencene, with 19% C2 alcohol and 52% of the corresponding ketone, the target product nootkatone. The total yield of C2 mono-oxidation product of 71% made RKA a good candidate variant for the production of nootkatone from valencene (Fig. 3.15).

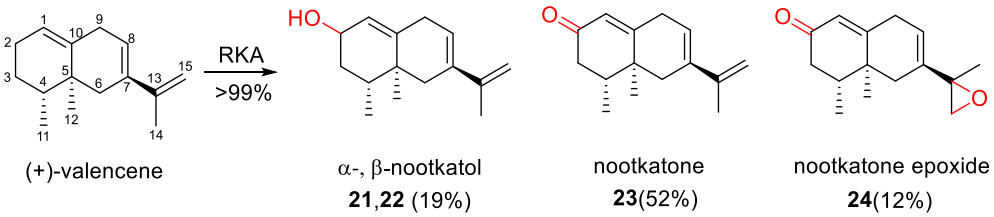
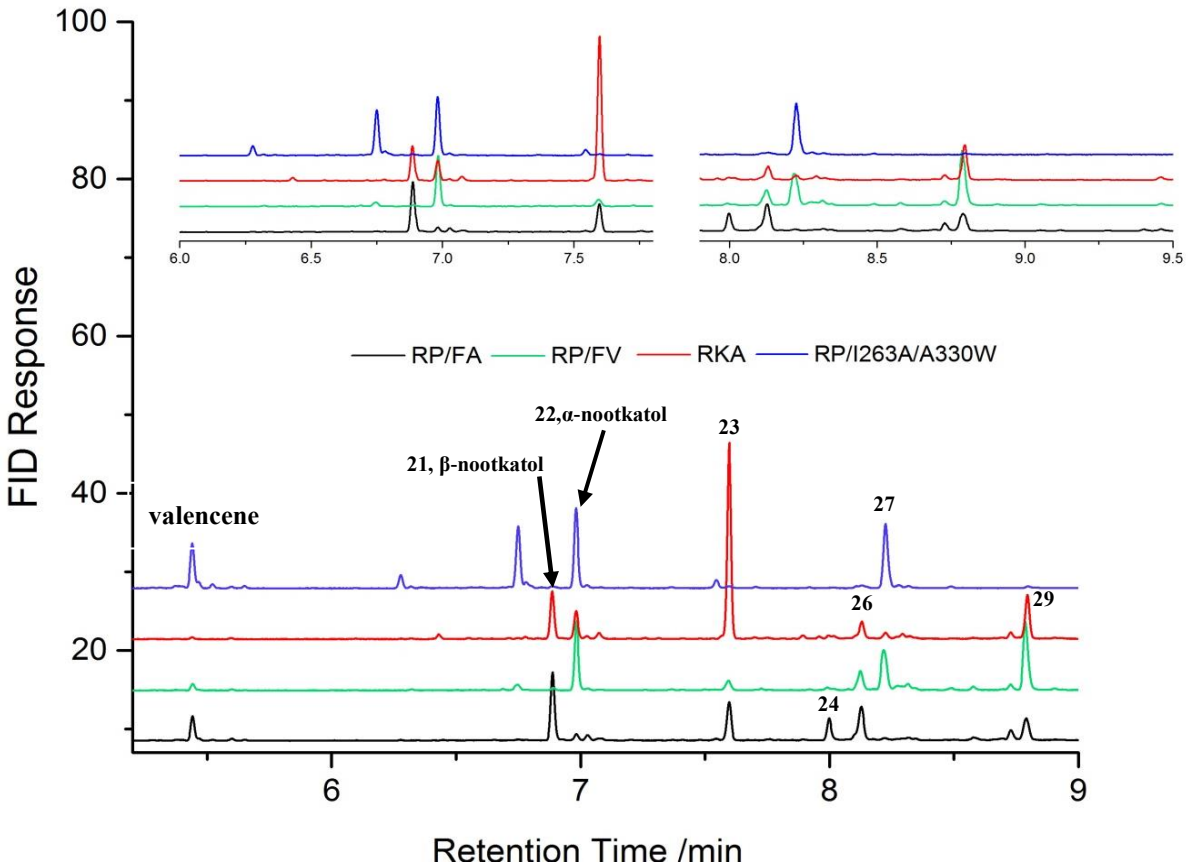



**Fig. 3.15** Valencene oxidation profile of the RKA variant. There are 15% of total products are other oxidation productsen.

Other mutants also gave interesting information on selectivity shifts (Fig. 3.16). Mutation of the F87 residue showed significant effects on manipulating the stereoselectivity for α- and β-nootkatol formation, while RP/FA with the F87A



mutation gave higher conversion against RP/FV with mutation F87V. RP/I263A/A330W gave only β-nootkatol, a further oxidation product **27** and a new product **20**, with a shorter retention time than the nootkatol, suggesting a new mono-oxygenated product of valencene.

**Fig. 3.16** Selected GC profile of valencene oxidation by $P450_{BM3}$ variants.

The effect of component mutations in RKA was investigated. The variant RLYF/K19 (abbreviated as R19) was used as the base variant and various mutations at A328 were introduced to generate $P450_{BM3}$ variants with F87 unsubstituted. In another experiment, RLYF was removed to study its effect, and I263A was introduced according to the result of increasing activity of I263G in GVQ/IG from screening experiments. Finally, R19/F87A with different A328 mutations were tested to find the best variant for further scale up process work. The results of GC analysis are presented in Fig. 3.17.

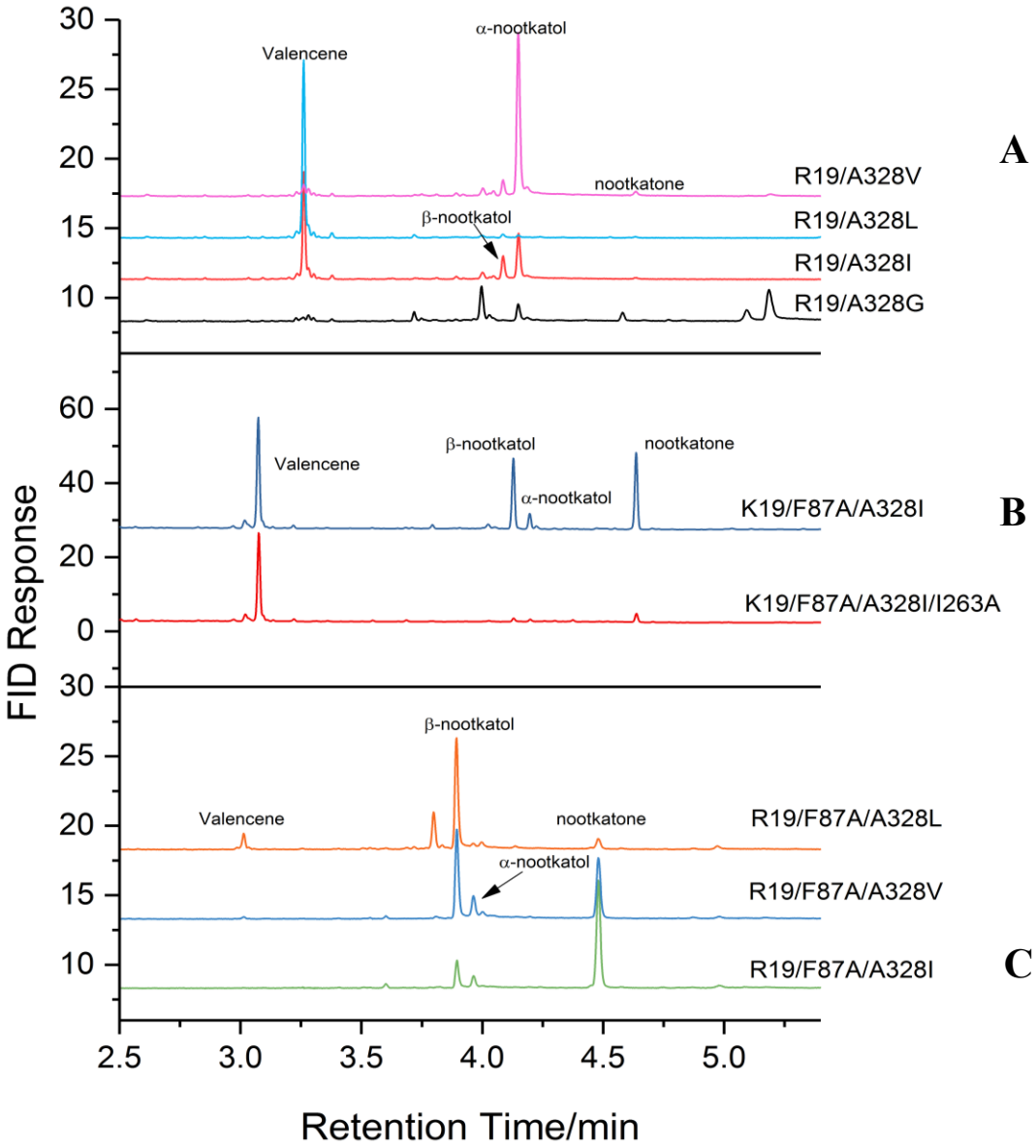



**Fig. 3.17** Mutation study of C2 oxidation of valencene to form nootkatone. **A**. Screening results of variants with different A328 mutations without F87A. **B**. Screening results of variants with F87A and A328I mutations without RLYF, I263A mutation was also tested. **C**. Screening results of variants with RLYF and F87A mutations but different amino acid substitution of A328.

Among variants carrying different A328 mutations based on R19, the A328G mutation was very active in promoting valencene oxidation to further oxidation products with longer retention times (data not shown). A328L mutation barely gave any activity while A328V and A328I showed activity to give nootkatol but no nootkatone (Fig. 3.17 A). Removing the RLYF double mutations from RKA caused a

drop in activity for K19/F87A/A328I but the C2 oxidation selectivity remained. I263G in GVQ/IG increased activity and gave more further oxidation products. Hence, the smaller side chain volume mutation of I263A was added to K19/F87A/A328I in order to increase the activity. However, this failed as the activity was shut down when introducing I263A into K19/F87A/A328I (Fig. 3.17 B). As RLYF and F87A mutations were necessary for C2 oxidation of valencene, various mutations of A328 were tested based on R19/F87A base variants (Fig. 3.17 C). R19/FA/A328L gave β-nootkatol as the major product but very little nootkatone. A328I and A328V gave very similar activity (>99% conversion) and selectivity pattern. However, A328I on R19/F87A gave higher nootkatone yield (52%) compared to A328V (43%). Therefore, RLYF/K19/F87A/A328I (RKA) was selected as the biocatalyst for further scale up research towards the synthesis of nootkatone from valencene oxidation.

# Chapter 4

# Cedrene Oxidation: Generating Diversity

## 4.1. α-Cedrene derivatives and the woody note

The sesquiterpenes α- and β-cedrene are the precursors to a family of cedranoid compounds found in cedarwood oil (Fig. 4.1) that have found applications such as food preservatives, disinfectants and fragrance components.[111-113]

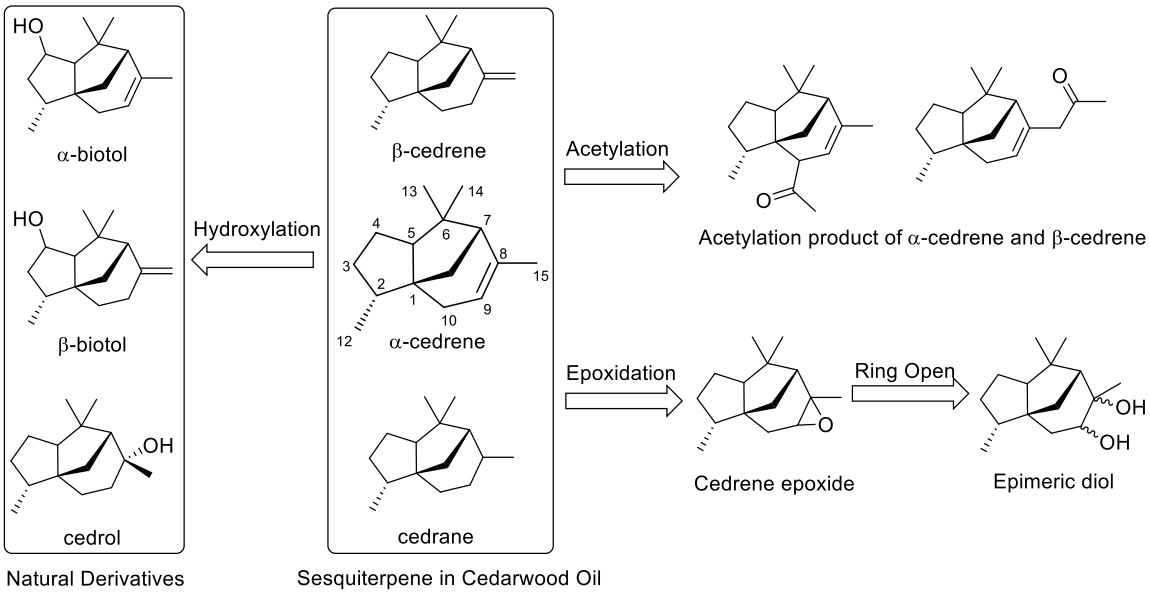



**Fig. 4.1** Main cedranoid compounds in cedar oil and acetalyation product of α-cedrene.

Cedarwood oil is widely used to impart a "warm woody note". As the main sesquiterpene component in cedarwood oil, α-cedrene and its derivatives, with their tricyclic structures, have been a popular target in total synthesis and functionalisation studies. Cedrol, the C8,C9 double bond hydration product of α-cedrene (Fig. 4.1), and biotol (4-hydroxy-α-cedrene) were isolated from the alcoholic fraction of cedarwood oil.[114] Cedrol is produced industrially; the methyl ether and the acetate have found use as woody notes. [115]

Chemical functionalisation studies showed the expected reactivity at the activated C8,C9 double bond and the allylic C10 and C15 positions. Chalchat *et al*. reported the hydroboration of α-cedrene and cedrol was obtained in 15% yield.[115] Qin *et al*. demonstrated selective primary allylic C–H bond activation at the C15 methyl group using dirhodium catalysts without a directing group on the substrate.[116] Kitchens *et al*. reported the acetylation of the hydrocarbon fraction from cedarwood oil and isolated 9-acetyl-α-cedrene which possessed a "warm woody smell".[117]

Total chemical syntheses of oxygenated cedrenes have been reported. Green *et al*. prepared α-pipitzol from oxidation and dearomatization of curcuphenol followed by a cycloaddition step (Fig. 4.2).[118] Similarly, Yates *et al*. used photocatalysed rearrangement of bicycle[2.2.2]octenone to prepare 4- and 12-hydroxy-α-cedrene. Each route contained more than 10 steps.[119,120]

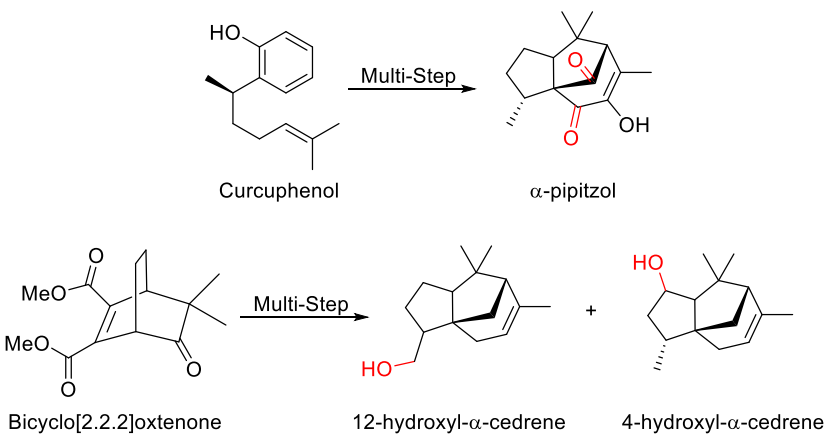



**Fig. 4.2** Chemical synthesis of oxygenated α-cedrene derivatives.

Functionalisation by microbial biotransformation has been widely studied. Lamare *et al*. reported the biohydroxylation of α-cedrene by the fungus *Beauveria sulfurescens* to the C3,C15-diol and C12,15-diols (Fig. 4.3).[121] 15-Hydroxy-α-cedrene

was shown to be the initial product which was further oxidised at C3 and C12. Similar results were reported by Abraham *et al*. for *Corynespora cassiicola* which was selected from screening a panel of 45 organisms.[122] Three products were characterised as diols and triols from C3, C10, C12 and C15 oxidation, probably via initial oxidation at the C15 allylic methyl group since cedrane was not oxidised. Takigawa *et al*. reported the allylic oxidation at C10 to form 10-hydroxy-α-cedrene (*sec*-cedrenol) and the corresponding ketone by *Rhodococcus* strain KSM-7358.[123]

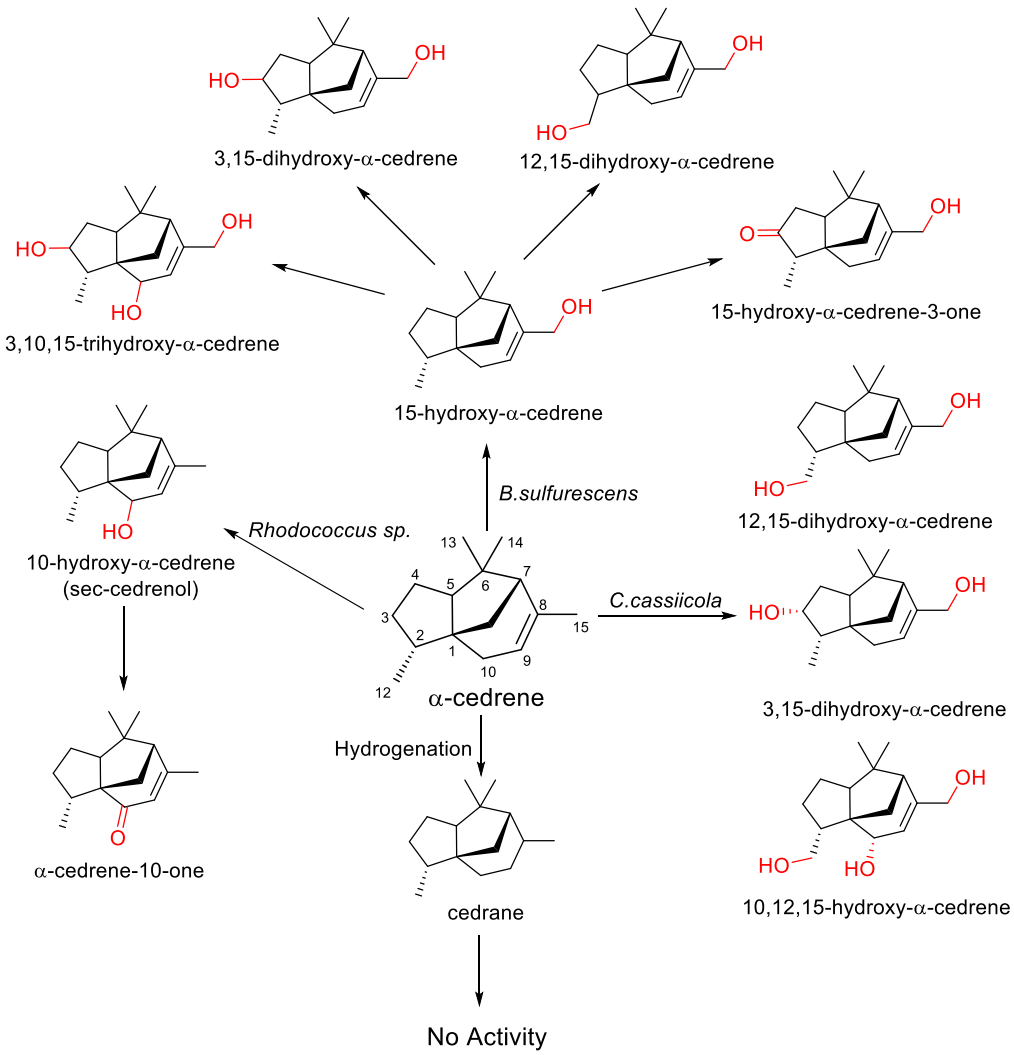



**Fig. 4.3** Functionisation of α-cedrene from microbial biotransformation.

With its complex tricyclic structure and the known woody aroma properties of its alcohol, ether and ester derivatives, α-cedrene was selected as a target substrate for oxidation by $P450_{BM3}$ variants. The aim was to generate not only known compounds but especially new products that may have desirable aromas or precursors to derivatives with such properties. α-Cedrene was screened for oxidation by a panel of $P450_{BM3}$ variants selected from the existing enzyme library on the basis of a broad coverage of active site topology variations. The oxidation products would be purified from preparative scale reactions and characterised in order to identify residues and mutation combinations that promoted cedrene conversion and altered product selectivity. New libraries would then be constructed by combining multiple mutations at two or three residues to investigate the potential to change the regioselectivity of cedrene oxidation.

## 4.2. Initial Screening of α-Cedrene

The panel of 24 $P450_{BM3}$ enzyme variants from Chapter 3 (Table 3.1) was screened for α-cedrene oxidation in 24-well plates following the conditions presented in Chapters 2 and 3. The enzymes included the WT, the four base variants RP, KSK19, GVQ and KT2 as well as their derivatives that contained additional mutations at residues located at different substrate recognition sites around the active site: S72W, A82M, F87A or F87V, I263G, E267V, and A330W.

Of the 24 tested variants, 9 showed >50% conversion (TON >250). The enzyme panel gave six major products (**30**, **33**, **36**, **37**, **44** and **45**, Fig. 4.4) along with 10 minor ones. The RKA variant was used in preparative scale reactions to isolate **30**,

**33**, **36** and **44**; RPFV was used for **45,** and GVQ/I263G for **37** (Fig. 4.5). Characterisation by NMR and MS data showed that the main oxidation sites were the C8,C9 double bond and the allylic C10 and C15 positions, giving the epoxide **30**, the C10 alcohol **33**, the C15 alcohol **36**, and the diols **44** and **45** from further oxidation.

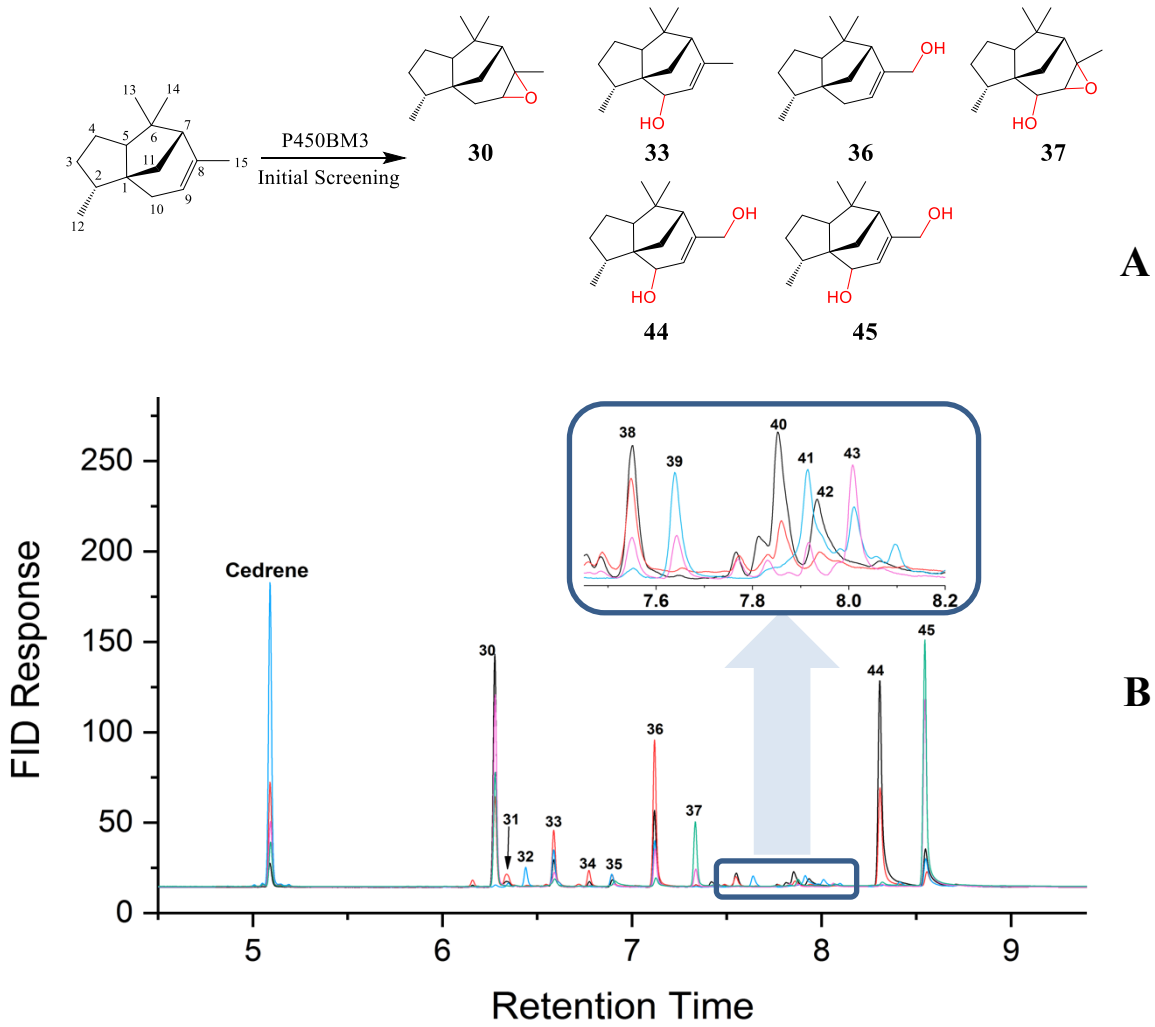



**Fig. 4.4 A**. Cedrene oxidation products characterised from initial screening and preparative scale reactions. **B**. The cedrene oxidation product profile for $P450_{BM3}$ variants.

WT $P450_{BM3}$ showed only 6% conversion (TON = 30) of the unnatural substrate cedrene and gave a small amount of the epoxide, **30**. The C10 allylic alcohol **33** was the major product (66%) and there was a third, unidentified product at 6.5 min (Table 4.1, Fig. 4.5). The RP and KT2 base variants did not contain active site mutations. RP showed higher activity than the WT, with 14% cedrene conversion. Unlike for the WT, however, RP gave more of the epoxide **30** (36%) and less of the C10 alcohol **33** (57%). KT2 was less active, though still more active than the WT with 6% conversion. Unexpectedly, the C15 alcohol **36** was the major product (75%) together with 25% of the C10 alcohol **33**.

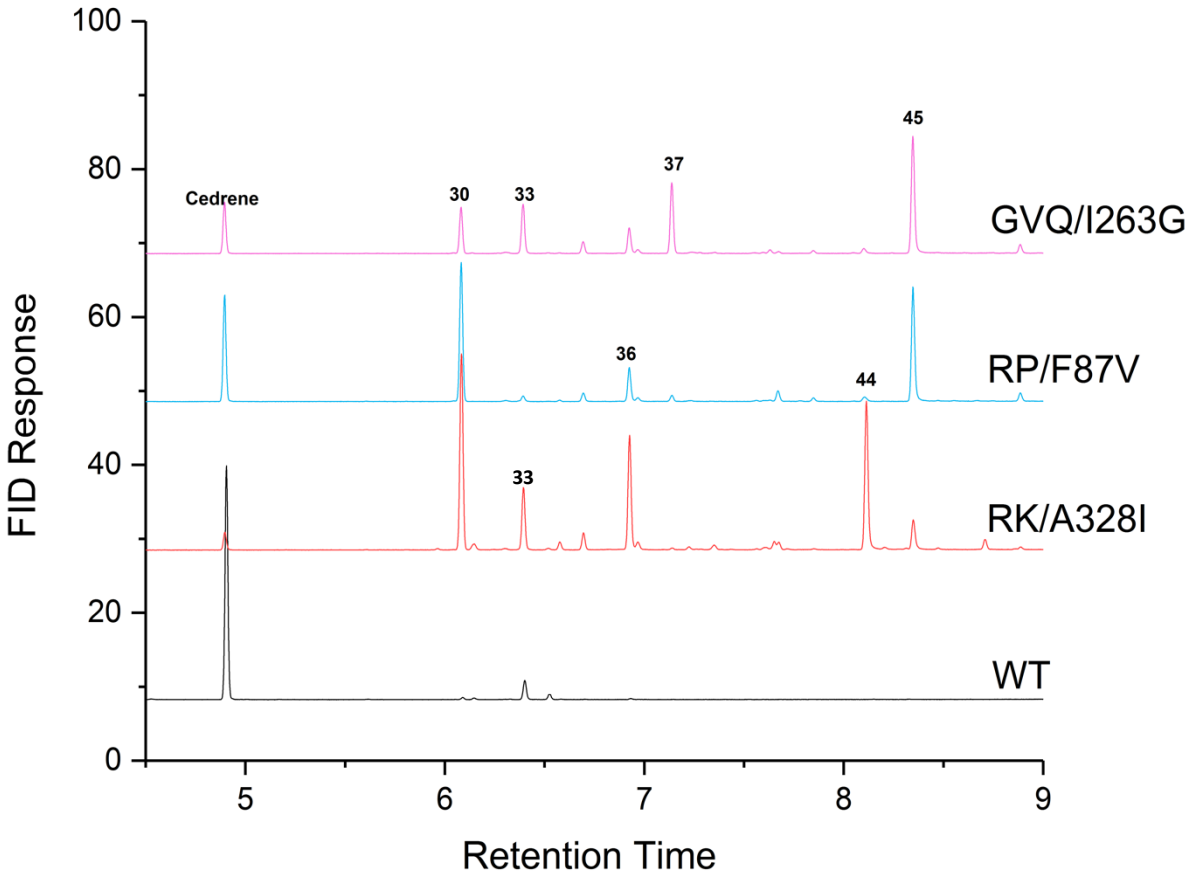
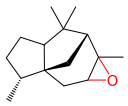
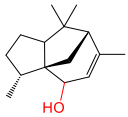
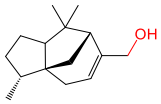



**Fig. 4.5** WT $P450_{BM3}$ compared with 3 other variants selected for preparative reactions to characterize the major products **30, 33, 36, 37, 44** and **45.**

**Table 4.1** Product distribution of WT, RP and KT2 variants

| Variant | Mutation | Conv. |  **30** |  **33** |  **36** |
|---|---|---|---|---|---|
| WT | None | 6% | 19% | 66% | |
| RP | R47L/Y51F/I401P | 14% | 30% | 57% | |
| KT2 | A191T/N239H/I259V/ A276T/L353I | 12% | | 25% | 75% |

The allylic alcohol 10-hydroxy-cedrene (*sec*-cedrenol), **33**, was identified by the low-field allylic C<u>H</u>OH proton (3.80 ppm) and its coupling to the alkene proton H9 at 5.36 ppm. The primary allylic alcohol **36** was identified by the AB-quartet resonance at 4.03 ppm assignable to the allylic C<u>H</u>$_2$OH protons (Fig. 4.6) and by the MS data which showed a molecular weight of M+16. The two products **44** and **45** at longer retention times were identified as the diastereomeric 10,15-diols. Both had a molecular weight of M+32, indicating di-oxygenation, and two resonances in the $^1$H NMR spectrum of 2:1 intensity ratio assignable to allylic alcohol protons. The resonance of intensity two in both spectra were AB-quartets assigned to the C15 alcohol protons. The stereochemistry at C10 was established by 1D nOe spectra by the presence or otherwise of an nOe between the C10 and C11 protons.

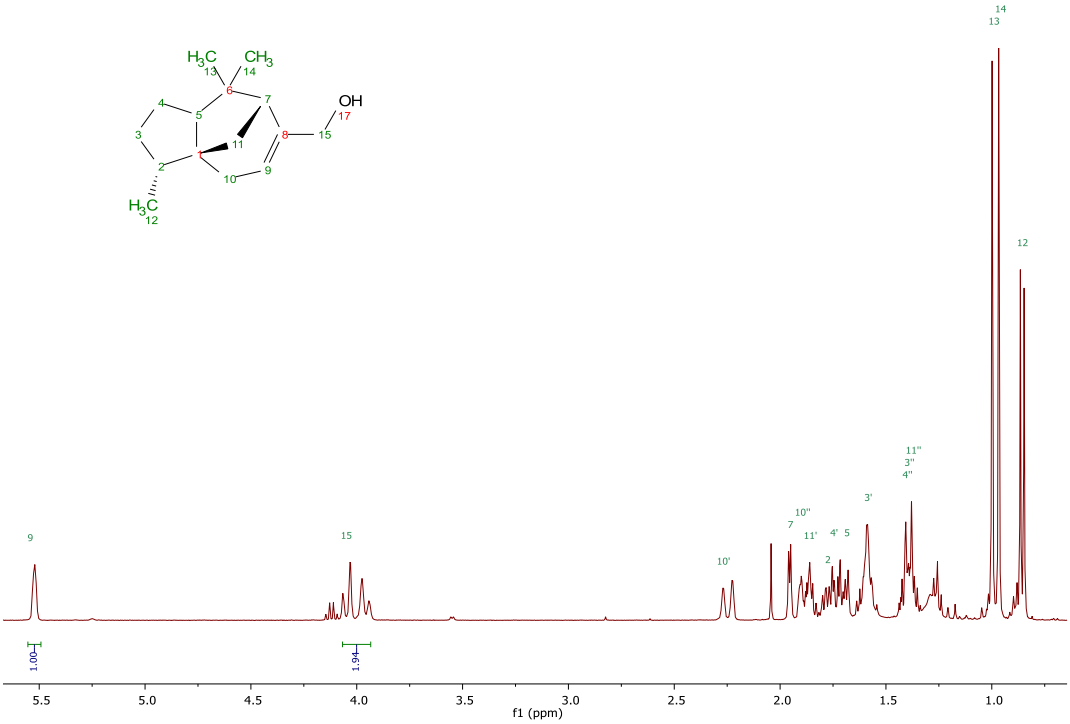



**Fig. 4.6** The $^{1}H$ NMR spectrum in $CDCl_3$ of the allylic alcohol 15-hydroxy-cedrene, **36.**

Mutations at F87 in the RP base variant to residues with smaller side chains, forming the RP/FA (F87A) and RP/FV (F87V) variants, dramatically increased the catalytic activity and altered product selectivity. RP/FA gave 73% conversion with the major products being **33** (25%) and **45** (23%), with significant amounts of **30** (11%) (Fig. 4.7). There were many products with retentions times (7.2 – 8.0 min) between those of the C15 allylic alcohol **36** and the diol **44**, that constituted 29% of the total products. The RP/F87V variant showed similar activity (77%) as the RP/FA but preferentially produced the epoxide **30** rather than the C10 allylic alcohol **33**.

The GVQ variant showed 92% conversion. As with the RP/FV variant, it formed mainly the epoxide **30** (22%) and the 10β,15-diol **45** (54%), with little of the C10 alcohol **33**, the C15 alcohol **36** or the 10α,15-diol **44** (Fig. 4.7). It appeared that there was more extensive further oxidation of **33** and **36** to form the diol **45.**

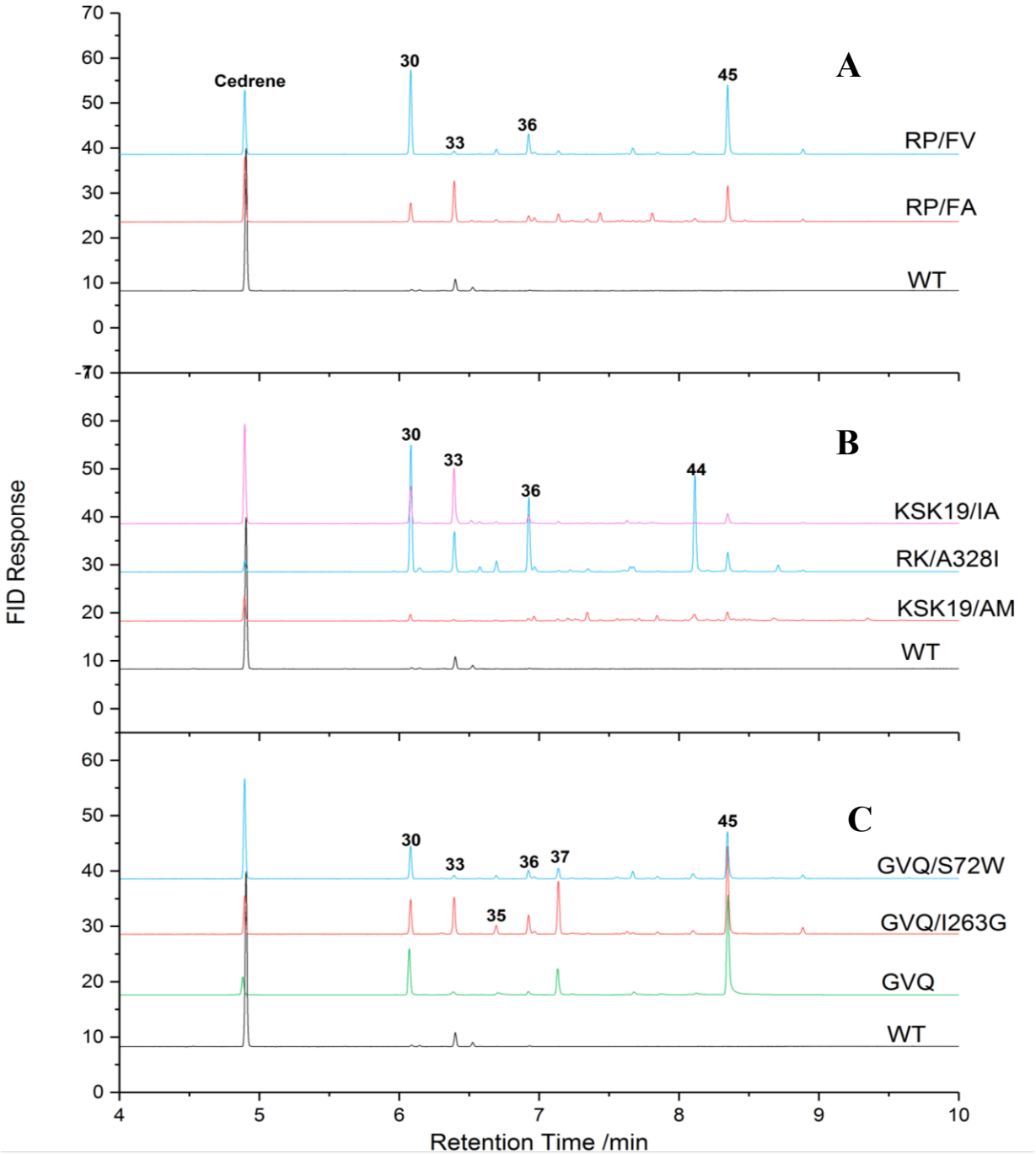



**Fig. 4.7** Product profile (GC) of variants containing F87 mutations (**A**), KSK19-based variants (**B**), and GVQ-based variants (**C**).

Another product, **37** (14%), was purified from a preparative scale reaction with GVQ/I263G and characterised as the di-oxygenation product arising from epoxidation and C10 oxidation. In the $^{1}$H NMR spectrum (Fig. 4.8), the absence of the low-field H9 proton indicated epoxidation of the C8,C9 double bond, with the epoxide proton at

2.40 ppm. The broad CHOH doublet at 3.67 ppm was assigned to H10 on the basis of its coupling to the epoxide proton H9. Its higher field position relative to the H10 resonance in the C10 allylic alcohol **33** further supported the assignment. The MS data of M+32 also supported the assignment of **37** as 10-hydroxy-cedrene oxide. Regardless of whether this was formed from the epoxide **30** or the C10 alcohol **33**, the formation of **37** was consistent with more extensive further oxidation for this mutant.

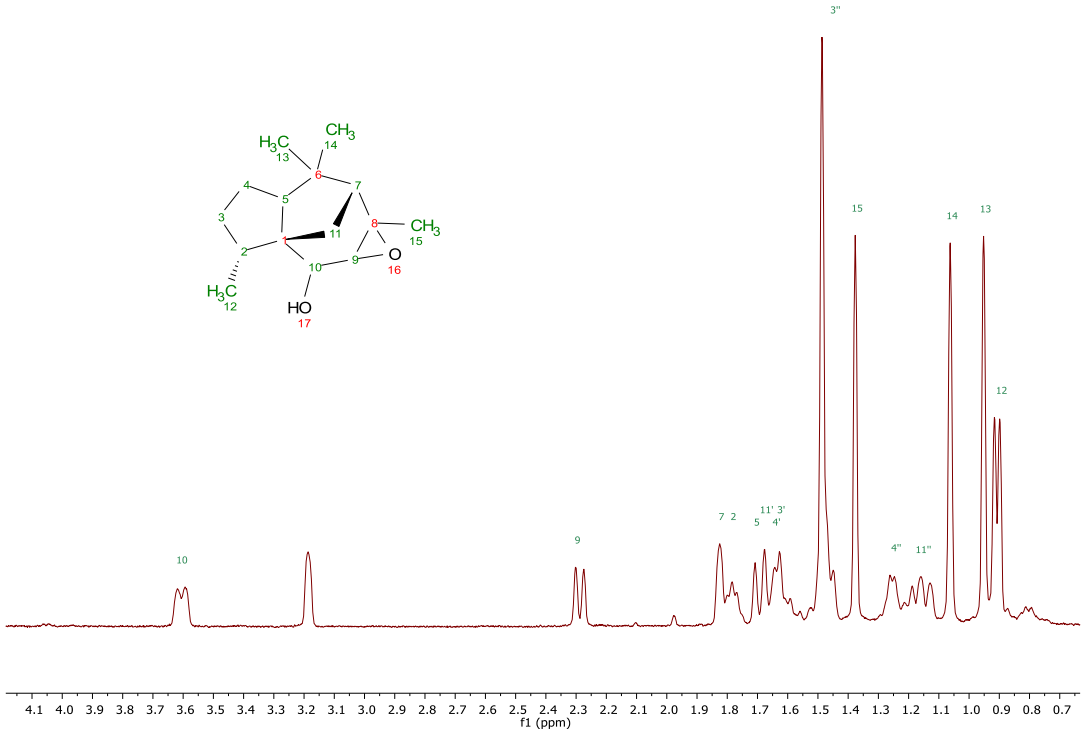



**Fig. 4.8** The $^{1}$H NMR spectrum of **37** in $CDCl_3$.

Among the three KSK19-based variants, RK/A328I which contained the R47L/Y51F couplet and the A328I mutation, gave 96% conversion (Fig 4.7B). The major products were **30**, **33**, **36**, and the 10α,15-diol **44** (35%) with much less of the 10β,15-diol **45** (5%). This variant was the only one in this initial screen to form **44** preferentially over **45**. The other two variants, KSK19/AM and KSK19/IA, gave

conversion of 81% and 59% respectively. Both the A82M and I263A mutations suppressed formation of the diols **44** and **45**. In the case of KSK19/IA this led to the C10 alcohol **33** (43%) being the major product. The A82M mutation gave many products within the 7.2–8.0 min retention time window between **36** and **44**. However, none of these could be purified for characterisation.

Introducing mutations at I263 and S72 to GVQ (Fig. 4.7C) increased activity to produce the epoxide-10-ol **37** and the 10β,15-diol **45** in GVQ/IG (98% conversion), but S72W retarded the reaction and did not give new products.

### 4.2.2 Summary

WT $P450_{BM3}$ barely showed any activity with cedrene but the four base variants and especially their F87A and F87V derivatives RP/FA, RKA (with A328I) and GVQ/IG (with I263G), showed high activity and some product diversity. Six major products were isolated and found to arise from oxidation at the C8,C9 double bond and the two allylic positions C10 and C15, as well as further oxidation products (the 10,15-diols **44** and **45**, and 10-hydroxy-cedrene oxide **37**). Their predominance is understandable as these are the most activated carbons positions in cedrene. However, other product peaks were also observed in the GC, some of these might arise from oxidation at other, less activated C–H bonds. Hence, to explore the possibility of oxidising every C–H bond in cedrene, further variants were constructed using KSK19 and GVQ as templates and by introducing other amino acid substitutions at F87, A328, A330 and I263.

## 4.3. Construction of new libraries of variants

### 4.3.1 Variants with F87, A328 and A330 substitutions

The RKA variant contained the activity-promoting mutation F87A and the A328I mutation that altered product selectivity. Hence, A328 and the close-by residue A330 (Fig. 1.9) were selected for hydrophobic substitutions in RKA. Glycine (G), valine (V), isoleucine (I), leucine (L), phenylalanine (F), and methionine (M) were substituted for A328. Proline (P), serine (S) and tryptophan (W) were substituted for A330. RKA is the abbreviation for RLYF/KSK19/A328I and contained the mutations R47L/Y51F/H171L/Q307H/N319Y (R19), F87A and A328I. Henceforth, variants like RKA are renamed R19/F87A/A328I.

**Table 4.2** List of $P450_{BM3}$ variants with mutations at A328 and A330 added to R19/FA.

| VARIANT | MUTATIONS |
|---|---|
| **R19/FA/AI (RK/A328I)** | R47L/Y51F/H171L/Q307H/N319Y/F87A/A328I |
| **R19/FA/AF** | R47L/Y51F/H171L/Q307H/N319Y/ F87A/A328F |
| **R19/FA/AG** | R47L/Y51F/H171L/Q307H/N319Y/ F87A/A328G |
| **R19/FA/AL** | R47L/Y51F/H171L/Q307H/N319Y/ F87A/A328L |
| **R19/FA/AM** | R47L/Y51F/H171L/Q307H/N319Y/ F87A/A328M |
| **R19/FA/AV** | R47L/Y51F/H171L/Q307H/N319Y/ F87A/A328V |
| **R19/FA/A330P** | R47L/Y51F/H171L/Q307H/N319Y/ F87A/A330P |
| **R19/FA/A330S** | R47L/Y51F/H171L/Q307H/N319Y/ F87A/A330S |
| **R19/FA/A330W** | R47L/Y51F/H171L/Q307H/N319Y/ F87A/A330W |

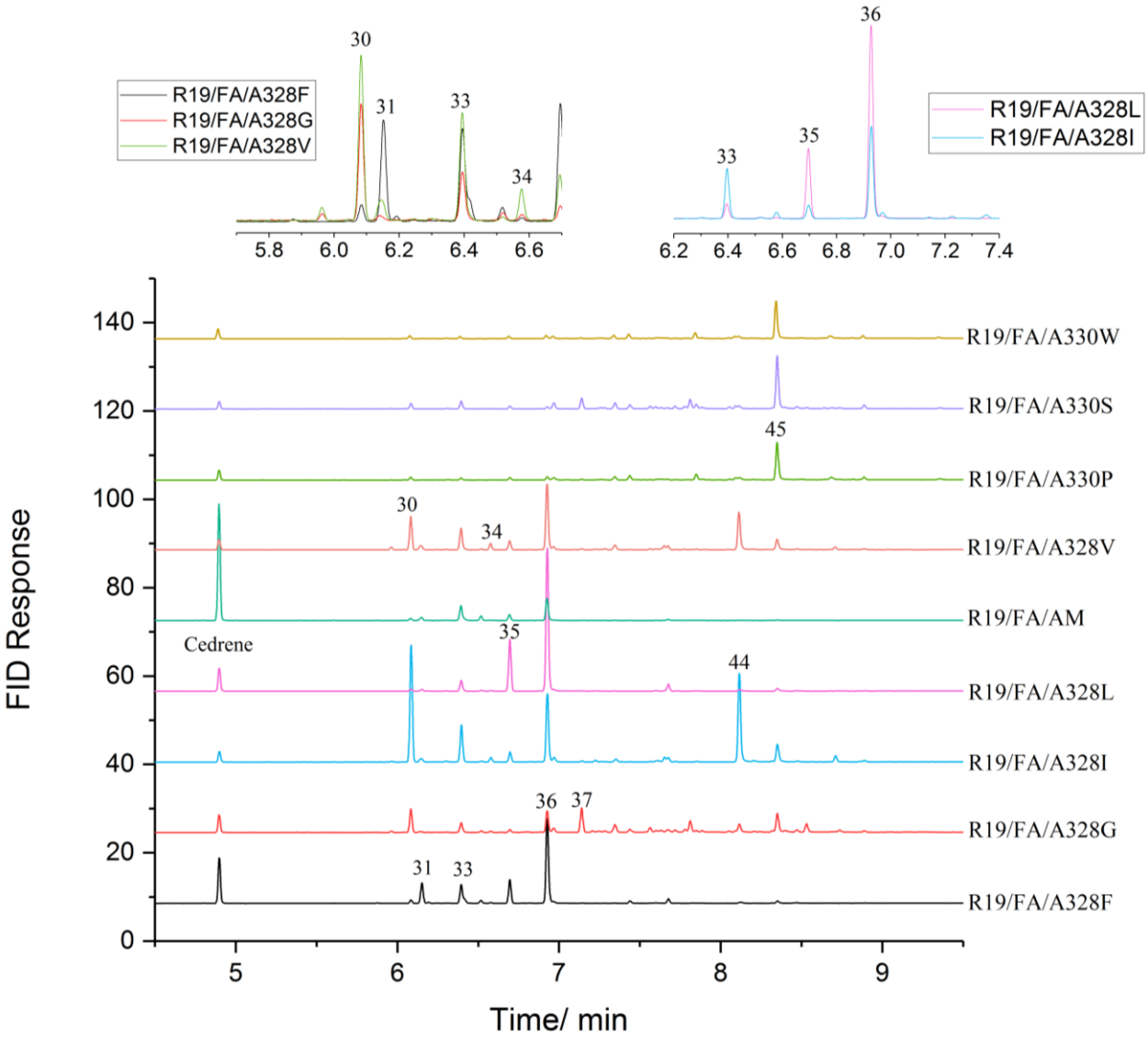



**Fig. 4.9** Product profile (GC) of variants containing A328 and A330 mutations based on R19/F87A.

All of these variants were successfully expressed. R19/FA showed similar conversion and product profile to RP/FA, as expected. The selectivity for two new products, **31** and **35**, was improved in variants R19/FA/A328L and R19/FA/A328F. All A330 mutations gave the 10β,15-diol **45** as the predominant product**.**

Compared to R19/FA/AI (RKA, in blue, Fig. 4.9), the R19/FA/AF variant suppressed epoxidation and formed a new product **31** (10%) which was purified from a preparative scale reaction and characterised as the tertiary alcohol arising from C2

oxidation (Fig. 4.10). H9 at 5.26 ppm showed clear coupling with the two H10 protons, while the presence of C13, C14 and C15 methyl resonances confirmed that no oxidation occured at the C8,C9 double bond, C10 or the methyl groups. The lower field chemical shift of the C12 methyl protons, and the loss of coupling to H2 from the COSY spectrum confirmed C2 hydroxylation to form 2-hydroxy-α-cedrene, **31**.

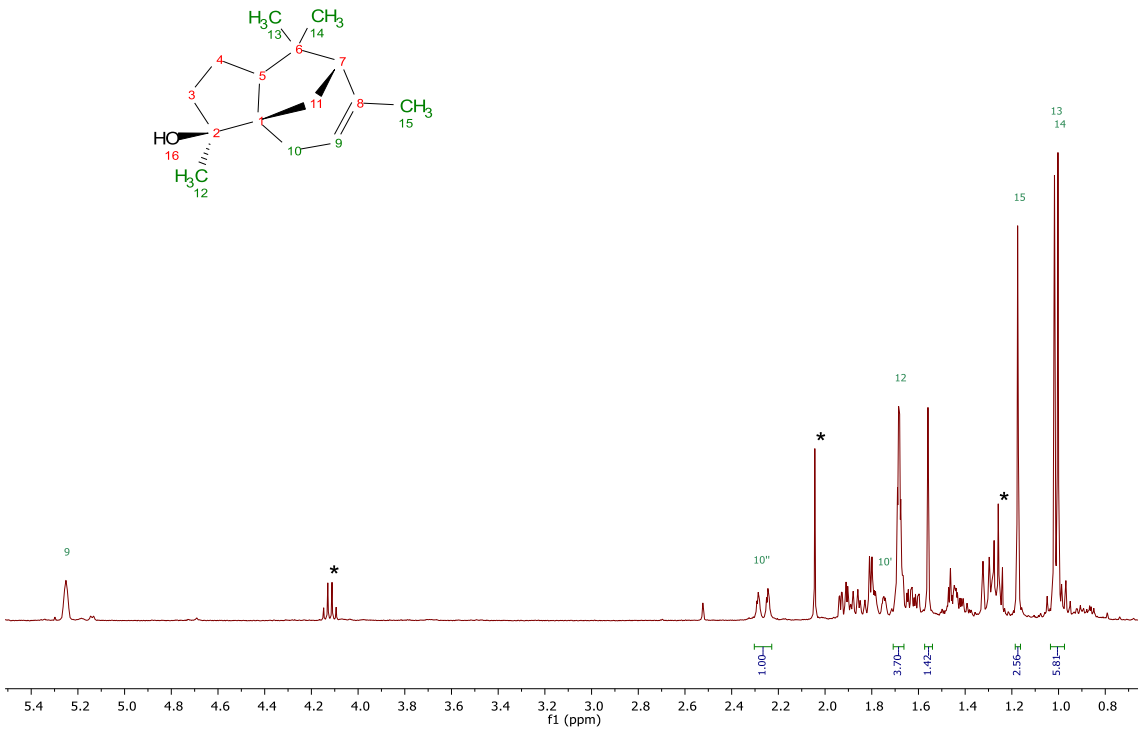



**Fig. 4.10** The $^{1}$H NMR spectrum of **31** in $CDCl_3$**.** Astensus demote resonaces due to EtoAc.

The R19/FA/AL also suppressed epoxide formation and gave as major products the C15 alcohol **36** (63%) and aldehyde **35** (23%), giving a total regioselectivity for C15 oxidation of 86%. Product **35** showed a low-field singlet (9.44 ppm) characteristic of an aldehyde while the resonance for the allylic C15-methyl group in the substrate was missing from the spectrum (Fig. 4.11). The NMR data were fully consistent with the α,β-unsaturated aldehyde, formed via further oxidation of the C15 alcohol **36**.

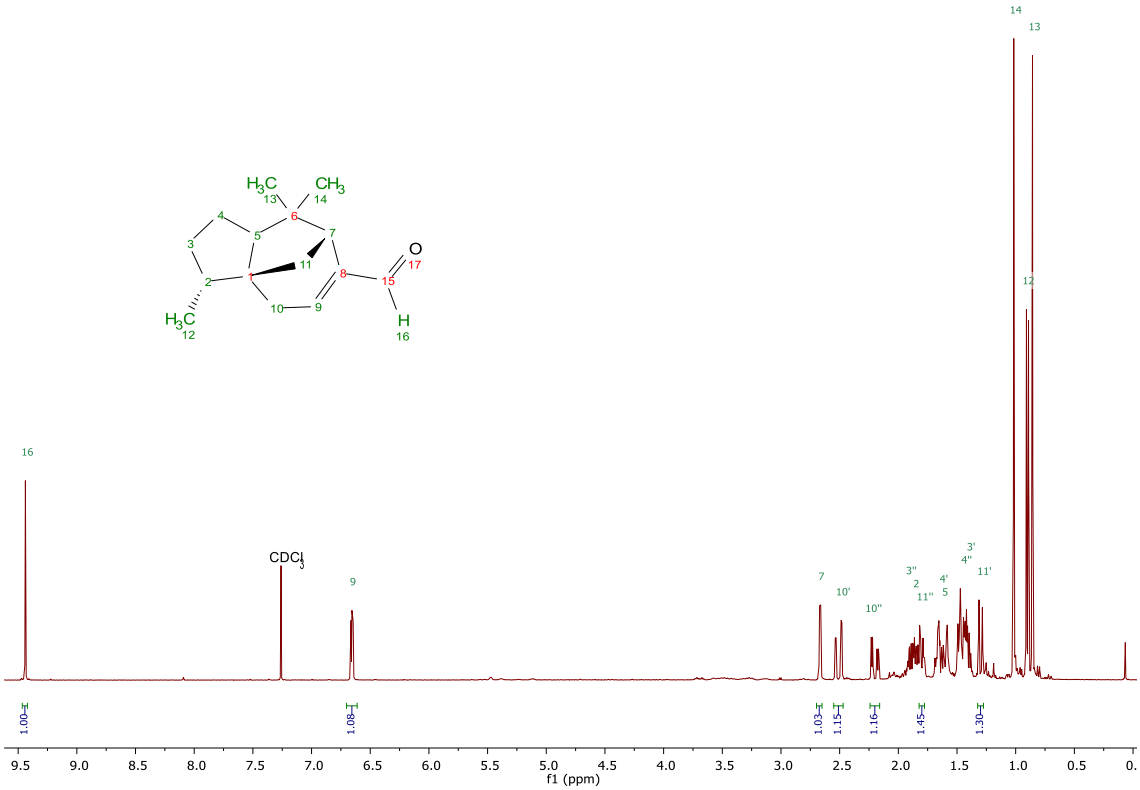



**Fig. 4.11** The $^{1}$H NMR spectrum of **35** in $CDCl_3$.

Like the R19/FA/AI variant, R19/FA/AV also formed the epoxide **30** and the C10 and C15 hydroxylation products **33**, **36**, **44** and **32** but there was a minor product, **34** (3%), with a retention time of 6.57 min. This was isolated and characterised as 9-hydroxy-β-cedrene, with an exocyclic double bond.

The resonance of the C15 methyl protons was absent from the spectrum. Two alkene protons were assigned to the $CH_2$ at C15 since theses couple to each other and to H9 4.27 ppm. H9 was coupled to the two C10 protons from the COSY spectrum. Given that there was <1% β-cedrene in the starting material (purity ≥95% of α-cedrene) this compound is likely formed *via* rearrangement of the primary allylic C15 radical of α-cedrene to the more stable secondary C9 radical of β-cedrene followed by radical rebound (Fig. 4.12). The fact that **34** was formed in smaller percentages with

other variants indicated longer lifetime of the C15 radical for this variant, probably as a result of greater mobility of the substrate so that the radical could move away from the $Fe^{IV}$–OH to avoid immediate radical rebound.

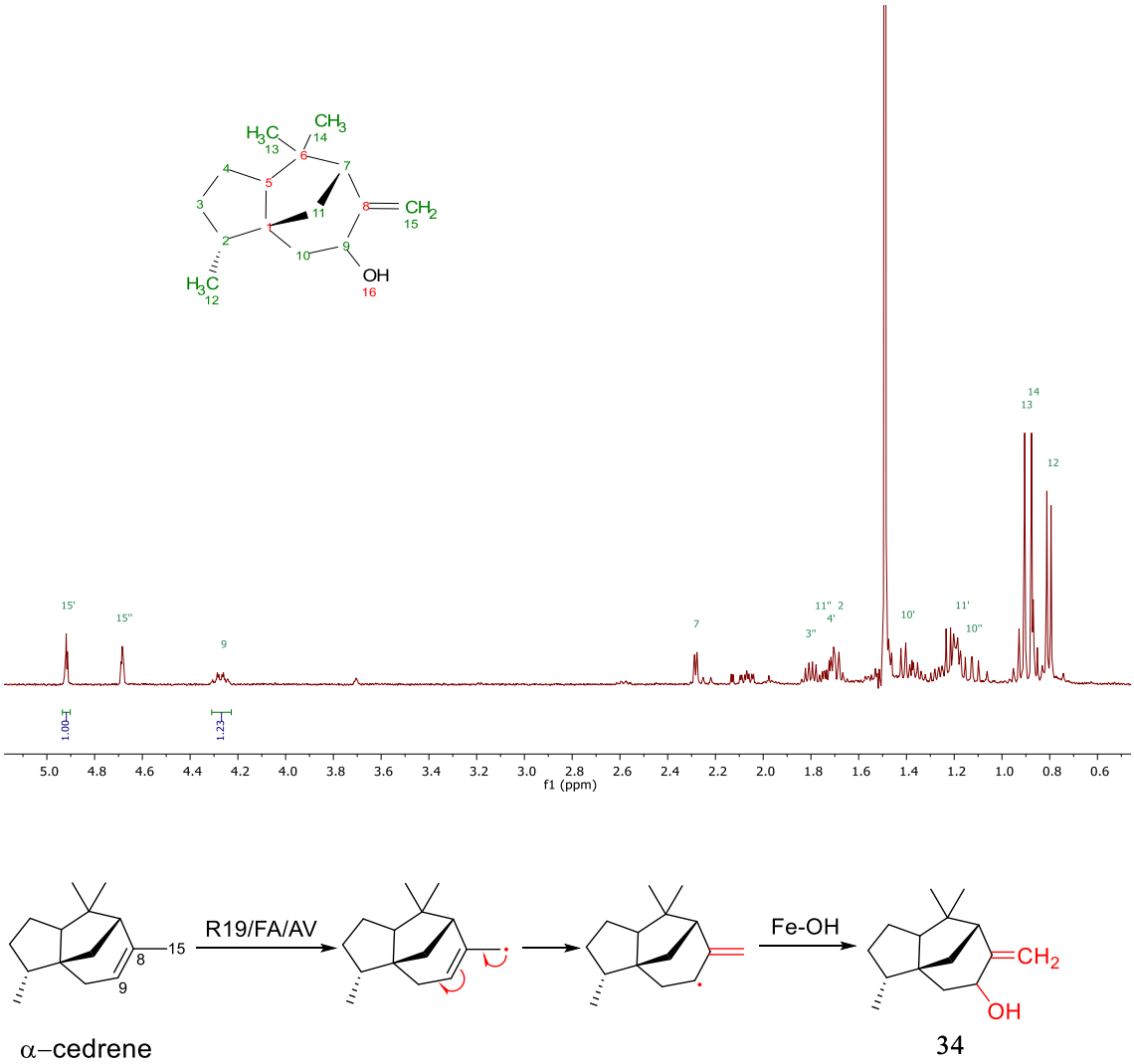



**Fig. 4.12** The $^{1}H$ NMR spectrum of **34** in $CDCl_3$ and the possible mechanism of its formation.

The R19/FA/AG gave many products with retention times between **37** and **44** but the yield was too low for pure samples to be obtained. Variants with A330 substitutions also gave many product peaks after **36** but the conversions were low and there were no major products that could be readily isolated.

Overall, the results suggested that a larger side chain at the 328 position could constrain a substrate to bind in a different orientation, leading to altered and often enhanced selectivity while multiple substitutions with smaller side chains tended to give more diverse products but the selectivity could be lower. Due to the strong effects of mutations of A328, further mutations were introduced at F87 in order to explore possible shifts in product selectivity.

### 4.3.2 A sub-library of F87/A328 variants

Based on the results of the initial screen and product characterisation, a two-residue combinatorial library was constructed by introducing five hydrophobic substitutions at both F87 and A328 of the R19 (R47L/Y51F/H171L/Q307H/N319Y) template. R19 was first generated by the A87F mutation on the R19/FA variant. A total of 36 variants was constructed via site-directed mutagenesis. Of these, 24 were successfully expressed; the exceptions were R19/FL/AV, R19/FL/AG, R19/FG/AI, R19/FG/AG, R19/FG/AV, R19/FG/AI and all variants containing R19/F87V.

The GVQ (A74G/F87V/L188Q) variant contained the F87V mutation and could be expressed at high levels. It was selected for completing the F87V–A328 two-residue library. In addition, due to the significantly different product profile of GVQ/IG in initial screening, the mutations I263G, I263A and I263V were introduced to the GVQ–A328 two-residue library variants to generate a three-residue library. Surprisingly, variants with the F87V/A328V mutations suffered from poor expression and were therefore not included in the final library used for testing.

#### 4.3.2.1 F87 variants containing the A328F mutation

In previous experiments, R19/FA/A328F and R19/FA/A328L were the only variants showing significant suppression of epoxidation. The R19/FA/AF was also the only variant that gave C2 oxidation, while the FA/AL combination favoured C15 oxidation. Since C2 is on the aliphatic 5-membered ring, the effect of combining A328F with other F87 mutations was explored. This mutation was added to R19 variants containing the F87I, F87L and F87G mutations (Fig. 4.13). As the side chain volume at residue 87 increased from glycine to isoleucine, there was a clear decrease in activity (Table 4.3).

**Table 4.3** Conversion of variants containing different F87 mutations combined with n A328F mutation.

| Variant Name | Mutations (except mutations in R19) | Conversion |
|---|---|---|
| R19/FA/AF | F87A/A328F | 89% |
| R19/FG/AF | F87G/A328F | 47% |
| R19/FI/AF | F87I/A328F | 15% |
| R19/FL/AF | F87L/A328F | 12% |

The R19/FG/AF variant gave another new product at a retention time (6.44 min) between the C2 alcohol **31** and the C10 alcohol **33**. This compound was eventually isolated and characterised as 4-hydroxy-α-cedrene (see section 4.4.1), another product from oxidation of an unactivated aliphatic C–H bond of the 5-membered ring. Moreover, despite the lower conversion than with R19/FA/AF, the R19/FG/AF variant also oxidised cedrene to products with retention times between the C15 alcohol **36** and the 10,15-diol **44**. This suggested that a smaller side chain the 87 position could diversify the range of products.

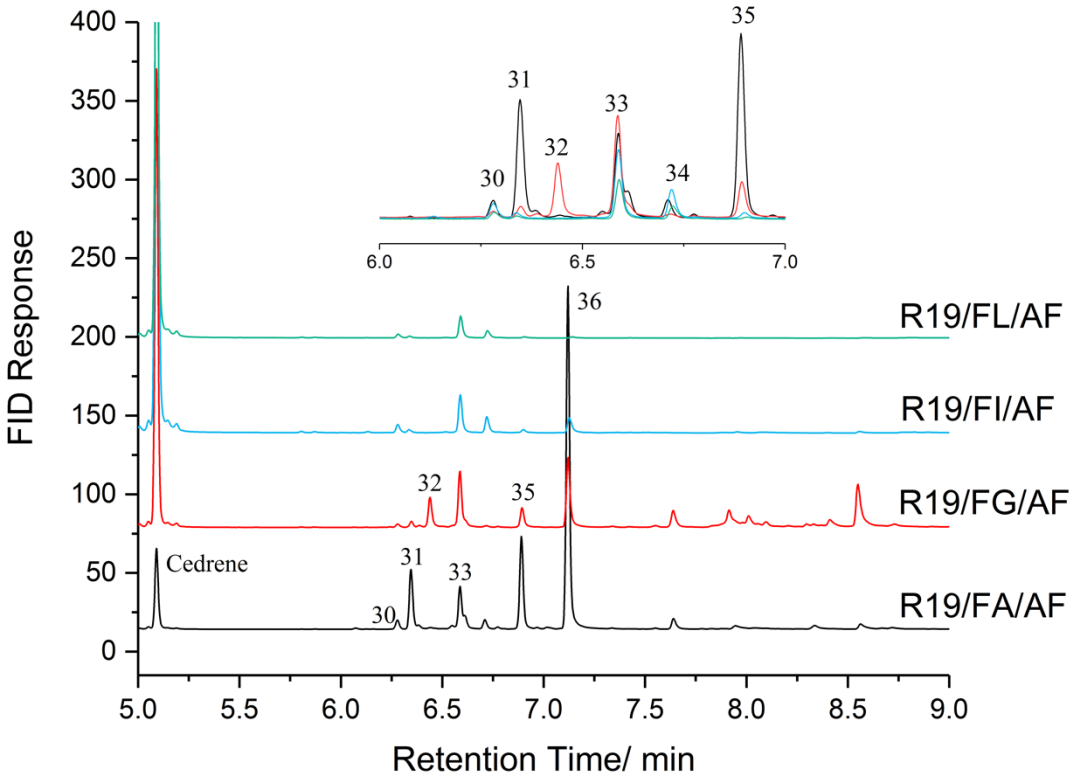
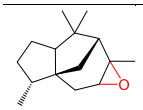
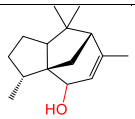
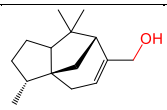
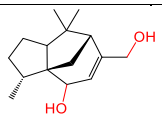



**Fig. 4.13** Product profile of variants containing A328F with different F87 mutations.

#### 4.3.2.2 Other leucine and isoleucine variants: effect of larger side chain volume

For most of the other variants in the R19-based library, the products were limited to oxidation at the C=C double bond and allylic position to form **30**, **33**, **36**, **44** and **45**, especially for those containing the mutations F87I and F87L. The activities showed that larger side chains at the 87 and 328 positions lowered the conversion. R19/FI/AG gave the highest conversion of 84% as glycine has a single hydrogen atom as side chain, while R19/FL/AL has the large side chains and showed the lowest activity. All F87L variants showed very low conversion and **33** was the major product. As the side chain volume at the 328 position increased the selectivity was shifted towards C10 oxidation to give **33** (Table 4.4, Fig 4.13).

**Table 4.4** The activity and selectivity of R19/F87I and R19/F87L based variants.

| Variant | Conv. | Selectivity | | | | |
|---|---|---|---|---|---|---|
| | |  **30** |  **33** |  **36** |  **45** | **Others** |
| **R19/F87I** | 78% | 39% | 4% | 13% | 42% | 2% |
| **R19/F87I/A328G** | 85% | 53% | 6% | – | 36% | 5% |
| **R19/F87I/A328V** | 57% | 46% | 16% | 26% | 5% | 7% |
| **R19/F87I/A328I** | 18% | 38% | 26% | 21% | – | 15% |
| **R19/F87I/A328L** | 8% | 11% | 63% | 6% | – | 20% |
| **R19/F87I/A328F** | 10% | 10% | 45% | 22% | – | 23% |
| **R19/F87L** | 8% | 24% | 46% | 13% | – | 18% |
| **R19/F87L/A328F** | 6% | 11% | 62% | – | – | 27% |
| **R19/F87L/A328I** | 10% | 29% | 32% | 18% | 10% | 11% |
| **R19/F87L/A328L** | 6% | 9% | 73% | – | – | 18% |

The F87I-based variants had higher activity than the F87L analogues and showed reasonable selectivity for C10 and C15 oxidation. Smaller side chains at the 328 position led to further oxidation and the formation of the 10β,15-diol **45**. The data suggested that the diol was formed from further oxidation of the C15 alcohol **36**. As the 328 side chain volume decreased, there was less C10 oxidation and more epoxidation and C15 oxidation (Fig. 4.14). This could be interpreted as the cedrene binding orientation moving away from C9 and C10 being closer to the ferryl oxygen to C8 and C15 being closer. The tendency for further oxidation of the C15 alcohol **36** as

side chain volume decreased is seen in the 28% of **36** and 5% of **45** in the R19/FI/AV variant, 13% of **36** and 42% of **45** for the R19/FI/A328, and 0% of **36** and 38% of **45** for the R19/FI/AG. Interestingly, the diastereoisomeric 10α,15-diol **44** was not detected, suggesting total diastereoselectivity for the oxidation of the C15 alcohol **36** at C10. The R19/FI/AG was the most active with 84% conversion. The A328G mutation also shifted the product selectivity slightly towards the epoxide **30** (56% *vs*. 39%).

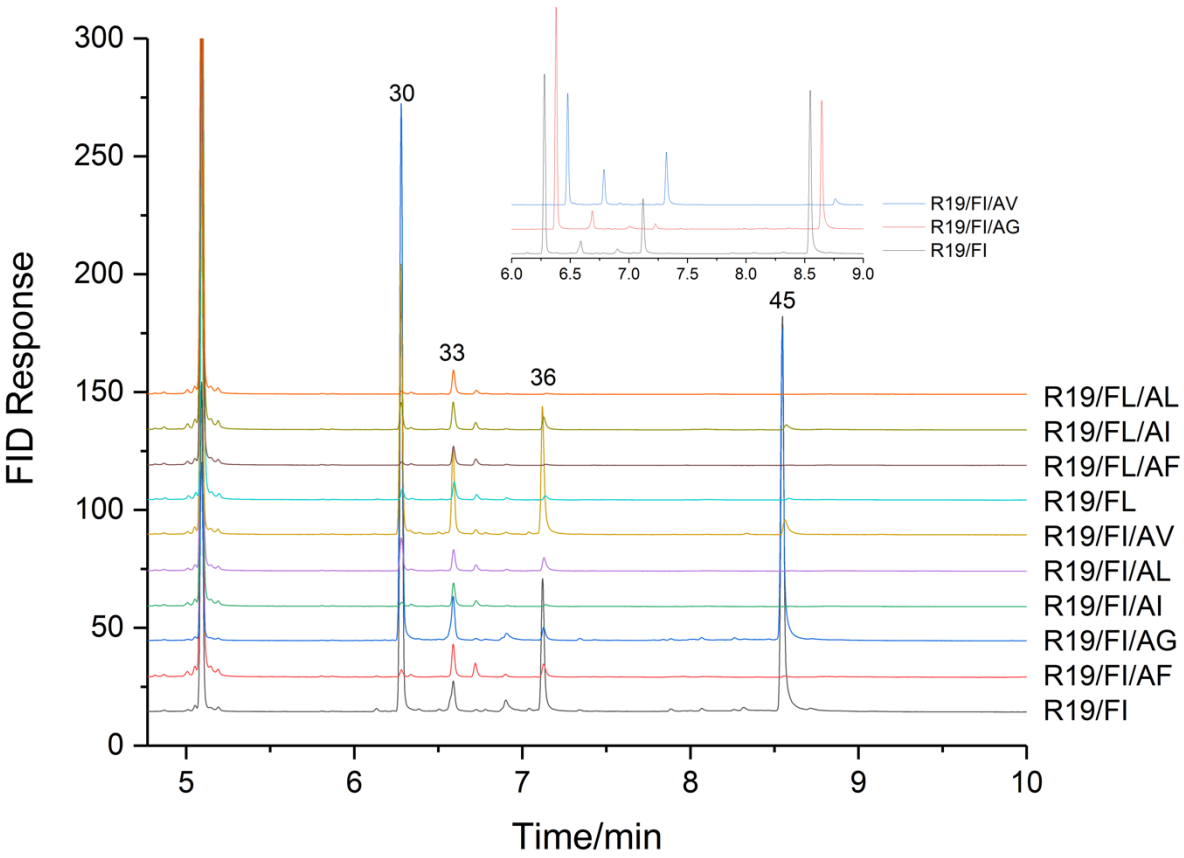
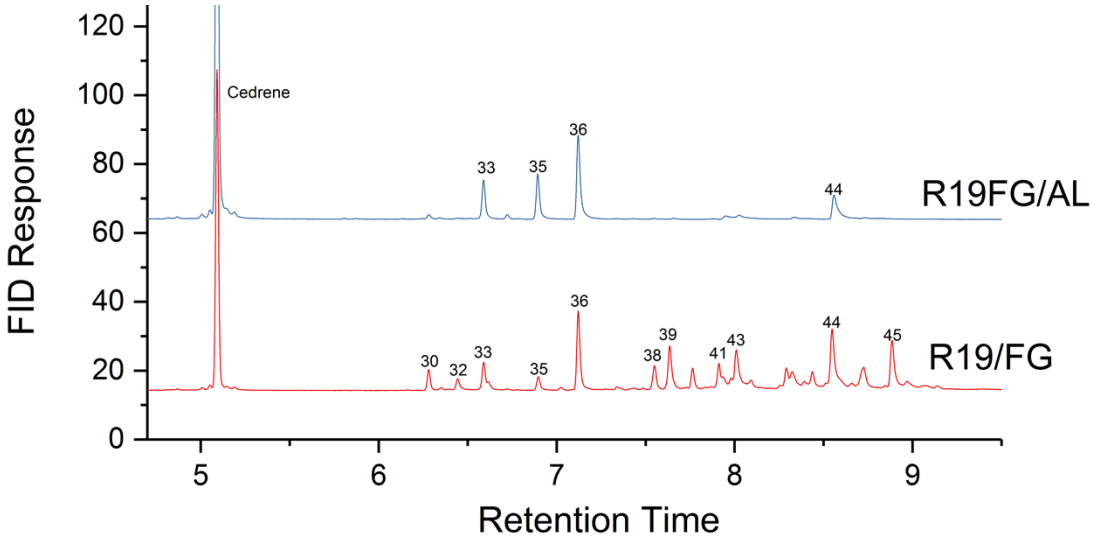



**Fig. 4.14** GC product profile of variants containing A328 mutations based on R19/F87I and R19/F87L.

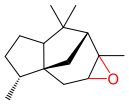
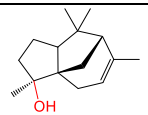
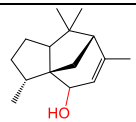
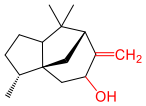
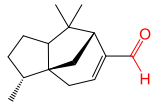
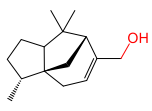
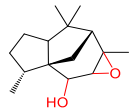
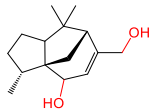
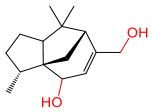
Of the F87G-based variants, only R19/FG, R19/FG/AL and R19/FG/AF were readily purified. They showed reasonable conversion (34% to 70%) and confirmed the significant effect of side chain volume at F87 and A328 (Fig. 4.15).

In R19/F87G, the smallest side chain of glycine dramatically increased the product diversity, giving 11 products. Increasing the side chain volume at the 328 position from alanine to leucine restored oxidation to only at allylic positions (C10 alcohol **33**, C15 alcohol **35,** C15 aldehyde **36**, C10,C15-diol **44**). However, the conversion of R19/FG and individual product yield were not high enough for product characterisation.




Fi**g. 4.15**. GC results of R19/F87G/A328L and R19/F87G. Smaller side chain volume at the 328 position dramatically increased the product diversity from 4 allylic oxidation products to more than 11 products.

**Table 4.5** Summarised results of R19/F87A and R19/F87G based variants.

| Variant | Conv. | Selectivity | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | |  |  |  |  |  |  |  |  |  | |
| | | **30** | **31** | **33** | **34** | **35** | **36** | **37** | **44** | **45** | |
| R19/F87A | 88% | 8% | – | 9% | – | – | – | 14% | – | 46% | 23% |
| R19/F87A/A328G | 86% | 7% | – | 4% | – | – | 1% | 5% | 3% | 15% | 64% |
| R19/F87A/A328V | 85% | 18% | 4% | 11% | 3% | 3% | 30% | – | 27% | 5% | 29% |
| R19/F87A/A328I | 97% | 28% | – | 4% | 1% | 1% | 11% | – | 36% | 8% | 10% |
| R19/F87A/A328L | 92% | 1% | 1% | – | – | 23% | 63% | – | – | 1% | 9% |
| R19/F87A/A328F | 89% | 2% | 11% | 7% | 2% | 16% | 58% | 2% | – | 1% | 2% |
| R19/F87G | 70% | 9% | – | 13% | – | – | 37% | – | – | 41 | 67% |
| R19/F87G/A328F | 47% | – | – | 25% | – | 9% | 29% | – | – | 24% | 39% |
| R19/F87G/A328L | 34% | – | – | 18% | – | 21% | 43% | – | – | 18% | 29% |

#### 4.3.2.3 GVQ-based variants containing F87, A328 and I263 substitutions

The GVQ variant was the template for combining F87V and A328 mutations. Due to the selectivity variations observed with KSK19/I263A and GVQ/I263G in initial screening, three mutations at I263, to alanine, valine and glycine, were introduced into each GVQ/A328X variant to generate a two-residue library with F87V mutation (Table 4.6 and Fig. 4.16). The F87V/A328V combination proved difficult to express, and the GVQ/A328V series of variants also failed to express.

**Table 4.6** Variants of the GVQ sub-library. The GVQ/A328V series of variants (in red) could not be expressed.

| | | | |
|---|---|---|---|
| GVQ/A328G | GVQ/A328V | GVQ/A328I | GVQ/A328L |
| GVQ/A328G/I263A | GVQ/A328V/I263A | GVQ/A328I/I263A | GVQ/A328L/I263A |
| GVQ/A328G/I263V | GVQ/A328V/I263V | GVQ/A328I/I263V | GVQ/A328L/I263V |
| GVQ/A328G/I263G | GVQ/A328V/I263G | GVQ/A328I/I263G | GVQ/A328L/I263G |

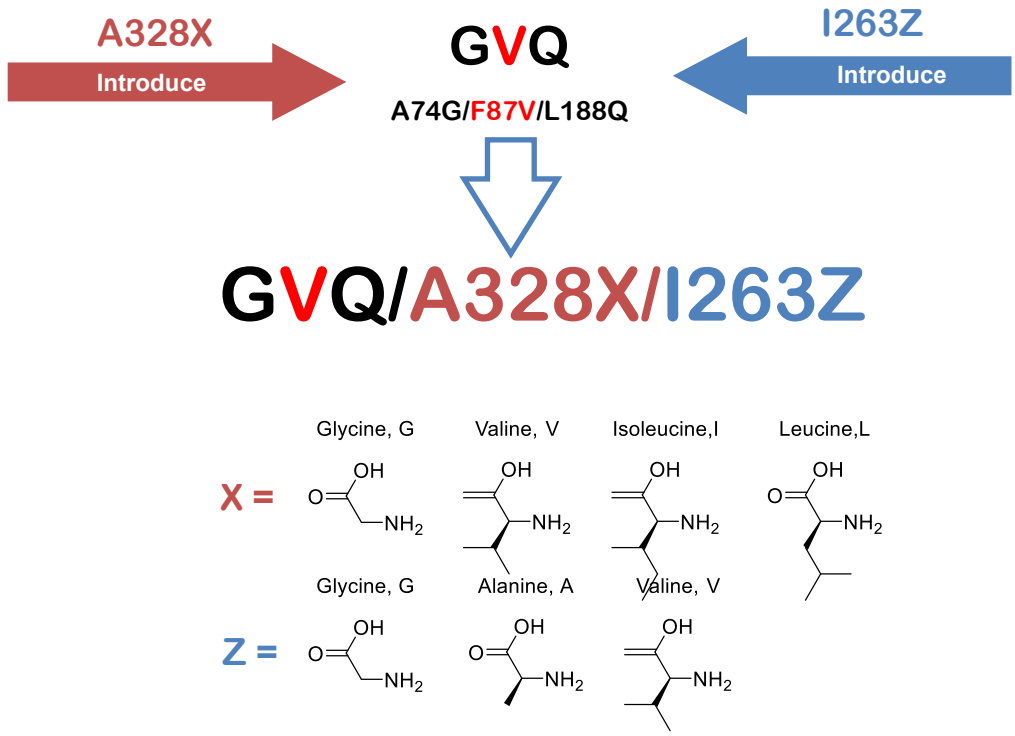



**Fig. 4.16** The construction cogitation of GVQ library and amino acids substituents for A328 and I263.

GC analysis of cedrene oxidation by these GVQ-based variants showed cleaner backgrounds compared to the R19/F87/A328 series, indicating tighter selectivity at certain preferred substrate positions. Allylic oxidation products **30**, **33**, **36**, **37** and **45** were the major products, with small amounts of other products (Fig. 4.17).

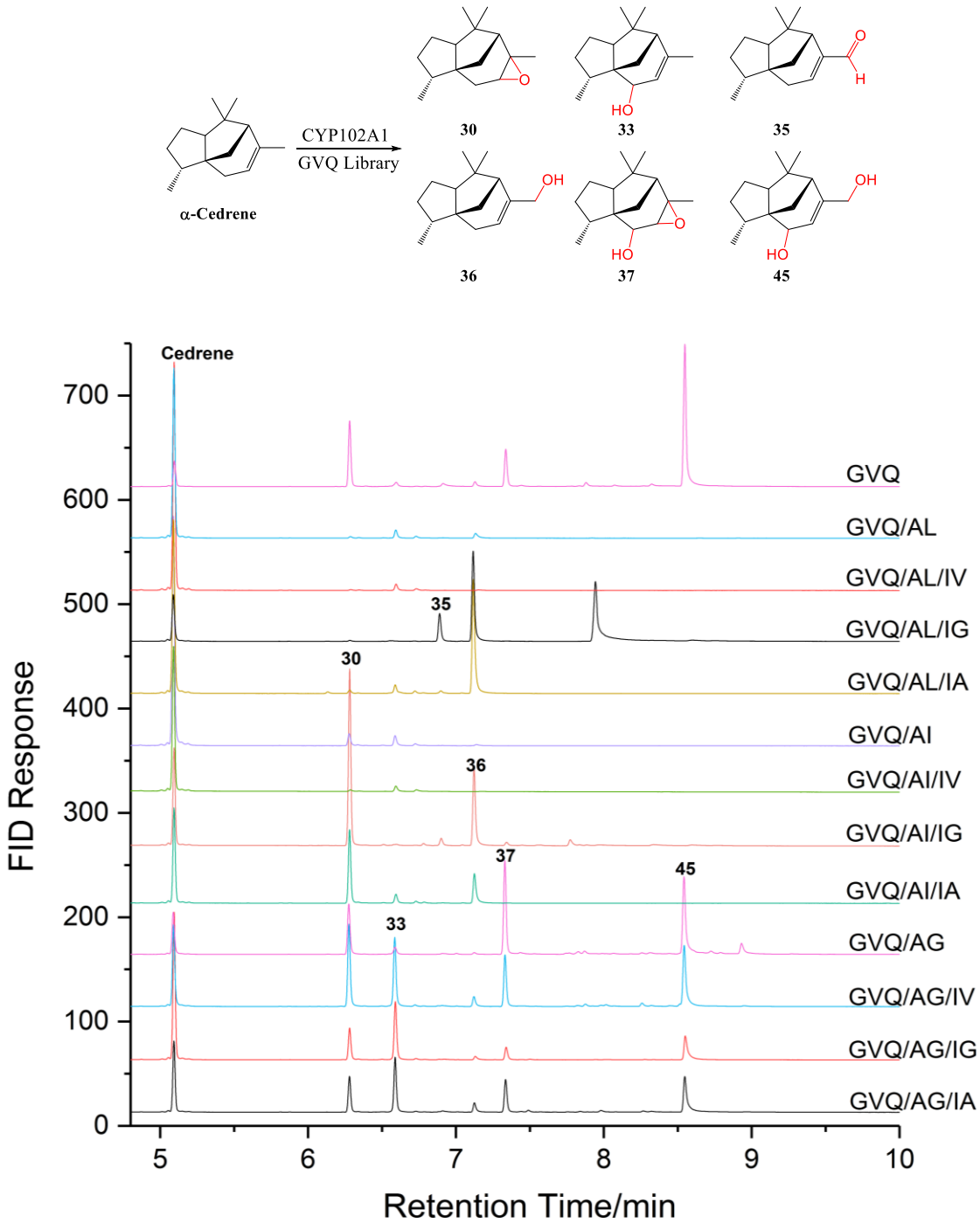



**Fig. 4.17** Major products from GVQ variants library and GC results of each variant.

With increased side chain volume at residue 328, the activity decreased and selectivity preference was shifted from the 10β,15-diol **45** (38% in GVQ/AG *vs*. 59% in GVQ) to **37** (38% for GVQ/AG *vs*. 15% for GVQ). Further increases in side chain volume led to lower conversion but also much reduced extents of further oxidation of the initial products such that the epoxide **30**, the C10-alcohol **33** and the C15-alcohol **36** were by far the major products.

### 4.3.2 A F87–I263 combinatorial library

In both the R19- and GVQ-based libraries, most variants containing A328 mutations gave allylic oxidation products, suggesting the main effect of A328 mutations was slight shifts in selectivity between oxidation of the activated allylic C–H bonds at C10 and C15, the C8,C9 double bond, and combinations among them. However, the A328F mutation stood out since it led to the oxidation of unactivated C–H bonds, at C2 by the R19/FA/A328F variant to form **44**, and at C4 by the R19/FG/A328F to give **45**. Combining A328 and I263 mutations with F87V in GVQ did not enhance selectivity. However, not all F87V-containing variants, and indeed other F87–A328 combinations, were successfully expressed. It was possible that some constituent mutations in the template were not compatible with substitutions at F87 and A328.

Parallel research by other members of the group showed that F87V-containing variants based on K19, i.e. the R47L/Y51F pair of mutations was removed from R19, were expressed to high levels. This led to the construction on a K19/F87/I263-based library in which variants containing the hydrophobic substitution of alanine (A), valine (V), glycine (G), leucine (L) and isoleucine (I) at F87 were combined with isoleucine (I), leucine (L) and glycine (G) at I263. Of this library of 20 variants, 19 were successfully expressed including those containing the F87V mutation. The only exception was the K19/F87G/I263A variant which was not expressed to a suitable level for further experiments (Table 4.7 and Fig 4.18).

Screening reactions showed that these K19/F87 variants maintained similar selectivity compared to R19/F87 variants but at slightly higher activity as more products with longer GC retention times were detected (Fig. 4.19).

**Table 4.7** Variants of the K19 library. K19 = H171L/Q307H/N329Y

| | | |
|---|---|---|
| K19/F87G | K19/F87I | K19/F87A |
| K19/F87G/I263G | K19/F87I/I263G | K19/F87A/I263G |
| K19F87G/I263A | K19/F87I/I263A | K19/F87A/I263A |
| K19/F87G/I263V | K19/F87I/I263V | K19/F87A/I263V |
| | | |
| K19/F87V | K19/F87L | |
| K19/F87V/I263G | K19/F87L/I263G | |
| K19/F87V/I263A | K19/F87L/I263A | |
| K19/F87V/I263V | K19/F87L/I263V | |

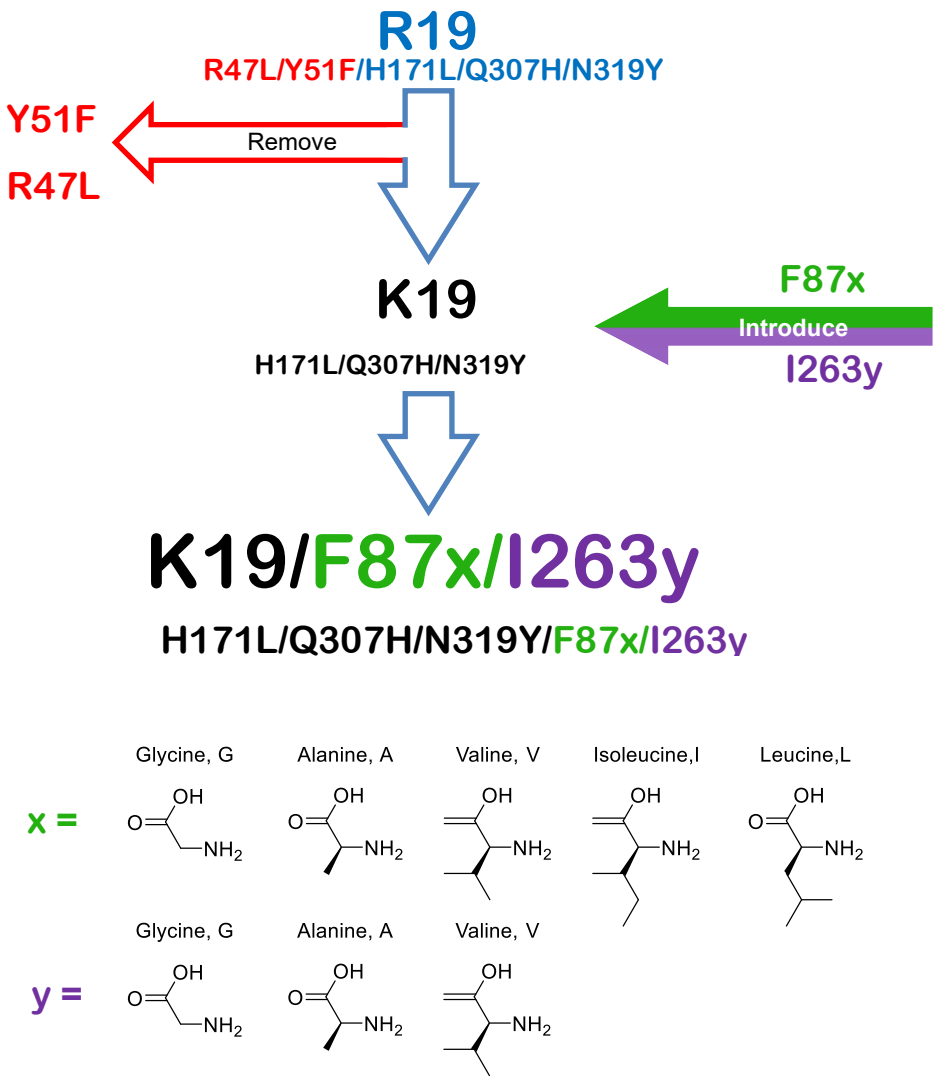



**Fig. 4.18** The construction cogitation of the K19 library and amino acids substituents for A328 and I263.

#### 4.3.2.1 Oxidation of unactivated C–H bonds

Apart from the major products, variants of the K19-based library produced many products with retention time between the epoxide alcohol **37** and the 10,15-diol **44**, indicating they were further oxidation products. Variants with the I263G and I263V in combination with F87A and F87G were particularly active (Fig. 4.19).

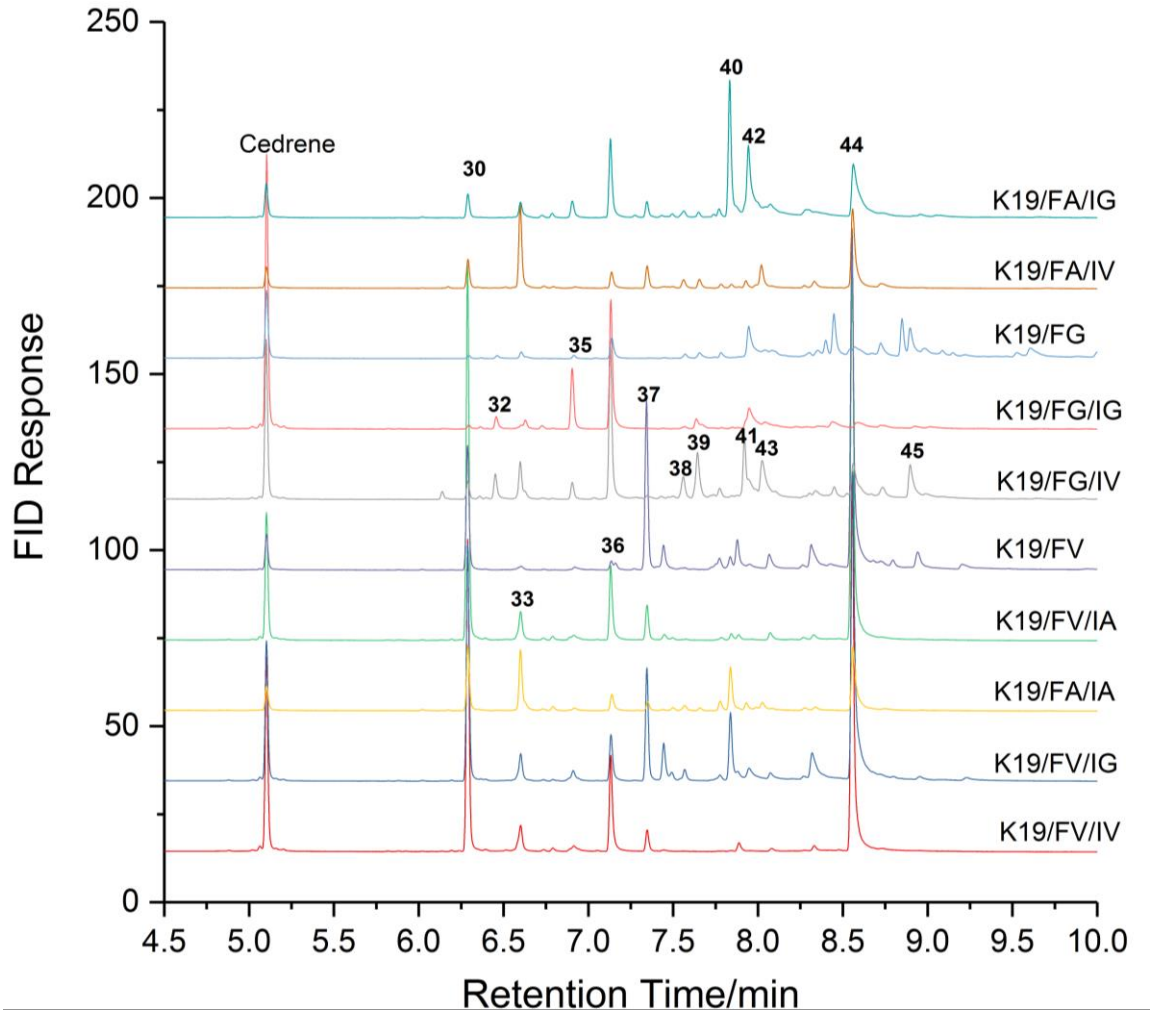



**Fig. 4.19** GC result of selected variants screening reactions which gave enough conversion and product yield for reparative scale reactions. Peaks located between **37** and **44** from K19/FA/IG and K19/FG/IV covered most of selectivity from this series of variants.

K19/F87G/I263V and K19/F87A/I263G were two stand-out variants that gave reasonable conversion (Fig 4.20), and were selected for preparative scale reactions to generate products for characterisation.

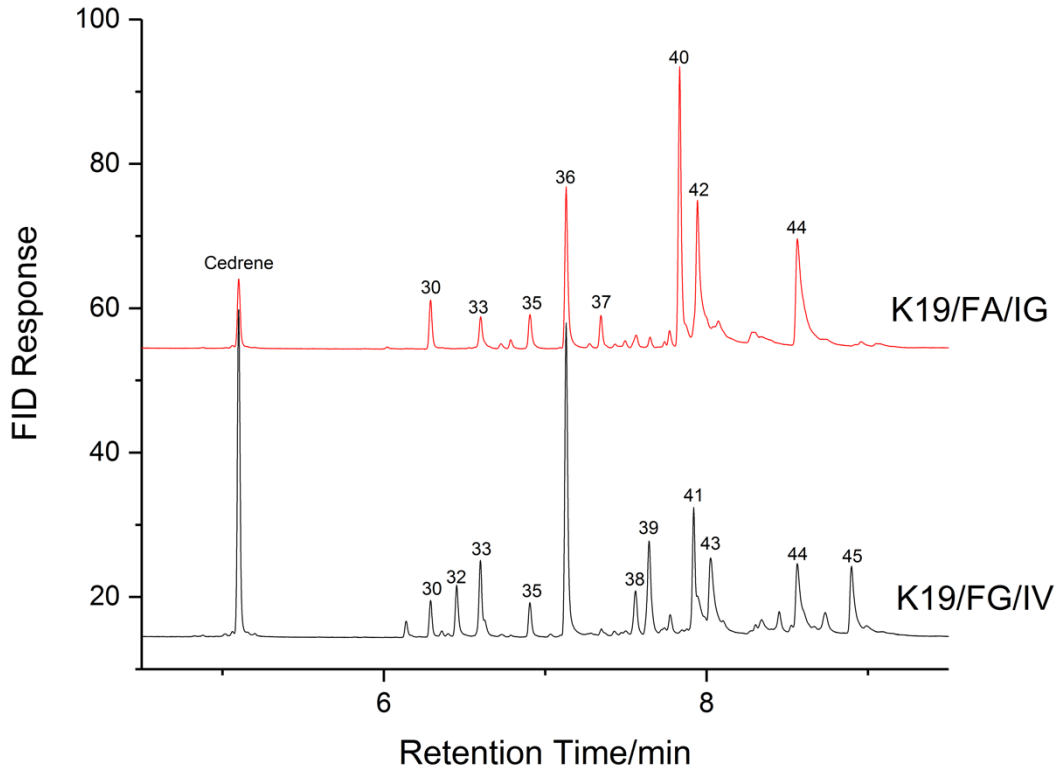



**Fig. 4.20** GC profile of K19/FG/IV and K19/FA/IG. Relevant products are labelled and numbered.

The product **32** was purified from the reaction with the K19/FG/IV variant. The conversion was 80% but **32** could only be isolated in 5% yield after repeated silica gel column chromatography. The MS data (M+16) indicated a monooxygenation product. In the $^{1}$H NMR spectrum (Fig. 4.21), the loss of couplings between H4 and H3 (identified by its coupling to the H2 multiplet) and between the C<u>H</u>OH and H3 suggested the product was 4-hydroxy-cedrene.

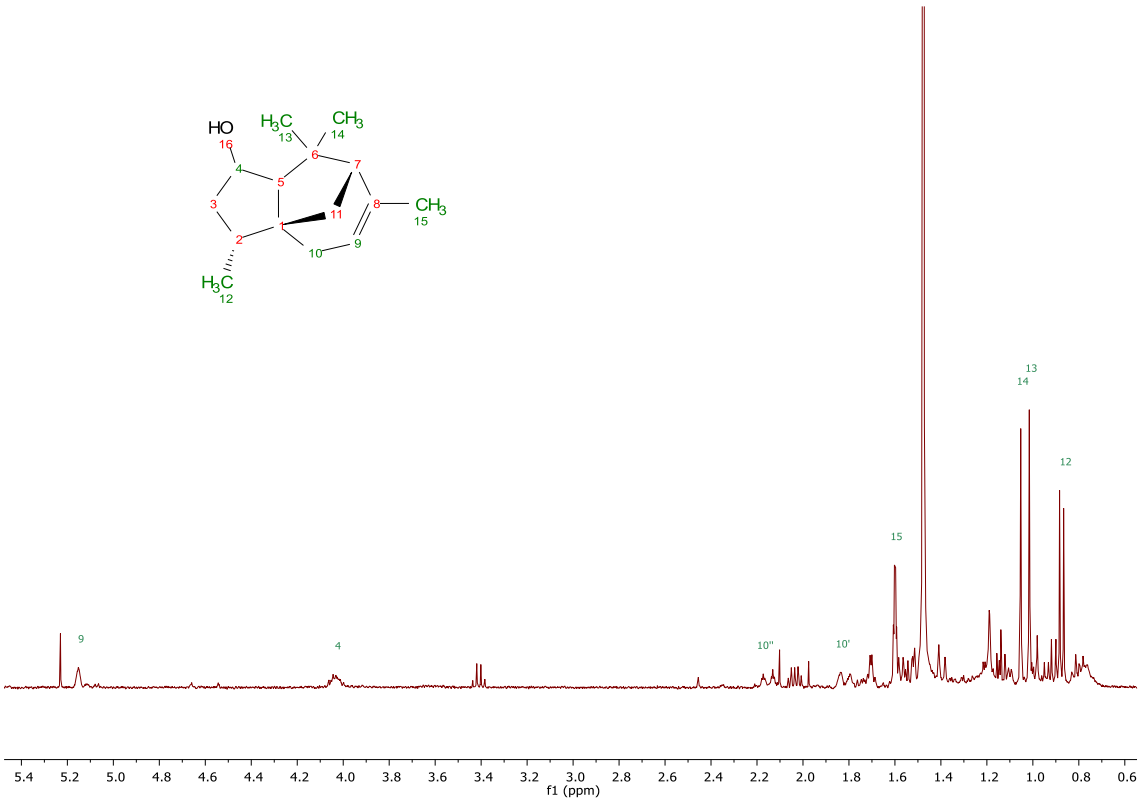



**Fig. 4.21** The $^{1}$H NMR spectrum of **32** in $CDCl_3$

In addition to **32**, K19/FG/IV gave 4 products with retention times within this range in a reasonable total yield of 38%. The K19/FA/IG variant showed >95% conversion and also gave two main products (43% in total) in this retention time window. In combination, these two variants gave sufficient yields for six new compounds, **38**, **39**, **40, 41**, **42** and **43** (Fig. 4.20) to be purified and characterised. Each preparative scale reaction was carried out to convert 100 mg of cedrene with 2 μM $P450_{BM3}$ enzyme.

The highest yield was product **41** (14%) isolated from the reaction with the K19/FG/IV and eluting with an increasing gradient solvent ratio of hexane : ethyl acetate from 10 : 1 to 3 : 1. The MS data of M+30 indicated dioxygenation with the loss of two hydrogens, i.e. a ketone/aldehyde-alcohol or a carboxylic acid. In the $^{1}$H NMR

spectrum the C15 allylic methyl resonance was absent and there was no resonance assignable to a CHOH proton. The alkene H9 proton had shifted to 7.0 ppm while retaining coupling to both H10 protons (Fig. 4.22). Such a strongly deshielded alkene proton could only arise from an α,β-unsaturated carboxylic acid. Hence, **41** was characterised as the C15 carboxylic acid.

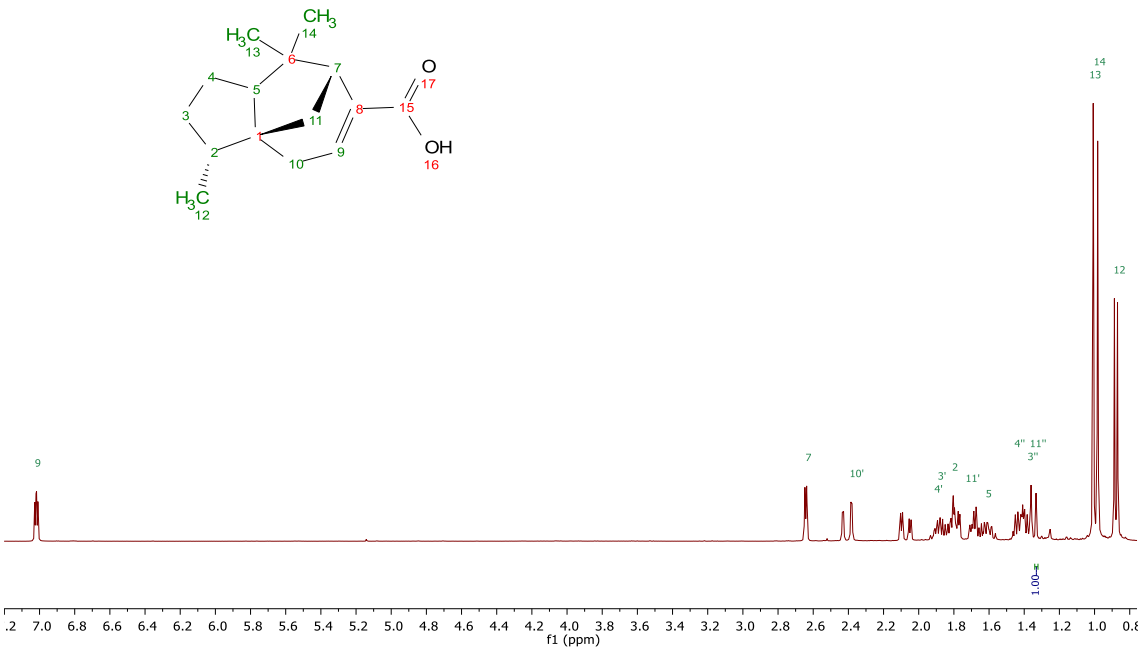



**Fig. 4.22** The $^{1}$H NMR spectrum of **41** in $CDCl_3$

Compounds **40** and **42** could be separated as mixture from the crude reaction extract. However, **42** was lost during the process to purify **40** further. The $^{1}$H NMR spectrum of **40** indicated epoxidation of the C8,C9 double bond since H9 had moved to 3.00 ppm while retaining coupling to both H10 protons. A CHOH proton was observed at 4.25 ppm. All four methyl groups were detected in the aliphatic region, suggesting the oxidation site could only be C3, C4, C7 or C11. C2 oxidation was ruled out as the H2

proton was identified at 1.70 ppm *via* its coupling with the C12 methyl protons. Both H3 protons were present and identified by their coupling to H2. Moreover, the H3 protons showed coupling to the CHOH resonance at 4.25 ppm, thus confirming C4 hydroxylation. Therefore, **40** was assigned as the double oxidation product of C8,C9 epoxidation and C4 oxidation, i.e. 4-hydroxy-cedrene oxide.

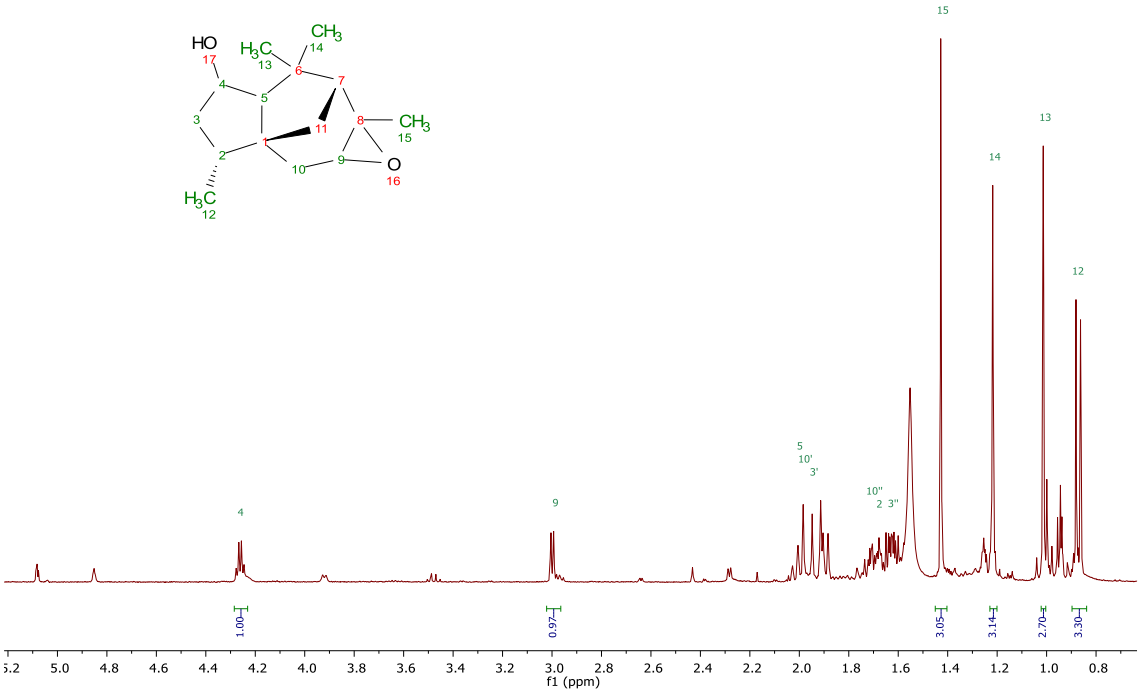



**Fig. 4.23** The $^{1}$H NMR spectrum of **40** in $CDCl_3$

Compound **39** was isolated from the reaction with K19/FG/IV. The $^{1}$H NMR spectrum showed similar resonance of C10 alcohol with the H9 proton at 5.38 ppm and one H10 proton at 4.03 ppm (Fig. 4.24). The C12 methyl protons moved slightly to low-field and retained coupling to H2 at 1.99 ppm. Only one H3 proton was identified via its coupling with H2 and two H4 protons. Hence **39** was assigned as 3,10-dihydroxyl-α-cedrene.

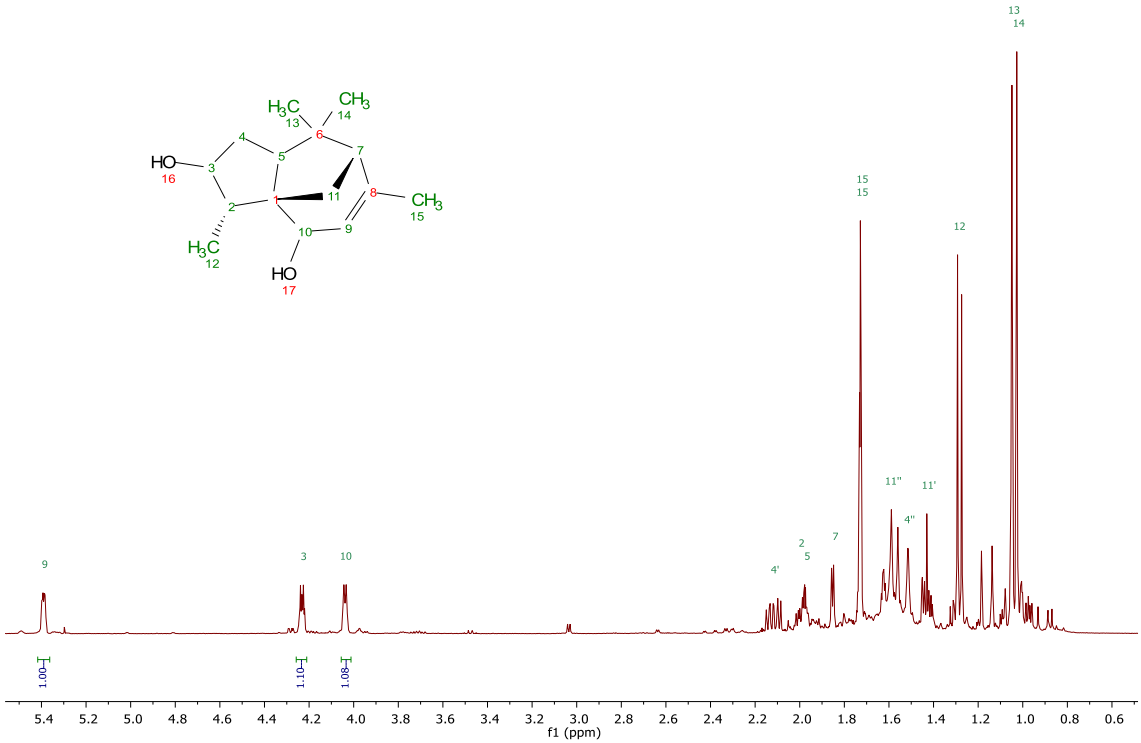



**Fig. 4.24** The $^{1}H$ NMR spectrum of **39** in $CDCl_3$

Similarly, **38** and **43** were assigned as 2,10-dihydroxyl-α-cedrene and 4,10-dihydroxyl-α-cedrene (NMR and MS data see appendix). Therefore, except for C5, the other aliphatic C–H bonds on the 5-member ring were activated either as mono-hydroxylated product (**31** and **32**) or further oxidation of allylic oxidation products (Fig. 4.24).

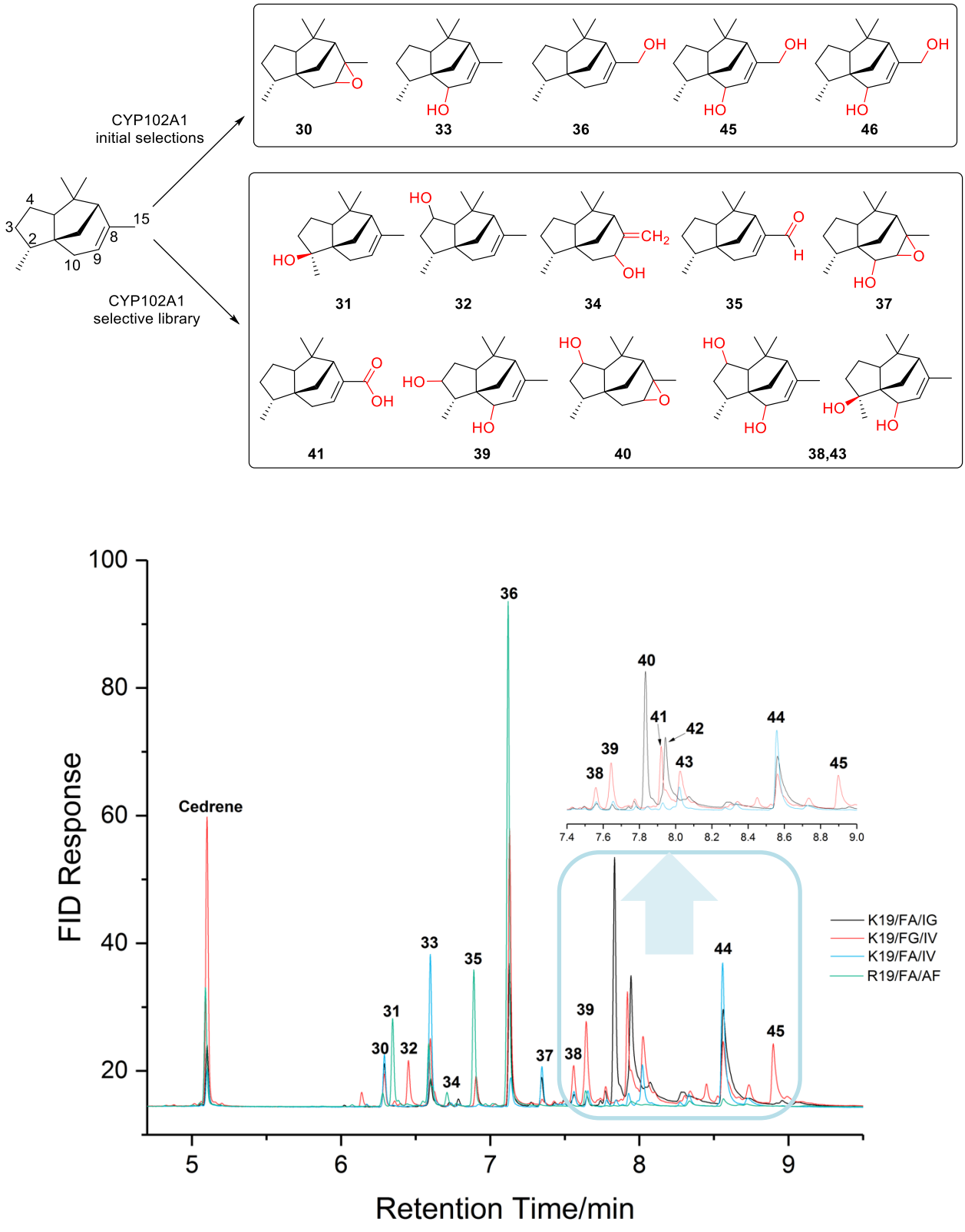



**Fig. 4.25** Products from K19/FG/IV and K19/FA/IG showed the oxidation on 5-member ring and formed the C15 carboxylic acid. Major allylic oxidation products were not shown.

## 4.4 Discussion

The two-residue and three-residue library approach, after key residues were identified from the initial screening, has enabled many new cedrene oxidation products to be isolated and characterised. High conversion was observed for many variants. Oxidation at unactivated aliphatic C–H bonds was found for C2, C3 and C4 oxidation in the five-membered ring. However, most of the new products were at longer retention times and arose from further oxidation of products from initial oxidation at the most activated positions, i.e. the C8,C9 double bond and the C10 and C15 allylic C–H bonds.

Further oxidation was particularly noticeable when the R47L/Y51F couplet of mutations was removed in order to generate the F87V and A328V libraries. This observation is consistent with these hydrophobic substitutions disfavouring entry and binding of more polar molecules (i.e. the product alcohols) while promoting oxidation of hydrophobic molecules (i.e. cedrene). It is also interesting that further oxidation could occur at C2 and C4 as well as the more activated allylic positions. Further work will be needed to investigate if other hydrophobic substitutions at Arg47 and Tyr51, and indeed other polar and charged residues at the entrance to the substrate access channel, could stabilise the fold of variants containing the F87V and A328V mutations to retard further oxidation while enabling the mutations close to the heme iron to bias cedrene binding towards the oxidation of unactivated C–H bonds.

The results showed that mutations to reduce side chain volume, e.g. phenylalanine to alanine and glycine (F87A, F87G), or isoleucine to glycine and valine

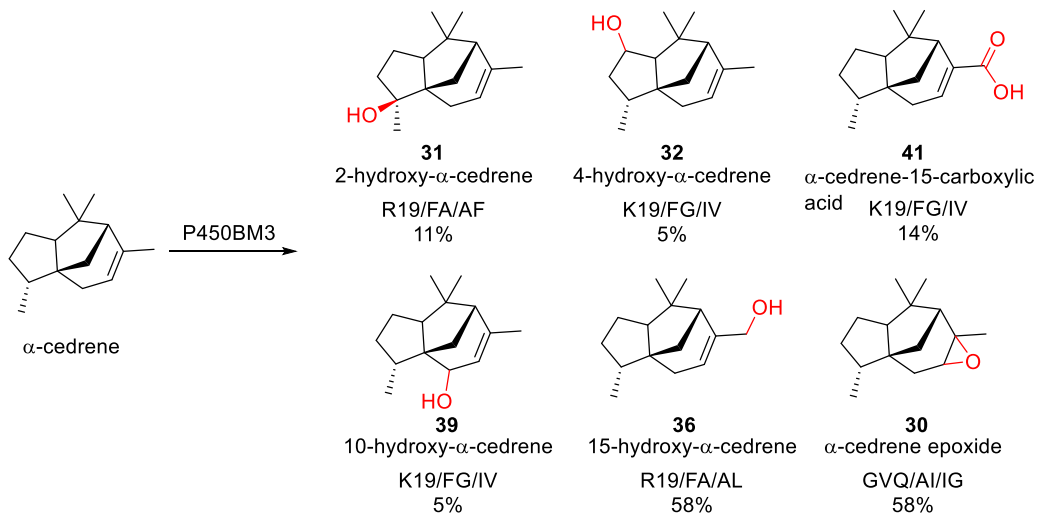



**Fig. 4.26** Mono-hydroxylated products and the corresponding variants with selectivity

(I263G, I263V), expanded the product scope from the allylic region to unactivated positions while larger side chains tended to narrow the range of products formed. For example, mutation of Ala328 to leucine in R19/FA/A328L led to the C15 alcohol and aldehyde being the dominant products (>85% of total products) while A328F in R19/FA/A328F almost abolished epoxidation (<2%) and also gave 10% of the C2 alcohol. Adding the A328F mutation to R19/FG also led to the formation of 18% of the C4 alcohol. Generating more space above the heme would likely lead to multiple cedrene binding orientations, effectively giving greater substrate mobility to present multiple carbon centres to the ferryl oxygen. The most activated positions, i.e. C=C double bonds and allylic C–H bonds, would be oxidised. Larger side chains could limit the number of binding orientations and give higher selectivity. Clearly, certain combinations of larger side chains will prevent cedrene from approaching the heme, leading to low turnover activity. Therefore, the combinatorial library approach is well suited to reveal the necessary combinations for higher activity and selectivity.

Since the F87 and A328 side chains are located directly above the heme in the substrate pocket it was not surprising that mutations at these two residues had such significant effects. Although I263 is in the I-helix and higher up in the substrate pocket, mutations at this residue also had a pivotal role in product selectivity, e.g. GVQ vs. GVQ/I263G, suggesting contact between cedrene and the I263 side chain, at least in some binding orientations. Overall, however, while the two- and three-residue libraries revealed mutation combinations that gave greater product diversity, the selectivity for a specific product remained low; the highest was the 86% total selectivity for C15 methyl oxidation. Any further increases in selectivity will have to come from mutations at other residues around the $P450_{BM3}$ substrate pocket being added to mutants with significant selectivity to a product. Such residues include L75, V78, A82, L437 and T438, amongst others. In principle, a single mutation can be added to the new variants to reveal any effects. Hits can be followed up by site-saturation mutagenesis or a more restricted range of substitutions.

# Chapter 5

# Valencene Oxidation: Scaling Up Study of $P450_{BM3}$ Biotransformation

## 5.1 The grapefruit flavour compound nootkatone

Nootkatone, sharing the same bicyclic skeleton with valencene, is a sesquiterpene ketone with significant applications in medicine and flavour and fragrance industry. Erdtman *et al.* investigated the heartwood constituents of Alaska yellow cedar, *callitropsis nootkatensis*, and extracted and isolated nootkatone from the neutral fraction.[124] Nootkatone has been found in trace amounts in different plants, e.g. 0.3% *v/v* in grapefruit oil but less than 0.01% *v/v* in other citrus oils such as from lemon and orange.[125]

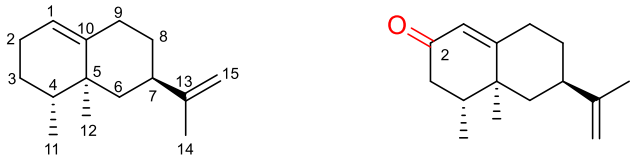



**Fig. 5.1** Structure of (+)-valencen and (+)-nootkatone

Stevens *et al.* studied the odour character and threshold value of nootkatone as well as three related compounds and confirmed nootkatone as the main component contributing to the grapefruit flavour.[126] Most compounds with olfactory properties are chiral and it is common for stereoisomers to have different odour quality and strength.[127,128] It was found that the single enantiomer sample (specific enantiotopic property not reported) possessed a similar odour threshold as the racemic nootkatone indicating there is a distinct difference in odour strength between the two enantiomers. Haring *et al.* carried out an olfactory study on the purified enantiomeric forms of nootkatone and showed (+)-nootkatone to be the main contributor to grapefruit odour with a threshold of 0.8 ppm against 600 ppm for (–)-nootkatone.[127]

### 5.1.1 Nootkatone synthesis

With its pleasant odour and attractive bitter taste, (+)-nootkatone is widely used in perfumery and as a food additive. However, approximately 400 tonnes of the fruit are required to produce 1 kg of ~85% purity natural *ex citrus* nootkatone, leading to a retail price of $7,000/kg. Numerous processes have been developed for the synthesis. Marshall *et al.* reported a total synthesis from dimethyl isopropylidene malonate to form α-vetivone.[129] However, as only an isomer of nootkatone, α-vetivone was renamed isonootkatone. Pesaro *et al.* reported the first total synthesis of nootkatone starting from 4-acetyl-1-ethoxycyclohexene to obtain racemic nootkatone in five steps (Fig 5.2).[130]

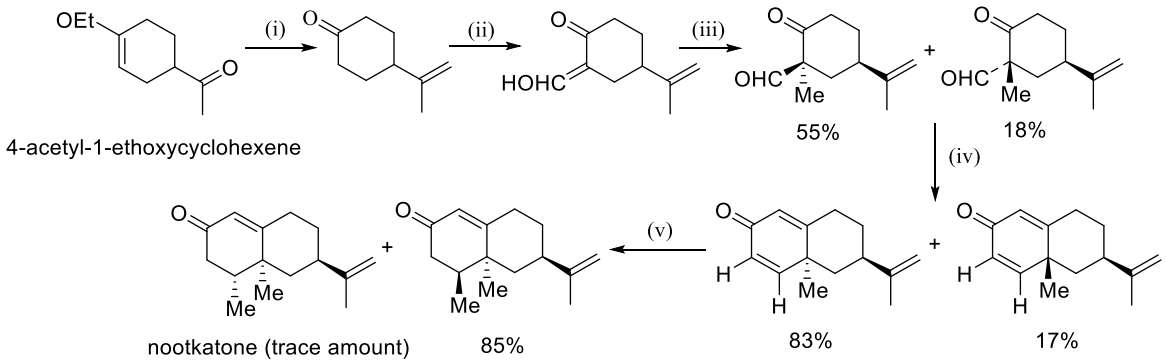



**Fig. 5.2** Total synthesis of racemic nootkatone.(i): triphenylmethyl phosphonium iodide, n-BuLi, THF, (ii): NaOMe, ethyl formate. (iii): MeI, acetone, (iv): piperidine-acetic acid, acetone, (v): $LiCuMe_2$.

In 1973, Dastur reported a five-step route to enantiopure (+)-nootkatone, via a Diels-Alder reaction starting with 1-methoxycyclohexa-1,4-diene (Fig. 5.3).[131] However, many steps were involved and complicated purification impeded its application.

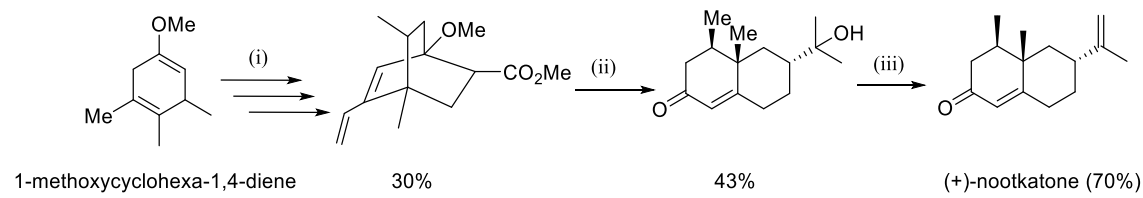



**Fig. 5.3** Total synthesis of (+)-nootkatone. i: methyl acrylate in dichloromaleic anhydride; $SeO_2$ in dioxane; $Ph_3PCH_2$, ii: formic acid; NaOH, *t*-BuOH, iii: collidine, 30% *w/w* alumina.

The oxidation of (+)-valencene to form (+)-nootkatone is an alternative. (+)-Valencene is present in 3% to 5% *v/v* in orange essential oil and even in years of poor orange harvest, valencene supply is plentiful. *t*-Butyl chromate,[132] sodium dichromate,[17] and t-butyl hydroperoxide[133] were reported to oxidise valencene to nootkatone. These oxidising agents are high energy compounds, some are carcinogenic, teratogenic or toxic. In addition, chemical oxidation of valencene also leads to low yields and complex procedures for product purification on a production scale.

Biological processes have been reported using bacteria, fungi and plant cell cultures to oxidise valencene to nootkatone. Valencene was first oxidised to 2-hydroxy-valencene, which is known as nootkatol, then further oxidised to nootkatone. Drawert *et al*. studied cell cultures of five different citrus species, three were found to convert valencene to nootkatone. Of these, the culture of *citrus paradisi* produced 1.79 mg/L nootkatone after 6 hours.[134] *Gynostemma pentaphyllum* was investigated for valencene conversion.[135] A 100 mL culture containing 4 g of callus culture converted 90 mg of valencene to 65 mg of nootkatone after 20 days. The yield from this plant cell culture process was as high as 650 mg/L, but the 20-day reaction time was too long for industrial production.

Microorganisms have been widely studied for converting valencene to nootkatone. Palmerín-Carreño screened 6 microorganisms for valencene oxidation; the best was *Botryodiplodia theobromae* 1368 which gave 230 mg/L nootkatone in 12 days but both substrate and product inhibition were observed.[136] Whole cells of *Yarrowia lipolytica* in a reaction system with orange essential oil to overcome the nootkatone

product inhibition achieved a yield of 850 mg/L in 4 days.[137] Submerged cultures of the ascomycete *Chaetomiu globosum* oxidised valencene to nootkatone and a number of flavour-inactive further oxidation products after days of fermentation.[138]

Although microorganisms having the ability to oxidise valencene to nootkatone have been identified and laboratory scale processes established, the long reaction times (normally days) and associated material and energy costs make the processes uneconomic. To overcome some of these issues, recombinant systems with *E. coli* and yeast expressing wild type as well as engineered P450 enzymes, more recently with co-expression of valencene synthase, have been developed. A lipoxygenase which converted valencene to nootkatone was purified from *Pleurotus sapidus.* The best final yield of 470 mg/L was achieved after 24 hours at 20°C.[139] CYP109B1 from *Bacillus subtills* was tested for valencene oxidation supported by putidaredoxin reductase (PdR) and putidaredoxin (Pdx) from the CYP101A1 system. The yield was 120 mg/L and 35% of total products in aqueous reactions were from undesired side products but 97% of products were nootkatone if an aqueous-organic system was used.[140]

As valencene and nootkatone are two components identified in Alaska cedar heartwood essential oil, the plant should have a biological route to nootkatone. Cankar *et al*. identified a P450 enzyme CYP706M1 from *C. nootkatensis* and combined it with valencene synthase to produce nootkatone *in vivo* by co-expression of both enzymes in the yeast strain WAT11.[94] The process, at 50 mL WAT11 yeast culture, gave 140 μg/L nootkatone in 3 days. This again confirmed that nootkatone may be present in trace amounts in plants due to the property of the relevant wild type enzyme such as the

expression level. An earlier study in our laboratory revealed that $P450_{BM3}$ could be engineered to show significant improvement in activity and selectivity for valencene oxidation to nootkatone but optimisation was needed.[57] In 2015, Schulz reported valencene oxidation to nootkatone by $P450_{BM3}$ mutants including the use of co-solvents and co-substrate screening, leading to a yield of 360 mg/L in 21 hours.[110]

In summary, all biological processes reported to date suffer from low product titres (<0.8 g/L), low substrate conversion (<40%), low turnover numbers (<500), and for fungal and plant cell cultures, long reaction times, rendering them uneconomical. This chapter describes research on valencene oxidation by a $P450_{BM3}$ variant to establish a method that has been scaled up to produce commercial quantities of natural grade nootkatone.

## 5.2 Oxidation of valencene to nootkatone catalysed by $P450_{BM3}$

### 5.2.1 Analysis method

As no detector is comprehensive enough to guarantee the response to all kinds of compounds while still maintain the high sensitivity. TCD (thermal conductivity detector) could detect all compounds but the sensitivity is relatively poorer than other detectors and is normally used in industrial scale analysis. ECD (Electron Capture Detector), FPD (Flame Photometric Detector) and FTD (flame thermionic detector) are selective detectors which are designed to response to compounds containing specific elements. FPD is sensitive with compounds containing S and P, while FTD is capable to detect

compounds with N and P and ECD is for halogen compounds. All of these three detectors are highly sensitive to specific compound but none could detect compounds without these elements. FID is sensitive to compound containing carbon and hydrogen and its sensitivity is higher than TCD but lower than FPD, FTD and ECD due to the wider adaptability of compound scope. Hence, FID response of valencene and oxidation products was used to calculate the conversion of each reaction.

100mM stock solution of valencene, nootkatol and nootkatone were made and different samples of mixture with equal ratio of these stock solution were made to test the difference of FID response and cross influence between compounds. The difference between each compound indicated that the FID does not have significant difference of sensitivity among valencene, nootkatol and nootkatone (Table 5.1).

Table 5.1 Percentage of valencene, nootkatol and nootkatone

| Ratio of compounds (valencene:nootkatol:nootkatone) | Percentage of Peak Area | | | |
|---|---|---|---|---|
| | Valencene | Nootkatol | nootkatone | Impurity |
| 1:1:0 | 41.1% | 41.9% | | 17% |
| 0:1:1 | | 45.3% | 45.1% | 9.6% |
| 1:0:1 | 49.44% | | 48.71% | 1.85% |
| 1:1:1 | 30.81% | 31.78% | 31.85% | 5.56% |

In order to measure the correlation between the FID response and the amount of compound, the calibration curve of nootkatone concentration was carried out. 1-undecanol was selected as internal standard and the ratio of nootkatone at different concentration to the internal standard was calculated according to the integrated peak area (Table 5.2). A standard cure was generated and the slope was calculated as 0.004412 which could be used as the standard reference for further GC analysis of reactions (Fig. 5.1).

Table 5.2 Peak area of internal standard and nootkatone

| Nootkatone Concentration Calibration | | | |
|---|---|---|---|
| Concentration/μM | Internal Standard Peak Area | Nootkatone Peak Area | Nootkatone / Internal Standard |
| 0 | 312546 | 0 | 0 |
| 100 | 316658 | 138339 | 0.44 |
| 150 | 313784 | 207124 | 0.66 |
| 200 | 301187 | 268695 | 0.89 |
| 250 | 306986 | 351624 | 1.15 |
| 300 | 319989 | 349208 | 1.09 |
| 400 | 313412 | 539633 | 1.72 |
| 500 | 314537 | 685847 | 2.18 |
| 600 | 316389 | 815187 | 2.58 |
| 700 | 322230 | 980861 | 3.04 |
| 800 | 307745 | 1102671 | 3.58 |
| 900 | 310975 | 1330784 | 4.28 |
| 1000 | 315075 | 1373499 | 4.36 |
| 1200 | 310396 | 1659953 | 5.35 |
| 1400 | 320609 | 1944419 | 6.06 |
| 1600 | 310747 | 2178628 | 7.01 |
| 1800 | 298822 | 2310087 | 7.73 |
| 2000 | 307493 | 2781668 | 9.05 |

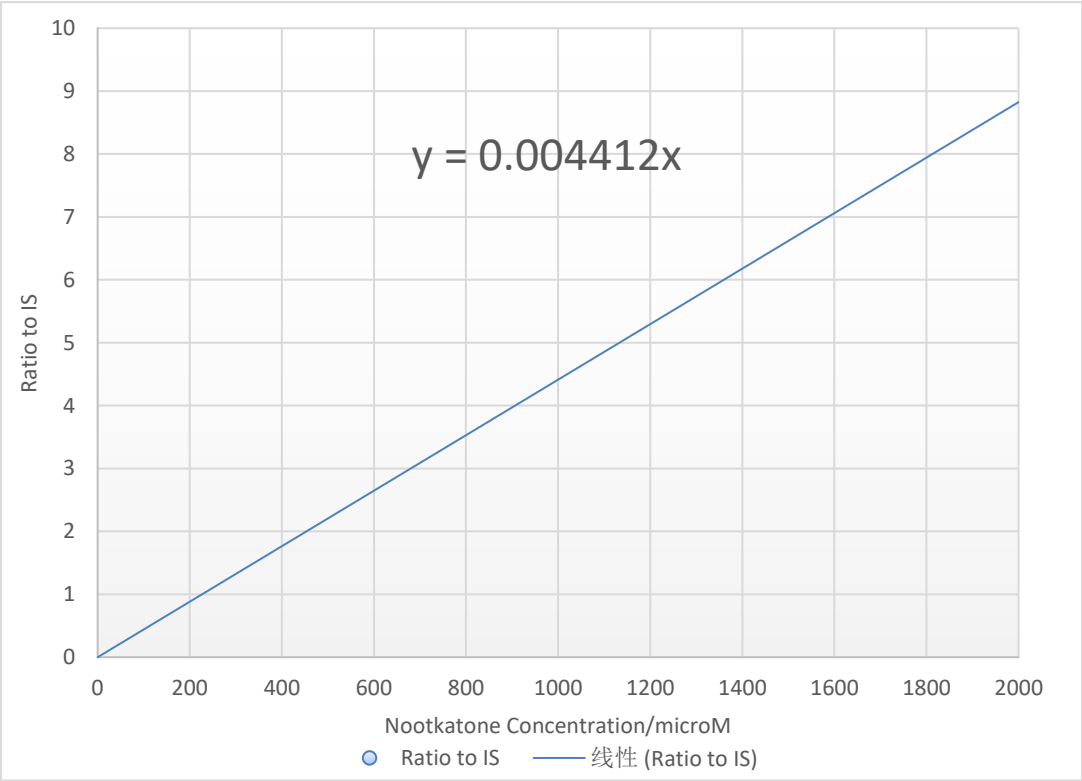



**Fig. 5.4** Standard curve of nootkatone concentration.

### 5.2.2 Variant selection and approach design

The sesquiterpene oxidation screening and engineering described in Chapter 3 identified the RKA (R47L/Y51F/H171L/Q307H/N319Y/F87A) mutant which showed, under screening conditions, ~90% valencene conversion (TON = 1,000) to give 32% nootkatol, 54% nootkatone and 10% of the further oxidation product nootkatone-13,14-epoxide (Fig. 4.1),[57] with <5% in total of other products.

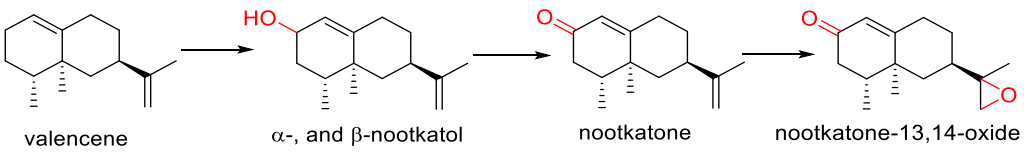



**Fig. 5.5** Oxidation of valencene to *cis*- and *trans*-nootkatol, nootkatone and nootkatone-13,14-epoxide by the $P450_{BM3}$ mutant R47L/Y51F/F87A/H171L/Q307H/N319Y/A328I (RKA).

This Chapter describes the approaches and results on scaling up this screening reaction. Although nootkatone was not the sole product it could be separated from nootkatol by fractional vacuum distillation. With specific reagents and conditions this enzymatically produced nootkatone can be labelled as natural grade under EU regulations, as it is derived from a substance (valencene) that is found in nature and purified by distillation steps only, and its conversion to nootkatone is carried out by an enzyme. The recovered nootkatols could be oxidised by a suitable dehydrogenase enzyme to give more natural-grade nootkatone. The reactions are carried out *in vitro* since products from a whole-cell process must be labelled as being derived from a genetically modified organism (GMO), which is less desirable in a consumer product.

The role and effect of organic co-solvents, second organic phases, aeration, phase transfer agents and valencene addition regimes on the product titre, turnover

number and selectivity were investigated. The conditions were then modified as the volume of reaction was increased to 200 mL and then to 3 litres. These experiments led to a method that gave >10 times higher product titre than the best values reported previously.

## 5.3 The effect of organic co-solvents

Glucose and glucose-dehydrogenase (GDH) were used to regenerate the NADPH cofactor from $NADP^+$ to support the RKA mutant in oxidising valencene to nootkatone (conditions detailed in Chapter 2). Valencene (70% from Sigma) was pre-mixed with buffer solution in a large volume to make the final concentration 1 mM. Initial reactions were carried out in 1 mL volume and valencene was added to a final concentration of 1 mM in the mixture, with 2 μM of RKA. The mixture was shaken at 120 rpm at room temperature for 2 hours, and extracted with 40% *v/v* ethyl acetate. After centrifuging at 14,000× g for at least 2 minutes, the organic phase was analysed by GC. With the increased loading of valencene from 0.5 mM in mutant screening to 1 mM, the activity decreased to less than 5% conversion and no further conversion was achieved through either a longer reaction time or by increasing the concentration of $NADP^+$ and GDH.

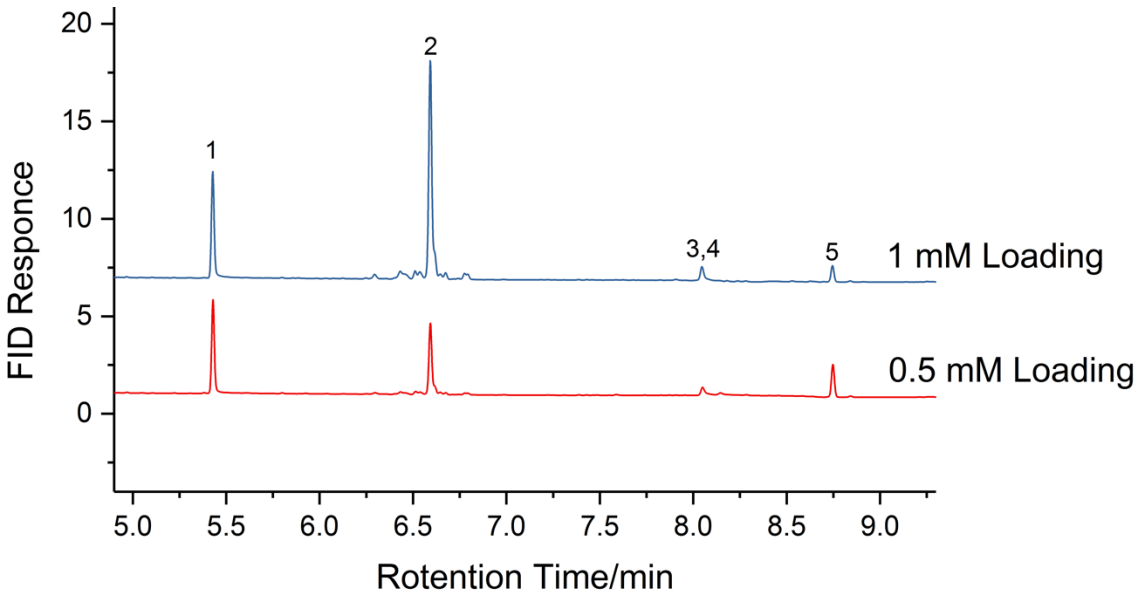



**Fig. 5.6** GC analysis of valencene oxidation at 1 mL scale; **1**: internal standard (1-hexanol), **2**: valencene, **3**,**4**: *cis*- and *trans*-nootkatol, **5**: nootkatone.

Careful observation of the reaction mixture revealed that a thin oil-like film was floating on top of the mixture after valencene was added and it persisted until the end of the reaction. Even under rigorous stirring the highly hydrophobic substrate was in the form of tiny droplets. The results suggested that valencene was not sufficiently well dispersed, indicating that the poor aqueous solubility of valencene could be one of the reasons for the low conversion.

To increase valencene accessibility to the enzyme, some solvents/chemicals and physical treatments were tested. Different concentrations of surfactants such as sodium laurate or solvents such as DMSO and ethanol were added to the reaction mixture prior to adding neat valencene. Sonication for 10 minutes in a sonication bath was also explored as a physical method to disperse valencene and increase its availability to the enzyme. The cofactor recycling system and RKA enzyme were then added to initiate the reaction.

The results in Table 5.1 show that better solubility and substrate dispersion could

improve the conversion but were insufficient to increase the total turnover. Ethanol at 15% *v/v* was inhibitory to the enzyme, while the reaction with 10% *v/v* DMSO gave only 6% higher conversion than that with 10% ethanol. The reaction with 20 mM sodium laurate gave 32% higher conversion than the reaction with 10% DMSO, but higher concentrations inhibited enzyme activity. Sonication showed interesting effects, gaining higher activity than DMSO and ethanol when it was used to disperse valencene into buffer, and even better if valencene was dispersed in a stock solution first.

**Table 5.3** The effect of co-solvents, surfactant and sonication on valencene oxidation.

| | Conv. | Valencene | Nootkatol | Nootkatone | *F.O.P. |
|---|---|---|---|---|---|
| **10% Ethanol** | 9% | 91% | 8% | 1% | 0% |
| **15% Ethanol** | 4% | 96% | 2% | 2% | 0% |
| **10% DMSO** | 15% | 85% | 10% | 5% | 0% |
| **2 mM sonicated in buffer** | 18% | 82% | 7% | 8% | 3% |
| **10 mM sonicated in stock** | 33% | 67% | 24% | 7% | 2% |
| **5 mM sodium laurate** | 36% | 64% | 15% | 17% | 4% |
| **20 mM Sodium laurate** | 41% | 59% | 17% | 21% | 3% |
| **40 mM Sodium laurate** | 15% | 85% | 6% | 7% | 2% |
| **60 mM Sodium laurate** | 6% | 94% | 3% | 1% | 2% |

***F.O.P.** stands for further oxidation products which are products **24-29** (Fig. 3.14)

It was clear that solubility of valencene in the reaction mixture was critical to high activity. Co-solvents did not increase the conversion sufficiently. Dispersion of valencene by sonication and adding a mild surfactant like sodium laurate increased conversion at least 2-fold compared with co-solvent. However, none of these combinations gave over 50% conversion and nootkatone was not the main product in most of these reactions.

Higher conversion could be achieved if valencene could be better dispersed in the aqueous phase, e.g. the large increase in conversion when valencene was sonicated with buffer to prepare a stock. Since sonication creates small droplets (dispersion) of valencene in water, an alternative is to dissolve valencene in an organic solvent. Acetone, ethanol, DMSO, ethyl acetate and isopropanol were tested as solvents for preparing valencene stock solutions which were added to different final volume concentrations to explore the effect on valencene conversion and solvent inhibition and toxicity. DMSO was found to be the most effective solvent (data not shown) and thus investigated further. As shown in Fig. 5.2A, with DMSO as the stock solvent, valencene conversion increased to 90% at 5% *v/v* DMSO but decreased thereafter. In an attempt to increase the product titre, the enzyme concentration was maintained at 2 μM and the reaction volume was kept the same, valencene was added in 1 mM aliquots from a 20 mM stock in DMSO, and a 0.5 mL sample was taken from the 20 mL reaction mixture for extraction with ethyl acetate and GC analysis 2 hours after each aliquot was added.

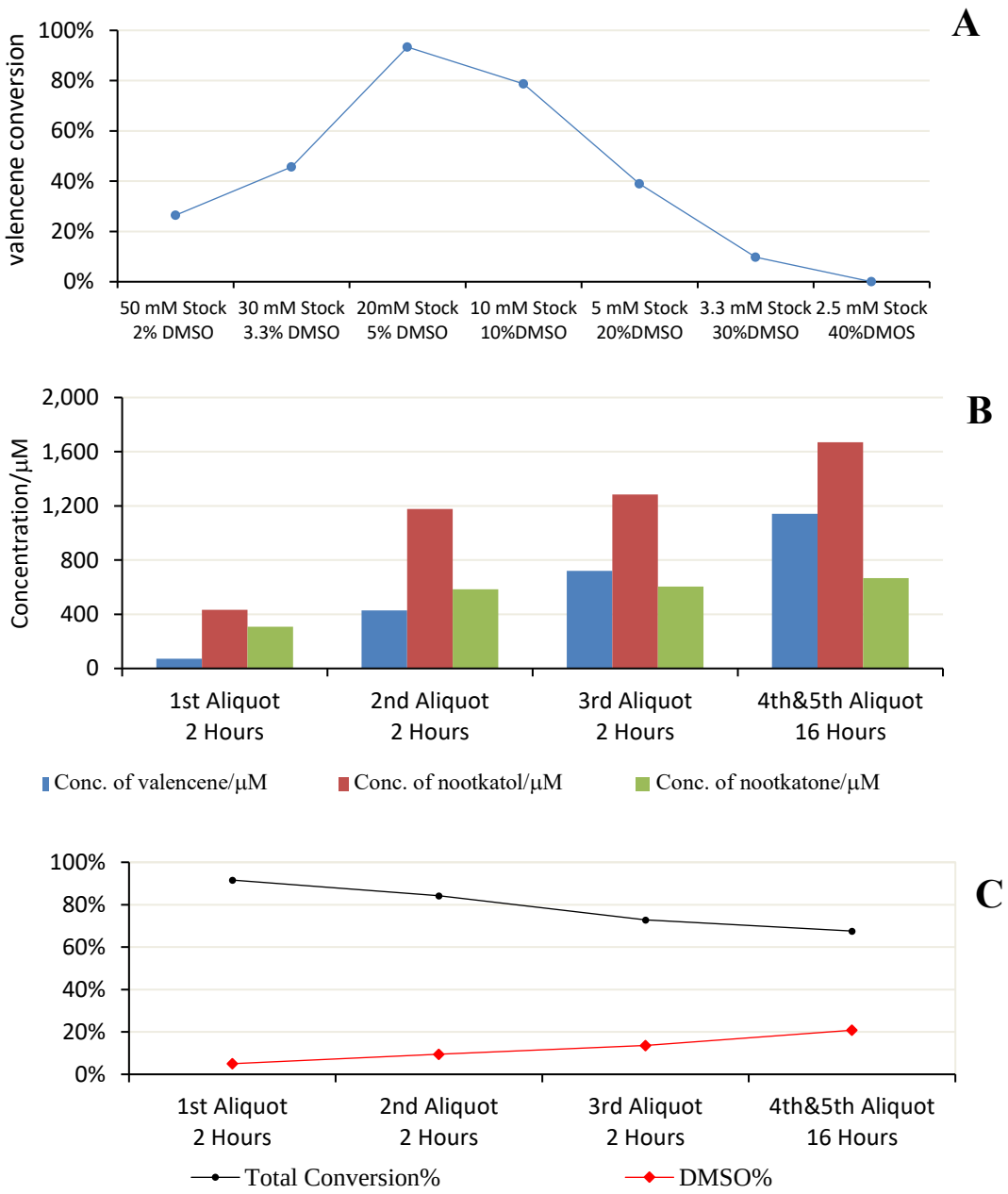



**Fig. 5.7** The effect of organic co-solvents. (**A**) The effect of DMSO concentration on valencene conversion; (**B**) and (**C**) The product concentration and total conversion with valencene added in 1 mM aliquots from a 20 mM DMSO stock solution. The final valencene concentration was 5 mM.

The results (Fig. 5.7.B and Fig. 5.7.C) showed that valencene was accumulating during the process while the nootkatone concentration stopped increasing after the 2$^{nd}$ aliquot. The total turnover number (TON) only increased slightly after the 2$^{nd}$ aliquot and, given as long as 16 hours and more substrate, very little further conversion (TON ~200) occurred compared to TON ~1,000 achieved for the first 3 aliquots. Hence, as

more aliquots of valencene were added the build-up of DMSO led to activity inhibition. DMSO had been identified as an inhibitor by occupying the heme axial position.[141] The data in Fig. 5.7A showed that DMSO strongly suppressed enzyme activity at 20% *v/v*; this concentration was reached at the 4$^{th}$ aliquot and little conversion occurred. It can be concluded that a co-solvent can greatly improve dispersion of the hydrophobic valencene but accumulation of solvent resulted in enzyme activity inhibition.

The solubility of valencene in DMSO was found to be ~50 mM. The maximum product concentration before solvent inhibition occurred would be ~10 mM. Although this was higher than any reported titre, the process would not be economically viable. Valencene was more soluble in ethanol (>1 M) but ethanol was less effective in dispersing valencene and low activity was observed. Attempts to use DMSO/ethanol mixtures did not lead to high activity (data not shown). Therefore, phase transfer catalysts were introduced to improve valencene dispersion.

## 5.4 Phase transfer catalysts

### 5.4.1 Polyethylene glycols

Phase transfer catalysts (PTC) are a class of molecules used to assist transfer of hydrophobic molecules between organic and aqueous phases. Ionic detergents and detergent-like compounds such as SDS and CTAB were ruled out because of their chaotropic properties. Non-ionic detergents such as Tween and Triton did not improve activity (data not shown). Polyethylene glycols (PEG) were selected as the PTC since the lighter PEGs are biocompatible and good solvents. PEGs of average MW between 200 and 10,000 were tested as solvents for valencene stock solutions. The melting point

of PEG increases with the MW, as does the solubility of valencene. PEG with MW >1000 are solids that have to be melted at 50°C to 70°C. The viscosity of PEG increases with the average MW, and extra effort was required to mix valencene and PEG when MW is over 400.

Different valencene–PEGs stocks were added to the reaction to a final valencene concentration of 1 mM, leading to >95% conversion under the standard reaction conditions for all PEG stocks tested. Since the solubility of valencene increases with the MW of PEGs, different stocks were made and each was tested for continuous addition to each reaction when the conversion reached 100%. The results were surprising: if measured by the maximum turnover number, the outcomes (TON ~2,000) were similar regardless of the stock concentration and the reaction would be gradually inhibited when the product concentration approached this threshold. However, it was difficult to control the effect of viscosity as the stocks made with heavier PEGs had higher valencene concentration and less volume was added to the reaction but the heavier PEGs led to high viscosity even at low concentrations. Among commercially available PEGs, PEG-400 (average MW 400) was a good compromise for further scaling-up efforts since it is the heaviest PEG that is a liquid at room temperature and in which velancene is soluble to ~100 mM.

Total conversion of 1 mM valencene by 2 μM RKA was observed at up to 20% *v/v* PEG-400 (Fig. 5.8). The conversion at 10% *v/v* PEG-400 was higher than reactions using a DMSO stock at final 10% *v/v* DMSO. In addition, there was less formation of nootkatone-13,14-oxide although the proportion of nootkatone was also reduced

at >20% *v/v* PEG-400 concentration. The conversion decreased to <10% at 40% *v/v* PEG-400 but the high viscosity at >30% *v/v* PEG-400 could be an critical factor.

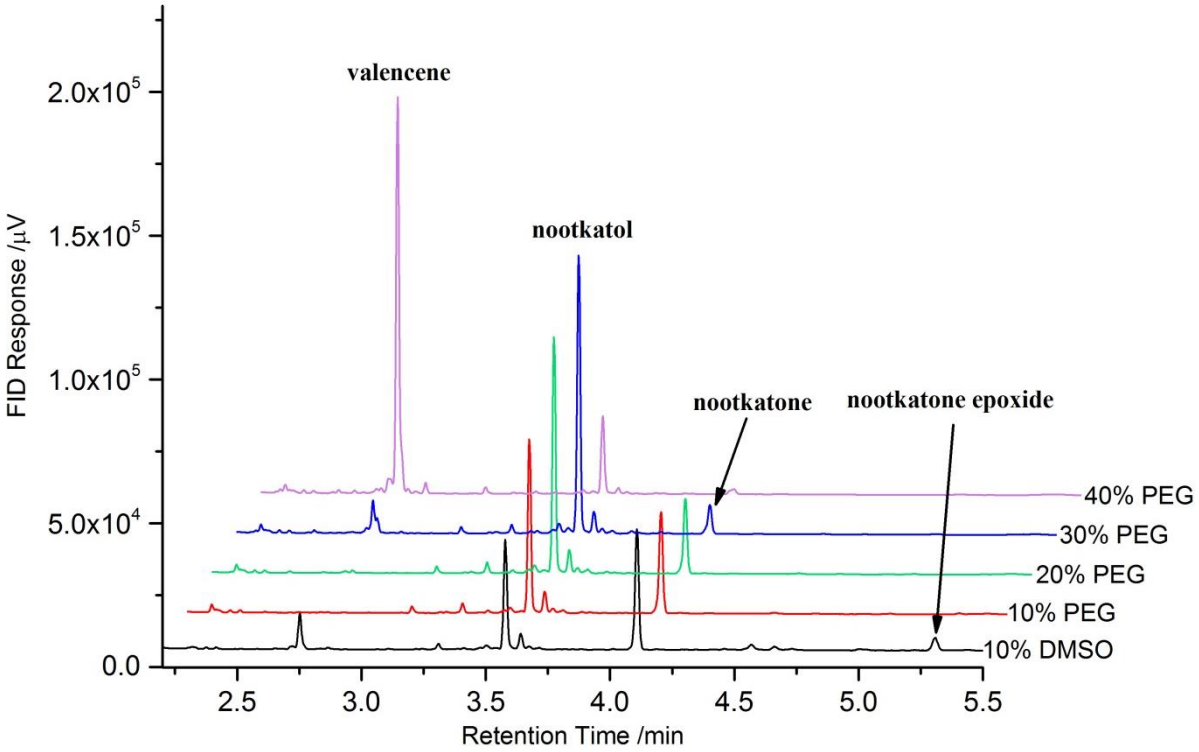
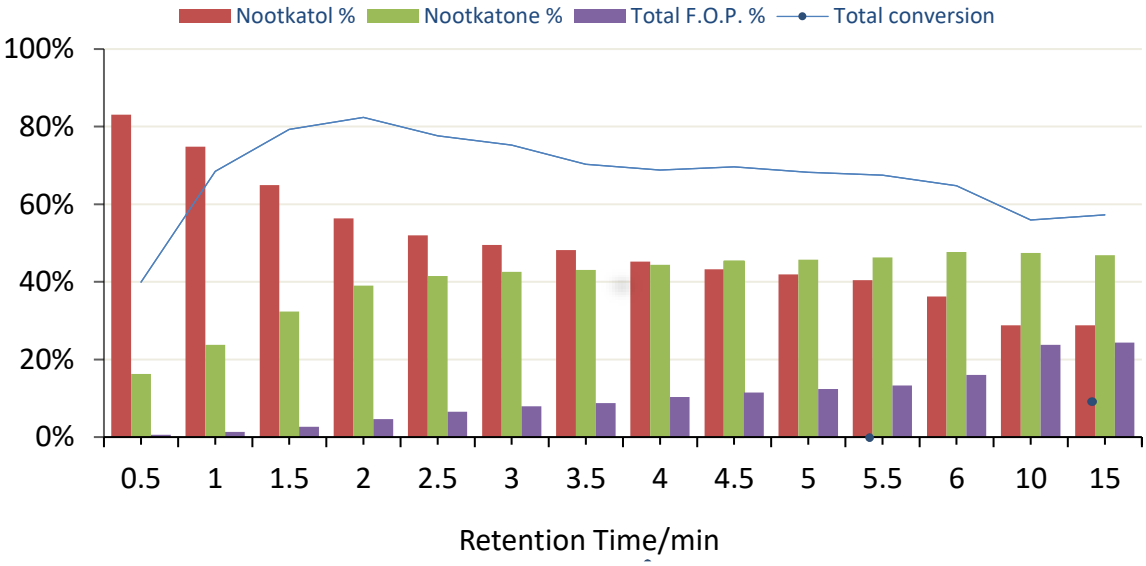



**Fig. 5.8** GC analysis of the effect of PEG-400 concentration on valencene oxidation compared with 10% *v/v* DMSO as control.

Encouraged by these results, valencene was added in 1 mM aliquots every 2 hours, and 90% conversion of 4 mM valencene by 2 μM $P450_{BM3}$ RKA at 20 mL scale was observed. The final PEG-400 concentration was 35% *v/v* and the turnover number was 1,800. Valencene was then continuously pumped into a reaction mixture into 5 μM RKA at a rate of 1 mM/hour for 10 hours and the reaction was allowed to proceed for a further 5 hours. The total valencene conversion dropped from ~90% at 2 hours to ~50% after 10 hours and no further conversion occurred during the ensuing 5 hours (Fig. 5.9). The final PEG-400 concentration was 20% *v/v* which made the mixture viscous and difficult to stir and presumably it also affected enzyme activity. It was notable that the

proportion of nootkatone epoxide increased steadily during the reaction and reached >20% after 10 hours. Hence, little was gained by continuous substrate addition




over adding aliquots since 10 mM substrate with ~50% conversion (TON = 1,000) was too low for a viable process.

**Fig. 5.9** The result of continuously feeding valencene from a PEG stock at 1 mM/hour. Feeding was stopped at $t$ = 10 hours and the reaction allowed to proceed for another 5 hours.

### 5.4.2 The effect of cyclodextrins

An alternative approach to dispersing the highly hydrophobic valencene using a stock solution in a solvent is to add a molecule that functions as a carrier or transporter to bring the substrate into the aqueous phase to be more available to the enzyme. Cyclodextrins are a class of cyclized oligosaccharides naturally generated from starch by glucosyltransferases. The three most common forms, α-, β- and γ-cyclodextrin consist of 6, 7 and 8 glucopyranoside units connected *via* glycosidic α-1,4 bonds (Fig. 5.10). The cyclic chain of cyclodextrins form a cup-like structure with the primary hydroxyl group located at the narrower edge and the secondary hydroxyl group at the wider edge. The interior cavity is relatively hydrophobic, creating an apolar inner

environment to bind lipophilic compounds while its water-soluble surface ensures aqueous solubility.

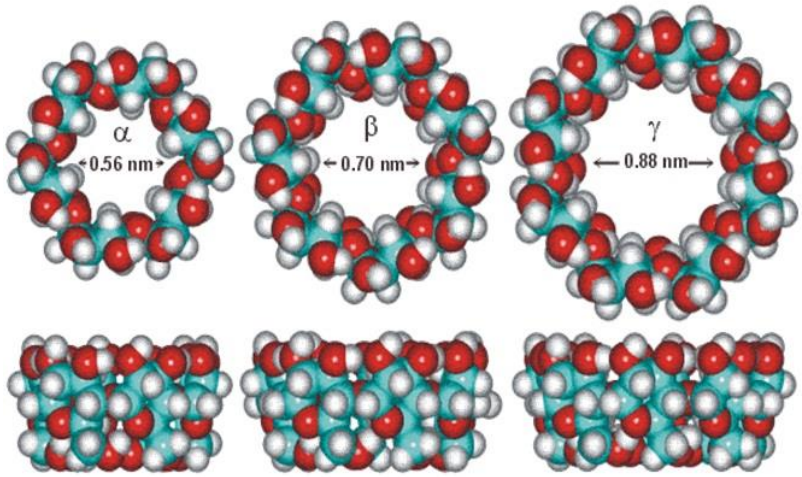



**Fig. 5.10** Structure of α-, β- and γ-cyclodextrin and their sizes of cavity.[142]

The size of guest molecules depends on the size of the cavity in the different types of cyclodextrin; α-cyclodextrin is suitable for aliphatic chains while β-cyclodextrin could better fit aromatic rings such as benzyl salicylate, paclitaxel,[143] and cholesterol,[144] which would also be suitable for valencene, nootkatol and nootkatone binding. However, β-cyclodextrin has surprisingly low solubility, only 2% *w/v* compared with 13% and 26% of the other two cyclodextrins, because of the intramolecular hydrogen bond between the secondary hydroxyl group at the wider edge. Chemically modified β-cyclodextrin at the secondary hydroxyl group could break the internal hydrogen bond to increase the solubility while maintaining the cup-like structure and bipolar property of β-cyclodextrin. 2-Hydroxylpropyl-β-cyclodextrin (HPβCD) was first synthesised by Pitha *et al*. in 1985 and the solubility was dramatically increased to 60% *w/w*.[145] HPβCD was therefore introduced into the reaction system for its high solubility in water.

HPβCD was initially tested at 1 mM valencene against 2 μM RKA enzyme at 1 mL scale. To avoid the cross influence of PEG-400 and cyclodextrin, the reaction was carried out with valencene added from a 20 mM DMSO stock and a reaction without HPβCD was carried out as positive control in parallell. HPβCD was tested at 5, 10 and 25 mM concentration. GC analysis showed that, after 2 hours, the first aliquot of valencene was completely converted in all 3 test reactions. Therefore, a second aliquot was added to each reaction which was allowed to proceed for another 2 hours; 100% conversion was achieved again in the reactions with 10 mM and 25 mM HPβCD. In order to test the capability HPβCD to solubilise valencene, extra neat valencene was added into the reactions with 10 mM and 25 mM HPβCD to make the final valencene concentration to 5 mM. These two reactions were left for a longer reaction time and then analysed by GC.

As shown in Fig. 5.11, the positive control finished with ~95% conversion of one aliquot of valencene (1 mM). The second aliquot added into the reaction with 5 mM HPβCD was not fully consumed as it gave the similar product yield compare to the positive control. Higher HPβCD concentration increased the total conversion to 65% in the reaction with 25 mM HPβCD after valencene was topped up to 5 mM from 2 aliquots of DMSO stock and 3 mM neat valencene, suggesting more HPβCD would increase the turnover number. The data also showed that a higher concentration of HPβCD lowered the extent of further oxidation and more of nootkatol and nootkatone accumulated.

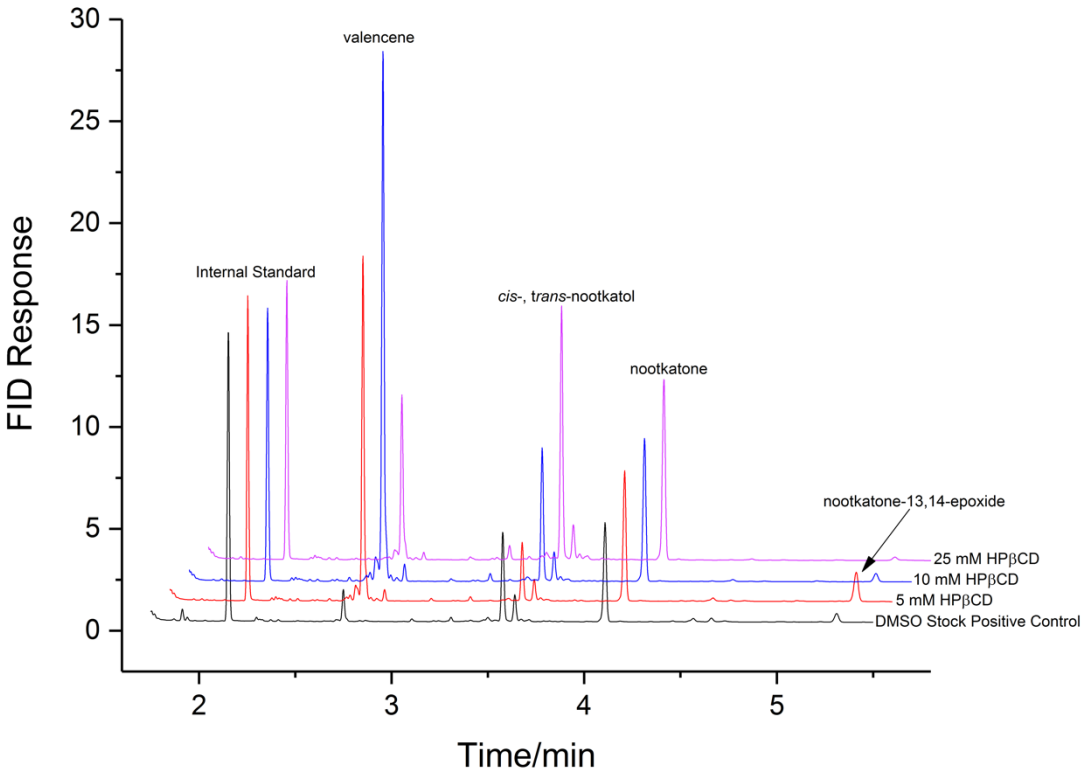



**Fig. 5.11** The effect of HPβCD on valencene conversion and product distribution. Internal standard is 1-hexanol. The final [valencene]:[enzyme] in different reactions were as follow:

| | |
|---|---|
| DMSO positive control: | 1 mM:2 μM |
| 5 mM HPβCD: | 2 mM:2 μM |
| 10 mM HPβCD: | 5 mM:2 μM |
| 25 mM HPβCD: | 5 mM:2 μM |

### 5.4.3 The effect of aeration and pH control

The reaction was scaled up to 20 mL with 3 μM RKA enzyme and 25 mM HPβCD. Valencene was added in 1 mM aliquots from a 20 mM PEG-400 stock. Samples were taken every 2 hours for GC analysis followed by the next 1 mM aliquot of valencene. The effect of increasing the oxygen concentration in the reaction was also studied. Two parallel reactions were set up under the same conditions but only one was aerated with pure oxygen.

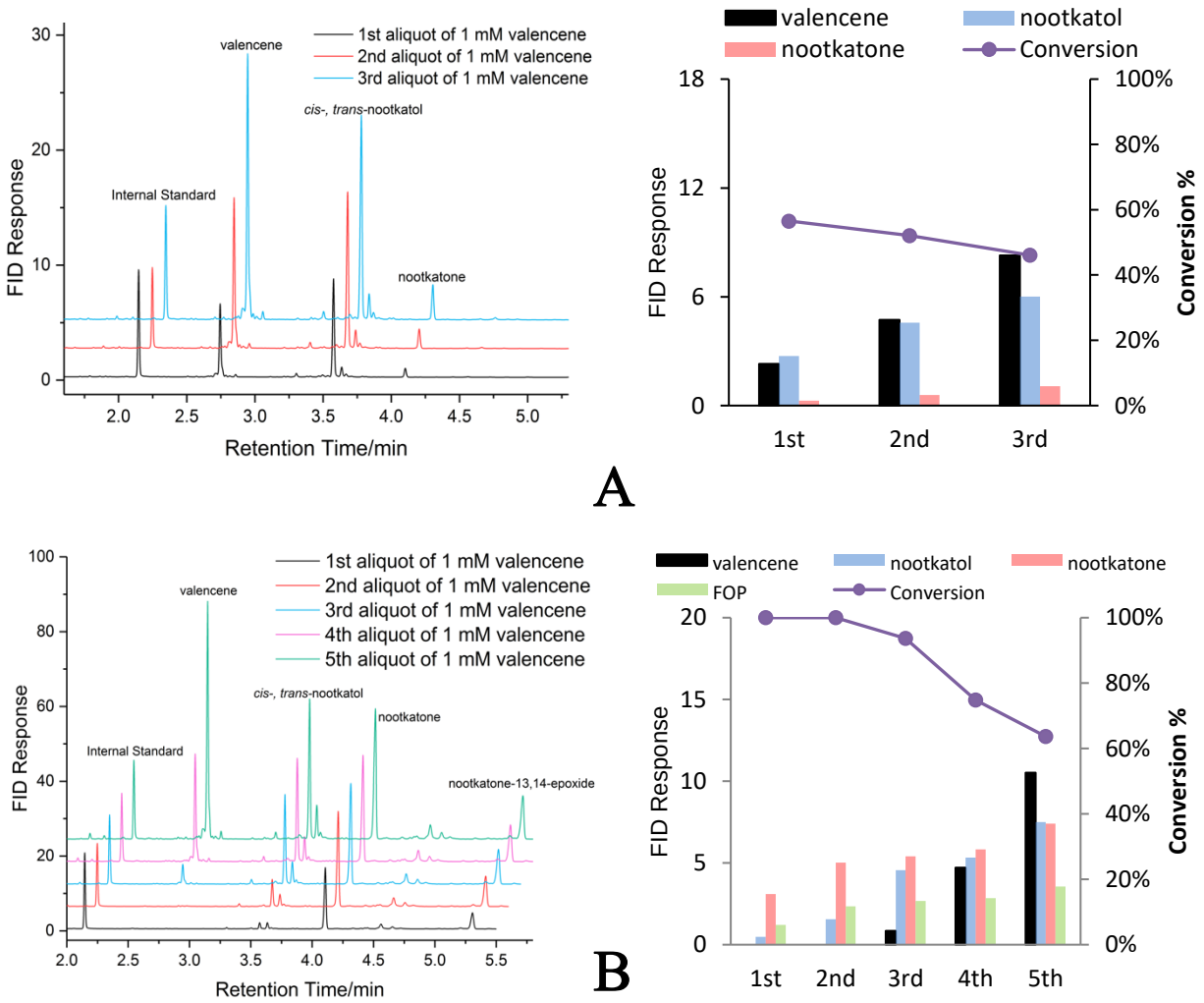



**Fig. 5.42** Effect of aeration. A. GC analysis of the reaction **without** aeration and the integration result after each aliquot. B. GC analysis of the parallel reaction **with** aeration and the integration result after each aliquot.

A total of five aliquots of valencene were added. GC analysis showed that the aerated reaction gave 100% conversion up to the 3$^{rd}$ aliquot (Fig. 5.12). Conversion was slower for the 4$^{th}$ and 5$^{th}$ aliquots but at the same time the proportion of nootkatone also increased to approximately 50% at the 4$^{th}$ aliquot (final TON>1000). On the other hand, the parallel reaction without aeration was much slower and had essentially stopped after 3 aliquots and the nootkatols were the main products with <10% nootkatone. There was no further conversion after another 16 hours (final TON ~500). The pH was measured at

the end of the reaction and found to have dropped from 7.8 to 6.0 in the aerated reaction. This significant decrease in pH of the reaction mixture would likely decrease the activity of the enzyme.

A 100 mL reaction was then carried out with 25 mM HPβCD with the RKA enzyme and valencene added as 1 μM and 2 mM aliquots, respectively. To avoid enzyme inhibition at >30% *v/v* PEG (see section 5.4.1), and since conversion of neat valencene was observed in the reaction with HPβCD (see section 5.4.2), 320 μL valencene (M.W. = 204.35 g/mol, density = 0.92 g/mL at 25°C), which would take the substrate concentration to 10 mM when added to a 100 mL reaction mixture, was mixed and well dispersed in 10 mL PEG-400 by vortexing. The volume of each aliquot of 2 mM valencene mixture would be 2 mL and the final percentage of PEG-400 would be 10%. The reaction mixture turned cloudy when valencene was added. The reaction was aerated by pure oxygen and the pH was monitored continuously and adjusted back to 7.8 by dropwise addition of 1 M KOH. A 0.5 mL sample was removed and analysed by GC when the reaction mixture turned clear. Extra aliquots of enzyme and substrate were added when GC confirmed the conversion.

After 8 hours, a total of 5 μM RKA enzyme and 10 mM valencene were added. GC analysis showed that, except for the first aliquot, all aliquots of enzyme gave almost 100% conversion of valencene (Fig. 5.13). Nootkatone was found to be the major product but the nootkatone epoxide also accumulated and reached a similar proportion as the nootkatols in the products. An extra 1 mM valencene was added after 10 hours without any extra enzyme. Total conversion was achieved after overnight reaction and

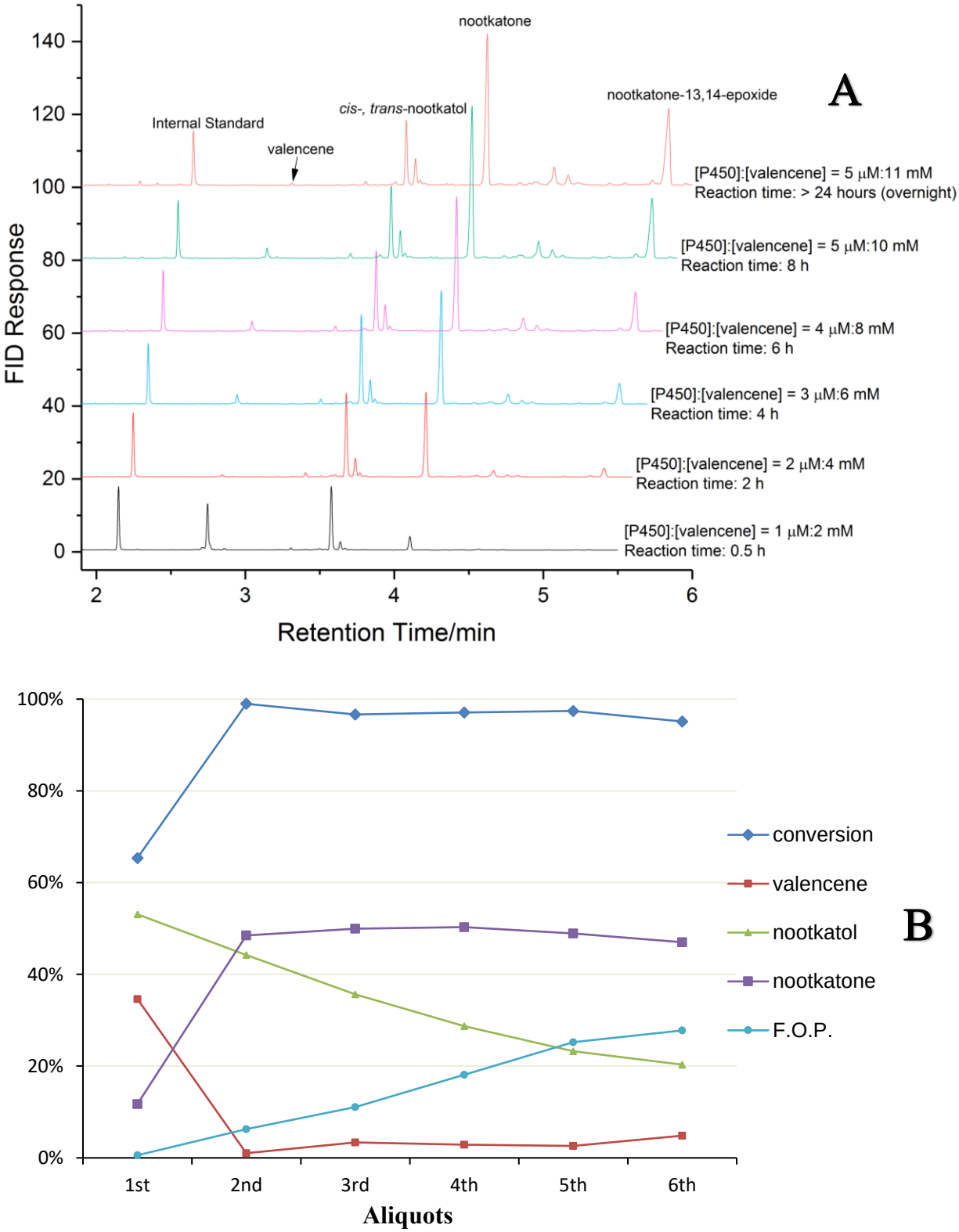



**Fig. 5.53** The GC plots and integration result of the reaction with aeration and pH adjustment. **A**. GC result of each aliquot of enzyme and substrate. **B**. The integration result of each aliquot conversion by the time the sample was analysed.

the proportion of the further oxidation product also increased. Overall, total conversion and a turnover number of 2,200 was achieved with aeration and pH adjustment. Valencene was not fully dissolved in the PEG-400 stock before being added to the reaction mixture, and yet total conversion was achieved.

Oxygen mass transport and pH were not previously noted as major factors in valencene oxidation. However, the solubility of oxygen in water at 20°C is only ~10 mg/L, (0.3 mM). Considering the catalytic cycle is not completely coupled to substrate oxidation, aeration becomes very important when the conversion is more than 5 mM. Moreover, because of the partial coupling, more glucose is oxidised to gluconic acid (pKa = 3.86), causing the pH to drop and a significant reduction in enzyme activity.[146]

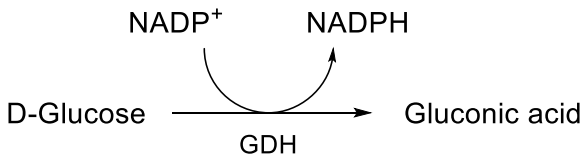



**Fig. 5.14** The formation of gluconic acid from NADPH regeneration system

## 5.5 Further optimisation

The reaction conditions were further investigated to improve the efficiency of valencene oxidation at 100 mL scale. An increase in the ratio of HPβCD to valencene to 2:1 was found to greatly increase the conversion and product titre at 100 mL scale. When the substrate was added in a single aliquot, >90% conversion of 10 mM substrate by 3 μM RKA was reliably obtained. Other changes to optimize the reaction system included the buffer strength, different concentrations of cofactor recycling system components, adding small amounts of PEG-400 in the mixture to increase the solubility of valencene,

more intensive mass transfer by increasing the stirring or shaking speed to increase the dissolved oxygen level, use of compressed air rather than pure oxygen, and further purification of the enzyme to improve activity, etc. The final optimised reaction conditions in 100 mL were as shown in Table 5.2

**Table 5.4** Reagents in 100 mL valencene oxidation reaction.

| Reagent | Final Concentration |
|---|---|
| $NADP^+$ | 10 μM |
| GDH | 5 U/mL |
| Glucose | 200 mM |
| Valencene | 10 mM |
| P450 enzyme | 3 μM |
| PEG-400 | 0.5% *v/v* |
| 2-Hydroxypropyl-β-cyclodextrin | 20 mM |
| Tris-HCl buffer, pH 7.8 | 50 mM |

The enzyme was reduced to 3 μM and this was added at the beginning of the reaction. The substrate was dissolved in ethanol to make a 1 M stock and it was added as aliquot to reach 10 mM final concentration. The consistent >90% conversion obtained proved that low percentage of ethanol, for example 1% *v/v*, did not affect enzyme activity. This made it possible to keep increasing the substrate concentration to test the limits of the reaction system.

Under these conditions, the reaction was scaled further to 3 litres with 30 mM valencene being oxidised by 10 μM enzyme in the presence of 60 mM HPβCD (Table 5.3). The final ethanol concentration was 3% when the substrate was added at 30 mM. The enzyme for these reactions was in the form of a cell-free extract at 50 to 100 μM

$P450_{BM3}$ stock concentration from high cell density *E. coli* fermentations and supplied by Oxford Biotrans Ltd.

**Table 5.5** Reagents in 3 L valencene oxidation reaction.

| Reagent | Final Concentration |
|---|---|
| $NADP^+$ | 20 μM |
| GDH | 10 U/mL |
| Glucose | 400 mM |
| Valencene | 30 mM |
| P450 enzyme | 10 μM |
| 2-Hydroxypropyl-β-cyclodextrin | 60 mM |
| Tris-HCl buffer, pH 7.8 | 50 mM |

Aeration by compressed air was controlled between 1 to 2 L/min and the impeller speed was set at 250 rpm. The pH adjustment was carried out using 1M KOH and 30 mM substrate and 10 μM enzyme were added at the beginning of the reaction. The NADPH regeneration system was scaled two times and glucose was increased to 400 mM. The HPβCD to substrate ratio was kept as 2:1 by using a final concentration of 60 mM. PEG-400 was no longer used in order to simplify the reaction system.

The turnover number was increased to close to 3,000 with >90% conversion and ~65% nootkatone in the products. The best published valencene oxidation activity produced 1.65 mM nootkatone from 10 mM valencene by 5 μM enzyme, a turnover number of 330.[110] The enzyme and method developed here gave dramatic improvements in conversion (>90%), product titre (>27 mM of nootkatol and nootkatone formed), and turnover number (~3,000).

**Table 5.6** Comparison between the conditions developed in this work with the best published data.

| | **Best published Result** | **This work** |
|---|---|---|
| **Reaction volume** | 10 mL | 3 L |
| **Enzyme (μM)** | 5 μM | 10 μM |
| **Substrate (mM)** | 10 mM | 30 mM |
| **Conversion** | 16.5% | >90% |
| **Yield of nootkatone (mM)** | 1.65 mM | >20 mM |

Following this work, the process was transferred to Oxford Biotrans Ltd. for pilot then commercial scale production research. This process was successfully repeated at much larger volumes and then further optimised to commercially viable production and sale of natural grade nootkatone.

# Chapter 6 Summary and Further Work

One aim of this project was to explore the oxidation of sesquiterpenes at as many positions as possible using CYP102A1 variants. The selectivity of each functionalised product would then be optimised by exploring the mutations at one residue or different mutation combinations within a promising variant discovered in initial screening. A selection of CYP102A1 variants was screened for their activity and selectivity against four sesquiterpenes, valencene, α-humulene, β-caryophyllene and α-cedrene. Variants were selected on the basis of residue location around the active site and binding pocket, and previous work in the group which identified highly active variants for unnatural substrate oxidation.

The screening results showed that, as the ring strain increased and the number of double bonds decreased along the series humulene (monocyclic), valencene and caryophyllene (bicyclic) and cedrene (tricyclic), both the activity and product diversity decreased. One variant, RKA, was found to give high valencene conversion (>90%) and high selectivity (>95%) for C2 oxidation to give a mixture of the nootkatols and nootkatone. In addition to the expected epoxide products, the substrate pocket topology of this initial library was also sufficiently diverse that some of the allylic alcohol derivatives of humulene and caryophyllene, which are difficult to synthesise by chemical methods, were readily accessible, in some cases with good selectivity. Most of these oxygenated sesquiterpenes possessed aromas that are reminiscent of varied combinations of plants and fruits. Their organoleptic properties should be evaluated by professional flavourists when sufficient quantities are these compounds become available.

A group of residues and mutations with significant effects were identified, revealing correlations between the side-chain volume of the substituted amino acid and regioselectivity of substrate oxidation. Three selectivity libraries were constructed by introducing different numbers and combinations of mutations on different base variants, and cedrene, the most inert substrate among the tested sesquiterpenes, was selected for further investigations with these libraries. A total of 84 variants with hydrophobic substitutions at three targeted residues (F87, A328 and I263) was prepared, there was clear evidence to show the manipulation of regioselectivity of cedrane oxidation.

In screening of cedrene with the initial selection of 24 variants, only products arising from oxidation at 3 sites, namely the C8=C9 double bonds and the allylic C10 and C15 C–H bonds was observed. The 3-residue library demonstrated shifts of regioselectivity from activated to unactivated C–H bonds. The A328F mutation in R19/F87A/A328F shifted the oxidation towards C2, giving 11% of the tertiary alcohol. The F87G/I263V combination in K19/F87G/I263V gave 5% of the C4 alcohol. Other products with C2, C3 and C4 oxidation were observed in further oxidation of the epoxide, the C10 and C15 allylic alcohols. Further oxidation was also observed in the conversion of the C15 alcohol to the corresponding aldehyde and carboxylic acid. The total number of products increased from 5 to 15, but no oxidation was observed at the aliphatic methyl groups, the tertiary C5 and C7 carbons, and the bridgehead C11. This approach shows the potential of significant impact on reaction selectivity with a relatively small number of variants compared to other strategies such as site-saturation mutagenesis, CAST and ISM. This “reduced library” approach could, however, fail to

cover sufficient combinations and thus not provide high selectivity; there is a trade-off between the number of mutants screened and the gain in the desired trait.

The other aim of this project is the development of scaleable processes. With this study, the reaction conditions could be optimised to increase the volumetric efficiency at laboratory scale for transfer to further scaling up of the process. P450 engineering has been extensively studied and extremely promising results obtained. However, there is still a huge gap between laboratory research and industrial production. Starting with screening of valencene oxidation to form nootkatone to identify the RKA variant as a promising biocatalyst, the process was scaled from 1 mL to 3 L and optimised to a scaleable protocol. Key parameters such as substrate dispersion and solubility, dissolved oxygen and pH, and the substrate-to-enzyme ratio were investigated at a 100 mL reaction scale. The final condition of valencene oxidation was completed at 3 L scale with 30 mM substrate fully converted. This protocol was optimised and scaled further, eventually to commercial production of natural grade nootkatone. This is the first example of fine chemical production by biotransformation catalysed by CYP102A1.

For further work, with the foundation of the scaleable process, any promising synthetic route designed via biotransformation with CYP102A1 variants could be scaled up further. Therefore, strategies of variant library design and more detailed knowledge of mutation-structure relationship are critical in further research on unnatural substrate oxidation. The 2-residue and 3-residue library consists of <100 variants and is limited by combinations of mutations at selected residues. There were mutations at other

residues that showed minor effects on selectivity in the initial screening, possibly because of the specific mutation in the initial selected variants were not optimal for the target substrate. The library should be expanded based on existing rational design strategy to search for higher selectivity for each product, and also to oxidise the remaining C–H bonds especially those of the aliphatic methyl groups in humulene, caryophyllene and cedrene. The rational design strategy and the libraries thus produced, showed potential and should be further studied and applied to the oxidation of more unnatural substrates.

# References


1. Ariens, E.J., *Trends Pharmacol. Sci.* **1993,** *14*, 68-75.
2. Rauws, A.G.; Groen, K., *Chirality* **1994,** *6*, 72-75.
3. Nadin, A.; Hattotuwagama, C.; Churcher, I., *Angew. Chem. Int. Ed. Engl.* **2012,** *51*, 1114-1122.
4. Reetz, M.T., *J. Am. Chem. Soc.* **2013,** *135*, 12480-12496.
5. Gal, J., *Chirality* **2008,** *20*, 5-19.
6. Renata, H.; Wang, Z.J.; Arnold, F.H., *Angew. Chem. Int. Ed. Engl.* **2015,** *54*, 3351-3367.
7. Bornscheuer, U.T.; Huisman, G.W.; Kazlauskas, R.J.; Lutz, S.; Moore, J.C.; Robins, K., *Nature* **2012,** *485*, 185-194.
8. Delcour, A.H., *Biochim. Biophys. Acta* **2009,** *1794*, 808-816.
9. Chartrain, M.; Greasham, R.; Moore, J.; Reider, P.; Robinson, D.; Buckland, B., *Journal of Molecular Catalysis B-Enzymatic* **2001,** *11*, 503-512.
10. Liang, J.; Mundorff, E.; Voladri, R.; Jenne, S.; Gilson, L.; Conway, A.; Krebber, A.; Wong, J.; Huisman, G.; Truesdell, S.; Lalonde, J., *Organic Process Research & Development* **2010,** *14*, 188-192.
11. Michel L. Belley, S.L., Marc Labelle, Patrick Roy, Yi B. Xiang, Daniel Guay US5565473A, 1996.
12. Liang, J.; Lalonde, J.; Borup, B.; Mitchell, V.; Mundorff, E.; Trinh, N.; Kochrekar, D.A.; Cherat, R.N.; Pai, G.G., *Organic Process Research & Development* **2010,** *14*, 193-198.
13. Weber, A.E., *J. Med. Chem.* **2004,** *47*, 4135-4141.
14. Thornberry, N.A.; Weber, A.E., *Curr. Top. Med. Chem.* **2007,** *7*, 557-568.
15. Hansen, K.B.; Balsells, J.; Dreher, S.; Hsiao, Y.; Kubryk, M.; Palucki, M.; Rivera, N.; Steinhuebel, D.; Armstrong, J.D.; Askin, D.; Grabowski, E.J.J., *Organic Process Research & Development* **2005,** *9*, 634-639.
16. Hansen, K.B.; Hsiao, Y.; Xu, F.; Rivera, N.; Clausen, A.; Kubryk, M.; Krska, S.; Rosner, T.; Simmons, B.; Balsells, J.; Ikemoto, N.; Sun, Y.; Spindler, F.; Malan, C.; Grabowski, E.J.; Armstrong, J.D., 3rd, *J. Am. Chem. Soc.* **2009,** *131*, 8798-8804.
17. Savile, C.K.; Janey, J.M.; Mundorff, E.C.; Moore, J.C.; Tam, S.; Jarvis, W.R.; Colbeck, J.C.; Krebber, A.; Fleitz, F.J.; Brands, J.; Devine, P.N.; Huisman, G.W.; Hughes, G.J., *Science* **2010,** *329*, 305-309.

18. Iwasaki, A.; Yamada, Y.; Kizaki, N.; Ikenaka, Y.; Hasegawa, J., *Appl. Microbiol. Biotechnol.* **2006,** *69*, 499-505.

19. Urlacher, V.; Schmid, R.D., *Curr. Opin. Biotechnol.* **2002,** *13*, 557-564.

20. Fer, M.; Dreano, Y.; Lucas, D.; Corcos, L.; Salaun, J.P.; Berthou, F.; Amet, Y., *Arch. Biochem. Biophys.* **2008,** *471*, 116-125.

21. Yueh, M.F.; Nguyen, N.; Famourzadeh, M.; Strassburg, C.P.; Oda, Y.; Guengerich, F.P.; Tukey, R.H., *Carcinogenesis* **2001,** *22*, 943-950.

22. Moody, S.C.; Zhao, B.; Lei, L.; Nelson, D.R.; Mullins, J.G.; Waterman, M.R.; Kelly, S.L.; Lamb, D.C., *FEBS J.* **2012,** *279*, 1640-1649.

23. Butler, C.F.; Peet, C.; McLean, K.J.; Baynham, M.T.; Blankley, R.T.; Fisher, K.; Rigby, S.E.; Leys, D.; Voice, M.W.; Munro, A.W., *Biochem. J* **2014,** *460*, 247-259.

24. Hasler, J.A.; Estabrook, R.; Murray, M.; Pikuleva, I.; Waterman, M.; Capdevila, J.; Holla, V.; Helvig, C.; Falck, J.R.; Farrell, G.; Kaminsky, L.S.; Spivack, S.D.; Boitier, E.; Beaune, P., *Mol. Aspects Med.* **1999,** *20*, 1-137.

25. Payne, A.H.; Hales, D.B., *Endocr. Rev.* **2004,** *25*, 947-970.

26. Jacobs, E.T.; Van Pelt, C.; Forster, R.E.; Zaidi, W.; Hibler, E.A.; Galligan, M.A.; Haussler, M.R.; Jurutka, P.W., *Cancer Res.* **2013,** *73*, 2563-2573.

27. Lamb, D.C.; Ikeda, H.; Nelson, D.R.; Ishikawa, J.; Skaug, T.; Jackson, C.; Omura, S.; Waterman, M.R.; Kelly, S.L., *Biochem. Biophys. Res. Commun.* **2003,** *307*, 610-619.

28. Nelson, D.R.; Kamataki, T.; Waxman, D.J.; Guengerich, F.P.; Estabrook, R.W.; Feyereisen, R.; Gonzalez, F.J.; Coon, M.J.; Gunsalus, I.C.; Gotoh, O.; et al., *DNA Cell Biol.* **1993,** *12*, 1-51.

29. Rittle, J.; Green, M.T., *Science* **2010,** *330*, 933-937.

30. Newcomb, M.; Zhang, R.; Chandrasena, R.E.; Halgrimson, J.A.; Horner, J.H.; Makris, T.M.; Sligar, S.G., *J. Am. Chem. Soc.* **2006,** *128*, 4580-4581.

31. Krest, C.M.; Onderko, E.L.; Yosca, T.H.; Calixto, J.C.; Karp, R.F.; Livada, J.; Rittle, J.; Green, M.T., *J. Biol. Chem.* **2013,** *288*, 17074-17081.

32. Whitehouse, C.J.; Bell, S.G.; Wong, L.L., *Chem. Soc. Rev.* **2012,** *41*, 1218-1260.

33. Trower, M.K.; Emptage, M.H.; Sariaslani, F.S., *Biochim. Biophys. Acta* **1990,** *1037*, 281-289.

34. Hanukoglu, I.; Jefcoate, C.R., *J. Biol. Chem.* **1980,** *255*, 3057-3061.

35. Gunsalus, I.C.; Wagner, G.C., *Methods Enzymol.* **1978,** *52*, 166-188.

36. Fleming, B.D.; Tian, Y.; Bell, S.G.; Wong, L.L.; Urlacher, V.; Hill, H.A., *Eur. J. Biochem.* **2003,** *270*, 4082-4088.

37. Roberts, G.A.; Grogan, G.; Greter, A.; Flitsch, S.L.; Turner, N.J., *J. Bacteriol.* **2002,** *184*, 3898-3908.

38. Miura, Y.; Fulco, A.J., *Biochim. Biophys. Acta* **1975,** *388*, 305-317.

39. Narhi, L.O.; Fulco, A.J., *J. Biol. Chem.* **1987,** *262*, 6683-6690.

40. Black, S.D.; Martin, S.T., *Biochemistry* **1994,** *33*, 12056-12062.

41. Jamakhandi, A.P.; Jeffus, B.C.; Dass, V.R.; Miller, G.P., *Arch. Biochem. Biophys.* **2005,** *439*, 165-174.

42. Zhang, H.; Yokom, A.L.; Cheng, S.; Su, M.; Hollenberg, P.F.; Southworth, D.R.; Osawa, Y., *J. Biol. Chem.* **2018,** *293*, 7727-7736.

43. Li, H.; Poulos, T.L., *Biochim. Biophys. Acta* **1999,** *1441*, 141-149.

44. Li, H.; Poulos, T.L., *Nat. Struct. Biol.* **1997,** *4*, 140-146.

45. Bruhlmann, F.; Fourage, L.; Ullmann, C.; Haefliger, O.P.; Jeckelmann, N.; Dubois, C.; Wahler, D., *J. Biotechnol.* **2014,** *184*, 17-26.

46. Whitehouse, C.J.; Yang, W.; Yorke, J.A.; Rowlatt, B.C.; Strong, A.J.; Blanford, C.F.; Bell, S.G.; Bartlam, M.; Wong, L.L.; Rao, Z., *ChemBioChem* **2010,** *11*, 2549-2556.

47. Haines, D.C.; Tomchick, D.R.; Machius, M.; Peterson, J.A., *Biochemistry* **2001,** *40*, 13456-13465.

48. Pflug, S.; Richter, S.M.; Urlacher, V.B., *J. Biotechnol.* **2007,** *129*, 481-488.

49. Kaluzna, I.; Schmitges, T.; Straatman, H.; van Tegelen, D.; Müller, M.; Schürmann, M.; Mink, D., *Organic Process Research & Development* **2016,** *20*, 814-819.

50. Li, Q.-S.; Schwaneberg, U.; Fischer, P.; Schmid, R.D., *Eur. J. Biochem.* **2000,** *6*, 1531-1536.

51. Jung, S.T.; Lauchli, R.; Arnold, F.H., *Curr. Opin. Biotechnol.* **2011,** *22*, 809-817.

52. Graham-Lorence, S.; Truan, G.; Peterson, J.A.; Falck, J.R.; Wei, S.; Helvig, C.; Capdevila, J.H., *J. Biol. Chem.* **1997,** *272*, 1127-1135.

53. Weber, E.; Seifert, A.; Antonovici, M.; Geinitz, C.; Pleiss, J.; Urlacher, V.B., *ChemComm* **2011,** *47*, 944-946.

54. Damsten, M.C.; van Vugt-Lussenburg, B.M.; Zeldenthuis, T.; de Vlieger, J.S.; Commandeur, J.N.; Vermeulen, N.P., *Chem. Biol. Interact.* **2008,** *171*, 96-107.

55. Noble, M.A.; Miles, C.S.; Chapman, S.K.; Lysek, D.A.; MacKay, A.C.; Reid, G.A.; Hanzlik, R.P.; Munro, A.W., *Biochem. J* **1999,** *339 ( Pt 2)*, 371-379.

56. Carmichael, A.B.; Wong, L.L., *Eur. J. Biochem.* **2001,** *268*, 3117-3125.

57. Sowden, R.J.; Yasmin, S.; Rees, N.H.; Bell, S.G.; Wong, L.L., *Org. Biomol. Chem.* **2005,** *3*, 57-64.

58. Huang, W.C.; Westlake, A.C.; Marechal, J.D.; Joyce, M.G.; Moody, P.C.; Roberts, G.C., *J. Mol. Biol.* **2007,** *373*, 633-651.

59. Glieder, A.; Farinas, E.T.; Arnold, F.H., *Nat. Biotechnol.* **2002,** *20*, 1135-1139.

60. Peters, M.W.; Meinhold, P.; Glieder, A.; Arnold, F.H., *J. Am. Chem. Soc.* **2003,** *125*, 13442-13450.

61. Meinhold, P.; Peters, M.W.; Chen, M.M.; Takahashi, K.; Arnold, F.H., *ChemBioChem* **2005,** *6*, 1765-1768.

62. Fasan, R.; Chen, M.M.; Crook, N.C.; Arnold, F.H., *Angew. Chem. Int. Ed. Engl.* **2007,** *46*, 8414-8418.

63. Landwehr, M.; Hochrein, L.; Otey, C.R.; Kasrayan, A.; Backvall, J.E.; Arnold, F.H., *J. Am. Chem. Soc.* **2006,** *128*, 6058-6059.

64. Otey, C.R.; Bandara, G.; Lalonde, J.; Takahashi, K.; Arnold, F.H., *Biotechnol. Bioeng.* **2006,** *93*, 494-499.

65. Sawayama, A.M.; Chen, M.M.; Kulanthaivel, P.; Kuo, M.S.; Hemmerle, H.; Arnold, F.H., *Chemistry (Easton)* **2009,** *15*, 11723-11729.

66. Rentmeister, A.; Brown, T.R.; Snow, C.D.; Carbone, M.N.; Arnold, F.H., *ChemCatChem* **2011,** *3*, 1065-1071.

67. Reinen, J.; Ferman, S.; Vottero, E.; Vermeulen, N.P.; Commandeur, J.N., *J. Biomol. Screen.* **2011,** *16*, 239-250.

68. van Vugt-Lussenburg, B.M.; Damsten, M.C.; Maasdijk, D.M.; Vermeulen, N.P.; Commandeur, J.N., *Biochem. Biophys. Res. Commun.* **2006,** *346*, 810-818.

69. van Vugt-Lussenburg, B.M.; Stjernschantz, E.; Lastdrager, J.; Oostenbrink, C.; Vermeulen, N.P.; Commandeur, J.N., *J. Med. Chem.* **2007,** *50*, 455-461.

70. Reinen, J.; van Leeuwen, J.S.; Li, Y.; Sun, L.; Grootenhuis, P.D.; Decker, C.J.; Saunders, J.; Vermeulen, N.P.; Commandeur, J.N., *Drug Metab. Dispos.* **2011,** *39*, 1568-1576.

71. Rea, V.; Kolkman, A.J.; Vottero, E.; Stronks, E.J.; Ampt, K.A.; Honing, M.; Vermeulen, N.P.; Wijmenga, S.S.; Commandeur, J.N., *Biochemistry* **2012,** *51*, 750-760.

72. Venkataraman, H.; Beer, S.B.; Bergen, L.A.; Essen, N.; Geerke, D.P.; Vermeulen, N.P.; Commandeur, J.N., *ChemBioChem* **2012,** *13*, 520-523.

73. Whitehouse, C.J.; Bell, S.G.; Tufton, H.G.; Kenny, R.J.; Ogilvie, L.C.; Wong, L.L., *Chem. Commun. (Camb.)* **2008,** *0*, 966-968.

74. Whitehouse, C.J.; Yang, W.; Yorke, J.A.; Tufton, H.G.; Ogilvie, L.C.; Bell, S.G.; Zhou, W.; Bartlam, M.; Rao, Z.; Wong, L.L., *Dalton Trans* **2011,** *40*, 10383-10396.

75. Whitehouse, C.J.; Bell, S.G.; Yang, W.; Yorke, J.A.; Blanford, C.F.; Strong, A.J.; Morse, E.J.; Bartlam, M.; Rao, Z.; Wong, L.L., *ChemBioChem* **2009,** *10*, 1654-1656.

76. Sevrioukova, I.F.; Li, H.; Zhang, H.; Peterson, J.A.; Poulos, T.L., *Proc. Natl. Acad. Sci. U.S.A* **1999,** *96*, 1863-1868.

77. Reetz, M.T.; Bocola, M.; Carballeira, J.; Zha, D.; Vogel, A., *Angew. Chem.* **2005,** *117*, 4264-4268.

78. Reetz, M.T.; Carballeira, J.D., *Nat. Protoc.* **2007,** *2*, 891-903.

79. Kille, S.; Zilly, F.E.; Acevedo, J.P.; Reetz, M.T., *Nat. Chem.* **2011,** *3*, 738-743.

80. Acevedo-Rocha, C.G.; Gamble, C.; Lonsdale, R.; Li, A.; Nett, N.; Hoebenreich, S.; Lingnau, J.B.; Wirtz, C.; Farès, C.; Hinrichs, H.; Deege, A.; Mulholland, A.J.; Nov, Y.; Leys, D.; McLean, K.J.; Munro, A.W.; Reetz, M.T., *ACS Catalysis* **2018,** *8*, 3395-3410.

81. Gotoh, O., *J. Biol. Chem.* **1992,** *267*, 83-90.

82. Seifert, A.; Vomund, S.; Grohmann, K.; Kriening, S.; Urlacher, V.B.; Laschat, S.; Pleiss, J., *ChemBioChem* **2009,** *10*, 853-861.

83. Zhang, K.; Shafer, B.M.; Demars, M.D., 2nd; Stern, H.A.; Fasan, R., *J. Am. Chem. Soc.* **2012,** *134*, 18695-18704.

84. Ren, X.; Yorke, J.A.; Taylor, E.; Zhang, T.; Zhou, W.; Wong, L.L., *Chemistry (Easton)* **2015,** *21*, 15039-15047.

85. O'Hanlon, J.A.; Ren, X.; Morris, M.; Wong, L.L.; Robertson, J., *Org. Biomol. Chem.* **2017,** *15*, 8780-8787.

86. Ren, X.; O'Hanlon, J.A.; Morris, M.; Robertson, J.; Wong, L.L., *ACS Catalysis* **2016,** *6*, 6833-6837.

87. Croteau, R.; Kutchan, T.M.; Lewis, N.G., American Society of Plant Physiologists: Maryland, 2000.

88. Fernandes, E.S.; Passos, G.F.; Medeiros, R.; da Cunha, F.M.; Ferreira, J.; Campos, M.M.; Pianowski, L.F.; Calixto, J.B., *Eur. J. Pharmacol.* **2007,** *569*, 228-236.

89. Fatima, A.; Abdul, A.B.; Abdullah, R.; Karjiban, R.A.; Lee, V.S., *Int. J. Mol. Sci.* **2015,** *16*, 2747-2766.

90. Schifrin, A.; Litzenburger, M.; Ringle, M.; Ly, T.T.; Bernhardt, R., *ChemBioChem* **2015,** *16*, 2624-2632.

91. Shamoto, T.; Matsuo, Y.; Shibata, T.; Tsuboi, K.; Nagasaki, T.; Takahashi, H.; Funahashi, H.; Okada, Y.; Takeyama, H., *Pancreas* **2014,** *43*, 396-404.

92. Kim, M.; Miyamoto, S.; Yasui, Y.; Oyama, T.; Murakami, A.; Tanaka, T., *Int. J. Cancer* **2009,** *124*, 264-271.

93. Takada, Y.; Murakami, A.; Aggarwal, B.B., *Oncogene* **2005,** *24*, 6957-6969.

94. Cankar, K.; van Houwelingen, A.; Goedbloed, M.; Renirie, R.; de Jong, R.M.; Bouwmeester, H.; Bosch, D.; Sonke, T.; Beekwilder, J., *FEBS Lett.* **2014,** *588*, 1001-1007.

95. Ro, D.K.; Paradise, E.M.; Ouellet, M.; Fisher, K.J.; Newman, K.L.; Ndungu, J.M.; Ho, K.A.; Eachus, R.A.; Ham, T.S.; Kirby, J.; Chang, M.C.; Withers, S.T.; Shiba, Y.; Sarpong, R.; Keasling, J.D., *Nature* **2006,** *440*, 940-943.

96. Whitehouse, C.J.C. University of Oxford, 2009.

97. Zheng, L.; Baumann, U.; Reymond, J.L., *Nucleic Acids Res.* **2004,** *32*, e115.

98. Omura, T.; Chem, S.-R.J., *J. biol. Chem* **1962,** *237*, 1375-1376.

99. Omura, T.; Sato, R., *J. Biol. Chem.* **1964,** *239*, 2370-2378.

100. Steenackers, B.; De Cooman, L.; De Vos, D., *Food Chem.* **2015,** *172*, 742-756.

101. Goiris, K.; De Ridder, M.; De Rouck, G.; Boeykens, A.; Van Opstaele, F.; Aerts, G.; De Cooman, L.; De Keukeleire, D., *J. Inst. Brew.* **2002,** *108*, 86-93.

102. Lam, K.C.; Foster, R.T.; Deinzer, M.L., *J. Agric. Food. Chem.* **1986,** *34*, 763-770.

103. Bahi, A.; Al Mansouri, S.; Al Memari, E.; Al Ameri, M.; Nurulain, S.M.; Ojha, S., *Physiol. Behav.* **2014,** *135*, 119-124.

104. Ghelardini, C.; Galeotti, N.; Di Cesare Mannelli, L.; Mazzanti, G.; Bartolini, A., *Farmaco* **2001,** *56*, 387-389.

105. Sung, B.; Prasad, S.; Yadav, V.R.; Aggarwal, B.B., *Nutr. Cancer* **2012,** *64*, 173-197.

106. Sithara, T.; Dhanya, B.P.; Arun, K.B.; Sini, S.; Dan, M.; Kokkuvayil Vasu, R.; Nisha, P., *J. Agric. Food. Chem.* **2018,** *66*, 602-612.

107. Kitayama, T.; Okamoto, T.; Hill, R.K.; Kawai, Y.; Takahashi, S.; Yonemori, S.; Yamamoto, Y.; Ohe, K.; Uemura, S.; Sawada, S., *J. Org. Chem.* **1999,** *64*, 2667-2672.

108. Ohe, K.; Miki, K.; Yanagi, S.; Tanaka, T.; Sawada, S.; Uemura, S., *Perkin Trans. 1* **2000**, 3627-3634.

109. Park, S.H.; Kim, D.H.; Kim, D.; Kim, D.H.; Jung, H.C.; Pan, J.G.; Ahn, T.; Kim, D.; Yun, C.H., *Drug Metab. Dispos.* **2010,** *38*, 732-739.

110. Schulz, S.; Girhard, M.; Gaßmeyer, S.K.; Jäger, V.D.; Schwarze, D.; Vogel, A.; Urlacher, V.B., *ChemCatChem* **2015,** *7*, 601-604.

111. Adams, R.P., *Econ. Bot.* **1987,** *41*, 48-54.

112. Adams, R.P., Essential Oils and Waxes, 1991; p 159-173.

113. Schatkowski, W.P.D. 1999.

114. Tomita, B.; Hirose, Y.; Nakatsuka, T., *Tetrahedron Lett.* **1968,** *9*, 843-848.

115. Chalchat, J.C.; Carry, R.P.; Michet, A.; Lecomte, E., *Flavour Fragrance J.* **1992,** *7*, 107-110.

116. Qin, C.; Davies, H.M., *J. Am. Chem. Soc.* **2014,** *136*, 9792-9796.

117. Kitchens, G.C.; Hochstet.Ar; Kaiser, K., *J. Org. Chem.* **1972,** *37*, 6-8.

118. Green, J.C.; Pettus, T.R., *J. Am. Chem. Soc.* **2011,** *133*, 1603-1608.

119. Peter Yates and , K.E.S., *Can. J. Chem.* **1982,** *60*, 825-834.

120. Yates, P.; Burnell, D.J.; Freer, V.J.; Sawyer, J.F., *Can. J. Chem.* **1987,** *65*, 69-77.

121. Lamare, V.; Fourneron, J.D.; Furstoss, R.; Ehret, C.; Corbier, B., *Tetrahedron Lett.* **1987,** *28*, 6269-6272.

122. Abraham, W.R.; Washausen, P.; Kieslich, K., *Z. Naturforsch. C Bio. Sc* **1987,** *42*, 414-419.

123. Takigawa, H.; Kubota, H.; Sonohara, H.; Okuda, M.; Tanaka, S.; Fujikura, Y.; Ito, S., *Appl. Environ. Microbiol.* **1993,** *59*, 1336-1341.

124. Erdtman, H.; Hirose, Y.; Theander, O., *Acta Chem. Scand.* **1962,** *16*, 1311-1314.

125. MacLeod, W.D.; Buigues, N.M., *J. Food Sci.* **1964,** *29*, 565-568.

126. Stevens, K.L.; Guadagni, D.G.; Stern, D.J., *J. Sci. Food Agric.* **1970,** *21*, 590-593.

127. Haring, H.G.; Rijkens, F.; Boelens, H.; Van der Gen, A., *J. Agric. Food. Chem.* **1972,** *20*, 1018-1021.

128. Friedman, L.; Miller, J.G., *Science* **1971,** *172*, 1044-1046.

129. Marshall, J.A.; Faubl, H.; Warne, T.M., *ChemComm* **1967,** *0*, 753-754.

130. Pesaro, M.; Bozzato, G.; Schudel, P., *ChemComm* **1968**, 1152-1154.

131. Dastur, K.P., *J. Am. Chem. Soc.* **1973,** *95*, 6509-6510.

132. Hunter, G.L.K.; Brogden, W.B., *J. Food Sci.* **1965,** *30*, 876-878.

133. Salvador, J.A.R.; Clark, J.H., *Green Chem.* **2002,** *4*, 352-356.

134. Drawert, F.; Berger, R.G.; Godelmann, R., *Plant Cell Rep.* **1984,** *3*, 37-40.

135. Sakamaki, H.; Itoh, K.; Taniai, T.; Kitanaka, S.; Takagi, Y.; Chai, W.; Horiuchi, C.A., *J Mol Catal B Enzym.* **2005,** *32*, 103-106.

136. Palmerin-Carreno, D.M.; Rutiaga-Quinones, O.M.; Calvo, J.R.V.; Prado-Barragan, A.; Huerta-Ochoa, S., *LWT.* **2015,** *64*, 788-793.

137. Palmerín-Carreño, D.M.; Castillo-Araiza, C.O.; Rutiaga-Quiñones, O.M.; Verde Calvo, J.R.; Trejo-Aguilar, G.M.; Dutta, A.; Huerta-Ochoa, S., *J. Chem. Technol. Biotechnol.* **2016,** *91*, 1164-1172.

138. Kaspera, R.; Krings, U.; Nanzad, T.; Berger, R.G., *Appl. Microbiol. Biotechnol.* **2005,** *67*, 477-483.

139. Fraatz, M.A.; Riemer, S.J.L.; Stober, R.; Kaspera, R.; Nimtz, M.; Berger, R.G.; Zorn, H., *J Mol Catal B Enzym.* **2009,** *61*, 202-207.

140. Girhard, M.; Machida, K.; Itoh, M.; Schmid, R.D.; Arisawa, A.; Urlacher, V.B., *Microb Cell Fact* **2009,** *8*, 36.

141. Kuper, J.; Tee, K.L.; Wilmanns, M.; Roccatano, D.; Schwaneberg, U.; Wong, T.S., *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2012,** *68*, 1013-1017.

142. Khalafi, L.; Rafiee, M., Gandhi Rádis Baptista, IntechOpen: 2013.

143. Davis, M.E.; Brewster, M.E., *Nat. Rev. Drug Discov.* **2004,** *3*, 1023-1035.

144. Klein, U.; Gimpl, G.; Fahrenholz, F., *Biochemistry* **1995,** *34*, 13784-13793.

145. Pitha, J.; Milecki, J.; Fales, H.; Pannell, L., *Int. J. Pharm.* **1986,** *29*, 73-82.

146. Pronk, J.T.; Levering, P.R.; Olijve, W.; Vandijken, J.P., *Enzyme Microb. Technol.* **1989,** *11*, 160-164.

# Appendix

## Appendix 1 Variants and mutations

**Appendix 1.1** Mutations in base variants

| Variants | Mutations |
|---|---|
| RLYF | R47L/Y51F |
| RP | R47L/Y51F/I401P |
| KT2 | A191T/N239H/I259V/A276T/L353I |
| RT2 | R47L/Y51F/A191T/N239H/I259V/A276T/L353I |
| GVQ | A74G/F87V/L188Q |
| K19 | H171L/Q307H/N319Y |
| R19 | R47L/Y51F/H171L/Q307H/N319Y |

**Appendix 1.2** Variants in R19 based library

| F87 / A328 | F87 | F87A | F87V | F87G | F87I | F87L |
|---|---|---|---|---|---|---|
| **A328** | R19 | R19/F87A | R19/F87V | R19/F87G | R19/F87I | R19/F87L |
| **A328I** | R19/A328I | R19/F87A/ A328I | R19/F87V/ A328I | R19/F87G/ A328I | R19/F87I/ A328I | R19/F87L/ A328I |
| **A328V** | R19/A328V | R19/F87A/ A328V | R19/F87V/ A328V | R19/F87G/ A328V | R19/F87I/ A328V | R19/F87L/ A328V |
| **A328G** | R19/A328G | R19/F87A/ A328G | R19/F87V/ A328G | R19/F87G/ A328G | R19/F87I/ A328G | R19/F87L/ A328G |
| **A328L** | R19/A328L | R19/F87A/ A328L | R19/F87V/ A328L | R19/F87G/ A328L | R19/F87I/ A328L | R19/F87L/ A328L |
| **A328F** | R19/A328F | R19/F87A/ A328F | R19/F87V/ A328F | R19/F87G/ A328F | R19/F87I/ A328F | R19/F87L/ A328F |

**Appendix 1.3** Variants in K19 based library containing mutations of F87 and I263

| F87 / I263 | F87 | F87A | F87V | F87G | F87I | F87L |
|---|---|---|---|---|---|---|
| **I263** | K19 | K19/F87A | K19/F87V | K19/F87G | K19/F87I | K19/F87L |
| **I263A** | K19/ I263A | K19/F87A/ I263A | K19/F87V/ I263A | K19/F87G/ I263A | K19/F87I/ I263A | K19/F87L/ I263A |
| **I263V** | K19/ I263V | K19/F87A/ I263V | K19/F87V/ I263V | K19/F87G/ I263V | K19/F87I/ I263V | K19/F87L/ I263V |
| **I263G** | K19/A328G | K19/F87A/ I263G | K19/F87V/ I263G | K19/F87G/ I263G | K19/F87I/ I263G | K19/F87L/ I263G |

**Appendix 1.4** Variants in GVQ based library containing mutations of A328 and I263 with the existence of F87V mutation

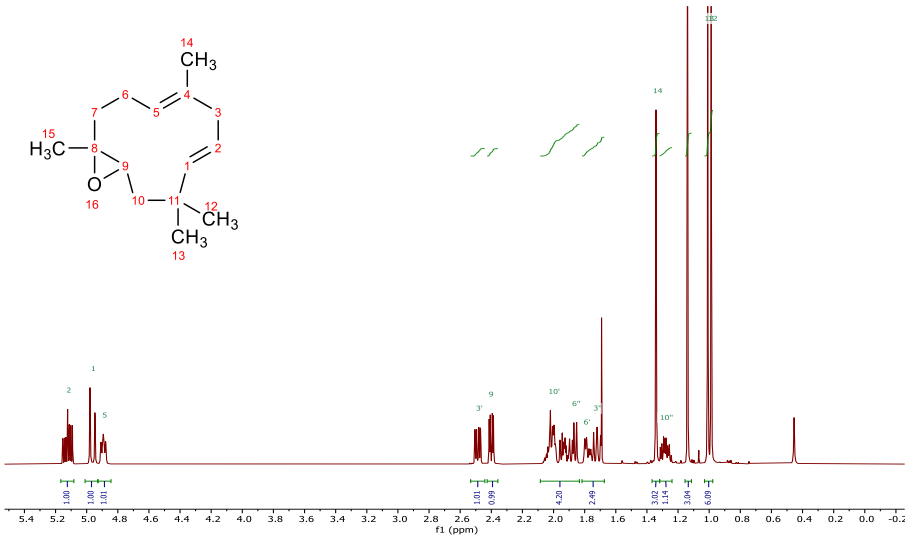
| A328 / I263 | A328 | A328I | A328G | A328V | A328L |
|---|---|---|---|---|---|
| **I263** | GVQ | GVQ/A328I | GVQ/A328G | GVQ/A328V | GVQ/A328L |
| **I263A** | GVQ/I263A | GVQ/A328I/ I263A | GVQ/A328G/ I263A | GVQ/A328V/ I263A | GVQ/A328L/ I263A |
| **I263V** | GVQ/I263V | GVQ/A328I/ I263V | GVQ/A328G/ I263V | GVQ/A328V/ I263V | GVQ/A328L/ I263V |
| **I263G** | GVQ/I263G | GVQ/A328I/ I263G | GVQ/A328G/ I263G | GVQ/A328V/ I263G | GVQ/A328L/ I263G |

**Appendix 1.5** Sequence of Mutations

| Mutations | Substitutions | Sequence of mutations | |
|---|---|---|---|
| R47L | Arg(R)47 → Leu(L)47 | 5'-cctggtcgtgtaacg-3' | → 5'-cctggtcTtgtaacg-3' |
| Y51F | Tyr(Y)51 → Phe(F)51 | 5'-cgccgctacttatca-3' | → 5'-cgccgctTcttatca-3' |
| A74G | Ala(A)74 → Gly(G)74 | 5'-agtcaagcgcttaaa-3' | → 5'-agtcaagGgcttaaa-3' |
| L188Q | Leu(L)188 → Gln(Q)188 | 5'-aacaagctgcagcga-3' | → 5'-aacaagcAgcagcga-3' |
| H171L | His(H)171 → Ile(L)171 | 5'-cagcctcatccattt-3' | → 5'-cagcctcTtccattt-3' |
| A191T | Ala(A)191 → Thr(T)191 | 5'-cagcgagcaaatcca-3' | → 5'-cagcgaAcaaatcca-3' |
| N239H | Asn(N)239 → His(H)239 | 5'-atgctaaacggaaaa-3' | → 5'-atgctaCacggaaaa-3' |
| I259V | Ile(I)259 → Val(V)259 | 5'caaattattacattc-3' | → 5'caaattGttacattc-3' |
| A276T | Ala(A)276 → Thr(T)276 | 5'-tcatttgcgctgtat-3' | → 5'-tcatttAcgctgtat-3' |
| Q307H | Gln(Q)307 → His(H)307 | 5'-tacaaacaagtcaaa-3' | → 5'-tacaaacaTgtcaaa-3' |
| N319Y | Asn(N)319 → Tyr(Y)319 | 5'-gtcttaaacgaagcg-3' | → 5'-gtcttaTacgaagcg-3' |
| L353I | Leu(L)353 → Ile(I)353 | 5'-gacgaactaatggtt-3' | → 5'-gacgaaAtaatggtt-3' |
| I401P | Ile(I)401 → Pro(P)401 | 5'-gcgtgtatcggtcag-3' | → 5'-gcgtgtCCcggtcag-3' |
| A330W/H/S/P | Ala(A)330 → Trp(W)330<br>His(H)330<br>Ser(S)330<br>Pro(P)330 | 5'-gctcctgcgttttcc-3' | → 5'-gctcctTGgttttcc-3'<br>5'-gctcctCATttttcc-3'<br>5'-gctcctTcgttttcc-3'<br>5'-gctcctCcgttttcc-3' |
| F87A/G/I/V/L | Phe(F)87 → Ala(A)87<br>Gly(G)87<br>Ile(I)87<br>Val(V)87<br>Leu(L)87 | 5'-gggttatttacaagc-3' | → 5'-gggttaGCtacaagc-3'<br>5'-gggttaGGtacaagc-3'<br>5'-gggttaAttacaagc-3'<br>5'-gggttaGttacaagc-3'<br>5'-gggttaCttacaagc-3' |

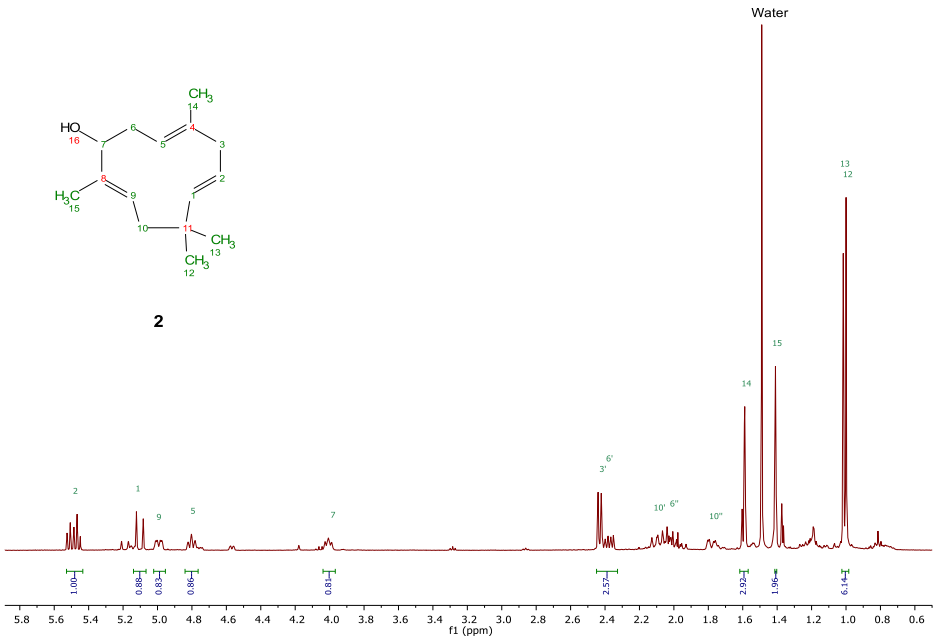
| | | | | |
|---|---|---|---|---|
| A328G/I/V/L/F | Ala(A)328 → | Gly(G)328 | 5’-ccaactgctcctgcg-3’ | → 5’-ccaactGgTcctgcg-3’ |
| | | Ile(L)328 | | 5’-ccaactATtcctgcg-3’ |
| | | Val(V)328 | | 5’-ccaactGtTcctgcg-3’ |
| | | Leu(L)328 | | 5’-ccaactCTtcctgcg-3’ |
| | | Phe(F)328 | | 5’-ccaactTTtcctgcg-3’ |
| I263A/G/V | Ile(I)263 → | Ala(A)263 | 5’-ttcttaattgcggga-3’ | → 5’-ttcttaGCtgcggga-3’ |
| | | Gly(G)263 | | 5’-ttcttaGGtgcggga-3’ |
| | | Val(V)263 | | 5’-ttcttaGttgcggga-3’ |

# Appendix 2 NMR result of products




**Appendix 2.1** The $^1$H NMR spectrum in $CDCl_3$ of the humulene-8,9-oxide, product **1.** The mass spectrum result is M+16.

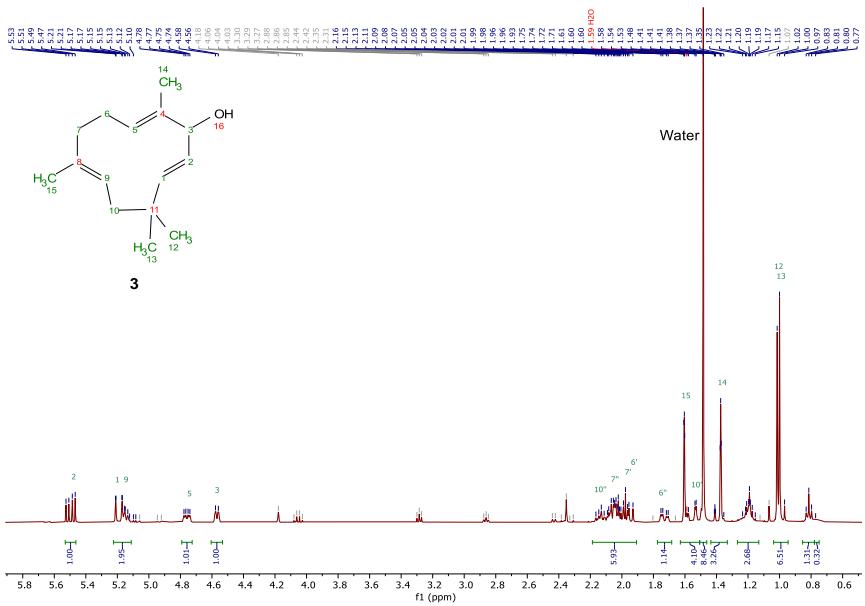
$^1$H NMR (500 MHz, $C_6D_6$) δ 5.12 (ddd, *J* = 15.6, 10.1, 5.3 Hz, 1H), 4.96 (d, *J* = 15.8 Hz, 1H), 4.93 – 4.84 (m, 1H), 2.49 (dd, *J* = 12.3, 5.3 Hz, 1H), 2.40 (dd, *J* = 10.2, 3.6 Hz, 1H), 2.07 – 1.97 (m, 2H), 1.97 – 1.90 (m, 1H), 1.87 (dd, *J* = 13.6, 8.9 Hz, 1H), 1.78 (dd, *J* = 13.8, 5.8 Hz, 1H), 1.72 (dd, *J* = 12.3, 10.1 Hz, 1H), 1.34 (t, *J* = 1.1 Hz, 3H), 1.32 – 1.24 (m, 1H), 1.14 (s, 3H), 1.00 (d, *J* = 11.1 Hz, 6H).




**2**

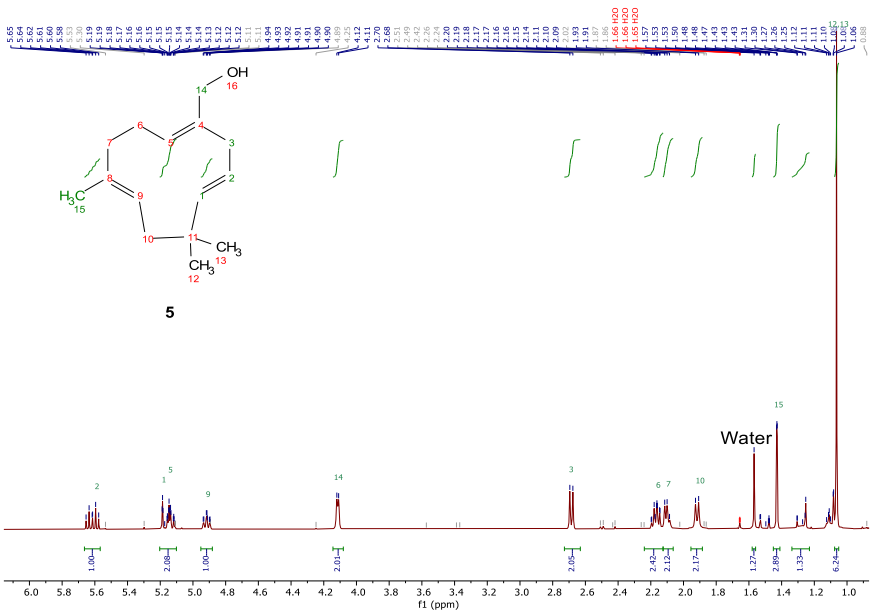
**Appendix 2.2** The 1H NMR spectrum in CDCl3 of the humulene-7-ol, product **2.** The mass spectrum result is M+16

1H NMR (400 MHz, CDCl3) δ 5.49 (dt, J = 15.8, 7.5 Hz, 1H), 5.10 (dt, J = 15.9, 0.9 Hz, 1H), 5.02 – 4.95 (m, 1H), 4.80 (t, J = 8.3 Hz, 1H), 4.01 (t, J = 7.7 Hz, 1H), 2.43 (d, J = 7.5 Hz, 2H), 2.37 (dd, J = 12.9, 6.3 Hz, 1H), 2.18 – 1.91 (m, 3H), 1.76 (td, J = 16.3, 14.9, 6.7 Hz, 1H), 1.60 (dt, J = 6.3, 1.2 Hz, 3H), 1.41 (t, J = 1.3 Hz, 2H), 1.01 (d, J = 6.5 Hz, 6H).



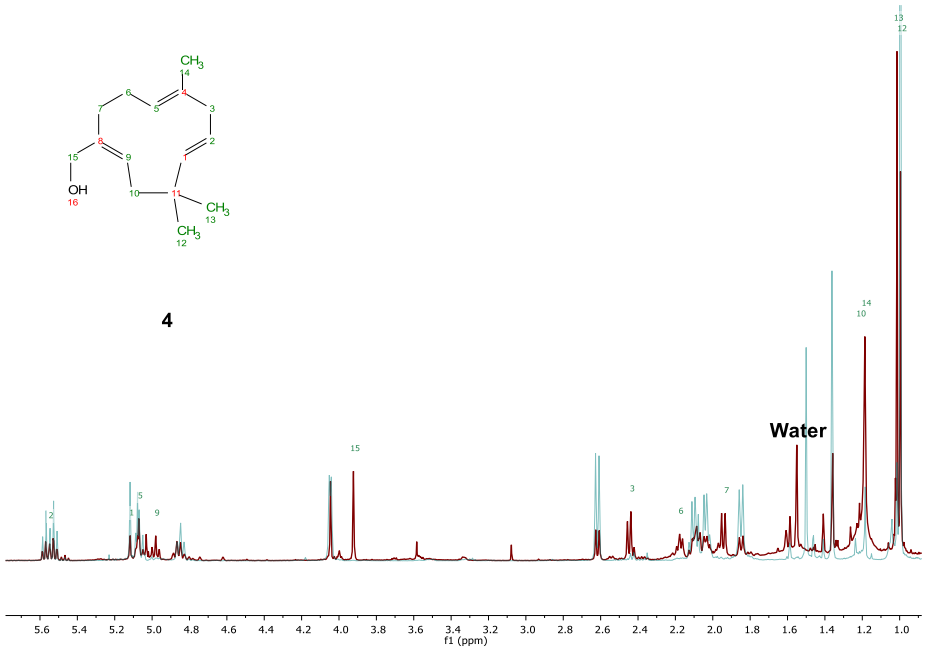

**Appendix 2.3** The 1H NMR spectrum in CDCl3 of the humulene-3-ol, product **3**. The mass spectrum result is M+16

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.50 (dd, *J* = 16.1, 7.3 Hz, 1H), 5.22 – 5.11 (m, 2H), 4.76 (dd, *J* = 10.4, 4.7 Hz, 1H), 4.57 (d, *J* = 7.3 Hz, 1H), 2.14 (q, *J* = 7.2 Hz, 1H), 2.09 – 2.01 (m, 2H), 2.01 – 1.89 (m, 2H), 1.73 (dd, *J* = 13.7, 4.7 Hz, 1H), 1.63 – 1.57 (m, 3H), 1.53 (d, *J* = 3.0 Hz, 1H), 1.37 (t, *J* = 1.3 Hz, 3H), 1.01 (d, *J* = 6.0 Hz, 6H).

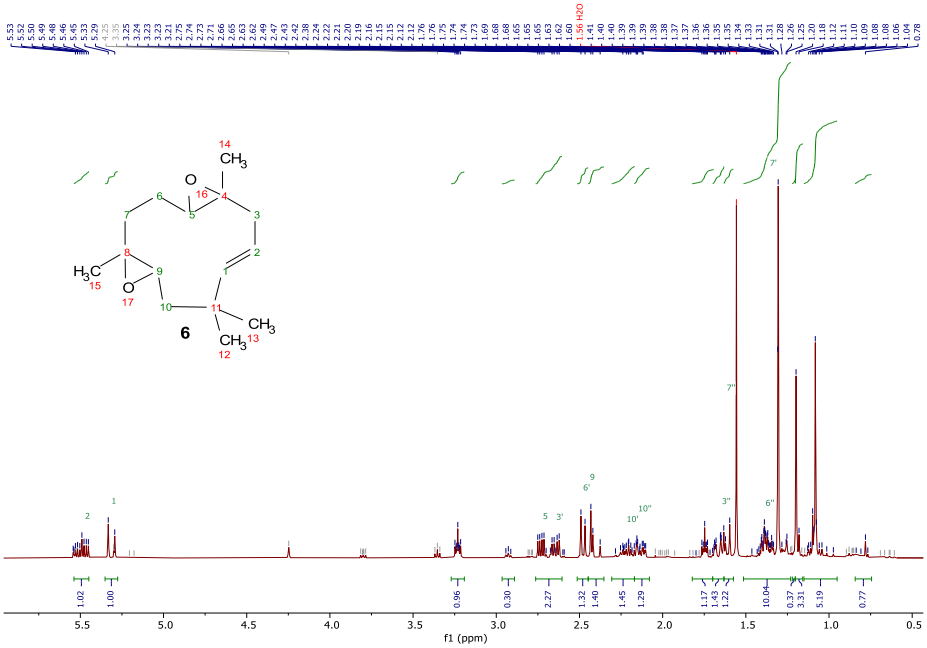



**Appendix 2.4** The 1H NMR spectrum in CDCl3 of the humulene-14-ol, product **5** The mass spectrum result is M+16

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.62 (dt, *J* = 16.0, 7.5 Hz, 1H), 5.20 – 5.10 (m, 2H), 4.92 (td, *J* = 7.6, 1.5 Hz, 1H), 4.11 (d, *J* = 4.4 Hz, 2H), 2.69 (d, *J* = 7.4 Hz, 2H), 2.20 – 2.13 (m, 2H), 2.13 – 2.07 (m, 2H), 1.92 (d, *J* = 7.7 Hz, 2H), 1.57 (s, 1H), 1.45 – 1.41 (m, 3H), 1.06 (s, 6H).

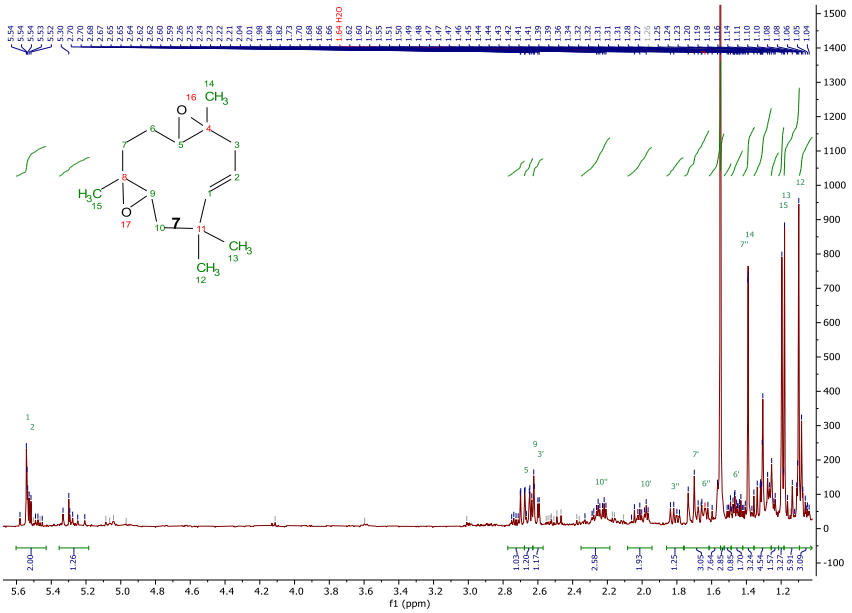



**Appendix 2.5** The 1H NMR spectrum in CDCl3 of the mixture of humulene-14-ol, product **5** and humulene-15-ol, product **4**. Product **5** could be partially isolated from the mixture by altering the ratio of hexane : ethyl acetate but not the product **4**. The 1H NMR spectrum of pure **5** (in light blue) was merged with the 1H NMR of **4** and **5** mixture (dark brown) to sketch out the proton resonance from product **4**. The mass spectrum result is M+16



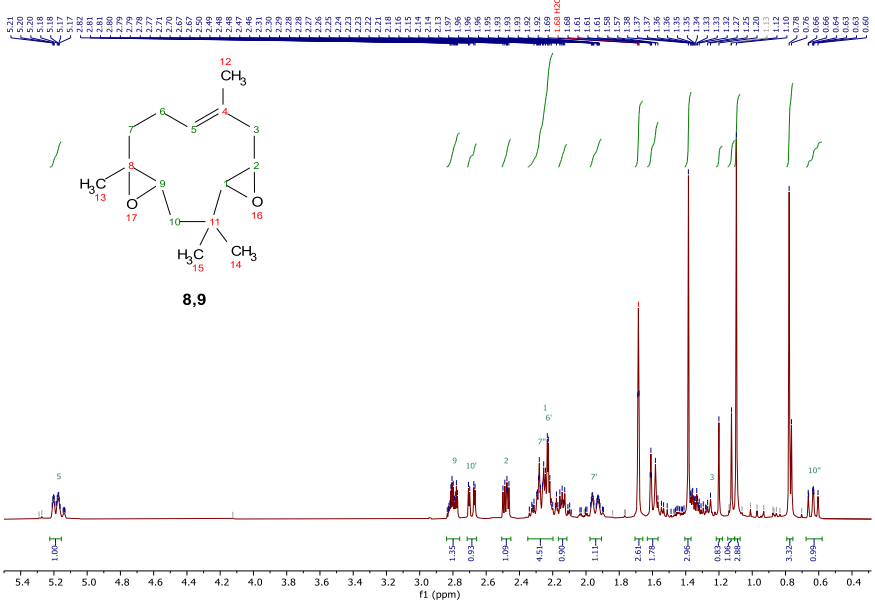

**Appendix 2.6** The 1H NMR spectrum in CDCl3 of one isomers of humulene-4,5,8,9-diepoxide isomers, product **6**. The mass spectrum result is M+32

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.50 (ddd, *J* = 20.6, 10.5, 5.2 Hz, 1H), 5.31 (d, *J* = 15.6 Hz, 1H), 2.76 – 2.60 (m, 2H), 2.48 (d, *J* = 9.9 Hz, 1H), 2.43 (d, *J* = 4.7 Hz, 1H), 2.29 – 2.17 (m, 1H), 2.17 – 2.07 (m, 1H), 1.77 – 1.72 (m, 1H), 1.70 – 1.59 (m, 3H), 1.56 (s, 4H), 1.44 – 1.32 (m, 3H), 1.31 (d, *J* = 1.1 Hz, 6H).



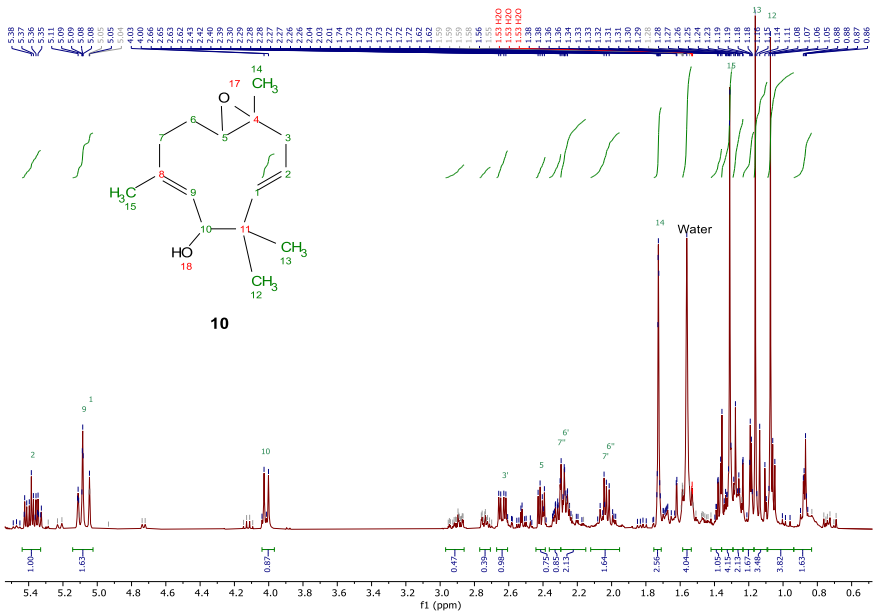

**Appendix 2.7** The 1H NMR spectrum in CDCl3 of one isomers of humulene-4,5,8,9-diepoxide isomers, product **7**. The mass spectrum result is M+32.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.60 – 5.43 (m, 2H), 5.35 – 5.18 (m, 1H), 2.77 – 2.66 (m, 1H), 2.70 – 2.62 (m, 1H), 2.66 – 2.57 (m, 2H), 2.35 – 2.19 (m, 1H), 2.08 – 1.94 (m, 1H), 1.81 (dd, *J* = 14.0, 6.9 Hz, 1H), 1.76 – 1.61 (m, 2H), 1.58 (d, *J* = 12.4 Hz, 1H), 1.55 (s, 12H), 1.53 – 1.45 (m, 1H), 1.49 – 1.40 (m, 1H), 1.39 (d, *J* = 0.8 Hz, 3H), 1.38 – 1.26 (m, 4H), 1.29 – 1.22 (m, 2H), 1.20 (s, 4H), 1.18 (s, 3H), 1.21 – 1.09 (m, 1H), 1.09 (d, *J* = 6.4 Hz, 5H), 1.09 – 0.98 (m, 1H), 0.92 – 0.83 (m, 1H), 0.83 (s, 2H), 0.82 – 0.67 (m, 1H), 0.07 (s, 3H).



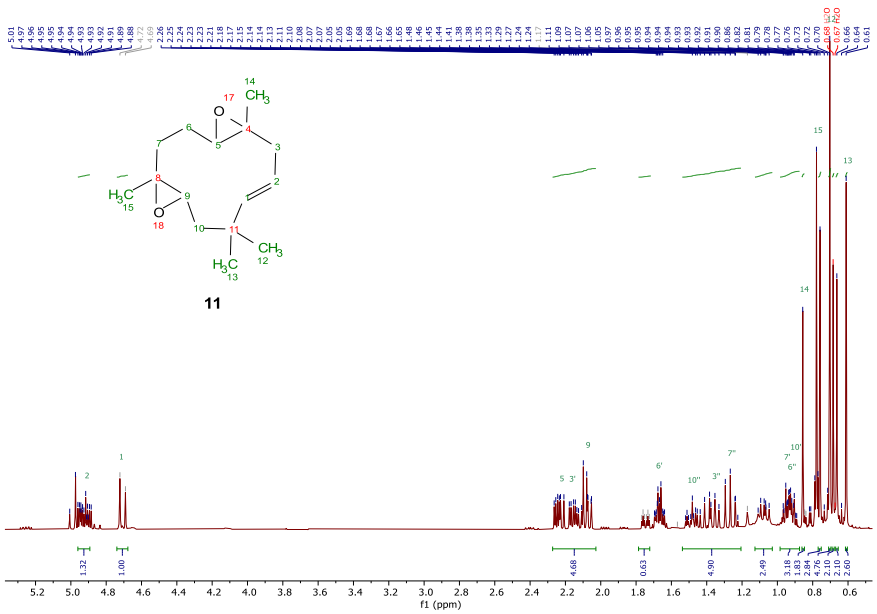

**Appendix 2.7** The 1H NMR spectrum in CDCl3 mixture of product 8 and 9 indicating they are stereoisomers of humulene-1,2,8,9-diepoxide. The mass spectrum result is M+32.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.19 (ddt, *J* = 10.7, 3.2, 1.5 Hz, 1H), 2.80 (ddt, *J* = 10.4, 5.3, 2.5 Hz, 1H), 2.69 (dd, *J* = 12.8, 3.0 Hz, 1H), 2.51 – 2.45 (m, 1H), 2.35 – 2.20 (m, 5H), 2.17 – 2.11 (m, 1H), 1.95 (ddq, *J* = 13.2, 3.1, 1.5 Hz, 1H), 1.68 (d, *J* = 3.1 Hz, 3H), 1.63 – 1.57 (m, 2H), 1.38 (s, 3H), 1.20 (s, 1H), 1.10 (s, 3H), 0.77 (d, *J* = 5.6 Hz, 3H), 0.68 – 0.58 (m, 1H).



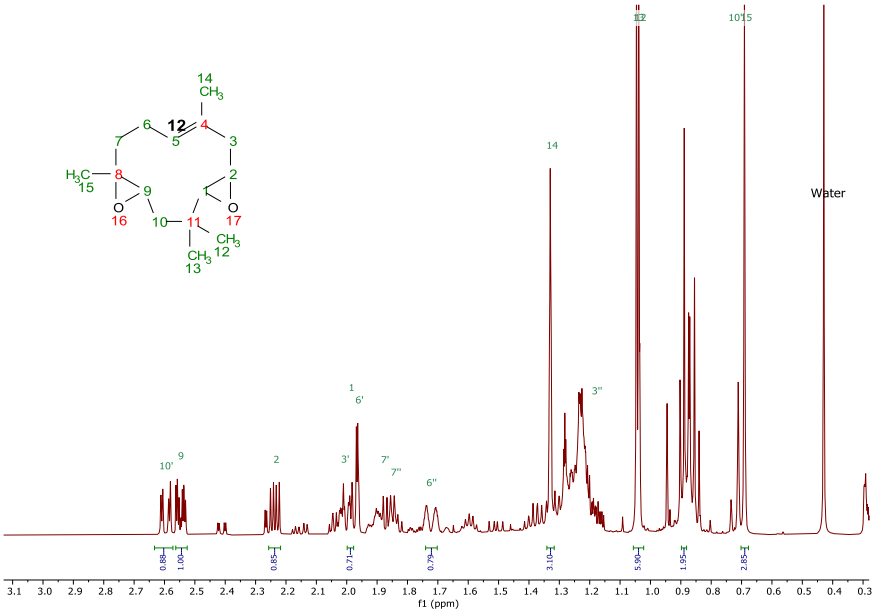

**Appendix 2.8** The 1H NMR spectrum in CDCl3 of the humulene-4,5-oxide-10-ol, product **10**. The mass spectrometry result is M+32

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.44 – 5.33 (m, 1H), 5.14 – 5.02 (m, 2H), 4.01 (d, *J* = 10.4 Hz, 1H), 2.97 – 2.86 (m, 1H), 2.64 (dd, *J* = 12.0, 4.7 Hz, 1H), 2.41 (dd, *J* = 9.9, 4.8 Hz, 1H), 2.35 – 2.20 (m, 3H), 2.09 – 1.95 (m, 1H), 1.75 – 1.71 (m, 3H), 1.31 (t, *J* = 1.7 Hz, 4H), 1.16 (s, 3H), 1.07 (d, *J* = 5.4 Hz, 3H).



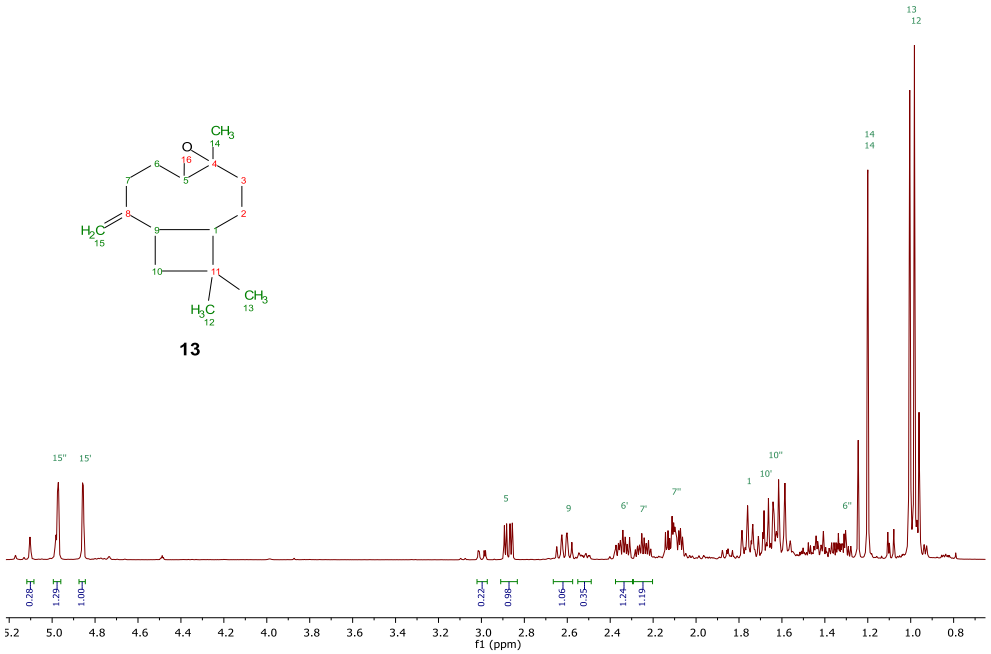

**Appendix 2.9** The 1H NMR spectrum in CDCl3 of humulene-4,5,8,9-dioxide, product **11**. The mass spectrum result is M+32.

$^1$H NMR (500 MHz, $C_6D_6$) δ 5.02 – 4.82 (m, 4H), 4.72 (s, 1H), 2.27 – 2.03 (m, 7H), 1.75 (dt, *J* = 13.8, 3.7 Hz, 1H), 1.70 – 1.63 (m, 3H), 1.54 – 1.42 (m, 3H), 1.41 – 1.32 (m, 2H), 1.32 – 1.20 (m, 2H), 1.00 – 0.88 (m, 5H), 0.86 (s, 3H), 0.76 (s, 5H), 0.70 (s, 8H), 0.66 (s, 3H), 0.61 (s, 4H).



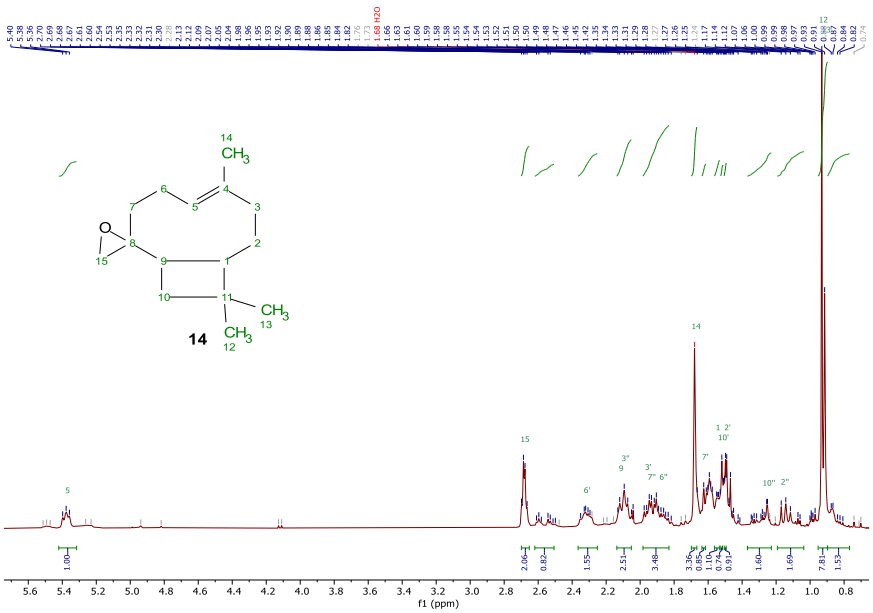

**Appendix 3.0** The 1H NMR spectrum in CDCl3 of humulene-1,2,8,9-dioxide, product **12** The mass spectrum result is M+32.

$^1$H NMR (500 MHz, $C_6D_6$) δ 4.97 (ddt, *J* = 10.5, 3.2, 1.6 Hz, 1H), 4.31 (dd, *J* = 12.7, 5.9 Hz, 1H), 2.60 (dd, *J* = 12.6, 3.0 Hz, 1H), 2.56 – 2.53 (m, 1H), 2.24 (dd, *J* = 9.6, 5.0 Hz, 1H), 2.00 – 1.98 (m, 1H), 1.74 – 1.70 (m, 1H), 1.33 (t, *J* = 1.5 Hz, 4H), 1.24 – 1.17 (m, 6H), 1.04 (d, *J* = 3.5 Hz, 7H), 0.89 (s, 2H), 0.69 (s, 3H).



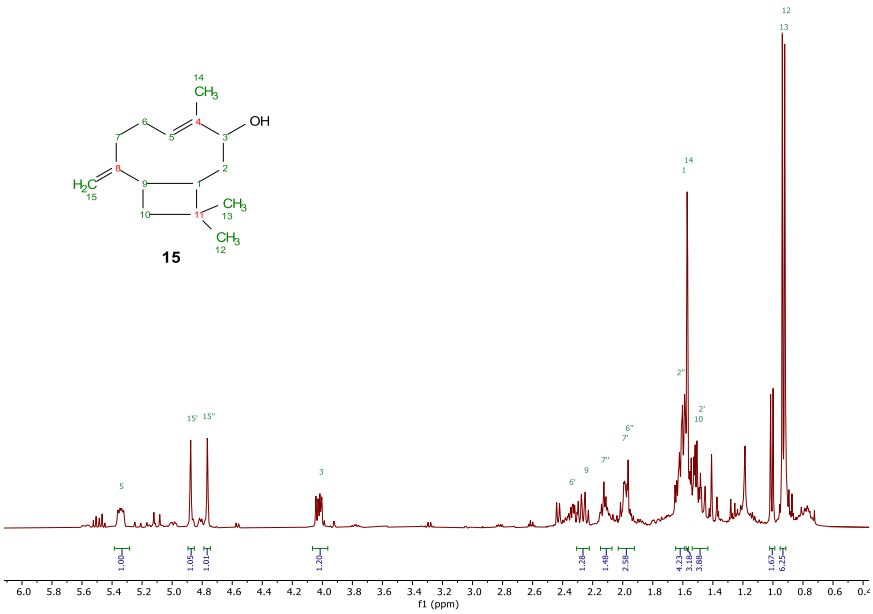

**Appendix 3.1** The 1H NMR spectrum in CDCl3 of caryophyllene-4,5-oxide, product **13**. The mass spectrum result is M+16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 4.98 (dd, *J* = 4.1, 1.4 Hz, 1H), 4.86 (dd, *J* = 1.6, 0.8 Hz, 1H), 2.87 (dd, *J* = 10.6, 4.2 Hz, 1H), 2.67 – 2.57 (m, 1H), 2.34 (ddd, *J* = 12.5, 8.1, 4.4 Hz, 1H), 2.25 (ddtd, *J* = 12.3, 8.6, 4.3, 0.9 Hz, 1H), 2.16 – 2.04 (m, 3H), 1.83 – 1.71 (m, 2H), 1.70 – 1.63 (m, 2H), 1.62 – 1.55 (m, 2H), 1.40 – 1.28 (m, 2H), 1.20 (d, *J* = 0.8 Hz, 3H).



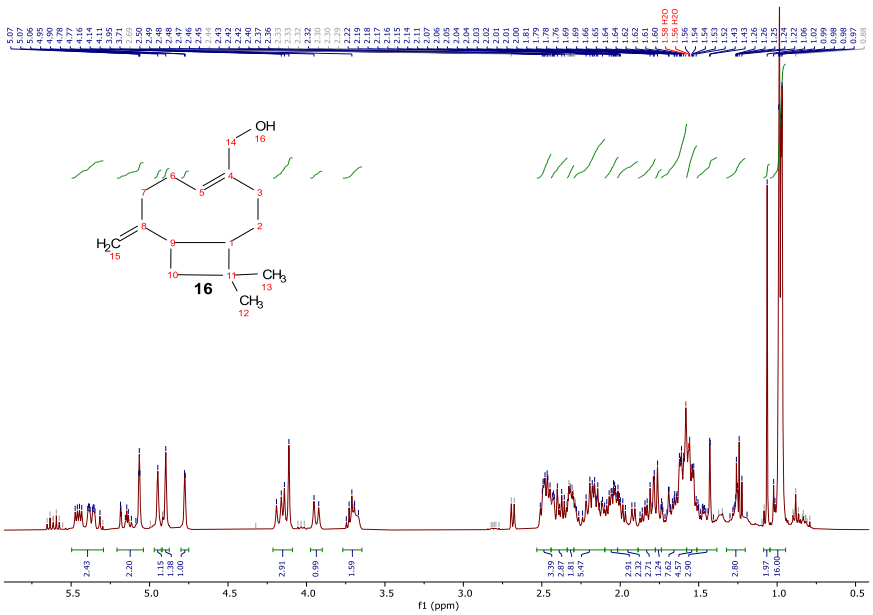

**Appendix 3.2** The 1H NMR spectrum in CDCl3 of caryophyllene-8,15-oxide, product **14**. The mass spectrum result is M+16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.38 (t, *J* = 8.0 Hz, 1H), 2.69 (t, *J* = 4.1 Hz, 2H), 2.56 (dt, *J* = 27.5, 6.0 Hz, 1H), 2.37 – 2.25 (m, 2H), 2.10 (q, *J* = 8.4, 6.4 Hz, 2H), 1.91 (dtd, *J* = 27.9, 11.3, 10.7, 5.8 Hz, 3H), 1.63 (s, 1H), 1.56 – 1.53 (m, 1H), 1.52 (s, 1H), 1.50 (d, *J* = 2.4 Hz, 1H), 1.33 – 1.19 (m, 1H), 1.19 – 1.05 (m, 1H), 0.92 (d, *J* = 6.7 Hz, 8H).



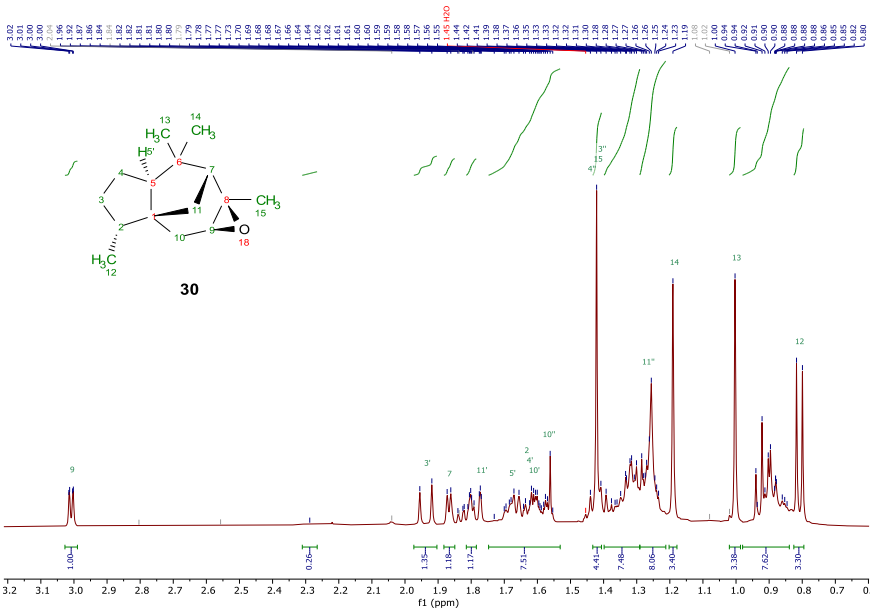

**Appendix 3.3** The 1H NMR spectrum in CDCl3 of caryophyllene-3-ol, product **15** The mass spectrum result is M+16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.34 (dd, *J* = 10.0, 5.3 Hz, 1H), 4.88 (d, *J* = 1.7 Hz, 1H), 4.76 (d, *J* = 1.7 Hz, 1H), 4.07 – 3.96 (m, 1H), 2.31 – 2.22 (m, 1H), 2.15 – 2.07 (m, 1H), 2.03 – 1.92 (m, 2H), 1.65 – 1.59 (m, 4H), 1.57 (d, *J* = 1.4 Hz, 3H), 1.54 – 1.43 (m, 4H), 1.01 (d, *J* = 6.5 Hz, 2H), 0.93 (d, *J* = 6.3 Hz, 6H).



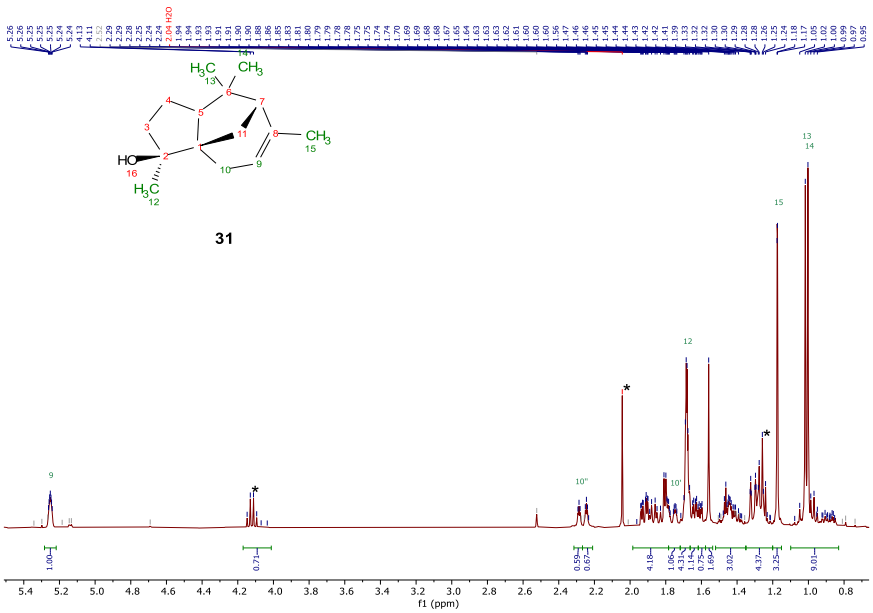

**Appendix 3.4** The 1H NMR spectrum in CDCl3 of caryophyllene-14-ol, product **16**. The mass spectrum result is M+16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.50 – 5.29 (m, 1H), 5.21 – 5.04 (m, 1H), 5.07 (s, 0H), 4.95 (s, 1H), 4.90 (s, 1H), 4.78 (d, *J* = 2.0 Hz, 1H), 4.21 – 4.09 (m, 2H), 3.94 (d, *J* = 12.2 Hz, 1H), 3.71 (td, *J* = 12.4, 10.7, 8.4 Hz, 1H), 2.53 – 2.41 (m, 2H), 2.45 – 2.34 (m, 1H), 2.32 (s, 1H), 2.30 – 2.08 (m, 2H), 2.11 – 2.01 (m, 1H), 2.04 – 1.89 (m, 1H), 1.89 – 1.77 (m, 1H), 1.81 – 1.71 (m, 1H), 1.75 – 1.58 (m, 2H), 1.58 – 1.48 (m, 1H), 1.52 – 1.38 (m, 1H), 1.32 – 1.20 (m, 1H), 1.19 (s, 1H), 1.06 (s, 1H), 0.98 (dd, *J* = 6.2, 2.4 Hz, 8H).



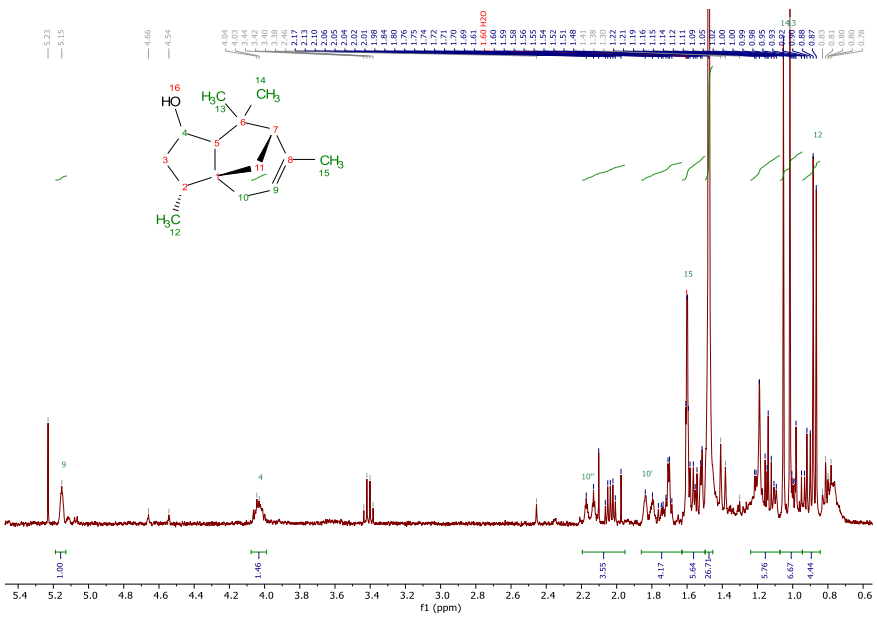

**Appendix 3.5** The 1H NMR spectrum in CDCl3 of cedrene-8,9-oxide, product **30**. The mass spectrum result is [M+K] +16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 3.01 (dd, *J* = 4.5, 0.9 Hz, 1H), 1.94 (d, *J* = 14.5 Hz, 1H), 1.87 (d, *J* = 4.4 Hz, 1H), 1.80 (dt, *J* = 4.1, 2.4 Hz, 1H), 1.71 – 1.64 (m, 2H), 1.65 – 1.55 (m, 3H), 1.42 (s, 4H), 1.19 (s, 3H), 1.00 (s, 3H), 0.81 (d, *J* = 7.2 Hz, 3H).



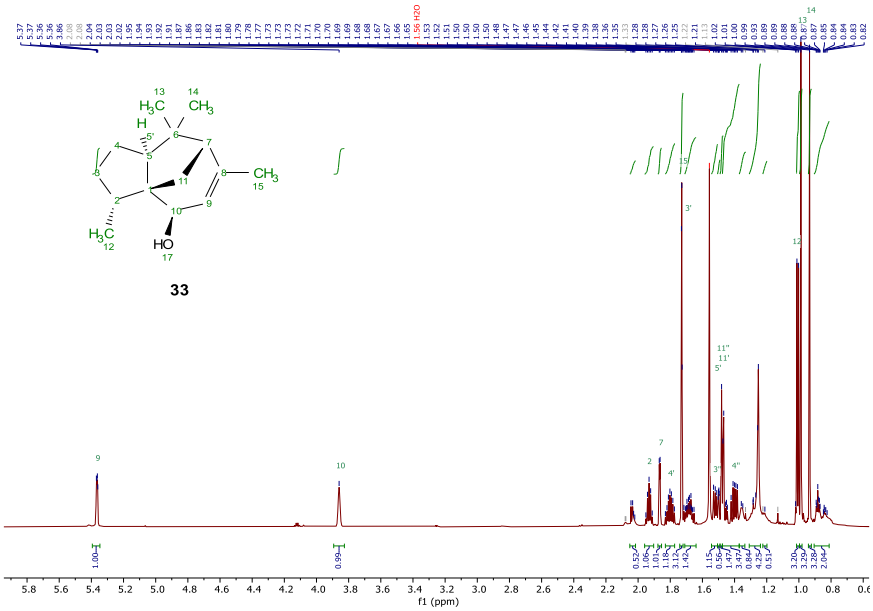

**Appendix 3.6** The 1H NMR spectrum in CDCl3 of cedrene-2-ol, product **31**. The mass spectrum result is [M+K] +16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.25 (tt, *J* = 2.8, 1.4 Hz, 2H), 4.12 (q, *J* = 7.2 Hz, 1H), 2.31 – 2.27 (m, 1H), 2.24 (q, *J* = 2.5 Hz, 1H), 1.98 – 1.76 (m, 7H), 1.80 – 1.72 (m, 1H), 1.68 (h, *J* = 3.4 Hz, 8H), 1.67 – 1.60 (m, 1H), 1.64 – 1.58 (m, 1H), 1.56 (s, 3H), 1.52 – 1.35 (m, 3H), 1.35 – 1.19 (m, 7H), 1.17 (d, *J* = 0.9 Hz, 5H), 1.10 – 0.83 (m, 14H), 0.88 (s, 2H).



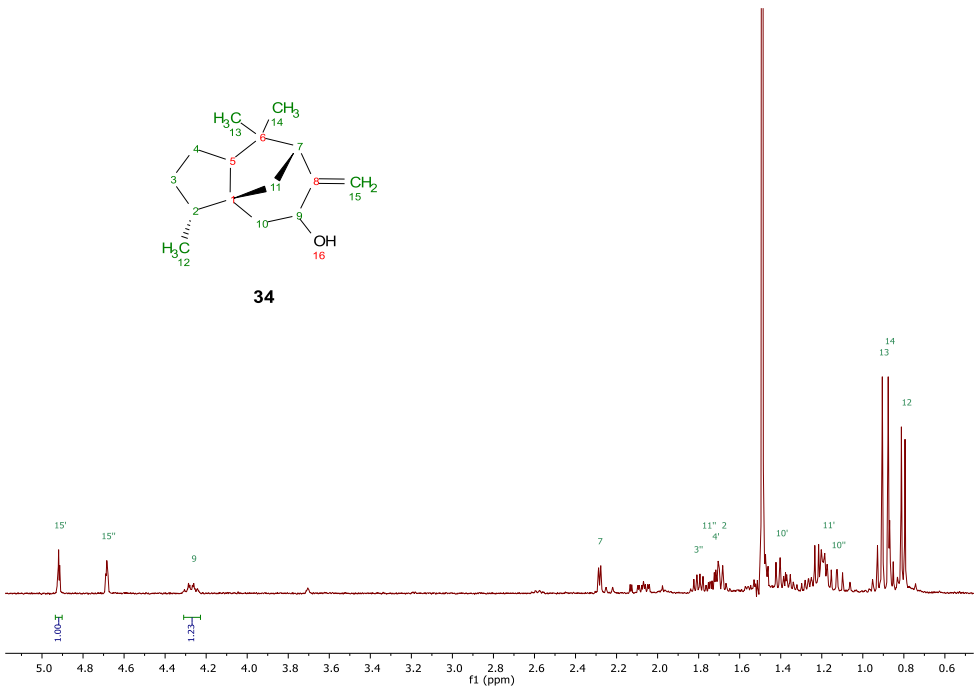

**Appendix 3.7** The 1H NMR spectrum in CDCl3 of cedrene-4-ol, product **32**. The mass spectrum result is [M+K] +16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.15 (s, 1H), 4.10 – 3.97 (m, 2H), 2.19 – 1.95 (m, 2H), 1.86 – 1.63 (m, 4H), 1.63 – 1.50 (m, 5H), 1.48 (s, 41H), 1.24 – 1.07 (m, 5H), 1.07 – 0.93 (m, 8H), 0.89 (dd, *J* = 13.9, 7.3 Hz, 5H).



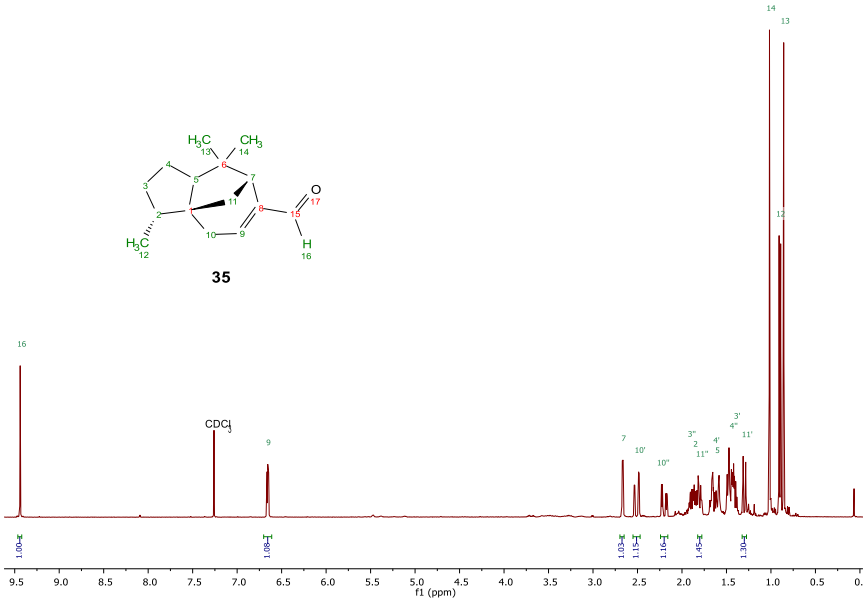

**Appendix 3.8** The 1H NMR spectrum in CDCl3 of cedrene-10-ol, product **33**. The mass spectrum result is [M+K] +16.

$^1$H NMR (700 MHz, $CDCl_3$) δ 5.36 (dd, *J* = 4.1, 1.8 Hz, 1H), 3.86 (s, 1H), 2.05 – 2.02 (m, 1H), 1.93 (p, *J* = 6.9 Hz, 1H), 1.87 (d, *J* = 3.6 Hz, 1H), 1.80 (tt, *J* = 13.1, 6.8 Hz, 1H), 1.73 (d, *J* = 1.4 Hz, 3H), 1.01 (d, *J* = 7.1 Hz, 3H), 0.99 (s, 3H), 0.93 (s, 3H).



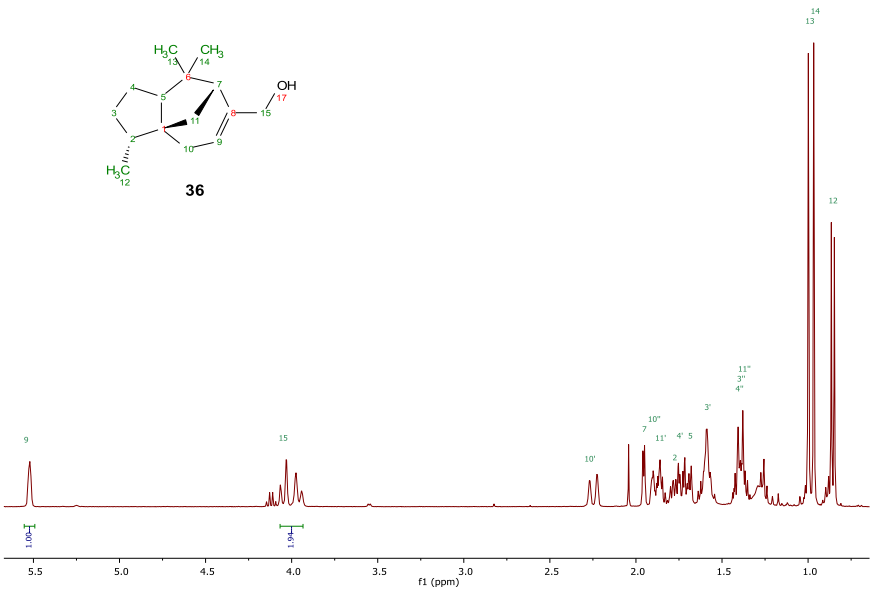

**Appendix 3.9** The 1H NMR spectrum in CDCl3 of β-cedrene-10-ol, product **34.** The mass spectrum result is [M+K] +16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 4.92 (t, *J* = 2.1 Hz, 1H), 4.68 (t, *J* = 2.3 Hz, 1H), 4.27 (dt, *J* = 10.0, 7.6 Hz, 1H), 2.32 – 2.25 (m, 1H), 1.81 (dt, *J* = 11.9, 6.0 Hz, 1H), 1.76 – 1.65 (m, 3H), 1.44 – 1.32 (m, 2H), 1.25 – 1.08 (m, 6H), 0.94 – 0.84 (m, 8H), 0.84 – 0.77 (m, 5H).



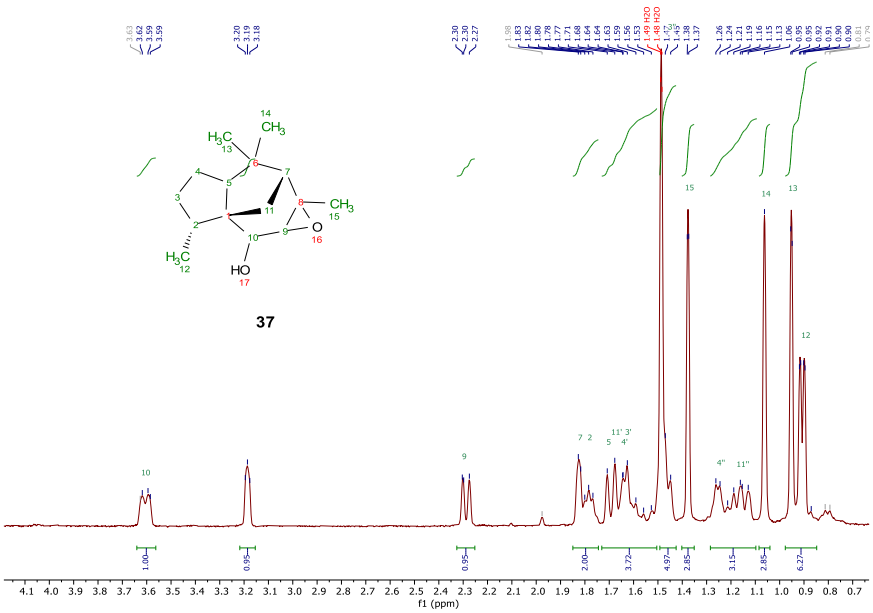

**Appendix 3.9** The 1H NMR spectrum in CDCl3 of cedrenal, product **35**. The mass spectrum result is [M+K] +14.

$^1$H NMR (400 MHz, $CDCl_3$) δ 9.44 (d, *J* = 0.7 Hz, 1H), 6.65 (dd, *J* = 4.2, 2.8 Hz, 1H), 2.67 (d, *J* = 3.9 Hz, 1H), 2.51 (dd, *J* = 19.7, 2.8 Hz, 1H), 2.20 (ddd, *J* = 19.7, 4.3, 1.4 Hz, 1H), 1.82 – 1.78 (m, 1H), 1.70 – 1.54 (m, 4H), 1.51 – 1.35 (m, 5H), 1.30 (d, *J* = 11.4 Hz, 1H), 1.02 (s, 4H), 0.92 – 0.84 (m, 8H).



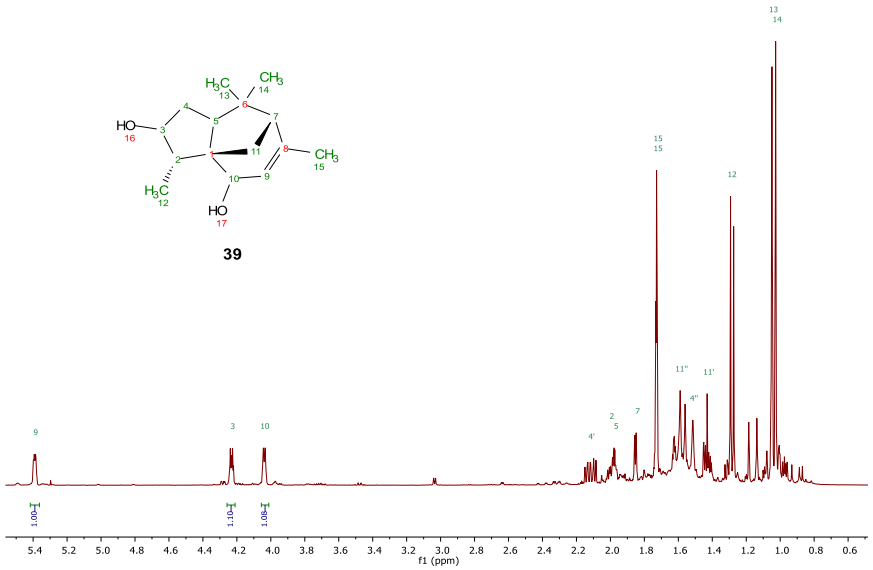

**Appendix 3.10** The 1H NMR spectrum in CDCl3 of cedrane-15-ol, product **36**. The mass spectrum result is [M+K] +16.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.52 (dq, *J* = 4.0, 1.5 Hz, 1H), 4.07 – 3.93 (m, 2H), 2.25 (dq, *J* = 17.0, 2.2 Hz, 1H), 1.95 (d, *J* = 3.9 Hz, 1H), 1.92 – 1.82 (m, 3H), 1.81 – 1.72 (m, 2H), 1.72 – 1.66 (m, 2H), 1.64 – 1.54 (m, 3H), 1.46 – 1.34 (m, 4H).



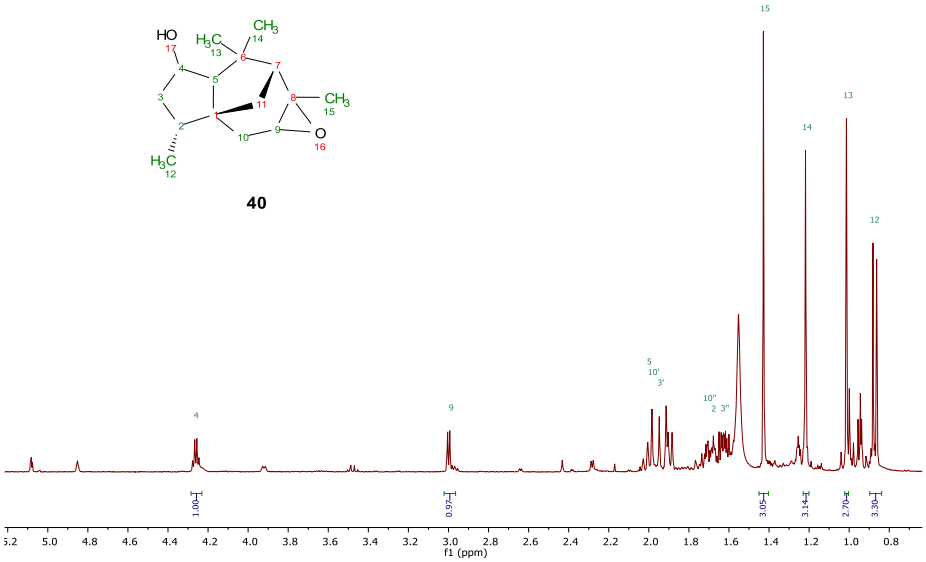
**Appendix 3.11** The 1H NMR spectrum in CDCl3 of cedrene-8,9-oxide-10-ol, product **37**. The mass spectrum result is [M+K] +32.

$^1$H NMR (400 MHz, $CDCl_3$) δ 7.19 (s, 1H), 3.64 – 3.56 (m, 1H), 3.18 (d, *J* = 3.9 Hz, 1H), 2.33 – 2.25 (m, 1H), 1.85 – 1.74 (m, 2H), 1.65 (dt, *J* = 21.2, 13.2 Hz, 4H), 1.46 (d, *J* = 8.8 Hz, 1H), 1.38 (d, *J* = 2.3 Hz, 3H), 1.28 – 1.10 (m, 3H), 1.06 (s, 3H), 0.98 – 0.85 (m, 7H).



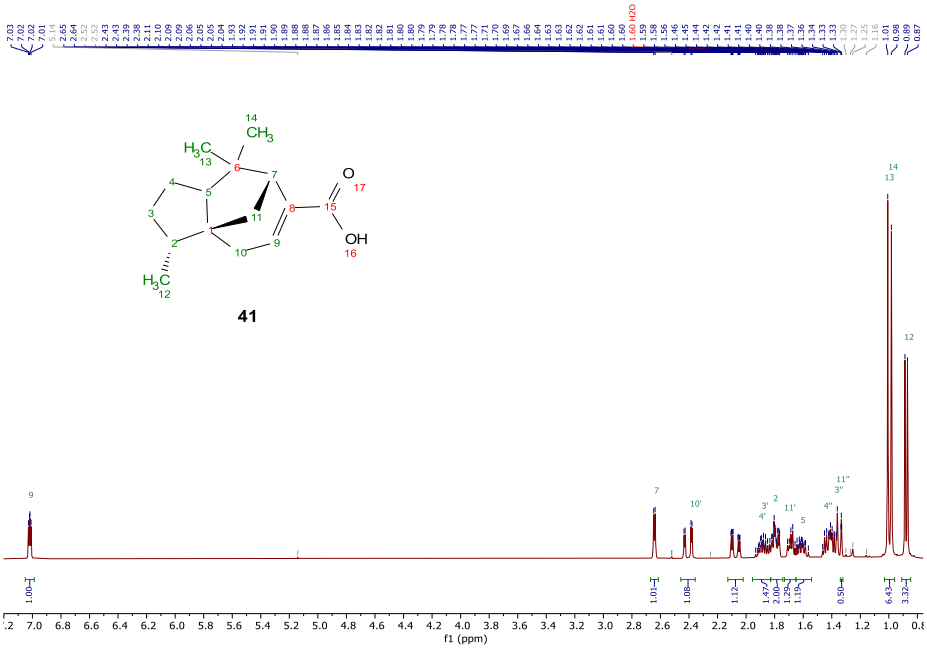
**Appendix 3.12** The 1H NMR spectrum in CDCl3 of cedrene-3,10-diol, product **39**. The mass spectrum result is [M+K] +32.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.42 – 5.36 (m, 1H), 4.23 (dt, *J* = 4.9, 2.2 Hz, 1H), 4.04 (dt, *J* = 4.3, 1.2 Hz, 1H), 2.12 (ddd, *J* = 13.7, 6.9, 5.3 Hz, 1H), 2.04 – 1.94 (m, 2H), 1.85 (d, *J* = 3.7 Hz, 1H), 1.76 – 1.70 (m, 4H), 1.66 – 1.47 (m, 11H), 1.47 – 1.39 (m, 2H), 1.33 – 1.25 (m, 5H), 1.05 (d, *J* = 2.1 Hz, 4H), 1.03 (s, 4H).



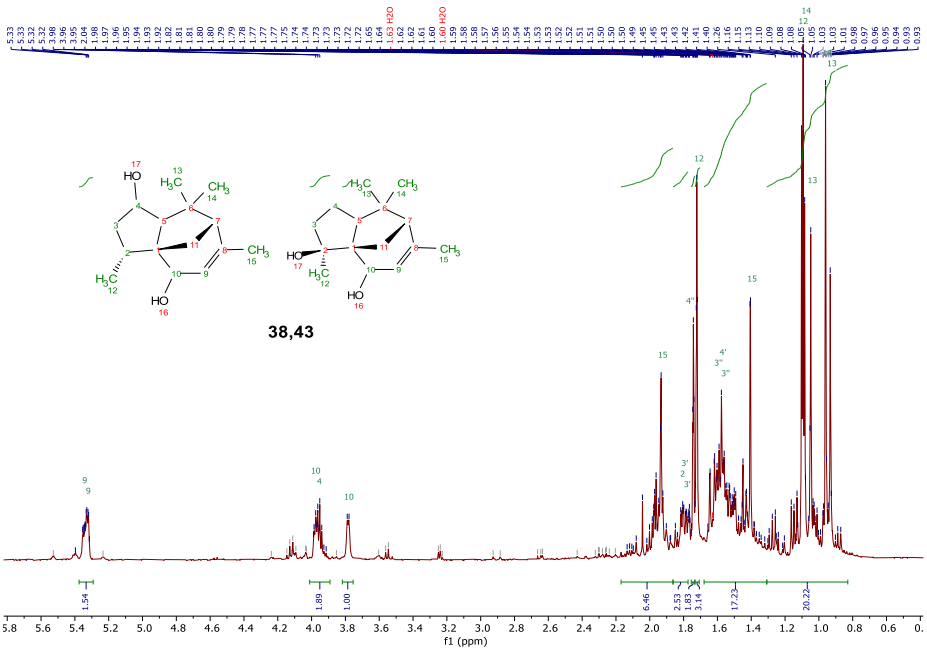

**Appendix 3.13** The 1H NMR spectrum in CDCl3 of cedrene-8,9-oxide-4-ol, product **40.** The mass spectrum result is [M+K] +32.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.30 (s, 1H), 4.26 (q, *J* = 4.2 Hz, 1H), 3.02 – 2.96 (m, 1H), 2.02 (d, *J* = 8.9 Hz, 1H), 1.98 (s, 1H), 1.95 (s, 1H), 1.71 (dd, *J* = 4.5, 2.0 Hz, 1H), 1.68 (ddd, *J* = 7.1, 4.8, 3.0 Hz, 1H), 1.63 (ddd, *J* = 9.2, 6.5, 4.2 Hz, 2H), 1.43 (s, 3H), 1.22 (s, 3H), 1.01 (s, 3H), 0.87 (d, *J* = 7.4 Hz, 3H).



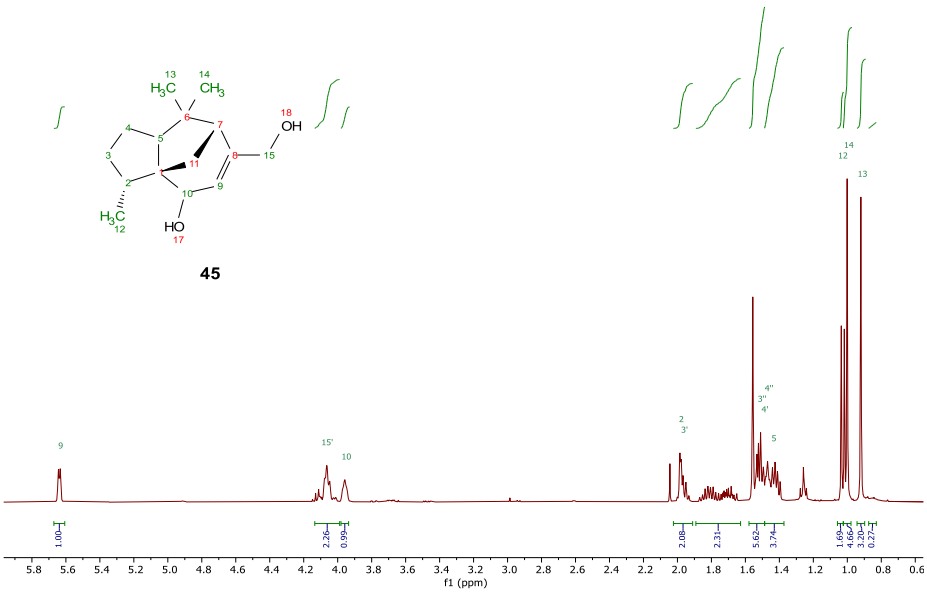

**Appendix 3.14** The 1H NMR spectrum in CDCl3 of cedrene-6-carboxylic acid, product **41.** The mass spectrum result is [M+K] +30.

$^1$H NMR (400 MHz, $CDCl_3$) δ 7.02 (dd, *J* = 4.3, 2.9 Hz, 2H), 2.64 (d, *J* = 3.9 Hz, 2H), 2.41 (dd, *J* = 19.4, 2.9 Hz, 2H), 2.07 (ddd, *J* = 19.4, 4.4, 1.3 Hz, 2H), 1.94 – 1.83 (m, 3H), 1.83 – 1.76 (m, 4H), 1.69 (dd, *J* = 9.3, 5.0 Hz, 2H), 1.66 – 1.56 (m, 2H), 1.47 – 1.39 (m, 4H), 1.37 (d, *J* = 3.2 Hz, 2H), 1.33 (d, *J* = 1.0 Hz, 1H), 1.01 (s, 7H), 0.98 (s, 7H).




**Appendix 3.15** The 1H NMR spectrum in CDCl3 of cedrene-4,10-diol and cedrene-2,10-diol, product **38 and 43**. The mass spectrum result is [M+K] +32.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.33 (ddt, *J* = 8.8, 4.5, 1.6 Hz, 2H), 3.97 (ddd, *J* = 11.1, 5.7, 3.0 Hz, 2H), 3.78 (d, *J* = 4.2 Hz, 1H), 1.97 – 1.90 (m, 3H), 1.74 (t, *J* = 1.4 Hz, 2H), 1.72 (t, *J* = 1.3 Hz, 3H), 1.65 – 1.48 (m, 8H), 1.42 (dd, *J* = 9.7, 1.1 Hz, 2H), 1.11 – 1.07 (m, 7H), 1.05 (d, *J* = 1.9 Hz, 3H), 0.96 (s, 3H), 0.93 (s, 2H).




**Appendix 3.16** The 1H NMR spectrum in CDCl3 of cedrene-10,15-diol, product **45**. The mass spectrum result is [M+K] +32.

$^1$H NMR (400 MHz, $CDCl_3$) δ 5.64 (dt, *J* = 3.6, 1.7 Hz, 1H), 4.13 – 3.99 (m, 2H), 3.96 (s, 1H), 1.97 (dd, *J* = 9.4, 4.9 Hz, 2H), 1.89 – 1.63 (m, 2H), 1.58 – 1.49 (m, 4H), 1.49 – 1.37 (m, 4H), 1.03 (d, *J* = 7.1 Hz, 3H), 1.00 (s, 3H), 0.92 (s, 3H), 0.85 (s, 0H).