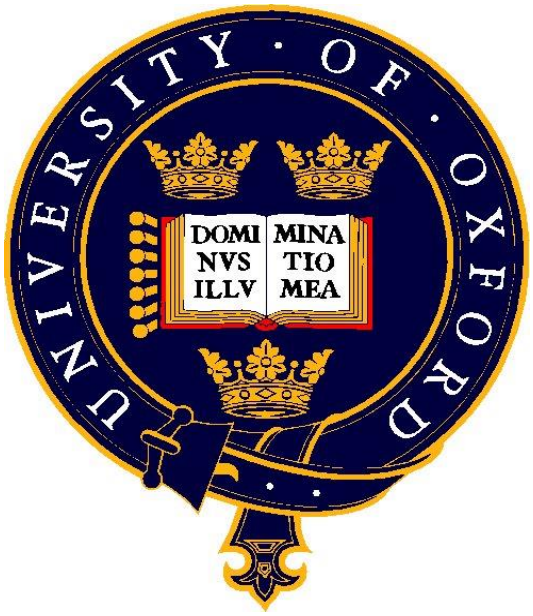
High-value wood aroma compounds via $P450_{BM3}$ catalysed sesquiterpene oxidation



A Thesis submitted to the Board of the Faculty of the Mathematical,

Physical and Life Sciences Division for the degree of

Doctor of Philosophy in the University of Oxford

Xiao Juie Wong

St Hilda's College

Hilary 2023

High-value wood aroma compounds via $P450_{BM3}$ catalysed sesquiterpene oxidation

Xiao Juie Wong D Phil. Thesis abstract

St Hilda's College Hilary, 2023

Cytochrome P450 enzyme (CYPs) catalyse the oxidation of C−H bonds via the insertion of an oxygen atom from atmospheric dioxygen and play key roles in the biosynthesis of endogenous compounds (alkaloids and steroids) and secondary metabolites (terpenoids). CYP102A1, also known as $P450_{BM3}$, has been widely studied for biotechnology application because of its unique structure of the electron transfer reductase domain being fused to the heme monooxygenase domain, which leads to high turnover frequency. Rational design and directed evolution are the main strategies that have been applied to engineer CYP102A1 to increase its activity for the oxidation of unnatural substrates.

Terpenes and terpenoid compounds have interesting properties such as aromas, flavours, pharmacological and biological activity. However, these compounds are often naturally low in abundance and their complex structures lead to complicated total synthesis pathways with low enantioselectivity and yield. This thesis reports the biosynthesis of valuable agarwood aroma compounds via biotransformation with engineered CYP102A1 variants.

Chapter 3 describes the engineering and application of CYP102A1 variants to develop high regio- and chemoselectivity synthesis of the guaiene skeleton compounds in agarwood by late-stage C−H functionalisation of α-bulnesene. The synthesis of (–)-guaia-1(10),11-dien-15,2-olide requires C−H functionalisations at both C2 and C15 of α-bulnesene and the development of a biocatalytic cascade reaction using two engineered CYP102A1 variants is described in Chapter 4. Chapter 5 details the reaction engineering of bulnesene oxidation for highly efficient synthesis of the target aroma compounds. In Chapter 6, the hydroxylation of α-guaiene is explored with engineered CYP102A1 variants to synthesise the peppercorn aroma compound rotundone which is also found in agarwood.

Acknowledgement

I wish to express my sincere gratitude to the University of Oxford and St Hilda's College for accepting me as a graduate student. Many thanks to Prof. Lorna Smith for her wise advice and kindness. This thesis would not be possible without the generous scholarship provided by the Ministry of Higher Education, Malaysia, for which I am eternally grateful.

I would like to thank Wenyu and Yuan for all the support they gave me, both academically and in life; Alethea, my first friend in Oxford, for always being there for me; Mathew, for everything and most importantly, for fixing my laptop for me and the princess card session. Many thanks to Melody, for going through the pain of grinding silica with me; Vicky for her enthusiasm and Jamie for his kindness and perpetual curiosity in Chemistry.

Also, I would like to thank Mary, Jiabao and Weiming (and of course, everyone in UK) for their support and for giving me the most memorable last few months in UK. Many thanks to Jen, Wei Hao, Shuoyi and Jiayin for supporting me through the years in UK. Special thanks to Jee, for always being there for me at the other side of the phone.

I would like to thank Oscar for putting up my bad tempers, calming me down when anxiety takes over and cheering me up throughout the journey. I miss the days with you in the UK but I am sure there are many more to come.

A special thanks to my supervisor, Prof. Luet Wong. I may not be a thousand-mile horse but thank you so much for being an amazing "Bo Le (discoverer)" and giving me a chance to pursue my DPhil study. I could not imagine myself, writing this acknowledgement here, without your constant support and guidance.

Last but not least, I would like to express my gratitude to my family. The unconditional love and support from my beloved parents and brother, to whom I am forever indebted. I am sure this thesis would never have been completed without your support and encouragement.

Table of contents

1 Introduction 2

1.1 C–H activation 2

1.2 Biocatalysts in synthesis 4

1.3 Cytochrome P450 enzymes: Direct C–H activation 5

1.3.1 The catalytic cycle of P450 enzymes 7

1.3.2 Electron transfer chains in P450 systems 10

1.4 $P450_{BM3}$ 10

1.4.1 Crystal structure of wild-type $P450_{BM3}$ 13

1.4.2 The active site and key residues of $P450_{BM3}$ 14

1.5 Protein engineering of $P450_{BM3}$ 16

1.5.1 Rational design and directed evolution 16

1.5.2 Combinatorial and iterative saturation mutagenesis 23

1.5.3 The mutant library approach 26

1.6 Molecular dynamics simulations and substrate docking 28

1.6.1 Molecular dynamics simulations 28

1.6.2 Molecular docking 31

1.6.3 Application in enzyme engineering 34

1.7 Terpenoid compounds 37

1.8 Agarwood 41

1.9 Thesis objectives 46

2 Materials and Methods 49

2.1 Materials, chemicals and solvents 49

2.2 Buffer, solution and growth media 49

2.3 Equipment and analysis method 50

2.4 DNA manipulation 52

2.4.1 Mutagenesis primer design 52

2.4.2 PCR site-directed mutagenesis 52

2.4.3 Agarose gel electrophoresis 52
2.4.4 Post-mutagenesis transformation 53
2.4.5 DNA purification and transformation 53
2.4.6 Enzyme production and purification 54
2.4.7 Enzyme quantification 55
2.5 *In vitro* screening, preparative scale reactions and product isolation 55
2.5.1 *In vitro* screening 55
2.5.2 Preparative scale reactions and product purification 57
2.5.3 *In vitro* screening for 15,2-lactone formation via a cascade oxidation of bulnesene 67
2.5.4 Preparative scale reactions for isolation of the 2-hydroxybulnesene-15-COOH, 1i 68
2.5.5 Preparative scale reactions for isolation of the 15,2-lactone, 1j 69
2.5.6 Preparative scale reactions for the isolation of rotundone (2f) 69
2.6 Biocatalytic synthesis of 15,2-lactone with bulnesene-enriched patchouli oil 70
2.7 Molecular dynamics simulations and molecular docking 71
3 The oxidation of α-bulnesene by CYP102A1 74
3.1 Purification of α-bulnesene and α-guaiene 74
3.2 Bulnesene oxidation by CYP102A1 76
3.2.1 Initial library screening and product characterisation 76
3.2.1.1 The effect of mutations 79
3.2.2 Sub-library screening of bulnesene 82
3.2.2.1 Summary 87
3.3 Synthesis of bulnesene-15-COOH 88
3.3.1 Chemoenzymatic synthesis of bulnesene-15-COOH 88
3.3.2 P450-meidated formation of bulnesene-15-COOH 90
3.4 Further mutagenesis for selective C15-oxidation of bulnesene 92
3.4.1 Progressive library guided mutagenesis 92
3.5 A third-generation library: docking-guided mutagenesis 96
3.5.1 Docking analysis of bulnesene with R19/F87I variant 98
3.5.1.1 The C15 poses 99

3.5.1.2 The importance of the F87I mutation 100
3.5.1.3 The NP poses 101
3.5.1.4 Other binding poses 103
3.5.1.5 Summary of docking analysis of bulnesene with variant R19/F87I 104
3.6 Docking guided mutagenesis study for bulnesene 105
3.6.1 Mutagenesis based on docking analysis 105
3.6.2 Fourth-generation variant library: combining mutations 107
3.7 Summary 111
4 Bulnesene-15,2-lactone synthesis by CYP102A1 114
4.1 Introduction 114
4.2 Oxidation of bulnesene-15-COOH with $P450_{BM3}$ 116
4.2.1 Initial and sub-library screening of bulnesene-15-COOH 116
4.2.2 Biocatalytic cascade reaction for 15,2-lactone formation 118
4.3 Optimisation of the biocatalytic cascade reaction 121
4.3.1 Second-generation variants for lactone production 122
4.3.2 Condition optimisation for lactone cyclisation 125
4.3.3 Minimising side product formation via reaction engineering 127
4.3.4 Preparative scale reaction for lactone production 131
4.4 Direct oxidation of bulnesene to bulnesene-15,2-lactone with a single CYP102A1 variant 132
4.4.1 Docking analysis with the R19/F87I/A82M/A184I/A330I variant 133
4.4.1.1 MD simulations for variant R19/F87I/A82M/A184I/A330I 134
4.4.1.2 General docking outputs 136
4.4.1.3 The C2 poses 137
4.4.1.4 The C11=C12 poses 138
4.4.1.5 The NP poses 139
4.5 Docking-guided mutagenesis for bulnesene-15-acid oxidation 140
4.5.1 Mutations designed for selectivity swap and activity enhancement 140
4.6 The synthesis of target compounds from bulnesene-enriched patchouli oil (BPO) 144
4.6.1 The synthesis of bulnesene-15-ol, bulnesene-15-al and bulnesene-15-COOH with BPO 144

4.6.2 The synthesis of bulnesene-15,2-lactone from BPO 145
4.6.3 Optimisation and titre test on the synthesis of bulnesene-15,2-lactone 147
4.7 Summary 149
5 Chemoselective bulnesene oxidation by reaction engineering 150
5.1 Chemoselectivity control of bulnesene oxidation 150
5.1.1 Reaction time 150
5.1.2 The substrate-to-enzyme ratio 153
5.2 Reaction optimisation 154
5.2.1 Methyl-β-cyclodextrin (MCD) concentration 154
5.2.2 Preparative scale reaction at higher titre with aeration 156
5.3 Summary of the biosynthesis of agarwood aroma compounds 158
6 Guaiene oxidation by engineered CYP102A1 163
6.1 Introduction 163
6.2 Oxidation of α-guaiene by CYP102A1 164
6.2.1 Initial library screening 164
6.2.2 Sub-library screening of guaiene 167
6.2.3 Summary 171
6.3 Biocatalytic synthesis of rotundone 172
6.3.1 Biocatalytic cascade synthesis of rotundone 172
6.3.2 A single CYP102A1 variant for direct oxidation of guaiene to rotundone 174
6.4 Summary 175
7 Conclusion 178
Reference 181

List of Abbreviations

ADH	alcohol dehydrogenase
COSY	correlation spectroscopy
CPR	cytochrome P450 reductase
CYP	cytochrome P450
DMAPP	dimethylallyl pyrophosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
FAD	flavin adenine dinucleotide
FID	flame ionisation detector
FMN	flavin mononucleotide
FPP	farnesyl pyrophosphate
GC	gas chromatography
GDH	glucose dehydrogenase
GGPP	geranylgeranyl pyrophosphate
GPP	geranyl pyrophosphate
HSQC	heteronuclear single quantum correlation spectroscopy
IPP	isopentenyl pyrophosphate
IPTG	isopropyl-β-D-thiogalactopyanoside
ISM	iterative saturation mutagenesis
Kan	kanamycin
CRED	carbonyl reductase
LB	lysogeny broth
MVA	mevalonate

$NAD(P)^+$	nicotine adenine dinucleotide (phosphate)
NAD(P)H	nicotine adenine dinucleotide (phosphate), reduced form
NMR	nuclear magnetic resonance
nOe	nuclear Overhauser effect
PAH	polycyclic aromatic hydrocarbon
PCR	polymerase chain reaction
PEO	patchouli essential oil
RPM	rotations per minute
SOC	super optimal broth with catabolite repression
TAE	tris-acetate-EDTA
TLC	thin layer chromatography
v/v	volume by volume
WT	wild type (i.e., mutation free)
TON	the total yield of product relative to enzyme

Chapter 1

1 Introduction

1.1 C–H activation

C–H activation is a powerful tool in organic synthesis as functionalisation of a chemically unreactive C–H bond with C–C, C–O and C–N bonds introduces versatile synthetic handles for structural elaboration of molecules. Extensive research in the past few decades have aimed at developing improved strategies for C–H bond functionalisation and palladium complexes are the most widely used catalysts in this endeavour.[1–3] In the presence of an oxidant, Pd(0)/Pd(II) could be used in a huge variety of organic reactions including but not limited to C–C/C–O bond formation. Neumann *et al.* reported one of the first C(sp^3)–H activation/C–N bond formation without the presence of a nitrene.[4] Under optimised conditions, a wide range of anilines with different functional groups could be converted to form varied indoline products. Pd(0) was reformed through reductive elimination from Pd(II), which could be used to catalyse a new reaction cycle. The formation of a range of 2-(trifluoromethyl)indoles which are important drug fragments, was reported by Cramer and co-workers via C(sp^3)–H functionalisation of trifluoroacetimidoyl chlorides by a Pd(0) catalyst (Fig. 1.1).[5]

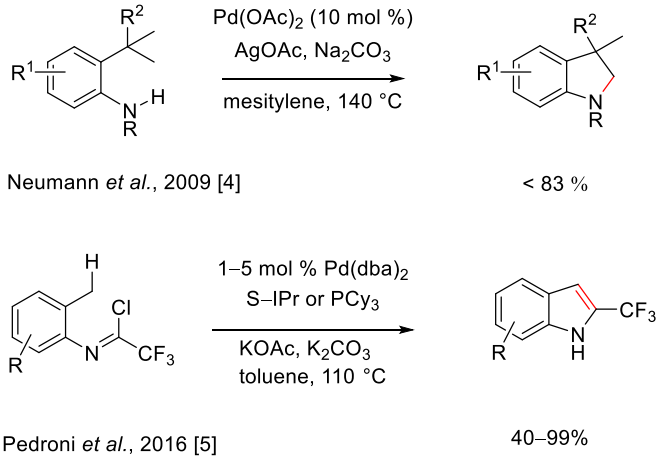


Figure 1.1 Examples of Pd(0)/Pd(II)-catalysed C–H activation reactions.

Although highly selective C–H activation could often be achieved by palladium catalysts, processes using this precious transition metal have higher carbon footprints due to its low abundance, which is also one of the major obstacles in large scale application.[6] With the need for sustainable synthesis and green chemistry becoming increasingly prominent, 3d metals are targeted as next-generation catalysts because they are more abundant and less toxic.[6–8] Chen *et al.* developed Fe-based catalysts that utilise hydrogen peroxide for aliphatic C(sp^3)–H activation.[8] The predictable selectivity based on the directing effect of a carboxylate group as well as the electronic and steric properties of the C–H bonds make the catalyst an attractive alternative for the late-stage functionalisation of organic compounds (Fig. 1.2).

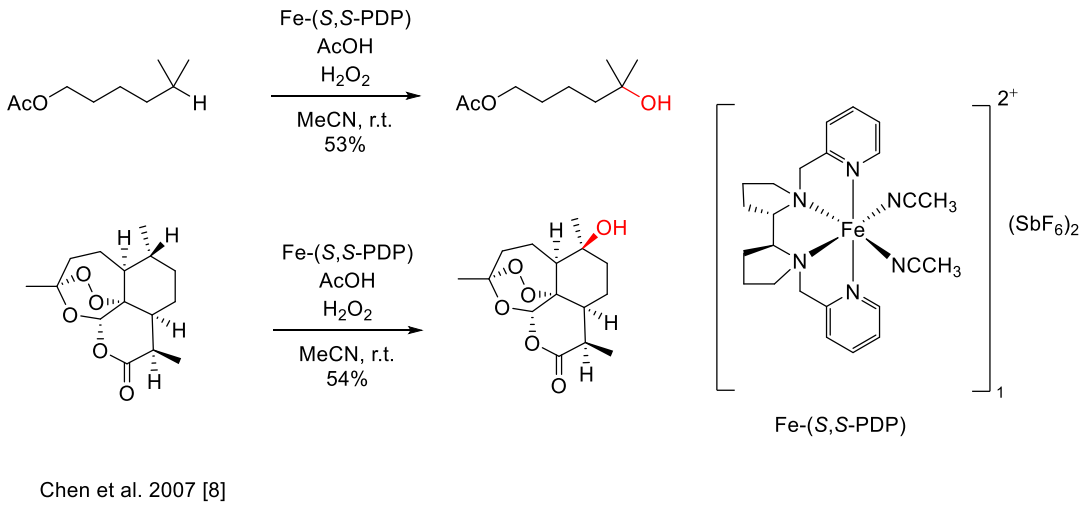


Figure 1.2 Examples of Fe-catalysed C-H activation.

Although catalysts with 3d metals are comparatively more sustainable, C–H activation under milder conditions with high efficiency remains highly desirable. Enzymes are well known and have been studied extensively for their ability to perform C–H activation with high regio- and enantioselectivity under ambient conditions.

1.2 Biocatalysts in synthesis

For thousands of years, enzymes have played important roles in the production of food such as bread and cheese through fermentation. More recently, enzymatic routes have gained popularity in chemical synthesis because of their advantages over conventional organic synthesis. Enzymes possess the combined benefits of a catalyst and high selectivity which usually give high atom efficiency under mild reaction conditions.[9–11]

Unlike chemical synthesis which needs to be done stepwise with the reaction conditions adjusted accordingly for each step, several enzymatic steps can be combined in a single aqueous reaction because of their inherent substrate specificity and regio- and stereoselectivity.[9] Besides being more atom-economic and environmentally friendly, a biocatalytic cascade is a much more convenient, streamlined process since the intermediate isolation steps are eliminated. The issue of having to isolate unstable intermediates and the accumulation of unstable intermediates could also be overcome with the linking of all the enzymatic reactions.[9,12,13]

With the advancement of biocatalytic process design and protein engineering, biocatalytic cascades are becoming more effective with highly selective enzymes and optimised conditions being reported to be used in the synthesis of complex molecules. For example, cycloketones, which are among the most important core structures in bioactive natural and synthetic compounds, can be produced via a dual enzyme cascade reaction involving both P450s and carbonyl reductase (CRED) from their corresponding cycloalkanes.[12]

An impressive example is the biocatalytic total synthesis of islatravir, a HIV drug, reported by Merck & Co. in 2019 (Fig. 1.3).[13] The cascade reaction was first planned with biocatalytic retrosynthesis to identify suitable biocatalysts, followed by the screening of a selection of enzymes. The chosen enzymes with high activity and selectivity were subjected to directed evolution to further improve the enzyme performance. Once the enzyme of each biocatalytic step had been identified, the

biocatalytic cascade was optimised to achieve a favourable equilibrium and maximize the pathway flux. Immobilisation of enzymes was introduced to simplify the reaction work up and allow the crystalline product to be filtered more easily. The number of steps required for the biocatalytic synthesis of islatravir were reduced significantly compared to conventional organic synthesis as product selectivity was achieved without the need of directing groups.

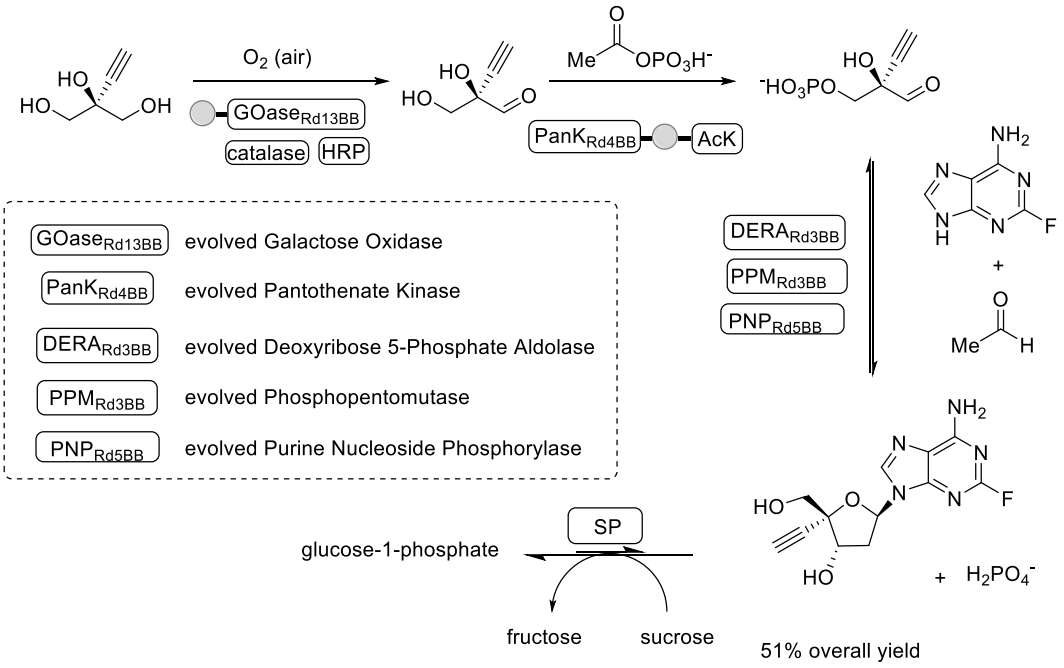


Figure 1.3 Fully assembled biocatalytic pathway of islatravir.

1.3 Cytochrome P450 enzymes: Direct C–H activation

The cytochromes P450 superfamily are heme-thiolate monooxygenases that catalyse C–H bond oxyfunctionalisation in many different types of compounds. First discovered by Martin Klingenberg in 1957, P450 was named after the characteristic absorption peak observed at 450 nm when the reduced, Fe^{II} form of the enzyme binds with CO.[14] P450 enzymes activate the dioxygen molecule and 'insert' one of the oxygen atoms into a C–H bond of the target compound. The other oxygen atom is reduced to water.

In addition to the signature reaction of C–H bond hydroxylation, P450 enzymes also catalyse the oxidation of alcohols, epoxidation of double bonds, *N*-dealkylation, C–C bond cleavage, *N*-oxidation, and etc (Fig. 1.4).[15] The development of P411, a serine substituted P450 at residue C400, opened new routes to various unnatural reactions such as carbene transfer and C–H amination.[16,17] The capability of P450s to perform these reactions on a wide range of different substrates under room temperature and pressure with high selectivity has made these enzymes promising tools for biocatalytic application.[15]

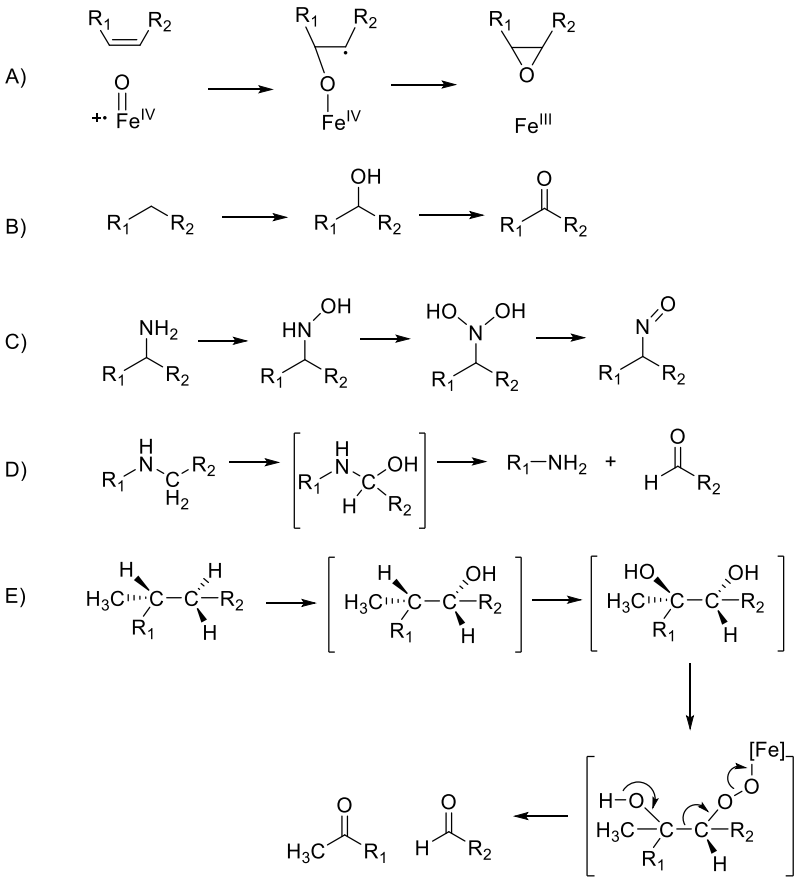


Figure 1.4 The varied activities of P450 enzymes: (a) epoxidation, (b) alcohol oxidation, (c) *N*-oxidation, (d) *N*-dealkylation, (e) C–C bond cleavage.

P450 enzymes are found in all kingdoms of life and play important roles in the production of second metabolites, biosynthesis of steroid in mammals,[18] herbicide metabolism in plants,[19] as well as the degradation of xenobiotics. One notable exception is *Escherichia coli*, which does not express a

P450 enzyme, making *E. coli* most widely used to over-express the enzyme. There are >400,000 *CYP* genes in databases (http://drnelson.uthsc.edu/CytochromeP450.html). The single uniquely conserved feature is a cysteine residue attached axially to the heme prosthetic group of the P450. The π-electron donation of the thiolate ligand was proposed to play an essential role in the heterolytic cleavage of the O–O bond of molecular dioxygen.[20]

The C–H bond oxy-functionalisation activity of P450 enzymes has been researched extensively for its application in synthesis. Drug metabolism is the most commonly known physiological role of P450 enzymes with over 90% of clinical drugs are metabolised by human P450 enzymes. The engineering of CYP102A1, also known as $P450_{BM3}$, from *Bacillus megaterium* for potential biocatalytic applications, is the subject of this thesis.

1.3.1 The catalytic cycle of P450 enzymes

The ferric heme iron (**I**) is hexa-coordinate and low spin (S=1/2) in the resting state of the enzyme, with the four porphyrin (P) nitrogens at the equatorial positions while a water molecule and the proximal cysteine thiolate ligand occupy the axial positions (Fig. 1.5). Binding of a substrate close to the heme displaces the water molecule and the penta-coordinate ferric iron (**II**) is high spin (S=5/2) due to the smaller ligand field splitting after the removal of a ligand. Loss of a ligand also destabilises the ferric state relative to the ferrous, leading to a more positive reduction potential, thus increasing the thermodynamic driving force for heme reduction compared to the low-spin, hexa-coordinated form in the absence of a bound substrate. In addition, since the ferrous form is also penta-coordinate, loss of the water ligand from the ferric state also lowers the reorganisation energy for electron transfer and accelerates heme reduction. This is the key mechanism for "gating" of the P450 catalytic cycle whereby binding of a viable substrate initiates the cycle by inducing rapid electron transfer reduction of the heme iron.

The two electrons from NAD(P)H required for P450 monooxygenase activity are transferred to the heme one at a time with the assistance of electron transfer partner protein(s). The ferric heme iron is reduced to the ferrous state (**III**) to which dioxygen binds and the oxy-complex $PFe^{III}(O_2)$ complex (**IV**) is formed. The Fe–O–O bond is bent because one component of the occupied π* orbitals of the dioxygen acts as a σ-donor, while the other π* orbital acts as a π-acceptor. Transfer of the second electron reduces the bound dioxygen to form the ferric peroxy complex (**V**). This is protonated at the terminal oxygen atom to give the ferric hydroperoxy complex (**VI**), also called Compound 0 (Fig. 1.5).

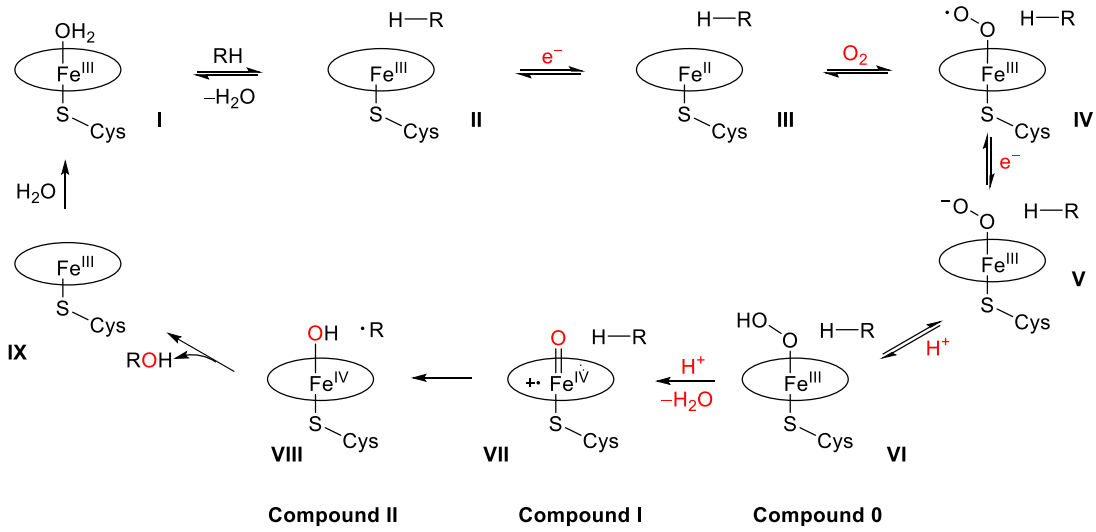


Figure 1.5 The catalytic cycle of P450 enzymes.

Protonation at the distal oxygen of Compound 0 results in water loss through heterolytic cleavage of the O–O bond and forms the formally iron(V), ferryl, Compound I species $[PFe^{V}{=}O]$ which is more appropriately formulated as $[P^{\bullet+}Fe^{IV}{=}O]$ (**VII**). Compound I is the powerful oxidising entity of P450. The π bond of the Fe=O group can be considered to be cleaved to form $[PFe^{IV}–O\bullet]$, with the radical localised on the oxygen. This oxygen abstracts a hydrogen atom of the substrate to form a substrate radical and the PFe^{IV}–OH species (**VIII**), also called Compound II. Radical rebound occurs which combines the substrate radical and ·OH, leading to overall hydroxylation of the substrate; the second electron in the Fe–OH bond of Compound II is returned to the iron centre to generate a ferric heme (**IX**).

The hydroxylated substrate leaves the active site and a water molecule re-coordinates to the heme and the enzyme returns to its resting state. Alternatively, a new substrate displaces the product, a new cycle of hydroxylation starts without the enzyme returning to the resting state.

P450 monooxygenase activity can be considered as a $2e^{-}/2H^{+}$ reduction of dioxygen to water and an O atom, which is stabilised by coordinating to the heme iron and then inserted into the substrate. NADPH provides two electrons and one proton to the reaction which requires an external proton in an overall reaction shown below.

$$RH + O_2 + 2e^{-} + 2H^{+} \rightarrow ROH + H_2O$$

There are three possible uncoupling pathways that might occur when the electrons supplied by NADPH are not fully utilised in the hydroxylation of substrate. The first uncoupling pathway is superoxide uncoupling which happens when the second electron transfer is too slow causing the oxy-complex (**IV**) to lose the superoxide and decompose back to the ferric form **II**. This may occur when the bound oxygen in the oxy-complex (**IV**) is sterically hindered by the substrate. The second uncoupling pathways is the loss of hydroperoxide ligand as hydrogen peroxide when a proton is attached to the iron-bound oxygen of ferric hydroperoxy complex **VI**. The third type of uncoupling occurs when the substrate is too far away from the ferryl oxygen. Instead of abstracting a hydrogen atom from the substrate, the ferryl complex **VII** is reduced back to the ferric form **II** by two electrons while releasing a water molecule at the same time.[21] The uncoupling processes can limit the applicability of P450 in hydroxylating the target compounds. It could also potentially lead to the degradation of the heme cofactor and the peptide backbone because of the release of reactive oxygen species.[15] Compared to fatty acid hydroxylation by CYP102A1 which almost fully utilized the electrons from NADPH, the coupling efficiency of a non-native substrate can be significantly lower.[22] Protein engineering has been widely researched and serves as a powerful strategy in increasing the coupling efficiency, together with the activity and selectivity of the enzyme.

1.3.2 Electron transfer chains in P450 systems

In general, P450s are divided into 4 different classes based on their electron transfer system.[23] The electron transfer system of Class I P450 consists of three separate components: FAD-containing ferredoxin reductase (FdR), an iron-sulphur ferredoxin (Fdx), and the P450 enzyme (Fig. 1.6). This type of electron transfer system is mainly found in bacteria and mitochondrial systems, such as CYP101A1 from *Pseudomonas putida*.[24,25] Class II P450s are mainly found in the endoplasmic reticulum. The electrons are transferred from NADPH to the heme of P450 in the microsomal system through cytochrome P450 reductase (CPR), which consists of separate FAD and FMN domains (Fig. 1.6).[26] Both Class III and IV enzymes are fused systems: FAD—FMN—heme for Class III and FMN—ferredoxin —heme for Class IV (Fig. 1.6).[21, 23]

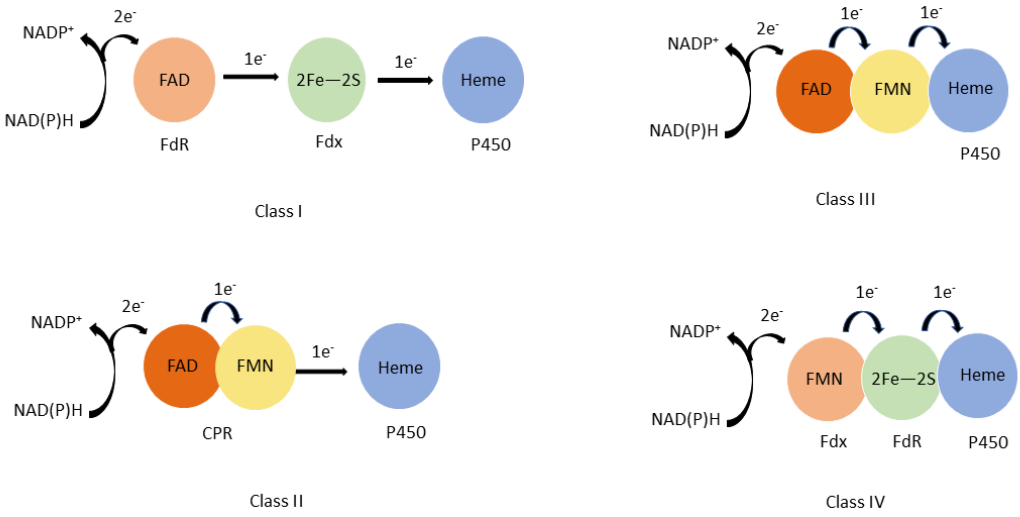


Figure 1.6 P450 electron transfer chains.

There has been increasing attention on Class III and IV enzymes because they involve the expression of a single gene due to their fused nature, unlike Class I and II which require separate expression of the cofactor protein(s). $P450_{BM3}$, the subject of this thesis, is a Class III system.[21]

1.4 $P450_{BM3}$

$P450_{BM3}$ (CYP102A1) was first discovered by Fulco in the early 1970s and it was identified as a soluble fatty acid hydroxylase.[27] It is more widely recognised with the name $P450_{BM3}$ as the third

P450 discovered from *Bacillus megaterium* and was the first prokaryotic enzyme discovered to have the electron transfer chains naturally attached to the P450 enzyme in a single polypeptide chain.[21,28] The exact physiological role of $P450_{BM3}$ remains unclear. Although straight chain unsaturated fatty acids have been widely recognised as the natural substrates of $P450_{BM3}$, they do not induce the expression of the enzyme in *B. megaterium*, unlike the expression of P450cam in *P. putida* that can be induced by its natural substrate, camphor.[21] In contrast, the enzyme can be induced by polyunsaturated fatty acids which interestingly, are harmful to *B. megaterium* if added exogenously and are produced only in small amounts as a transient response to low temperature.[29,30] This suggests that detoxifying xenobiotic unsaturated fatty acids could possibly be the physiological role of $P450_{BM3}$.

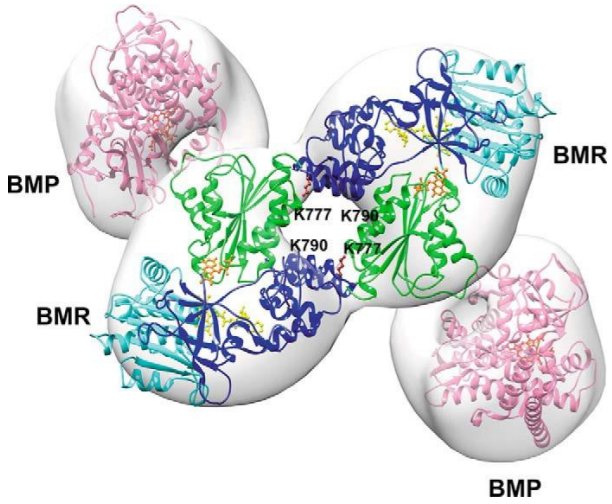


Figure 1.7 A 3D model of dimeric $P450_{BM3}$ showing the close view of the dimerization interface. The P450 domains of BMP are shown in magenta. FMN, FAD, and nucleotide-binding domains of BMR are shown in green, blue and cyan respectively. The FMN, FAD and heme co-factors are shown in orange, yellow and red.[31]

The molecular weight of the enzyme is 120 kDa, with about 55 kDa contributed by the P450 domain (BMP) and another 65 kDa by the FAD-FMN reductase domain (BMR) fused at its C-terminus.[31] By studying the enzyme with sedimentation velocity experiment, it was discovered that $P450_{BM3}$ exists as a mixture of monomers, dimers, trimers and tetramers with the majority in dimer form with a molecular weight of 236 kDa.[32] Zhang *et al*. studied the structure of the $P450_{BM3}$ dimer with various

methods including negative stain EM and molecular modelling and concluded that the CYP102A1 dimerises at BMR domains while BMP domains showed various conformations which promote high activity of the enzyme (Fig. 1.7).[33]

The dimeric enzyme dissociates to the inactive monomeric form at low ionic strength, indicating that either or both electron transfer steps (FAD to FMN or FMN to heme) happen across the dimer interface (Fig. 1.7).[34, 35]

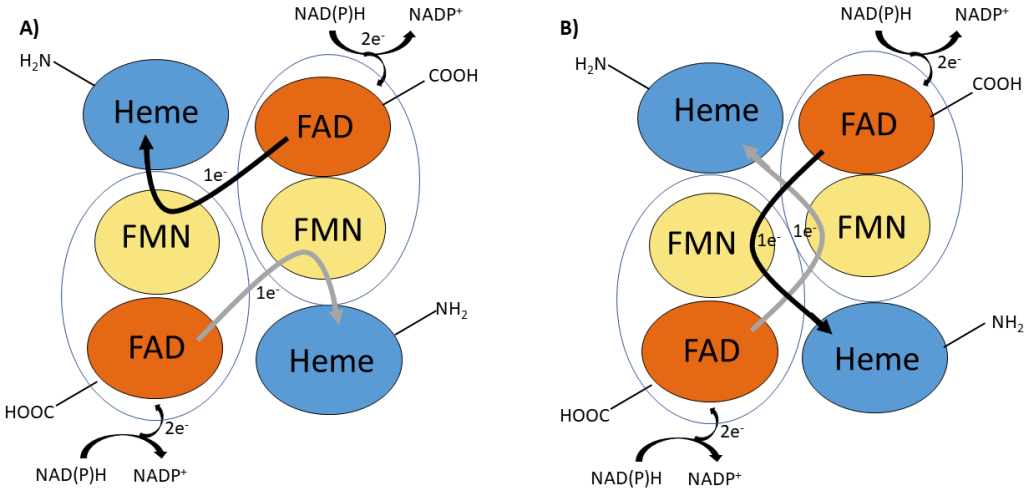


Figure 1.8 Schematic representation of the electron transfer in FAD—FMN—Heme fused system of $P450_{BM3}$. The electron transfer from FAD to FMN between the monomers followed by **A**) the electron transfer from FMN to Heme within the monomers ($FAD_A \rightarrow FMN_B \rightarrow HEME_B$) **B**) the electron transfer from FMN to Heme between the monomers ($FAD_A \rightarrow FMN_B \rightarrow HEME_A$)

A set of experiments was designed by eliminating each individual domains through site-directed mutagenesis and measuring the activity to study the electron transfer pathway of $P450_{BM3}$ (Fig. 1.8). The result indicated that the electron transfer from FAD to FMN domain occurs between the monomers ($FAD_A \rightarrow FMN_B$) and the electron could then be passed from the reduced FMN domain to the heme domain of either monomer ($FMN_B \rightarrow HEME_{A/B}$).[36]

$P450_{BM3}$ has been widely studied for biotechnology application because of its catalytic self-sufficiency. No cofactor protein is needed in the monooxygenase reaction. $P450_{BM3}$ shows a much higher activity compared to other fatty acid hydroxylases because all electron transfer steps occur

within the functional dimeric form and do not require collision encounters and adaptation of catalytically relevant binding orientations between the P450 enzyme and its electron transfer partner protein.[21,36] High activity is generally observed with medium-chain (C_{12} – C_{16}) fatty acids and the highest was recorded in the hydroxylation of palmitic acid (k_{cat} =122 ± 5.9 s^{-1}).[21, 37] The hydroxylation of fatty acid mostly occurred at sub-terminal positions – palmitic acid was oxidised at the ω–1 (31%), ω–2 (48%) and ω–3 (21%) positions (Fig. 1.9).[21]

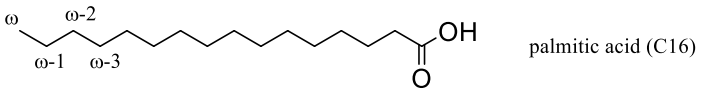


Figure 1.9 The oxidation position of palmitic acid.

An efficient method for $P450_{BM3}$ expression is important because a huge amount of enzyme is needed for industrial scale biotransformation. A fed-batch production of $P450_{BM3}$ with expression level of 12.6 μmoles L^{-1} (1.5 g L^{-1}) was reported in 2007 by Pflug *et al*.[38] Kaluzna *et al.* reported high expression levels of WT $P450_{BM3}$ (30 μmols L^{-1}) with NADPH-dependent glucose dehydrogenase in *E. coli* in a 1,000 L scale production. The production of 4-hydroxy-α-isophorone from isophorone was successfully conducted at a substrate load of 10 g/L with 61% conversion (TON = 13,200).[39] The kilogram-scale biotransformation and high-level expression demonstrated the potential of $P450_{BM3}$ in biotechnology applications.

1.4.1 Crystal structure of wild-type $P450_{BM3}$

The structure of catalytically active dimeric form of $P450_{BM3}$ has only been reported at low resolution with a cryo-EM image.[33] The crystal structures of both substrate-free (SF, Fig. 1.10-A) and substrate-bound (SB, Fig. 1.11) forms of the separately expressed BMP domain of $P450_{BM3}$ have been studied. There are 15 major helices and other β sheets and loops surrounding the heme domain which are located at the centre of the structure.

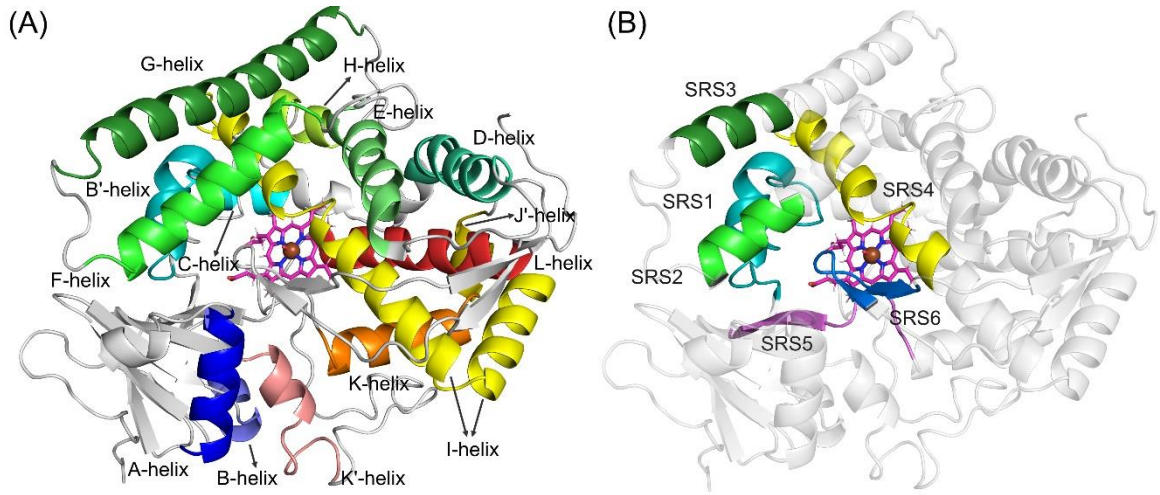


Figure 1.10 (A) The crystal structure of SF WT $P450_{BM3}$ (PDB code: 1BU7). (B) The location of SRSs.

The six substrate recognition sites (SRSs) determined by sequence alignment techniques across the P450 superfamily are scattered around the heme and contain roughly 16% of the total residues collectively (Fig. 1.10-B). SRS1 (69–92) is mainly aligned especially at C-terminus with some disruptions at N-terminus. SRS2 (181–188) and SRS5 (325–335) have some noticeable deviations while there is little deviation observed in SRS6 (434–441). Unlike SRS3 (200–208) which has changed almost entirely from the 15 sub-family members, SRS4 (253–271) has retained the conformation. The important residues that have been reported to bring significant substrate specificity changes of $P450_{BM3}$ are located at or around the SRSs. [21,40]

1.4.2 The active site and key residues of $P450_{BM3}$

By analysing the SB and SF structure of $P450_{BM3}$, some of the important residues for substrate recognition and binding were identified which are the basis of many new variants discovered later with enhanced oxidation activity for unnatural substrates. Based on the SB crystal structure, the carboxylate group of the *N*-palmitoylglycine (NPG) substrate is near the entrance of the substrate channel and interacts with Arg47 and Tyr51 (Fig. 1.11).[21] The aliphatic chain of the fatty acid substrate is extended and in contact with residues V78 and A82 across the heme. Interestingly, when mutations were introduced to residues Arg47 and Tyr51 individually, the activity of fatty acid

hydroxylation decreased with the increase of binding constant being observed. Together, the two resides were suggested as the 'gate keeper' of the substrate access channel and thus, double substitutions at both residues have been introduced.[21] Mutations R47L/Y51F increased polycyclic aromatic hydrocarbon (PAH) oxidation activity by up to 40-fold, compared to the WT which gave very low product formation rate.[41]

The bulky residue F87 is located above and almost perpendicular to the porphyrin ring. It was proposed that the access of substrate to the heme for hydroxylation was limited by the presence of residue F87 which in turn regulated the regioselectivity.[42] Carmichael *et al*. reported that mutation F87A increased the PAH oxidation activity drastically by creating more space above the heme domain and the effect was even more profound when combined with mutation R47L/Y51F.[41] Mutating the residue to valine (F87V) has also been studied and used extensively in enhancing the hydroxylation of various non-natural substrates by $P450_{BM3}$.[42]

Sitting directly opposite to residue F87, residue A328 is located close to the heme iron and thus, has been targeted for mutagenesis for activity and selectivity changes. Peters *et al.* reported the drastic effect of mutation A328V on the regio- and enantioselectivity for the hydroxylation of various alkanes.[43]

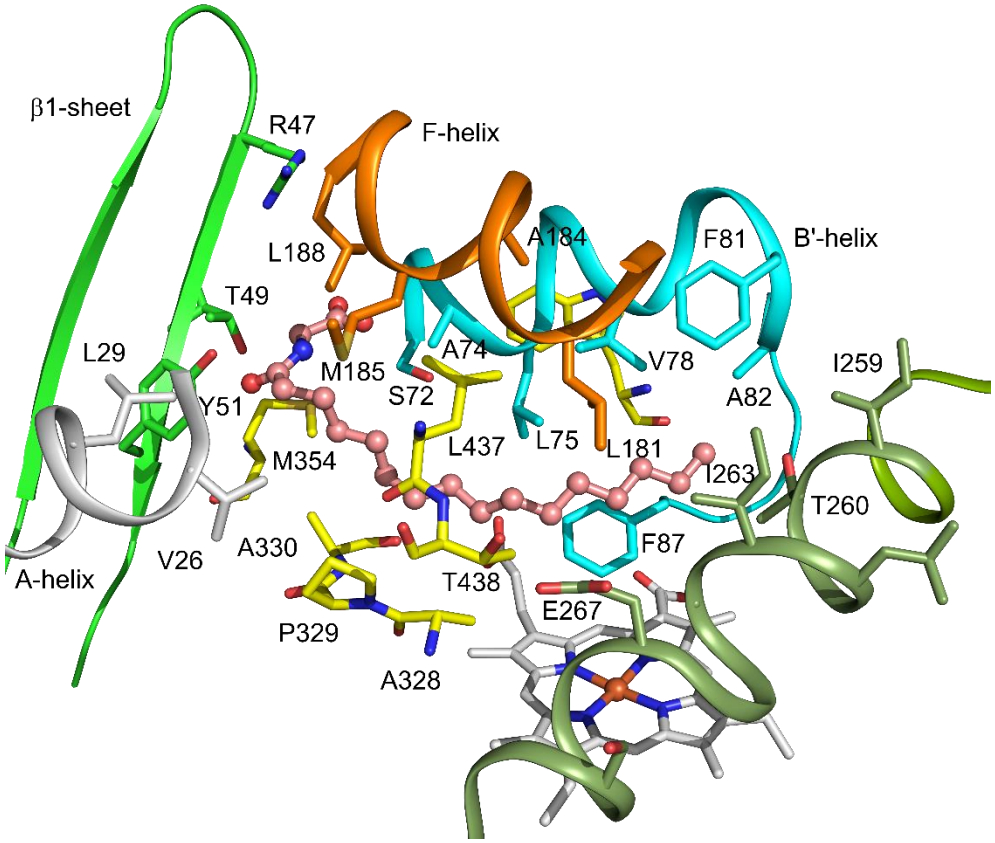


Figure 1.11 The active site of *N*-palmitoylglycine (NPG)-bound WT $P450_{BM3}$ (PDB code: 1JPZ) with selected residues.

The end of the aliphatic chain of NPG extended across the substrate pocket and was in contact with residue A82, as shown in the SB crystal structure of $P450_{BM3}$ (Fig. 1.11). The idea of "filling a hole" was proposed to introduce larger hydrophobic residues at A82 to fill in the pocket, which could lead to a smaller substrate pocket for improved unnatural substrate oxidation. Enhanced substrate binding of up to 800-fold and improved catalytic efficiency observed for variants A82W and A82F suggested that this residue was indeed crucial in increasing oxidation activity.[44]

1.5 Protein engineering of $P450_{BM3}$

1.5.1 Rational design and directed evolution

Over the years, $P450_{BM3}$ has been studied extensively because of its potential biotechnological applications. However, even though the natural substrates of the enzyme could be used as the precursors for some desirable compounds, there are many more compounds of interest which are not

the natural substrates of $P450_{BM3}$. The WT enzyme usually gives very little activity, if any, because of the structural, polarity and size difference between the natural and non-natural substrates.[21] Rational design and directed evolution are two different enzyme engineering approaches that have been widely used to design a mutant that gives desirable activity and selectivity towards target compounds with other important characteristics including high expression, good stability and robustness.

Rational design requires comprehensive understanding of the structure and function of the protein, usually based on the crystal structure or computational modelling, before deciding on the target regions and residues to be mutated. For example, residues R47, Y51, A82, F87 and A328 were the very first residues targeted for mutagenesis based on the SB crystal structure of $P450_{BM3}$ and all of them have been found to increase the activity and selectivity for numerous unnatural substrates.[41-44] However, crystal structure might not be the actual reflection of the transition state of the reaction and the detailed structure of a protein is not always available, thus making the mutagenesis process more difficult. Besides, the residues targeted based on rational design are usually in close proximity to substrate pocket and substrate channel while other residues would seldom be explored.

Remote residues, which do not appear to be important by rational design, could be explored by directed evolution. Site-saturation and various random mutagenesis methods, such as DNA shuffling and error-prone PCR are used to introduce random mutations at random residues which generate a huge number of mutants. For example, substituting a protein of 500 amino acids with all 20 amino acids would generate 20^{500} mutants; it is impossible to test and analyse each mutant individually. Therefore, a highly efficient and cost-effective method is needed to streamline and automate the process of the high-throughput screening and data analysis.

Li and co-workers had observed the formation of blue pigments at some colonies several hours after being induced.[45] As this phenomenon has not been observed in the expression of WT, the formation

of indigo, which was the hydroxylated product of indole from the breaking down of tryptophan in bacteria by P450, was hypothesised to be an indicator for the activity of the new P450 mutants.[21,45] Interestingly, the sequencing results of all the mutants which produced blue pigment were found to have mutations at A74, F87 or L188. Successive site-saturation mutagenesis with the formation of indigo as activity indicator was performed at the three residues and a high activity triple-mutation mutant, GVQ (A74G/F87V/L188Q) was discovered. The activity of the GVQ variant on various polycyclic aromatic hydrocarbons, including fluorene and acenaphthylene, was up to 10000-fold higher than the WT enzyme.[46]

Using a screening assay similar to that for esterase directed evolution, Arnold and co-workers used an analogue of octane, *p*-nitrophenyl octyl ether (8-pnpane) as the substrate in a colorimetric assay used to screen a large number of mutants developed for the hydroxylation of medium-chain alkanes.[47–49] Upon hydroxylation, 8-pnpane decomposed to form octanal and *p*-nitrophenolate which could be monitored by its yellow colour (Fig. 1.12).

After five rounds of random mutagenesis, mutants with high activity for the oxidation of short--chain alkanes were evolved. The variant 139-3 (V78A/H138Y/T175I/V178I/A184V/H236Q/ E252G/R255S/A290V/A295T/L353V) showed enhance activity for all substrates tested. The turnover frequency for octane and propane oxidation increased to ~3000 min^{-1} and ~1000 min^{-1} from 80 min^{-1} and 15 min^{-1}, respectively, by the WT enzyme. Propane and butane were oxidised selectively to a single product.[48] Interestingly, many of the mutations were remote from the active site, except for V78A, emphasising the importance of directed evolution in tracking beneficial mutations away from the substrate pocket. After more rounds of random mutagenesis, variant 35E11 was developed which demonstrated high activity for propane oxidation (total turnover number, TON ~6000).[50]

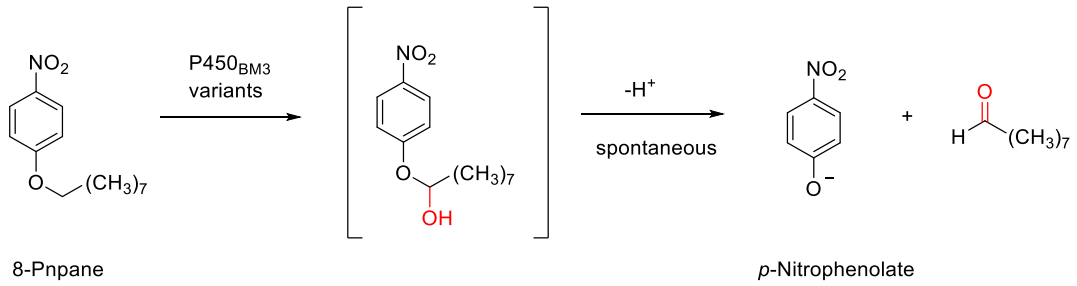


Figure 1.12 Schematic mechanism of the activity screening assay using *p*-nitrophenyl octyl ether (8-pnpane).

A domain-based protein engineering strategy was introduced to further increase the catalytic efficiency for propane oxidation (Fig. 1.13). In this approach, heme, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) domains of variant 35E11 were evolved separately via random and site-saturation mutagenesis. Beneficial mutations within the three domains were combined and a highly active propane monooxygenase ($P450_{PMO}$) was evolved with improvement of the TON to 45,000.[22, 51] These evolved mutants were also found to oxidise numerous drug molecules, including verapamil,[52] buspirone,[53] and astemizole.[52]

Further mutagenesis has led to the discovery of mutation C400S which shifted the characteristic ferrous-CO Soret peak of P450 from 450 nm to 411 nm.[16] The new variant was named cytochrome 'P411'. Its unique serine-heme ligation enabled efficient carbene insertion activity, such as olefin cyclopropanation by P411-CIS[16] and indole alkylation by P411-HF (Fig. 1.14).[54]

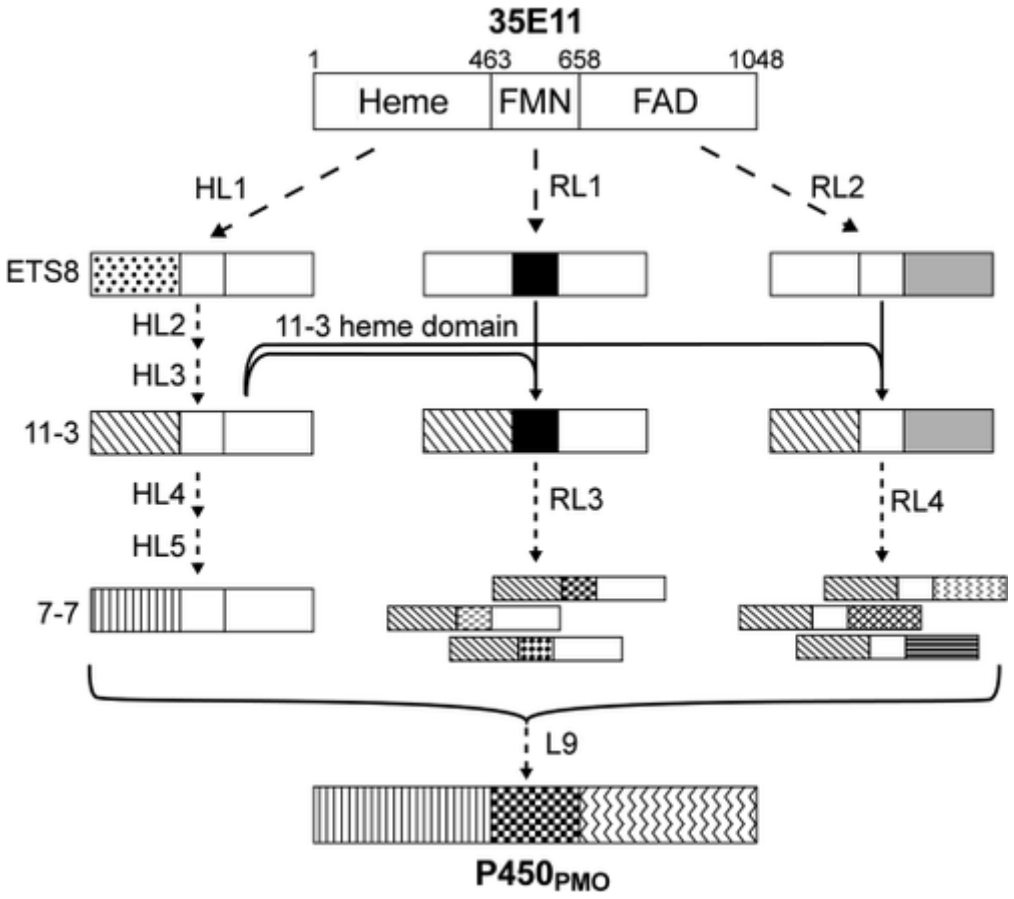


Figure 1.13 The schematic representation of the flow chart for domain engineering strategy with HL and RL representing heme and reductase domain libraries respectively.[22]

Rational design and directed evolution have different but complementary advantages and the combination of them as rational evolution has been studied widely in protein engineering. By utilising indigo formation as preliminarily activity screen, Wong *et al.* were able to identify high activity variants generated by ep-PCR on the BMP region of $P450_{BM3}$. Error prone PCR (Ep-PCR) introduces mutations into a piece of DNA by allowing the Taq polymerase to make mistakes in the base pairing during DNA synthesis, normally through the alternation of the composition of reaction buffer. Gene shuffling were performed on selected variants which were then subjected to the indigo assay again. There were four variants identified, A330P, A191T/N239H/I259V/A276T/L353I (KT2); F87A/H171L/Q307H/N319Y (KSK19); and F87A/A330P/E377A/D425N (KT5). The new variants showed enhanced activity on various substrates, even higher compared to GVQ variant. For example, the highest product formation rate of propylbenzene was 2688 min^{-1} by R47L/Y51F/KT2 compared

to GVQ at 1015 min^{-1} and WT at 606 min^{-1}. Variant R47L/Y51F/A330P demonstrated the highest product formation rate of naphthalene at 666 min^{-1} while WT showed almost no activity to the substrate. The new variants that gave significantly improved activity, could be used as the basis for further mutagenesis.[55]

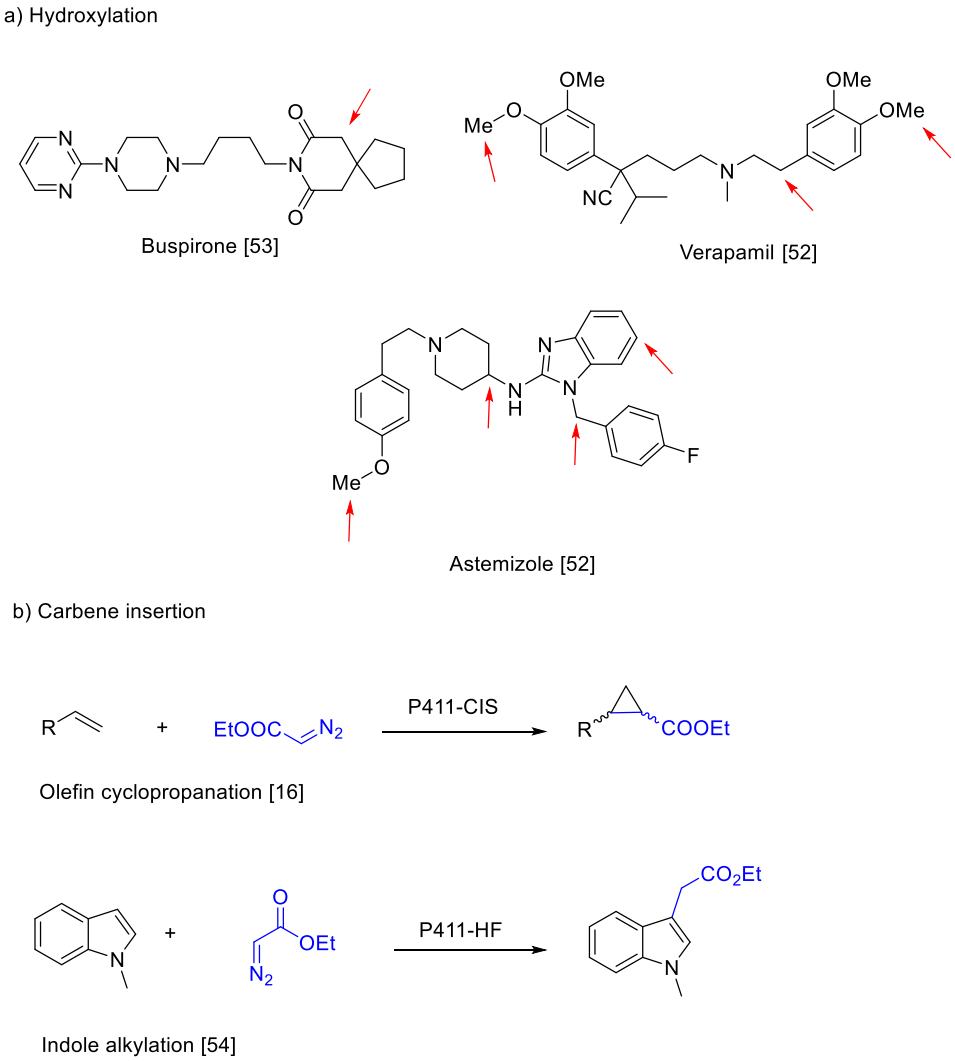


Figure 1.14. (a) Examples of the hydroxylation of various drug molecules by evolved $P450_{BM3}$. Red arrows represent the hydroxylated positions by the variants. (b) Examples of carbene insertion activity catalysed by P411 variants.

Although proline is structurally disruptive and could cause misfolding in protein, the discovery of the single mutation variant, A330P with high activity has inspired the introduction of proline residues at a few selected residues of $P450_{BM3}$. Among the tested variants, I401P showed significantly improved product formation rate of propylbenzene at 3578 min^{-1}, much higher compared to

R47L/Y51F/KT2 and GVQ.[56] To study the effect of the mutations in greater detailed, both variants A330P and I401P were crystallised and the crystal structures were overlaid with the SF and SB structures of the WT enzyme (Fig. 1.15 and 1.16).

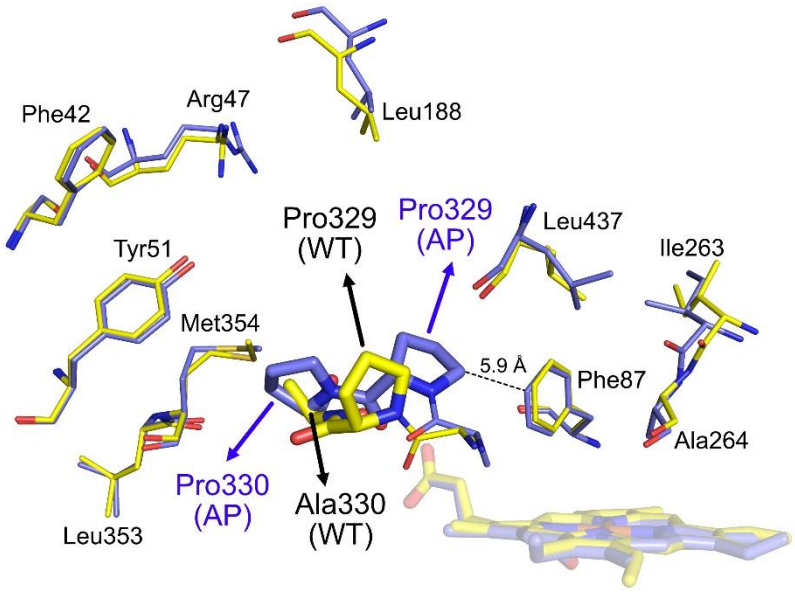


Figure 1.15 Overlay structures of the active-site residues of variant A330P (in blue, PDB: 3M4V) and WT (in yellow, PDB: 1BU7).

The introduction of mutation A330P has resulted in the shifting of the proline ring of residue P329, which is part of the access channel lining, towards the heme iron (Fig. 1.15).[57] The distance between residue P329 and the side chain of residue F87 was reduced to 5.9 Å compared to 9 Å in WT. With the substrate pocket becoming more constrained, the mutant showed increased activity towards smaller, non-natural substrates, such as propylbenzene and naphthalene, with altered products profile compared to WT.[55]

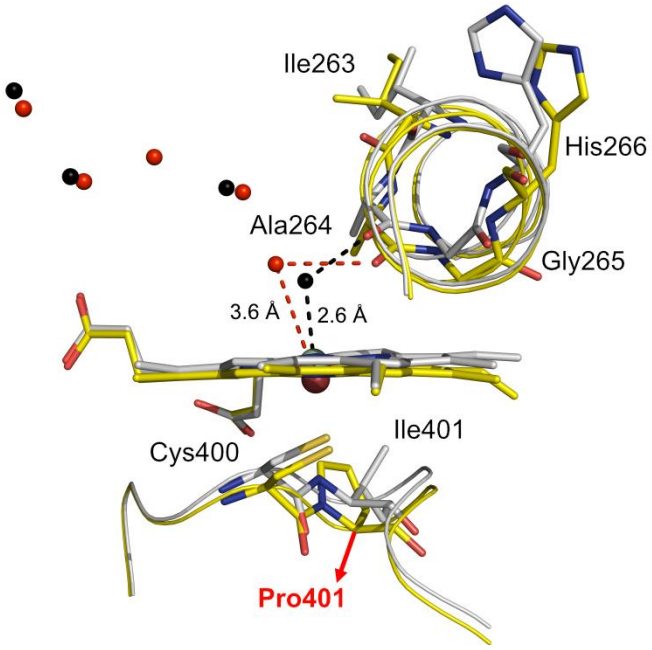


Figure 1.16 Overlay structures of the active-site residues of variant I401P (in yellow, PDB: 3HF2) with water molecules in red and SF WT (in grey, PDB: 1BU7) with water molecules in black.

Compared to SF WT, the I401P mutation resulted in the sinking of adjacent residue C400 (Fig. 1.16). The heme iron moved away slightly from the porphyrin ring and the distance between the iron atom and the axial water ligand increased from 2.6 to 3.6 Å. The displacement of this water molecule upon substrate binding became much easier compared to WT. This conformation showed remarkable resemblance with SB WT which explained the enhanced activity properties of variant I401P; the mutant was in a "catalytically ready" conformation when in the SF form.[56]

1.5.2 Combinatorial and iterative saturation mutagenesis

One of the major drawbacks of using enzymes as biocatalysts compared to conventional synthetic organic chemistry is that enzyme is usually evolved for certain groups of substrates unlike organic catalysts which could be used for a wide range of different compounds. Common directed evolution techniques like ep-PCR, site-saturation and DNA shuffling have been found useful in expanding the substrate acceptance but these methods usually produce a large number of variants which would require a highly efficient screening and analysing methods. In order to reduce the library size,

Combinatorial Active-site Saturation Test (CAST) has been developed by Reetz and co-workers to narrow down the targets by generating a more active-site-focused variant library guided by the crystal structure of the enzyme.[58,59] The idea of the method is to select and randomise several sets of two or more amino acids which are in close proximity to each other. For example, if the selected residue is at residue n, then the other one in the pair will be at (*n*+1) in a loop, (*n*+2) in a β sheet, (*n*+3) in a 3_{10} helix, and (*n*+4) in an α helix, depending on the secondary structure of the selected residues. The synergistic effect of mutating the residue sets could potentially bring significant effect in the oxidation of substrates which would otherwise not be observed with single mutation.

Once effective sets of mutations have been identified, the activity could be further improved by combining the improved mutants. It is a rather effective and quick method, but there is a limit for the degree of activity improvement because of the restricted structural diversity of the mutants.[59] Reetz and co-workers have then developed another strategy, iterative saturation mutagenesis (ISM) which would normally be combined with CAST for further mutagenesis. Once a mutation with improved activity has been identified, it would be used as the template for the next round of saturation mutagenesis at other sites, and the process is continued until desired activity has been achieved.[59,60]

With the combination of CAST and ISM, Reetz and co-workers have successfully evolved new mutants that gave 96%–97% regioselectivity of 2β and 15β oxidation of testosterone (TST) with good conversion, starting from variant F87A which gave a roughly 1:1 mixture of the 2β- and 15β-alcohols. Twenty residues around the active site were selected and grouped into nine sites, each with two or three residues to increase the likelihood of synergistic effect. Variant A330W/F87A, which was evolved by introducing NNK codon degeneracy at residues S72, A328 and A330, gave 97% selectivity for the 2β-alcohol at 79% conversion. The most selective mutant for the 15β-alcohol was V78L/A82F/F87A at 91% selectivity and 86% conversion from library B which V78 and A82 were

randomised by using NNK codon degeneracy. Variant R47Y/T49F/F87A, the most active and highest 15β selectivity variant from library A (R47, T49 and Y51), was selected as template for randomisation at site B with ISM. Selectivity for the 15β-alcohol was increased to 96% by the mutant R47Y/T49F/V78L/A82M/F87A.[60]

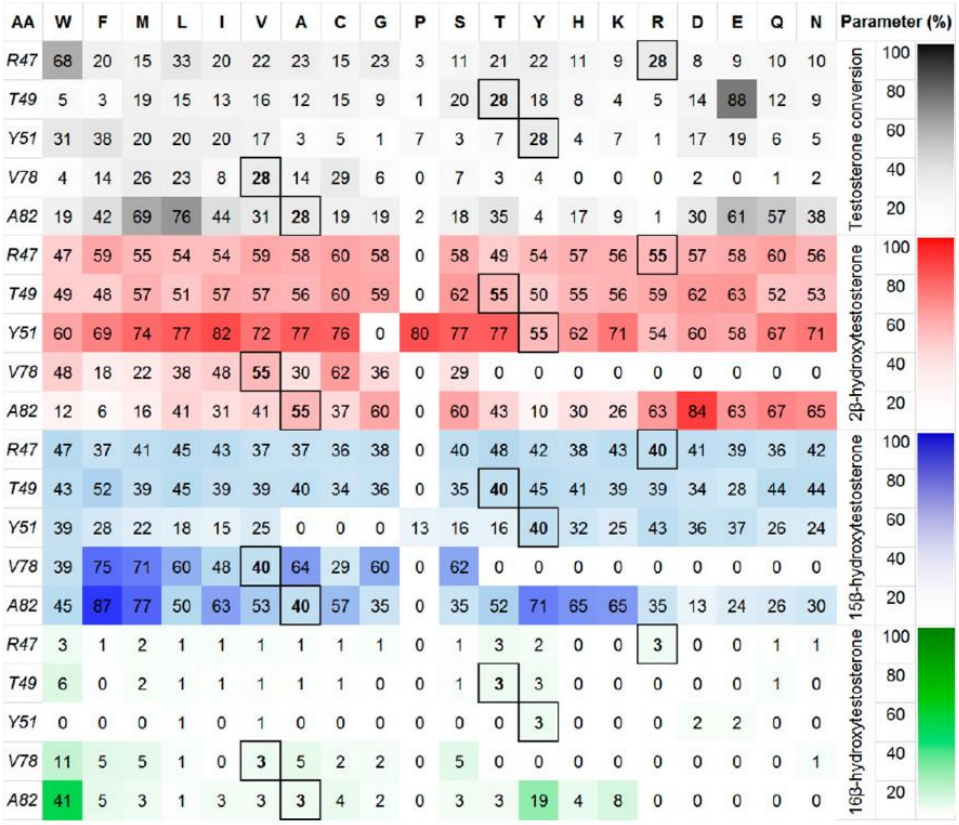


Figure 1.17 Mutability landscape of variant F87A of $P450_{BM3}$ towards testosterone (TST). The five target residues are shown on the left while the scales of each parameter are displayed on the right. Black represents TST conversion while red, blue and green demonstrate the selectivity for 2β-, 15β- and 16β-hydroxytestosterone respectively. The values of the parent enzymes are shown in squares.[61]

Trade-off effects is one of the most commonly observed side effects in enzyme engineering where one desired parameter was achieved with the compromise of other parameters including selectivity, conversion, activity, enzyme stability, expression, *etc*. Many different methods have been studied to tackle the problems, but enormous enzyme libraries and screening effect are usually required.[61] Inspired by Rost and co-workers, Reetz and co-workers have further improved the process by incorporating mutability landscapes (MLs) in selecting mutations for ISM, guided by molecular

dynamics computations.[61–63] Unlike conventional methods which usually study the positive mutations only, MLs provide a fuller picture of the effect of each amino acid substitution on various parameters (selectivity, conversion, enzyme stability, *etc.*) by including all the information systematically in a map, regardless of whether they are beneficial, neutral or detrimental (Fig. 1.17). The data is then used as a guidance for further mutagenesis and mutations which are more likely to have the trade-off effect could also been identified and avoided. The strategy has drastically decreased the number of variants screened (~3000 variants) and high selectivity for 16α (96% by R47W/Y51W/S72I/A82F/F87I/L181C) and 16β (92% by R47W/A82W/F87V/L181Q/T436R/M177S) oxidation of testosterone was observed.[61]

Using the uracil-specific excision reagent (USER) method, 15 residues were randomised to 2 amino acids each and a large library with roughly 30,000 variants were created. Variant LG-23 (R47W/S72W/F77Y/V78L/F81I/A82L/T88S/M177T/M185Q/L188Q/I209T) showed 90% selectivity for 7β oxidation of TST with almost full conversion.[64]

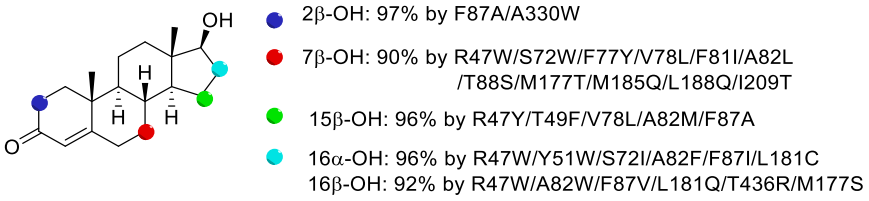


Figure 1.18 Hydroxylation sites of testosterone (TST) by engineered $P450_{BM3}$ variants developed by Reetz and co-workers through the combination of CAST, ISM, MLs and USER.[60,61,64]

1.5.3 The mutant library approach

Although many different protein engineering approaches have been developed, enormous efforts on mutagenesis and screening are still required to develop mutants with high activity and product selectivity. Seifert *et al.* reported improved regio-, stereo- and chemoselectivity for a range of different terpene substrates with a variant library of only 24 variants. This minimal and highly

enriched mutant library was constructed by mutating the side chains of residues F87 and A328 systematically with five different hydrocarbon amino acids. As the side chains varying in size, it demonstrated the effect of tuning the binding site shape to accommodate different substrates. Generally, variants with single mutation showed improved selectivity on smaller acyclic substrates while double-mutation variants were needed for cyclic or bicyclic substrates which were bigger in size.[65]

Wong and co-workers have developed a screening library of roughly 100 mutants which has been demonstrated to be highly active and selective for the oxidation of a wide range of substrates including steroids (testosterone and androstenedione),[66] sesquiterpenes (valencene), drug molecules (lidocaine and amitriptyline),[67] core motifs in drug molecules (quinolines and tetrahydroquinolines)[68] and diterpenes (eleuthoside).[69] The highly diverse mutant library was based on five variants, A330P, K19 (H171L/Q307H/N319Y), RP (R47L/Y51F/I401P), RT2 (R47L/Y51F/A191T/N239H/I259V/A276T/L353I), and GVQ (A74G/F87V/L188Q).[66–70] Various mutations at the active sites were then introduced to the base mutants, resulting in the generation of a panel of variants with different binding pocket sizes that would match with various unnatural substates for oxidation.

Fasan and co-workers have developed a three-step methodology in designing a mutant library that has roughly 75 variants only but gave high activity and product selectivity to the target compound, artemisinin. One of the biggest implications of this strategy is that it could potentially be used as a general method for a wide range of substrates. Starting with active sites mutagenesis and high-throughput P450 fingerprinting which are independent of the substrate, the final step, fingerprint-driven P450 reactivity predictions could be applied on target compound easily.[71,72] Variants II-H10 and IV-H4, which showed 100% selectivity to 7(*R*)- and 7(*S*)-alcohols of artemisinin respectively, were identified from the method, starting from the parent mutant which only gave 83% and 10%.[71] Since the hydroxylation of artemisinin at C7 remains inaccessible with chemical approaches, the

production of 7(*R*)-alcohol of artemisinin served as an alternative for the synthesis of enantiopure C7-fluorinated derivatives of the clinical antimalarial drugs artesunate and artemether (Fig. 1.19).[71,72]

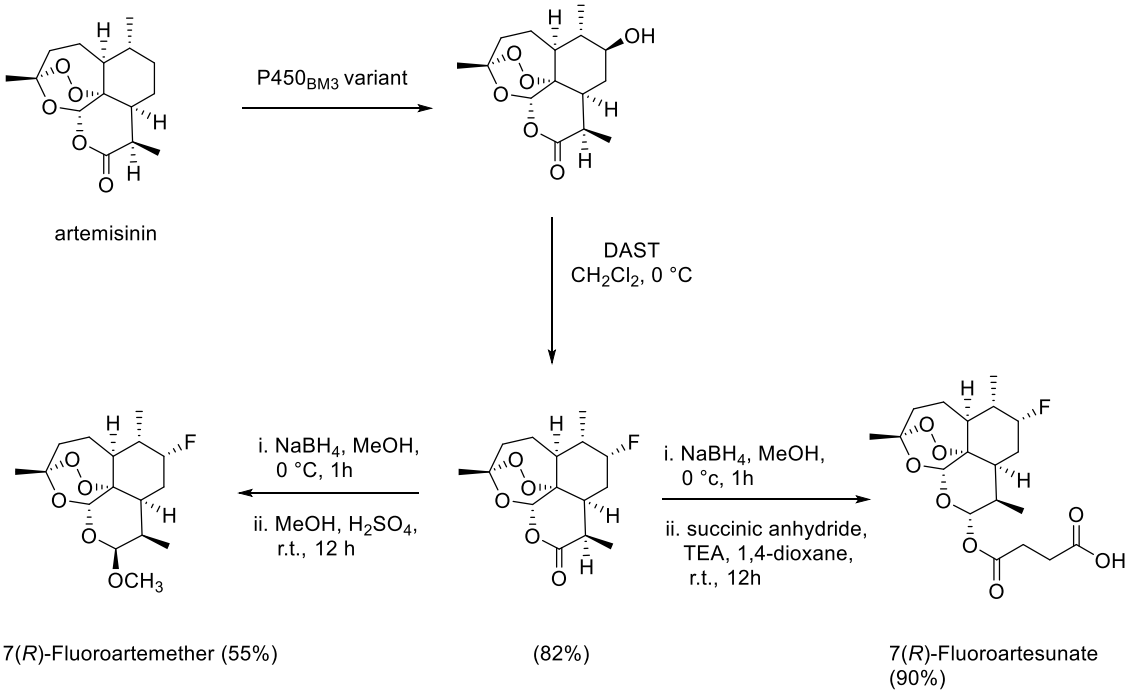


Figure 1.19 Chemoenzymatic synthesis of (*7R*)-fluoroartemether and (*7R*)-fluoroartesunate via selective C7 oxidation of artemisinin by engineered $P450_{BM3}$ variants.[71,72]

1.6 Molecular dynamics simulations and substrate docking

1.6.1 Molecular dynamics simulations

Although significant advances in protein engineering and directed evolution have been made, many of these approaches still require huge efforts in screening and mutagenesis which could be reduced by a deeper understanding of the enzyme-substrate binding interaction to guide mutation design. Many important mutations were discovered through the analysis of crystal structures, but its potential is limited because enzymes are dynamic systems and may undergo conformational changes during catalysis.[73] The conventional method of understanding the conformational changes of a macromolecule is by gathering experimental structures covering the conformational space. For

example, the crystal structures of a protein with different conformations could be generated under various conditions and environments or by collecting the conformational ensembles through NMR experiments.[73] These methods are laborious, and the effort required escalates for a more complete picture of the macromolecular dynamic properties. With the increasing power of computational hardware, molecular dynamics (MD) simulations have evolved as a powerful and convenient alternative to provide insight into the structural features of enzymes and enzyme-substrate interaction.

MD simulations predict how every atom in a molecular system, or a protein, moves over time, based on a classical mechanics treatment of interatomic interactions.[73–75] The first MD simulation was conducted on gas molecules in 1957 by Alder and Wainwright.[76] MD methods were first applied in the late 70s to biological systems to study the dynamics of globular proteins such as lysozyme and bovine pancreatic trypsin inhibitor.[77,78] Evolving from the simulation of a few atoms in the gas phase to complex biological systems solvated with solvent, simulations of 50,000–100,000 atoms are now routine.[73]

MD simulations calculate the forces exerted on each atom by all other atoms in a biomolecular system. The initial model is obtained with various experimental methods, such as X-ray crystallography, cryo-EM and NMR, or homology modelling. As the environment plays a vital role in determining the properties and reactivity of the biomolecular system, the initial model is then subjected to solvation. Explicit-solvent methods like Particle mesh Ewald (PME) with TIP3P water is the most reliable method by far because they include all pairs of solute and solvent atoms in the calculation.[79,80] Although reasonable dynamic and thermodynamic properties could be achieved with this approach, it is computationally intensive and most of the required time is for the calculation of solvent-solvent interactions instead of the solute of interest. To overcome the limitations of explicit-solvent methods, many implicit-solvent methods have been developed, such as the generalised Born model, which treats the discrete solvent molecules as continuums and speed up the simulation

drastically.[81–83] However, with the introduction of GPU-accelerated methods and hardware, explicit solvent treatments are now routine.

Once the solvation step is completed, force field equations are applied to obtain the forces acting on all atoms and deduce the potential energy from the molecular structure. Many different force fields with various parameters have been used but they share similar features such as spring-like terms for bond length, Coulomb's law for electrostatic interactions, Lennard-Jones potentials for van der Waals interactions, and periodic functions for bond rotation. By applying Newton's laws of motion, the spatial position of each atom is predicted and a trajectory which shows the conformation of the system at every atom during the simulations is generated (Fig. 1.20). As biochemical processes normally take place in ns, μs or longer timescales, each time step in MD simulations is usually set to be a few femtoseconds, shorter than the fastest movement in molecules to avoid instability. Many different software have been developed for MD simulations, such as OpenMM,[84] NAMD,[85] CHARMM,[86] and GROMACS.[87]

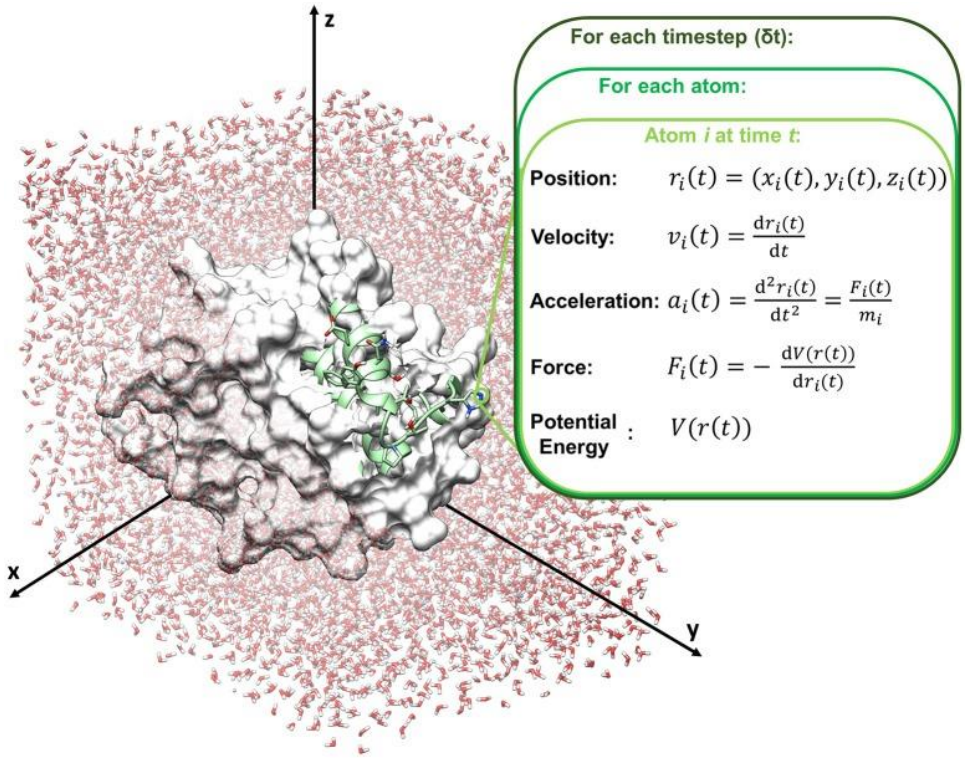


Figure 1.20 Schematic representation of a molecular dynamics cycle.[88]

1.6.2 Molecular docking

Molecular docking is used to model the binding of a small molecule (ligand) to a protein, which allows researchers to study the basic biochemical process by understanding the behaviour and binding orientations of the ligand in the target protein. There are two main steps in molecular docking: the generation of several possible orientations of the ligand within the protein binding sites by using a sampling algorithm, followed by a scoring function which ranks each predicted conformation based on a score given.[88,89]

The first sampling algorithm was developed by Kuntz *et al.* in 1982.[90] The samplings were made based on the steric overlap between receptor and binding site which both have been approximated by a set of spheres respectively.[90] This method is classified under the group of rigid docking, according to a categorisation which divides all docking methods into groups based on the degrees of flexibility of the molecules, including rigid docking (both ligand and protein are considered as rigid), semi-flexible docking (only the ligand is considered as flexible) and flexible docking (both ligand and protein are considered as flexible).[88] Most early docking approaches are rigid due to the limited computational power while most approaches are semi-flexible docking, such as Autodock Vina,[91] GOLD,[92] and DARWIN.[93] Although flexible docking could be more representative of what is happening in nature, flexible docking is particularly challenging as it requires much longer computational time. More recently, docking programs such as ADFR have been developed to define several selected residues to be flexible, depending on their roles in interacting with the ligands.[94]

Various sampling algorithms have been developed since then. By representing both protein and ligand as pharmacophores, the ligand is mapped into the binding pocket of a protein in terms of shape features and chemical information through matching algorithms (MA) method. Each distance within the protein and ligand is calculated for a match while new ligand conformations are governed by the distance matrix between the pharmacophore and the corresponding ligand atoms.[95–97] With the incremental construction (IC) method, the rotatable bonds of a ligand are broken to divide the

ligand into different fragments. The largest fragment is docked into the active site of the protein while the remaining fragments are added incrementally to generate ligand with different orientations.[98–100] Both multiple copy simultaneous search (MCSS) and LUDI are fragment-based approaches with the former focusing on fragments with different functional groups while the latter prioritises the formation of hydrogen bonds or by filling a hydrophobic pocket. Once different fragments are fitted into the binding sites, they would be connected to form a new single molecule.[101,102] Through bond rotation, rigid-body translation or rotation, Monte Carlo (MC) methods generate various poses which would be tested with an energy-based selection criterion. The process is iterative where the poses that 'survived' (i.e., exceeded a threshold) would then be subjected to another round of modification until the predefined quantity of conformations is collected.[103,104] The idea of genetic algorithms (GA) is derived from Darwin's theory of evolution. Degrees of freedom of the ligand are encoded as binary strings called genes while these genes make up the poses of the ligand, known as chromosomes. There are two kinds of genetic operators which would affect the genes and result in new ligand structure: random mutation of the genes and crossover exchanges of the genes between two chromosomes. New structures will be subjected to scoring function and the poses which have exceeded the criteria would be used for the next generation.[105–107] Examples of applying the GA methods include AUTODOCK,[91] GOLD,[92] DIVALI,[108] and DARWIN.[93] A refined GA has recently been developed and implemented in the software AutoDock for Flexible Receptors (ADFR).[94] Molecular dynamics has also been used in molecular docking by moving each atom separately in the field of the other atoms and it provides larger flexibility to both the ligand and protein compared to other algorithms.[109,110]

Scoring functions are used to differentiate the correct poses from all the poses generated in a reasonable computation time. It involves estimation of the binding affinity between the protein and ligand with various assumptions depending on the software used. In general, scoring functions could be divided into three types, which are force-field based, empirical and knowledge-based.[88,89]

Classical force-field-based scoring functions assess the binding energy by calculating the sum of the non-bonded (electrostatics and van der Waals) interactions. It has relatively slow computational speed and thus cut-off distance is used to handle the non-bonded interactions which might bring adverse effect to the accuracy of the long-range binding effects.[111–113] The extended versions, which are commonly used in GOLD[92] and AutoDock,[91] also take hydrogen bonds, solvation and entropy contributions into account. Empirical scoring functions decompose the binding energy into several energy components, such as hydrogen bonding, ionic interactions, the hydrophobic effect and binding entropy. The final score recorded is the sum of each component multiplied by a coefficient obtained from regression analysis fitted to a test set of ligand-protein complexes with known binding affinities.[114–118] Examples of software with empirical scoring functions include LUDI,[102] PLP,[114,115,119] and ChemScore.[120] With knowledge-based methods, the crystal structures of the ligand-protein complexes are analysed statistically to obtain the interatomic contact frequencies and the distances between ligand and protein.[121–125] The fundamental assumption is that the more favourable an interaction is, the greater the frequency of occurrence will be. The score is calculated by weighting more the favoured contacts and penalising the repulsive interactions between each atom in the ligand and protein within the pre-defined cut off. PMF[121], SMoG,[123] and Bleep[122] are based on knowledge-based functions. Consensus scoring is a strategy that combines several different scores, and a pose would be considered as correct if it scores well under a number of different scoring schemes.[126]

1.6.3 Application in enzyme engineering

Being one of the most widely used methods in structure-based drug design, the combined application of MD simulations and molecular docking have been explored recently to aid in enzyme engineering. With the discoveries of two asymmetric units of the crystal structure of $P450_{BM3}$, Paulsen and Ornstein performed the first MD simulated structure of $P450_{BM3}$ in 1995 using the SF crystal structure. The result demonstrated that the substrate access channel of $P450_{BM3}$ is highly mobile, and the range of substrates could be accepted is broader than expected.[127]

To date, molecular docking has mainly been used to investigate the origins of improved enzymatic activity and product selectivity while its application as guidance in protein engineering has only started to be explored in the past few years. With the discovery of highly selective (>95%) mutants for 2β- and 15β- oxidation of testosterone, Reetz and co-workers performed MD simulations and molecular docking of the variants KSA-2 and KSA-14, together with the precursor variant, F87A.[60] The docking result of F87A showed two active positional orientations for 2β- and 15β-hydroxylation which agrees with the experimental data of 1:1 mixture of 2β- and 15β-alcohols. On the other hand, the docking of testosterone into variant KSA-2 gave the binding pose with the 2β hydrogen pointed to the haem iron as the only pose produced while variant KSA-14 showed the substrate to be aligned differently with the 15β hydrogen in ideal proximity to the haem; both showed good agreement with the experimental result.

By overlaying the crystal structure of WT $P450_{BM3}$ with androstenedione (AD) bound structure of steroid C19-demethylase, CYP19A1, Wong and co-workers identified regions which blocked the binding orientation for C19 oxidation of AD in WT $P450_{BM3}$. A panel of variants was designed, mainly focusing on I helix and β_1 strand. The molecular docking of androstenedione (AD) with one of the most selective variants for 2β oxidation of AD, K19/F87A/A82M/I263G/A264G (59%), revealed that the dominant 2β poses showed significant resemblance to the AD-bound structure of

CYP19A1. As C2 and C19 of AD are remarkably close in proximity, Chen *et al*. have demonstrated that the variants developed from nature-inspired mutagenesis, mimicked the binding features of CYP19A1 (Fig. 1.21).[66]

In 2014, Pleiss developed a general strategy which combined data mining of enzyme families and molecular modelling of enzyme-substrate interactions to design a focused library for improved regio- and stereoselectivity.[128] Based on the analysis of the sequence of over 6300 *CYP* genes and the relevant MD simulations, three hotspots (A264, A328 and L437) were identified in $P450_{BM3}$. After two subsequent rounds of molecular modelling and variant screening, a highly selective variant, A328V/L437F/A264V was generated, demonstrating 97% selectivity for the C7 oxidation of (*4R*)-limonene to perillyl alcohol.

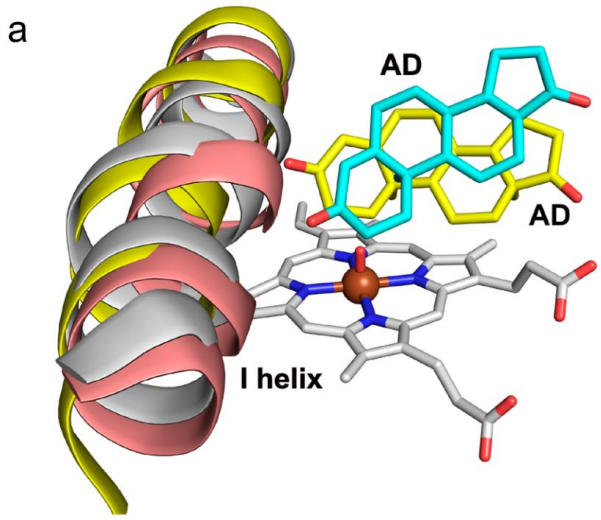


Figure 1.21 The overlay MD energy minimised structures of WT $P450_{BM3}$ (salmon), CYP19A1 (yellow) bound with AD (yellow), and K19/A82M/F87A/I263G/A264G (grey) with AD (cyan) docked in a 2β pose.

To obtain an enzyme for C11 oxidation of 9-dehydronandrolone, Li and co-workers performed MD simulations and molecular docking on the previously generated LG-23 mutant (R47W/S72W/F77Y/V78L/F81I/A82L/F87G/T88S/M177T/M185Q/L188Q/I209T/A328G/

A330W).[64] By comparing the productive and non-productive poses, they discovered that the side-chain methyl group of T438 moved inwardly and disrupted the hydrogen bonds between 9-dehydronandrolone and L439 and T327. As a result, the binding orientation of 9-dehydronandrolone located C11 away from the haem iron. A new variant, LG-23/T438S with improved 11α selectivity (78% to 92%) was developed by subjecting the precursor mutant to NNK-based site saturation mutagenesis.[129]

Zhang *et al.* developed an attractive methodology to increase enantioselectivity and enzyme activity by utilising the result of molecular docking as guidance in designing mutations.[130] The basic idea of the approach is to design mutations to disfavour binding poses leading to unwanted enantiomers to increase enantioselectivity and disfavour non-productive poses for enzyme activity enhancement. Each generation of iterative docking guided mutagenesis (IDGM) required roughly 10 variants to achieve significant improvement in enantioselectivity and enzyme activity. High enantioselectivity (>90% e.e.) was achieved at all unactivated C–H bonds of various single-ring, bicyclic and spirocyclic amines with engineered $P450_{BM3}$ variants (Fig. 1.22). The alcohol products are important synthetic intermediates for drugs and functionalised fragment molecules.

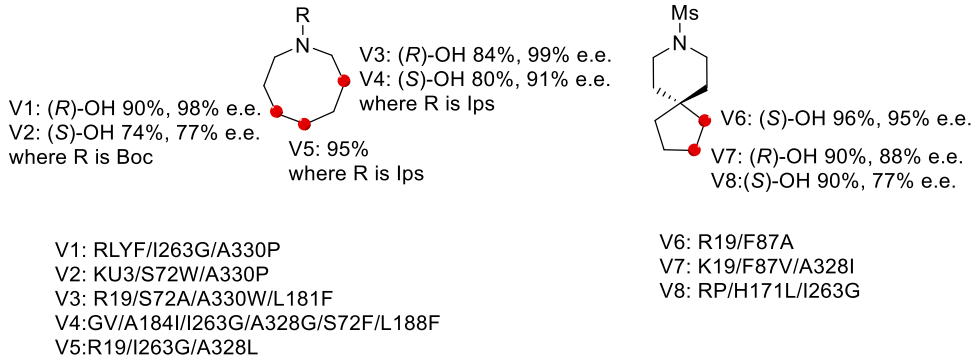


Figure 1.22 The oxidation products of various cyclic amines modified with different protecting groups by engineered $P450_{BM3}$.

1.7 Terpenoid compounds

Terpenes and their derivatives are produced as metabolites during the metabolic processes of plants, animals and microorganisms. They are the largest class of natural products and the major constituents of secondary metabolites.[131] More than 70,000 terpenoids have been recorded in the Dictionary of Natural Products with a wide range of different hydrocarbon skeletons and oxygenated functional groups.[132] Many of them show interesting biological activities and have been studied extensively for centuries for different purposes (Fig. 1.23). As the major aromatic components in plant essential oils, they are often used as aroma sources in the cosmetics and perfume industry as well as flavourings in food, e.g., nootkatone for the grapefruit flavour, rose oxide for the rose fragrance and patchoulol in patchouli oil. Many of them show pharmaceutical and physiological activity, for example, artemisinin, a widely used antimalaria drug and the anticancer compound Taxol.[133,134]

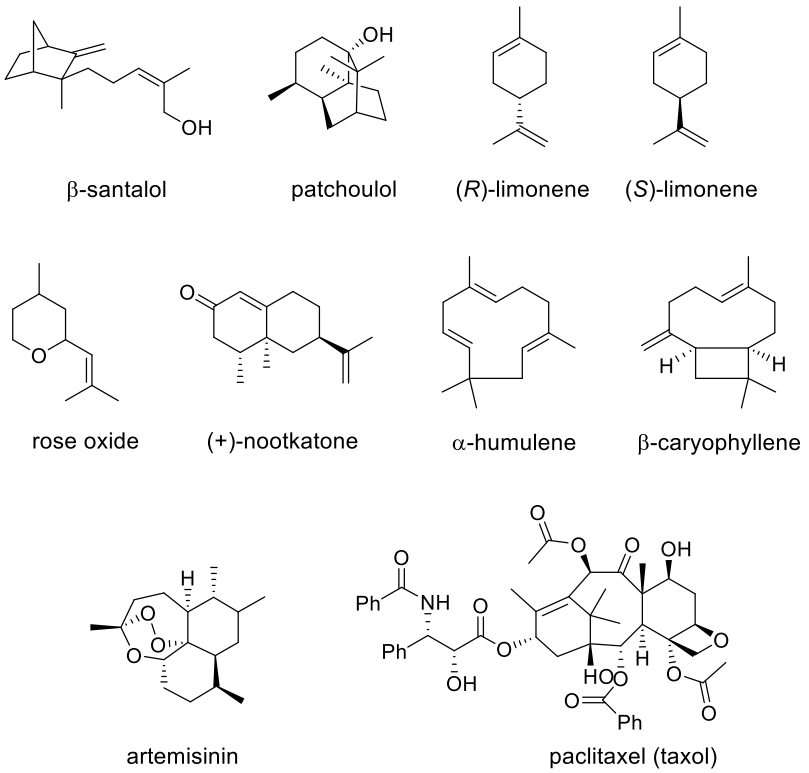


Figure 1.23 Examples of terpenoid compounds.

Isopentenyl pyrophosphate (IPP) is the precursor for all terpenes. The 2-*C*-metyl-D-erythritol-4-phosphate pathway (MEP), also known as the DXP pathway, is normally present in bacteria and

plant plastids. In this pathway, IPP is synthesised from D-glyceraldehyde-3-phosphate and pyruvate. DXP pathway enzymes are targets for drugs because eukaryotes do not have these enzymes. In eukaryotes and archaea, IPP is biosynthesised by the mevalonate pathway (MVA) which produces mevalonate from acetyl CoA (Fig. 1.24a). In the presence of an isomerase, IPP isomerises to form dimethylallyl pyrophosphate (DMAPP). Loss of pyrophosphate from DMAPP followed by the 'head-to-tail' coupling with another IPP molecule gives geranyl pyrophosphate (GPP), precursor of all monoterpenes. The coupling of another IPP unit to GPP forms the precursor of all sesquiterpenes (C15), farnesyl pyrophosphate (FPP). Loss of pyrophosphate and coupling continues to form prenyl diphosphate molecules of different chain lengths which are the precursors of all terpenes based on the number of isoprene (C5) units (Fig. 1.24b).[131]

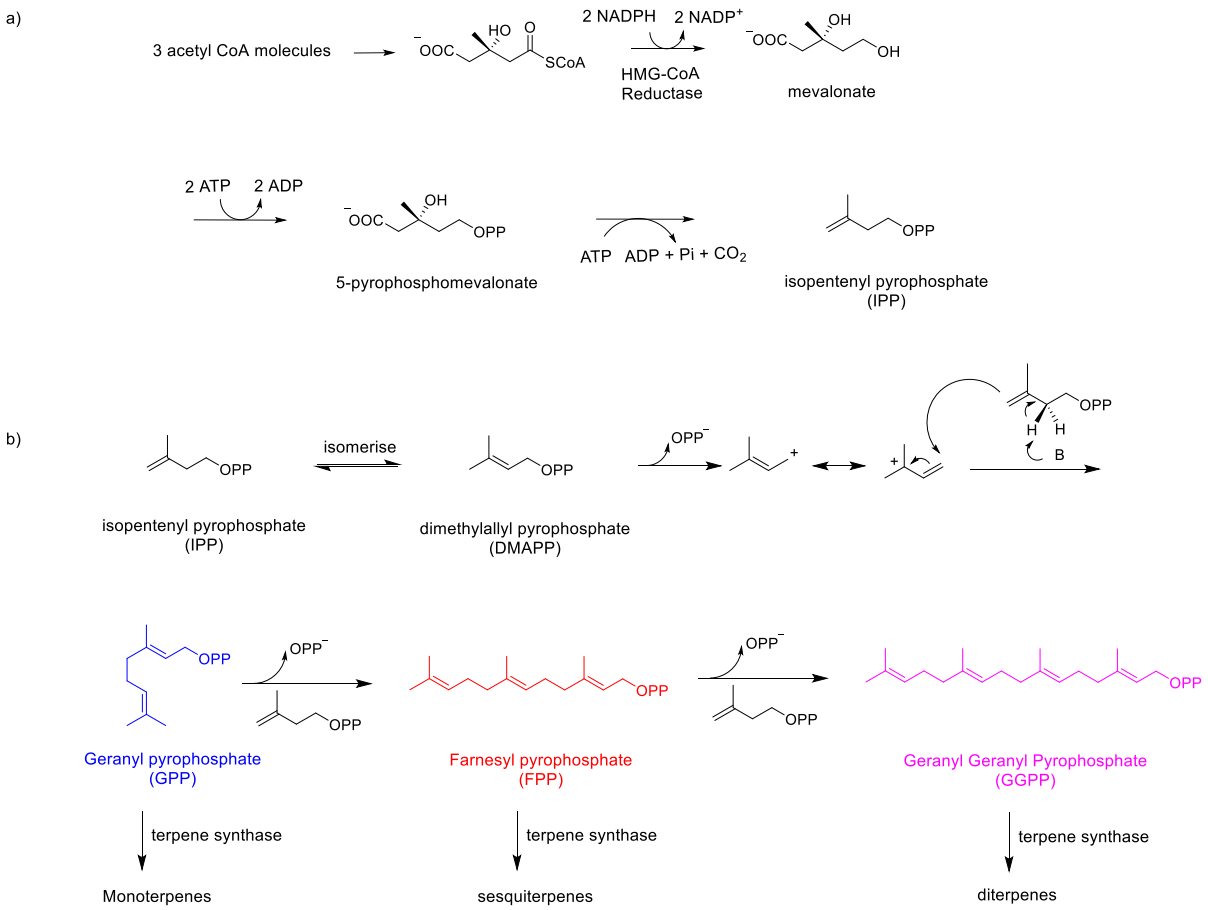


Figure 1.24 Biosynthesis of terpenes. a) The synthesis of isopentenyl pyrophosphate (IPP) from acetyl CoA via the mevalonate pathway (MVA). b) The synthesis of different terpenes from the universal precursor, IPP.

Catalysed by a terpene synthase, the pyrophosphate group on FPP is lost and a carbocation intermediate is generated. The carbocation intermediate then undergoes rearrangement followed by deprotonation or water capture to form a wide variety of terpenoids in nature. Deguerry *et al.* have reported that a patchouli synthase contributes to the synthesis of 14 different sesquiterpenes in patchouli oil including the major component, patchoulol, as well as α-guaiene and α-bulnesene, the precursors to a family of aromatic compounds found in agarwood (Fig. 1.25).[135]

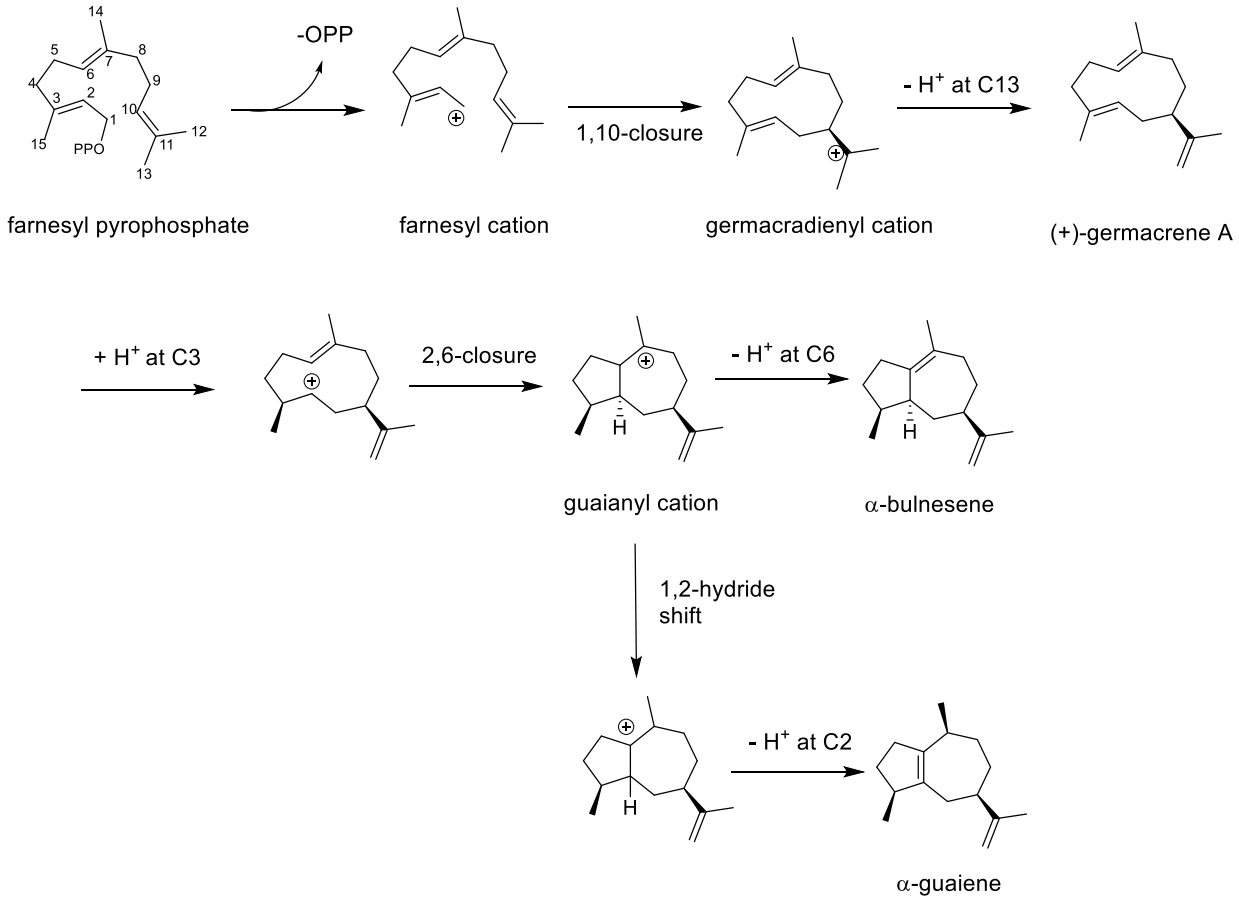


Figure 1.25 Different products synthesised by patchoulol synthase from farnesyl pyrophosphate.

The carbocation formed after the ionisation of farnesyl pyrophosphate undergoes 1,10-closure to form the germacradienyl cation (Fig. 1.25). Deprotonation of this cation forms (+)-germacrene A, one of the aromatic sesquiterpenes found in patchouli oil. Protonation of germacrene A at C3 allows the carbocation intermediate formed to undergo ring closure between C2 and C6 to form the guaianyl carbocation. Deprotonation of the guaianyl carbocation at C6 will from α-bulnesene or the guaianyl

carbocation undergoes a 1,2-hydride shift followed by deprotonation at C2 to form α-guaiene (Fig. 1.25).[135]

Late-stage oxy-functionalisation of terpenes by P450 has been increasingly researched because of the potential use of the products as fine chemicals, drugs and fragrances. Classical organic synthesis approaches are often less desirable because it is difficult to functionalise complex molecules with high selectivity, especially at non-activated C–H bonds.[136]

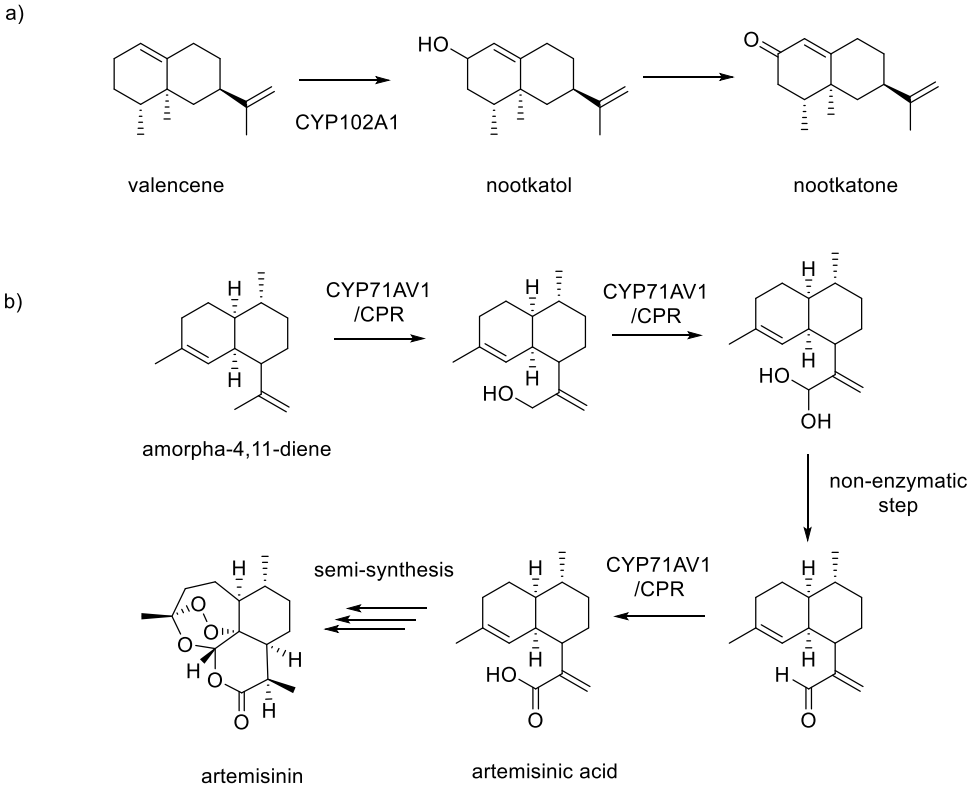


Figure 1.26 The synthesis of nootkatone and artemisinin via the hydroxylation by engineered $P450_{BM3}$.

The valuable sesquiterpene flavouring of grapefruit, nootkatone, is synthesised by the oxidation of valencene at position C2 to form the ketone (Fig. 1.26).[137] Such a process for producing nootkatone has been operating at commercial scale. The hydroxylation of steroid molecules by P450s provides the precursors for steroid drugs as well as metabolites with different pharmaceutical activities. An interesting example of synthetic biology and metabolic engineering involving P450 is the *de novo* synthesis of artemisinic acid, the precursor of artemisinin which has been used for thousands of years

for its antimalaria properties. A terpene biosynthesis pathway was assembled in yeast to synthesise amorpha-4,11-diene from glucose. CYP71AV1 was introduced for the oxyfunctionalisation of amorpha-4,11-diene to artemisinic acid via the intermediates artemisinic alcohol and artemisinic aldehyde (Fig. 1.26).[138]

1.8 Agarwood

Distributed in Southeast Asia, agarwood is a type of valuable resinous heartwood obtained mainly from *Aquilaria* species.[139] It has been used for centuries as an important incense material for religious worship purposes. It also plays essential roles in traditional medicine because of the various physiological effects it shows such as cough and high fever relief, antimicrobial and antitumor activity.[139,140] More recently, the essential oil of agarwood has also been widely used in perfumery and the fragrance industry for its unique and desirable fragrance. Among many different types of agarwood, *Aquilaria agollocha* (Kanakoh) is considered to be the highest quality.[141,142] Due to the increasing demand, excessive logging has caused agarwood to become endangered and it was listed in the 2006 IUCN Red List of Threatened Species.[140]

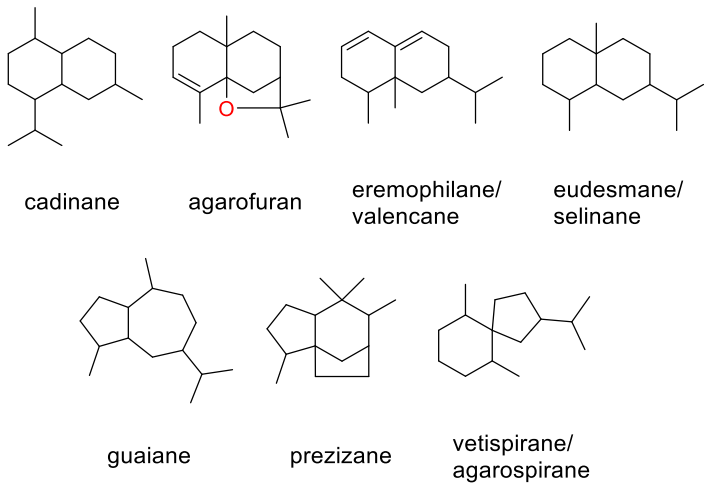


Figure 1.27 The seven main groups of sesquiterpenes present in agarwood organised based on their skeleton structures.

The constituents of agarwood essential oil vary, depending on the geographical origins and extraction modes, such as hydrodistillation, extraction with supercritical CO_2 or different types of organic

solvents such as benzene, acetone, petroleum ether, and ethyl acetate. It mainly contains many different sesquiterpenes and chromones. The sesquiterpenes present in agarwood oil are divided into seven mains groups based on their skeletons: cadinanes, agarofurans, eremophilanes and valencanes, eudesmanes and selinanes, guaianes, prezizanes and vetispiranes (Fig. 1.27).[139,140,143]

Dihydroagarofurans, α-agarofurans and β-agarofurans are among the first compounds that were isolated from agarwood originating from *Aquillaria agallocha* Roxb by Maheshwari *et al.* in 1963 (Fig. 1.28).[144] The compounds, which were named after agarwood, are likely to be the biosynthesis products of 10-epi-γ-eudesmol, a tertiary alcohol with the eudesmane core.[139] More compounds with the agarofuran skeleton were identified later. α-Agarofuran and β-agarofuran are reported to be two of the major components found in high-quality agarwood oil. Barrett *et al.* reported multi-step synthesis of α-agarofuran but the yield was low.[145]

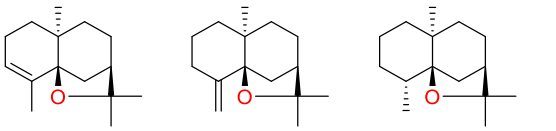


Figure 1.28 Compounds with agarofuran skeleton found in agarwood.

Nakanishi *et al.* first isolated Jinkohol and Jinkohol II in 1981 and 1983 respectively from an agarwood originating from *Aquilaria malaccensis Benth*, a low-quality agarwood (Fig. 1.29).[146,147] Both Jinkohol and Jinkohol II make up a significant amount of the agarwood oil (Jinkohol: 5.2%; Jinkohol II: 5.6%) and are considered as the trademark of this particular type of agarwood.[148] Jinkohol was described as a storm of the noblest woods while Jinkohol II brings the smell of the woody smoke of agarwood.[149] Nagashima *et al*. reported a 3-step chemical synthesis of Jinkohol II from prezizaene.[150]

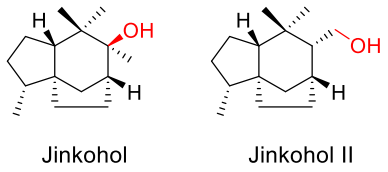


Figure 1.29 Compounds with the prezizane skeleton found in agarwood.[139]

(+)-Dihydrokaranone and (+)-karanone are important compounds in agarwood and are found in the majority of studied oils and extracts with the following exceptions: the absence of (+)-karanone from *Aquilaria agallocha* and *Aquilaria malaccensis*, and of dihydrokaranone from *Aquilaria malaccensis* (Fig. 1.30).[148] They are highly conjugated ketones with an eremophilane core with a warm scent of woodiness.[149] Eudesmane derivatives make up a significant portion of the acetone extract of Kanakoh (0.2–7 w/w%).[139] (+)-Selina-3,11-dien-9-ol and (–)-selina-3,11-dien-9-one were among the first discovered components in Kanakoh by Ishihara in 1990 (Fig. 1.30).[141] (–)-Selina-3,11-dien-9-one has very flowery and fresh odour like new blooms while the odour of (+)-selina-3,11-dien-9-ol is more woody and relatively weaker in comparison. The tertiary eudesmane alcohol derivative, 10-epi-γ-eudesmol, has not only been identified as the precursor of agarofuran but valencenes, an aroma component of citrus fruit (Fig. 1.30).[139]

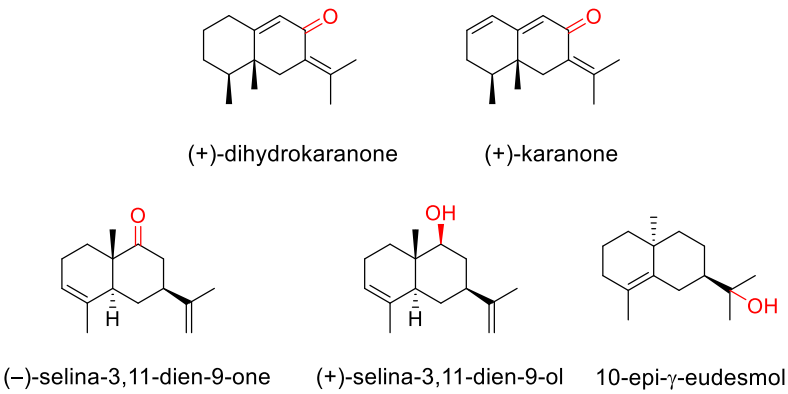


Figure 1.30 Eremophilane and eudesmane derivatives found in agarwood.[139]

In 1990, (–)-guaia-1(10),11-diene-15-al was first isolated together with two different compounds of selinane skeleton from Kanankoh, originating from *Aquilaria agallocha* Roxb, by Ishihara *et al.* (Fig. 1.31).[141] It was reported to be the most important component among those isolated from Kanankoh

as it has a beautiful scent of β-damascenone and is only found in Kanakoh.[149] Another type of Kanankoh was later studied by the same group and seven new guaiane-skeleton sesquiterpenes were identified. Guaiane sesquiterpenes were discovered as the primary components of the Kanakoh as they made up 5% of the ether extracts.[151] (–)-Guaia-1(10),11-dien-15,2-olide has been reported to be a major contributor to the elegant and unique fragrance of agarwood because of its long-lasting woody and ambery aroma with orris sub notes (Fig. 1.31).[149,151]

One of the guaiane skeleton molecules found in agarwood, rotundone, was identified by Wood *et al*. as an important compound contributing to the peppercorn aroma in many different kinds of herbs, especially pepper and also grapes (Fig. 1.31)[152].

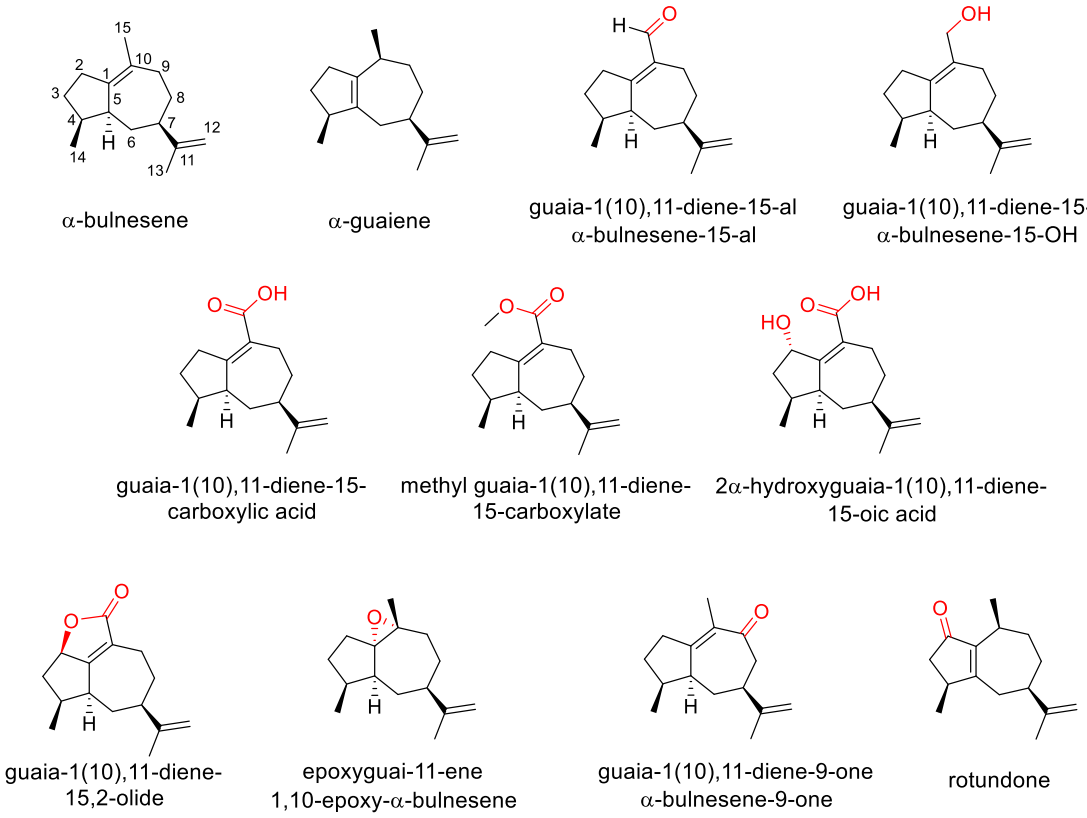


Figure 1.31 Compounds with guaiane skeleton found in agarwood.[139]

The presence of chromones plays an essential role in the rich, warm and long-lasting aroma of agarwood when it is burnt. These compounds can only be found in the extracted products from agarwood but not the distilled products.[139] It was reported by Michael Zviely in a review that "*the*

burning of all the unique sesquiterpenes discussed above together with chromones is what brings the aroma a sense of spiritual peace, the reason why it has been used for centuries as incense."[149] Around 60% of the acetone extract of Kanakoh is made up of chromone, much higher as compared to that of Jinkoh (low-quality agarwood) which contains around 1.5%. Most chromones can be categorised into two groups based on their skeleton structure, 2-(2-phenylethyl)-chromones and 5,6,7,8-tetrahydro-2-(2-phenylethyl)-chromones with different substituents (Fig. 1.32).[153] Besides sesquiterpene and chromones, various different simple volatile aromatic compounds which are most likely the degradation product of chromones and also a few different compounds such as fatty acid can also be found in agarwood.[139]

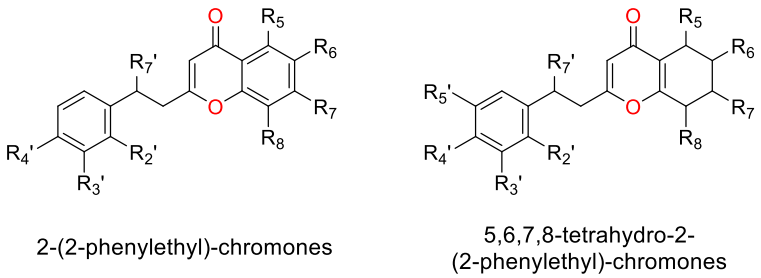


Figure 1.32 The basic skeleton structures of two main types of chromones with different substituents at R, either -Cl, -OH, $-OCH_3$ or $-OCOCH_3$.

1.9 Thesis objectives

Agarwood has been used for centuries as an incense material for religious worship and more recently, as a fragrance. Guaiane-derived compounds are significant contributors to the aroma of agarwood with (–)-guaia-1(10),11-diene-15-al (bulnesene-15-al) and (–)-guaia-1(10),11-dien-15,2-olide (15,2-lactone) being identified as the major contributors to the highly desirable woody aroma of agarwood. Rotundone, which attributes to the special peppercorn aroma in Australian Shiraz wine, is also found in agarwood essential oil.

Due to high demand, excessive logging has caused agarwood to become endangered and alternative approach of sourcing theses valuable compounds is needed. Examples of the synthesis of these valuable compounds are very limited especially for bulnesene-15-al and 15,2-lactone which usually use other guaiene-derived compounds from agarwood as precursors and thus, equally unsustainable. The presence of α-bulnesene and α-guaiene in patchouli oil as major components opens up new possibilities for the synthesis of bulnesene-15-al, 15,2-lactone and rotundone as bulnesene and guaiene have been proposed to be the precursors of the compounds.

Chapter 3 describes the development of the biocatalytic synthesis of the desirable fragrance molecule, bulnesene-15-al via the oxidation of α-bulnesene by engineered CYP102A1 variants. Besides the target compound, all oxidation products are of interest because of their potential flavour, fragrance, and biological properties. Further mutagenesis, guided by progressive library screening and molecular docking analysis were carried out to increase the activity and selectivity with the view to scale up toward possible pre-pilot process development.

In Chapter 4, a biocatalytic cascade reaction was designed to synthesise bulnesene-15,2-lactone. The system is more challenging and complicated as it requires a CYP102A1 variant to oxidise bulnesene to bulnesene-15-COOH, followed by another C–H bond oxidation to form (*2R*)-hydroxy-bulnesene-

15-COOH which cyclises spontaneously to form lactone. Besides the development of highly selective variants via protein engineering, reaction engineering was equally crucial to develop a system with high efficiency. The synthesis of bulnesene-15-al and 15,2-lactone by using bulnesene-enriched patchouli oil was also explored and discussed in this chapter.

Chapter 5 describes experiments to optimise the parameters such as substrate-to-enzyme ratio to develop biocatalytic routes to the valuable bulnesene-derived compounds with high regio- and chemoselectivity and TON. The biocatalytic synthesis of rotundone was explored in Chapter 6. α-Guaiene was screened against the libraries following the strategy established for bulnesene oxidation. The biocatalytic synthesis of the target compound was performed by utilising engineered CYP102A1 to functionalise guaiene at C2 position, followed by the oxidation of the C2-alcohol to rotundone by a CRED enzyme.

Chapter 2

2 Materials and Methods

2.1 Materials, chemicals and solvents

General reagents and chemicals were from Acros Organics, Alfa-Aesar, Fischer Scientific or Sigma-Aldrich. Patchouli oil was from DeMonchy Aromatics, UK.

HPLC-grade solvents were from Fischer Scientific or Sigma-Aldrich. Media components were from Melford Laboratories, Sigma-Aldrich and Fisher Scientific. Kanamycin and IPTG were supplied by Melford Laboratories, UK. Hen egg white lysozyme was from Sigma-Aldrich. $NADP^+$ disodium salt was from Prozomix, UK; glucose dehydrogenase (GDH) was from Codexis, California, USA. The carbonyl reductase, CRED-231 was a gift from Oxford Biotrans Ltd. Oligonucleotides were supplied by Eurofins Genetic Service, UK. Site-directed mutagenesis was carried out using the Q5 High-Fidelity DNA Polymerase from NEB.

2.2 Buffer, solution and growth media

Type 1 water ($\rho \geq 10$ MΩ.cm^{-1}) from a Millipore ELIX-5 system was used for the preparation of growth media and buffers for biotransformation and enzyme purification. Type 2 water from Millipore a Milli Q system ($\rho \geq 18$ MΩ.cm^{-1}) was used for PCR, general DNA manipulation, and for the preparation of SOC media, TFB1 and TFB2 buffers. The components of each media or solution are shown in Table 2.1.

Liquid media for growth were sterilised by autoclaving at 120 °C for 20 min. Temperature-sensitive media components and induction solution were filter-sterilised by using a 0.22 μM filter unit (Merck) before use. Kanamycin was added after the LB media for both agar plate and growth media had cooled to approximately 40 °C.

Table 2.1 Media and buffer components

Media and Buffer	Components (per litre unless stated otherwise)
LB for agar plate	10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g bacto-agar, 30 mg kanamycin
LB	10 g tryptone, 5 g yeast extract, 10 g NaCl, 30 mg kanamycin
$FB_{Glycerol}$	14.7 g K_2HPO_4, 3.6 g NaH_2PO_4, 2.5 g $(NH_4)_2SO_4$, 2.0 g Na_2SO_4, 1.2 g Na_3citrate.$2H_2O$, 0.5 g NH_4Cl, 5 mL glycerol
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 $CuSO_4 \cdot 5H_2O$
Inoculation solution	15 mg kanamycin, 250 mg $MgSO_4 \cdot 7H_2O$, 50 mg thiamine·HCl, add trace element solution to 1.5 mL
Induction solution/40 mL	360 mg IPTG, 672 mg δ-ALA·HCl, 112 mg $FeSO_4 \cdot 7H_2O$, 600 mg kanamycin in 40 ml H_2O
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.
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 (pH 6.5), 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol (pH adjusted to 6.5 with 1 M KOH)
200 mM KPi (Buffer P)	2.4 g KH_2PO_4, 31.7 g K_2HPO_4

2.3 Equipment and analysis method

PCR reactions were carried out in a Techne thermocycler or a Cetus thermocycler from Perkin-Elmer. Sonication was carried out using a Misonix 3000 sonicator with a Fisher Scientific FB-505 ultrasonic processor. Beckman Coulter Allegra X-22R refrigerated bench top centrifuges were used for enzyme purification using a combination of 4 × 250 mL swing-out and 6 × 50 mL and 8 × 40 mL fixed angle rotors. Thermo Scientific Heraeus Pico microcentrifuges were used for smaller scale centrifugation in microcentrifuge tubes. Plasmids were purified using the GeneJet kit from Fisher Scientific. UV-

vis spectra were acquired on a Varian CARY50 spectrophotometer using 1 cm pathlength quartz cuvettes.

Gas chromatographic 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 MS fused silica column (30 m × 0.25 mm diameter, 0.25 μm film thickness, Agilent Technologies). The oven temperature was held at 120 °C for 1 min, raised at 15 °C min^{-1} to 240 °C, and kept at this temperature for 4 min. The retention times were, bulnesene, **1**: 5.93 min; bulnesene-15-ol, **1a**: 7.82 min; bulnesene-15-al, **1b**: 8.09 min; (2*S*)-hydroxy-bulnesene, **1c**: 7.17 min; (2*R*)-hydroxy-bulnesene, **1d**: 7.19 min; 15-hydroxy-11,12-epoxy-bulnesene, **1e**: 9.02 min; 9-hydroxy-bulnesene, **1f**: 7.55 min; bulnesene-15-COOH, **1g**: 8.58 min; 11,12-epoxy-bulnesene-15-COOH, **1h**: 9.71 min; 11,12-epoxy-bulnesene-15-COOH, **1h(a)**: 9.68 min; bulnesene-15,2-lactone, **1j**: 9.51 min; guaiene, **2**: 5.92 min; 11,12-epoxy-guaiene, **2a**: 7.22 min; (*2S*)-rotundol, **2b**: 7.27 min; (*9S*)-hydroxy-guaiene, **2c**: 7.41 min; 2-hydroxy-11,12-epoxy-guaiene, **2d**: 8.52 min; (*2R,9S*)-dihydroxy-guaiene, **2e**: 8.81 min; rotundone, **2f**: 7.84 min.

Silver-nitrate-impregnated silica (Geduran Silica 60, 40–63 μm, Fisher Scientific) was used for flash column chromatography for the isolation of bulnesene and guaiene while the oxidation products were purified by flash chromatography with unmodified silica, in both cases eluting with mixtures of petroleum ether (b.p. 40–60 °C) and ethyl acetate. Analytical thin-layer chromatography was performed, and bands were visualised with permanganate or phosphomolybdic acid staining agents. ^{1}H, ^{13}C, COSY, HSQC, HMBC, NOESY and 1D nOe NMR spectra were collected on Bruker AVIII-500 (500/125 MHz), Bruker AVIII-400 38 (400/100 MHz), or Bruker AV-400 (400/100 MHz) spectrometers. High-resolution mass spectra (HRMS) were obtained on a Bruker Daltonics MicroTOF spectrometer; mass-to-charge ratios (*m/z*) are reported in Daltons.

2.4 DNA manipulation

2.4.1 Mutagenesis primer design

To introduce mutations into the template DNA, two staggered complementary oligonucleotides were designed by using SnapGene Viewer. Each oligonucleotide had 30 to 50 bases including at least 10 bases to the 5′ side of the mutated nucleotides and 15 bases from its 3′ side. The melting temperature (T_m) of the primer pairs were kept between 55 °C to 65 °C with <1 °C difference. Oligonucleotides with double or triple mutations were designed with a similar strategy. A list of the oligonucleotide sequences is included in the Appendix 5.

2.4.2 PCR site-directed mutagenesis

The template DNA (isolated from miniprep plasmid preparation experiment) and primers used in mutagenesis by PCR were 10× diluted with Milli-Q water before being combined in a PCR microcentrifuge tube. The amount needed for each component for a 25 µL PCR run is listed in Table 2.2. After the initial denaturation at 95 °C for 2 min, the following thermal cycle was run: denaturation at 95 °C for 20 s, annealing at 55 °C to 65 °C for 10 s, extension at 70 °C for 3 min 20 s. The cycle was repeated 30 times before the final extension step at 70 °C for 5 mins. Once the cycle was complete, the reaction mixture was cooled to 4 °C before the next step.

2.4.3 Agarose gel electrophoresis

A mixture of agarose powder (0.5 g) and 1× TAE (Tris-Acetate-EDTA) buffer (50 mL) in a 100 mL conical flask was heated in microwave for 40 s until the suspension became clear. RedSafe™ staining solution (5 µL) was added and the solution was swirled until the solution was homogenous. The solution was poured into a gel tray with a well comb in place and left to set. Once solidified, the agarose gel was placed in the electrophoresis apparatus and 1× TAE buffer was added until the gel was fully submerged in the buffer. 10 µL of the PCR mixture and 2 µL of loading dye were mixed

thoroughly before being pipetted carefully into wells in the gel. 4 µL of a 1 kbp DNA reference ladder (Promega) was also loaded into one of the wells. A voltage difference of 100 V was applied across the gel for 20 min and a long wavelength UV lamp was used to visualise the nucleic acid bands. A band with the same mobility as a linearised 8 500 bp fragment indicated that the PCR step was successful. DpnI (0.5 µL) was added to the remaining PCR mixture to digest the template DNA at 4 °C for 15 min.

Table 2.2 Components of a standard PCR mixture for site-directed mutagenesis

Component	Volume in a 25 µL reaction
5× Reaction buffer	5 µL
10 mM dNTPs	0.5 µL
10 µM forward and reverse primer	1.25 µL
Template DNA stock	0.5 µL
Q5 High-fidelity DNA Polymerase	0.25 µL
5× Q5 High GC Enhancer (optional)	5 µL
Milli-Q water	12.5 µL (up to 25 µL)

2.4.4 Post-mutagenesis transformation

A 10 µL aliquot of the digested PCR product was mixed with 40 µL *E. coli* DH5α or Acella competent cells freshly thawed from storage at −80 °C in a 1.5 mL microcentrifuge tube. The mixture was incubated on ice for 30 min. The cells were subjected to heat-shock at 42 °C for 45 s and then incubated on ice for 5 min. Filter-sterilised SOC media (800 µL) was added to the mixture and the culture was shaken at 250 rpm and 37 °C for 2 h. The cells were spread on a LB agar plate with kanamycin and placed in 37 °C incubator oven overnight.

2.4.5 DNA purification and transformation

A single colony was picked and inoculated into 5 mL of LB_{KAN} media. The cell culture was incubated at 120 rpm and 37 °C for 16 h. The culture was centrifuged at 9 000 *g* at 4 °C for 10 min and the

supernatant was removed. Plasmid DNA was then purified using a GeneJET plasmid miniprep kit from ThermoFisher Scientific. The purified plasmid DNA was recovered with 40 µL elution buffer. 15 µL of the DNA was sent for DNA sequencing by Source BioScience using two primers, T7F and BM3FP2 (Table 2.3) to confirm the outcome of mutagenesis. 1 µL of the plasmid DNA was transformed into 40 µL *E. coli* Acella or BL21 (DE3) competent cells for recombinant protein production with the same transformation protocol as described in section 2.4.4. The plasmid DNA solution was stored at –80 °C.

Table 2.3 Nucleotide sequences of the T7F and BM3FP2 primers.

Primer	Sequence
T7F	TAATACGACTCACTATAGG
BM3FP2	CTGCAGCGAGCAAATCCAGACGAC

2.4.6 Enzyme production and purification

A single colony of transformed *E. coli* BL21 (DE3) or Acella from a LB_{KAN} plate was picked and inoculated into 5 mL of LB_{KAN}. The culture was incubated at 37 °C and 160 rpm. After 6 h, 200 µL of the pre-culture and 1.5 mL of inoculation solution was added into a 2 L conical flask with 500 mL of $FB_{glycerol}$. The culture was incubated at 37 °C and 120 rpm for 16 h. The temperature of the incubator was decreased to 20 °C and the culture was left to shake for 1 h before adding 1 mL of induction solution and 2 mL of sterilised tryptone stock (0.25 g/mL). The induction steps were repeated every 24 h.

Cells were harvested after 72 h of growth by centrifuging at 5 250 *g* for 20 min. After discarding the supernatant, the cells were resuspended with 30 mL of buffer P per 500 mL culture with 10 mg of lysozyme added. The resuspended cells were chilled on ice for 10 min and then lysed on ice with a 5-minute cycle of 15-second pulses and 30-second cooling period, at 45% amplitude. The lysate was transferred to a 50 mL centrifuge tube and centrifuged for 20 min at 9 500 *g*, 4 °C to remove the cell

debris. The supernatant was subjected to ammonium sulfate fractionation. The amount of ammonium sulfate required for the supernatant to achieve 40% saturation (ammonium sulfate calculator http://encorbio.com/protocols/AM-SO4.htm) was added and the solution was shaken gently until all the salt had fully dissolved. The solution was centrifuged at 9 500 *g* for 20 min. The supernatant was transferred to a new 50 mL centrifuge tube and more ammonium sulfate was added to achieve 60% saturation. Once the salt had fully dissolved, the mixture was centrifuged at 9 500 *g* for 20 min. The supernatant was discarded, and 5 mL of buffer P was added to dissolve the enzyme pellet.

2.4.7 Enzyme quantification

The enzyme concentration was determined by the CO-difference assay. A 50 µL aliquot of the enzyme was diluted with 950 µL of buffer P and a few crystals of sodium hydrosulphite was added to reduce the enzyme. The UV-vis spectrum (200–500 nm) of the sample was recorded as baseline. Carbon monoxide (~20 bubbles) was slowly bubbled through the sample and the difference spectrum recorded. The characteristic peak of the $Fe^{II}(CO)$ form of P450 enzymes at 450 nm would be observed and the absorbance was used to calculate the enzyme concentration with the following equation:

$$[P450]\ \mu M = \frac{20\ (A_{450} - A_{490})}{91000} \times 10^6$$

where $\varepsilon_{450–490}$ = 91 000 M^{-1} cm^{-1} is the extinction coefficient.[154]

2.5 *In vitro* screening, preparative scale reactions and product isolation

2.5.1 *In vitro* screening

Bulnesene **1** and guaiene **2** were dissolved in EtOH to make 400 mM stock solutions. Screening reaction mixtures in 200 mM phosphate buffer (pH 7.9) contained 2 mM substrate and 2 µM P450 enzyme in a total volume of 0.5 mL in 24-well plates. Glucose (100 mM) and glucose dehydrogenase

(4 U mL^{-1}) were used to regenerate the NADPH cofactor from $NADP^+$ (40 μM). Methyl-β-cyclodextrin was added to 10 mM to increase the aqueous solubility of the substrates. Screening plates were shaken at 120 rpm for 6 h at 20 °C. The organics were extracted with 300 μL ethyl acetate. After centrifugation at 14 300 *g* to separate the phases, the organic extracts were analysed by GC and the conversion, product selectivity and TON of each variant were calculated based on the results. Conversion is the percentage of substrate being converted; product selectivity is the percentage of a certain product over all the products formed while TON is the total yield of the product(s) relative to enzyme.

Table 2.4 Reaction components for screening of α-bulnesene **1** and α-guaiene **2**.

Reagent	**Stock concentration**	**Final concentration**
Substrate	400 mM	2 mM
Glucose	1 M	100 mM
GDH	200 U/mL	4 U/mL
$NADP^+$	1 M	40 μM
Enzyme	10 μM	2 μM
Methyl-β-cyclodextrin	200 mM	10 mM
200 mM buffer P	200 mM	N/A

Bulnesene-15-al **1b** and bulnesene-15-COOH **1g** were dissolved in EtOH to make 200 mM stock solutions. Screening reaction mixtures in 200 mM phosphate buffer (pH 7.9) contained 1 mM substrate and 2 μM P450 enzyme in a total volume of 0.5 mL in 24-well plates. Glucose (100 mM) and glucose dehydrogenase (4 U mL^{-1}) were used to regenerate the cofactor NADPH from $NADP^+$ (40 μM). Methyl-β-cyclodextrin was added to 10 mM to increase the aqueous solubility of the substrates. Screening plates were shaken at 120 rpm for 16 h at 20 °C. The organics were extracted with 300 μL ethyl acetate. After centrifugation at 14 300 *g* to separate the phases, the organic extracts were analysed by GC.

Table 2.5 Reaction components for screening of bulnesene-15-al **1b** and bulnesene-15-COOH **1g**.

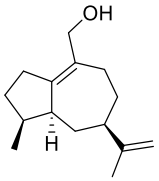
Reagent	Stock concentration	Final concentration
Substrate	200 mM	1 mM
Glucose	1 M	100 mM
GDH	200 U/mL	4 U/mL
$NADP^+$	1 M	40 µM
Enzyme	10 µM	2 µM
Methyl-β-cyclodextrin	200 mM	10 mM
200 mM buffer P	200 mM	N/A

2.5.2 Preparative scale reactions and product purification

Preparative scale reactions were carried out with the same conditions and concentration for each reagent as described in Section 2.5.1 for the initial screening but in larger volumes. For example, 100 mL scale reactions were carried out in 500 mL conical flasks and 200 mL scale reactions in 1 litre flasks, *etc*. The flasks were shaken in an orbital shaker at 20 °C and 120 rpm. Aliquots (400 µL) were removed periodically to monitor the progress of the reaction by GC. Once the reaction was complete, the reaction mixture was extracted three times with an equal volume of ethyl acetate. The combined extracts were washed with water and then brine before drying over Na_2SO_4. The solvent was removed *in vacuo*.

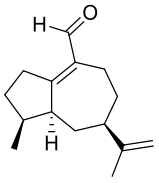
Thin layer chromatography (TLC) was used to identify a suitable solvent system for product separation and flash column chromatography purification was carried out. Each fraction was analysed by TLC and selected fractions were further analysed with GC. Fractions with the same compound were combined and solvents were removed *in vacuo*. Isolated yields were calculated based on the mass of bulnesene converted.

15-Hydroxy-bulnesene (bulnesene-15-ol, 1a)



A preparative scale reaction (200 mL) was carried out with the R19/F87I/T88F variant (2 μM, 0.4 μmol) and bulnesene (2 mM of 90% pure bulnesene, actual concentration = 1.8 mM, 73.4 mg, 0.36 mmol). After 7 h, GC analysis showed that the reaction had reached 87% bulnesene conversion with 82% total selectivity for C15 oxidation and 43% chemoselectivity for bulnesene-15-ol (**1a**). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (7:1 petroleum ether/ethyl acetate) to give bulnesene-15-ol (**1a**) as a yellow oil (12.2 mg, 18% yield). R_f 0.17 (7:1 petroleum ether/ethyl acetate). **^{1}H NMR** (500 MHz, $CDCl_3$) δ 4.70 – 4.60 (m, 2H), 4.16 – 4.03 (m, 2H), 2.56 – 2.41 (m, 2H), 2.38 – 2.08 (m, 5H), 1.79 (dddd, *J* = 13.3, 7.2, 3.2, 1.5 Hz, 1H), 1.73 – 1.61 (m, 5H), 1.43 – 1.32 (m, 1H), 1.30 – 1.18 (m, 1H), 1.04 (dt, *J* = 13.0, 11.8 Hz, 1H), 0.92 (d, *J* = 7.0 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 152.13, 147.44, 132.26, 108.20, 65.67, 51.06, 46.39, 38.22, 33.05, 32.50, 32.18, 30.43, 29.50, 20.80, 15.30. The NMR spectroscopic data are consistent with those reported in the literature.[151]

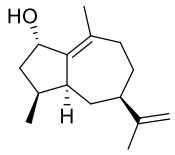
Guaia-1(10),11-diene-15-al (Bulnesene-15-al, 1b)



A preparative scale reaction (200 mL) was carried out with the R19/F87I/T88F variant (2 μM, 0.4 μmol) and bulnesene (2 mM of 90% pure bulnesene, actual concentration = 1.8 mM, 73.4 mg, 0.36 mmol). After 7 h, GC analysis showed that the reaction had reached 87% bulnesene conversion with 82% total selectivity for C15 oxidation and 39% chemoselectivity for bulnesene-15-al (**1b**). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel

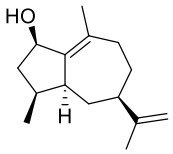
column chromatography (9:1 petroleum ether/ethyl acetate) to give bulnesene-15-al (**1b**) as a yellow oil (9.8 mg, 14% yield). R_f 0.44 (7:1 petroleum ether/ethyl acetate). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 9.92 (s, 1H), 4.71 – 4.61 (m, 2H), 3.09 – 2.94 (m, 2H), 2.86 – 2.71 (m, 2H), 2.20 (tdd, *J* = 11.3, 6.3, 3.2 Hz, 2H), 1.94 – 1.72 (m, 4H), 1.69 (s,3H), 1.59 – 1.45 (m, 1H), 1.21 – 1.03 (m, 2H), 0.98 (d, *J* = 7.0 Hz, 3H). **^{13}C NMR** (400 MHz, $CDCl_3$) δ 191.49, 173.79, 151.54, 136.00, 108.87, 51.03, 49.28, 37.03, 33.14, 31.59, 31.38, 29.08, 23.59, 20.85, 15.13. The NMR spectroscopic data are consistent with those reported in the literature.[141] **HRMS** (ESI+): Calc'd for $C_{15}H_{23}O^+$ $[M+H]^+$: 219.1743, found: 219.1744.

(2*S*)-Hydroxy-α-bulnesene (1c)



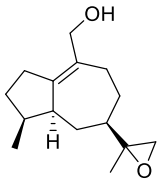
A preparative scale reaction (100 mL) was carried out with the R19/F87A/A264G/A328I variant (2 μM, 0.2 μmol) and bulnesene (2 mM of 75% pure bulnesene, actual concentration = 1.5 mM, 30.6 mg, 0.15 mmol). After 23 h, GC analysis showed that the reaction had reached 88% bulnesene conversion with 12% selectivity of (2*S*)-hydroxy-α-bulnesene (**1c**). The reaction mixture was extracted with ethyl acetate and the products were separated by $AgNO_3$-SiO_2 silica gel column chromatography (100% petroleum ether) to give (2*S*)-hydroxy-α-bulnesene (**1c**) as a colourless oil (2.9 mg, 10% yield). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.89 – 4.46 (m, 3H), 2.60 (s, 1H), 2.52 – 2.38 (m, 1H), 2.28 (q, *J* = 13.0 Hz, 1H), 2.12 (tt, *J* = 11.6, 3.3 Hz, 1H), 2.05 – 1.95 (m, 1H), 1.86 (d, *J* = 1.7 Hz, 3H), 1.81 – 1.64 (m, 5H), 1.63 – 1.42 (m, 2H), 1.29 – 1.14 (m, 1H), 0.95 (d, *J* = 7.0 Hz, 3H), 0.87 (d, *J* = 12.3 Hz, 1H). **^{13}C NMR** (400 MHz, $CDCl_3$) δ 152.24, 146.11, 136.77, 108.46, 72.41, 51.52, 44.96, 42.96, 35.18, 35.02, 31.46, 31.20, 21.78, 20.89, 15.52. The NMR spectroscopic data are consistent with those reported in the literature.[141]

(2*R*)-Hydroxy-α-bulnesene (1d)



A preparative scale reaction (100 mL) was carried out with the R19/F87A/V78I/A328I variant (2 μM, 0.2 μmol) and bulnesene (2 mM of 75% pure bulnesene, actual concentration = 1.5 mM, 30.6 mg, 0.15 mmol). After 24 h, GC analysis showed that the reaction had reached 97% bulnesene conversion with 29% selectivity of (2*R*)-hydroxy-α-bulnesene (**1d**). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (15:1 petroleum ether/ethyl acetate) to give (2*R*)-hydroxy-α-bulnesene (**1d**) as a yellow oil (7.2 mg, 23% yield). R_f 0.23 (7:1 petroleum ether/ethyl acetate).**^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.70 – 4.60 (m, 3H), 2.49 (d, *J* = 1.8 Hz, 1H), 2.24 (s, 1H), 2.19 – 1.95 (m, 4H), 1.86 (dt, *J* = 1.9, 0.8 Hz, 3H), 1.80 – 1.64 (m, 5H), 1.60 – 1.45 (m, 1H), 1.36 – 1.22 (m, 1H), 1.22 – 1.09 (m, 1H), 1.03 (d, *J* = 7.2 Hz, 3H). **^{13}C NMR** (400 MHz, $CDCl_3$) δ 152.24, 144.99, 136.31, 108.39, 74.20, 51.04, 46.42, 43.23, 36.10, 35.61, 33.82, 31.95, 22.15, 20.96, 17.01. The NMR spectroscopic data are consistent with those reported in the literature.[155]

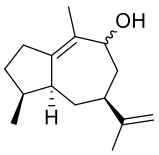
15-Hydroxy-11,12-epoxy-α-bulnesene (1e)



A preparative scale reaction (100 mL) was carried out with the GLQ/I263G/A328L variant (2 μM, 0.2 μmol) and bulnesene (2 mM of 75% pure bulnesene, actual concentration = 1.5 mM, 30.6 mg, 0.15 mmol). After 18 h, GC analysis showed that the reaction had reached 98% bulnesene conversion with 41% selectivity of 15-hydroxy-11,12-epoxy-α-bulnesene (**1e**). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography

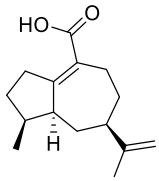
(3:1 petroleum ether/ethyl acetate) to give 15-hydroxy-11,12-epoxy- α-bulnesene (**1e**) as a yellow oil (7.3 mg, 21% yield). R_f 0.22 (3:1 petroleum ether/ethyl acetate). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.28 – 3.92 (m, 2H), 2.63 – 2.57 (m, 2H), 2.51 – 2.41 (m, 2H), 2.41 – 2.32 (m, 1H), 2.31 – 2.23 (m, 1H), 2.19 – 2.07 (m, 2H), 1.83 (s, 1H), 1.78 (dddt, *J* = 10.8, 6.9, 2.5, 1.4 Hz, 1H), 1.69 (ddt, *J* = 11.8, 8.0, 5.7 Hz, 1H), 1.38 (dd, *J* = 12.1, 7.7 Hz, 1H), 1.28 – 1.22 (m, 1H), 1.19 (s, 3H), 1.17 – 1.09 (m, 1H), 1.02 (dd, *J* = 11.7, 0.9 Hz, 1H), 0.95 (d, *J* = 7.0 Hz, 3H). **^{13}C NMR** (400 MHz, $CDCl_3$) δ 147.90, 132.17, 65.81, 60.98, 55.13, 50.83, 46.25, 38.33, 33.23, 30.57 (2C), 29.76, 29.04, 17.42, 15.51. **HRMS** (ESI+): Calc'd for $C_{15}H_{24}O_2Na^+$ $[M+Na]^+$: 259.1669, found: 259.1671.

9-Hydroxy-α-bulnesene (1f)



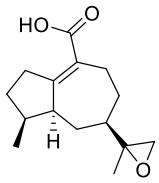
A preparative scale reaction (100 mL) was carried out with the R19/F87A/A264G/A328G variant (2 μM, 0.2 μmol) and bulnesene (2 mM of 75% pure bulnesene, actual concentration = 1.5 mM, 30.6 mg, 0.15 mmol). After 24 h, GC analysis showed that the reaction had reached 100% bulnesene conversion with 37% selectivity of 9-hydroxy-α-bulnesene (**1f**). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (7:1 petroleum ether/ethyl acetate) to give 9-hydroxy-α-bulnesene (**1f**) as a yellow oil (8.9 mg, 27% yield). R_f 0.14 (7:1 petroleum ether/ethyl acetate). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.67 (dq, *J* = 5.5, 1.4 Hz, 2H), 4.46 (d, *J* = 10.3 Hz, 1H), 2.39 (dddt, *J* = 12.6, 6.2, 2.9, 1.5 Hz, 2H), 2.31 (s, 1H), 2.24 – 2.14 (m, 1H), 2.14 – 2.06 (m, 1H), 1.78 – 1.64 (m, 8H), 1.62 – 1.44 (m, 2H), 1.36 – 1.28 (m, 1H), 0.94 (d, *J* = 6.9 Hz, 3H), 0.91 – 0.85 (m, 1H). **^{13}C NMR** (400 MHz, $CDCl_3$) δ 151.41, 140.85, 131.72, 108.95, 73.25, 48.57, 45.45, 41.69, 38.87, 32.64, 30.80, 30.69, 20.84, 15.66, 14.63.

Bulnesene-15-COOH (1g)



A preparative scale reaction (150 mL) was carried out with the R19/F87I/A184I/V26L variant (2 μM, 0.3 μmol) and bulnesene (2 mM of 90% pure bulnesene, actual concentration = 1.8 mM, 55.1 mg, 0.27 mmol). After 16 h, GC analysis showed that the reaction had reached 97% bulnesene conversion with 70% selectivity of bulnesene-15-COOH (**1g**). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (10:1 petroleum ether/ethyl acetate) to give bulnesene-15-COOH (**1g**) as a yellow oil (13.5 mg, 22% yield). R_f 0.31 (5:1 petroleum ether/ethyl acetate). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.66 (d, *J* = 1.3 Hz, 2H), 3.02 – 2.85 (m, 2H), 2.81 – 2.58 (m, 2H), 2.18 (dddd, *J* = 23.1, 13.8, 7.1, 4.6 Hz, 3H), 1.85 – 1.71 (m, 3H), 1.69 (d, *J* = 1.2 Hz, 3H), 1.40 (dtd, *J* = 12.1, 9.6, 7.7 Hz, 1H), 1.22 (dtd, *J* = 13.2, 11.7, 1.4 Hz, 1H), 1.12 – 1.01 (m, 1H), 0.98 (d, *J* = 7.0 Hz, 3H) **^{13}C NMR** (400 MHz, $CDCl_3$) δ 174.39, 170.31, 151.63, 125.85, 108.84, 51.43, 49.32, 37.87, 34.38, 33.30, 31.65, 30.15, 28.19, 20.77, 15.25. NMR spectroscopic data are consistent with those reported in the literature.[151] **HRMS** (ESI+): Calc'd for $C_{15}H_{22}O_2Na^+$ $[M+Na]^+$: 257.1512, found: 257.1513.

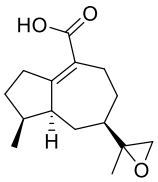
11,12-Epoxy-bulnesene-15-COOH (1h)



A preparative scale reaction (100 mL) was carried out with the R19/F87I/A82M/A184I/A330V variant (2 μM, 0.2 μmol) and bulnesene (2 mM of 90% pure bulnesene, actual concentration = 1.8 mM, 36.7 mg, 0.18 mmol). After 16 h, GC analysis showed that the reaction had reached 92% bulnesene conversion with 56% selectivity of 11,12-epoxy-bulnesene-15-COOH (**1h**). The reaction

mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (1:1 petroleum ether/ethyl acetate) to give 11,12-epoxy-bulnesene-15-COOH (**1h**) as a yellow oil (11.2 mg, 27% yield). R_f 0.25 (1:1 petroleum ether/ethyl acetate). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 3.01 – 2.87 (m, 2H), 2.68 (ddd, *J* = 11.6, 9.5, 7.5 Hz, 2H), 2.63 – 2.57 (m, 2H), 2.14 (dd, *J* = 14.5, 10.7 Hz, 2H), 1.98 (dt, *J* = 13.1, 5.0 Hz, 1H), 1.84 – 1.67 (m, 2H), 1.48 – 1.31 (m, 2H), 1.27 – 1.14 (m, 4H), 1.05 – 0.87 (m, 4H). **^{13}C NMR** (400 MHz, $CDCl_3$) δ 173.04, 169.61, 125.92, 60.61, 54.66, 50.67, 48.99, 37.79, 34.35, 33.26, 28.52, 27.98, 27.79, 17.67, 15.28. **HRMS** (ESI+): Calc'd for $C_{15}H_{23}O_3^+$ $[M+H]^+$: 251.1642, found: 251.1642.

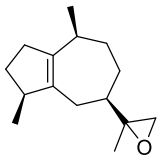
11,12-Epoxy-bulnesene-15-COOH [1h(a)]



A preparative scale reaction (50 mL) was carried out with the R19/F87I/A330F variant (2 μM, 0.1 μmol) and bulnesene (2 mM of 90% pure bulnesene, actual concentration = 1.8 mM, 18.4 mg, 0.09 mmol). After 16 h, GC analysis showed that the reaction had reached 82% bulnesene conversion with only 4% selectivity of **1h(a)** formed. Another aliquot of enzyme was added. After a total of 32 h, the reaction was completed with 100% conversion of bulnesene and 40% selectivity of **1h(a)** was formed. The total enzyme concentration of the reaction is 4 μM (0.2 μmol). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (1:1 petroleum ether/ethyl acetate) to give another diastereomer of 11,12-epoxy-bulnesene-15-COOH (**1h(a)**) as a yellow oil (4.2 mg, 19% yield). R_f 0.28 (1:1 petroleum ether/ethyl acetate). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 2.94 (dddd, *J* = 14.9, 7.2, 5.3, 1.6 Hz, 2H), 2.68 (dddd, *J* = 13.8, 9.0, 4.3, 2.2 Hz, 2H), 2.61 (d, *J* = 1.0 Hz, 2H), 2.15 (ddd, *J* = 14.5, 9.0, 5.2 Hz, 2H), 1.91 (dq, *J* = 11.7, 1.7 Hz, 1H), 1.78 (td, *J* = 12.0, 6.5 Hz, 2H), 1.48 – 1.23 (m, 2H), 1.19 (d, *J* = 6.1 Hz, 4H), 1.09 – 0.95 (m, 4H). **^{13}C NMR** (400 MHz, $CDCl_3$) δ 173.30, 170.45, 125.49, 60.62, 55.06, 50.97,

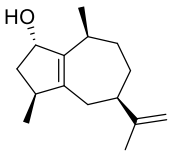
48.90, 37.86, 34.43, 33.31, 29.51, 28.13, 26.90, 17.27, 15.29. **HRMS** (ESI+): Calc'd for $C_{15}H_{23}O_3^+$ $[M+H]^+$: 251.1642, found: 251.1642.

11,12-Epoxy-α-guaiene (2a)



A preparative scale reaction (100 mL) was carried out with the GLQ/I263G/A328L variant (2 μM, 0.2 μmol) and guaiene (2 mM, 40.8 mg, 0.2 mmol). With an extra aliquot of enzyme added at 16 h, GC analysis showed that the reaction had reached 97% conversion with 44% selectivity for 11,12-epoxy-α-guaiene (**2a**) after 34 h. The total enzyme concentration of the reaction is 4 μM (0.4 μmol). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (7:1 petroleum ether/ethyl acetate) to give 11,12-epoxy-α-guaiene (**2a**) as a colourless oil (4.2 mg, 10% yield). R_f 0.25 (7:1 petroleum ether/ethyl acetate). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 2.59 (d, *J* = 0.8 Hz, 3H), 2.48 – 2.36 (m, 1H), 2.32 (s, 1H), 2.21 – 2.04 (m, 3H), 2.02 – 1.89 (m, 1H), 1.76 – 1.58 (m, 4H), 1.35 – 1.15 (m, 5H), 1.00 (d, *J* = 7.2 Hz, 3H), 0.95 (d, *J* = 6.9 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 140.69, 138.15, 60.87, 55.04, 46.41, 46.05, 36.44, 33.81, 33.53, 31.20, 29.82, 28.44, 19.85, 18.59, 17.13.

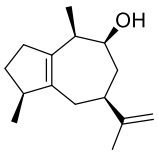
(*2S*)-Hydroxy-α-guaiene [(*2S*)-rotundol, 2b]



A preparative scale reaction (200 mL) was carried out with the RT2/S72G/A330W/L437LA variant (2 μM, 0.4 μmol) and guaiene (2 mM, 81.6 mg, 0.4 mmol). After 5 h, GC analysis showed that the reaction had reached 89% guaiene conversion with 57% selectivity of rotundol (**2b**). The reaction

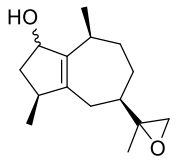
mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (7:1 petroleum ether/ethyl acetate) to give rotundol (**2b**) as a yellow oil (8.8 mg, 10% yield). R_f 0.28 (7:1 petroleum ether/ethyl acetate). **^{1}H NMR** (500 MHz, $CDCl_3$) δ 4.72 – 4.57 (m, 3H), 2.68 (dd, *J* = 9.5, 3.7 Hz, 1H), 2.62 (d, *J* = 1.7 Hz, 1H), 2.20 (ddq, *J* = 15.1, 11.9, 1.5 Hz, 1H), 2.16 – 2.08 (m, 1H), 2.05 – 1.97 (m, 1H), 1.88 (ddd, *J* = 13.8, 7.8, 3.8 Hz, 2H), 1.81 – 1.69 (m, 6H), 1.62 (dq, *J* = 11.2, 3.9 Hz, 1H), 1.04 (d, *J* = 7.2 Hz, 3H), 0.95 (d, *J* = 7.0 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 152.12, 144.13, 143.40, 108.38, 79.28, 46.66, 43.14, 42.39, 34.34, 33.85, 31.10, 30.47, 20.51, 20.30, 18.11. The NMR spectroscopic data are consistent with those reported in the literature.[156,157]

(*9S*)-Hydroxy-α-guaiene (2c)



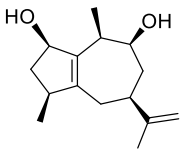
A preparative scale reaction (100 mL) was carried out with the GVQ/A328G variant (2 μM, 0.2 μmol) and guaiene (2 mM, 40.8 mg, 0.2 mmol). After 9 h, GC analysis showed that the reaction had reached 96% guaiene conversion with 5% selectivity of (*9S*)-hydroxy-α-guaiene (**2c**). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (7:1 petroleum ether/ethyl acetate) to give (*9S*)-hydroxy-α-guaiene (**2c**) as a colourless oil (1.5 mg, 4% yield). R_f 0.25 (7:1 petroleum ether/ethyl acetate). **^{1}H NMR** (500 MHz, $CDCl_3$) δ 4.73 – 4.64 (m, 2H), 3.87 – 3.78 (m, 1H), 2.59 (d, *J* = 6.9 Hz, 1H), 2.46 (ddd, *J* = 9.5, 6.4, 2.9 Hz, 2H), 2.31 – 1.86 (m, 6H), 1.82 – 1.68 (m, 4H), 1.31 (dddd, *J* = 9.5, 8.5, 4.8, 2.2 Hz, 1H), 1.06 (dd, *J* = 7.1, 1.0 Hz, 3H), 0.93 (dd, *J* = 6.9, 1.0 Hz, 3H). **^{13}C NMR** (500 MHz, $CDCl_3$) δ 151.30, 140.00, 136.61, 108.83, 72.85, 46.53, 44.96, 40.52, 39.83, 37.13, 33.03, 31.09, 20.51, 19.79, 11.44. The configuration of C9 was assigned as (*S*) by the nOe data.

2-Hydroxy-11,12-epoxy-α-guaiene (2d)



A preparative scale reaction (100 mL) was carried out with the GLQ/I263G/A328L variant (2 μM, 0.2 μmol) and guaiene (2 mM, 40.8 mg, 0.2 mmol). With an extra aliquot of enzyme added at 16 h, GC analysis showed that the reaction had reached 97% conversion with 33% selectivity for 2-hydroxy-11,12-epoxy-α-guaiene (**2d**) after 34 h. The total enzyme concentration of the reaction is 4 μM (0.4 μmol). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (3:1 petroleum ether/ethyl acetate) to give 2-hydroxy-11,12-epoxy-α-guaiene (**2d**) as a colourless oil (3.5 mg, 8% yield). R_f 0.25 (3:1 petroleum ether/ethyl acetate). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.65 (s, 1H), 2.71 (d, *J* = 6.5 Hz, 1H), 2.60 (s, 3H), 2.28 – 2.08 (m, 2H), 1.91 – 1.62 (m, 4H), 1.35 – 1.13 (m, 6H), 1.03 (d, *J* = 7.2 Hz, 3H), 0.97 (d, *J* = 6.9 Hz, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 143.63, 143.48, 79.23, 60.81, 55.21, 46.15, 42.90, 42.37, 33.95, 30.37, 30.17, 28.45f, 20.23, 17.99, 16.99.

(*2R*,9*S*)-Dihydroxy-α-guaiene (2e)



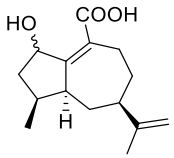
A preparative scale reaction (100 mL) was carried out with the GVQ/A328G variant (2 μM, 0.2 μmol) and guaiene (2 mM, 40.8 mg, 0.2 mmol). After 9 h, GC analysis showed that the reaction had reached 96% guaiene conversion with 35% selectivity of (*2R*,9*S*)-dihydroxy-α-guaiene (**2e**). The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (3:1 petroleum ether/ethyl acetate) to give (*2R*, *9S*)-dihydroxy-α-guaiene (**2e**) as a colourless oil (4.8 mg, 11% yield). R_f 0.3 (3:1 petroleum ether/ethyl acetate). **^{1}H NMR** (500 MHz,

$CDCl_3$) δ 4.74 – 4.62 (m, 2H), 4.54 (d, *J* = 7.3 Hz, 1H), 3.77 (dt, *J* = 10.8, 3.5 Hz, 1H), 2.65 (dt, *J* = 7.4, 3.6 Hz, 1H), 2.53 – 2.39 (m, 2H), 2.22 (ddd, *J* = 13.8, 11.4, 2.1 Hz, 1H), 2.15 – 2.01 (m, 2H), 1.95 (dt, *J* = 13.4, 11.0 Hz, 1H), 1.86 – 1.78 (m, 1H), 1.74 (s, 3H), 1.29 – 1.22 (m, 1H), 1.18 (d, *J* = 7.2 Hz, 3H), 1.05 (d, *J* = 6.5 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ 150.90, 145.45, 139.40, 109.12, 80.66, 72.43, 44.57, 42.50, 41.82, 39.73, 38.36, 32.60, 21.13, 20.38, 12.86. The configuration of C2 and C9 were assigned as (*2R*) and (9*S*) respectively by the nOe data.

2.5.3 *In vitro* screening for 15,2-lactone formation via a cascade oxidation of bulnesene

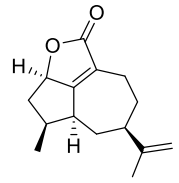
Bulnesene (**1**) was dissolved in EtOH and added as a 400 mM stock solution to a final concentration of 2 mM, with 2 μM of enzyme E1 and 4 μM of enzyme E2 in a volume of 0.5 mL in 24-well plates. Glucose (100 mM) and glucose dehydrogenase (4 U mL^{-1}) were used to regenerate the cofactor NADPH from $NADP^+$ (40 μM). Methyl-β-cyclodextrin was added to 10 mM to increase the aqueous solubility of bulnesene. Screening plates were shaken at 120 rpm for 16 h at 20 °C. An internal standard and 100 μL of 2 M HCl were added to each reaction mixture which was heated at 40 °C for 1 h. 400 μL of the reaction mixture was extracted with 300 μL of ethyl acetate. After centrifugation at 14300 *g* to separate the phases, the organic extracts were washed with water and brine before being analysed by GC.

2.5.4 Preparative scale reactions for isolation of the 2-hydroxybulnesene-15-COOH, 1i



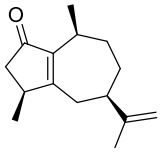
A preparative scale reaction (100 mL in a 250 mL conical flask) was performed at 20 °C and 120 rpm in phosphate buffer (200 mM, pH 7.9) with variant R19/F87I/A82M/A330V (2 μM, 0.2 μmol) and bulnesene (2 mM of 90% pure bulnesene, actual concentration = 1.8 mM, 36.7 mg, 0.18 mmol). As soon as bulnesene was fully converted to bulnsene-15-COOH **1g**, F87A/A184I/I263G/A264G/A330P (4 μM, 0.4 μmol) was added to further oxidise **1g**. The reaction was terminated once **1g** has been fully consumed. The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (100% ethyl acetate) to give 2-hydroxybulnesene-15-COOH (**1i**) as a colourless oil (0.5 mg, 1% yield). R_f 0.12 (100% ethyl acetate). **^{1}H NMR** (600 MHz, $CDCl_3$) δ 4.84 (t, *J* = 7.4 Hz, 1H), 4.67 (s, 2H), 2.94 (dd, *J* = 15.1, 7.1 Hz, 1H), 2.72 (dd, *J* = 11.9, 7.7 Hz, 1H), 2.34 – 2.25 (m, 1H), 2.24 – 2.07 (m, 3H), 1.86 – 1.57 (m, 6H), 1.29 – 1.16 (m, 2H), 1.06 (t, *J* = 7.2 Hz, 3H). **^{13}C NMR** (151 MHz, $CDCl_3$) δ 171.31, 151.32, 109.13, 73.54, 51.55, 49.20, 41.05, 34.35, 31.53, 30.30, 28.6, 20.75, 15.79 (C-15 and C-10 are not detected due to low intensity of the spectrum recorded). The NMR spectroscopic data are consistent with those reported in the literature.[158] **HRMS** (ESI+): Calc'd for $C_{15}H_{22}O_3^+$ $[M+Na]^+$: 273.1461, found: 273.1461.

2.5.5 Preparative scale reactions for isolation of the 15,2-lactone, 1j



A preparative scale reaction (200 mL in a 1 litre conical flask) was performed at 20 °C and 120 rpm in phosphate buffer (200 mM, pH 7.9) with the bulnesene, (2 mM of 90% pure bulnesene, actual concentration = 1.8 mM, 73.4 mg, 0.36 mmol) R19/F87I/A82M/A330V (2 μM, 0.4 μmol) and F87A/A184I/I263G/A264G/A330P (4 μM, 0.8 μmol) variants. GC analysis showed that bulnesene conversion reached >95% after 24 h. The reaction mixture was adjusted to pH 1 with 6 M HCl and heated in a water bath at 40 °C for 4 h, then extracted three times with an equal volume of ethyl acetate. The combined extracts were washed with water and then brine before drying over Na_2SO_4. Solvent was removed *in vacuo*. The crude mixture was isolated with good purity with 38% yield (32 mg). **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.92 (ddd, *J* = 11.2, 6.6, 3.1 Hz, 1H), 4.68 – 4.63 (m, 2H), 2.98 – 2.84 (m, 1H), 2.72 – 2.55 (m, 2H), 2.33 – 2.11 (m, 3H), 1.87 (dddt, *J* = 26.7, 13.0, 5.3, 1.5 Hz, 2H), 1.73 (t, *J* = 1.1 Hz, 3H), 1.41 (dddd, *J* = 13.7, 12.2, 10.0, 3.6 Hz, 1H), 1.27 – 1.11 (m, 2H), 1.06 (d, *J* = 7.2 Hz, 3H). **^{13}C NMR** (400 MHz, $CDCl_3$) δ 176.62, 176.10, 151.98, 125.67, 108.52, 82.15, 47.62, 39.74, 38.19, 36.88, 35.62, 33.52, 23.48, 20.48, 15.33. The NMR spectroscopic data are consistent with those reported in the literature.[151] **HRMS** (ESI+): Calc'd for $C_{15}H_{20}O_2Na^+$ $[M+Na]^+$: 255.13555, found: 255.13560.

2.5.6 Preparative scale reactions for the isolation of rotundone (2f)



The preparative scale reaction was performed at 20 °C and 120 rpm in 100 mL of 200 mM phosphate buffer, pH 7.9, with the RT2/S72G/A330W/L437LA variant. In 6 h, the 76% of the guaiene was

converted with 55% selectivity for rotundol (**2b**). The pH of reaction mixture was adjusted to 8.5 and the temperature was raised to 37 °C before adding CRED-231. After 24 h, guaiene conversion reached 100% and rotundol was also fully converted, with 45% of rotundone. The reaction mixture was extracted with ethyl acetate and the products were separated by silica gel column chromatography (20:1 petroleum ether/ethyl acetate) to give rotundone (**2f**) as a yellow oil (5.2 mg, 12% yield). R_f 0.45 (20:1 petroleum ether/ethyl acetate). **^{1}H NMR** (500 MHz, $CDCl_3$) δ 4.76 – 4.67 (m, 2H), 2.99 (dq, *J* = 7.4, 3.7 Hz, 1H), 2.77 – 2.68 (m, 1H), 2.64 – 2.56 (m, 2H), 2.35 (dt, *J* = 15.9, 2.0 Hz, 1H), 2.08 – 1.93 (m, 2H), 1.91 – 1.74 (m, 6H), 1.52 (m, 1H), 1.12 (d, *J* = 7.1 Hz, 3H), 1.01 (d, *J* = 7.3 Hz, 3H).**^{13}C NMR** (400 MHz, $CDCl_3$) δ 208.08, 176.94, 151.08, 145.43, 109.06, 46.13, 43.01, 37.90, 36.79, 32.67, 30.72, 26.88, 20.27, 19.19, 17.55. The NMR spectroscopic data are consistent with those reported in the literature.[151] **HRMS** (ESI+): Calc'd for $C_{15}H_{23}O^+$ $[M+H]^+$: 219.17434, found: 219.17444.

2.6 Biocatalytic synthesis of 15,2-lactone with bulnesene-enriched patchouli oil

The reaction (100 mL) was performed at 20 °C and 120 rpm in 200 mM phosphate buffer, pH 7.9, with the R19/F87I/A82M/A330V and F87A/A184I/I263G/A264G/A330P variants. A bulnesene-enriched patchouli oil fraction containing 50% bulnesene (by GC) was added as 400 mM stock solution in ethanol to a final bulnesene concentration of 2 mM. Two 400 µL aliquots were removed periodically for GC analysis. Methyl benzoate (1 M stock in methanol) was added as internal standard to help quantify the relative proportion of the 15,2-lactone formed. One aliquot was acidified with 100 µL of 2 M HCl and heated at 40 °C to promote lactone formation. The internal standard (4 µL) was added to both aliquots and the mixtures extracted with 300 µL of ethyl acetate. After centrifuging at 14 300 *g* for 1 min to separate phases, the organic layers were removed for GC analysis. Once the amount of lactone formed had stopped increasing, the reaction mixture was adjusted to pH 1 and heated in a water bath at 40 °C for 4 h. The relative amount of **1j** formed was

calculated based on the relative peak area of **1j** formed to the relative peak area of internal standard on GC.

2.7 Molecular dynamics simulations and molecular docking

MD simulations were performed on $P450_{BM3}$ variants which showed interesting selectivity and activity for sesquiterpene oxidation. *In silico* mutagenesis was performed in Pymol on the structure of the WT enzyme with *N*-palmitoylglycine (NPG) bound within the active site (PDB: 1jpz).[159] The NPG ligand was removed, and the heme prosthetic group was replaced with a ferryl species using a set of parameters from Shahrokh *et al.*[160] This hypothetical ferryl form of the variant enzyme was then subjected to MD simulation by GROMACS 2018.6 for 100 ns.[161] All simulations were carried out in triplicate. The protein was prepared using the Amber 99SB*-ILDN force field and TIP3P water.[162,163] The protein was placed into the centre of a truncated octahedral box, with a minimum distance of 10 nm to any box edge. Approximately 17 500 water molecules were added to solvate the protein and Na^+ ions were added to neutralise the net charge. Steepest-decent energy minimisation was performed until the maximum force was <500 kJ mol^{-1} nm^{-2}. Periodic boundary conditions were used to model the system. The bond lengths involving H-atoms were constrained with the LINCS algorithm while the electrostatic interactions were treated by the particle mesh Ewald method.[164,165] Short-range non-bonded interactions were calculated within a 1 nm cut-off and a timestep of 2 fs. For equilibration steps, all the protein backbone Cα atoms were restrained with a positional restraint force constant of 1 000 kJ mol^{-1} nm^{-2}. Using the modified Berendsen thermostat, a NVT equilibration step was performed at 298 K for 100 ps followed by an NPT equilibration step for 1 250 ps with a time constant of 1 ps, using the Berendsen barostat.[166,167] Each replica MD simulation was performed for 100 ns, a snapshot structure was recorded every 10 ps. The output for each replica was clustered using the Daura algorithm with a RMSD cut-off of 1.2 Å.[168] The top 3 clusters, which typically accounted for >80% of all structures, were compared to ensure that no drastic and unreasonable changes had occurred. Substrates were docked into the clustered structures

(typically 9 clusters from the 3 replica simulations) by Autodock Vina. All water molecules were removed from the structure prior to the docking calculation. The docking site was defined as a 30 × 30 × 30 Å box, centred on the ferryl oxygen. The output poses were ranked using the scoring function in Autodock Vina.[91] Each substrate docking generated 9 poses which were classified as productive or non-productive. A productive pose is one with a carbon atom within 4 Å of the ferryl oxygen while a non-productive pose has the substrate bound too far away from the ferryl oxygen.

Chapter 3

3 The oxidation of α-bulnesene by CYP102A1

3.1 Purification of α-bulnesene and α-guaiene

Patchouli essential oil (PEO) was chosen as the source of bulnesene and guaiene because, apart from the major component, patchoulol, which makes up 27–35%, the other volatile components are sesquiterpene hydrocarbons (Fig. 3.1). Of these, α-guaiene constitutes 11–16% and α-bulnesene 13–21%. PEO is one of the most important plant essential oils in the aroma and flavour industry, with annual production exceeding 1,400 tonnes. The price depends on the proportion of patchoulol in the oil, with material containing >30% patchoulol commanding higher prices. Materials containing less patchoulol, and thus higher proportions of sesquiterpenes, are therefore an economic source of α-bulnesene, **1**, and α-guaiene, **2**.

PEO was firstly subjected to fractional distillation to separate patchoulol from the hydrocarbons and potentially to produce fractions with differential enrichment of the sesquiterpenes. The oil (50 mL) was placed in a 100 mL round bottom flask fitted with a Vigreux column and the system was brought to reflux at 125 °C and 1 mbar. A fraction was collected when the observed vapour temperature started to stabilise. The temperature was increased in 10 °C steps after collecting each fraction until a new stable vapour temperature was established. The first three fractions were collected at vapour temperatures of 70 °C, 70–73 °C, and 82 °C. The fractions were analysed with gas chromatography (Fig. 3.1). The percentage of α-guaiene (**2**) increased from 14.3% in the PEO sample to 26.3% in Fraction 1, with enrichment over the proportion of α-bulnesene (**1**). There was less enrichment of α-guaiene over α-bulnesene in Fraction 2, but both fractions contained greatly reduced proportions of patchoulol (**6**). It was clear that Fraction 1 was slightly enriched in the lower boiling sesquiterpenes while Fraction 2 had more of the sesquiterpenes with intermediate boiling points. Fraction 3 was enriched in α-bulnesene, to 38.2% from 19.5% in PEO although it also contained more patchoulol.

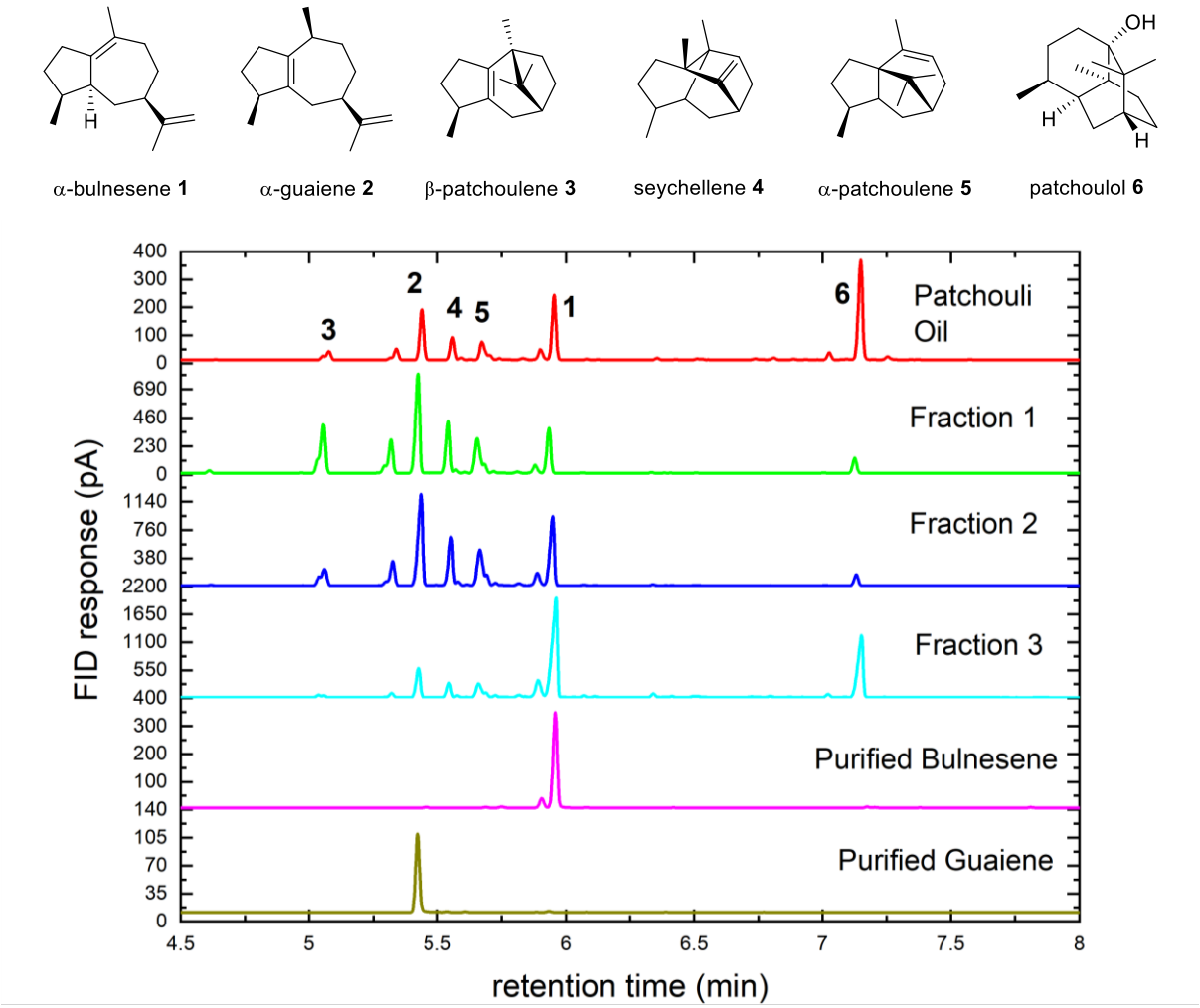


Figure 3.1 GC analysis of patchouli essential oil (PEO), the distillation fractions of PEO, and α-bulnesene (**1**) and α-guaiene (**2**) purified with silver nitrate-impregnated silica.

Fractions 1 and 3, which had the highest amount of α-guaiene and α-bulnesene respectively, were further purified by column chromatography before being used in enzymatic reactions. As sesquiterpenes have very similar polarity, their separation by silica gel chromatography was poor; α-bulnesene (**1**) and α-guaiene (**2**) were obtained in 60% and 49% purity respectively. However, the purity of both sesquiterpenes increased to 95% for α-guaiene and 92% for α-bulnesene when silver nitrate-impregnated silica gel was used (Fig. 3.1). The silver ions interacted weakly with the C=C double bonds of sesquiterpenes, allowing better separation of the hydrocarbons. Moreover, β-patchoulene (**3**), seychellene (**4**), and α-patchoulene (**5**) were also obtained in high purity (>90%) by $AgNO_3$-SiO_2 column chromatography (Table 3.1).

Table 3.1 The isolated yield of α-bulnesene (**1**), α-guaiene (**2**), β-patchoulene (**3**), seychellene (**4**) and α-patchoulene (**5**) based on their respective quantities in PEO distillation fractions purified by $AgNO_3$-SiO_2 column chromatography.

Sesquiterpenes	Distillation Fraction	Isolated yield (%)
Bulnesene (**1**)	F3	40
Guaiene (**2**)	F1	34
β-patchoulene (**3**)	F1	17
Seychellene (**4**)	F1	20
α-patchoulene (**5**)	F1	19

3.2 Bulnesene oxidation by CYP102A1

3.2.1 Initial library screening and product characterisation

The purified α-bulnesene (**1**) was screened with a library of 44 CYP102A1 variants. Such screening scale reactions (0.5 mL in 24-well plates) were carried out at a 1000:1 ratio, with 2 mM α-bulnesene and 2 μM P450 enzyme. Glucose (100 mM) and glucose dehydrogenase (4 U mL^{-1}) were used to regenerate the cofactor NADPH (40 μM). Methyl-β-cyclodextrin was added to 10 mM to increase the solubility of α-bulnesene in buffer. After shaking at 120 rpm, 20 °C for 6 hours, the organics in each well were extracted with ethyl acetate and analysed by GC. Full activity and selectivity data are given in Appendix 2.

The screening library was constructed from five CYP102A1 template variants, GVQ (A74G/F87V/L188Q), RT2 (R47L/Y51F/A191T/N239H/I259V/A276T/L353I), K19 (H171L/Q307H/N319Y), R19 (R47L/Y51F/K19), and RP (R47L/Y51F/I401P).[21,66–70] Additional mutations were introduced at two or more active site residues including R47, Y51, S72, V78, F81, A82, F87, A184, I263, A264, E267, A328, A330, and L437, to generate a library of variants with different substrate pocket sizes and shapes to search for matches with unnatural substrates for their oxidation.

No activity was observed for the wild-type enzyme (WT). Of the variants, 25 gave >50% conversion of **1** (turnover number, TON, >500) which indicated that α-bulnesene was a good substrate for CYP102A1. Five major products were isolated by silica column chromatography from preparative scale reactions and characterised by their NMR and MS data as the α-bulnesene C15-alcohol (**1a**), the C15-aldehyde (**1b**), (2*S*)-hydroxy-α-bulnesene (**1c**), (2*R*)-hydroxy-α-bulnesene (**1d**) and 11,12-epoxy-15-hydroxy-α-bulnesene (**1e**). The NMR data for products **1a**, **1b**, **1c** and **1d** were consistent with those reported in the literature while **1e** is a new metabolite of α-bulnesene.[141,151,155] Full characterisation data of the products are included in Chapter 2. Interestingly, while **1a** and **1b** were isolated through the extraction of agarwood, **1c** was transformed through the ring opening of (–)-1,10-epoxyguai-11-ene catalysed by aluminium isopropoxide.[151] On the other hand, the C2-alcohols were obtained through the allylic oxidation of α-bulnesene with CrO_3 and 3,5-dimethyl-1W-pyrazole to form bulnesene-2-one which was then reduced to (2*S*)- and (2*R*)-hydroxybulnesene via Luche reduction. The oxidation of α-bulnesene-rich olefinic fraction with laccase has also been reported and formed a mixture with a range of different terpenoids including (2*S*)- and (2*R*)-hydroxybulnesene.[155]

In the ^{1}H NMR spectrum of **1a**, the resonance of the allylic methyl C15 in the substrate was missing, indicating C15 oxidation to a monooxygenated product. The two-proton AB-quartet at 4.10 ppm confirmed the structure of **1a** as the C15-alcohol.[151] The presence of a characteristic low-field aldehyde singlet (9.92 ppm), the absence of the C15-methyl resonance and the MS data (+14) demonstrated that product **1b** was the C15-aldehyde (Fig. 3.3).[141]

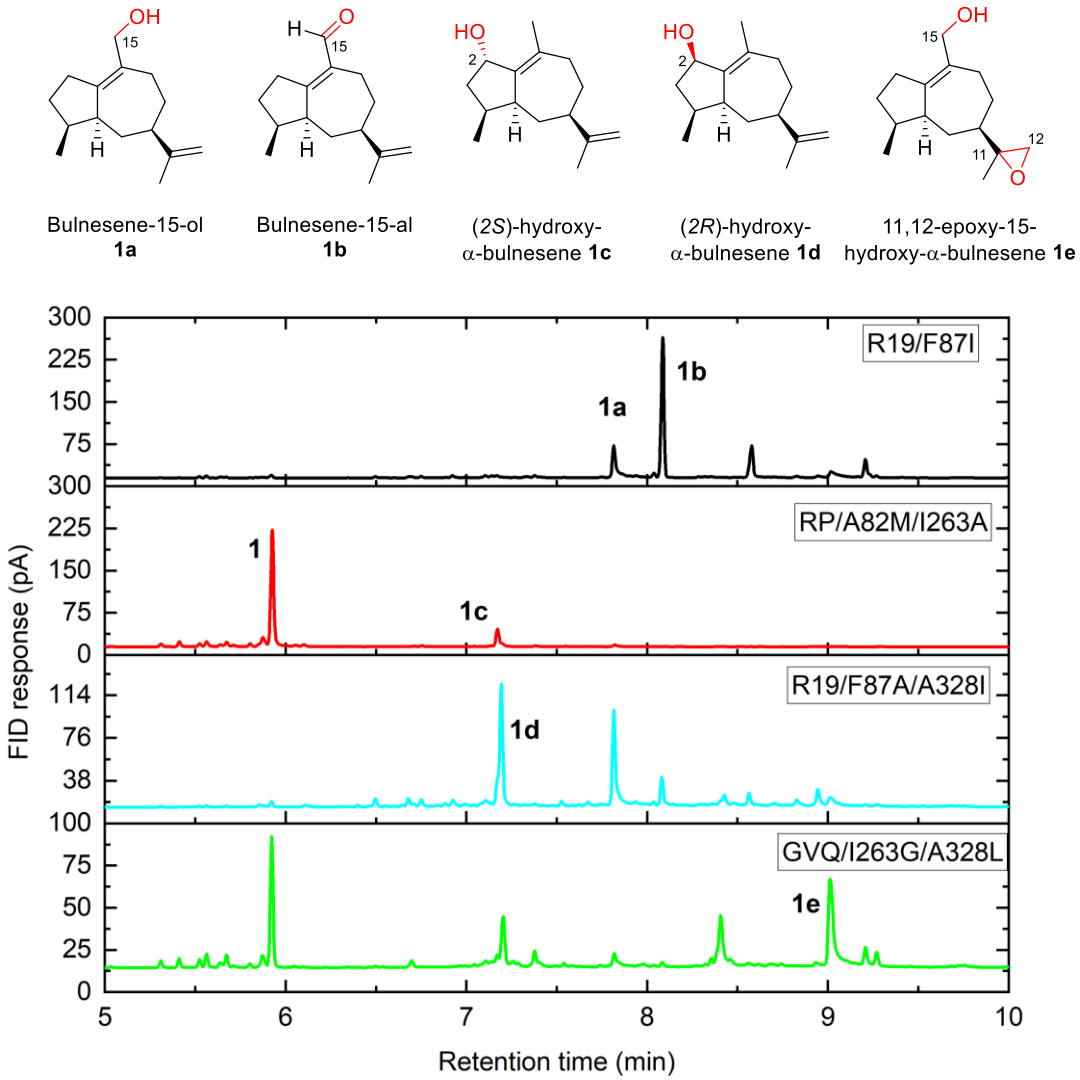


Figure 3.2 GC analysis of selected variants giving α-bulnesene oxidation products **1a** to **1e**.

The diastereomeric (2*S*)-hydroxy-α-bulnesene (**1c**) and (2*R*)-hydroxy-α-bulnesene (**1d**) had very similar NMR spectra. The resonance of the allylic C2-alcohol of both stereoisomers had similar chemical shift (around 4.66 ppm), as did the C12 methylene protons (around 4.66 ppm).[141, 155]

A peak at 259 m/z corresponding to $(M + 32 + Na)^+$ was observed in the mass spectrum of product **1e** which indicated that two oxygens were added to the substrate (RMM 204 g mol^{-1}). The two-proton resonance at 4.09 ppm in ^1H NMR spectrum indicated a C15 hydroxyl group. The C12 olefinic protons in α-bulnesene, which appeared at 4.70 ppm, had shifted upfield to around 2.60 ppm. Hence, the double bond at C11 and C12 had undergone epoxidation and product **1e** was characterised as 11,12-epoxy-15-hydroxy-α-bulnesene.

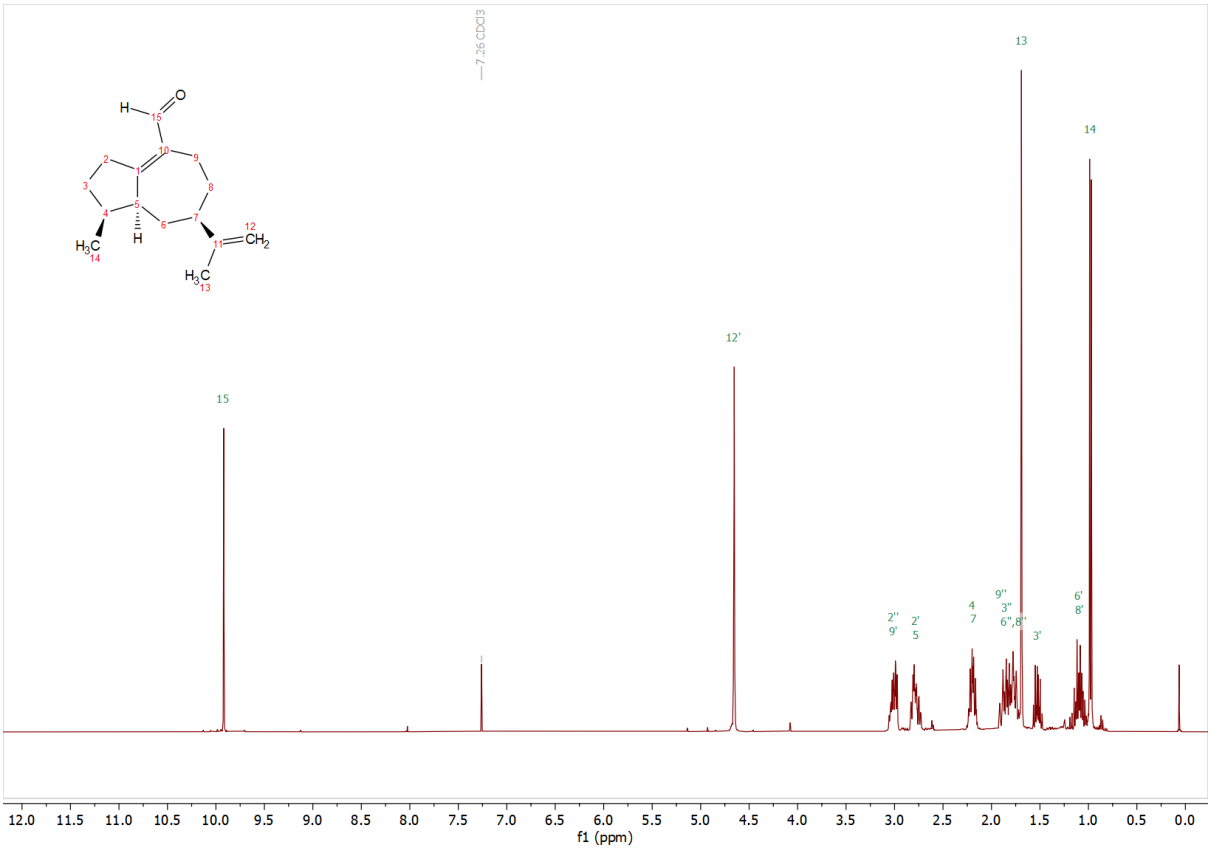


Figure 3.3 The ^{1}H NMR spectrum of the C15-aldehyde, bulnesene-15-al (**1b**) in $CDCl_3$.

3.2.1.1 The effect of mutations

In general, most variants without mutation of residue F87 had low conversion except the variant RP/V78I/E267V and a few RT2-based variants. This is most likely because the phenyl side chain of F87 is directly above the heme and prevents the substrate from binding sufficiently close to the ferryl oxygen for oxidation. For example, RP-based variants gave low conversion in general (<40%) including variant RP/A82M/I263A which gave the highest selectivity of (2*S*)-hydroxy-α-bulnesene, **1c** (38%, Table 3.2, Entry 2). Mutating F87 to residues with a smaller side chain, such as valine (V), alanine (A), and isoleucine (I), created space close to the heme for substrate binding and oxidation. RT2-based variants had been found to show increased activity for unnatural substrates oxidation.[55] Even without a F87 mutation, most of the RT2 series surprisingly showed high conversion but low product selectivity (Entries 7–10).

Variants with the F87V mutation were in three groups, GVQ, K19/F87V and RP/F87V. The GVQ series of variants had low to moderate conversion except GVQ/I263G/A328L which produced the most 11,12-epoxy-15-hydroxy-α-bulnesene, **1e** (34%), after K19/F87A/I263G (41%) among the variants screened (Entries 14 and 29). Variant GVQ/I263G, which also had a mutation at I263, showed low conversion (Entry 12). This demonstrated the importance of the I263G/A328L combination for α-bulnesene oxidation. Without the L188Q mutation, the conversion and activity of bulnesene by variant GV/A184I increased dramatically (100%, Entry 15). All these variants produced many products which had longer GC retention times, suggesting that multiply oxidised compounds were formed with little selectivity.

Most RP/F87V-based variants showed higher conversion to give similar patterns of multiply-oxygenated products (Entries 16–18). K19/F87V-based variants also showed high conversion generally and formed numerous products except for the K19/F87V/E267V and K19 /F87V/V78I/E267V variants which had low conversion (<25%) suggesting that the E267V mutation hindered the oxidation of α-bulnesene (Entries 19–24). This could be because the acid group on residue E267 is involved in the proton delivery steps of the catalytic cycle and thus brings adverse impact on the enzyme activity once being mutated.[21]

Compared to the K19/F87V-based variants, the K19/F87A-based variants had lower conversion (Entries 25–29). However, α-bulnesene conversion increased dramatically with the addition of mutations R47L/Y51F, for example, the R19/F87A variant showed very high activity although it gave a range of different products with low selectivity (Entry 30). This is likely because mutation F87A has widened the substrate pocket which allows different substrate binding orientations to present different substrate C–H bonds for attack by the ferryl oxygen. On the other hand, mutation of A328 to residues with bulkier side chains, such as isoleucine (I), phenylalanine (F), and leucine (L), re-shaped the substrate pocket and led to more selective oxidation, mainly at the C2 and C15

positions. Variants R19/F87A/A328F and R19/F87A/A328I gave the highest selectivity for (2*R*)-hydroxy-α-bulnesene (**1d**) at 39% and 34% respectively (Entries 32 and 33).

Table 3.2 Substrate conversion and product selectivity of selected variants from initial library for the oxidation of α-bulnesene, **1**. The variants are grouped based on the base variants and mutations on residue F87 for better understanding of their effect respectively.

Entry	CYP102A1 variants	Base Variant	F87	Conv.	1a	1b	1c	1d	1e
1	RP/V78I/E267V	RP	F87	97%	7%	34%	2%	–	19%
2	RP/A82M/I263A			31%	6%	–	**38%**	–	–
3	RP/A82W/I263A			28%	–	–	13%	8%	–
4	RP/H171L/I263G			36%	8%	–	9%	15%	8%
5	RP/H171L/A184I/I263G			39%	6%	–	16%	13%	6%
6	RP/I263A/E267V			34%	29%	–	16%	4%	–
7	RT2	RT2		95%	12%	–	14%	12%	13%
8	RT2/S72G/A330W			84%	20%	–	32%	15%	7%
9	RT2/S72W/A330W			64%	21%	–	14%	10%	6%
10	RT2/V78I/A184I/A330P			89%	11%	2%	22%	16%	12%
11	GVQ	GVQ	F87V	32%	27%	–	14%	17%	–
12	GVQ/I263G			21%	–	–	–	–	–
13	GVQ/A264G			62%	**31%**	4%	27%	10%	6%
14	GVQ/I263G/A328L			79%	–	–	–	11%	34%
15	GV/A184I			100%	–	–	3%	–	15%
16	RP/F87V	RP		96%	4%	25%	2%	–	17%
17	RP/F87V/V78I			35%	16%	–	25%	16%	–
18	RP/F87V/V78I/H171L/I263G			67%	8%	3%	4%	8%	27%
19	K19/F87V	K19		100%	5%	20%	5%	–	20%
20	K19/F87V/A264G			91%	14%	5%	23%	6%	15%
21	K19/F87V/E267V			23%	12%	9%	–	–	–
22	K19/F87V/V78I/E267V			22%	12%	–	–	–	–
23	K19/F87V/A328V			91%	17%	17%	10%	18%	26%
24	K19/F87V/Q403P			83%	21%	24%	13%	8%	17%
25	K19/F87A/F81W		F87A	39%	28%	–	9%	12%	18%
26	K19/F87A/A82M/I263G			30%	18%	–	19%	–	–
27	K19/F87A/A82M/E267F			27%	21%	–	–	12%	–
28	K19/F87A/I263A			32%	26%	–	8%	19%	12%
29	K19/F87A/I263G			68%	17%	–	6%	11%	**41%**
30	R19/F87A	R19		100%	12%	5%	3%	–	30%
31	R19/F87A/A184I			94%	7%	2%	3%	–	28%
32	R19/F87A/A328F			50%	23%	–	3%	**39%**	7%
33	R19/F87A/A328I			100%	29%	8%	4%	34%	5%
34	R19/F87I		F87I	100%	15%	**51%**	–	–	6%

The selectivity for the most selective variants for each product is in bold.

Of greatest interest for agarwood aroma synthesis was the variant R19/F87I which showed total α-bulnesene conversion and high selectivity for C15 oxidation (Entry 34). Importantly, it also showed further oxidation of C15-alcohol **1a** to the C15-aldehyde **1b** (51%), which is one of the key contributors to the most desirable agarwood aroma. By contrast, the R19/F87A variant gave the 11,12-epoxide of bulneseone-15-ol (**1e**) (30%, Entry 30). The F87I mutation could be crucial in biasing not only the regioselectivity towards C15 oxidation but in promoting the oxidation of bulnesene-15-ol **1a** to aldehyde **1b**.

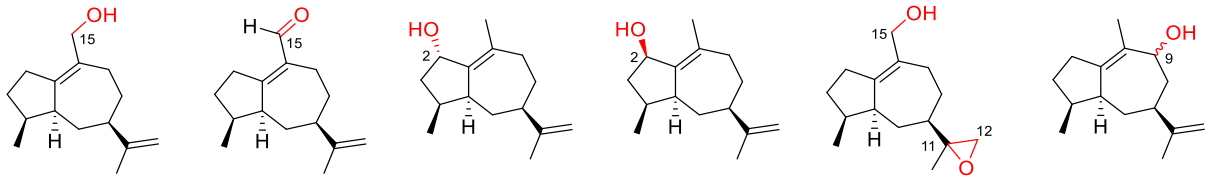
3.2.2 Sub-library screening of bulnesene

The initial screening of α-bulnesene with a small library of just 44 variants identified residues and mutations of CYP102A1 that promoted oxidation of this unnatural substrate and affected the product profiles, i.e., regioselectivity and chemoselectivity (monooxygenation *vs*. multiple oxygenations). The results indicated the crucial role of the residue at the 87 position; the RT2 and RP series of variants which did not have a mutation at F87 gave either low conversion or poor product selectivity in general. The F87A-based variants generally lacked selectivity unless being combined with mutations on residue A328. The F87V series showed the effect of other mutations in modulating the product profile (for example, formation of the C2 alcohols and epoxides by the GVQ/I263G/A328L variant). The F87I mutation favoured C15 oxidation to both the alcohol and aldehyde. Based on these results, the full library of more than 500 CYP102A1 variants available in the group enzyme collection was scrutinised for variants derived from these "hits" within the initial screening library.

Five sub-panels of variants were selected from the whole library. Three panels were based on the R19, RP and F87I variants that included additional mutations at various active site residues such as those in the B′ helix (S72 – A82), F helix (A184, L188), I helix (T260, I263, A264), and the β1 strand (A328, P329, A330). The fourth panel was based on GVQ/I263G/A328L and contained different combinations of hydrophobic amino acid substitutions at F87 and A328, while the fifth panel was

an F87A-based series of variants with additional mutations at residues in the I helix and β1 strand. A total of 78 additional variants were screened, and the activity and selectivity of selected variants are summarised in Table 3.3.

Table 3.3 Substrate conversion and product selectivity of selected variants for the oxidation of α-bulnesene, **1**.



Bulnesene-15-ol **1a** | Bulnesene-15-al **1b** | (2*S*)-hydroxy-α-bulnesene **1c** | (2*R*)-hydroxy-α-bulnesene **1d** | 11,12-epoxy-15-hydroxy-α-bulnesene **1e** | 9-hydroxy-α-bulnesene **1f**

Entry	CYP102A1 variants	Conv.	1a	1b	1c	1d	1e	1f	TON
1	KT2/N70S/F87I	61%	**45%**	22%	–	–	–	–	275
2	K19/F87I	79%	**52%**	18%	4%	–	6%	–	415
3	R19/A184I/T260G	80%	**63%**	–	3%	7%	7%	–	505
4	R19/A82M/T260G	69%	**66%**	–	6%	–	–	–	455
5	R19/A82M/A184I/T260G	93%	**54%**	14%	–	–	6%	–	505
6	R19/T260G	9%	–	–	–	–	–	–	–
7	R19/P329G/A330W	51%	14%	–	17%	**38%**	–	–	195
8	RP/A82M/H171L/I263G	61%	12%	–	**57%**	–	–	–	350
9	GLQ/I263G	75%	**42%**	–	21%	13%	–	–	315
10	GAQ/I263G/A328L	97%	–	–	–	–	**34%**	–	330
11	GLQ/I263G/A328L	99%	8%	–	–	–	**52%**	–	515
12	GGQ/I263G/A328L	87%	–	–	–	–	–	–	–
13	R19/F87A/A264G	87%	10%	–	7%	–	–	**26%**	230
14	R19/F87A/A264G/A328G	77%	10%	–	5%	–	–	**36%**	280

Screening scale reactions were in 200 mM phosphate buffer containing 2 μM CYP102A1 variants, 2 mM α-bulnesene, 40 μM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 6 hours. The selectivity for the most selective variants for each product is in red. TON refers to the yield of the product in bold relative to enzyme.

The F87I panel

It was slightly disappointing to find that most variants of the F87I sub-panel had little to no α-bulnesene oxidation activity other than variants KT2/N70S/F87I and K19/F87I. Variant KT2/N70S/F87I showed similar profile to R19/F87I, converting bulnesene to C15-alcohol **1a** (45%) and C15-aldehyde **1b** (22%, Table 3.3, Entry 1) with moderate conversion (61%). KT2/F87I did not convert bulnesene under the same conditions, which indicated that the N70S mutation increased

bulnesene oxidation activity. Variant K19/F87I showed 70% total C15 oxidation but the conversion was significantly lower (79%) as compared to variant R19/F87I (100%). This showed the importance of mutations R47L/Y51F in promoting the oxidation of more hydrophobic substrates while retarding the oxidation of more polar compounds (Table 3.3, Entry 2; Table 3.2, Entry 34).[21,41]

The R19 panel (R47L/Y51F/H171L/Q307H/N319Y)

The R19 library of variants were assembled as R19/F87I was found to give high C15 selectivity and it also offered comparisons with variants of F87 mutations. Variants with A82M and/or A184I mutations were found to increase selectivity or conversion. The variants R19/A184I/T260G (63%), R19/A82M/T260G (66%) and R19/A82M/A184I/T260G (54%) showed high selectivity for C15-alcohol **1a**, with little further oxidation to the C15-aldehyde **1b** (Table 3.3, Entries 3–5), which demonstrated that mutations A82M, A184I and T260G could be crucial in controlling the chemoselectivity of C15 oxidation. This is important as the alcohol product could be separated from the hydrocarbon substrate by simple vacuum distillation. However, unlike mutations A82M and A184I which generally increased variant activity, T260G was found to give adverse effect on bulnesene oxidation if not combined with mutations A82M and A184I. For example, variant R19/T260G gave <10% conversion of bulnesene (Entry 6).

Variant R19/P329G/A330W gave similar selectivity for (*2R*)-alcohol **1d** (38%, Table 3.3, Entry 7) to the most selective variant from the initial library, R19/F87A/A328F (39%, Table 3.2, Entry 32), both at roughly 50% conversion. This demonstrated that mutations at the β1 strand (A328, P329, A330) could be crucial in orientating bulnesene for *2R* oxidation.

The RP panel (R47L/Y51F/I401P)

Variants based on the template variant RP, which did not contain a mutation of F87, favoured the formation of (2*S*)-hydroxy-α-bulnesene (**1c**), especially when combined with mutations I263A and A82M, with variant RP/A82M/I263A giving 38% of **1c** (Table 3.2, Entry 2). Variant RP/A82M/H171L/I263G from the RP panel gave higher selectivity for **1c** (57%) with significantly higher conversion (61%, Table 3.3, Entry 8).

The GVQ panel (A74G/F87V/L188Q)

The majority of variants in this panel showed moderate to high activity and gave a high diversity of products, some with reasonable selectivity; the GLQ/I263G variant, for example, gave C15-alcohol **1a** at 42% (Table 3.3, Entry 9). However, the wide range of products indicated that variants based on the GVQ variant might not be suitable for further engineering for selective bulnesene oxidation. This was demonstrated with the GXQ/I263G/A328L series of variants where the residue at the 87 position was varied from F to G, A, V and L. Both the GAQ/I263G/A328L and GVQ/I263G/A328L variants gave 34% of 15-hydroxy-11,12-epoxy-α-bulnesene **1e** (Table 3.3, Entry 10; Table 3.2, Entry 14) while GLQ/I263G/A328L gave the higher selectivity of 52% (Table 3.3, Entry 11) and the F87G analogue did not give any **1e** (Entry 12). A time-course reaction of GLQ/I263G/A328L variant was performed to study the formation of **1e** with variant GLQ/I263G/A328L (Fig. 3.4). Bulnesene **1** was almost fully converted in 2 hours with three main products, (*2R*)-alcohol **1d**, C15-alcohol **1a**, and the epoxide-alcohol **1e**. As the reaction continued, the concentration of **1a** decreased gradually with a concomitant increase in **1e** concentration, which demonstrated that C15-alcohol **1a** was the precursor to this double oxygenation product.

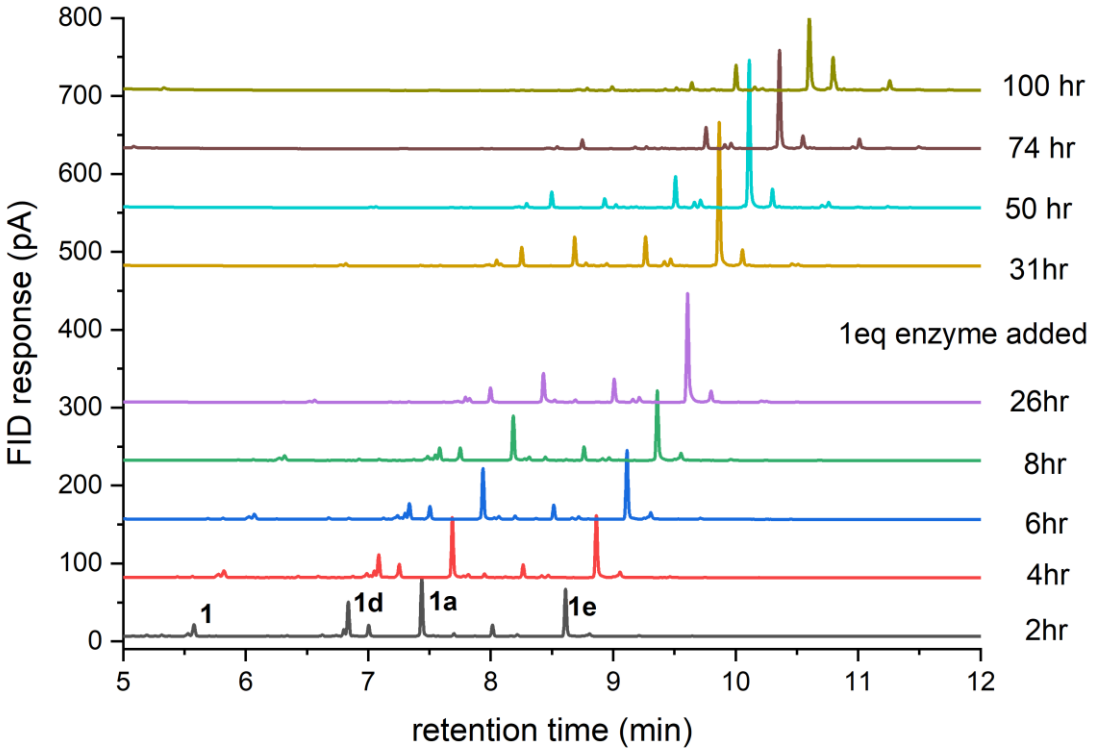


Figure 3.4 A time-course reaction of the GLQ/I263G/A328L variant with bulnesene **1**.

The F87A panel

The results of the initial screening showed that the addition of A328 mutations altered the shape and size of the substrate pocket of variant R19/F87A and promoted more selective oxidation of bulnesene. Addition of mutations of I-helix residues led to a wide range of products, with little selectivity. However, a new product was formed with variants R19/F87A/A264G and R19/F87A/A264G/A328G at moderate selectivity (26% and 36% respectively, Table 3.3, Entries 13 and 14) which was isolated and characterised as the C9-alcohol **1f**. The ^{1}H NMR spectrum showed an allylic alcohol proton at 4.45 ppm, slightly more upfield compared to the C2-alcohols. A review of all screening data showed that **1f** was formed in low selectivity by variant GVQ/A264G (14%) and K19/F87V/A264G (13%). Moreover, the R19/F87A base variant gave 12% of **1f**. These observations showed that the A264G mutation could be crucial for C9 oxidation of bulnesene.

3.2.2.1 Summary

Screening of the initial panel of variants followed by a broader range of CYP102A1 variants selected on the basis of selectivity trends from the initial screen showed bulnesene to be a good substrate for the engineered enzymes.

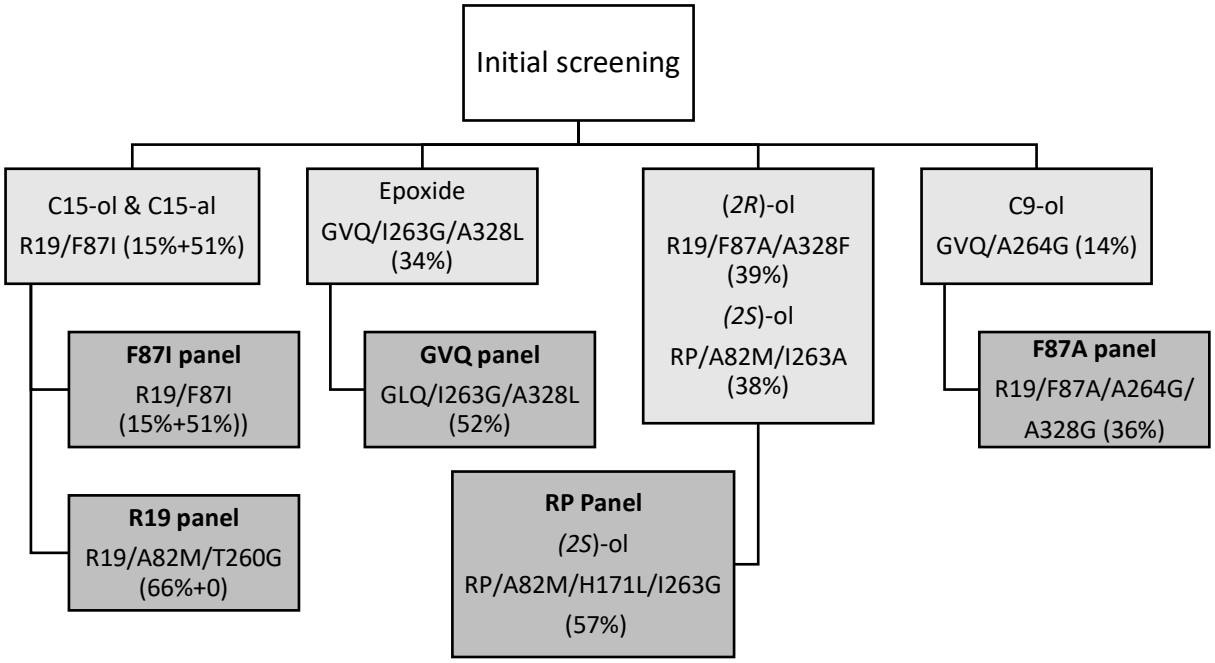


Figure 3.5 Progressive library screening for each oxidation product of bulnesene **1** from initial screening. The selectivity of each variant is in bracket. Light grey boxes represent variants from the initial library while dark grey boxes represent variants from each sub-library, with the name of each library in bold.

Almost full conversion was observed with many variants while the selectivity data showed the pivotal role of the residues at positions 87 and 328. C–H bond oxidations were observed at C2, C9 and C15, but also epoxidation of the C11=C12 double bond. Variants without a mutation of F87 and with the A82M and I263G/A mutations favoured the (2*S*)-alcohol **1c** whereas the F87A/A328I or F combination gave more of the (2*R*)- diastereomer **1d**. When the A264G and A328G mutations were added to the R19/F87A template, the selectivity was switched towards the C9-alcohol **1f**. Variants based on the F87V mutation tended to show lower product selectivity. On the other hand, the F87I mutation was found to favour C15 oxidation strongly, to the extent that the R19/F87I variant was able to oxidise bulnesene to C15-alcohol **1a** and then the C15-aldehyde **1b**. This latter product is an

important agarwood aroma compound and the initial screen had identified a variant with high C15 selectivity, offering a good template for further enzyme engineering and process optimisation.

Progressive screening led to the identification of existing variants with high product selectivity. Instead of having to screen all the variants in the full library of >500 variants, significant improvement in product selectivity was achieved with merely 78 variants. More importantly, residues that play a role in product selectivity could be identified more easily to provide a good basis for further mutagenesis. For example, from the screening results of the R19/A82M/T260G and R19/A184I/T260G variants, the A82M and A184I mutations were found to be crucial for C15 selectivity while T260G retarded bulnesene oxidation; the information which would not have been discovered without screening the sub-libraries.

3.3 Synthesis of bulnesene-15-COOH

Bulnesene-15-COOH, which is also present in agarwood, was not observed as a major product in the initial screening. It is a proposed precursor to (–)-guaia-1(10),11-dien-15,2-olide, also known as the 15,2-lactone, and is the major contributor to the agarwood aroma. The biosynthesis pathway for bulnesene-15-COOH has not yet been discovered but it is likely to involve a combination of several enzymes, including a P450 enzyme which performs the C–H activation at C15 of bulnesene, an alcohol dehydrogenase or another P450 which oxidises the alcohol to the aldehyde and finally an aldehyde dehydrogenase to oxidise the aldehyde to the carboxylic acid. In this section, the synthesis of bulnesene-15-COOH was explored with several different strategies such as chemoenzymatic synthesis and biosynthesis by engineered CYP102A1 variants.

3.3.1 Chemoenzymatic synthesis of bulnesene-15-COOH

The chemoenzymatic synthesis of bulnesene-15-COOH (**1g**) from bulnesene (**1**) was established by first using the P450 variant, R19/F87I, for C15 oxidation of bulnesene to a mixture of the alcohol and aldehyde. The reaction was set up with less glucose (25 mM) since excess glucose would

continuously recycle $NADP^+$ to NADPH, reducing its availability for the CRED enzyme which was to be used to convert the alcohol to the aldehyde. The amount of added glucose required for maximum bulnesene conversion by the R19/F87I variant without influencing the performance of CRED in the next step was determined by control experiments.

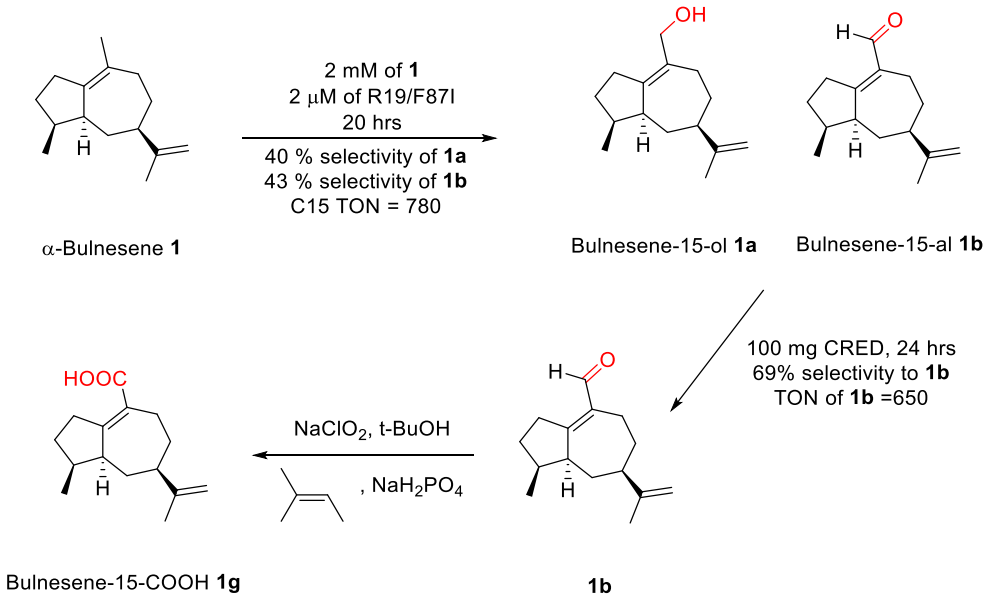


Figure 3.6 The chemoenzymatic synthesis of bulnesene-15-COOH **1g**.

The major products were bulnesene-15-ol (**1a**) and bulnesene-15-al (**1b**) after the overnight reaction with variant R19/F87I reached almost full conversion (94%). The pH of the phosphate buffer was then adjusted to 8.5 and the temperature was raised to 37 °C since the equilibrium for the CRED reaction lies on the aldehyde side at alkaline pH and optimum CRED activity was observed at 37 °C. A carbonyl reductase, CRED-231 and $NADP^+$ (40 μM) were used to convert **1a** to **1b**, which was then oxidised to the C15-carboxylic acid **1g** by the Pinnick reaction. CRED-231 used was from Oxford Biotrans and sourced from an undisclosed supplier.

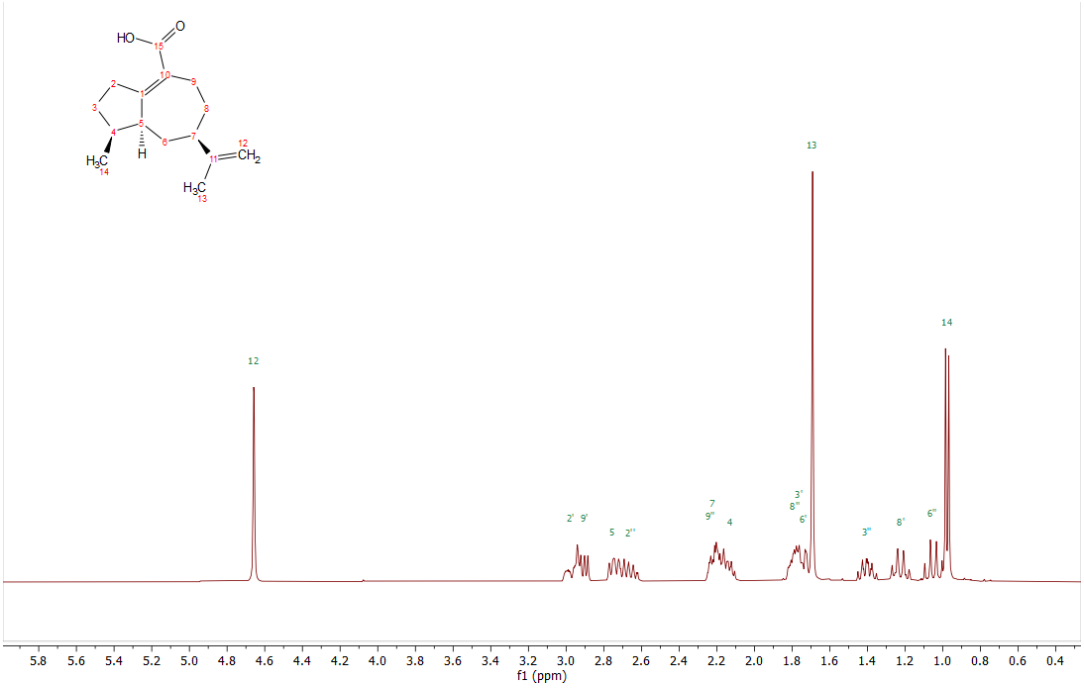


Figure 3.7 The ^{1}H NMR spectrum of the C15-carboxylic acid, bulnesene-15-COOH **1g** in $CDCl_3$.

The product was isolated by silica column chromatography and characterised by the NMR and MS data which were in agreement with reported data in the literature.[151] The MS data (+30) showed that it was a carboxylic acid and the loss of the C15 methyl resonance in the ^{1}H NMR spectrum further demonstrated that it was bulnesene-15-COOH (Fig. 3.7). However, chemical synthesis is less desirable due to the increasing demand for natural flavours and fragrances and thus, the biosynthesis of **1g** was explored.

3.3.2 P450-meidated formation of bulnesene-15-COOH

The precursor of **1g**, bulnesene-15-al **1b**, was isolated from the cascade reaction of R19/F87I and CRED-231. The purified **1b** was screened with a library of 45 CYP102A1 variants (1 mM substrate, 2 μM enzyme) with a longer reaction time of 16 hours to increase the conversion. The library was constructed with mutants that had combinations of different base variants and mutations to identify which mutations are important in bulnesene-15-al oxidation. Full activity and selectivity data are given in Appendix 2.

All variants showed relatively high conversion (<50%) with one major product which was isolated and characterised as bulnesene-15-COOH **1g**. Only variants R19/F87A/V78I/A328I, K19/F87V/E267V and GVQ/I263G/A328L gave a new metabolite at a longer retention time which was not characterised as the target compound was the C15-acid. Variant R19/F87I also showed relatively high conversion (70%) and selectivity (71%) for the oxidation of bulnesene-15-al to the C15-acid **1g**, which indicated that a longer reaction time could enable this mutant to oxidise bulnesene directly to **1g**. A time-course reaction was performed to study the biosynthesis of **1g** with the R19/F87I variant (Fig. 3.8).

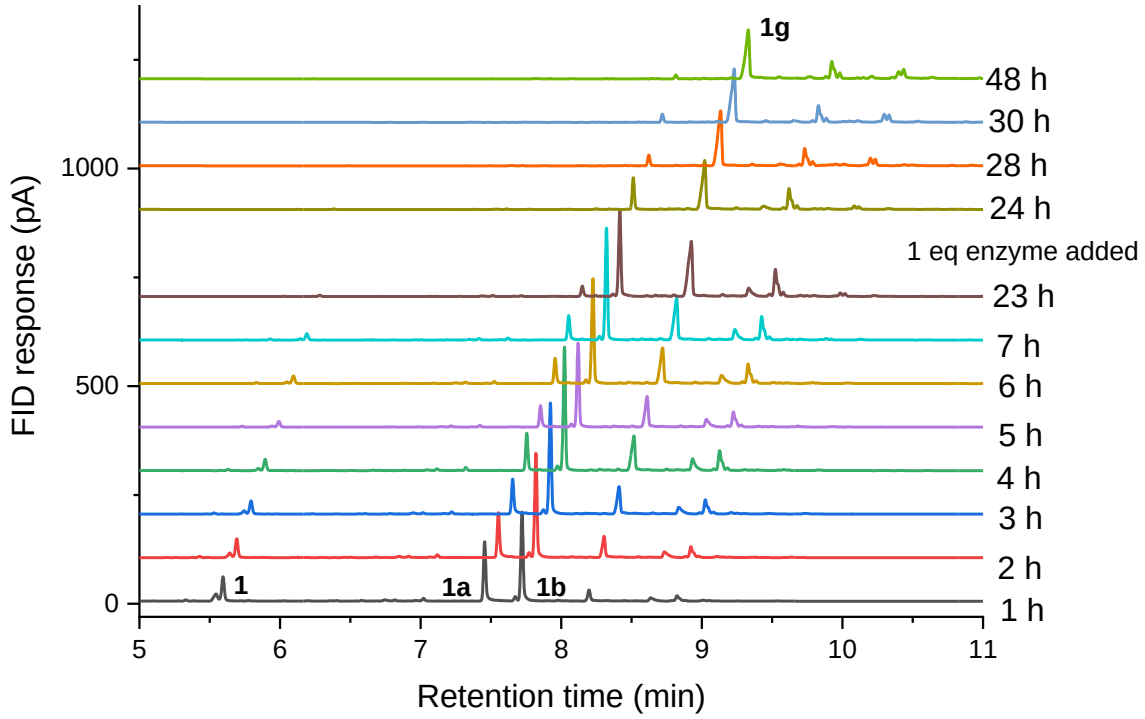


Figure 3.8 Time-course reaction of R19/F87I with bulnesene **1**.

Bulnesene **1** was almost fully converted in one hour to two major products, bulnesene-15-ol **1a** (31%) and bulnesene-15-al **1b** (47%) with C15-acid **1g** as the minor product at 7% selectivity. Bulnesene-15-ol **1a** was then further oxidised until it was almost fully converted to bulnesene-15-al **1b** (36%) and bulnesene-15-COOH **1g** (40%) at 23 hours. Another aliquot of the R19/F87I variant was added to further observe the reaction profile. The aldehyde was fully converted at 48 hours to C15-acid **1g**

as the major product (56%). A new product was formed at longer retention time but not isolated due to the low selectivity (15%).

Table 3.4 Substrate conversion and product selectivity of variants that give bulnesene-15-COOH, **1g**.

Entry	CYP102A1 variants	Conv.	1a	1b	1c	1d	1e	1f	1g
1	R19/F87I	100%	15%	51%	–	–	6%	–	**15%**
2	RP/F87V	96%	4%	25%	2%	–	17%	–	**12%**
3	GV/A184I	100%	–	–	3%	–	15%	–	**13%**
4	R19/F87A/S72A/A328I	96%	20%	12%	–	19%	9%	–	**10%**
5	RP/V78I/E267V	97%	7%	34%	2%	–	19%	5%	**9%**
6	K19/F87V	100%	5%	20%	5%	–	20%	–	**7%**

Screening scale reactions were in 200 mM phosphate buffer containing 2 μM CYP102A1, 2 mM α-bulnesene, 40 μM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 6 hours.

By cross-checking with the screening result, six variants were found to give bulnesene-15-COOH **1g** as a minor product, as shown in Table 3.4. Besides RP/V78I/E267V, all variants had mutation at F87 to either alanine (A), isoleucine (I) or valine (V). However, the selectivity was low (<25%), including R19/F87I that gave the most **1g** (15%, Table 3.4, Entry 1) at 1000:1 substrate:enzyme ratio in six hours. Variant R19/F87I could be used to synthesis **1g** with higher enzyme concentration (overall 4 μM, 500:1) and longer reaction time (48 h) as demonstrated in Figure 3.8 but the selectivity for **1g** was low (56%). Therefore, mutagenesis based on the R19/F87I variant was explored to generate a more active and selective variant for the formation of **1g**.

3.4 Further mutagenesis for selective C15-oxidation of bulnesene

3.4.1 Progressive library guided mutagenesis

Further mutagenesis was carried out based on the selectivity trend from the progressive screening library to increase the variant activity and selectivity so bulnesene **1** could be oxidised to bulnesene-15-COOH **1g** in less time and with higher selectivity. Mutants that give high total C15 selectivity (>60%) were selected as templates for the second-generation library.

Variants R19/A82M/T260G and R19/A184I/T260G showed very little selectivity for **1b** even though both were highly selective for **1a** (>60%) and had high total C15 selectivity (Table 3.5, Entries 3 and 4) which demonstrated that the A82M, A184I and T260G mutations were important in C15-alcohol production. Variant K19/F87I showed 70% total C15 selectivity with significantly lower conversion compared to R19/F87I (100%, Entries 1 and 2) which demonstrated the importance of mutations R47L/Y51F in promoting bulnesene oxidation.

Bulnesene-15-al (**1b**) and total C15 selectivity of variant R19/F87I were the highest among all variants, and with total conversion (Entry 1). Therefore, R19/F87I was the most active variant for C15 oxidation which made it a good starting point for mutagenesis. Residues A82, A184 and T260 were substituted with hydrophobic residues of different sizes. R19 was also used as the template for mutagenesis to study the effect of mutation at each residue without the presence of mutation F87I.

All variants were successfully expressed. Although the combination of A184I and/or A82M with T260G gave high selectivity for C15 oxidation in variants R19/A82M/T260G and R19/A184I/T260G (Entries 3 and 4), the C15 selectivity dropped significantly when the mutations were introduced individually to the R19 variant. Variant R19/A82M promoted oxidation at the C2 and C15 positions with high conversion while R19/T260G showed almost no conversion of bulnesene (Entries 7 and 8).

The addition of A82M, A184I and T260G to variant R19/F87I increased C15 oxidation drastically compared to their corresponding variants with no F87I mutation. Both R19/F87I/A82M and R19/F87I/A184I had almost fully converted bulnesene **1** to bulnesene-15-COOH, **1g**, with a total C15 selectivity of 77% and 75%, respectively, in 6 hours (Entries 9 and 11). This is a significant improvement compared to the precursor R19/F87I which showed much lower production of **1g** in 6 hours (Entry 1).

Table 3.5 Substrate conversion and product selectivity of selected variants for the oxidation of α-bulnesene, **1**.

Entry	CYP102A1 variants	Conv.	1a	1b	1c	1d	1e	1f	1g	C15	TON
1	*R19/F87I	100%	15%	51%	–	–	6%	–	15%	81%	810
2	*K19/F87I	79%	52%	18%	4%	–	6%	–	–	70%	555
3	*R19/A82M/T260G	69%	66%	–	6%	–	–	–	–	66%	455
4	*R19/A184I/T260G	80%	63%	–	3%	7%	7%	–	–	63%	505
5	*R19/A82M/A184I/T260G	93%	54%	14%	–	–	6%	–	–	68%	635
6	*R19	55%	19%	–	21%	28%	–	–	–	19%	105
7	R19/A82M	86%	42%	7%	31%	–	–	10%	–	49%	425
8	*R19/T260G	9%	–	–	–	–	–	–	–	–	_
9	R19/F87I/A82M	86%	–	7%	–	–	–	–	70%	77%	665
10	R19/F87I/A82I	86%	14%	32%	–	–	–	–	35%	81%	700
11	R19/F87I/A184I	91%	–	3%	–	–	–	–	72%	75%	685
12	R19/F87I/A184F	78%	34%	36%	–	–	–	–	9%	79%	620
13	R19/F87I/A184M	93%	8%	15%	–	–	–	–	47%	70%	655
14	R19/F87I/T260G	33%	72%	–	–	–	–	–	–	72%	240
15	R19/F87I/T260L	81%	44%	30%	–	–	–	–	4%	78%	635

C15 refers to the sum of selectivity for bulnesene-15-ol, **1a**, bulnesene-15-al, **1b** and bulnesene-15-COOH, **1g**. TON refers to the total yield of **1a**, **1b** and **1g** relative to enzyme. *: variants from previous library. Screening scale reactions were in 200 mM phosphate buffer containing 2 μM CYP102A1, 2 mM α-bulnesene, 40 μM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 6 hours.

Although variants with T260G mutation showed high selectivity for **1a** with little further oxidation to **1b** and **1g**, the conversion decreased drastically if not combined with the A82M or A184I mutation. For example, the conversion of variant R19/F87I/T260G was 33% while no activity was observed with its corresponding variant without the F87I mutation, R19/T260G (Entries 8 and 14).

Besides the mutations mentioned above, key residues were also selected for hydrophobic substitution with side chains of different sizes. Compared to methionine (M), substituting the A82 residue for a smaller residue like isoleucine (I) decreased the selectivity of acid **1g** from 70% to 35% (Entries 9 and 10). This could be because of the idea of "filling a hole" by using bulky substitution at A82 to block the active site that sequestered the aliphatic chain in fatty acid-bound structures of the WT.[44] The A82M mutation would lead to a smaller substrate pocket and thus enhance bulnesene oxidation. On the other hand, substituting A184I with bulkier residues like methionine (M) and phenylalanine (F) decreased the substrate pocket size which could be too small for a bicyclic substrate like bulnesene and thus reduced the variant activity (Entries 12 and 13).

Significant improvements were achieved by a small library with seven variants, developed first through progressive variants library screening to identify 'hit' residues followed by enzyme engineering based on the knowledge built up from previous screening. The new variants developed were highly active and catalysed three cycles of bulnesene oxidation selectively to form bulnesene-15-COOH **1g** in 6 hours.

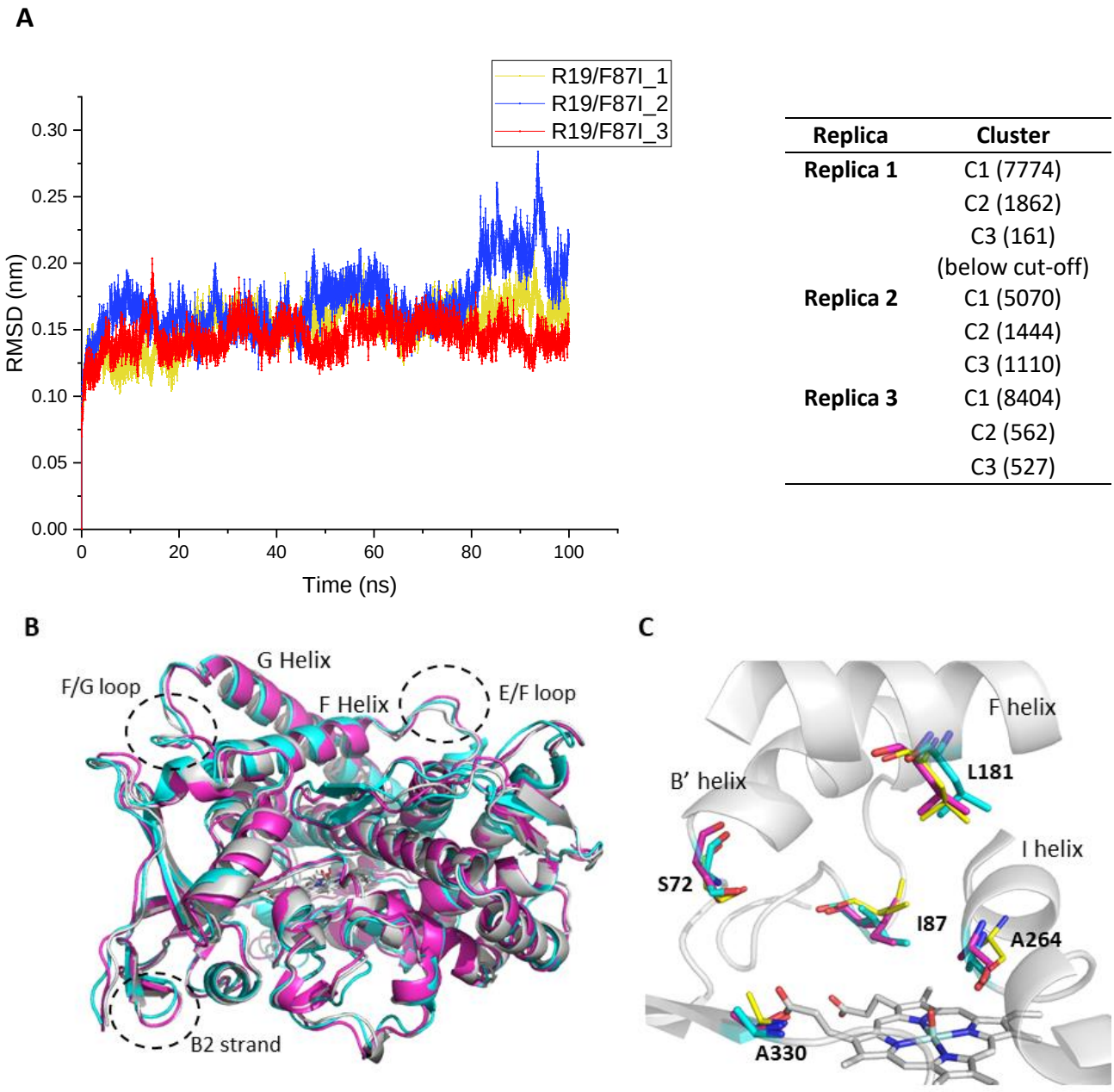
3.5 A third-generation library: docking-guided mutagenesis

Although mutagenesis based on progressive screening increased bulnesene oxidation activity and product selectivity, there is a limit to the degree of improvement because the residues explored were constrained to those available in the library. Another strategy was introduced to identify residues which had not been targeted.

MD simulations and computational substrate docking studies provide insight into substrate binding orientations and substrate–residue interactions which could guide mutation design. The background work leading to substrate docking was performed by Matthew Fisher. *In silico* mutagenesis on the structure of WT CYP102A1 with *N*-palmitoyl glycine (NPG) bound within its active site was performed in PyMol. The NPG molecule was removed and the heme prosthetic group was replaced with a ferryl species using a set of parameters from the literature.[160] This hypothetical ferryl form of the variant was then subjected to MD simulation by GROMACS for 100 ns. Structural snapshots were taken every 10 ps, leading to 10000 snapshots per simulation which was carried out in triplicate. For the R19/F87I variant, the RMSD of Cα atoms of all the recorded structures throughout the three 100 ns simulation runs were 1–2.5 Å (Fig. 3.9-A) which showed that, during each replica simulation, most of the record conformations only had small deviations from one another. The output for each replica was clustered with a RMSD cut-off of 1.2 Å. The top 3 clusters, with a population cut-off of 250 (2.5% of 10000 snapshots) typically accounted for >80% of all structures (Fig. 3.9-A). The structures were compared to ensure that no drastic and unreasonable changes had occurred (Fig. 3.9-B) before being used as receptors for docking studies by Autodock Vina (Section 3.5.1).[91]

The three most populated clusters from each replica were overlaid and compared for any deviation in structure. The structural variations were mostly found in loop regions between helices, e.g., the E/F and F/G loops, as well as the β2 strand. The helices were mainly well aligned, apart from the beginning of F helix and the end of G helix of cluster R2C1 (Figure 3.9-B). Most residues showed

good alignment except I87 where the rotamer conformations were observed for the side chains (Fig. 3.9-C).



Replica	Cluster
Replica 1	C1 (7774)
	C2 (1862)
	C3 (161)
	(below cut-off)
Replica 2	C1 (5070)
	C2 (1444)
	C3 (1110)
Replica 3	C1 (8404)
	C2 (562)
	C3 (527)

Figure 3.9. A) The RMSD plot and cluster population of each replica MD simulation of the R19/F87I variant. **B**) Overlay of the structure of the most populated clusters from 3 replica MD simulations of the structure of variant R19/F87I. Circles highlight the regions showing large deviations between replicas R1C1 (grey), R2C1 (cyan) and R3C1 (magenta). **C**) Overlay of selected residues and helices from R1C1 (cyan), R2C1 (magenta) and R3C1 (yellow) of MD-simulated structures of R19/F87I.

3.5.1 Docking analysis of bulnesene with R19/F87I variant

Bulnesene **1** was docked into the eight most populated clusters of the three replica simulations of R19/F87I by Autodock Vina which typically generated 9 poses each with different affinity energies. Among the 72 poses generated, 29 of them were non-productive (NP), which were generally too far away for oxidation. *i.e.*, all carbon centres were more than 4 Å from the ferryl oxygen (Table 3.6). There were 43 productive poses and 15 of them (35%) indicated C15 oxidation. Although C15 selectivity based on docking is considerably lower than the experimental result (81%), C15 oxidation was the dominant pose.

Table 3.6 Pose distribution with cluster populations of the docking of bulnesene **1** with variant R19/F87I by Autodock Vina.

		C2	C9	C11=C12	C15	C3	C4	C5	C6	C7	C8	C14	NP
Replica 1	C1 (7774)	2			1							1	5
	C2 (1862)			2	3							3	1
Replica 2	C1 (5070)			1	2								6
	C2 (1444)	3	2		1							2	1
	C3 (1110)	1	1	1	2		1				1	1	1
Replica 3	C1 (8404)				3	1							5
	C2 (562)	1			2	1						1	4
	C3 (527)		1	1	1								6
		7	4	5	15	2	1	0	0	0	1	8	29

The docking output demonstrated more NP poses than productive ones, while experimentally bulnesene was oxidised with 100% conversion by this variant. The presence of a large number of NP poses might indicate low activity as the substrate could be bound far away from the heme, reducing the probability of oxidation. On the other hand, if the NP poses were in similar locations to productive ones, such that these could readily adopt a productive pose, high activity would be possible. Another factor was the affinity energies of the poses. As shown in Figure 3.10, eleven NP poses had affinity energies higher than –7 kcal/mol whereas the affinity energies of most of the C15 poses were lower than –7 kcal/mol. It appeared that the location of the NP poses was the more important factor.

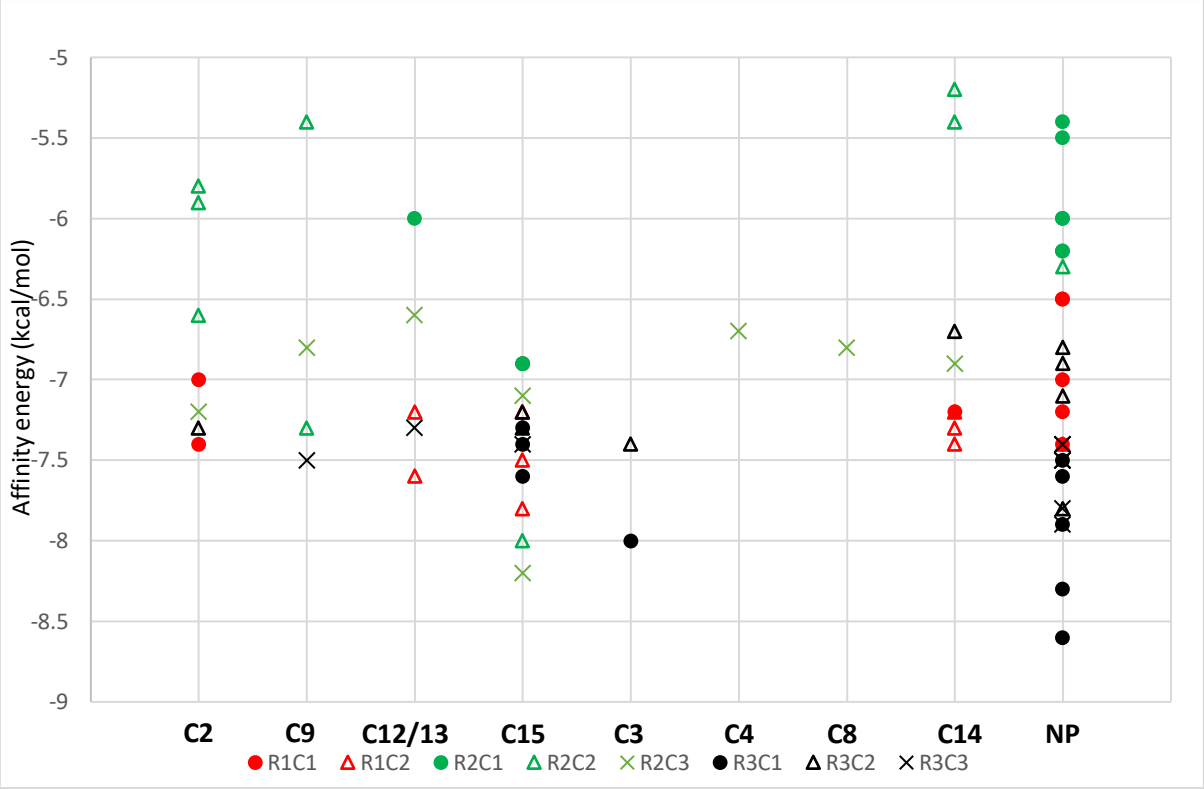


Figure 3.10 The affinity energies of output poses from docking of bulnesene **1** with variant R19/F87I.

3.5.1.1 The C15 poses

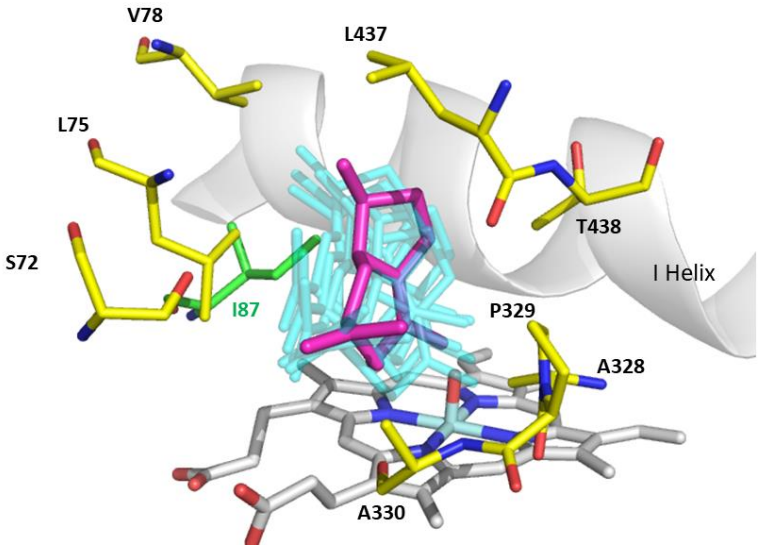


Figure 3.11 Overlays of simulation structure of R19/F87I (grey) with all C15 poses (cyan) from the most populated group and the lowest energy C15 pose among them (magenta). Residues that contacted the representative bulnesene substrate are in yellow while I87 is in green.

The C15 poses were divided into groups based on their conformations and the most populated group had eight out of the 15 poses generated (Fig. 3.11). All the poses from the most populated group, including the pose with the lowest affinity energy (–7.6 kcal/mol) in the group were overlaid with R3C1 to study the substrate binding interactions. As shown in Figure 3.11, the poses were mainly in van der Waals contact with residues in the β1 strand (A328, P329 and A330), the β4 strand (L437 and T438) and the B′ helix (S72, L75, V78 and I87). Residue I87 was 3.8 Å away from the lowest energy C15 pose which demonstrated the importance of the residue in holding bulnesene in the orientation which favoured C15 oxidation. This pose (from R3C1) was considered as a representative binding orientation for C15 oxidation of bulnesene.

3.5.1.2 The importance of the F87I mutation

Only variant R19/F87I out of the screening library gave relatively high selectivity for C15 oxidation of bulnesene. Variants R19/F87A, K19/F87V and R19 gave <25% C15 oxidation selectivity. The docking results demonstrated the importance of I87. The lowest energy C15 pose from the most populated cluster was overlaid with the simulated structures of the WT and variants R19 and R19/F87A to study the role of residue 87.

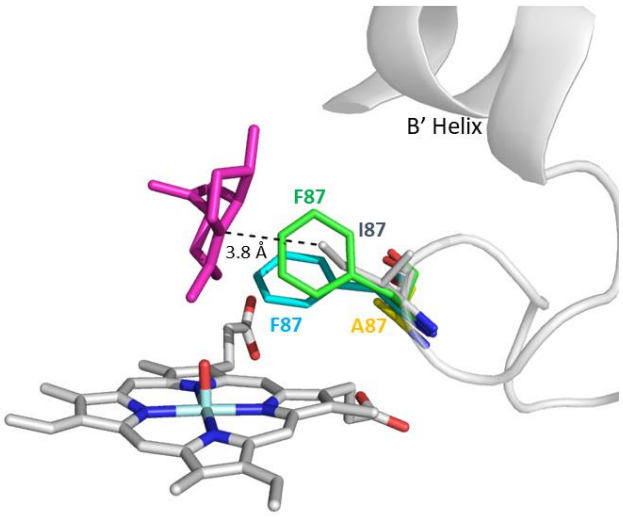


Figure 3.12 Overlays of the lowest energy C15 pose from the most populated cluster (magenta) with the MD simulation structures of R19/F87I (grey), WT (green), R19 (cyan) and R19/F87A (yellow) showing the role of the side chain of residue 87.

As shown in Figure 3.12, the phenyl side chain of F87 in the WT and R19 variant were located directly above the heme and would block bulnesene from binding close to the ferryl oxygen due to severe steric clashes with the C15 pose in the R19/F87I variant, which agreed with the experimental data that these enzymes showed low bulnesene conversion and very little or no C15 oxidation. The selectivity was shifted to C2 oxidation for variant R19. On the other hand, the F87A mutation expanded the substrate pocket which allowed many different orientations for bulnesene oxidation; for example, the A87 side chain is 6.2 Å from the C15 pose. Therefore, variant R19/F87A oxidised bulnesene at high conversion and gave a series of different products, each with low selectivity. The I87 side chain in the R19/F87I variant offered the appropriate contacts for the C15 binding pose.

To summarise, the C15 poses generated from the R19/F87I docking were clustered together. By overlaying the representative C15 pose with the simulation structures of different variants, we could further understand how different side chains on a single residue could have dramatic effects on product selectivity. Distances between residues and the C15 pose could serve as a useful guide in designing mutation that do not clash with the C15 pose to improve enzymatic activity or product selectivity without affecting the C15 selectivity.

3.5.1.3 The NP poses

The lowest energy C15 pose from the most populated cluster was overlaid with the NP poses (Fig. 3.13). The C15 pose was located near the centre of the NP cluster. Residues that were close to the NP poses (shown in yellow in Figure 3.13) but far away from the C15 pose could be targeted for mutagenesis to disfavour the NP poses.

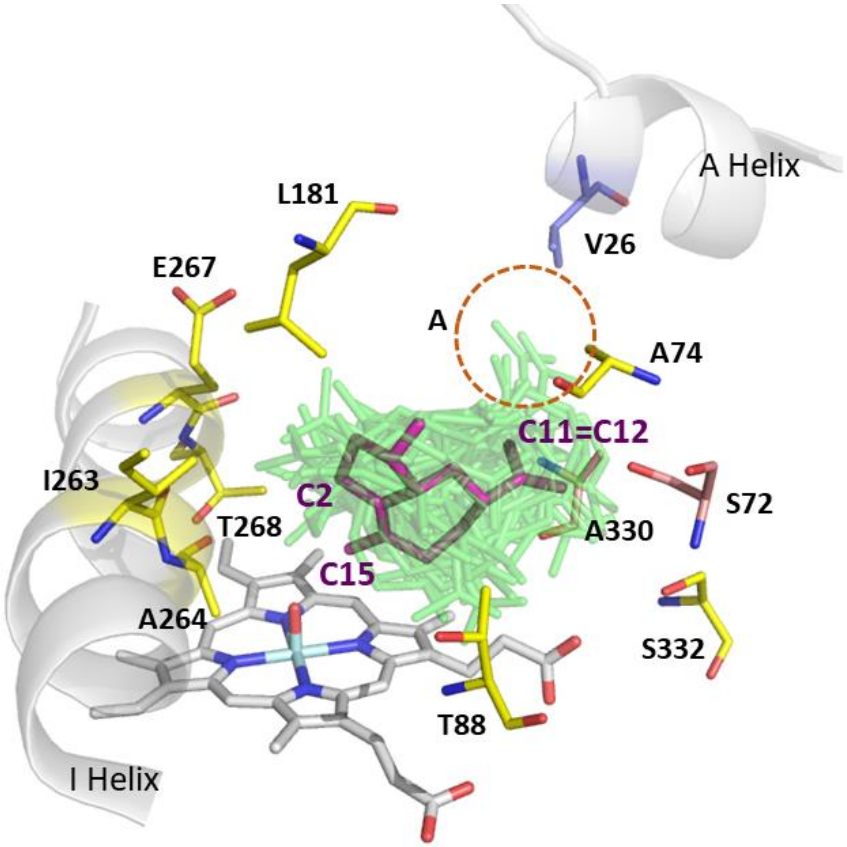


Figure 3.13 Overlays of simulation structure of R19/F87I (grey) with all NP poses (green) and the lowest energy C15 pose from the most populated cluster (magenta). Residues shown in yellow are close to NP poses while being far away from the C15 pose. Residues in salmon are close to both the NP poses and the C15 pose. Residue targeted to fill in region A is in purple.

Among these residues, T88 and S332 were especially interesting because neither had been mutated before. Most NP poses were bound in the region between the B′-helix residues A74 and T88, the β_1-strand residues A328 and A330 and I-helix residues I263 and A264 (Fig. 3.13). If these and other residues in this region were mutated to ones with a bulkier side chain, some NP poses would be disfavoured, and activity would be increased. Despite being far away from the NP poses (4.7 Å), residue V26 on the A helix was targeted for mutagenesis as it was pointing directly to a region (**A**) where some of the NP poses gathered (Fig. 3.13).

3.5.1.4 Other binding poses

The C2 and C11=C12 poses were significantly different from the C15 pose as they were located more towards the B′ helix (Fig. 3.14). Similar to the NP poses, residues A74 and T88 on the B′ helix were in contact with these poses while being far away from the C15 poses. The C11=C12 poses were ~4 Å from the T88 side chain, while the C15 pose was 6.9 Å away. Similarly, the C2 and C11=C12 poses were ~4 Å away from residue A74 which was 5.6 Å from the C15 pose. It could be hypothesised that, mutating residues A74 and T88 to residues with bulkier side chains could disfavour C2 and C11=C12 poses without affecting the C15 poses, potentially improving the C15 selectivity and activity for bulnesene oxidation.

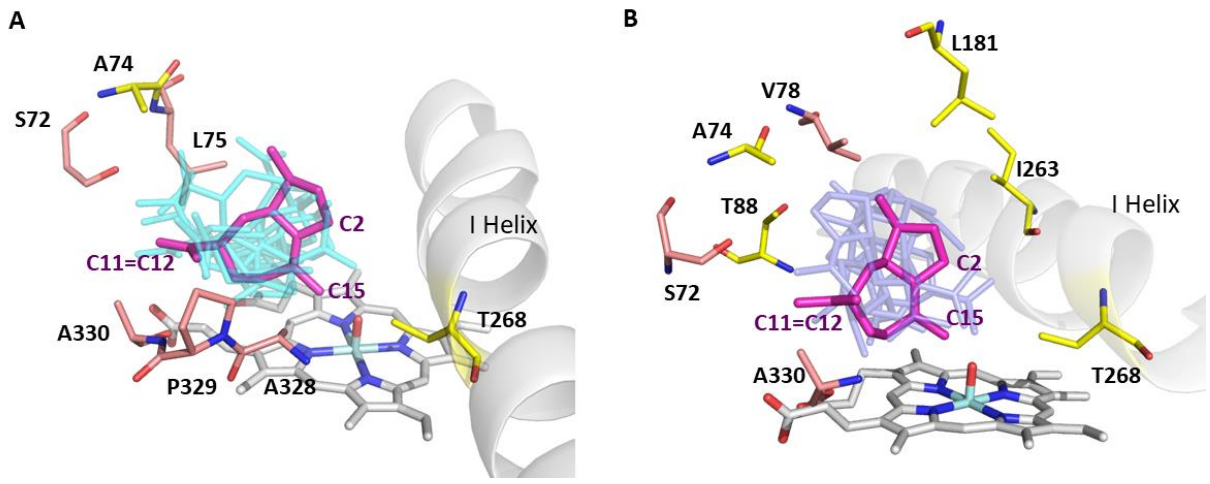


Figure 3.14. Overlays of simulation structure of R19/F87I (grey) with **A)** all C2 poses (cyan), and **B)** all C11=C12 poses (purple) and the lowest energy C15 pose from the most populated cluster (magenta). Residues shown in yellow are close to either C2 or C12 poses while being far away from the C15 pose. Residues in salmon are close to both the undesired poses (C2 and C11=C12) and the C15 pose.

Based on the analysis of the undesired poses, residues S72 and A330, which were ~3.7 Å away from the representative C15 pose, were found to be in contact with all the clusters of the undesired NP, C2 and C11=C12 poses. Overlaying each of the C15 poses with their respective clusters revealed that eight and ten out of fifteen C15 poses were >4 Å away from residues S72 and A330, respectively. Therefore, mutations at residues A330 and S72 were also proposed since, although some C15 poses might be sacrificed, these mutations could possibly disfavour more of the NP and side products poses.

3.5.1.5 Summary of docking analysis of bulnesene with variant R19/F87I

In summary, comparisons between different poses provided insight into the binding interactions and could be used to guide mutation design. A cluster of NP poses was located in the region around the I helix, B′ helix and β_1 strand. Significant differences were observed between the location and binding of C2 and C11=C12 poses compared to C15 poses. The residues which were in van der Waals contact to the undesired poses (<4 Å), but further away from the C15 poses (>4 Å), were selected for mutagenesis. Despite being far away from the centre of substrate pocket, residue V26 was also targeted for mutagenesis as it was pointing directly to a region where a cluster of NP poses gathered. Residues S72 and A330 were discovered to be in contact with all three undesired clusters of NP, C2 and C11=C12 poses. Although both residues S72 and A330 in R3C1 were ~3.7 Å away from the representative C15 pose, the residues were further away for more than half of the other C15 poses and thus were also targeted for mutagenesis. This analysis, using a relatively crude rigid-body docking routine, had enabled the design of new mutations that might have been overlooked if screening data only were used.

3.6 Docking guided mutagenesis study for bulnesene

3.6.1 Mutagenesis based on docking analysis

The 3rd generation variant library showed good conversion activity; only 4 out of 17 variants showed <40 % conversion which are variants with mutations at residues I263, A264 and S332.

Following on from the discovery of F87I as an important mutation in promoting C15 selectivity of bulnesene oxidation, mutation T88F increased C15 selectivity to 90% compared to 81% for variant R19/F87I and TON to 840 (Table 3.7, Entries 1 and 12). The residue A330 played an important role in the formation of bulnesene-15-COOH, **1g**. By substituting alanine (A) with valine (V), isoleucine (I) and phenylalanine (F), the selectivity for **1g** was increased (Entries 17–19). The highest was observed with R19/F87I/A330V that gave 30% of **1g** at 93% conversion (Entry 17). The A330I mutation, with a larger side chain, decreased both the conversion and C15 selectivity and the adverse effect was more drastic with the A330F mutation (Entries 18 and 19).

Although the selectivity for **1a** were increased slightly to roughly 70% by variants R19/F87I/S72V and R19/F87I/S332I, the conversions were lower at 60% and 32% respectively compared to R19/A82M/T260G (Entries 2, 7 and 21). Generally, variants which gave more of **1a** with low selectivity for **1b** had lower conversion (<70 %) as they were less active. On the other hand, variants that gave high conversion were more active and would further oxidise **1a** to C15-aldehyde **1b**, eventually to C15-acid **1g**.

Despite demonstrating slightly lower **1b** chemoselectivity compared to the variant R19/F87I, variants R19/F87I/V26M, R19/F87I/V26L and R19/F87I/A74L gave lower selectivity for **1g** while maintaining good **1b** selectivity which demonstrated that these mutations hindered the further oxidation of C15-aldehyde **1b** to C15-acid **1g** (Entries 5, 6 and 10).

Comparison of the docking results and the activity of the new variants showed that mutations that enhanced bulnesene oxidation were at residues that were not in close proximity of the C15 poses.

Most notable were mutations of residue T88 which increased C15 selectivity and TON, and of A74 and V26 that demonstrated good chemoselectivity for **1b**. These residues were at least 5 Å away from the C15 poses.

Table 3.7 Substrate conversion and product selectivity of selected variants for the oxidation of α-bulnesene, **1**.

Entry	CYP102A1 variants	Conv.	1a	1b	1c	1d	1e	1f	1g	C15	TON
1	*R19/F87I	100%	15%	51%	–	–	6%	–	15%	81%	810
2	*R19/A82M/T260G	69%	66%	–	6%	–	–	–	–	66%	455
3	*R19/F87I/A82M	86%	–	7%	–	–	–	–	70%	77%	665
4	*R19/F87I/A184I	91%	–	3%	–	–	–	–	72%	75%	685
5	R19/F87I/V26L	92%	16%	48%	–	–	–	–	10%	74%	685
6	R19/F87I/V26M	93%	20%	46%	–	–	–	–	10%	76%	710
7	R19/F87I/S72V	60%	70%	6%	9%	–	–	–	–	76%	460
8	R19/F87I/S72I	84%	53%	17%	10%	–	3%	–	–	70%	590
9	R19/F87I/S72L	76%	61%	13%	10%	4%	–	–	–	74%	565
10	R19/F87I/A74L	91%	31%	43%	–	–	–	–	5%	79%	720
11	R19/F87I/T88L	85%	55%	25%	5%	–	–	–	–	80%	680
12	R19/F87I/T88F	93%	65%	25%	–	–	–	–	–	90%	840
13	R19/F87I/T88M	96%	21%	31%	–	–	–	–	17%	69%	665
14	R19/F87I/L181F	84%	18%	30%	–	–	–	–	21%	69%	580
15	R19/F87/I263W	28%	36%	–	–	–	–	–	–	36%	105
16	R19/F87I/A264F	19%	59%	–	–	–	–	–	–	59%	115
17	R19/F87I/A330V	93%	10%	45%	–	–	–	–	30%	85%	795
18	R19/F87I/A330I	90%	11%	45%	–	–	–	–	26%	82%	740
19	R19/F87I/A330F	87%	12%	13%	–	–	–	–	24%	49%	430
20	R19/F87I/S332L	17%	58%	–	–	–	–	–	–	58%	100
21	R19/F87I/S332I	32%	71%	–	–	–	–	–	–	71%	230

C15 refers to the sum of selectivity for bulnesene-15-ol, **1a**, bulnesene-15-al, **1b** and bulnesene-15-COOH, **1g**. TON refers to the total yield of **1a**, **1b** and **1g** relative to enzyme. *: variants from previous library. Screening scale reactions were in 200 mM phosphate buffer containing 2 μM CYP102A1, 2 mM α-bulnesene, 40 μM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 6 hours.

One exception was the residue A330, which significantly enhanced activity, being 3.7 Å away from the representative C15 pose. The residue was first explored as it was frequently observed to be in contact with the undesired poses and further analysis showed that more than half of the C15 poses were >4 Å away from this residue. Similarly, the residue S72 which was targeted for the same reason, also increased the **1a** chemoselectivity significantly. Therefore, despite having to sacrifice some of

the C15 poses, mutations of residues S72 and A330 blocked more of the undesired poses than the C15 poses.

Mutations of I263 and A264 led to low conversion despite them being far away from the C15 poses. Unlike other residues which were tested with amino acids of various sizes, only mutations with large side chains like phenylalanine (F) and tryptophan (W) were introduced at both residues. The drastic difference in size could be the reason of the adverse effect observed with variants R19/F87I/I263W and R19/F87I/A264F. In addition, the residue I263 was only in contact with two NP poses of bulnesene which were from R3C2 and R3C3. As both clusters had very small populations (6% each), the poses generated were perhaps less representative.

In summary, docking of bulnesene with the MD simulation structure of the active R19/F87I variant provided useful insights into the substrate–enzyme interactions leading to C15 selectivity. Assessment of the location of poses leading to other products as well as non-productive (NP) poses enabled mutations to be designed that led to material improvements in bulnesene conversion and product selectivity. Residues that are further away from the desirable poses could be the major targets in mutagenesis, and residues which are in contact with NP poses and poses indicating experimentally observed products should be prioritised.

3.6.2 Fourth-generation variant library: combining mutations

In the second- and third-generation libraries, different mutations were introduced separately to the variant R19/F87I to study the effect of each on C15 oxidation of bulnesene. The mutations changed the shape of the substrate pocket in various ways, gradually tuning it for better fitting of bulnesene. The logical next step was to combine mutations that promoted activity with good C15 selectivity with mutations A184I and A82M which had been found to increase bulnesene oxidation activity from the second-generation library. The results are summarised in Figure 3.15.

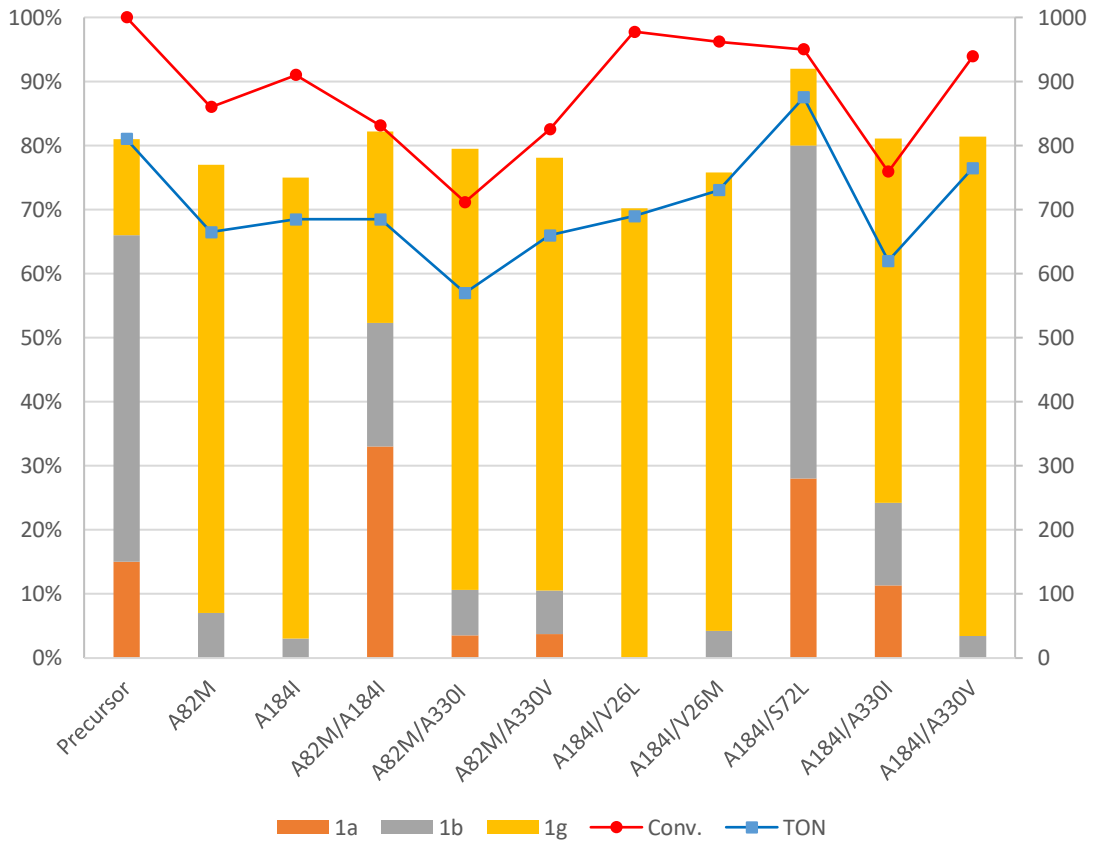


Figure 3.15 Effect of combining effective mutations at residues V26, S72, A82, A184 and A330 with R19/F87I.

Despite having the highest **1g** selectivity among the single-mutation additions to the R19/F87I variant, combining mutations A184I and A82M decreased the selectivity for **1g** (30%) with slightly higher total C15 selectivity (82%). Combining mutations A330I/V with R19/F87I/A82M gave a similar trend compared to the precursor R19/F87I/A82M at a slightly lower conversion. Variant R19/F87I/A184I/A330V increased the selectivity of **1g** to 78% with improved TON (765) compared to R19/F87I/A184I (685), the highest from the previous library. On the other hand, just by substituting A330 with isoleucine (I), the TON dropped significantly (620); the selectivity of **1g** was only 57% and the conversion was lower (76%).

As R19/F87I/A184I was the most active variant with just one mutation added to the R19/F87I variant, the mutations that showed high selectivity for C15-alcohol **1a** and C15-aldehyde **1b** were added separately to the variant R19/F87I/A184I to increase the selectivity of each product. Mutation S72L,

which showed high selectivity for C15-alcohol **1a** on the R19/F87I base variant, (61%, Table 3.7, Entry 9) increased the selectivity of C15-aldehyde **1b** to 52% when being combined with variant R19/F87I/A184I. The new variant, R19/F87I/A184I/S72L also improved the TON (875) and the overall C15 selectivity (92%) which were the highest among all variants screened. When the V26L/M mutations that promoted aldehyde formation were introduced, the TON was increased for variant R19/F87I/A184I/V26M (730) with C15-acid **1g** as the major product (72%).

In general, combining mutations which promoted high chemoselectivity (V26 and S72) for **1a** and **1b** with mutation A184I did not increase the chemoselectivity for the products. The combination of these high activity mutations, A82M, A184I and A330I/V increased the activity further and changed the chemoselectivity towards C15-acid formation. Hence, high activity and high **1g** selectivity have been achieved by combining effective mutations. However, high chemoselectivity for **1a** and **1b** remained a challenge which could be improved by protein engineering or reaction engineering. The effective mutations were combined again to generate triple-mutation additions to variant R19/F87I to refine the substrate pocket further.

For all triple-mutation variants, a new product was observed at a longer retention time. This was isolated and characterised by NMR and MS data. The loss of the C15 methyl resonance and shifting of the C12 olefinic protons in bulnesene from 4.70 ppm to around 2.60 ppm in the ^{1}H NMR spectrum suggested an epoxide with oxidation at C15. A peak at 251 m/z corresponding to $(M + 46 + H)^{+}$ indicated the addition of three oxygens to the substrate (RMM 204 g mol^{-1}) and the loss of two hydrogens. Hence, the product **1h** was characterised as 11,12-epoxy-bulnesene-15-COOH – the further oxidation product of bulnesene-15-COOH **1g**. Triple-mutation variants with mutation at V26 gave less **1h** (<20%) compared to the R19/F87I/A82M/A184I/A330I variant which gave 49%.

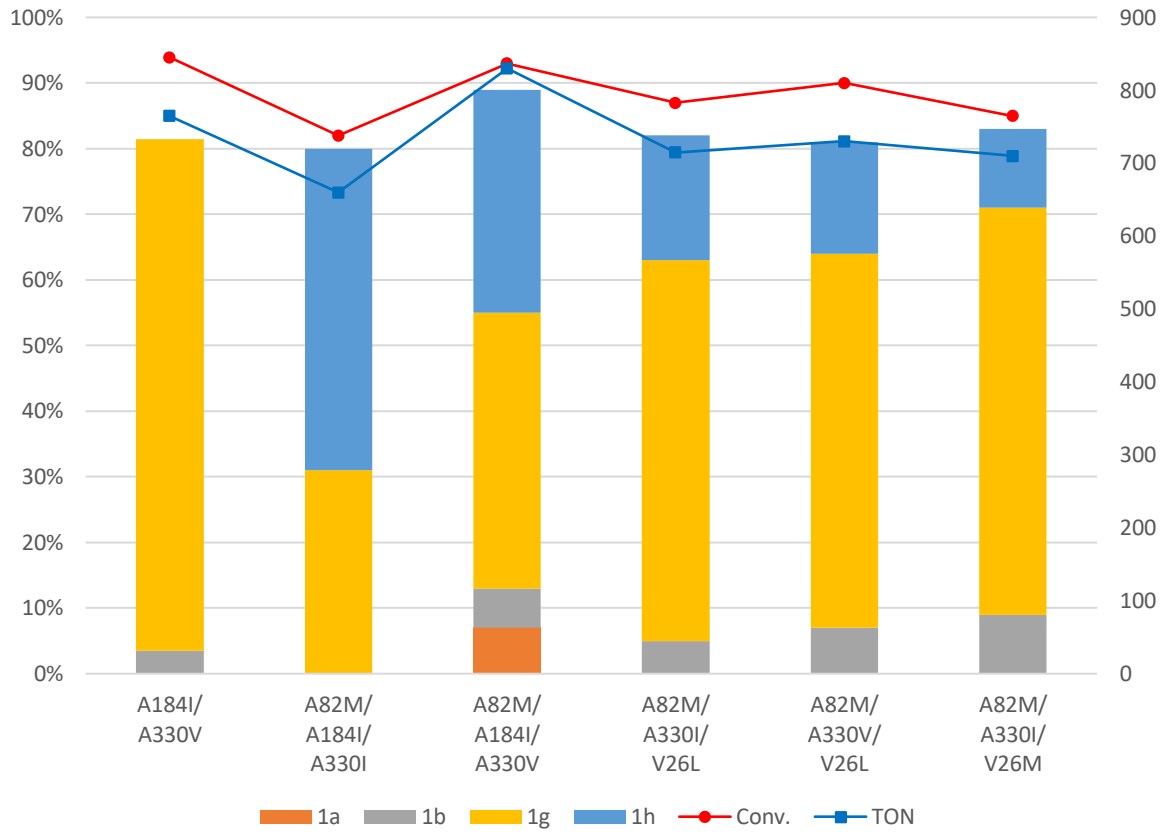


Figure 3.16 Effect of triple-mutation additions to the R19/F87I variant.

All the new variants showed >80% C15 selectivity; the highest was the 89% for the R19/F87I/A82M/A184I/A330V variant which also had the highest TON (830) among all the triple-mutation variants tested. If the chemoselectivity could be controlled while maintaining the high conversion and high activity, these engineered CYP102A1 variants would be highly efficient for the synthesis of each valuable agarwood aroma compound.

3.7 Summary

The biosynthesis of the precious wood aroma compounds bulnesene-15-ol (**1a**), bulnesen-15-al (**1b**) and bulnesene-15-COOH (**1g**) have been demonstrated. In addition, five other metabolites have also been characterised. All oxidation products are of interest because of their potential flavour, fragrance and biological properties.

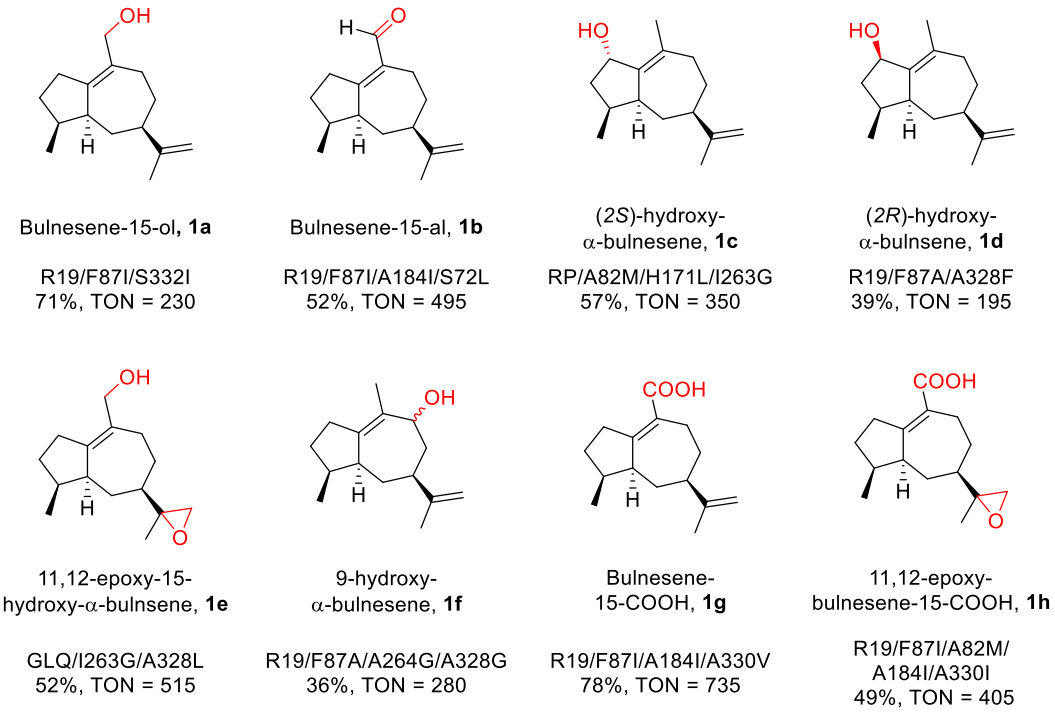


Figure 3.17 Highest selectivity and TON achieved for all oxidation products of α-bulnesene and relevant variants developed by progressive library screening and docking guided mutagenesis.

Instead of screening all the variants in the full library of more than 500 variants, the progressive library screening method provided an alternative route for identifying high-performing variants using a subset of less than 80 variants which also revealed residues that play important roles in bulnesene oxidation. MD simulations and substrate docking have provided insight into substrate–enzyme interactions which were utilised to guide mutation design, in particular the introduction of mutations at new residues which had not been targeted in the existing library. For example, T88 which was found to give high C15 selectivity, and V26 that increased activity.

The combination of the two different approaches has increased the activity substantially; the new-generation variants were highly active and could catalyse three cycles of bulnesene oxidation at the C15 position with high selectivity to form bulnesene-15-COOH which was only observed in very small quantities in the initial screening. The synthesis of the C15-carboxylic acid, which required a multi-step synthesis, could now be accomplished with a single variant in a 6-hour reaction.

Combining the effective mutations had the effect of tuning the substrate pocket for better substrate fitting. All the triple-mutation variants were highly active and had good C15 selectivity which make them good candidates for the biosynthesis of the valuable aroma compounds in agarwood. However, chemoselectivity control remained a challenge for the selective synthesis of the C15-alcohol **1a** and C15-aldehyde **1b**.

Chapter 4

4 Bulnesene-15,2-lactone synthesis by CYP102A1

4.1 Introduction

Sesquiterpene lactones (STLs) are a subclass of sesquiterpenoids which contain a lactone ring. They are the characteristic compounds of plants from the family *Asteraceae* and also found in other angiosperm families.[169,170] There are four main classes of STLs: the eudesmanolides, guaianolides, germacranolides and pseudoguaianolides.[169] STLs are known for their varied biological activities, such as antioxidant, anticancer, antitumour and anti-inflammatory activities and thus they are important lead compounds in the development of many therapeutic agents.[170–172] One of the most notable examples is artemisinin, a sesquiterpene lactone isolated from *Artemisia annua* that is active against *Plasmodium falciparum* (malaria). Artemisinin and its derivatives are a group of drugs widely used for the treatment of malaria.

(–)-Guaia-1(10),11-dien-15,2-olide, also known as bulnesene-15,2-lactone, is a sesquiterpene lactone derived from α-bulnesene and is a major contributor to the desirable and unique aroma of agarwood because of its long-lasting woody and ambery aroma with orris sub notes.[149] The biosynthesis pathway of this sesquiterpene lactone is yet to be identified but they are possibly the oxidation product of α-bulnesene and α-guaiene which can also be found in agarwood.[139,151] Although the compound could be extracted and isolated from agarwood, a more sustainable and efficient method is needed because it is only present in low amount in agarwood.

The chemical synthesis of 15,2-lactone has been reported by Ishihara and co-workers.[158] Starting from (–)-epoxybulnesene, the eight-step synthesis involved magnesium amide-catalysed selective oxirane ring opening as the key step. The final step was the selective α,β-unsaturated ketone reduction of (–)-methyl-2-oxoguaia-1(10),11-dien-15-oate through Luche reduction, followed by cyclisation under acidic condition to form the 15,2-lactone (Fig. 4.1). Surprisingly, one of the

intermediates, (–)-2α-hydroxyguaia-1-(10),11-dien-15-oic acid, which was thermally sensitive, was converted to the 15,2-lactone slowly in $CDCl_3$ over 2 weeks.

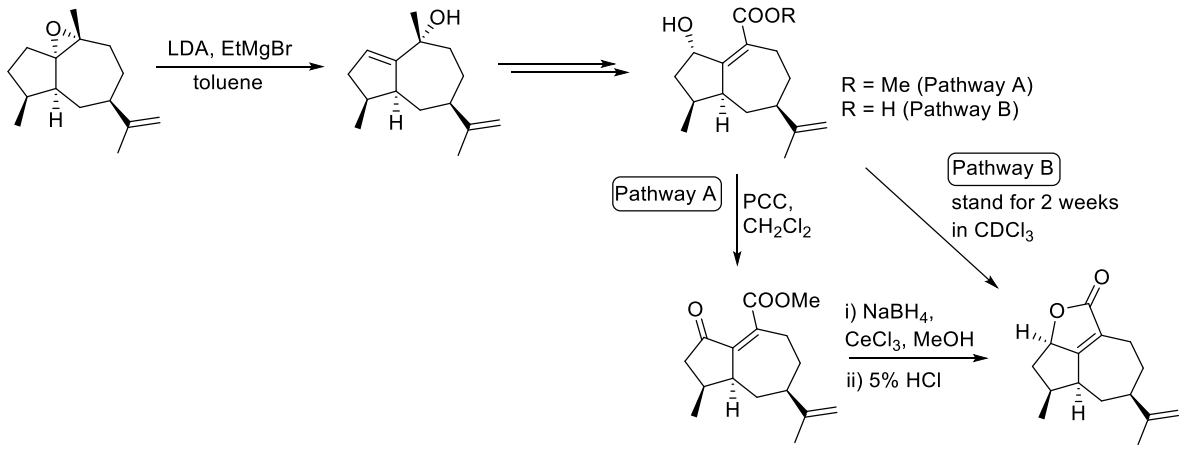


Figure 4.1 The multi-step chemical synthesis of bulnesene-15,2-lactone from (–)-epoxybulnesene.

This multi-step synthesis utilised various harsh chemicals and conditions and circumvented the challenging construction of the fused-ring system with multiple stereogenic centres by starting with (–)-epoxybulnesene from patchouli oil. Such convoluted pathways were necessary because of the difficulties in classical methods of direct C–H functionalisation to form (–)-2α-hydroxyguaia-1-(10),11-dien-15-oic acid from bulnesene. The subject of this chapter is to develop the biocatalytic synthesis of bulnesene-15,2-lactone via an enzymatic cascade reaction in which bulnesene is oxidised to the C15 carboxylic acid using a CYP102A1 variant, followed by another C–H bond oxidation by the same or a separate variant to form 2-hydroxy-bulnesene-15-COOH, the precursor of the 15,2-lactone. Bulnesene-15-COOH (**1g**) was synthesised by the oxidation of bulnesene catalysed by the variant R19/F87I/A330V/A184I, which gave acid **1g** with 78% selectivity and almost total conversion in 6 hours, and screened for oxidation with the CYP102A1 enzyme library.

4.2 Oxidation of bulnesene-15-COOH with $P450_{BM3}$

4.2.1 Initial and sub-library screening of bulnesene-15-COOH

A library of 47 CYP102A1 variants (Appendix 3) was screened for the oxidation of bulnesene-15-COOH (**1g**) (1 mM **1g**; 2 µM enzyme). This initial library was similar to that for bulnesene oxidation screening and was based on five template variants GVQ, K19, R19, RP and RT2.

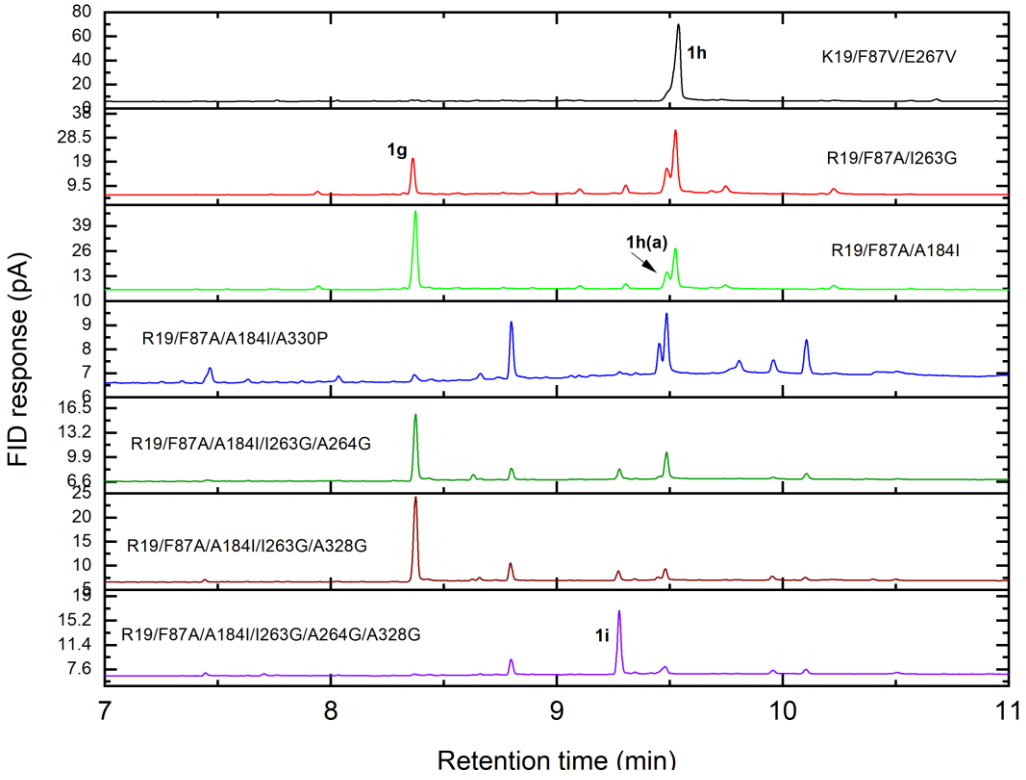


Figure 4.2 GC analysis of selected variants giving bulnesene-15-COOH oxidation products **1h** and **1h(a)** to **1i**.

GC analysis of the screening reactions showed that bulnesene-15-COOH **1g** was a poorer substrate for CYP102A1 compared to bulnesene. Only eight variants gave conversion higher than 40% after 16 hours. All the variants with reasonable conversion (>40%) gave 11,12-epoxy-bulnesene-15-COOH (**1h**) as the only product except for R19/F87A/A184I and R19/F87A/I263G (Table 4.1, Entries 1 & 2, Fig. 4.2). Variant R19/F87A/A184I gave a minor product which was isolated and characterised as the other diastereomer of 11,12-epoxy-bulnesene-15-COOH **1h(a)**. However, the

absolute configuration of the epoxide could not be determined by nOe because of the free-rotating nature of the C7–C11 bond. Variant R19/F87A/I263G gave four minor products besides the two 11,12-epoxide diastereomers; these minor products could not be isolated due to the low selectivity. Overall, the screening results provided useful information that mutations A184I and I263G could bias the regioselectivity of the oxidation of C15-acid **1g** away from the C11=C12 double bond.

Based on the initial screening results, three sub-libraries, based on the K19/F87A/A82M, R19/F87A and R19/F87A/A184I template variants, were assembled from the existing collection of CYP102A1 variants from our group. The selected variants had additional mutations within the I helix (T260, I263, A264) and the β_1 strand (A328, P329, A330, S332).

The R19/F87A panel (R47L/Y51F/F87A/H171L/Q307H/N319Y)

The majority of variants in the R19/F87A library had little to no activity for bulnesene-15-COOH, **1g**. Variants with reasonable conversion (>20%) gave 11,12-epoxy-bulnesene-15-COOH (**1h**) as the major product, similar to the results of the initial library. Variant R19/F87A/T260G/I263G/A264G/A328G gave many product peaks on the GC which indicated that the four glycine mutations had greatly enlarged the substrate pocket which promoted oxidation at multiple sites with little selectivity.

The K19/F87A/A82M panel (A82M/F87A/H171L/I263G/Q307H/N319Y)

Similar results were demonstrated by the K19/F87A/A82M library where most variants showed little activity and gave one major product, the 11,12-epoxide **1h**. Variants K19/F87A/A82M/A264G and K19/F87A/A82M/A264G/A328G had higher conversion (97% and 67%, Table 4.1, Entries 3 and 4) which indicated that mutation A264G and A328G could promote C15-acid oxidation.

The R19/F87A/A184I panel (R47L/Y51F/F87A/H171L/A184I/Q307H/N319Y)

The variants of the R19/F87A/A184I library showed higher conversion of bulnesene-15-COOH **1g** with a more diverse product profile. Two new major products were observed (Fig. 4.2). The first product was observed at 8.801 min with variant R19/F87A/A184I/A330P at 18% selectivity (Table 4.1, Entry 5, Fig. 4.2). However, it was not isolated due to the poor expression of this variant. Another new peak was observed at 9.271 min with variant R19/F87A/A184I/I263G/A264G/A328G giving 55% selectivity at high conversion (98%, Table 4.1, Entry 8, Fig. 4.2). Due to the low expression of this variant, R19/F87A/A184I/I263G/A264G was used instead for the preparative scale reaction, despite the lower conversion (53%) and selectivity (14%, Table 4.1, Entry 6).

The 100 mL preparative scale reaction contained 1 mM of **1g** and twice the concentration of variant R19/F87A/A184I/I263G/A264G (4 μM) compared to screening scale reactions (2 μM) to promote conversion. The product **1i** was isolated in very poor yield (1%) but the amount was sufficient for its characterisation by NMR and MS data as 2-hydroxy-bulnesene-15-COOH. The MS data of **1i** showed a peak at 273 m/z corresponding to $(M + 16 + Na)^+$ and the one-proton resonance at 4.84 ppm in the ^1H NMR demonstrated that an alcohol group has been added at the C2 position of **1g** (RMM 234 g mol^{-1}).[158] The poor yield could be the result of the low product selectivity and the poor extraction of the compound from the aqueous reaction mixture as multiple hydrogen bonds are likely to be formed between water molecules and the alcohol and carboxylic acid groups of **1i**. However, the discovery of 2-hydroxy-bulnsene-15-COOH was a big milestone for this project, as it is the precursor for the target compound, 15,2-lactone.

4.2.2 Biocatalytic cascade reaction for 15,2-lactone formation

Although purified C15-carboxylic acid **1g** was used as the starting material to produce 2-hydroxy-bulnesene-15-COOH **1i**, a more efficient method for its production was needed, especially for larger scale reactions. Biocatalytic cascade reactions have gained popularity because they are simpler, one-

pot reactions and the intermediate isolation steps could be avoided which could result in higher yield.[9,12,13] Therefore, a cascade reaction was designed by using a CYP102A1 enzyme (E1) to oxidise bulnesene to acid **1g** which would be oxidised to 2-hydroxy-bulnesene-15-COOH (**1i**) by another CYP102A1 enzyme (E2) *in situ* in a one-pot reaction. Any variant discussed in Chapter 3 that could oxidise bulnesene to acid **1g** with high activity and selectivity could be used as E1.

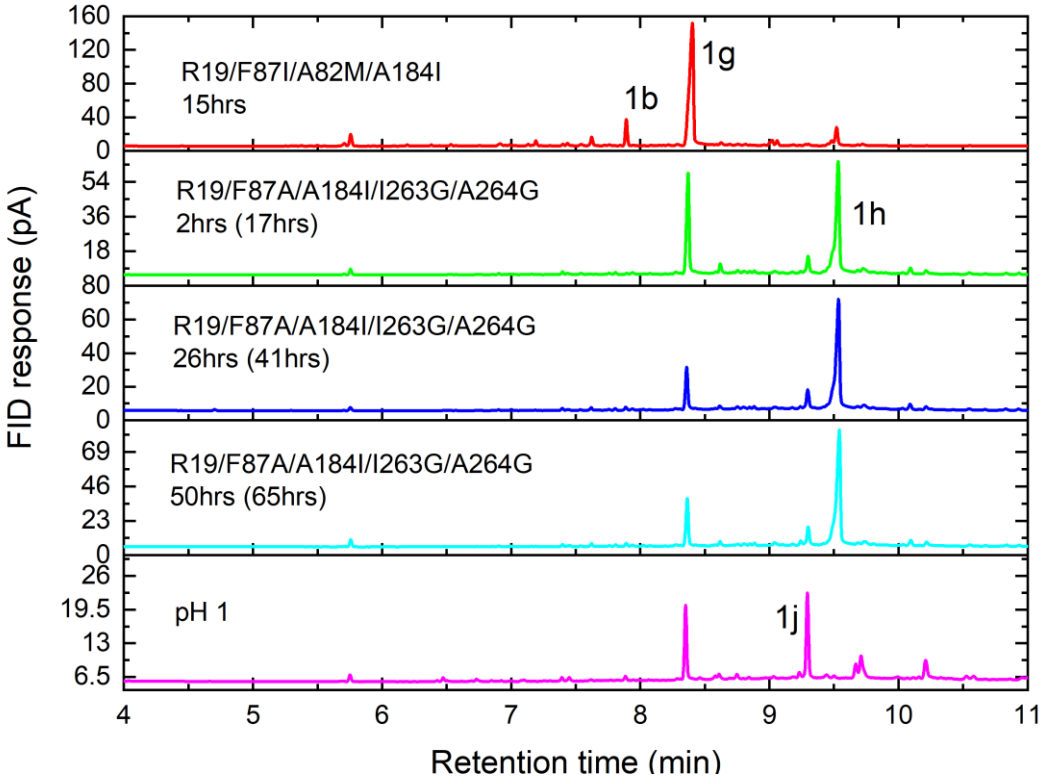


Figure 4.3 The time course of the CYP102A1 cascade reaction for the formation of 15,2-lactone **1j**. The times in brackets represented the take taken since the beginning of reaction. The 2-hydroxy-bulnesene-15-COOH **1i** was not observed as it was converted to the 15,2-lactone **1j** upon heating in the injector of the GC.

A 100 mL reaction was set up with bulnesene (2 mM) and variant R19/F87I/A82M/A184I (E1, 2 μM, 1000:1 ratio) and monitored by GC (Fig. 4.3). Variant R19/F87A/A184I/I263G/A264G (E2, 4 μM) was added as soon as bulnesene **1** had been fully converted to acid **1g** by E1 (15 hours). After 26 hours, approximately 84% of **1g** was converted to **1i** (9%) and the 11,12-epoxide **1h** (86%). No further conversion was observed with longer reaction time. The proportion of epoxide **1h** formed was much higher than anticipated for E2 activity from the screening results (41%, Table 4.1, Entry 6); probably because R19/F87I/A82M/A184I (E1) left in the reaction mixture was competing with

R19/F87A/A184I/I23G/A264G (E2) and oxidised **1g** to epoxide **1h**. As extraction of **1i** from aqueous buffer was poor, the reaction mixture was adjusted to pH 1 by 6 M HCl to assist in the formation of the neutral 15,2-lactone, which should be easier to be extracted. Surprisingly, adjusting the reaction mixture to lower pH decomposed most of the epoxides **1h** and **1h(a)**, while the acid **1g** and **1i** remained. No new peak was observed on the GC as might be expected from the formation of the 15,2-lactone. The large increase of the relative peak intensity of **1i** compared to **1g** suggested that acidic conditions had improved the extraction of **1i** from aqueous reaction.

The product was isolated by silica column chromatography. Surprisingly, the NMR of the sample was different from that of 2-hydroxy-bulnesene-15-COOH **1i**. The distinctive doublet of doublets of doublets at 4.92 ppm for the C2 proton and the loss of C15 methyl resonance were consistent with the literature data of the 15,2-lactone.[151] A peak at 255 m/z corresponding to $(M + 28 + Na)^+$ further demonstrated that 15,2-lactone **1j** was formed (1 mg, 2%) from bulnesene (RMM 204 g mol^{-1}).

The carboxylic acid **1g** formed by variant E1 was hydroxylated at C2 to form 2-hydroxy-bulnesene-15-COOH **1i**, which cyclised under acidic conditions to form lactone **1j**. The identical peaks on the GC for both compounds was likely because the acid-alcohol **1i** was cyclised to form lactone **1j** upon heating in the injector (240 °C) on the GC column and thus, the peak observed on the GC was due to the 15,2-lactone instead of 2-hydroxy-bulnesene-15-COOH. The increased relative peak area of **1j** after acidification demonstrated that the 15,2-lactone **1j**, which was much easier to be extracted compared to 2-hydroxy-bulnesene-15-COOH, was formed more efficiently after heating and acidification. However, the isolated yield of lactone **1j** was very low (2%) and therefore the biocatalytic cascade reaction needs to be optimised for more efficient synthesis of this product.

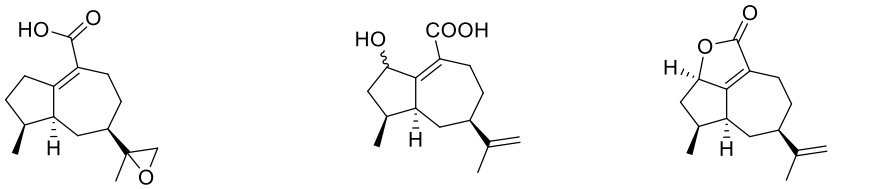


Figure 4.4 The oxidation products of bulnesene-15-COOH **1g**.

4.3 Optimisation of the biocatalytic cascade reaction

Three main issues were identified, with different solutions to improve the yield of the 15,2-lactone from the one-pot cascade reaction.

a) Better performing variants

   Although variant R19/F87A/A184I/I263G/A264G/A328G gave moderate selectivity for **1j** (55%, Table 4.1, Entry 8), the expression was low. A higher expression variant such as R19/F87A/A184I/I263G/A264G could be used but it had low selectivity for **1j** (14%, Table 4.1, Entry 6). A new variant with high C2 selectivity and expression was needed to enhance the efficiency of the cascade reaction.

b) Conditions for lactone cyclisation

   Since the peak intensity of **1j** increased after acidification, it was likely that only a small proportion of 2-hydroxy-bulnesene-15-COOH **1i** was extracted under normal conditions and this cyclised to form 15,2-lactone **1j** upon heating on GC analysis but most of the uncyclized **1i** remained in the aqueous phase. Once the reaction mixture was acidified, lactonisation of **1i** occurred and the more readily extracted 15,2-lactone **1j** increased the relative peak intensity of **1j**. The conditions for lactone cyclisation need to be optimised to maximise lactonisation and increase the yield of the 15,2-lactone.

c) Hinder the side products

The side product, 11,12-epoxy-bulnesene-15-COOH (**1h**), was produced in higher proportions in the cascade reaction (86%) compared to the screening results (41%, Table 4.1, Entry 6,). This could be because both E1 and E2 were competing for the oxidation of acid **1g** to form different products, which would result in the decrease in **1j** production. Therefore, the reaction needs to be optimised by reaction engineering to determine the optimal conditions for the biocatalytic cascade reaction.

4.3.1 Second-generation variants for lactone production

Protein engineering was used in an effort to develop better performing variants with improved activity, regioselectivity and expression for more efficient lactone synthesis. Variant R19/F87A/A184I/I263G/A264G/A328G was used as the template for further mutagenesis as it was the only variant which gave reasonable C2 oxidation selectivity of C15-acid **1g**.

Table 4.1 Substrate conversion and product selectivity of selected variants for the oxidation of bulnesene-15-COOH, **1g**.

Entry	CYP102A1 Variants	Conv.	1j	1h(a)	1h	1h(a)+1h	TON
1	R19/F87A/I263G	83%	6%	20%	47%	67%	25
2	R19/F87A/A184I	52%	5%	22%	52%	74%	15
3	K19/F87A/A82M/A264G	97%	–	–	95%	95%	–
4	K19/F87A/A82M/A264G/A328G	67%	–	–	86%	86%	–
5	R19/F87A/A184I/A330P	97%	–	9%	21%	30%	–
6	R19/F87A/A184I/I263G/A264G	53%	14%	1%	41%	42%	40
7	R19/F87A/A184I/I263G/A328G	42%	15%	5%	19%	24%	35
8	R19/F87A/A184I/I263G/A264G/A328G	98%	55%	–	9%	9%	270
9	R19/F87A/A184I/I263G/A330P	98%	19%	10%	26%	36%	95
10	R19/F87A/A184I/I263G/A264G/A330P	100%	61%	6%	14%	20%	305

TON refers to the yield of **1j** relative to enzyme. Screening scale reactions were in 200 mM phosphate buffer containing 2 μM CYP102A1, 1 mM bulnesene-15-COOH (**1g**), 40 μM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 16 hours. Residues in red represent new mutations introduced to the precursors.

Mutations at the I helix (I263 and A264) and the β_1 strand (A328 and A330) seemed to be important in changing the product selectivity. The oxidation site changed from solely C11=C12 epoxidation to C2 and other unidentified minor products with different combinations of mutations at residues I263,

A264, A328 and A330. Variants R19/F87A/A184I/A330P, R19/F87A/A184I/I263G/A328G and R19/F87A/A184I/I263G/A264G/A328G had the lowest selectivity (≤30%) for the 11,12-epoxides **1h(a)** and **1h** which demonstrated the importance of A330P and A328G in changing the substrate binding orientation in the substrate pocket (Table 4.1, Entries 5, 7 and 8). Mutation A330P also increased the conversion significantly; bulnesene-15-COOH (**1g**) was almost fully converted by variant R19/F87A/A184I/A330P in 16 hours (Entry 5). Therefore, new variants were generated by combining these important mutations.

Addition of mutation I263G to the high-conversion variant R19/F87A/A184I/A330P led to the formation of **1j** (19%) while maintaining the conversion at 98% (Entry 9), which again demonstrated the importance of the I263G mutation for C2 selectivity and the A330P mutation for conversion. Compared to R19/F87A/A184I/I263G/A264G/A328G, the selectivity for 15,2-lacone **1j** increased from 55% to 61% for the R19/F87A/A184I/I263G/A264G/A330P variant with full conversion (Entries 8 and 10). However, the expression levels of both variants were low and new variants with higher expression levels and selectivity for the lactone were needed.

Table 4.2 Substrate conversion and product selectivity of selected variants for the oxidation of bulnesene-15-COOH, **1g**.

Entry	CYP102A1 variants	Conv.	1j	1h(a)	1h	1h(a)+1h	TON
1	F87A/I263G	37%	–	30%	47%	77%	–
2	F87A/A330P	–	–	–	–	–	–
3	F87A/I263G/A264G	12%	–	14%	41%	55%	–
4	F87A/I263G/A330P	100%	5%	31%	25%	56%	25
5	F87A/A184I/A330P	76%	–	24%	39%	63%	–
6	F87A/A184I/I263G/A330P	100%	10%	17%	28%	45%	50
7	F87A/A184I/I263G/A264G/A328G	51%	51%	4%	11%	15%	130
8	F87A/A184I/I263G/A264G/A330P	100%	65%	6%	14%	19%	325

TON refers to yield of **1j** relative to enzyme. Screening scale reactions were in 200 mM phosphate buffer containing 2 μM CYP102A1, 1 mM bulnesene-15-COOH (**1g**), 40 μM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 16 hours.

At this point of the project, it seemed that further mutagenesis based on the best variants identified thus far would be challenging because of the poor expression of some of the enzymes. Therefore, instead of adding new mutations to the variants, the effect of each constituent mutation was studied

by a bottom-up approach. Mutations F87A, A184I, I263G, A264G, A328G and A330P were the major targets in this study.

Generally, variants with mutations A184I and A330P had higher conversion than their precursors. Whereas the F87A/I263G variant (Table 4.2, Entry 1) showed some conversion of acid **1g** to the epoxides **1h** and **1h(a)**, the F87A/A330P variant (Entry 2) was inactive towards **1g**. On the other hand, addition of the A184I mutation raised the conversion of the F87A/A330P variant to 76% (Entry 5) although epoxides **1h** and **1h(a)** remained the major products. The first appearance of the 15,2-lactone (5%) was found with the F87A/I263G/A330P variant which showed high conversion (100%, Entry 4). Addition of the A184I mutation raised the lactone selectivity to 10% for the F87A/A184I/I263G/A330P variant (Entry 6). Previous results had shown the importance of the combination A184I/I263G/A264G/A330P in biasing the product selectivity towards lactone **1j** (Table 4.1, Entry 10). When these mutations were combined, the variant F87A/A184I/I263G/A264G/A330P gave 65% of the 15,2-lactone with full conversion (Table 4.2, Entry 8). This variant had increased the lactone selectivity with much better expression compared to the most selective variant from the previous library, R19/F87A/A184I/I263G/A264G/A330P (Entry 10, Table 4.1). This was an interesting observation because the R19 combination of mutations (R47L/Y51F/H171L/Q307H/N319Y) had been identified as the key mutations in increasing the activity of CYP102A1 towards many unnatural substrates.[21,41] This could be because the polar carboxylic acid group required/favoured hydrogen bond formation with the polar side chains of R47 and Y51. Another important property of the F87A/A184I/I263G/A264G/A330P variant was its low activity for bulnesene oxidation, which made it an excellent candidate for E2 (to convert the acid **1g** to the lactone **1j** in the cascade.

4.3.2 Condition optimisation for lactone cyclisation

Inefficient lactonisation has been identified as one of the major obstacles to increase the yield of 15,2-lactone **1j**. Besides acidification, heating was also crucial because 15,2-lactone was formed from 2-hydroxy-bulnesene-15-COOH (**1i**) upon heating in GC analysis. In order to find the optimum conditions for lactone cyclisation, a preparative scale reaction was performed with R19/F87I/A82M/A330V as E1 and the new variant, F87A/A184I/I263G/A264G/A330P, as E2. Once bulnesene had been fully converted, several identical aliquots (400 µL) were removed and treated with different amounts of 2 M HCl at different temperatures for 1.5 hours. As 2-hydroxy-bulnesene-15-COOH was reported to decompose at high temperatures,[158] 70 °C was used as the starting point. As a result of the decomposition of **1h** and **1h(a)** upon acidification, it was impossible to calculate the selectivity for **1j**; the relative areas of **1j** in the GC analysis of each reaction were used as indicators of the amount of **1j** formed.

Table 4.3 Relative areas of the GC peak for 15,2-lactone **1j** formed after being treated with different amounts of 2 M HCl at various temperatures for 1.5 hours. * The relative area of an unidentified product.

Amount of 2 M HCl added (µL)	50	100	150
Final Concentration of HCl (M)	0.22	0.4	0.55
r.t.	50%	73%	72%
70 °C	62%	17% + 49%*	4% + 56%*

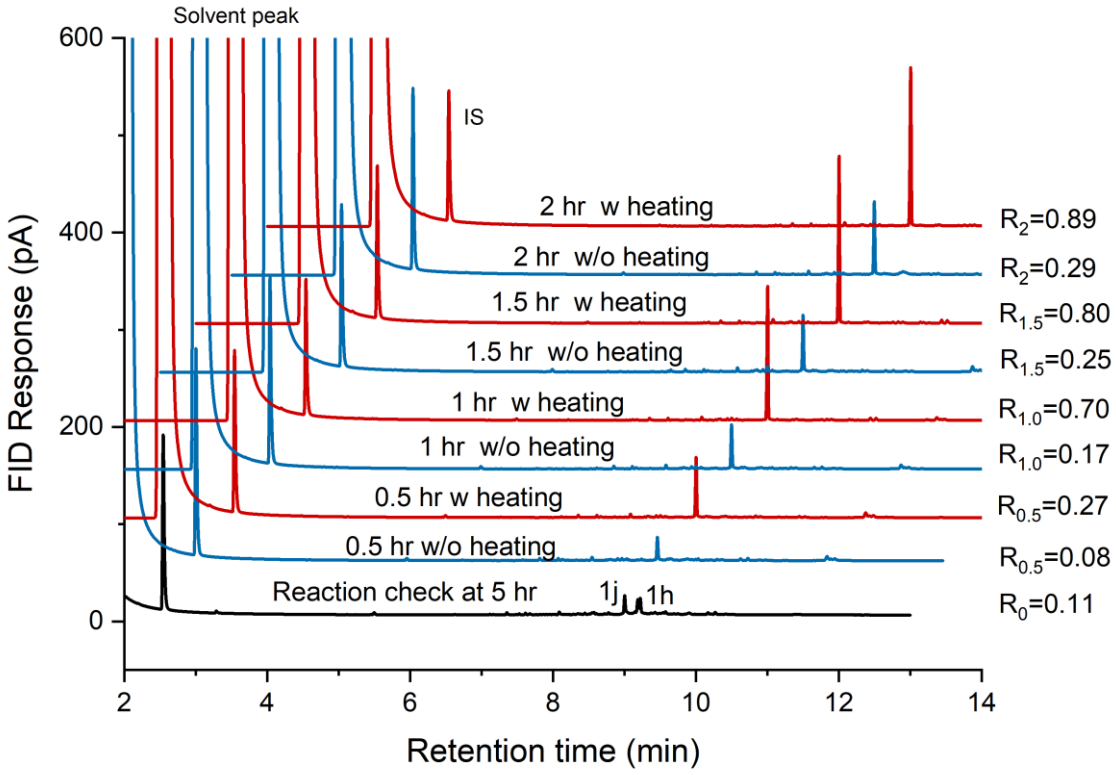
The cyclisation of 2-hydroxy-bulnesene-15-COOH **1i** was more efficient at higher temperature (70 °C) compared to room temperature when 50 µL of 2 M HCl was added in the reaction (final concentration of 0.22 M, Table 4.3). However, as the volume of 2 M HCl added increased (100 µL and 150 µL), **1j** decomposed into a new product which had a longer retention time. The highest amount of **1j** (73%) was observed with 100 µL of 2 M HCl (final concentration of HCl is 0.4 M and final pH is ~1) at room temperature. Although the relative areas of the GC peaks could serve as indicators of how much **1j** had been formed, the analysis might not be accurate because the extraction

efficiencies may vary. Therefore, an internal standard (IS) was used to quantify the amount of 15,2-lactone formed.

The optimum pH was determined as ~1 (0.4 M HCl) and further optimisation was designed by varying the temperature and reaction time at pH 1. As 15,2-lactone **1j** was found to decompose at 70 °C, a temperature between 70 °C and room temperature was tested (40 °C). The biocatalytic cascade reaction was repeated and adjusted to pH 1 as soon as the reaction had completed. Eight identical aliquots were removed and treated with various conditions (Fig. 4.5). The internal standard was added to each sample before acid and heat treatment to quantify the amount of **1j** formed.

Heating increased the relative amount of **1j** formed to IS (ratio, R, of the peaks are of **1j** to IS) by three- to four-fold while more **1j** was observed with longer acidification time especially after one hour, with or without heating. The ratio R increased from 0.11 to 0.89 after the reaction mixture was heated at pH 1 and 40 °C for 2 hours. Interestingly, the optimised cyclisation conditions had decomposed most of the side products, including **1h** and **1h(a)**, leaving **1j** as the only product in the reaction mixture, with a clean background (Fig. 4.5). This could be a highly beneficial feature to further development towards a viable process for the synthesis of lactone **1j** as purification would be much simpler.

In summary, the pH, temperature and reaction time were found to be crucial in enhancing the efficiency of lactonisation to form 15,2-lactone **1j**. Although the amount of **1j** formed was found to increase with longer reaction time and lower pH, most lactonisation occurred in the first hour of the reaction at pH 1 and 40 °C. Higher temperature (70 °C) led to the decomposition of **1j** to an unidentified product.



	R (r.t.)	R (40 °C)
0.5 hr	0.08	0.27
1 hr	0.17	0.70
1.5 hr	0.25	0.80
2 hr	0.29	0.89

Figure 4.5 The GC analysis and R value of each reaction aliquot after being treated with various conditions. Aliquots represented by red lines were treated at 40 °C while lines in blue showed aliquots without heating. Black line showed the reaction without being acidified and heated.

4.3.3 Minimising side product formation via reaction engineering

More side products were observed in the cascade reaction compared to the screening reaction with bulnesene-15-COOH **1g**. This is most likely because E1 was competing with E2 in oxidising **1g** to form epoxides **1h** and **1h(a)** which were the side products of the reaction. The new variant F87A/A184I/I263G/A264G/A330P, which gave higher conversion and selectivity for 15,2-lactone **1j**, could help in competing with E1 for **1g** oxidation. However, in order to minimise epoxide

formation, reaction engineering was introduced to optimise the condition. All the optimisation experiments were done at screening scale (0.5 mL in 24-well plates).

a) Timing for the addition of E2

Previously, enzyme E2 was added once bulnesene had been fully converted to the C15-acid. Although this approach could minimise the oxidation of bulnesene and the intermediates (bulnesene-15-ol, **1a**, and bulnesene-15-al, **1b**) by E2, it increased the chances of **1g** being oxidised by E1 to the side products **1h** and **1h(a)**. An appropriate timing for the addition of E2 was needed to find the balance between the two variables.

A set of experiments was designed by adding E2 at different times at intervals of one hour (Fig. 4.6). Each reaction was analysed after six hours by GC with and without being heated and acidified to monitor the synthesis of both **1j** and **1h**. The product profiles demonstrated an interesting trend; the later E2 was added, the more **1h** and the less **1j** were formed. The highest production of **1j** was observed when E2 was added together with E1 at the start of the reaction (T_0). This could be rationalised from a reaction with just R19/F87I/A82M/A330V where bulnesene was usually fully converted to **1g** at high selectivity first, before **1g** was further oxidised to **1h**. This means that bulnesene (**1**), the C15-alcohol (**1a**) and the C15-aldehyde (**1b**) were better substrates for the R19/F87I/A82M/A330V variant compared to C15-acid **1g**. On the other hand, the F87A/A184I/I263G/A264G/A330P variant was much less active towards **1**, **1a** and **1b** but highly active towards **1g**. Therefore, with both E1 and E2 in the reaction, E1 would be occupied by the oxidation of **1**, **1a** and **1b** to **1g** while E2 would oxidise **1g** to **1j** as soon as **1g** was formed in the reaction and thus less **1h** and **1h(a)** would be formed. Besides reaction engineering, variants with extremely high product selectivity and substrate specificity also played crucial role in this reaction to maintain the remarkable quadruple oxidation of bulnesene with increased efficiency.

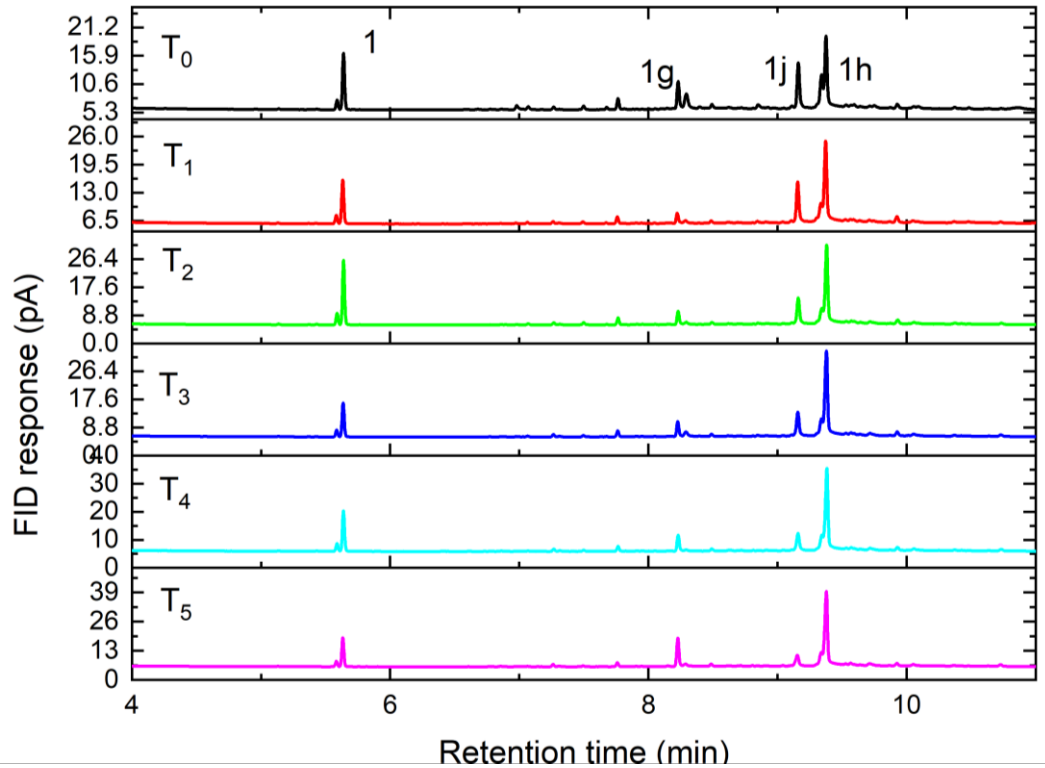


Figure 4.6 GC analysis of the cascade reaction of CYP102A1 for the synthesis of 15,2-lactone **1j**. E1 (R19/F87I/A82M/A330V) was added at the beginning of the reaction (T_0) while E2 (F87A/A184I/I263G/A264G/A330P) was added at different times with an hourly interval (T_0-T_5).

b) Choice of E1 and the E1:E2 ratio

All the variants that could oxidise bulnesene to bulnesene-15-COOH with high conversion and selectivity were tested with the C15-acid **1g** as substrate to identify which would oxidise **1g** to the epoxides **1h** and **1h(a)**. All the variants tested oxidised **1g** to the epoxides **1h** and **1h(a)** as the only products. R19/F87I/A82M was selected as potential E1 candidate as it was one of the least active variants for **1g** (20% conversion, Table 4.4, Entry 4). The most active variant, R19/F87I/A82M/A330V oxidised **1g** at 48% conversion and was also chosen as a potential E1 candidate to compare their efficiency in the cascade reaction (Entry 1). Both variants were used with different concentration of E2 to determine the optimum condition for the cascade reaction. Both sets of experiments were run in parallel. Once the reaction had gone to completion, the internal standard

was added followed by acidification (pH 1) and heating (40 °C) for one hour to promote 15,2-lactone formation. The organics were extracted with 300 µL ethyl acetate and analysed by GC.

Table 4.4 Substrate conversion and product selectivity of selected variants for the oxidation of bulnesene-15-COOH, **1g**.

Entry	CYP102A1 variants	Conv.	1h	1h(a)
1	R19/F87I/A82M/A330V	48%	3%	97%
2	R19/F87I/A184I/V26M	3%	–	100%
3	R19/F87I/A82M/A330I	18%	6%	94%
4	R19/F87I/A82M	20%	5%	95%
5	R19/F87I/A184I	9%	20%	80%

Screening scale reactions were in 200 mM phosphate buffer containing 2 µM CYP102A1, 1 mM bulnesene-15-COOH (**1g**), 40 µM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 6 hours.

The relative amount of lactone **1j** formed with R19/F87I/A82M as E1 was lower compared to R19/F87I/A82M/A330V (Table 4.5). This was because R19/F87I/A82M was less active comparatively and bulnesene was not fully converted to carboxylic acid **1g** in a 6-hour cascade reaction. Interestingly, as the concentration of E2 increased, bulnesene oxidation by E1 to **1g** became more efficient as fewer intermediate products like the alcohol **1a** and aldehyde **1b** were observed. This demonstrated the inhibiting nature of carboxylic acid **1g** and the presence of E2 helped in oxidising **1g** which in turn promoted bulnesene oxidation by E1.

The amount of lactone was expected to be higher as the concentration of E2 increased because more E2 was available for C15-carboxylic acid oxidation. This was true when a less active variant, R19/F87I/A82M was used as E1. However, when R19/F87I/A82M/A330V was used as E1, the relative amount of lactone formed decreased as the concentration of E2 increased once the relative amount of **1j** reached the highest at an E1:E2 ratio of 1:1.5. This could be because E2 was also oxidising bulnesene (**1**), bulnesene-15-ol (**1a**), and bulnesene-15-al (**1b**) to different side products when it was used in excess. Therefore, the ratio of E1 and E2 was found to be the best at 1:1.5 with R19/F87I/A82M/A330V as E1.

Table 4.5 The amount of each product formed under various conditions relative to the internal standard (IS) after being acidified and heated. Screening scale reactions were in 200 mM phosphate buffer containing 2 µM E1, 1 mM bulnesene-15-COOH (**1g**), 40 µM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 6 hours.

Entry	E1	E1:E2	1	1a	1b	1g	1j
1	R19/F87I/A82M	1:1	0.07	0.08	0.27	0.99	1.76
2		1:1.5	–	–	0.08	0.49	1.96
3		1:2	–	–	–	0.17	2.01
4		1:2.5	–	–	–	0.09	2.13
5		1:3	–	–	–	–	2.35
6	R19/F87I/A82M/A330V	1:1	0.1	–	0.11	1.24	2.24
7		1:1.5	0.07	–	–	0.27	2.42
8		1:2	0.05	–	–	–	2.31
9		1:2.5	–	–	–	–	2.31
10		1:3	0.06	–	–	–	2.09

4.3.4 Preparative scale reaction for lactone production

A preparative scale cascade reaction (200 mL) was set up with the optimised conditions. Both R19/F87I/A82M/A330V (E1) and F87A/A184I/I263G/A264G/A330P (E2) were added at the beginning of the reaction (T_0) at 1:1.5 ratio. Two identical aliquots were removed hourly to monitor the progress of the reaction by GC. The internal standard was added to both aliquots; one was treated with acid at 40 °C for an hour while the other was left untreated to monitor both the target compound **1j** and side product **1h**. The results are summarised in Figure 4.7.

About 90% of bulnesene was converted in eight hours and full conversion was reached after 24 hours. The intermediates, bulnesene-15-ol (**1a**) and bulnesene-15-al (**1b**), were observed in small amounts at the first hour of the reaction and were fully consumed in three hours. The relative amount of C15-acid **1g** reached a maximum at two hours which was the same time when the amount of **1a** and **1b** started to decrease significantly. An interesting trend was shown by the C15-acid **1g** where it was first synthesised in the first hour, reached a maximum at two hours, decreased and was fully consumed in 24 hours. The concentration of 15,2-lactone **1j** and 11,12-epoxides **1h/1h(a)** increased steadily throughout the reaction.

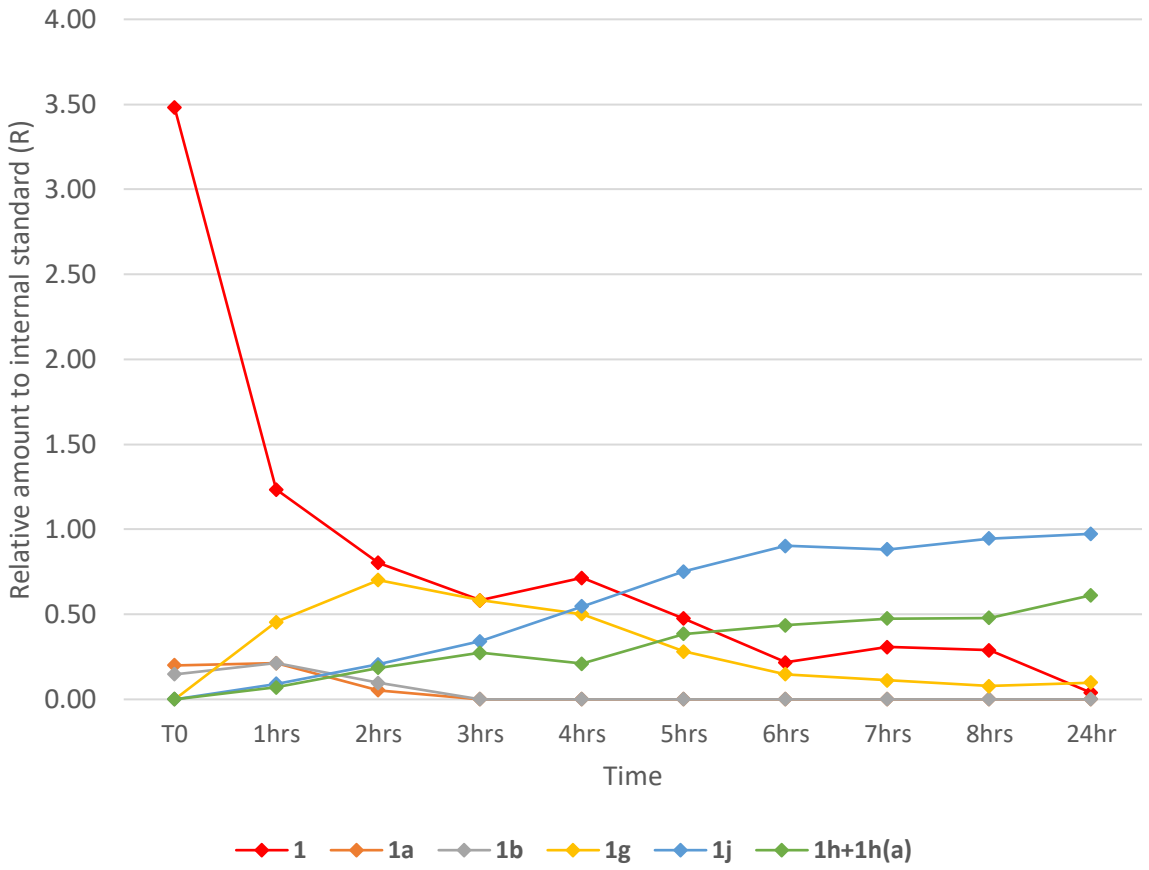


Figure 4.7 The time-course of the relative amount of each compound formed in a CYP102A1 cascade reaction.

Once the reaction was complete, the reaction mixture was adjusted to pH 1, heated at 40 °C for four hours and then extracted three times with an equal volume of ethyl acetate. The combined extracts were washed with water and brine, dried over $Na_2(SO_4)$ and the solvent was removed *in vacuo*. The 15,2-lactone **1j** was isolated with good purity as the side products had decomposed upon acidification. The yield was 38% (32 mg) based on the amount of bulnesene converted, similar to the result calculated based on GC analysis (28%).

4.4 Direct oxidation of bulnesene to bulnesene-15,2-lactone with a single CYP102A1 variant

Once the CYP102A1 cascade reaction for the biosynthesis of bulnesene-15,2-lactone (**1j**) had been established, the possible engineering of a single CYP102A1 variant for the quadruple oxidation of

bulnesene to the 15,2-lactone was explored. This was challenging because the variant not only had to possess high activity but also high selectivity for three cycles of bulnesene oxidation at C15 to form acid **1g** followed by oxidation of **1g** at C2 to give the precursor to the 15,2-lactone.

As the peak of lactone **1j** had been identified on the GC, the screening results of all the variants that had been studied were examined to see if there were any variants that gave this 15,2-lactone. Both the R19/F87I/A82M/A184I/A330I and R19/F87I/A82M/A184I/A330V variants oxidised bulnesene with high activity, firstly to the C15-acid **1g**, then mostly to the 11,12-epoxides **1h/1h(a)**; however, they also gave the 15,2-lactone **1j** as a minor product (≤5%). As both variants gave very similar conversion and product profile, R19/F87I/A82M/A184I/A330I which gave slightly higher **1j** selectivity (5%) was selected for mutagenesis with the aim of switching the selectivity from epoxidation of the C11=C12 double bond to C2 hydroxylation.

4.4.1 Docking analysis with the R19/F87I/A82M/A184I/A330I variant

The potential of docking-guided mutagenesis for increasing the activity of CYP102A1 for bulnesene oxidation was demonstrated in Chapter 3.6. Switching selectivity will be much more challenging compared to activity enhancement, especially for the synthesis of 15,2-lactone **1j** from bulnesene **1**. This is because the substrate pocket needs to be tuned to switch the binding orientation of C15-carboxylic acid **1g** that favoured C11=C12 double bond epoxidation to C2 oxidation without interfering with the C15 oxidation of bulnesene.

4.4.1.1 MD simulations for variant R19/F87I/A82M/A184I/A330I

The variant R19/F87I/A82M/A184I/A330I was chosen for docking analysis for bulnesene-15-COOH (**1g**) as it was one of the very few variants that gave 15,2-lactone **1j** directly from bulnesene. Three replica MD simulations were carried out for this variant.

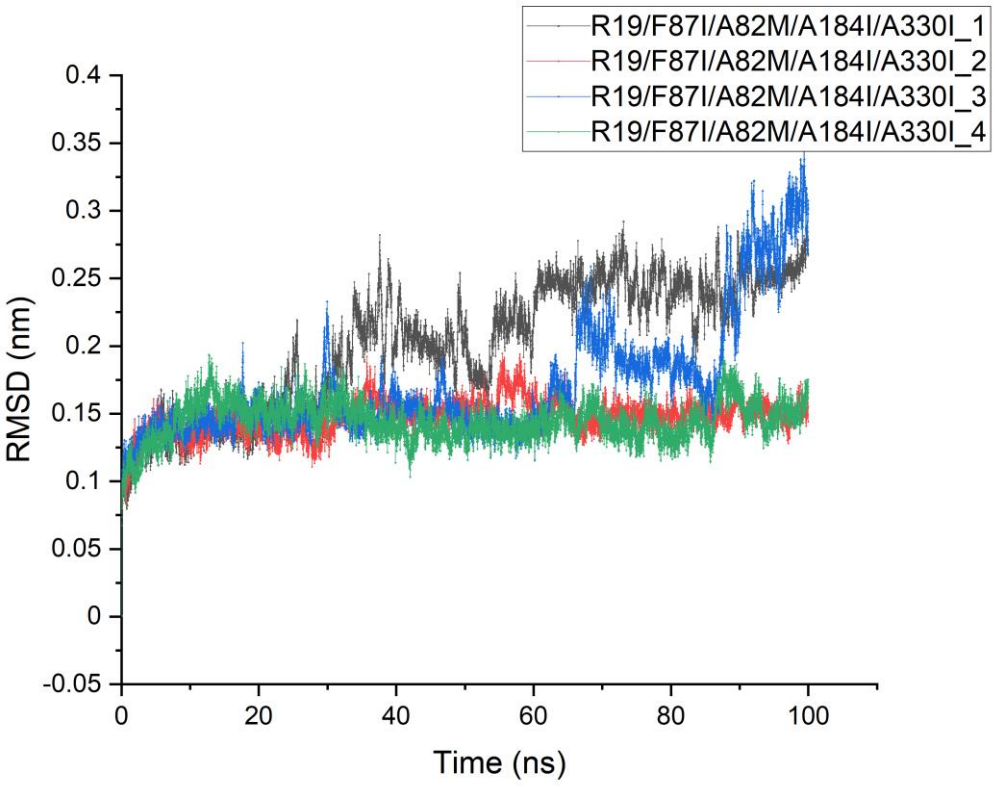


Figure 4.8 The RMSD plot of each replica of mutant R19/F87I/A82M/A184I/A330I.

The RMSD values of replicas 1 and 3 indicated relatively unstable structures as the deviations were much larger (up to 3.5 Å) as the 100 ns simulations progressed (Fig. 4.8). Replica 2 was significantly more stable compared to replicas 1 and 3. Since unstable simulated structures might lead to less reliable docking results, a further simulation was performed. The RMSD values of replica 4 demonstrated a stable structure, similar to the RMSD values of replica 2. The cluster populations for the replicas are summarised in Table 4.6.

Table 4.6 Cluster population of the four replica MD simulated structures of R19/F87I/A82M/A184I/A330I.

Replica	Cluster population
Replica 1	R1C1 (3353)
	R1C2 (2439)
	R1C3 (1411)
Replica 2	R2C1 (8031)
	R2C2 (1016)
	R2C3 (443)
Replica 3	R3C1 (5647)
	R3C2 (1820)
	R3C3 (635)
Replica 4	R4C1 (8179)
	R4C2 (593)
	R4C3 (419)

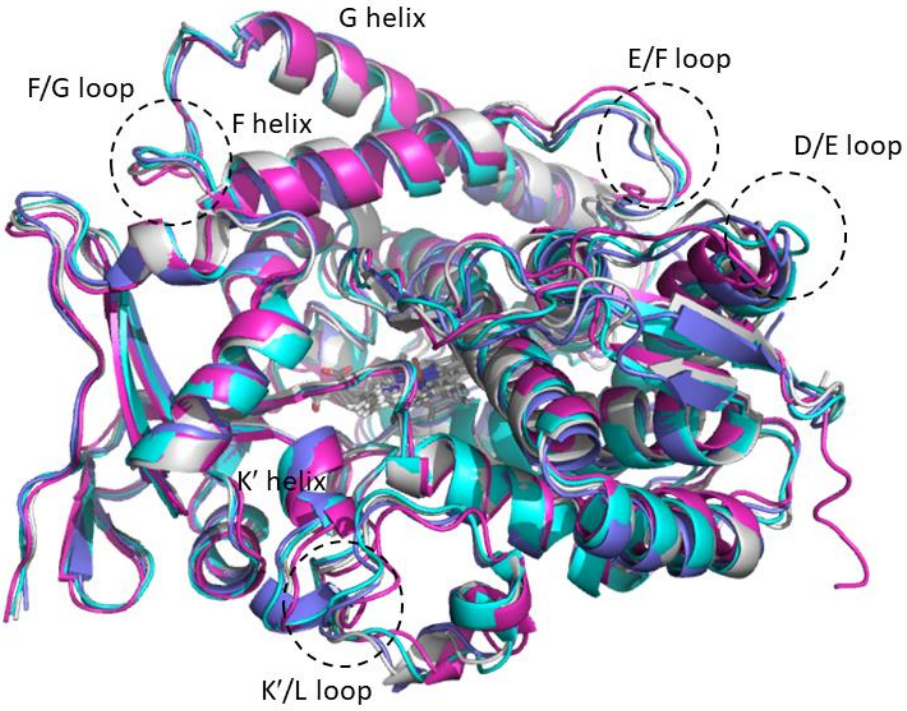


Figure 4.9 Overlay of the structures of the most populated cluster from four replica simulations of the structure of variant R19/F87I/A82M/A184I/A330I. Circles highlight the regions showing large deviations between replicas, R1C1 (magenta), R2C1 (cyan), R3C1 (grey) and R4C1 (purple).

The most populated clustered structures from each replica were overlaid and compared for deviations in structure. Most of the residues showed good alignment, with small deviations in the loop regions (Fig. 4.9). Of the mutated residues, rotamers were observed for residues M82 and I87 (Fig. 4.10A). The MD-simulated structure of R19/F87I and WT were then overlaid with these clustered structures

for comparison. The mutations of A82M, A184I and A330I had constrained and reshaped the substrate pocket, providing a better fitting for bulnesene in the substrate pocket (Fig. 4.10B).

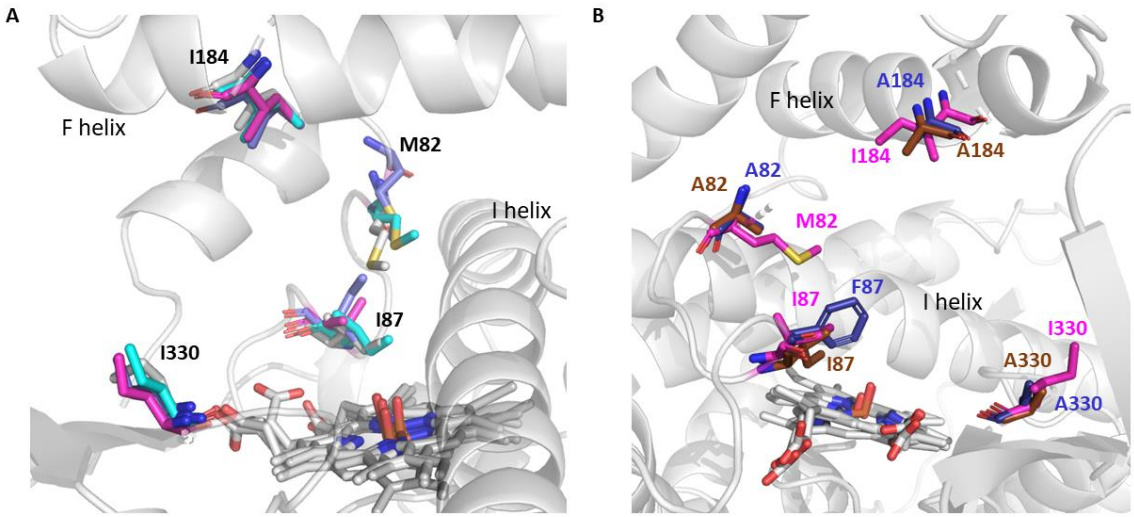


Figure 4.10 Overlay of selected residues and helices from (**A**) R1C1 (magenta), R2C1 (cyan), R3C1 (grey) and R4C1 (purple); and (**B**) The MD-simulated structure of R1C1(magenta), R19/F87I (brown) and WT (blue).

4.4.1.2 General docking outputs

Bulnesene-15-COOH (**1g**) was docked into the 12 clusters of R19/F87I/A82M/A184I/A330I by Autodock Vina and summaries of the results are shown in Table 4.7 and Figure 4.11. There were 9 poses generated from each cluster which resulted in a total of 108 poses with C14 poses as the most dominant (30%). Poses for C11=C12 oxidation, which was the major product of the oxidation of C15-acid **1g** by variant R19/F87I/A82M/A184I/A330I, were the second most dominant (27%). As C14 was a non-activated primary methyl group, the C11=C12 double bond was more likely to be oxidised, in agreement with the experimental result. There were only 5 out of 44 productive poses (11%) indicating C2 oxidation, compared to 4% observed experimentally. By overlaying the C2 poses with C11=C12 poses and non-productive poses (NP), mutations would be introduced to hinder these undesired poses without affecting the C2 poses. The residues which formed hydrogen bonds with C12 poses would be targeted specially to hinder C11=C12 oxidation and hopefully switch the binding orientation to C2 poses.

Table 4.7 Pose distribution of the docking of bulnesene-15-COOH **1g** with variant R19/F87I/A82M/A184I/A330I with Autodock Vina.

		C2	C9	C12/13	C3	C4	C5	C6	C7	C8	C14	NP
Replica 1	C1 (3353)			1	2						1	5
	C2 (2439)			2	1						1	5
	C3 (1411)	1	1	2							2	3
Replica 2	C1 (8031)	1	1								2	5
	C2 (1016)											9
	C3 (443)			1								8
Replica 3	C1 (5647)	1		2							1	5
	C2 (1820)	2	2		2						1	2
	C3 (635)		1	2		1				1	2	2
Replica 4	C1 (8179)					1				1		7
	C2 (593)			2							1	6
	C3 (419)										2	7
		5	5	12	5	2	0	0	0	2	13	64

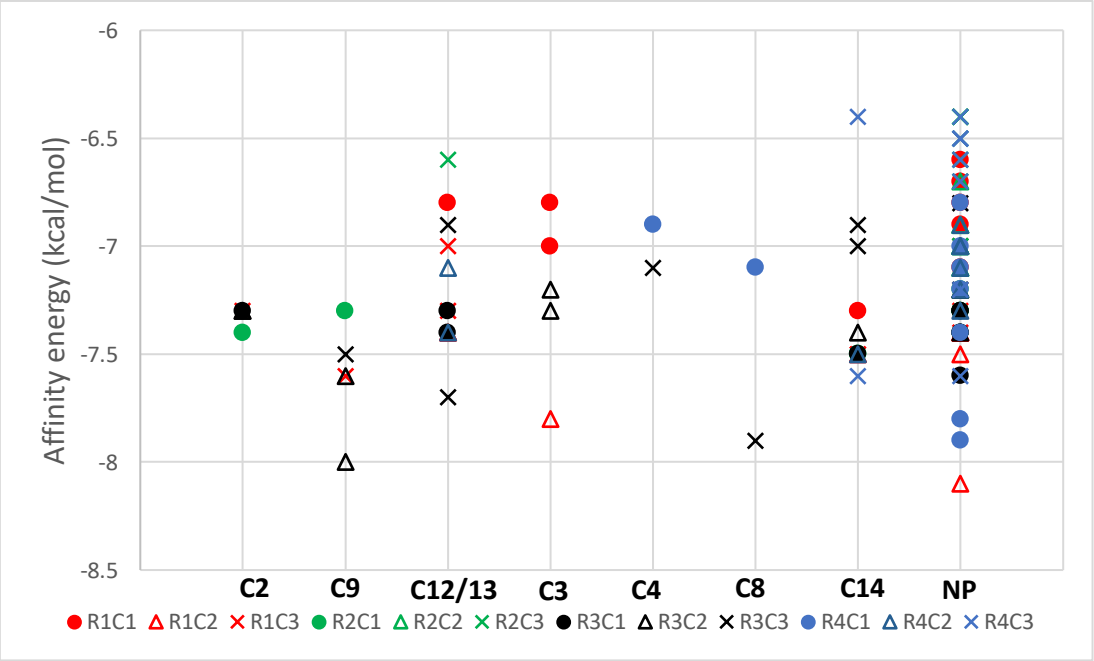


Figure 4.11 The affinity energies of output poses from docking of bulnesene-15-COOH **1g** with variant R19/F87I/A82M/A184I/A330I with Autodock Vina.

4.4.1.3 The C2 poses

All C2 poses had similar orientation (Fig. 4.12A). The exocyclic double bond pointed towards the B′ helix and was in contact with residues S72, A74, L75 and V78. The carboxylic acid group was close to the I helix and formed a hydrogen bond with A264. The presence of this hydrogen bond

could be crucial in holding **1g** at the specific orientation for C2 oxidation. The lowest energy C2 pose was aligned with the other C2 poses but shifted slightly away from S72 in the B′ helix (Fig. 4.12B).

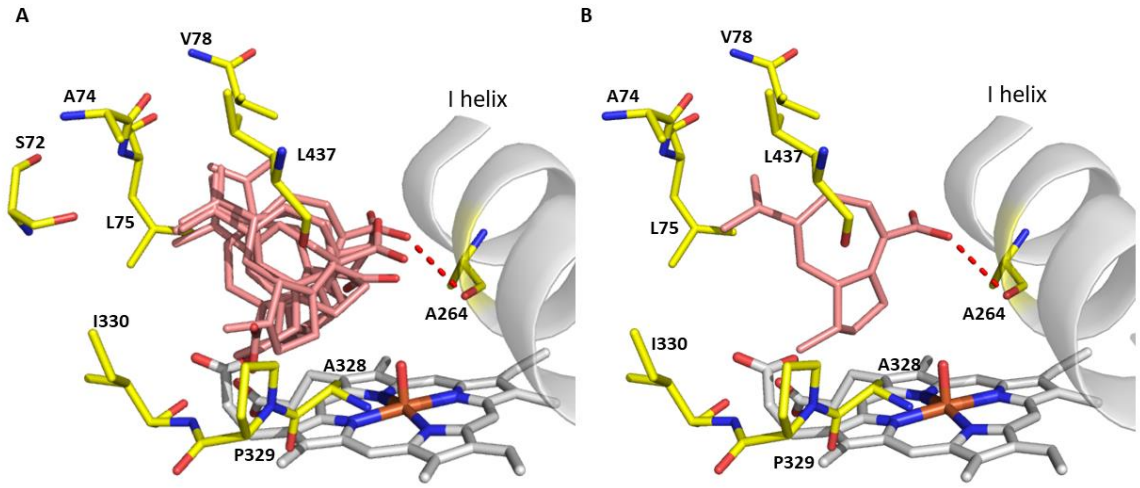


Figure 4.12 Overlays of simulation structure of R19/F87I/A82M/A184I/A330I (grey) with **(A)** all C2 poses (salmon) and **(B)** the lowest energy C2 pose (salmon). Residues that contacted the bulnesene substrate are in yellow. Red dashes indicate hydrogen bonds.

4.4.1.4 The C11=C12 poses

The C11=C12 poses were scattered in the middle of the substrate pocket, close to where the C2 poses were located (Fig. 4.13). A group of C11=C12 poses was between the β_1-strand residues I330 and S332 and the B′-helix residues L71 and S72, with hydrogen bonds to the hydroxyl groups of S72 and S332. These poses could be disfavoured by mutations to remove these hydrogen bonds and by introducing residues with bulkier side chains. These and other poses were also in contact with residues like V78, L181 and E267, and some poses formed a hydrogen bond with the side chain of T438.

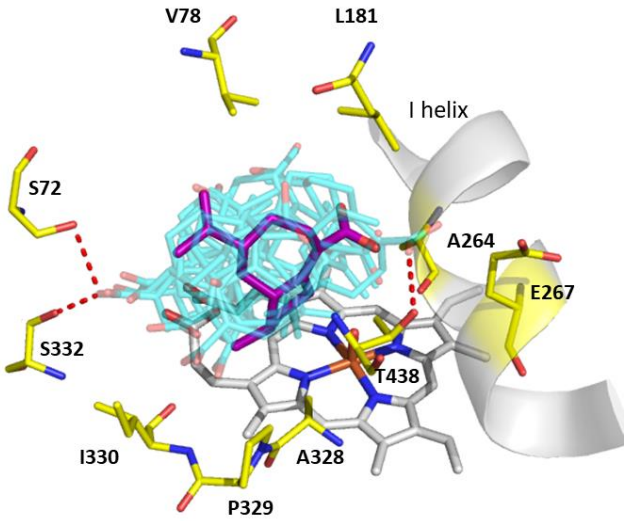


Figure 4.13 Overlays of MD-simulation structure of R19/F87I/A82M/A184I/A330I (grey) with C11=C12 poses (cyan) and the lowest energy C2 pose (magenta). Residues which are in contact with C11=C12 poses are in yellow. Red dashes denote hydrogen bonding with active site residues.

4.4.1.5 The NP poses

The non-productive (NP) poses were clustered in the middle of the substrate pocket and when overlaid with the representative C2 pose, two main regions were targeted for mutagenesis to disfavour these poses. The first region (**A**, Fig. 4.14) was near the β_1-strand residues I330 and S332 and the B′-helix residues L71 and S72; this was the same region targeted to disfavour the C11=C12 poses. Both residues S72 and S332 formed hydrogen bonds with the NP poses so mutations could be introduced to remove this binding orientation.

The other cluster of NP poses were gathered below the second region (**B**, Fig. 4.14) at the start of the F helix, further away from the heme. As mutating residue A184 on the F helix to Ile was found to enhance bulnesene oxidation, the F-helix residues V178, L181 and L188 in this region and which pointed towards the substrate pocket, were selected for mutagenesis.

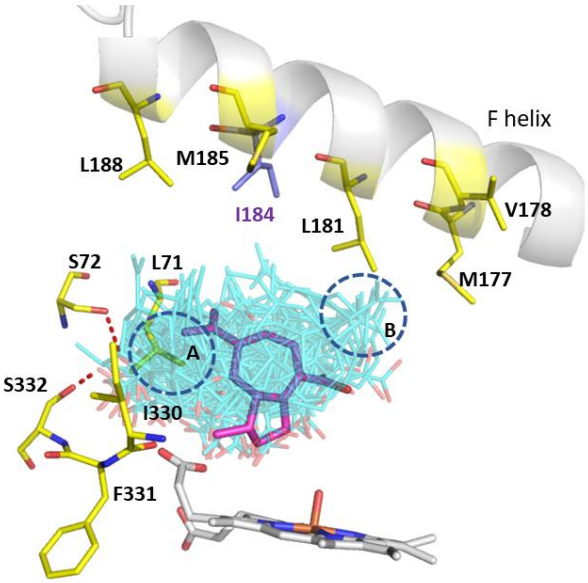


Figure 4.14 Overlays of the MD-simulation structure of R19/F87I/A82M/A184I/A330I (grey) with all NP poses (cyan) and the lowest energy C2 pose (magenta). Residues which are in contact with the NP poses are in yellow. Red dashes denote hydrogen bonding with active site residues.

4.5 Docking-guided mutagenesis for bulnesene-15-acid oxidation

4.5.1 Mutations designed for selectivity swap and activity enhancement

Mutations were introduced to the template variant R19/F87I/A82M/A184I/A330I to eliminate hydrogen bonding and increase side-chain volume. The target residues were L71 and S72 on the B′ helix, S332 on the β_1 strand, V178 and L188 of the F helix as well as L437 and T438 as they were in the main regions where NP and C11=C12 poses were clustered. Hydrogen bonds were also observed between these undesired poses and S72, S332 and T438. The new variants were screened for bulnesene oxidation activity, and the results are summarised in Figure 4.15. Both epoxides of C15-acid **1h** and **1h(a)** were taken into account for C15 selectivity as they were the further oxidation products of C15-acid **1g**.

Bulnesene conversion by the new variants were comparable or higher than the template variant and, in most cases, with increased C15 selectivity. The selectivity of C2 oxidation was not increased; the highest was the 5% from the R19/F87I/A82M/A184I/A330I precursor and the new variant

R19/F87I/A82M/A184I/A330I/L171H/V178F which also showed higher conversion (97%) and increased total C15 selectivity (86%). Notably, however, the proportion of 15,2-lactone relative to the 11,12-epoxides was increased, indicating a small shift in regioselectivity of C15-acid oxidation. The V178F mutation increased side chain volume and was designed to block NP poses by filling the gap beneath the F helix. The L171 mutation in the template was reverted to histidine (H) as the V178F mutation was added. Due to insufficient time, the intended variant was not prepared. L188F was another F helix mutation designed to block NP poses. There was a slight increase in conversion and the total C15 selectivity; however, chemoselectivity switched to 67% of bulnesene-15-ol **1a** and 23% bulnesene-15-al **1b** without further oxidation to the C15-acid. Whilst no 15,2-lactone was formed, the L188F mutation may be applicable in high-conversion variants favouring the C15-alcohol **1a** and C15-aldehyde **1b**.

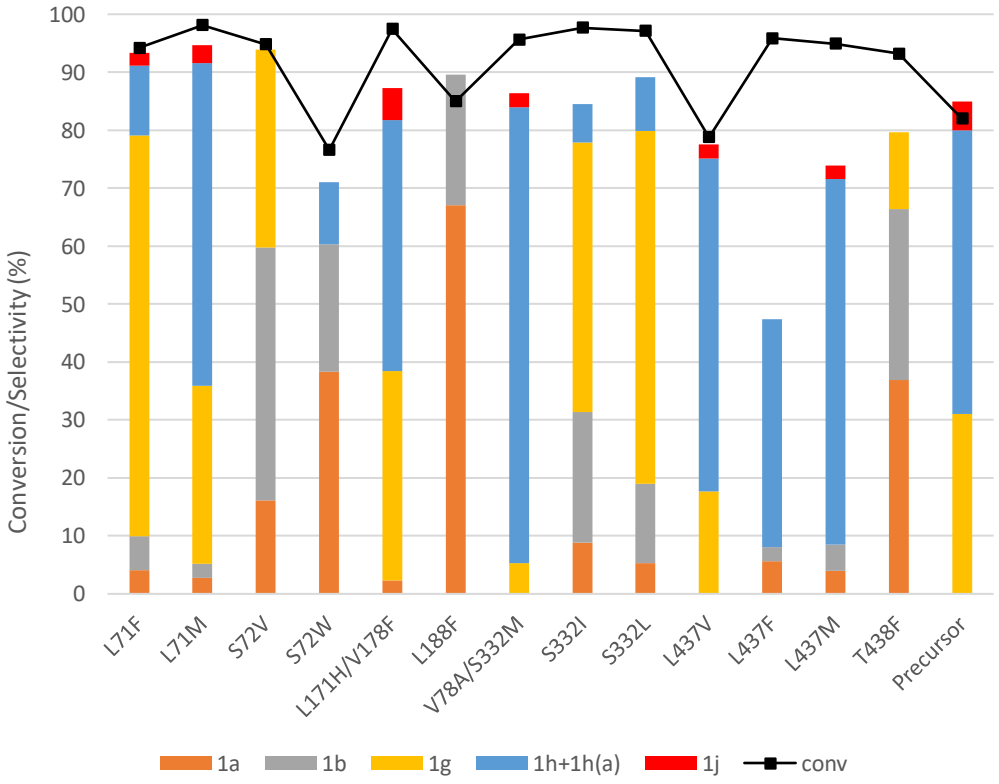


Figure 4.15 The effect of new mutations on variant R19/F87I/A82M/A184I/A330I.

The L437 mutations lowered C15 selectivity, especially for L437F (48%), but increased bulnesene conversion. Interestingly, these mutations increased the conversion of the C15-acid, to the extent

that the acid was not observed with the L437F and L437M variants even if the 11,12-epoxides were the major products with less of the 15,2-lactone formed than the template variant. These mutations could be effective in promoting C15-acid oxidation activity by CYP102A1, for example, once better regioselectivity control is established.

The T438F mutation eliminated the possibility of hydrogen bonding with the C15-acid and decreased the activity significantly. Mutations of this residue could be relevant for engineering chemoselectivity for the C15-alcohol and C15-aldehyde without further oxidation to the acid.

Mutation of L71 and S72 on the B′ helix to residues with bulkier side chains increased the total C15 selectivity to >90%, except for S72W which lowered it to 70%. The S72V mutation retarded C15 oxidation, with less of the acid formed and no further oxidation to the 11,12-epoxide. The L71F mutation had the interesting effect of giving more of the C15-acid but less of the 11,12-epoxide while forming some of the 15,2-lactone; in other words, this mutation switched the selectivity of C15-acid oxidation away from the C11=C12 double bond towards the C2 position. The L71M mutation led to more of the 11,12-epoxides while maintaining the 15,2-lactone proportion. Therefore, the L71M mutation increased C15 selectivity but had little effect on the selectivity of C15-acid oxidation.

Mutations S332I, S332L and S332M increased bulnesene conversion to >95% compared to 82% for the template variant and maintained the total C15 selectivity. The S332I and S332L mutations changed the chemoselectivity back towards the C15-alcohol, C15-aldehyde and C15-acid, with less oxidation of the acid. The variant with the S332M mutation also had the V78A mutation, most likely an adventitious mutation; it showed significantly higher activity, oxidising bulnesene with 95% conversion to mostly the 11,12-epoxide of the C15-acid (79%) with 2% 15,2-lactone and just 5% of the C15-acid remained. However, it was not clear whether these changes were due to the S332M or V78A mutation, and there was not sufficient time for further investigations.

As heating and acidification were found to be important in lactonisation, the screening reactions were repeated with internal standard added, and they were adjusted to pH 1 with 2 M HCl and heated at 40 °C for one hour. Surprisingly, the 15,2-lactone **1j** was observed at higher selectivity for variant R19/F87I/A82M/A184I/A330I/V78A/S332M with the R value (ratio of the relative area of the peaks of **1j** to IS) increased to 0.14 compared to 0.06 by the precursor variant R19/F87I/A82M/A184I/A330I. Variant R19/F87I/A82M/A184I/A330I/L171H/V178F gave similar R value to the precursor variant (~0.06) while all the other variants gave little (<0.05) or no 15,2-lactone, as observed from the screening results without acid and heat treatment (Fig. 4.15). Mutations of S332, which were introduced to disfavour the C11=C12 and NP poses, had shown significant increase in activity and a significant selectivity change towards C2 oxidation.

Overall, the results demonstrated that docking-guided mutation design could lead to changes in activity and selectivity even if the outcomes might not follow the rationale for the design of the mutation. Thus, C2 selectivity for C15-acid oxidation was not materially enhanced except with the V78A/S332M combination. However, most of the mutations designed to disfavour NP poses in order to increase activity led to increased conversion and, indeed, other desirable traits such as increased C15 selectivity, increased proportions of the C15-acid (e.g. L71F), and chemoselectivity change to favour the C15-alcohol and C15-aldehyde (e.g. L188F). Further work should firstly establish whether the L171H reversion played a significant role in the activity of the V178F mutant, and the role of the V78A mutation in the S332M combination. It would be most interesting to carry out further rounds of docking-guided mutagenesis with docking of the C15-acid into MD-simulation structures of the S332M (with and without V78A) and V178F variants for mutation design to disfavour side products.

Since the selectivity for the 15,2-lactone remained rather low, it appeared that an acceptable solution was the one-pot cascade enzyme system in which one enzyme E1, such as the R19/F87I/A82M/A330V variant, would generate the C15-acid **1g** with high conversion and

selectivity while a second enzyme E2, such as the F87A/A184I/I263G/A264G/A330P variant which has little activity for bulnesene oxidation, would convert acid **1g** to 15,2-lactone **1j**.

4.6 The synthesis of target compounds from bulnesene-enriched patchouli oil (BPO)

4.6.1 The synthesis of bulnesene-15-ol, bulnesene-15-al and bulnesene-15-COOH with BPO

Although high efficiency processes for the synthesis of these target compounds have been established, the starting material bulnesene was purified by chromatography using $AgNO_3$-SiO_2 which is not applicable at scale. An alternative source of bulnesene was required. One possibility is to synthesise bulnesene *de novo* using synthetic biology and metabolic pathway engineering. Such systems require optimisation. Another possibility arose when bulnesene-enriched patchouli oil (BPO) obtained by distillation was made available by Endeavour Chemicals. Bulnesene makes up roughly 50% of BPO, patchoulol is the other major component (~15%) while the rest are other sesquiterpenes (total ~35%). It was expected that a variant with high substrate specificity would be needed to selectively oxidise bulnesene out of the pool of different compounds to synthesise the target compounds efficiently. Nevertheless, BPO presented an interesting alternative source of bulnesene worthy of investigation.

Variant R19/F87I/A82M/A330V was found to oxidise bulnesene to C15-carboxylic acid **1g** selectively at 1000:1 S:E ratio. Therefore, the variant was first used to oxidise BPO to determine if it would show the same behaviour as when pure bulnesene was used as substrate. Both purified bulnesene and BPO were fully converted in 6 hours under the same conditions (1000:1, Fig. 4.16). BPO was oxidised to C15-alcohol **1a**, C15-aldehyde **1b** and C15-carboxylic acid **1g** in roughly 1:2:1 ratio while purified bulnesene was converted to mainly **1g**. Another reaction of BPO which was run in parallel with a 500:1 ratio gave mainly **1g** with some **1b**. The results demonstrated that, even though the selectivity of the variant remained similar, the bulnesene conversion activity was lower

when BPO was used as substrate which might be because the other compounds in BPO, such as α-guaiene **2**, α-patchoulene **4** and seychellene **5**, were also oxidised by the variant as their corresponding peaks had disappeared. Interestingly, patchoulol **6**, which may also be inhibiting the enzymatic activity, was barely oxidised by the enzyme and remained in the reaction mixture. The hydroxylated compounds and patchoulol could be separated by distillation. A CRED could also be used to oxidise **1a** to **1b** for the synthesis of the agarwood aroma compound, bulnesene-15-al **1b**. Despite the decrease in activity, the reaction had shown that BPO could be used as feedstock to synthesis agarwood aroma compounds.

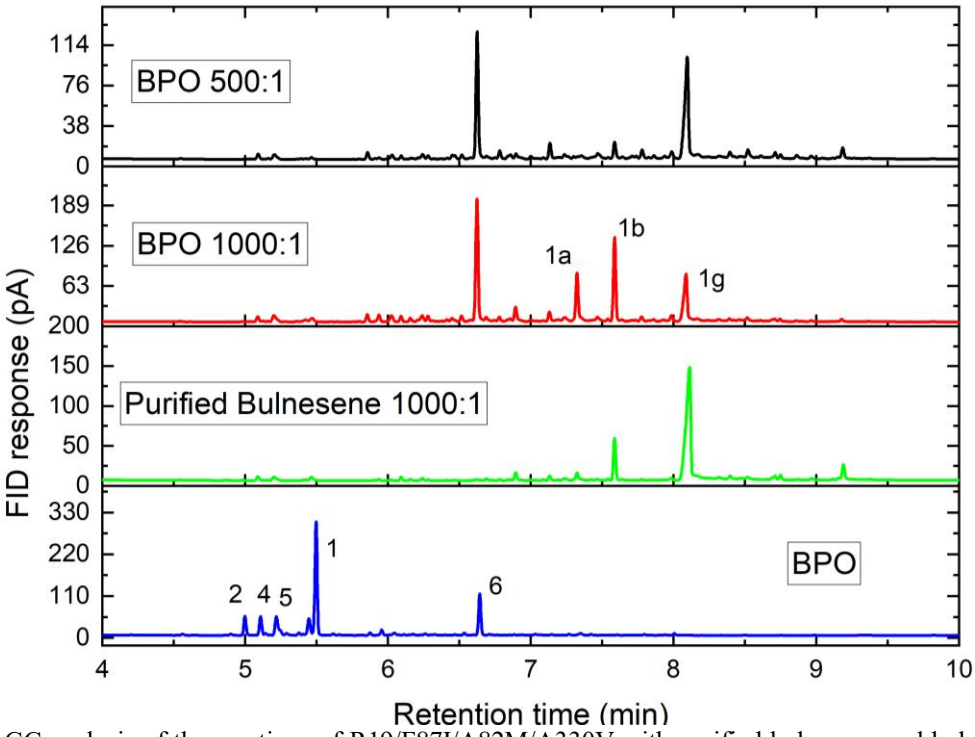


Figure 4.16 GC analysis of the reactions of R19/F87I/A82M/A330V with purified bulnesene and bulnesene-enriched patchouli oil (BPO) at different substrate-to-enzyme ratios.

4.6.2 The synthesis of bulnesene-15,2-lactone from BPO

As the conditions for the synthesis of 15,2-lactone **1j** with purified bulnesene had been established, the synthesis of 15,2-lactone from BPO under similar conditions was attempted. R19/F87I/A82M/A330V (E1) and F87A/A84I/I263G/A264G/A330P (E2) were added at the beginning of the reaction (T_0). Two aliquots were removed periodically (every 1.5 hours) for GC analysis, with one being heated and acidified to encourage lactone formation. Interestingly,

bulnesene was almost fully converted in the first 1.5 hours with only **1g** and **1j** being observed in the reaction mixture (Fig. 4.17). The lactone **1j** grew gradually as **1g** decreased and the reaction ended in 24 hours.

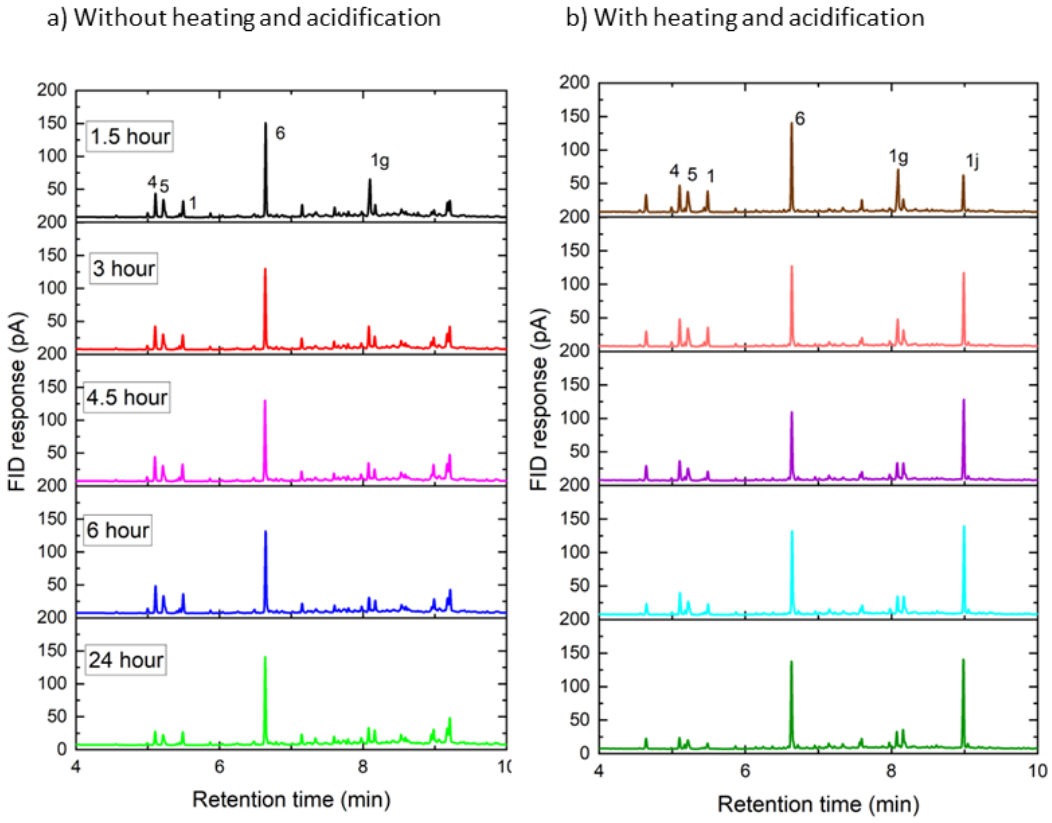


Figure 4.17 Time-course reaction for the synthesis of 15,2-latone **1j**, with and without heating and acidification, using bulnesene-enriched patchouli oil (BPO).

The results were promising as the reaction demonstrated similar results as when purified bulnesene was used. Only **1j** and patchoulol **6** remained as the main components in the reaction mixture which could be easily separated by distillation.

4.6.3 Optimisation and titre test on the synthesis of bulnesene-15,2-lactone

The use of BPO simplified the process of obtaining bulnesene and allowed the synthesis of 15,2-lactone. Besides R19/F87I/A82M/A330V （E1a） which had been used as E1, a parallel reaction with R19/F87I/A82M/A184I/A330I/V78A/S332M (E1b) was run as this variant was shown to give high activity and produce **1j** as a minor product.

Table 4.8 The relative amount of bulnesene **1**, before and after reaction, and 15,2-lactone **1j** formed by cascade of different CYP102A1 variants E1 and the F87A/A184I/I263G/A264G/A330P functioning as E2. The bulnesene:E1 ratio was 1000:1 (2 mM bulnesene and 2 μM of E1). The reaction time was 24 hours. The concentrations of bulnesene and its metabolites are expressed as the ratio of the peak area of each compound to the peak area of an added internal standard. Production of **1j** was the percentage of **1j** to the amount of bulnesene, **1**, in BPO at the beginning of the reaction.

CYP102AI variants	E1:E2	1 (before)	1 (after)	1j	Production of 1j (%)
R19/F87I/A82M/A330V	1:1	1.73	–	0.42	24
	1:1.5	1.45	–	0.57	39
	1:2	1.40	0.05	0.67	48
R19/F87I/A82M/A184I/A330I/V78A/ S332M	1:1	1.17	–	0.35	30
	1:1.5	1.11	0.05	0.46	42
	1:2	1.33	0.07	0.75	56

Previous results had shown that the bulnesene oxidation activity was lowered when BPO was used as substrate and thus, the ratio E1:E2 was adjusted for more efficient synthesis of 15,2-lactone **1j**. In general, the amount of **1j** produced increased with higher ratio of E1:E2 (Table 4.8). The production of **1j** was higher when R19/F87I/A82M/A184I/A330I/V78A/S332M was used as E1; the highest was observed at a 1:2 ratio. This might be because variant R19/F87I/A82M/A184I/A330I/V78A/S332M is more active and gives **1j** as one of the products so using this variant as E1 would promote the production of **1j**.

Table 4.9 The relative amounts of bulnesene, **1**, before and after the reaction, and its metabolites formed by cascades of different CYP102A1 variants E1 and the F87A/A184I/I263G/A264G/A330P variant functioning as E2 at different titres.

CYP102A1	S:E	1 (Before)	1(After)	Conv.	1g	1j	1g+1j
E1a	2000:1	3.93	0.97	75	0.49	0.75	1.24
	3000:1	8.26	5.86	29	1.46	1.10	2.56
E1b	2000:1	3.30	0.74	78	0.51	0.99	1.50
	3000:1	5.12	2.55	50	0.89	0.82	1.71

***1j** produced at 1000:1 was 0.67 and 0.75 for E1a and E1b respectively from Table 4.8.

The bulnesene:E1 ratios (S:E) were 2000:1 and 3000:1 (4/6 mM bulnesene and 2 μM of each of the E1). The E1:E2 ratio was 1:2. The reaction time was 24 hours. The concentrations of bulnesene and its metabolites are expressed as the ratio of the peak area of each compound to the peak area of an added internal standard.

With the optimised conditions, a set of experiments was designed at higher titre (2000:1 and 3000:1) as both R19/F87I/A82M/A330V and R19/F87I/A82M/A184I/A330I/V78A/S332M were shown to give high TON at higher S:E ratio. When the S:E ratio increased to 2000:1 and 3000:1, the amount of **1j** produced was expected to increase by roughly two and three times respectively. However, even though the increase of **1j** had been observed when R19/F87I/A82M/A330V (E1a) was used, it was only 1.1 and 1.7 times higher. This is most likely because, as the S:E ratio increased, both the conversion of bulnesene **1** and C15-carboxylic acid **1g** decreased, which resulted in less **1j** being formed.

Unlike 1000:1 reaction that gave almost 100% conversion (Table 4.8), both 2000:1 reactions by E1a and E1b gave good conversion (>75%) but this dropped significantly at 3000:1 ratio. A significant amount of **1g** was left in the reaction mixture especially for the 3000:1 reactions which signified that more E2 was needed for higher production of **1j**. This is consistent with the screening result of purified **1g** with F87A/A184I/I263G/A264G/A330P which was fully converted at 500:1 but decreased significantly once the titre was increased. Therefore, further optimisation could be done by increasing the E2 concentration for a reaction with higher titre. However, due to the time constraint, the optimisation was not able to be done but improvements are feasible if there is interest in the 15,2-lactone, the agarwood aroma compound.

4.7 Summary

The biocatalytic synthesis of 15,2-lactone (**1j**) from bulnesene (**1**) has been established by a one-pot cascade enzyme system. A CYP102A1 variant enzyme (E1) is used to catalyse three cycles of oxidation of bulnesene at C15 to form acid **1g** which is then oxidised at the C2 position by a second CYP102A1 variant enzyme (E2) in a one-pot reaction. Through the combined effects of enzyme engineering and reaction engineering, the 15,2-lactone was isolated with significantly increased yield and good purity which could be beneficial to the development of a viable scaled process for the synthesis of lactone **1j**. The biosynthesis of the agarwood aroma compound was demonstrated by using bulnesene-enriched patchouli oil as feedstock. Such a process could only be achieved with variants of high substrate specificity because there were numerous different sesquiterpenes present in the bulnesene-enriched patchouli oil. Further effort is needed to optimise the reaction conditions for the large-scale production of 15,2-lactone from bulnesene-enriched patchouli oil at higher titre and to develop the direct oxidation of bulnesene to bulnesene-15,2-lactone with a single CYP102A1 variant.

5 Chemoselective bulnesene oxidation by reaction engineering

5.1 Chemoselectivity control of bulnesene oxidation

Although the bulnesene oxidation products could be separated by vacuum distillation, high chemoselectivity for the target compounds simplifies downstream processing. All variants developed were tested again with different reaction time and substrate-to-enzyme ratio (S:E) to study if the chemoselectivity could be controlled by reaction engineering. Variants showing >50% bulnesene conversion were divided into four groups based on their major products; Group A, B, C and D with the major products of C15-OH, C15-al, C15-COOH and the 11,12-epoxide of C15-COOH, respectively.

Table 5.1 The criteria for grouping the variants for chemoselectivity test.

Group	Criteria
A	Selectivity for bulensene-15-ol, **1a** >50%
B	Selectivity for bulnesene-15-al, **1b** >40%
C	Selectivity for bulnesene-15-al, **1g** >50%
D	Mutants that oxidised bulnesene to 11,12-epoxy-bulnesene-15-COOH, **1h**

5.1.1 Reaction time

The reactions were repeated with the same conditions as the screening experiments at various reaction times (2, 4, 6 and 16 hours) and the data were analysed to study how the product profile varied as the reaction progressed (Fig. 5.1).

Group A variants were R19/F87I-based, with additional mutations at residues S72 and T88. Bulnesene conversion increased with time and typically plateaued at 6 hours while the chemoselectivity remained similar regardless of the reaction time (Fig. 5.1). For example, selectivity

for C15-alcohol **1a** of variant R19/F87I/T88F remained at ~65% after 2, 4 and 6 hours although it decreased slightly (59%) for the overnight reaction; the proportion of aldehyde **1b** increased by 4% and a small amount of the carboxylic acid **1g** was also formed (2%).

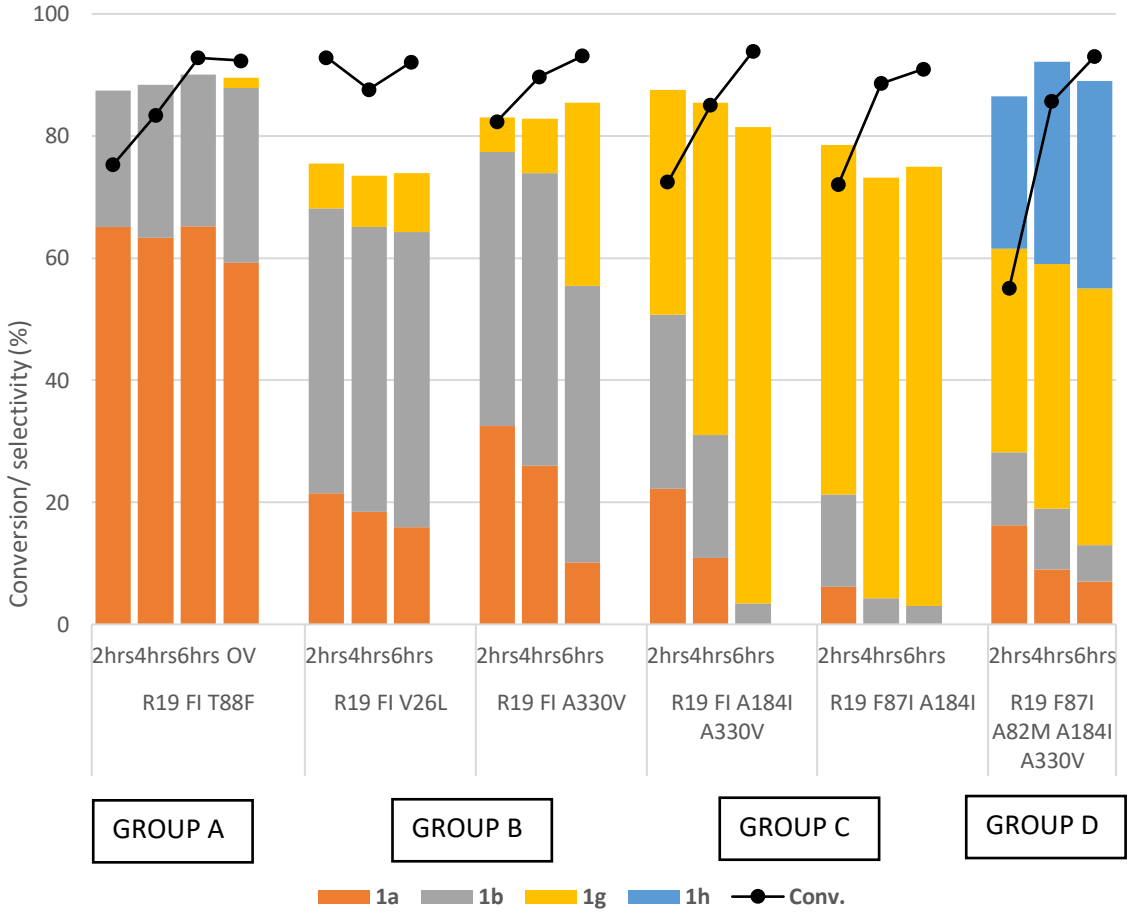


Figure 5.1 Effect of reaction time on the chemoselectivity of bulnesene oxidation.

Variants with C15-aldehyde **1b** selectivity higher than 40% were categorised as group B and two types of behaviour were observed. Variant R19/F87I with additional mutations at V26 and A74 showed similar chemoselectivity at 2, 4 and 6 hours with very slight increase in C15-carboxylic acid **1g** and a decrease in C15-alcohol **1a**. The conversion remained similar throughout. Variants with mutations at A330 were more active, as demonstrated by R19/F87I/A330V (Fig. 5.1). They showed a similar trend in the first 4 hours but the proportion of **1g** increased significantly after 6 hours. Interestingly, the decrease in the proportion of **1a** was similar to the increase in the formation of **1g**

while the change of **1b** percentage was little throughout the reaction for both R19/F87I/A330V and R19/F87I/A330I. The behaviour of **1b** was similar to an intermediate in a steady-state reaction, i.e., the rate of formation of **1b** was similar to that of its oxidation to **1g**.

For group C variants there was a steady increase in the proportion of acid **1g** as the amount of **1a** decreased while that of **1b** changed little, e.g. the R19/F87I/A184I/A330V variant. The R19/F87I/A184I variant was highly active, with bulnesene being almost fully converted to the carboxylic acid **1g** in just 4 hours.

Triple-mutation variants based on R19/F87I from Group D were extremely active and they oxidised bulnesene to 11,12-epoxy-bulnesene-15-COOH **1h** in 6 hours. It was proposed that if the reactions could be stopped before **1h** was formed, then acid **1g** could be produced at high selectivity within a short time. However, group D variants showed similar chemoselectivity throughout the reaction, with gradual decreases in **1a** and **1b**, and **1h** was formed after just 2 hours even before **1a** was fully consumed. The results demonstrated that the C15-acid **1g** was as good a substrate as **1**, **1a** and **1b** for group D mutants that it was further oxidised as soon as it was produced after a certain amount of **1b** and **1g** had been formed.

To summarise, increased conversion was observed with longer reaction time but varying the reaction time did not change the chemoselectivity as much as expected. Different substrate-to-enzyme ratios were explored as an alternative to control the chemoselecivity of each product.

5.1.2 The substrate-to-enzyme ratio

A new set of experiments was designed to study the chemoselectivity of the variants with different substrate-to-enzyme (S:E) ratios.

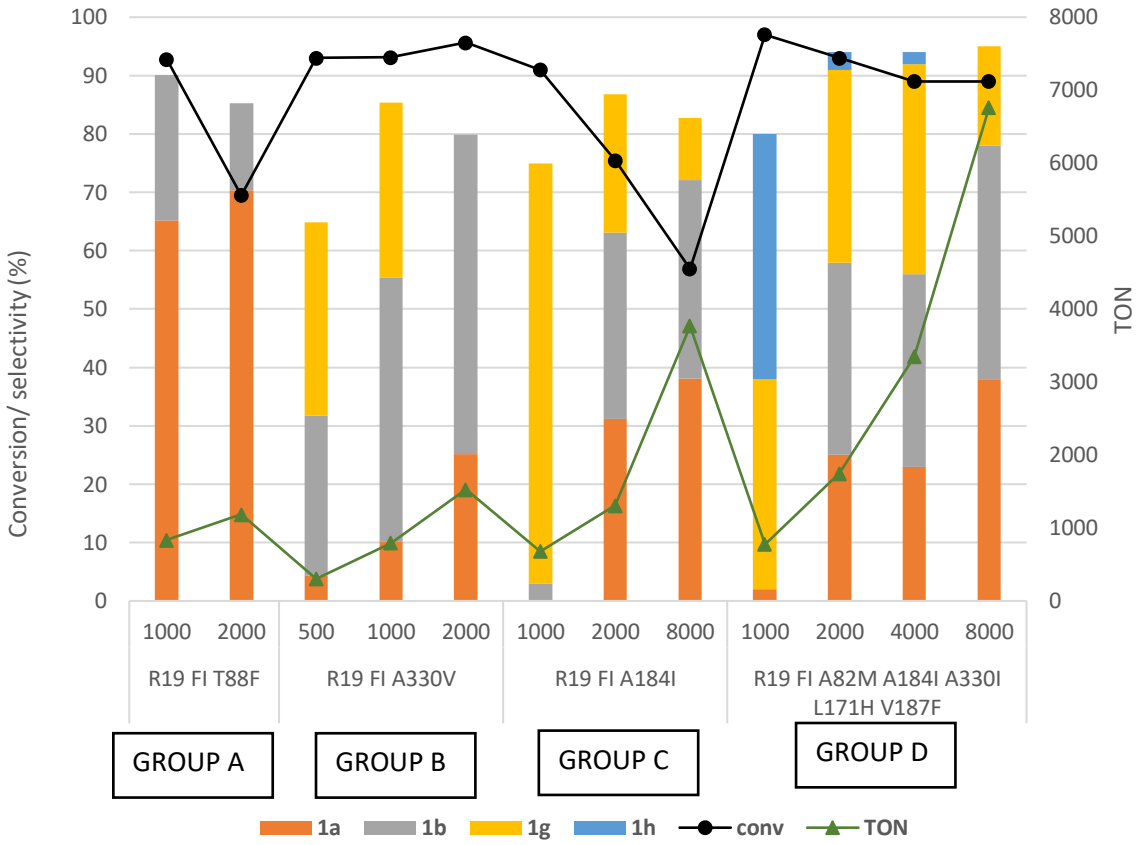


Figure 5.2 Chemoselectivity control by varying the substrate-to-enzyme (S:E) ratio.

In general, increasing the S:E ratio had the effect of tuning the chemoselectivity of bulnesene-15-ol **1a**. As shown in Figure 5.2, the chemoselectivity of **1a** increased as the S:E increased for most of the variants. Despite the decrease in percentage conversion, the TON grew which demonstrated that high activity variants from groups C and D could be used to synthesise **1a** at higher titre; the selectivity for **1a** of variant R19/F87I/A82M/A184I/A330I/L171H/V178F increased from 2% at 1000:1 to 38% at 8000:1. The TON for bulnesene conversion was 6770 which was the highest among all variants tested. Higher S:E ratios could possibly increase **1a** selectivity further. Decreasing the

S:E ratio to 500:1 had adverse effects on the overall C15 selectivity as further oxidation was observed and many side products were produced.

Changing the S:E ratio did not have much of an effect on increasing the selectivity for **1b** and **1g**. Similar to the observation before, the selectivity of **1b** varied slightly at different S:E ratios regardless of the selectivity change of other products, which agreed with the earlier hypothesis that the amount of **1b** was maintained at a steady-state concentration. The highest **1g** chemoselectivity remained as 72% by variant R19/F87I/A184I and any increase in S:E ratio gave a mixture of **1a**, **1b** and **1g** as the products.

Interesting trends were observed with high activity variants from group C and D. For example, the conversion of variant R19/F87I/A82M/A184I/A330I/L171H/V178F dropped slightly as expected when the S:E ratio increased but it remained high (85%) even at 8000:1 ratio. For 2000, 4000 and 8000:1 S:E ratios, the selectivity trends for all the products remained similar with only slight decrease in **1g** at higher ratios. The results demonstrated that the S:E ratio could be further increased and variant R19/F87I/A82M/A184I/A330I/L171H/V178F could be an excellent candidate for a process which oxidises bulnesene to **1a** selectively or to a mixture of **1a** and **1b** that could then be further oxidised chemically or enzymatically, for example, with a CRED, to give exclusively **1b** with high selectivity.

5.2 Reaction optimisation

5.2.1 Methyl-β-cyclodextrin (MCD) concentration

Methyl-β-cyclodextrin (MCD) was added in the reaction system as a solubiliser and transporter to bring the highly hydrophobic substrate, bulnesene, into the aqueous phase and make it more available to the enzyme. It has a cup-like structure with a hydrophobic interior cavity that allows the lipophilic substrate to bind within it and a water-soluble surface which increases its water solubility.

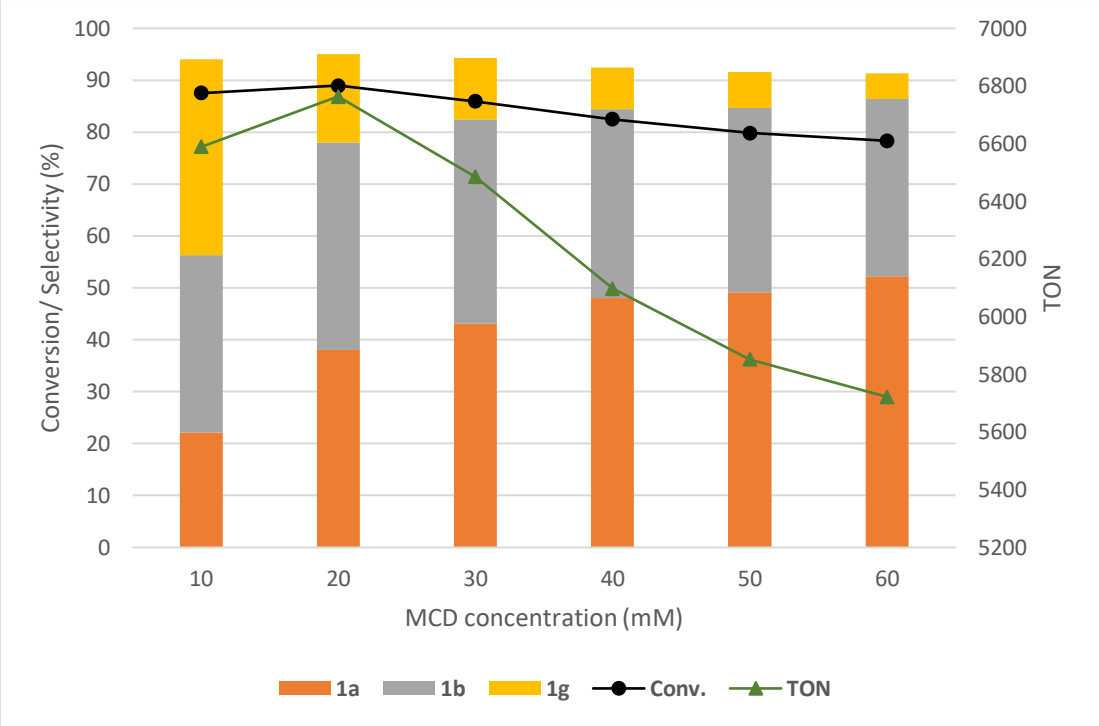


Figure 5.3 The conversion, TON and product selectivity of the R19/F87I/A82M/A184I/A330I/L171H/V178F variant at 8000:1 S:E ratio (8 mM bulnesene to 1µM enzyme) with different MCD concentration.

In the initial screening, MCD was added at 10 mM concentration in a reaction system of 2 mM bulnesene and 2 µM CYP102A1 at 0.5 mL scale. The substrate concentration was increased to 8 mM and different MCD concentrations (10 mM to 60 mM) were tested to determine whether the concentration of MCD had to be increased accordingly (Fig. 5.3).

The effect of MCD concentration on conversion was relatively minor. The conversion decreased slightly as the MCD concentration increased. However, because of the high bulnesene concentration (8 mM), a small change in conversion and selectivity would have a huge impact on the TON. The highest TON (6770) was achieved at 20 mM MCD, which was the MCD concentration used in the initial screening of 8 mM bulnesene, followed by 6590 at 10 mM.

Interestingly, despite the similarly in the total C15 selectivity for all reactions, the selectivity of individual products varied. The selectivity for bulnesene-15-COOH **1g** was the highest (38%) at 10 mM MCD and decreased significantly as the MCD concentration increased. On the other hand, the

proportion of bulnesene-15-ol **1a** increased with increased MCD concentration. This could be because the oxidation of polar compounds like C15-alcohol **1a** and C15-aldehyde **1b** are more probable at lower MCD concentration while higher MCD concentration promotes the oxidation of bulnesene up to a certain level. The MCD concentration could potentially be reduced further to increase the **1g** selectivity without sacrificing the conversion of bulnesene.

5.2.2 Preparative scale reaction at higher titre with aeration

Although selective synthesis of **1g** (~70%) was achieved, the reaction was carried out at 1000:1 S:E ratio (TON for **1g** formation: 700). A reaction at a S:E ratio of 8000:1 was performed with the most active variant, R19/F87I/A82M/A184I/A330I/L171H/V178F, with aeration at 25 mL scale by bubbling air through the reaction mixture to determine if aeration would improve the activity for higher TON synthesis of **1g** at high chemoselectivity. The MCD concentration was 10 mM which was found to promote **1g** synthesis at 8000:1 ratio (Fig. 5.3). A parallel reaction with the flask being shaken without aeration was set up as control.

The reactions were almost identical for the first 6 hours, with the conversion of bulnesene at approximately 40% (Fig. 5.4). The product profiles were very similar, with an approximate 1:2:1 ratio of **1a**:**1b**:**1g** at 6 hours. The effect of aeration was more significant over a longer period; bulnesene was almost fully converted (97%) after 24 hours in the aerated reaction while the conversion was only about 83% for the other reaction. Interestingly, despite the increase in conversion, aeration did not change the product profile which could be a useful method in accelerating the reaction.

An aliquot of enzyme was added to the aerated reaction at 24 hours. Bulnesene-15-ol **1a** and bulnesene-15-al **1b** were slowly converted to the C15-carboxylic acid **1g**. The reaction was stopped at 48 hours when the selectivity of **1g** reached 72% with 15% of **1b** left as the further oxidation product **1h** had started to be formed. Starting with an 8000:1 S:E ratio, the conversion reached 97%

with 35% of **1g** in 24 hours. The addition of the second aliquot of enzyme, which changed the overall S:E ratio to 4000:1, fully converted bulnesene with 72% of **1g** formed. The reaction outcome was different for a 4000:1 reaction (Fig. 5.2) which gave a mixture of **1a**, **1b** and **1g**. High chemoselectivity synthesis of **1g**, which could only be achieved at 1000:1 ratio previously, could be performed at 4000:1 ratio with the addition of a second aliquot of enzyme.

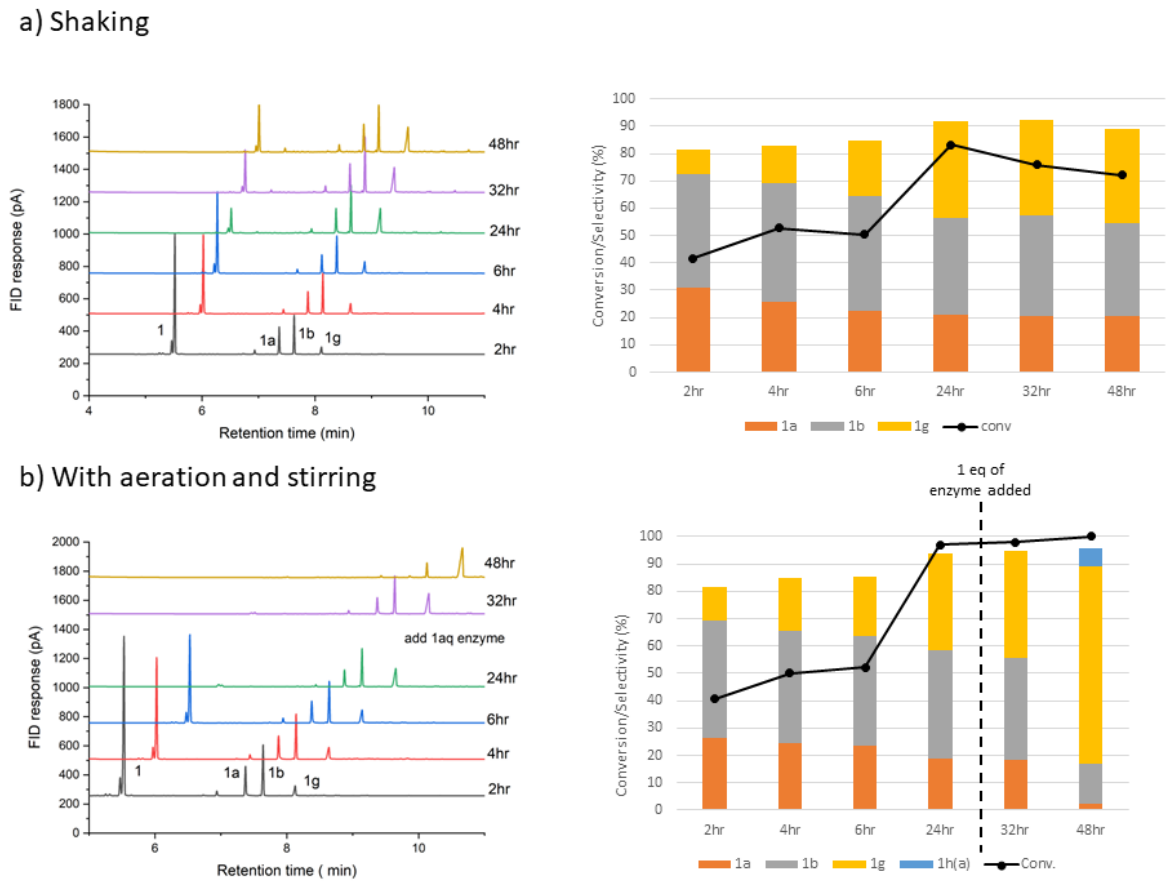


Figure 5.4 The time course of the reactions with variant R19/F87I/A82M/A184I/A330I/L171H/V187F at 8000:1 S:E ratio (8 mM bulnesene and 1 μM enzyme) with either (a) shaking (control) or (b) aeration and stirring.

5.3 Summary of the biosynthesis of agarwood aroma compounds

There are 12 guaiane skeleton compounds reported in literature, which were isolated and characterised from agarwood, including bulnesene and guaiene.[139,141] Two of these, rotundone and guaiol, are synthesised from guaiene, the others are the derivatives of bulnesene (Fig. 5.5).

Table 5.2 The highest chemoselectivity achieved for **1a**, **1b** and **1g** respectively at 8000:1 ratio. TON represent the turnover number for each target compound respectively. Asterisk (*) denoted that the S:E ratio in total for the whole system.

Compounds	Variants	Conv.	S:E ratio	Chemoselectivity			TON
				1a	1b	1g	
Bulnesene-15-ol, **1a**	R19/F87I	87	8000	**70%**	20%	–	4880
Bulnesene-15-al, **1b**	R19/F87I/V26M/A82M /A330I	82	8000	33%	**47%**	12%	3090
Bulnesene-15-COOH, **1g**	R19/F87I/A82M/A184I/ A330I/L171H/V178F	100	4000*	2%	15%	**72%**	2880

Although many variants oxidised bulnesene to **1a**, **1b** and **1g**, effective systems have to be developed through a combination of enzyme engineering and reaction engineering to synthesise these compounds more efficiently, with high turnover number, S:E ratio, regioselectivity and chemoselectivity. As a proof-of-concept, the highest S:E ratio tested was 8000:1 in this work but the potential of the system is huge as the variants were highly active and selective.

The highest **1a** chemoselectivity was 70 – 75% throughout the enzyme library. Variants which gave high **1a** selectivity at a 1000:1 S:E ratio were less active and had low conversion as the S:E ratio and substrate concentration was increased. The apparent high proportion of C15-alcohol **1a** might simply reflect the low bulnesene oxidation activity such that further oxidation to the aldehyde and beyond did not occur. With high activity variants, on the other hand, the proportion of **1a** increased with higher S:E ratio, likely reflecting slow oxidation of **1a** and **1b** until more bulnesene had been consumed. Therefore, higher activity variants are better options for the synthesis of **1a** if high **1a** chemoselectivity could be achieved. Highest TON for bulnesene-15-ol **1a** was 4880 found with

R19/F87I at 70% selectivity (Table 5.2), which makes this variant a good candidate for the biocatalytic synthesis of **1a**.

Chemoselectivity control for bulnesene-15-al **1b** was much more difficult and the highest selectivity achieved was 48% by variant R19/F87I/V26L at a 1000:1 S:E ratio. As observed with selective oxidation of bulnesene to C15-alcohol **1a**, variants that showed high activity for the direct oxidation of bulnesene to derivatives of the C15-acid, formed C15-alcohol and C15-aldehyde as major products at high S:E ratios. Therefore, an alternative would be using a higher activity variant to oxidise bulnesene to a mixture of **1a** and **1b**. For example, similar **1b** chemoselectivity (47%) was demonstrated by variant R19/F87I/V26M/A82M/A330I at 8000:1 ratio, with a reasonable TON for **1b** of 3090 (Table 5.2). Some C15-acid was produced (12%) but selective synthesis of aldehyde **1b** could be achieved by oxidising **1a** to **1b** with a CRED in a cascade reaction. Assuming **1a** could be fully converted to **1b** with a CRED, the total **1b** selectivity would be 80% with a hypothetical TON of 5250.

High chemoselectivity for **1g** could be achieved by oxidising bulnesene with Group C variants such as R19/F87I/A184I (72%) and R19/F87I/A184I/A330V (78%). However, such selective synthesis could only be achieved at a 1000:1 S:E ratio. At higher ratios, more of the precursors **1a** and **1b** were produced, likely reflecting the slow oxidation of either the alcohol **1a** or aldehyde **1b**. A high **1g** turnover number synthesis was demonstrated with variant R19/F87I/A82M/A184I/A330I/L171H/V178F, which at screening scale gave the 11,12-epoxide of the C15-acid (**1h**), by adding the enzyme in aliquots. For a 8000:1 reaction, two aliquots of enzyme were added after the activity of the first was observed to decrease significantly (4000:1 overall S:E ratio). After 48 hours, bulnesene was fully converted with **1g** as the major product (72%) and a TON for **1g** of 2880.

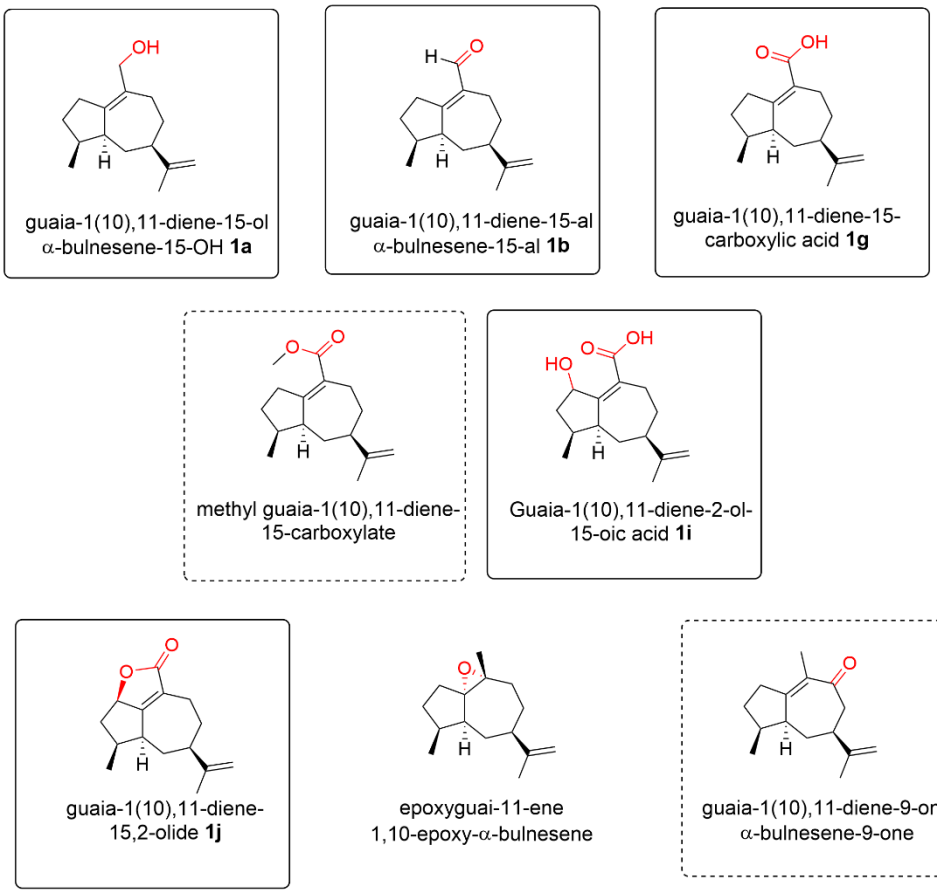


Figure 5.5 Guaiene skeleton compounds in agarwood derived from bulnesene. The solid boxes represent the compounds which have been successfully synthesised from bulnesene via engineered CYP102A1 variants. Dashed boxes represent compounds which could be synthesised with the metabolites of bulnesene as precursors.

The biocatalytic cascade reaction for synthesising 15,2-lactone **1j** produced this highly desirable agarwood aroma compound at roughly 30% yield based on the amount of bulnesene converted. The **1j** produced had high purity as all the side products decomposed under acid and heat treatment for lactonisation. Therefore, no further purification step was required which greatly simplified the process. The precursor of **1j**, 2-hydroxy-bulnesene-15-COOH (**1i**) was isolated and characterised before being cyclised to form **1j**, but the yield was poor due to the inefficient extraction of the polar compound from aqueous solution.

Among the eight bulnesene-derived agarwood compounds (Fig. 5.5), five have been successfully synthesised from bulnesene with engineered CYP102A1 variants. Both methyl guai-1(10),11-diene-15-carboxylate (methylbulnesene-15-carboxylate) and bulnesene-9-one could be synthesised from C15-acid **1g** and bulnesene-9-ol **1f** respectively. For example, 9-alcohol **1f** could be oxidised to bulnesene-9-one with a CRED while the methylbulnesene-15-carboxylate could be synthesised via Fischer esterification with methanol under acidic conditions. Only 1,10-epoxy-guai-11-ene has not been observed with the variant library, which could be because the tetra-substituted C1=C10 double bond is more hindered. Interestingly, the 11,12-epoxide of the C15-alcohol (**1e**), which is a common metabolite of bulnesene by the engineered CYP102A1 variants, has not been found in agarwood. Nevertheless, this novel compound is still of interest for its potential flavour, fragrance and biological properties.

Chapter 6

6 Guaiene oxidation by engineered CYP102A1

6.1 Introduction

One of the guaiane skeleton molecules found in agarwood, rotundone, was identified by Wood *et al*. as an important compound contributing to the peppercorn aroma in many different kinds of herbs, especially pepper and also grapes.[152] First discovered in Australian Shiraz wine, the special peppercorn aroma in the wine was attributable to rotundone. Due to its very low odour threshold level (8 ng/L in water and 16 ng/L in wine), the flavour of the beverage would be markedly changed even with the addition of a small amount of rotundone. α-Guaiene, a sesquiterpene found mainly in patchouli oil and also agarwood, has been proposed to be the precursor of rotundone. Huang *et al*. suggested that the formation of rotundone is the consequence of air contact of α-guaiene in plants rather than by biosynthesis through a specific enzymatic pathway.[173] They demonstrated the formation of rotundone by aerial oxidation of α-guaiene on cellulose filter paper. However, CYP71BE5 was recently discovered as the α-guaiene oxidase in grapevine (*Vitis vinifera*).[174] There is continuing interest in chemical oxidation of the sesquiterpene. Ishihara demonstrated the production of rotundone from α-guaiene by a 2-step organic synthesis.[151] A biomimetic catalyst, consisting of a non-heme Fe^{2+}-chelate catalyst and ferric-chelate reductase, has been developed to synthesise rotundone and other oxygenated terpenoids.[175] However, none of these has high enough selectivity and yield for producing rotundone as a food additive.

The subject of this chapter is to develop a biocatalytic synthesis of rotundone via C2 oxidation of α-guaiene, the sesquiterpene precursor, by engineered CYP102A1 variants. Moreover, all guaiene oxidation products are of interest because of their potential flavour, fragrance and biological properties, including medicinal effects.

6.2 Oxidation of α-guaiene by CYP102A1

6.2.1 Initial library screening

α-Guaiene, **2**, was first screened for oxidation with a library of 44 CYP102A1 variants in 24-well plates, shaking at 20 °C and 120 rpm. The 500 µL screening-scale reactions were carried out in 200 mM phosphate buffer (pH 7.9) with 2 mM of substrate (0.4 M ethanol stock) and 2 µM of the CYP102A1 variant. Glucose (100 mM) and glucose dehydrogenase (4 U mL^{-1}) were used to regenerate the cofactor NADPH (40 µM). Methyl-β-cyclodextrin was added to 10 mM to increase the solubility of guaiene in buffer. The organics were extracted with ethyl acetate after 6 hours and analysed by GC. Full activity and selectivity data are given in Appendix 4.

The initial screening library of variants (Table 6.1) were similar to those for bulnesene oxidation screening. This panel of variants showed good activity for α-guaiene oxidation, with 35 variants giving more than 50% conversion (TON >500). Screening reactions that showed high activity were selected for preparative scale reactions, followed by column chromatography for product purification. Five major products were isolated and characterised as 11,12-epoxy-α-guaiene, **2a**, (2*S*)-hydroxy-α-guaiene (rotundol), **2b**, (9*S*)-hydroxy-α-guaiene, **2c**, 2-hydroxy-11,12-epoxy-α-guaiene, **2d**, and (*2R*,*9S*)-dihydroxy-α-guaiene, **2e** (Fig. 6.1).

In the ^{1}H NMR spectrum of **2a**, the resonances of the C12 methylene protons (4.65 ppm) in α-guaiene were absent while a two-proton multiplet appeared at 2.59 ppm, indicating epoxidation of the C11=C12 double bond to form 11,12-epoxy-α-guaiene. The resonance of the C2 allylic proton of 2-hydroxy-α-guaiene (**2b**) had the same chemical shift as the C12 methylene protons at 4.66 ppm, and the configuration at C2 was assigned as (*S*) by comparison with reported data in the literature.[156] Compared to the allylic C2-alcohol **2b**, the C<u>H</u>(OH) resonance of the C9-alcohol **2c** was at higher field (3.83 ppm) and the configuration of C9 was assigned as (*S*) by the nOe data.

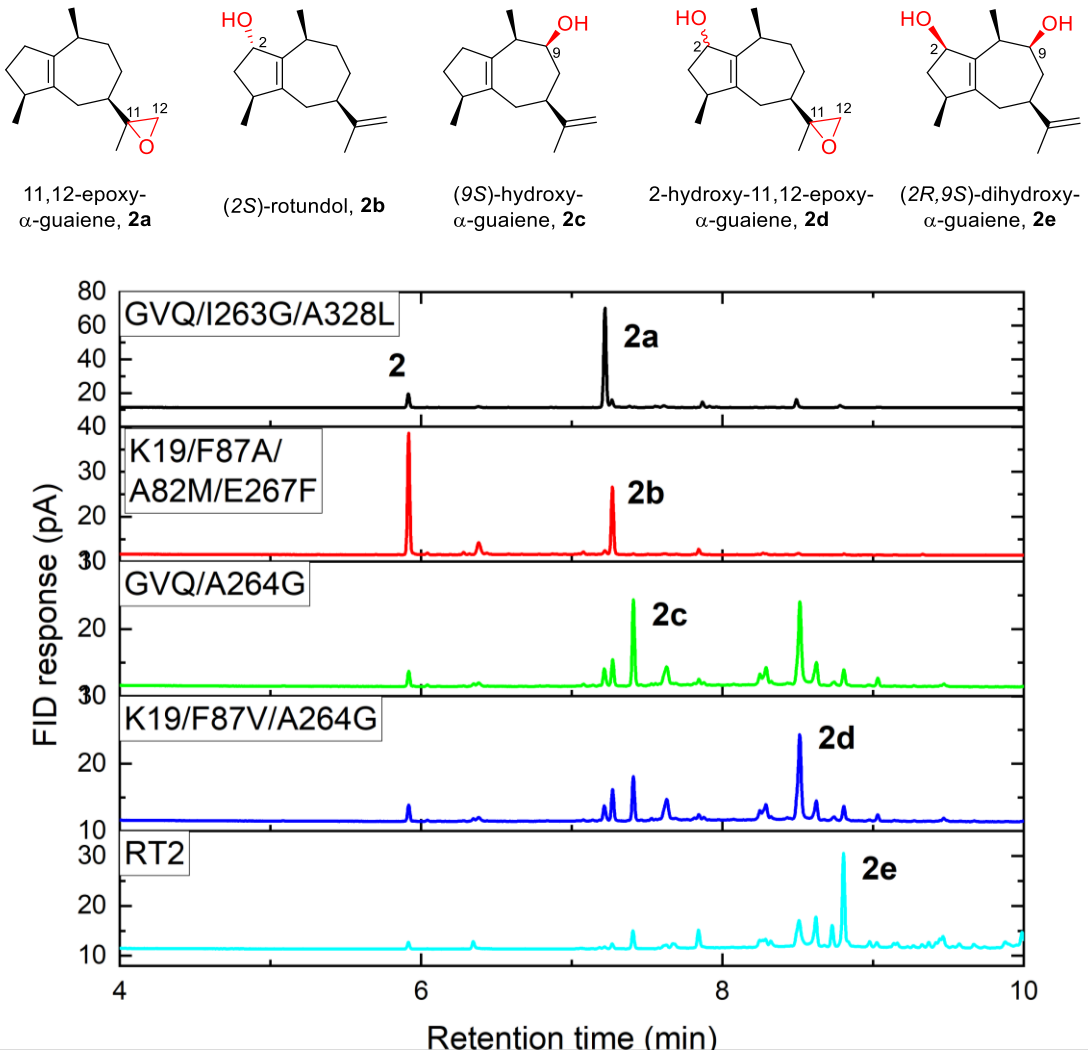


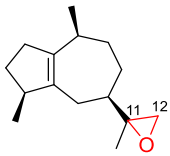
Figure 6.1 GC analysis of the highest selectivity variant from the screening library of each product giving α-guaiene oxidation products **2a** to **2e**.

As observed in the ^{1}H NMR spectrum of **2a**, the exocyclic C12 methylene protons of **2d** shifted upfield to 2.60 ppm and a new resonance was observed at 4.65 ppm which corresponded to an allylic alcohol group at the C2 position. Hence, product **2d** was characterised as 2-hydroxy-11,12-epoxy-α-guaiene. Two resonances were found at 4.54 ppm and 3.77 ppm in the ^{1}H NMR spectrum of **2e** which demonstrated that the product had an allylic as well as an aliphatic alcohol group; the 2D NMR data identified **2e** as 2,9-dihydroxy-α-guaiene. The configuration of C2 and C9 were assigned as (*2R*) and (*9S*) respectively by nOe data.

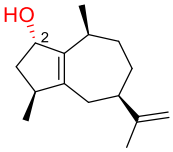
The target compound, rotundone, was not observed from this initial screening panel of enzymes. On the other hand, many variants oxidised α-guaiene to 2-hydroxy-α-guaiene (rotundol, **2b**), the

precursor to rotundone, with reasonable conversion and selectivity (Table 6.1). Nine out of 44 variants had >40% selectivity for rotundol. The highest selectivity of 68% was observed with the variant K19/F87A/A82M/E267F but at low conversion (45%) and TON (310, Table 6.1, Entry 1), followed by variant RP/A82M/I263A (58%) at 28% conversion (Entry 2). High turnover numbers of **2b** were shown by the RT2-based variants, RT2/S72W/A330W (515, Entry 3) and RT2/S72G/A330W (490, Entry 4).

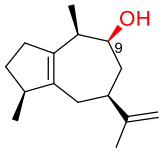
Table 6.1 Substrate conversion and product selectivity of selected CYP102A1 variants for the oxidation of α-guaiene, **2**. The selectivity for the most selective variants for each product is in red. TON refers to the yield of the product in bold relative to enzyme.



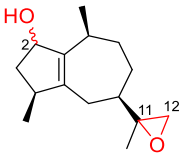
11,12-epoxy-α-guaiene, **2a**



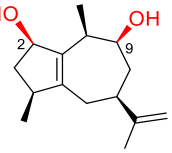
(2*S*)-rotundol, **2b**



(9*S*)-hydroxy-α-guaiene, **2c**



2-hydroxy-11,12-epoxy-α-guaiene, **2d**



(2*R*,9*S*)-dihydroxy-α-guaiene, **2e**

Entry	CYP102A1 Variants	Conv.	2a	2b	2c	2d	2e	TON
1	K19/F87A/A82M/E267F	45%	4%	**68%**	–	–	–	310
2	RP/A82M/I263A	28%	7%	**58%**	–	–	–	165
3	RT2/S72W/A330W	90%	10%	**57%**	–	23%	–	515
4	RT2/S72G/A330W	98%	5%	**50%**	–	14%	–	490
5	GVQ /I263G/A328L	90%	**78%**	6%	–	7%	–	705
6	GVQ/A264G	96%	5%	7%	**22%**	33%	5%	215
7	R19/F87I	75%	17%	31%	**21%**	6%	–	160
8	GVQ/I263G	94%	12%	7%	–	**45%**	1%	425
9	K19/F87V/A264G	96%	5%	9%	13%	**38%**	5%	365
10	RT2	100%	–	–	5%	11%	**28%**	280
11	GV/A184I	97%	–	–	5%	11%	**27%**	265
12	GVQ	74%	9%	6%	–	**15%**	–	115

The 11,12-epoxide **2a** was another common product. High selectivity was observed with variant GVQ/I263G/A328L (78% of **2a**) and with high conversion (90%) and TON (705, Entry 5). The mutations I263G and A328L were important in promoting the formation of **2a** as the template variant GVQ gave a range of different products including **2a** and **2b** (Entry 12).

C9 is an unactivated aliphatic carbon and requires specific binding orientation(s) that position this C–H bond close to the ferryl oxygen so that oxidation of the less reactive sp^3 C–H bond can compete with oxidation at other, more activated C–H bonds which are also accessible to the ferryl oxygen. Therefore, selective oxidation of the C9 position is more challenging; there were only four variants that gave >20% of C9-alcohol **2c**. The variant GVQ/A264G gave the highest selectivity for **2c** (22%) with 2-hydroxy-11,12-epoxy-α-guaiene **2d** as another major product (33%, Entry 6).

2-Hydroxy-11,12-epoxy-α-guaiene (**2d**) could be the further oxidation product of either the C2-alcohol or the 11,12-epoxide. Variant GVQ/I263G showed the highest **2d** selectivity at 45% with reasonable TON (425, Entry 8). Interestingly, variants K19/F87V/A264G and GVQ/A264G, both with the F87V/A264G combination of mutations, gave reasonable selectivity for both **2c** and **2d** with almost full conversion (Entries 9 and 6). Variants RT2 and GV/A184I gave the highest proportions of 2,9-dihydroxy-α-guaiene (**2e**) which might be the further oxidation product of the C2- or C9-alcohol (Entries 10 and 11). It was also notable that the RT2 template variant showed total conversion but gave a mixture of products including the further oxidation products **2d** and **2e**. Therefore, in addition to altering the regioselectivity of bulnesene oxidation, the A330W and S72G/W mutations also retarded the further oxidation of the initial oxidation products (Entries 3 and 4).

6.2.2 Sub-library screening of guaiene

The initial screening of α-guaiene with 44 CYP102A1 variants identified residues and mutations that promoted the oxidation of the sesquiterpene. Variants with high rotundol (**2b**) selectivity mainly had K19/F87A/A82M, RP and RT2 as base variants while the highest selectivity variants for the epoxide **2a** and C9-alcohol **2c** were GVQ-based. Therefore, CYP102A1 variants with mutations added to the K19/F87A/A82M, RP, RT2 and GVQ templates within the main library of approximately 500 enzymes previously generated within the group were assembled into four new sub-libraries and their guaiene oxidation activity was investigated.

The K19/F87A/A82M panel (A82M/F87A/H171L/Q307H/N319Y)

The K19/F87A/A82M base variant had much lower conversion (13%) but higher selectivity for **2b** (78%, Table 6.2, Entry 2) than K19/F87A/A82M/E267F, the most selective variant from the initial screening library (68%, Entry 1). The sub-library based on K19/F87A/A82M mainly had mutations of I-helix residues – E267 is also in the I helix. Significantly increased conversion was observed for variant K19/F87A/A82M/T260G (75%) with a slight decrease in **2b** selectivity to 62% (Entry 3). The TON increased from 310 for the K19/F87A/A82M/E267F variant to 465 (Entries 1 and 3).

The RP panel (R47L/Y51F/I401P)

The RP-based sub-library was selected because the RP/A82M/I263A variant showed good selectivity for **2b** (58%) despite the low conversion of 28% (Entry 4). The C2-alcohol was the major product of variants containing the A82M, I263A/G and A330W mutations (Entries 4–8). Two variants showed higher C2 selectivity than K19/F87A/A82M/E267F – 80% **2b** for RP/A82M/I263G at 36% conversion; 73% for RP/A82M/H171L/I263G at 58% conversion (Entries 5 and 6). Although the RP-based variants had the activity-promoting I401P mutation, the H171L mutation was important for increased conversion (Entry 5 *vs*. Entry 6) while the I263G mutation was effective in enhancing both conversion and C2 selectivity (Entry 4 *vs*. Entry 5). Overall, however, the RP-based variants showed lower conversion, and thus lower TON, and were less active than the K19/F87A/A82M/T260G variant for rotundol formation.

The RT2 Panel (R47L/Y51F/A191T/N239H/I259V/A276T/L353I)

The RT2 template variant showed 100% conversion but gave mostly further oxidation products (Table 6.2, Entry 19). Variants RT2/S72G/A330W and RT2/S72W/A330W showed reasonable selectivity for **2b** (~50%) but the high conversions led to higher TONs (~500) for **2b** formation (Entries 9 and 10). Variants with different mutation combinations added to the RT2 template and

different mutations of S72 on the RT2/S72X/A330W template were selected from the larger enzyme library and screened for guaiene oxidation.

RT2 gave the highest proportion of the 2,9-diol **2e** (28%) from the initial library, which surprisingly was not observed in the RT2 sub-library (Entry 19). The high conversion of the RT2 template highlighted the effect of additional active site mutations. The A330W mutation lowered the conversion (70%) and hindered the further oxidation, forming 21% of 11,12-epoxide **2a** and 48% of rotundol **2b** (Entry 16). The RT2/L437F variant in the new sub-library showed 100% conversion and gave 57% of **2b** (TON=570) while variant RT2/I263G gave more of the 2-hydroxy-11,12-epoxy-guaiene **2d** (59%, Entries 17 and 18). The effect of different mutations of S72 in the RT2/S72X/A330W series of variants were also of interest; mutation of S72 to the neutral or polar residues Gly, Val, His, Tyr and Trp increased both the conversion (>90%) and the rotundol selectivity at ~60% (Entries 9–13) compared to variant RT2/A330W (Entry 16). Variant RT2/S72V/A330W showed, by a small margin, the highest rotundol selectivity (63%) and TON (595, Entry 11). Both activity and selectivity were lost with the S72R mutation (Entry 14). Addition of the L437LA insertion mutation to the S72G/A330W combination increased the selectivity for **2b** from 50% to 64% but the conversion was lowered to 81% (Entries 10 and 15).

Table 6.2 Substrate conversion and product selectivity of selected variants for the oxidation of α-guaiene, **2**.

Entry	CYP102A1 variants	Conv.	2a	2b	2c	2d	2e	TON
1	*K19/F87A/A82M/E267F	45%	4%	**68%**	–	–	–	310
2	K19/F87A/A82M	13%	–	**78%**	–	–	–	105
3	K19/F87A/A82M/T260G	75%	–	**62%**	–	3%	–	465
4	*RP/A82M/I263A	28%	7%	**58%**	–	–	–	165
5	RP/A82M/I263G	36%	6%	**80%**	–	–	–	290
6	RP/A82M/H171L/I263G	58%	5%	**73%**	–	5%	–	425
7	RP/A82M/I263A/A330W	42%	5%	**68%**	–	–	–	290
8	RP/A82M/A330W	35%	5%	**66%**	–	–	–	235
9	*RT2/S72W/A330W	90%	10%	**57%**	–	23%	–	515
10	*RT2/S72G/A330W	98%	5%	**50%**	–	14%	–	490
11	RT2/S72V/A330W	94%	11%	**63%**	–	16%	–	595
12	RT2/S72Y/A330W	94%	15%	**59%**	–	18%	–	555
13	RT2/S72H/A330W	97%	7%	**59%**	–	9%	–	575
14	RT2/S72R/A330W	23%	–	**22%**	–	–	–	55
15	RT2/S72G/A330W/L437LA	81%	19%	**64%**	–	12%	–	520
16	RT2/A330W	70%	21%	**48%**	–	4%	–	340
17	RT2/L437F	100%	2%	**57%**	–	37%	–	570
18	RT2/I263G	100%	–	4%	–	**59%**	–	590
19	*RT2	100%	–	–	5%	11%	**28%**	280
20	*GVQ /I263G/A328L	90%	**78%**	6%	–	7%	–	705
21	GVQ/I263G/A328I	100%	10%	6%	–	**76%**	2%	760
22	GLQ/I263G	96%	28%	**52%**	–	17%	–	500
23	GLQ/I263G/A328G	100%	7%	**48%**	–	16%	7%	480
24	GLQ/I263G/A328I	100%	8%	**45%**	–	43%	–	450
25	*GVQ	74%	9%	6%	–	**15%**	–	115
26	*GVQ/A264G	96%	5%	7%	**22%**	33%	5%	215
27	GVQ/A328G	90%	2%	16%	**31%**	7%	11%	280

Screening scale reactions were in 200 mM phosphate buffer containing 2 μM CYP102A1, 2 mM α-guaiene, 40 μM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 6 hours. The selectivity for the most selective variants of each product is in red. TON refers to the yield of the products in bold relative to enzyme. *: variants from previous library.

*The GVQ panel (*A74G/F87V/L188Q)

Variants with the highest selectivity for 11,12-epoxide **2a** and the C9-alcohol **2c** were based on the GVQ template variant. Therefore, the GVQ sub-library was assembled with variants of different hydrophobic substitutions at residues 87 and 328. GVQ/I263G/A328L remained the variant that gave the highest selectivity for **2a** (78%, Table 6.2, Entry 20). Some variants including GVQ/I263G/A328I had significantly improved activity and gave the further oxidation product 2-hydroxy-11,12-epoxy-guaiene, **2d**, as major product (76%, Entry 21). The regioselectivity was

switched to mainly C2 oxidation when F87 was mutated to leucine; the GLQ/I263G variant had reasonable selectivity for rotundol **2b** (52%) followed by GLQ/I263G/A328G and GLQ/I263G/A328I at 48% and 45% respectively (Entries 22–24). Variant GVQ/A328G produced the most C9-alcohol **2c** at 31% among all the variants screened, slightly higher compared to GVQ/A264G (22%) from the initial screening (Entries 26–27).

6.2.3 Summary

In this chapter, a screening library was firstly used to identify template variants, residues and mutations that affect guaiene oxidation activity and product selectivity. This was followed by a selection of variants within the existing, extensive enzyme library that contained other mutations at key residues for further screening (Fig. 6.2). This approach could reduce the time and effort required to engineer some desired activity.

It was found that variants of the templates K19/F87A/A82M, RP and RT2 gave (*2S*)-alcohol **2b** as the major product with the highest **2b** selectivity achieved mainly by RP-variants. However, high TONs of **2b** were generally observed with RT2-based variants because of their high conversion (>90%). The TON of **2b** was increased to 595 at 94% conversion with variant RT2/S72V/A330W (Table 6.2, Entry 11), compared to 515 by variant RT2/S72W/A330W from the initial library (Entry 9). Only slight increases in **2c** selectivity were observed by GVQ/A328G variant at 31% (Entry 27). The highest selectivity for 11,12-epoxide **2a** remained the 78% observed with the GVQ/I263G/A328L variant from the initial library (Entry 20). However, increased activity was observed with variant GVQ/I263G/A328I from GVQ panel with the further oxidation product, 2-hydroxy-11,12-epoxy-guaiene **2d** as the major product at 76% (Entry 21).

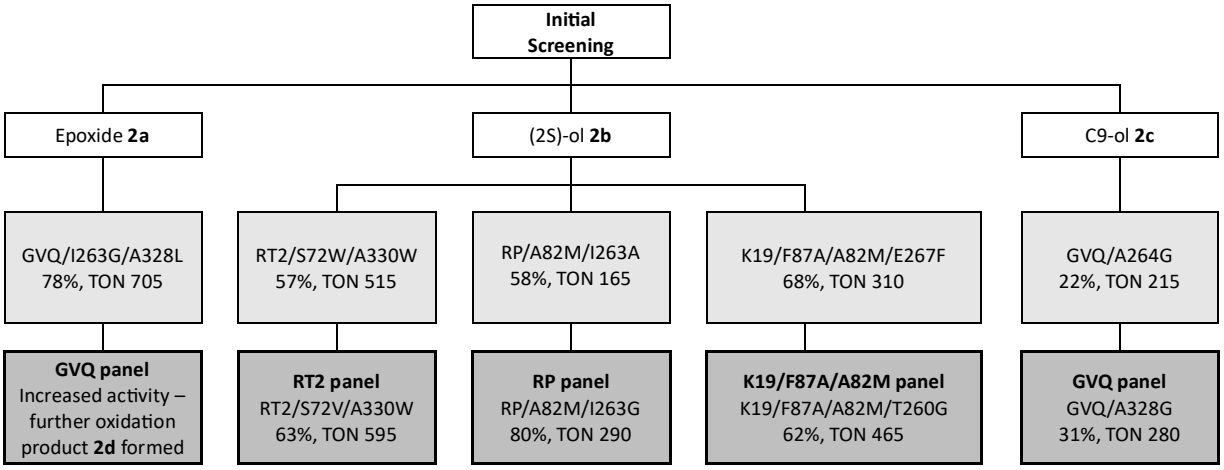


Figure 6.2 The progressive library screening for each monooxygenated product of guaiene **2** starting from the initial library. Light grey boxes represent variants from the initial library while dark grey boxes represent variants from each sub-library, with the name of library in bold.

The major products **2a** – **2e** arose from the oxidation at the activated allylic C2 position, the C11=C12 double bond and the unactivated aliphatic C9 position. It was notable that no new product was formed in any significant proportions from these rounds of screening and variant selection. This suggested that even the larger library was not sufficiently diverse and thus no variant possessed the active site architecture to enable oxidation at other positions. However, the selectivity patterns suggested possible mutation combinations to increase selectivity for rotundol, for example, by adding the A82M and L437F mutations to the RT2/S72V/A330W variant. Although time constraints precluded exploration of these possibilities, the screening approach provided variants with good selectivity for rotundol and suggested mutations for improving this further.

6.3 Biocatalytic synthesis of rotundone

6.3.1 Biocatalytic cascade synthesis of rotundone

Rotundone had not been isolated from the libraries, suggesting that the CYP102A1 variants did not oxidise rotundol **2b** to the ketone. An alternative to direct P450-catalysed oxidation of guaiene to rotundone would be a two-step reaction – α-guaiene oxidation to rotundol by a CYP102A1 variant followed by oxidation of rotundol to rotundone by a carbonyl reductase (CRED, Fig. 6.3).

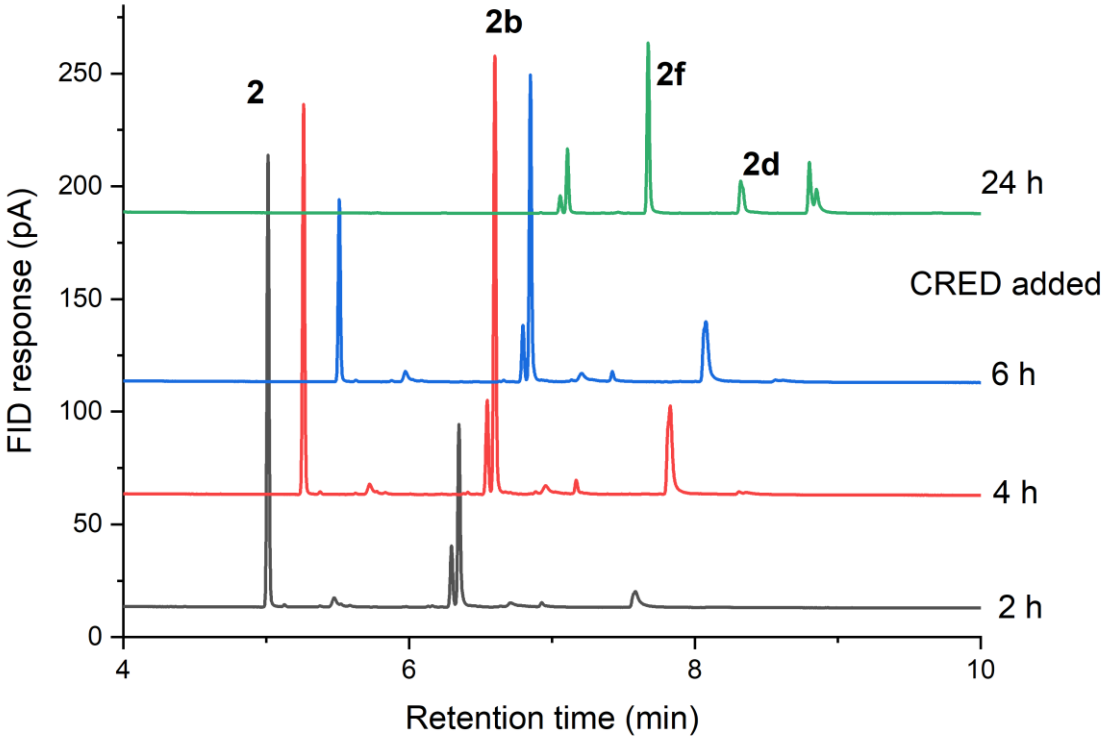


Figure 6.3 Time-course reaction of the synthesis of rotundone **2f** with guaiene **2**.

A 100 mL scale reaction was carried out with the RT2/S72G/A330W/L437LA variant because of its good selectivity (64%) for rotundol and high conversion (81%, Table 6.2, Entry 15). Similar to the reaction discussed in Chapter 3 for the cascade oxidation of bulnesene (bulnesene-15-ol **1a** to bulnesene-15-al **1b** by CRED), the amount of glucose used in the reaction was optimised (25 mM) to maximise the performance of CRED without influencing the reaction of CYP102A1. Guaiene was roughly 76% converted in 6 hours, with rotundol **2b** as the major product (55%). The pH of the reaction mixture was adjusted to 8.5 and the temperature was raised to 37 °C before adding CRED (100 mg). Most of the rotundol was converted to a new product **2f** in 24 hours.

The new product **2f** was isolated by column chromatography and characterised by NMR and MS data. A peak at 219 m/z corresponding to $(M + 14 + H)^+$ was observed in the mass spectrum of product **2f** which indicated that the product was a ketone. The presence of a characteristic peak at

208 ppm of ^{13}C NMR and the absence of resonances of C2 protons demonstrated that C2 had been oxidised to a ketone and product **2f** was confirmed to be the target compound, rotundone.[151]

6.3.2 A single CYP102A1 variant for direct oxidation of guaiene to rotundone

Once the peak corresponding to rotundone had been identified on the GC, the screening data was checked to determine if there were any variants that gave rotundone directly from guaiene. Interestingly, only variants from the K19/F87A/A82M sub-library gave TON>100 for rotundone formation. The most active variants were K19/F87A/A82M/I263G/A328G and K19/F87A/A82M/A264G/A328G (TON>160, Table 6.3, Entries 2 and 3). However, the combination variant K19/F87A/A82M/I263G/A264G/A328G showed lower activity (TON=130, Table 6.3, Entry 4). As all the variants that gave higher TON (>100) for rotundone had different combinations of glycine mutations on residues I263, A264 and A328, new variants with these mutations on different template variants were tested for any increase in the TON of rotundone.

Although certain mutation combinations on the GV/A184I, R19/F87A/A184I and GVQ template variants gave reasonable TON (>100) for rotundone formation, variants based on K19/F87A/A82M still produced the most rotundone and the K19/F87A/A82M/I263G/A328G variant remained the most active (Table 6.3, Entry 2).

Table 6.3 Substrate conversion and product selectivity of selected variants for the oxidation of α-guaiene, **2**.

Entry	CYP102A1 variants	Conv.	2a	2b	2c	2d	2e	2f	TON
1	K19/F87A/A82M/I263G/A264G	89%	6%	35%	–	6%	–	**11%**	100
2	K19/F87A/A82M/I263G/A328G	88%	–	31%	–	–	–	**19%**	170
3	K19/F87A/A82M/A264G/A328G	90%	–	23%	–	5%	–	**18%**	165
4	K19/F87A/A82M/I263G/A264G /A328G	79%	5%	32%	–	5%	–	**16%**	130
5	GV/A184I/I263G/A264G	92%	16%	23%	–	15%	–	**6%**	55
6	GV/A184I/I263G/A328G	90%	9%	20%	–	7%	–	**11%**	100
7	GV/A184I/A264G/A328G	73%	4%	24%	7%	8%	–	**6%**	45
8	GV/A184I/I263G/A264G/A328G	73%	7%	21%	–	4%	4%	–	–
9	R19/F87A/A184I/I263G/A264G	91%	5%	18%	–	7%	–	**12%**	110
10	R19/F87A/A184I/I263G/A328G	95%	4%	15%	–	16%	–	**3%**	30
11	R19/F87A/A184I/A264G/A328G	69%	2%	23%	–	3%	–	**13%**	90
12	R19/F87A/A184I/I263G/A264G /A328G	79%	6%	18%	–	11%	–	**5%**	40
13	GVQ/I263G/A264G	90%	22%	21%	–	11%	–	**5%**	45
14	GVQ/I263G/A328G	78%	17%	44%	15%	4%	–	–	–
15	GVQ/I263G/A264G/A328G	80%	14%	32%	–	5%	–	**14%**	115

Screening scale reactions were in 200 mM phosphate buffer containing 2 μM CYP102A1, 2 mM α-guaiene, 40 μM $NADP^+$, 100 mM glucose, 4 U mL^{-1} GDH and 10 mM methyl-β-cyclodextrin. Plates were shaken at 120 rpm at 20 °C for 6 hours. TON refers to the yield of **2f** relative to enzyme.

It was notable that most variants with reasonable rotundone selectivity also gave good rotundol **2b** selectivity (>20%) which demonstrated that longer reaction time might increase **2f** selectivity by promoting the further oxidation of rotundol **2b** (Table 6.3). As varying the template variants did not bring any significant increase in the TON for **2f**, further mutagenesis could target other residues on the I helix and β_1 strand. Due to time constraints, further mutagenesis was not pursued. However, the CYP–CRED cascade would still provide an effective route to synthesis rotundone.

6.4 Summary

As the engineered CYP102A1 variants from the existing library showed low selectivity for rotundone, an alternative route to direct P450-catalysed oxidation of guaiene to rotundone was developed, i.e., α-guaiene oxidation to rotundol by a CYP102A1 variant followed by oxidation of rotundol to rotundone by a carbonyl reductase (CRED). The initial library of 44 variants identified

important template variants and mutations for guaiene oxidation. This was followed by the assembly of focused sub-panels within the existing library. Significant increase in TON of rotundol was observed by screening of an additional 49 variants in the extended libraries. The selectivity trends observed from the progressive screening also suggested possible mutation combinations in guiding the further mutagenesis to increase the selectivity for rotundol.

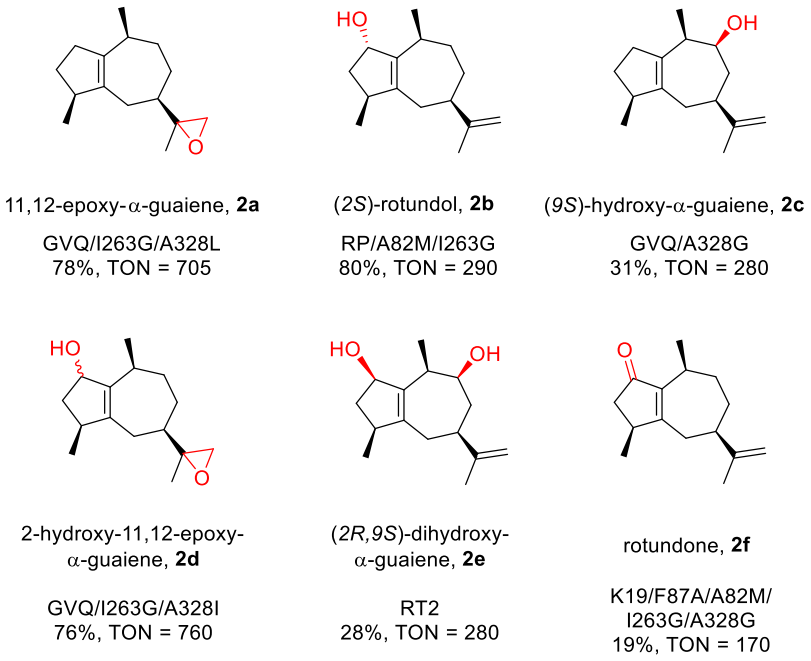


Figure 6.4 Highest selectivity and TON achieved for all oxidation products of α-guaiene and relevant variants developed by progressive library screening.

Chapter 7

7 Conclusion

The main goal of this thesis was to synthesise some of the desirable wood aroma compounds from agarwood by engineered variants of CYP102A1. The combined strategies of progressive screening library, docking-guided mutagenesis and reaction engineering were utilised to develop systems that produced the target compounds with high efficiency.

Chapter 3 demonstrated the use of CYP102A1 variants in the biosynthesis of the aroma compounds, bulnesene-15-ol, bulnesene-15-al and bulnesene-15-COOH. Variant R19/F87I from the initial library displayed good total C15 selectivity (81%) with full conversion, producing the target compounds as a mixture of bulnesene-15-ol and bulnsene-15-al. In addition, five other metabolites have also been characterised which are also of interest because of their potential flavour, fragrance and biological properties. The progressive screening approach improved the selectivity for each metabolite significantly with a small subset of variants. 'Hit' residues identified from the combined screening results were subjected to further mutagenesis. MD simulation and substrate docking were also applied to guide mutation design, targeting the residues which had not been targeted in the existing libraries and introducing mutations that might disfavour undesired binding poses. Further improvement of enzymatic activity and product selectivity were achieved with the combination of both approaches. The new-generation variants were highly active and could catalyse three cycles of bulnesene oxidation at the C15 position with high selectivity to form bulnesene-15-COOH.

The discovery of the highly selective variants for bulnesene-15-COOH provided a foundation to the biosynthesis of the 15,2-lactone. Most variants in the initial library gave very little activity for bulnesene-15-COOH oxidation, and 11,12-epoxy-bulnesene-15-COOH was the dominant product. Guided by progressive screening library, variants which oxidised C15-acid at the C2 position to produce 2-hydroxy-bulnesene-15-COOH with good selectivity were developed after two rounds of mutagenesis. The biocatalytic synthesis of bulnesene-15,2-lactone from bulnesene was established

by a one-pot cascade enzyme system, *i.e.*, a CYP102A1 variant enzyme was used to oxidise bulnesene at C15 to form the C15-acid which was then oxidised at the C2 position by a second CYP102A1 variant. Besides increasing the efficiency of the system, reaction engineering also promoted the decomposition of the side products, and the 15,2-lactone was isolated with high purity which would be beneficial towards further scaling-up of the reaction. The biosynthesis of this aroma compound was also demonstrated using a bulnesene-enriched distillation fraction of patchouli oil as feedstock. Despite a decrease in activity, the result was promising as it showed similar results as when purified bulnesene was used. Direct oxidation of bulnesene to the 15,2-lactone was attempted and a variant was developed with improved activity and some increase in 15,2-lactone selectivity. Further efforts are needed to establish optimised systems for the large-scale production of the aroma compounds from bulnesene-enriched patchouli oil and to develop the direct oxidation of bulnesene to 15,2-lactone with a single CYP102A1 variant.

Chapter 5 summarised the optimisation efforts in chemoselective α-bulnesene oxidation by reaction engineering. The substrate-to-enzyme ratio was found to be crucial for chemoselectivity control and effective systems were developed to synthesise the target compounds with high turnover number, substrate-to-enzyme ratio, regioselectivity and chemoselectivity. High chemoselectivity (>70%) syntheses of bulnesene-15-ol and bulnesene-15-COOH were demonstrated at TONs of 4880 and 2880 respectively. The chemoselectivity control of bulnesene-15-al was more challenging as it was maintained at a steady-state concentration. The highest chemoselectivity of bulnesene-15-al was 47% with a TON of 3090. Selective synthesis of this aldehyde could be achieved by oxidising the C15-alcohol to C15-aldehyde with a carbonyl reductase in a cascade reaction. High regioselectivity (~90%) and conversion (>80%) were achieved for all the optimised reactions.

Chapter 6 described the enzymatic synthesis of rotundone via α-guaiene oxidation by CYP102A1 variants. Significant increases in TON of all the monooxygenated products of guaiene were observed via the progressive screening library approach. As rotundone was only observed in small amounts

from the libraries, an alternative route was developed, *i.e.*, the biocatalytic cascade synthesis of rotundone where α-guaiene was oxidised to rotundol by a CYP102A1 variant followed by the oxidation of rotundol to rotundone by an alcohol dehydrogenase. Further mutagenesis could be designed based on the data from the progressive screening to generate variants with higher rotundol and rotundone selectivity for more efficient synthesis of the target compound.

Reference

1 C. G. Newton, S. G. Wang, C. C. Oliveira and N. Cramer, *Chem Rev*, 2017, **117**, 8908–8976.
2 T. Rogge, N. Kaplaneris, N. Chatani, J. Kim, S. Chang, B. Punji, L. L. Schafer, D. G. Musaev, J. Wencel-Delord, C. A. Roberts, R. Sarpong, Z. E. Wilson, M. A. Brimble, M. J. Johansson and L. Ackermann, *Nat Rev Methods Primers*, 2021, **1**, 1–31.
3 R. H. Crabtree and A. Lei, *Chem Rev*, 2017, **117**, 8481–8482.
4 J. J. Neumann, S. Rakshit, T. Droge and F. Glorius, *Angew Chem, Int Ed*, 2009, **48**, 6892–6895.
5 J. Pedroni and N. Cramer, *Org Lett*, 2016, **18**, 1932–1935.
6 T. Dalton, T. Faber and F. Glorius, *ACS Cent Sci*, 2021, **7**, 245–261.
7 R. Shang, L. Ilies and E. Nakamura, *Chem Rev*, 2017, **117**, 9086–9139.
8 M. S. Chen and M. C. White, *Science*, 2007, **318**, 783–787.
9 E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner and S. L. Flitsch, *Nat Rev Methods Primers*, 2021, **1**, 1–21.
10 N. J. Turner and E. O'Reilly, *Nat Chem Biol*, 2013, **9**, 285–288.
11 M. T. Reetz, *J Am Chem Soc*, 2013, **135**, 12480–12496.
12 R. J. Li, Z. Zhang, C. G. Acevedo-Rocha, J. Zhao and A. Li, *Green Synth Catal*, 2020, **1**, 52–59.
13 M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra-Garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. Mclaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel, A. Rodriguez-Granillo, S. A. Robaire, E. C. Sherer, M. D. Truppo, A. M. Whittaker, D. Verma, L. Xiao, Y. Xu and H. Yang, *Science (1979)*, 2020, **368**, 1255–1259.
14 M. Klingenberg, *Arch Biochem Biophys*, 2003, **409**, 2–6.
15 R. Bernhardt and V. B. Urlacher, *Appl Microbiol Biotechnol*, 2014, **98**, 6185–6203.
16 P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold and E. M. Brustad, *Nat Chem Biol*, 2013, **9**, 485–487.
17 J. A. McIntosh, P. S. Coelho, C. C. Farwell, Z. J. Wang, J. C. Lewis, T. R. Brown and F. H. Arnold, *Angew Chem, Int Ed*, 2013, **52**, 9309–9312.
18 A. H. Payne and D. B. Hales, *Endocr Rev*, 2004, **25**, 947–970.
19 B. Siminszky, *Phytochem Rev*, 2006, **5**, 445–458.
20 S. Yoshioka, T. Tosha, S. Takahashi, K. Ishimori, H. Hori and I. Morishima, *J Am Chem Soc*, 2002, **124**, 14571–14579.
21 C. J. C. Whitehouse, S. G. Bell and L. L. Wong, *Chem Soc Rev*, 2012, **41**, 1218–1260.
22 R. Fasan, M. M. Chen, N. C. Crook and F. H. Arnold, *Angew Chem, Int Ed*, 2007, **46**, 8414–8418.
23 G. A. Roberts, G. Grogan, A. Greter, S. L. Flitsch and N. J. Turner, *J Bacteriol*, 2002, **184**, 3898–3908.
24 W. Yang, S. G. Bell, H. Wang, W. Zhou, N. Hoskins, A. Dale, M. Bartlam, L. L. Wong and Z. Rao, *J Biol Chem*, 2010, **285**, 27372–27384.
25 I. C. Gunsalus and G. C. Wagner, *Methods Enzymol*, 1978, **52**, 166–188.
26 I. Hanukoglu, *Adv Mol Cell Bio*, 1996, **14**, 29–56.
27 Y. Miura and A. J. Fulco, *J Biol Chem*, 1974, **249**, 1880–1888.
28 L. O. Narhi and J. Fulco, *J Biol Chem*, 1986, **261**, 7160–7169.
29 N. English, V. Hughes and C. R. Wolf, *J Biol Chem*, 1994, **269**, 26836–26841.
30 C. N. A. Palmer, E. Axen, V. Hughes and C. R. Wolf, *J Biol Chem*, 1998, **273**, 18109–18116.
31 L. O. Narhi and A. J. Fulco, *J Biol Chem*, 1987, **262**, 6683–6690.
32 S. D. Black and S. T. Martin, *Biochemistry*, 1994, **33**, 12056–12062.
33 H. Zhang, A. L. Yokom, S. Cheng, M. Su, P. F. Hollenberg, D. R. Southworth and Y. Osawa, *J Biol Chem*, 2018, **293**, 7727–7736.
34 A. P. Jamakhandi, B. C. Jeffus, V. R. Dass and G. P. Miller, *Arch Biochem Biophys*, 2005, **439**, 165–174.
35 T. Kitazume, D. C. Haines, R. W. Estabrook, B. Chen and J. A. Peterson, *Biochemistry*, 2007, **46**, 11892–11901.
36 R. Neeli, H. M. Girvan, A. Lawrence, M. J. Warren, D. Leys, N. S. Scrutton and A. W. Munro, *FEBS Lett*, 2005, **579**, 5582–5588.
37 Y. Miura and A. J. Fulco, *Biochim Biophys Acta*, 1975, **388**, 305–317.
38 S. Pflug, S. M. Richter and V. B. Urlacher, *J Biotechnol*, 2007, **129**, 481–488.

39 I. Kaluzna, T. Schmitges, H. Straatman, D. Van Tegelen, M. Müller, M. Schürmann and D. Mink, *Org Process Res Dev*, 2016, **20**, 814–819.
40 O. Gotoh, *J Biol Chem*, 1992, **267**, 83–90.
41 A. B. Carmichael and L. L. Wong, *Eur J Biochem*, 2001, **268**, 3117–3125.
42 S. Graham-Lorence, G. Truant, J. A. Peterson, J. R. Falck, S. Wei, C. Helvig and J. H. Capdevila, *J Biol Chem*, 1997, **272**, 1127–1135.
43 M. W. Peters, P. Meinhold, A. Glieder and F. H. Arnold, *J Am Chem Soc*, 2003, **125**, 13442–13450.
44 W. C. Huang, A. C. G. Westlake, J. D. Maréchal, M. G. Joyce, P. C. E. Moody and G. C. K. Roberts, *J Mol Biol*, 2007, **373**, 633–651.
45 Q. S. Li, U. Schwaneberg, P. Fischer and R. D. Schmid, *Eur J Chem*, 2000, **6**, 1531–1536.
46 Q. S. Li, J. Ogawa, R. D. Schmid and S. Shimizu, *Appl Environ Microbiol*, 2001, **67**, 5735–5739.
47 L. Giver, A. Gershenson, P. O. Freskgard and F. H. Arnold, *Proc Natl Acad Sci U S A*, 1998, **95**, 12809–12813.
48 A. Glieder, E. T. Farinas and F. H. Arnold, *Nat Biotechnol*, 2002, **20**, 1135–1139.
49 E. T. Farinas, U. Schwaneberg, A. Glieder and F. H. Arnold, *Adv Synth Catal*, 2001, **343**, 601–606.
50 P. Meinhold, M. W. Peters, M. M. Y. Chen, K. Takahashi and F. H. Arnold, *ChemBioChem*, 2005, **6**, 1765–1768.
51 R. Fasan, Y. T. Meharenna, C. D. Snow, T. L. Poulos and F. H. Arnold, J Mol Biol, 2008, 383, 1069- 1080.
52 A. Sawayama, M. M. Y. Chen, P. Kulanthaivel, M.-S. Kuo, H. Hemmerle and F. H. Arnold, *Chemistry (Easton)*, 2009, **15**, 11723–11729.
53 M. Landwehr, L. Hochrein, C. R. Otey, A. Kasrayan, J. E. Bäckvall and F. H. Arnold, *J Am Chem Soc*, 2006, **128**, 6058–6059.
54 O. F. Brandenberg, K. Chen and F. H. Arnold, *J Am Chem Soc*, 2019, **141**, 8989–8995.
55 C. J. C. Whitehouse, S. G. Bell, H. G. Tufton, R. J. P. Kenny, L. C. I. Ogilvie and L. L. Wong, *Chem comm*, 2008, **1**, 966–968.
56 C. J. C. Whitehouse, S. G. Bell, W. Yang, J. A. Yorke, C. F. Blanford, A. J. F. Strong, E. J. Morse, M. Bartlam, Z. Rao and L. L. Wong, *ChemBioChem*, 2009, **10**, 1654–1656.
57 C. J. C. Whitehouse, W. Yang, J. A. Yorke, B. C. Rowlatt, A. J. F. Strong, C. F. Blanford, S. G. Bell, M. Bartlam, L. L. Wong and Z. Rao, *ChemBioChem*, 2010, **11**, 2549–2556.
58 M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha and A. Vogel, *Angew Chem, Int Ed*, 2005, **44**, 4192–4196.
59 M. T. Reetz, *Angew Chem, Int Ed*, 2011, **50**, 138–174.
60 S. Kille, F. E. Zilly, J. P. Acevedo and M. T. Reetz, *Nat Chem*, 2011, **3**, 738–743.
61 C. G. Acevedo-Rocha, C. G. Gamble, R. Lonsdale, A. Li, N. Nett, S. Hoebenreich, J. B. Lingnau, C. Wirtz, C. Fares, H. Hinrichs, A. Deege, A. J. Mulholland, Y. Nov, D. Leys, K. J. McLean, A. W. Munro and M. T. Reetz, *ACS Catal*, 2018, **8**, 3395–3410.
62 M. Hecht, Y. Bromberg and B. Rost, *J Mol Biol*, 2013, **425**, 3937–3948.
63 J. Y. van der Meer, L. Biewenga and G. J. Poelarends, *ChemBioChem*, 2016, **17**, 1792–1799.
64 A. Li, C. G. Acevedo-Rocha, L. D'Amore, J. Chen, Y. Peng, M. Garcia-Borràs, C. Gao, J. Zhu, H. Rickerby, S. Osuna, J. Zhou and M. T. Reetz, *Angew Chem, Int Ed*, 2020, **59**, 12499–12505.
65 A. Seifert, S. Vomund, K. Grohmann, S. Kriening, V. B. Urlacher, S. Laschat and J. Pleiss, *ChemBioChem*, 2009, **10**, 853–861.
66 W. Chen, M. J. Fisher, A. Leung, Y. Cao and L. L. Wong, *ACS Catal*, 2020, **10**, 8334–8343.
67 X. Ren, J. A. Yorke, E. Taylor, T. Zhang, W. Zhou and L. L. Wong, *Chem Eur J*, 2015, **21**, 15039–15047.
68 Y. Li and L. L. Wong, *Angew Chem, Int Ed*, 2019, **58**, 9551–9555.
69 L. D. Syntrivanis, L. L. Wong and J. Robertson, *European J Org Chem*, 2018, **2018**, 6369–6378.
70 X. Ren, J. A. O'Hanlon, M. Morris, J. Robertson and L. L. Wong, *ACS Catal*, 2016, **6**, 6833–6837.
71 K. Zhang, B. M. Shafer, M. D. Demars, H. A. Stern and R. Fasan, *J Am Chem Soc*, 2012, **134**, 18695–18704.
72 K. Zhang, S. el Damaty and R. Fasan, *J Am Chem Soc*, 2011, **133**, 3242–3245.
73 A. Hospital, J. R. Goñi, M. Orozco and J. L. Gelpí, *Adv Appl Bioinforma Chem*, 2015, **8**, 37–47.
74 S. A. Hollingsworth and R. O. Dror, *Neuron*, 2018, **99**, 1129–1143.
75 M. Karplus, *Acc Chem Res*, 2002, **35**, 321–323.
76 B. J. Alder and T. E. Wainwright, *J Chem Phys*, 1957, **27**, 1208–1209.
77 A. Warshel and M. Levitt, *J Mol Biol*, 1976, **103**, 227–249.
78 M. Karplus, J. A. McCammon and B. R. Gelin, *Nature*, 1977, **267**, 585–590.
79 M. Kawata and U. Nagashima, *Chem Phys Lett*, 2001, **340**, 165–172.
80 T. Lazaridis and M. Karplus, *Proteins: Struct Funct Genet*, 1999, **35**, 133–152.
81 M. Orozco and F. J. Luque, *Chem Rev*, 2000, **100**, 4187–4225.
82 B. Roux and T. Simonson, *Biophys Chem*, 1999, **78**, 1–20.
83 B. Donald and D. A. Case, *Annu Rev Phys Chem*, 2002, **51**, 129–52.

84 P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L. P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks and V. S. Pande, *PLoS Comput Biol*, 2017, 13, e1005659.
85 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J Comput Chem*, 2005, **26**, 1781–1802.
86 B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J Comput Chem*, 2009, **30**, 1545–1614.
87 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J Chem Theory Comput*, 2008, **4**, 435–447.
88 V. Salmaso and S. Moro, *Front Pharmacol*, 2018, **9**, 1–16.
89 X.-Y. Meng, H.-X. Zhang, M. Mezei and M. Cui, *Curr Comput Aided Drug Des,* 2011, **7**, 146–157.
90 I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge and T. E. Ferrin, *J Mol Biol*, 1982, **161**, 269–288.
91 O. Trott and A. J. Olson, *J Comput Chem*, 2009, **31**, 455-461.
92 M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins: Struct Funct Genet*, 2003, **52**, 609–623.
93 J. S. Taylor and R. M. Burnett, *Proteins: Struct Funct Genet*, 2000, **41**, 173–191.
94 P. A. Ravindranath, S. Forli, D. S. Goodsell, A. J. Olson and M. F. Sanner, *PLoS Comput Biol*, 2015, **11**, 1–28.
95 A. T. Brint and P. Willett, *J Chem Inf Comput Sci*, 1987, **27**, 152–158.
96 D. Fischer, R. Norel, H. Wolfson and R. Nussinov, *Proteins: Struct Funct Genet*, 1993, **16**, 278–292.
97 R. Norel, D. Fischer, H. J. Wolfson and R. Nussinov, *Protein Eng. Des. Sel.*, 1994, **7**, 39–46.
98 M. Rarey, B. Kramer, T. Lengauer and G. Klebe, *J Mol Biol*, 1996, **261**, 470–489.
99 R. L. DesJarlais, I. D. Kuntz, R. P. Sheridan, R. Venkataraghavan and J. S. Dixon, *J Med Chem*, 1986, **29**, 2149–2153.
100 A. R. Leach and I. D. Kuntz, *J Comput Chem*, 1992, **13**, 730–748.
101 A. Miranker and M. Karplus, *Proteins: Struct Funct Genet*, 1991, **11**, 29–34.
102 H. J. Böhm, *J Comput Aided Mol Des*, 1992, **6**, 593–606.
103 D. S. Goodsell, H. Lauble, C. D. Stout and A. J. Olson, *Proteins: Struct Funct Genet,* 1993, **17**, 1–10.
104 T. N. Hart and R. J. Read, *Proteins: Struct Funct Genet*, 1992, **13**, 206–222.
105 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J Comput Chem*, 1998, **19**, 1639–1662.
106 G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J Mol Biol*, 1997, **267**, 727–748.
107 C. M. Oshiro, I. D. Kuntz and J. S. Dixon, *J Comput Aided Mol Des*, 1995, **9**, 113–130.
108 K. P. Clark and Ajay, *J Comput Chem*, 1995, **16**, 1210–1226.
109 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J Am Chem Soc*, 1995, **117**, 5179–5197.
110 S. J. Weiner, P. A. Kollman, U. C. Singh, D. A. Case, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *J Am Chem Soc*, 1984, **106**, 765–784.
111 P. Kollman, *Chem Rev*, 1993, **93**, 2395–2417.
112 J. Aqvist, V. B. Luzhkov and B. O. Brandsdal, *Acc Chem Res*, 2002, **35**, 358–365.
113 H. A. Carlson and W. L. Jorgensen, *J Phys Chem B*, 1995, **99**, 10667–10673.
114 D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. R. Fogel, L. J. Fogel and S. T. Freer, *Chem Biol*, 1995, **2**, 317–324.
115 G. M. Verkhivker, D. Bouzida, D. K. Gehlhaar, P. A. Rejto, S. Arthurs, A. B. Colson, S. T. Freer, V. Larson, B. A. Lutty, T. Marrone and P. W. Rose, *J Comput Aided Mol Des*, 2000, **14**, 731–751.
116 A. N. Jain, *J Comput Aided Mol Des*, 1996, **10**, 427–440.
117 R. D. Head, M. L. Smythe, T. I. Oprea, C. L. Waller, S. M. Green and G. R. Marshall, *J Am Chem Soc*, 1996, **118**, 3959–3969.
118 H. J. Böhm, *J Comput Aided Mol Des*, 1998, **12**, 309.
119 D. K. Gehlhaar, K. E. Moerder, D. Zichi, C. J. Sherman, R. C. Ogden and S. T. Freer, *J Med Chem*, 1995, **38**, 466–472.
120 M. D. Eldridge, C. W. Murray, T. R. Auton, G. v. Paolini and R. P. Mee, *J Comput Aided Mol Des*, 1997, **11**, 425–445.
121 I. Muegge and Y. C. Martin, *J Med Chem*, 1999, **42**, 791–804.
122 J. B. O. Mitchell, R. A. Laskowski, A. Alex, M. J. Forster and J. M. Thornton, *J Comput Chem*, 1999, **20**, 1177–1185.
123 R. S. DeWitte, A. v. Ishchenko and E. I. Shakhnovich, *J Am Chem Soc*, 1997, **119**, 4608–4617.
124 M. Feher, E. Deretey and S. Roy, *J Chem Inf Comput Sci*, 2003, **43**, 1316–1327.
125 A. Wallqvist, R. L. Jernigan and D. G. Covell, *Protein Science*, 1995, **4**, 1881–1903.
126 P. S. Charifson, J. J. Corkery, M. A. Murcko and W. P. Walters, *J Med Chem*, 1999, **42**, 5100–5109.

127 M. D. Paulsen and R. L. Ornstein, *Proteins: Struct Funct Genet*, 1995, **21**, 237–243.
128 J. Pleiss, *ChemCatChem*, 2014, **6**, 944–950.
129 Y. Peng, C. Gao, Z. Zhang, S. Wu, J. Zhao and A. Li, *ACS Catal*, 2022, **12**, 2907–2914.
130 Y. Zhang, Z. Xiong, Y. Li, M. Wilson, K. E. Christensen, E. Jaques, P. Hernández-Lladó, J. Robertson and L. L. Wong, *Nature Synthesis*, 2022, 1, 936–945.
131 F. M. Schempp, L. Drummond, M. Buchhaupt and J. Schrader, *J Agric Food Chem*, 2018, 66, 2247–2258.
132 M. Chen, N. Al-Lami, M. Janvier, E. L. D'Antonio, J. A. Faraldos, D. E. Cane, R. K. Allemann and D. W. Christianson, *Biochemistry*, 2013, **52**, 5441–5453.
133 D. L. Klayman, *Science*, 1985, **228**, 1049–1055.
134 R. Croteau, R. Ketchum, R. Long, R. Kaspera and M. Wildung, *Phytochem Rev*, 2006, **5**, 75–97.
135 F. Deguerry, L. Pastore, S. Wu, A. Clark, J. Chappell and M. Schalk, *Arch Biochem Biophys*, 2006, **454**, 123–136.
136 V. B. Urlacher and M. Girhard, *Trends Biotechnol*, 2019, **37**, 882–897.
137 R. J. Sowden, S. Yasmin, N. H. Rees, S. G. Bell and L. L. Wong, *Org Biomol Chem*, 2005, **3**, 57–64.
138 D. K. Ro, E. M. Paradise, M. Quellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong and J. D. Keasling, *Nature*, 2006, **440**, 940–943.
139 R. Naef, *Flavour Fragr J*, 2011, **26**, 73–87.
140 H. Q. Chen, J. H. Wei, J. S. Yang, Z. Zhang, Y. Yang, Z. H. Gao, C. Sui and B. Gong, *Chem Biodivers*, 2012, **9**, 236–250.
141 M. Ishihara, T. Tsuneya, M. Shiga and K. Uneyama, *Pytochemistry*, 1991, **30**, 563–566.
142 M. T. N. Ismail, N. Ali, M. Jamil, M. Rahiman, S. Tajuddin, *J Teknol*, 2014, **1**, 37–42.
143 D. T. T. Thuy, T. T. Tuyen, T. T. T. Thuy, P. T. H. Minh, Q. T. Tran, P. Q. Long, D. C. Nguyen, L. G. Bach and N. Q. Chien, *Processes*, 2019, **7**, 432.
144 M. L. Maheshwari, T. C. Jain, R. B. Bates and S. C. Bhattacharyya, *Tetrahedron*, 1963, **19**, 1079–1090.
145 H. C. Barrett and G. Büchi, *J Am Chem Soc*, 1967, **89**, 5665–5667.
146 T. Nakanishi, E. Yamagata, K. Yoneda and I. Miura, *Phytochemistry*, 1981, **20**, 1597–1599.
147 T. Nakanishi, E. Yamagata and K. Yoneda, *J. Chem. Soc. Perkin Trans I*, 1983, 601–604.
148 K. Yoneda, E. Yamagata, T. Nakanishi, T. Nagashima, I. Kawasaki, T. Yoshida, H. Mori and I. Miura, *Phytochemistry*, 1984, **23**, 2068–2069.
149 M. Zviely and A. Boix-camps, *Isr Chem Eng*, 2015, 27–31.
150 US Pat., 4444982, 1984.
151 M. Ishihara, T. Tsuneya and K. Uneyama, *Phytochemistry*, 1991, **30**, 3343–3347.
152 C. Wood, T. E. Siebert, M. Parker, D. L. Capone, G. M. Elsey, A. P. Pollnitz, M. Eggers, M. Meier, T. Vössing, S. Widder, G. Krammer, M. A. Sefton and M. J. Herderich, *J Agric Food Chem*, 2008, **56**, 3738–3744.
153 N. Baldovini, C. Delasalle and D. Joulain, *Flavour Fragr J*, 2011, **26**, 7–26.
154 F. P. Guengerich and A. W. Munro, *J Biol Chem*, 2013, **288**, 17065–17073.
155 WIPO, WO 2012/001018 A1, 2012.
156 A. C. Huang, M. A. Sefton and D. K. Taylor, *J Agric Food Chem*, 2015, **63**, 1932–1938
157 A. C. Huang, M. A. Sefton, C. J. Sumby, E. R. T. Tiekink and D. K. Taylor, *J Nat Prod*, 2015, **78**, 131–145.
158 M. Ishihara, Y. Masatsugu and K. Uneyama, *Tetrahedron*, 1992, **48**, 10265–10276.
159 D. C. Haines, D. R. Tomchick, M. Machius and J. A. Peterson, *Biochemistry*, 2001, **40**, 13456–13465.
160 K. Shahrokh, A. Orendt, G. S. Yost and T. E. Cheatham, *J Comput Chem*, 2012, **33**, 119–133.
161 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1–2**, 19–25.
162 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct Funct Genet*, 2010, **78**, 1950–1958.
163 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J Chem Phys*, 1983, **79**, 926–935.
164 B. Hess, *J Chem Theory Comput*, 2008, **4**, 116–122.
165 T. Darden, D. York and L. Pedersen, *J Chem Phys*, 1993, **98**, 10089–10092.
166 G. Bussi, D. Donadio and M. Parrinello, *Chem Phys*, 2007,**126**, 04101.
167 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola and J. R. Haak, *J Chem Phys*, 1984, **81**, 3684–3690.
168 X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew Chem, Int Ed* 1999, **38**, 236–240.
169 A. K. Picman, *Biochem Syst Ecol*, 1986, **14**, 255–281.
170 M. Shoaib, I. Shah, N. Ali, A. Adhikari, M. N. Tahir, S. W. A. Shah, S. Ishtiaq, J. Khan, S. Khan and M. N. Umer, *BMC Complement Altern Med*, 2017, **17**, 1–11.
171 J. Y. Cho, K. U. Baik, J. H. Jung and M. H. Park, *Eur J Pharmacol*, 2000, **398**, 399–407.
172 D. W. Knight, *Nat Prod Rep*, 1995, **12**, 271–276.

173 A. C. Huang, S. Burrett, M. A. Sefton and D. K. Taylor, *J Agric Food Chem*, 2014, **62**, 10809–10815.
174 H. Takase, K. Sasaki, H. Shinmori, A. Shinohara, C. Mochizuki, H. Kobayashi, G. Ikoma, H. Saito, H. Matsuo, S. Suzuki and R. Takata, *J Exp Bot*, 2016, **67**, 787–798.
175 S. Umezawa, S. Konishi and K. Kino, *Biosci Biotechnol Biochem*, 2019, **83**, 1875–1883.

Appendices

Appendix 1: List of CYP102A1 variants for progressive screening

Table 1.1 Variants and mutations in the initial screening and progressive screening plates

	Variants	Mutations
	Initial screening_V1 (for bulnesene, **1**, and guaiene, **2**) (Chapter 3.2 and Chapter 6.2)	
1	GV/A184I	A74G/F87V/A184I
2	GVQ	A74G/F87V/L188Q
3	GVQ/A264G	A74G/F87V/L188Q/A264G
4	GVQ/I263G	A74G/F87V/L188Q/I263G
5	GVQ/I263G/A328L	A74G/F87V/L188Q/I263G/A328L
6	K19/F87A/A82M/E267F	A82M/F87A/H171L/E267F/Q307H/N319Y
7	K19/F87A/A82M/I263G	A82M/F87A/H171L/I263G/Q307H/N319Y
8	K19/F87A/F81W	F81W/F87A/H171L/Q307H/N319Y
9	K19/F87A/I263A	F87A/I263A/H171L/Q307H/N319Y
10	K19/F87A/I263G	F87A/I263G/H171L/Q307H/N319Y
11	K19/F87V	F87V/H171L/Q307H/N319Y
12	K19/F87V/A264G	F87V/H171L/A264G/Q307H/N319Y
13	K19/F87V/A328V	F87V/H171L/Q307H/N319Y/A328V
14	K19/F87V/E267V	F87V/H171L/E267V/Q307H/N319Y
15	K19/F87V/V78I/E267V	V78I/F87V/H171L/E267V/Q307H/N319Y
16	K19/F87V/Q403P	F87V/H171L/Q307H/N319Y/Q403P
17	KU3/A330P	N239H/I259V/A276T/A330P
18	KU3/A328I/A330P	N239H/I259V/A276T/A328I/A330P
19	R19	R47L/Y51F/H171L/Q307H/N319Y
20	R19/A184I/A328L	R47L/Y51F/H171L/A184I/Q307H/N319Y/A328L
21	R19/F87A	R47L/Y51F/F87A/H171L/Q307H/N319Y
22	R19/F87A/A184I	R47L/Y51F/F87A/H171L/A184I/Q307H/N319Y
23	R19/F87A/A328F	R47L/Y51F/F87A/H171L/Q307H/N319Y/A328F
24	R19/F87A/A328I	R47L/Y51F/F87A/H171L/Q307H/N319Y/A328I
25	R19/F87A/S72W/A328I	R47L/Y51F/S72W/F87A/H171L/Q307H/N319Y/A328I
26	R19/F87A/V78I/A328I	R47L/Y51F/V78I/F87A/H171L/Q307H/N319Y/A328I
27	R19/F87A/A328L	R47L/Y51F/F87A/H171L/Q307H/N319Y/A328L
28	R19/F87A/I263G	R47L/Y51F/F87A/H171L/I263G/Q307H/N319Y
29	R19/F87I	R47L/Y51F/F87I/H171L/Q307H/N319Y
30	R47L/Y51F/I263R	R47L/Y51F/I263R
31	RP/A82M/I263A	R47L/Y51F/A82M/I263A/I401P
32	RP/F87V	R47L/Y51F/F87V/I401P
33	RP/F87V/V78I	R47L/Y51F/V78I/F87V/I401P
34	RP/H171L/I263G	R47L/Y51F/H171L/I263G/I401P
35	RP/H171L/A184I/I263G	R47L/Y51F/H171L/A184I/I263G/I401P

36	RP/F87V/V78I/H171L/I263G	R47L/Y51F/V78I/F87V/H171L/I263G/I401P
37	RP/I263A/E267V	R47L/Y51F/I263A/E267V/I401P
38	RP/V78I/E267V	R47L/Y51F/V78I/E267V/I401P
39	RP/A82W/I263A	R47L/Y51F/A82W/I263A/I401P
40	RT2	R47L/Y51F/A191T/N239H/I259V/A276T/L353I
41	RT2/V78I/A184I/A330P	R47L/Y51F/V78I/A184I/A191T/N239H/I259V/A276T/A330P/L353I
42	RT2/S72G/A330W	R47L/Y51F/S72G/A191T/N239H/I259V/A276T/A330W/L353I
43	RT2/S72W/A330W	R47L/Y51F/S72W/A191T/N239H/I259V/A276T/A330W/L353I
44	WT	
	Initial Screening_V2 (for bulnesene-15-COOH, **1g**) (Chapter 4.2)	
1	A330P	A330P
2	GLQ/I263G/A328G	A74G/F87L/L188Q/I263G/A328G
3	GQ/I263G/A328L/A330L /S332F	A74G/L188Q/I263G/A328L/A330L/S332F
4	GV/A184I	A74G/F87V/A184I
5	GV/A184I/I263G/A328G	A74G/F87V/A184I/I263G/A328G
6	GVQ	A74G/F87V/L188Q
7	GVQ/A264G	A74G/F87V/L188Q/A264G
8	GVQ/A328G	A74G/F87V/L188Q/A328G
9	GVQ/A330W	A74G/F87V/L188Q/A330W
10	GVQ/I263G	A74G/F87V/L188Q/I263G
11	GVQ/I263G/A328L	A74G/F87V/L188Q/I263G/A328L
12	K19/F87A/A82M/I263G	A82M/F87A/H171L/I263G/Q307H/N319Y
13	K19/F87A/I263A	F87A/H171L/I263A/Q307H/N319Y
14	K19/F87A/I263G	F87A/H171L/I263G/Q307H/N319Y
15	K19/F87V	F87V/H171L/Q307H/N319Y
16	K19/F87V/A264G	F87V/H171L/A264G/Q307H/N319Y
17	K19/F87V/E267V	F87V/H171L/E267V/Q307H/N319Y
18	K19/F87V/I263G	F87V/H171L/I263G/Q307H/N319Y
19	K19/F87V/Q403P	F87V/H171L/Q307H/N319Y/Q403P
20	K19/F87V/V78I/E267V	V78I/F87V/H171L/E267V/Q307H/N319Y
21	KT2/L188G/I263G	L188G/A191T/N239H/I259V/I263G/A276T/L353I
22	KU3/I263A/A330P	N239H/I259V/I263A/A276T/A330P
23	KU3/S72W/A330P	S72W/N239H/I259V/A276T/A330P
24	R19	R47L/Y51F/H171L/Q307H/N319Y
25	R19/A330W	R47L/Y51F/H171L/Q307H/N319Y/A330W
26	R19/F87A	R47L/Y51F/F87A/H171L/Q307H/N319Y
27	R19/F87A/A184I	R47L/Y51F/F87A/H171L/A184I/Q307H/N319Y
28	R19/F87A/A184I/T260G /A328G	R47L/Y51F/F87A/H171L/A184I/T260G/Q307H/N319Y/A328G
29	R19/F87A/A328G	R47L/Y51F/F87A/H171L/Q307H/N319Y/A328G
30	R19/F87A/A328G/P329G /A330G	R47L/Y51F/F87A/H171L/Q307H/N319Y/A328G/P329G/ A330G
31	R19/F87A/A328L	R47L/Y51F/F87A/H171L/Q307H/N319Y/A328L

32	R19/F87A/F81W/T260G /A328G	R47L/Y51F/F81W/F87A/H171L/T260G/Q307H/N319Y/A328G
33	R19/F87A/I263G	R47L/Y51F/F87A/H171L/I263G/Q307H/N319Y
34	R19/F87A/S72W/A328I	S72W/R47L/Y51F/F87A/H171L/Q307H/N319Y/A328I
35	R19/F87A/T260G	R47L/Y51F/F87A/H171L/T260G/Q307H/N319Y
36	R19/F87A/V78I/A328I	R47L/Y51F/V78I/F87A/H171L/Q307H/N319Y/A328I
37	R19/F87I	R47L/Y51F/F87I/H171L/Q307H/N319Y
38	R19/F87I/L437F	R47L/Y51F/F87I/H171L/Q307H/N319Y/L437F
39	R19/F87I/L437LV	R47L/Y51F/F87I/H171L/Q307H/N319Y/L437LV
40	R19/I263A/A328L	R47L/Y51F/H171L/I263A/Q307H/N319Y/A328L
41	R19/S72G/A330W	S72G/R47L/Y51F/H171L/Q307H/N319Y/A330W
42	RP/A82M/I263A	R47L/Y51F/A82M/I263A/I401P
43	RP/H171L/A184I/I263G	R47L/Y51F/H171L/A184I/I263G/I401P
44	RP/H171L/I263G	R47L/Y51F/H171L/I263G/I401P
45	RT2	R47L/Y51F/A191T/N239H/I259V/A276T/L353I
46	VQ/S72G/A330W	S72G/F87V/L188Q/A330W
47	WT	
	GVQ panel (Chapter 3.2 and Chapter 6.2)	
1	GVQ/A328G	A74G/F87V/L188Q/A328G
2	GVQ/I263G/A328F	A74G/F87V/L188Q/I263G/A328F
3	GVQ/I263G/A328I	A74G/F87V/L188Q/I263G/A328I
4	GAQ/I263G	A74G/F87A/L188Q/I263G
5	GAQ/I263G/A328G	A74G/F87A/L188Q/I263G/A328G
6	GAQ/I263G/A328L	A74G/F87A/L188Q/I263G/A328L
7	GLQ/I263G	A74G/F87L/L188Q/I263G
8	GLQ/I263G/A328I	A74G/F87L/L188Q/I263G/A328I
9	GLQ/I263G/A328G	A74G/F87L/L188Q/I263G/A328G
10	GLQ/I263G/A328L	A74G/F87L/L188Q/I263G/A328L
11	GLQ/I263G/A328F	A74G/F87L/L188Q/I263G/A328F
12	GGQ/I263G	A74G/F87G/L188Q/I263G
13	GGQ/I263G/A328F	A74G/F87G/L188Q/I263G/A328F
14	GGQ/I263G/A328G	A74G/F87G/L188Q/I263G/A328G
15	GGQ/I263G/A328L	A74G/F87G/L188Q/I263G/A328L
	F87I panel (Chapter 3.2)	
1	K19/F87I	F87I/H171L/Q307H/N319Y
2	KT2/N70S/F87I	R47L/Y51F/N70S/F87I/A191T/N239H/I259V/A276T/L353I
3	R19/F87I	R47L/Y51F/F87I/H171L/Q307H/N319Y
4	KT2/F87I	F87I/A191T/N239H/I259V/A276T/L353I
5	RP/F87I/E267V/A328I	R47L/Y51F/F87I/E267V/A328I/I401P
6	KT2/F87I/I263F	F87I/A191T/N239H/I259V/I263F/A276T/L353I
7	KT2/F87I/I263V	F87I/A191T/N239H/I259V/I263V/A276T/L353I
8	KT2/F87I/I263F/A328I	F87I/A191T/N239H/I259V/I263F/A276T/A328I/L353I

9	KT2/L75W/F87I	L75W/F87I/A191T/N239H/I259V/A276T/L353I
10	KT2/F87I/A82M	F87I/A82M//A191T/N239H/I259V/A276T/L353I
11	KT2/F87I/M237L	F87I/A191T/M237L/N239H/I259V/A276T/L353I
12	KT2/F87I/A328V	F87I/A191T/N239H/I259V/A276T/A328V/L353I
13	KT2/F87I/A328V/I263V	F87I/A191T/N239H/I259V/I263V/A276T/A328V/L353I
14	KT2/F87I/A328V/I263F	F87I/A191T/N239H/I259V/I263F/A276T/A328V/L353I
15	KT2/F87I/I263V/M237L	F87I/A191T/M237L/N239H/I259V/I263V/A276T/L353I
16	KT2/F87I/M237L/A328V	F87I/A191T/M237L/N239H/I259V/A276T/A328V/L353I
17	KT2/F87I/M237L/I263F/ A328V	F87I/A191T/M237L/N239H/I259V/I263F/A276T/A328V/L353I
18	RT2/F87I /263F/A328V	R47L/Y51F/F87I/A191T/N239H/I259V/I263F/A276T/A328V/L353I
19	KT2/N70S/F87I/M237L/ I263F/A328V	N70S/F87I/A191T/M237L/N239H/I259V/I263F/A276T/A328V /L353I
20	RT2/F87I/E267F	R47L/Y51F/F87I/A191T/N239H/I259V/E267F/A276T/L353I
21	KT2/N70S/L75Q/F87I/ N237L/I263F/A328V	N70S/L75Q/F87I/A191T/N237L/N239H/I259V/I263F/A276T/A328V /L353I
	R19/F87A panel_V1 (for bulnesene, **1**) (Chapter 3.2)	
1	R19/F87A/P329G/A330W	R47L/Y51F/F87A/H171L/Q307H/N319Y/P329G/A330W
2	R19/F87A/P329G/A330P	R47L/Y51F/F87A/H171L/Q307H/N319Y/P329G/A330P
3	R19/F87A/P329V/A330P	R47L/Y51F/F87A/H171L/Q307H/N319Y/P329V/A330P
4	R19/F87A/I263G	R47L/Y51F/F87A/H171L/I263G/Q307H/N319Y
5	R19/F87A/I263G/P329G /A330W	R47L/Y51F/F87A/H171L/I263G/Q307H/N319Y/P329G/A330W
6	R19/F87A/I263G/P329G /A330P	R47L/Y51F/F87A/H171L/I263G/Q307H/N319Y/P329G/A330P
7	R19/F87A/I263G/V127I	R47L/Y51F/F87A/H171L/I263G/V127I/Q307H/N319Y
8	R19/F87A/A191P	R47L/Y51F/F87A/H171L/A191P/Q307H/N319Y
9	R19/F87A/A328G	R47L/Y51F/F87A/H171L/Q307H/N319Y/A328G
10	R19/F87A/A264G	R47L/Y51F/F87A/H171L/A264G/Q307H/N319Y
11	R19/F87A/A264G/A328G	R47L/Y51F/F87A/H171L/A264G/Q307H/N319Y/A328G
12	R19/F87A/Q403P	R47L/Y51F/F87A/H171L/Q307H/N319Y/Q403P
13	R19/F87A/I263G/A264G	R47L/Y51F/F87A/H171L/I263G/A264G/Q307H/N319Y
14	R19/F87A/I263G/A264G /A328G	R47L/Y51F/F87A/H171L/I263G/A264G/Q307H/N319Y/A328G
15	R19/F87A/A264G/A328I	R47L/Y51F/F87A/H171L/A264G/Q307H/N319Y/A328I
16	R19/F87A/S72A/A328I	R47L/Y51F/S72A/F87A/H171L/Q307H/N319Y/A328I
	R19/F87A panel_V2 (for bulnesene-15-COOH, **1g**) (Chapter 4.2)	
1	R19/F87A/A264G	R47L/Y51F/F87A/H171L/A264G/Q307H/N319Y
2	R19/F87A/G265GG	R47L/Y51F/F87A/H171L/G265GG/Q307H/N319Y
3	R19/F87A/I263G/A328G	R47L/Y51F/F87A/H171L/I263G/Q307H/N319Y/A328G
4	R19/F87A/L262G	R47L/Y51F/F87A/H171L/L262G/Q307H/N319Y
5	R19/F87A/L262G/A328G	R47L/Y51F/F87A/H171L/L262G/Q307H/N319Y/A328G
6	R19/F87A/P329G/A330W	R47L/Y51F/F87A/H171L/Q307H/N319Y/P329G/A330W
7	R19/F87A/P329V/A330P	R47L/Y51F/F87A/H171L/Q307H/N319Y/P329V/A330P

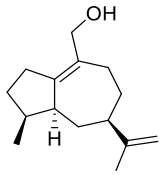
8	R19/F87A/Q403P	R47L/Y51F/F87A/H171L/Q307H/N319Y/Q403P
9	R19/F87A/S270G	R47L/Y51F/F87A/H171L/S270G/Q307H/N319Y
10	R19/F87A/S270G/A328G	R47L/Y51F/F87A/H171L/S270G/Q307H/N319Y/A328G
11	R19/F87A/T260G/A264G	R47L/Y51F/F87A/H171L/T260G/A264G/Q307H/N319Y
12	R19/F87A/T260G/A264G /A328G	R47L/Y51F/F87A/H171L/T260G/A264G/Q307H/N319Y /A328G
13	R19/F87A/T260G/A328G	R47L/Y51F/F87A/H171L/T260G/Q307H/N319Y/A328G
14	R19/F87A/T260G/I263G	R47L/Y51F/F87A/H171L/T260G/I263G/Q307H/N319Y
15	R19/F87A/T260G/I263G /A264G	R47L/Y51F/F87A/H171L/T260G/I263G/A264G/Q307H/N319Y
16	R19/F87A/T260G/I263G /A264G/A328G	R47L/Y51F/F87A/H171L/T260G/I263G/A264G/Q307H/N319Y /A328G
17	R19/F87A/T260G/I263G /A328G	R47L/Y51F/F87A/H171L/T260G/I263G/Q307H/N319Y/A328G
18	R19/F87A/T260G/P329G /A330W	R47L/Y51F/F87A/H171L/T260G/Q307H/N319Y/P329G /A330W
19	R19/F87A/T260G/T268S /A328G	R47L/Y51F/F87A/H171L/T260G/T268S/Q307H/N319Y /A328G
20	R19/F87A/T269G	R47L/Y51F/F87A/H171L/T269G/Q307H/N319Y
21	R19/F87A/T269G/A328G	R47L/Y51F/F87A/H171L/T269G/Q307H/N319Y/A328G
	R19 panel (Chapter 3.2)	
1	R19/A184I/T260G	R47L/Y51F/H171L/A184I/T260G/Q307H/N319Y
2	R19/I263G/A264G/A328G	R47L/Y51F/H171L/I263G/A264G/Q307H/N319Y/A328G
3	R19/I263W/A328G	R47L/Y51F/H171L/I263W/Q307H/N319Y/A328G
4	R19/A328G/P329G/A330G	R47L/Y51F/H171L/Q307H/N319Y/A328G/P329G/A330G
5	R19/A328G/P329G/A330G /T260G/Y278H	R47L/Y51F/H171L/T260G/Y278H/Q307H/N319Y/A328G /P329G/A330G
6	R19/T260G/A328G	R47L/Y51F/H171L/T260G/Q307H/N319Y/A328G
7	R19/W130C/G265GG	R47L/Y51F/W130C/H171L/G265GG/Q307H/N319Y
8	R19/I263G	R47L/Y51F/H171L/I263G/Q307H/N319Y
9	R19/I263G/A328G	R47L/Y51F/H171L/I263G/Q307H/N319Y/A328G
10	R19/P329G/A330W	R47L/Y51F/H171L/Q307H/N319Y/P329G/A330W
11	R19/T260G	R47L/Y51F/H171L/T260G/Q307H/N319Y
12	R19/A82M/A184I/T260G	R47L/Y51F/A82M/H171L/A184I/T260G/Q307H/N319Y
13	R19/A82M/I263G/A264G /A328G	R47L/Y51F/A82M/H171L//I263G/A264G/Q307H/N319Y/A328G
14	R19/I263W/A82M/A328G	R47L/Y51F/A82M/H171L/I263W/Q307H/N319Y/A328G
15	R19	R47L/Y51F/H171L/Q307H/N319Y
16	R19/A82M/T260G/A328G/P329G/ A330G	R47L/Y51F/A82M/H171L/T260G/Q307H/N319Y/A328G /P329G/A330G
	RP panel (Chapter 3.2 and 6.2)	
1	RP/A82M/A330W	R47L/Y51F/A82M/A330W/I401P
2	RP/A82M/H171L/I263G	R47L/Y51F/A82M/H171L/I263G/I401P
3	RP/A82M/I263A/A330W	R47L/Y51F/A82M/I263A/A330W/I401P
4	RP/A82M/I263A/E267F	R47L/Y51F/A82M/I263A/E267F/I401P
5	RP/A82M/I263G	R47L/Y51F/A82M/I263G/I401P

6	RP/H171L/E267F	R47L/Y51F/H171L/E267F/I401P
7	RP/I263A/A330W	R47L/Y51F/I263A/A330W/I401P
8	RP/S72G/H171L/I263G	R47L/Y51F/S72G/H171L/I263G/I401P
9	RP/S72W/H171L/I263G	R47L/Y51F/S72W/H171L/I263G/I401P
	R19/F87A/A184I panel (Chapter 4.2)	
1	R19/F87A/A184I/A264G	R47L/Y51F/F87A/H171L/A184I/A264G/Q307H/N319Y
2	R19/F87A/A184I/A330P	R47L/Y51F/F87A/H171L/A184I/Q307H/N319Y/A330P
3	R19/F87A/A184I/A82M /L181F/T260G/A328G	R47L/Y51F/A82M/F87A/H171L/L181F/A184I/T260G/Q307H/N319Y/A328G
4	R19/F87A/A184I/A82M /T260G	R47L/Y51F/A82M/F87A/H171L/A184I/T260G/Q307H/N319Y
5	R19/F87A/A184I/A82M /T260G/A328G/A330I	R47L/Y51F/A82M/F87A/H171L/A184I/T260G/Q307H/N319Y /A328G/A330I
6	R19/F87A/A184I/A82M /T260G/A328G/A330I/S332F	R47L/Y51F/A82M/F87A/H171L/A184I/T260G/Q307H/N319Y/ A328G/A330I/S332F
7	R19/F87A/A184I/A82M /T260G/A328G/A330L	R47L/Y51F/A82M/F87A/H171L/A184I/T260G/Q307H/N319Y/ A328G/A330L
8	R19/F87A/A184I/A82M /T260G/A328G/S332M	R47L/Y51F/A82M/F87A/H171L/A184I/T260G/Q307H/N319Y/ A328G/S332M
9	R19/F87A/A184I/I263G /A264G	R47L/Y51F/F87A/H171L/A184I/I263G/A264G/Q307H/N319Y
10	R19/F87A/A184I/I263G /A264G/A328G	R47L/Y51F/F87A/H171L/A184I/I263G/A264G/Q307H/N319Y /A328G
11	R19/F87A/A184I/I263G /A328G	R47L/Y51F/F87A/H171L/A184I/I263G/Q307H/N319Y/A328G
12	R19/F87A/A184I/I263W	R47L/Y51F/F87A/H171L/I263W/Q307H/N319Y
13	R19/F87A/A184I/S72F/A82M /T260G/A328G	S72F/R47L/Y51F/A82M/F87A/H171L/A184I/T260G/Q307H/ N319Y/A328G
14	R19/F87A/A184I/S72F /T260G/A264G/A328G	S72F/R47L/Y51F/F87A/H171L/A184I/T260G/A264G/Q307H/ N319Y/A328G
15	R19/F87A/A184I/S72W /T260G/A264G/A328G	S72W/R47L/Y51F/F87A/H171L/A184I/T260G/A264G/Q307H/ N319Y/A328G
16	R19/F87A/A184I/T260G	R47L/Y51F/F87A/H171L/T260G/Q307H/N319Y
17	R19/F87A/A184I/T260G /A264G	R47L/Y51F/F87A/H171L/A184I/T260G/A264G/Q307H/N319Y
18	R19/F87A/A184I/T260G /A264G/A328G	R47L/Y51F/F87A/H171L/A184I/T260G/A264G/Q307H/N319Y/ A328G
19	R19/F87A/A184I/T260G /I263G	R47L/Y51F/F87A/H171L/A184I/T260G/I263G /Q307H/N319Y
20	R19/F87A/A184I/T260G /I263G/A264G	R47L/Y51F/F87A/H171L/A184I/T260G/I263G/A264G/Q307H/ N319Y
21	R19/F87A/A184I/T260G /I263G/A328G	R47L/Y51F/F87A/H171L/A184I/T260G/I263G/Q307H/N319Y/ A328G
22	R19/F87A/A184I/T260G /T268S	R47L/Y51F/F87A/H171L/T260G/T268S/Q307H/N319Y
	K19/F87A/A82M panel (Chapter 4.2 and 6.2)	
1	K19/F87A/A82M	A82M/F87A/H171L/Q307H/N319Y
2	K19/F87A/A82M/A264G	A82M/F87A/H171L/A264G/Q307H/N319Y
3	K19/F87A/A82M/A264G /A328G	A82M/F87A/H171L/A264G/Q307H/N319Y/A328G
4	K19/F87A/A82M/AW	A82M/F87A/H171L/Q307H/N319Y/A330W

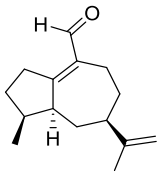
5	K19/F87A/A82M/F261G	A82M/F87A/H171L/F261G/Q307H/N319Y
6	K19/F87A/A82M/I263G /A264G	A82M/F87A/H171L/I263G/A264G/Q307H/N319Y
7	K19/F87A/A82M/I263G /A264G/A328G	A82M/F87A/H171L/I263G/A264G/Q307H/N319Y/A328G
8	K19/F87A/A82M/I263G /A264G/A330W	A82M/F87A/H171L/I263G/A264G/Q307H/N319Y/A330W
9	K19/F87A/A82M/I263G /A328G	A82M/F87A/H171L/I263G/Q307H/N319Y/A328G
10	K19/F87A/A82M/I263G /A330W	A82M/F87A/H171L/I263G/Q307H/N319Y/A330W
11	K19/F87A/A82M/I263W	A82M/F87A/H171L/I263W/Q307H/N319Y
12	K19/F87A/A82M/T260G	A82M/F87A/H171L/T260G/Q307H/N319Y
	RT2 panel (Chapter 6.2)	
1	RT2/S72G/A330W	S72G/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I
2	RT2/S72W/A330W	S72W/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I
3	RT2 /S72V/A330W	S72V/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I
4	RT2 /S72Y/A330W	S72Y/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I
5	RT2 /S72H/A330W	S72H/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I
6	RT2/S72R/A330W	S72R/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I
7	RT2/S72G/A330W/L437LA	S72G/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I /L437LA
8	RT2/S72G/A330W/L437LM	S72G/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I /L437LM
9	RT2/S72G/A330W/L437LV	S72G/R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I /L437LV
10	RT2/S72G/A330W/I263G	S72G/R47L/Y51F/A191T/N239H/I259V/I263G/A276T/A330W /L353I
11	RT2/A330W	R47L/Y51F/A191T/N239H/I259V/A276T/A330W/L353I
12	RT2/L437F	R47L/Y51F/A191T/N239H/I259V/A276T/L353I/L437F
13	RT2/I263G	R47L/Y51F/A191T/N239H/I259V/I263G/A276T/L353I

Appendix 2: The activity and selectivity of bulnesene oxidation

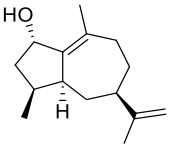
Table 2.1 Conversion (%) and selectivity (%) for each bulnesene metabolite by variants from the progressive screening library. Variants showing <10% bulnesene conversion are not included.



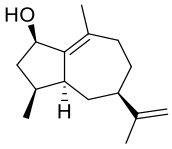
1a



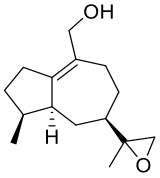
1b



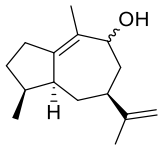
1c



1d



1e



1f

Initial screening_V1								
	Variants	**Conv.**	**1a**	**1b**	**1c**	**1d**	**1e**	**1f**
1	GV/A184I	100	–	–	3	–	15	–
2	GVQ	32	27	–	14	17	–	–
3	GVQ/A264G	62	31	4	27	10	6	14
4	GVQ/I263G	21	–	–	–	–	–	–
5	GVQ/I263G/A328L	79	–	–	–	11	34	–
6	K19/F87A/A82M/E267F	27	21	–	–	12	–	–
7	K19/F87A/A82M/I263G	30	18	–	19	–	–	–
8	K19/F87A/F81W	39	28	–	9	12	18	–
9	K19/F87A/I263A	32	26	–	8	19	12	–
10	K19/F87A/I263G	68	17	–	6	11	41	–
11	K19/F87V	100	5	20	5	–	20	–
12	K19/F87V/A264G	91	14	5	23	6	15	13
13	K19/F87V/A328V	91	17	17	10	18	26	–
14	K19/F87V/E267V	23	12	9	–	–	–	–
15	K19/F87V/V78I/E267V	22	12	–	–	–	–	–
16	K19/F87V/Q403P	83	21	24	13	8	17	7
17	KU3/A330P	55	12	–	37	16	5	–
18	KU3/A330P/A328I	25	10	–	11	8	–	–
19	R19	41	18	–	19	15	–	–
20	R19/A328L/A184I	58	38	–	5	24	8	–
21	R19/F87A	100	12	5	3	–	30	12
22	R19/F87A/A184I	94	7	2	3	–	28	–
23	R19/F87A/A328F	50	23	–	3	39	7	–
24	R19/F87A/A328I	100	29	8	4	34	5	–
25	R19/F87A/A328I/S72W	55	39	7	4	28	–	–
26	R19/F87A/A328I/V78I	90	56	7	–	32	3	–
27	R19/F87A/A328L	100	9	3	–	37	11	–
28	R19/F87A/I263G	95	7	–	–	–	32	–
29	R19/F87I	100	15	51	–	–	6	–
30	R47L/Y51F/I263R	19	–	–	–	–	–	–

	Variants	Conv.	1a	1b	1c	1d	1e	1f
31	RP/A82M/I263A	31	6	–	38	–	–	–
32	RP/F87V	96	4	25	2	–	17	–
33	RP/F87V/V78I	35	16	–	25	16	–	–
34	RP/H171L/I263G	36	8	–	9	15	8	–
35	RP/H171L/A184I/I263G	39	6	–	16	13	6	–
36	RP/F87V/V78I/H171L/I263G	67	8	3	4	8	27	–
37	RP/I263A/E267V	34	29	–	16	4	–	–
38	RP/V78I/E267V	97	7	34	2	–	19	5
39	RP/A82W/I263A	28	–	–	13	8	–	–
40	RT2	95	12	–	14	12	13	–
41	RT2/V78I/A184I/A330P/	89	11	2	22	16	12	–
42	RT2/S72G/A330W	84	20	–	32	15	7	–
43	RT2/S72W/A330W	64	21	–	14	10	6	–
44	WT	20	–	–	–	–	–	–
	GVQ panel							
	Variants	**Conv.**	**1a**	**1b**	**1c**	**1d**	**1e**	**1f**
1	GVQ/A328G	79	16	13	–	–	12	7
2	GVQ/I263G/A328F	84	6	–	9	15	11	–
3	GVQ/I263G/A328I	38	11	–	–	15	–	–
4	GAQ/I263G	51	16	–	–	15	13	–
5	GAQ/I263G/A328G	66	14	–	–	14	10	–
6	GAQ/I263G/A328L	97	–	–	–	–	34	–
7	GLQ/I263G	75	42	–	21	13	–	–
8	GLQ/I263G/A328I	95	28	–	5	7	28	–
9	GLQ/I263G/A328G	88	17	–	11	8	22	–
10	GLQ/I263G/A328L	99	8	–	–	–	52	–
11	GLQ/I263G/A328F	93	31	–	14	17	12	–
12	GGQ/I263G	49	13	–	8	–	–	–
13	GGQ/I263G/A328F	97	–	–	–	–	18	–
14	GGQ/I263G/A328G	48	10	–	6	11	–	–
15	GGQ/I263G/A328L	87	–	–	–	–	–	–
	F87I panel							
	Variants	**Conv.**	**1a**	**1b**	**1c**	**1d**	**1e**	**1f**
1	K19/F87I	79	52	18	4	–	6	–
2	KT2/N70S/F87I	61	45	22	–	–	–	–
	R19/F87A panel_V1							
	Variants	**Conv.**	**1a**	**1b**	**1c**	**1d**	**1e**	**1f**
1	R19/F87A/P329G/A330W	54	14	–	9	8	–	–
2	R19/F87A/P329G/A330P	59	24	–	–	11	–	–
3	R19/F87A/P329V/A330P	45	–	–	–	–	–	–
4	R19/F87A/I263G	90	10	–	–	5	–	–

	Variants	Conv.	1a	1b	1c	1d	1e	1f
4	R19/F87A/I263G	90	10	–	–	5	–	–
5	R19/F87A/I263G/P329G/A330W	69	11	–	–	8	9	–
6	R19/F87A/I263G/P329G/A330P	72	13	–	–	–	–	–
7	R19/F87A/I263G/V127I	84	12	–	–	8	27	–
8	R19/F87A/A191P	74	21	–	9	10	8	8
9	R19/F87A/A328G	30	–	–	–	–	–	–
10	R19/F87A/A264G	87	10	–	7	–	–	26
11	R19/F87A/A264G/A328G	77	10	–	5	–	–	36
12	R19/F87A/Q403P	65	22	–	7	6	–	–
13	R19/F87A/I263G/A264G	89	14	–	–	–	18	–
14	R19/F87A/I263G/A264G/A328G	87	19	–	–	9	–	10
15	R19/F87A/A264G/A328I	99	8	–	7	28	–	–
16	R19/F87A/S72A/A328I	96	20	12	–	19	9	–

R19 panel

	Variants	Conv.	1a	1b	1c	1d	1e	1f
1	R19/A184I/T260G	80	63	–	3	7	7	–
2	R19/I263G/A264G/A328G	82	19	–	11	21	16	–
3	R19/I263W/A328G	25	–	–	–	–	–	–
4	R19/A328G/P329G/A330G	41	22	–	–	31	–	–
5	R19/A328G/P329G/A330G/T260G/Y278H	19	–	–	–	–	–	–
6	R19/T260G/A328G	20	46	–	–	–	–	–
7	R19/I263G	97	–	–	7	–	14	–
8	R19/I263G/A328G	97	4	–	10	4	20	–
9	R19/P329G/A330W	51	14	–	17	38	–	–
10	R19/A82M/A184I/T260G	93	54	14	–	–	6	–
11	R19/A82M/I263G/A264G/A328G	83	45	2	16	15	–	–
12	R19/I263W/A82M/A328G	34	14	–	–	22	–	–
13	R19	55	19	–	21	28	–	–
14	R19/A82M/T260G/A328G/P329G/A330G	18	–	–	–	–	–	–
15	R19/A330W	35	20	–	21	31	–	–
16	R19/A82M/G265GG	10	–	–	–	–	–	–
17	R19/A82M/I263G	25	–	–	–	–	–	–
18	R19/A82M/I263G/A328G	81	47	–	27	–	–	7
19	R19/A82M/P329G/A330W	51	29	–	26	13	–	–
20	R19/A82M/T260G	69	66	–	6	–	–	–

RP panel

	Variants	Conv.	1a	1b	1c	1d	1e	1f
1	RP/A82M/A330W	39	8	–	47	–	–	–
2	RP/A82M/H171L/I263G	61	12	–	57	–	–	–
3	RP/A82M/I263A/A330W	49	10	–	36	9	–	–
4	RP/A82M/I263A/E267F	23	–	–	–	–	–	–

5	RP/H171L/E267F	29	19	–	15	18	–	–
6	RP/I263A/A330W	24	–	–	–	–	–	–
7	RP/S72G/H171L/I263G	55	11	–	10	21	13	–
8	RP/S72W/H171L/I263G	61	15	–	–	22	10	–

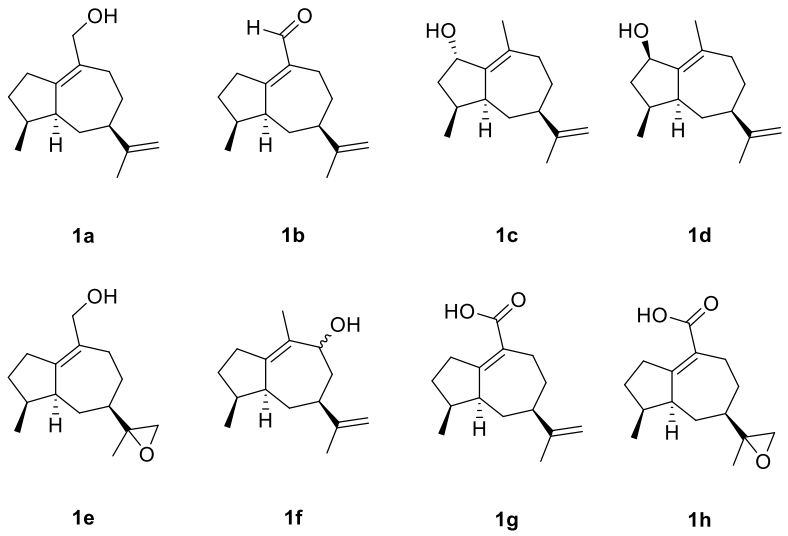
Table 2.2 Substrate conversion (%) and product selectivity (%) of selected variants for the oxidation of bulnesene-15-al **1g**.

	Variants	**Conv.**	**1g**		**Variants**	**Conv.**	**1g**
1	R19/A330W	64	76	**24**	RP/I263A/E267V	70	65
2	K19/F87A/I263G	68	62	**25**	WT	82	49
3	GVQ	60	69	**26**	GVQ/I263G	61	65
4	R19/F87I	70	71	**27**	GVQ/A264G	61	67
5	K19/F87V	70	62	**28**	K19/F87V/A264G	68	56
6	K19/F87V/E267V	100	28	**29**	K19/F87V/V78I/E267V	82	54
7	K19/F87A/A82M/I263G	73	53	**30**	A330P	77	58
8	K19/F87V/A328V	82	50	**31**	K19/F87V/I263G	70	60
9	K19/F87A/F81W	67	56	**32**	R19/F87A/F81W/T260G /A328G	56	100
10	K19/F87A/I263A	59	58	**33**	R19/F87A/V78I/A328I	96	3
11	KU3/I263G/A330P	74	68	**34**	R19/F87I/L437LV	66	63
12	KU3/S72W/A330P	65	68	**35**	GVQ/A330W	72	75
13	R19/A184I/A328L	90	76	**36**	R19/S72G/A330W	47	100
14	R19/F87A/A328G	65	62	**37**	RP/A82M/I263A	84	47
15	R19	59	56	**38**	KU3/I263G/A330P	82	66
16	R19/F87A/A184I	72	48	**39**	RP/H171L/A184I/I263G	67	63
17	GVQ/V78L/I263G	88	61	**40**	I263G	100	57
18	R19/F87A	71	61	**41**	GVQ/I263G/A328L	100	6
19	GV/A184I	88	41	**42**	R19/F87A/A328G/ P329G/A330G	82	46
20	K19/F87V/Q403P	88	44	**43**	GLQ/I263G/A328G	63	59
21	R19/F87A/A184I/ T260G/A328G	65	58	**44**	R19/I263A/A328L	77	59
22	RT2/V78I/A184I/A330P	82	78	**45**	K19/F87A/A82M/A330W	58	69
23	RT2/S72G/A330W	62	58				

Table 2.3 Variants and mutations in the new generation library.

	Variants	Mutations
1	R19/A82M	R47L/Y51F/H171L/Q307H/N319Y/A82M
2	R19/F87I/A82M	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M
3	R19/F87I/A82I	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82I
4	R19/F87I/A184I	R47L/Y51F/F87I/H171L/Q307H/N319Y/A184I
5	R19/F87I/A184F	R47L/Y51F/F87I/H171L/Q307H/N319Y/A184F
6	R19/F87I/A184M	R47L/Y51F/F87I/H171L/Q307H/N319Y/A184M
7	R19/F87I/V26L	R47L/Y51F/F87I/H171L/Q307H/N319Y/V26L
8	R19/F87I/V26M	R47L/Y51F/F87I/H171L/Q307H/N319Y/V26M
9	R19/F87I/S72V	R47L/Y51F/F87I/H171L/Q307H/N319Y/S72V
10	R19/F87I/S72I	R47L/Y51F/F87I/H171L/Q307H/N319Y/S72I
11	R19/F87I/S72L	R47L/Y51F/F87I/H171L/Q307H/N319Y/S72L
12	R19/F87I/A74L	R47L/Y51F/F87I/H171L/Q307H/N319Y/A74L
13	R19/F87I/T88L	R47L/Y51F/F87I/H171L/Q307H/N319Y/T88L
14	R19/F87I/T88F	R47L/Y51F/F87I/H171L/Q307H/N319Y/T88F
15	R19/F87I/T88M	R47L/Y51F/F87I/H171L/Q307H/N319Y/T88M
16	R19/F87I/L181F	R47L/Y51F/F87I/H171L/Q307H/N319Y/L181F
17	R19/F87I/T260G	R47L/Y51F/F87I/H171L/Q307H/N319Y/T260G
18	R19/F87I/T260L	R47L/Y51F/F87I/H171L/Q307H/N319Y/T260L
19	R19/F87I/I263W	R47L/Y51F/F87I/H171L/Q307H/N319Y/I263W
20	R19/F87I/A264F	R47L/Y51F/F87I/H171L/Q307H/N319Y/A264F
21	R19/F87I/A330V	R47L/Y51F/F87I/H171L/Q307H/N319Y/A330V
22	R19/F87I/A330I	R47L/Y51F/F87I/H171L/Q307H/N319Y/A330I
23	R19/F87I/A330F	R47L/Y51F/F87I/H171L/Q307H/N319Y/A330F
24	R19/F87I/S332L	R47L/Y51F/F87I/H171L/Q307H/N319Y/S332L
25	R19/F87I/S332I	R47L/Y51F/F87I/H171L/Q307H/N319Y/S332I
26	R19/F87I/A82M/A184I	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I
27	R19/F87I/A82M/A330I	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A330I
28	R19/F87I/A82M/A330V	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A330V
29	R19/F87I/A184IV26L	R47L/Y51F/F87I/H171L/Q307H/N319Y/A184I/V26L
30	R19/F87I/A184I/V26M	R47L/Y51F/F87I/H171L/Q307H/N319Y/A184I/V26M
31	R19/F87I/A184I/S72L	R47L/Y51F/F87I/H171L/Q307H/N319Y/A184I/S72L
32	R19/F87I/A184I/A330I	R47L/Y51F/F87I/H171L/Q307H/N319Y/A184I/A330I
33	R19/F87I/A184I/A330V	R47L/Y51F/F87I/H171L/Q307H/N319Y/A184I/A330V
34	R19/F87I/A82M/A184I/A330I	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I
35	R19/F87I/A82M/A184I/A330V	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330V
36	R19/F87I/A82M/A330I/V26L	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A330I/V26L
37	R19/F87I/A82M/A330V/V26L	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A330V/V26L
38	R19/F87I/A82M/A330I/V26M	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A330I/V26M

Table 2.4 Conversion (%) and selectivity (%) of each bulnesene metabolite by variants from new generation library.



	CYP102A1 variant	Conv.	1a	1b	1c	1d	1e	1f	1g	1h	C15
1	R19/A82M	86	42	7	31	–	–	10	–	–	49
2	R19/F87I/A82M	86	–	7	–	–	–	–	70	–	77
3	R19/F87I/A82I	86	14	32	–	–	–	–	35	–	81
4	R19/F87I/A184I	91	–	3	–	–	–	–	72	–	75
5	R19/F87I/A184F	78	34	36	–	–	–	–	9	–	79
6	R19/F87I/A184M	93	8	15	–	–	–	–	47	–	70
7	R19/F87I/V26L	92	16	48	–	–	–	–	10	–	74
8	R19/F87I/V26M	93	20	46	–	–	–	–	10	–	76
9	R19/F87I/S72V	60	70	6	9	–	–	–	–	–	76
10	R19/F87I/S72I	84	53	17	10	–	3	–	–	–	70
11	R19/F87I/S72L	76	61	13	10	4	–	–	–	–	74
12	R19/F87I/A74L	91	31	43	–	–	–	–	5	–	79
13	R19/F87I/T88L	85	55	25	5	–	–	–	–	–	80
14	R19/F87I/T88F	93	65	25	–	–	–	–	–	–	90
15	R19/F87I/T88M	96	21	31	–	–	–	–	17	–	69
16	R19/F87I/L181F	84	18	30	–	–	–	–	21	–	69
17	R19/F87I/T260G	33	72	–	–	–	–	–	–	–	72
18	R19/F87I/T260L	81	44	30	–	–	–	–	4	–	78
19	R19/F87I/I263W	28	36	–	–	–	–	–	–	–	36
20	R19/F87I/A264F	19	59	–	–	–	–	–	–	–	59
21	R19/F87I/A330V	93	10	45	–	–	–	–	30	–	85
22	R19/F87I/A330I	90	11	45	–	–	–	–	26	–	82
23	R19/F87I/A330F	87	12	13	–	–	–	–	24	–	49
24	R19/F87I/S332L	17	58	–	–	–	–	–	–	–	58
25	R19/F87I/S332I	32	71	–	–	–	–	–	–	–	71
26	R19/F87I/A82M/A184I	83	33	19	–	–	–	–	30	–	82
27	R19/F87I/A82M/A330I	71	4	7	–	–	–	–	69	–	80

28	R19/F87I/A82M/A330V	83	4	7	–	–	–	–	68	–	79
29	R19/F87I/A184I/V26L	98	–	–	–	–	–	–	70	–	70
30	R19/F87I/A184I/V26M	96	–	4	–	–	–	–	72	–	76
31	R19/F87I/A184I/S72L	95	28	52	–	–	–	–	12	–	92
32	R19/F87I/A184I/A330I	76	11	13	–	–	–	–	57	–	81
33	R19/F87I/A184I/A330V	94	–	3	–	–	–	–	78	–	81
34	R19/F87I/A82M/A184I/A330I	82	–	–	–	–	–	–	31	49	80
35	R19/F87I/A82M/A184I/A330V	93	7	6	–	–	–	–	42	34	89
36	R19/F87I/A82M/A330I/V26L	87	–	5	–	–	–	–	58	19	82
37	R19/F87I/A82M/A330V/V26L	90	–	7	–	–	–	–	57	17	81
38	R19/F87I/A82M/A330I/V26M	85	–	9	–	–	–	–	62	12	83

Bulnesene-15-ol, **1a**

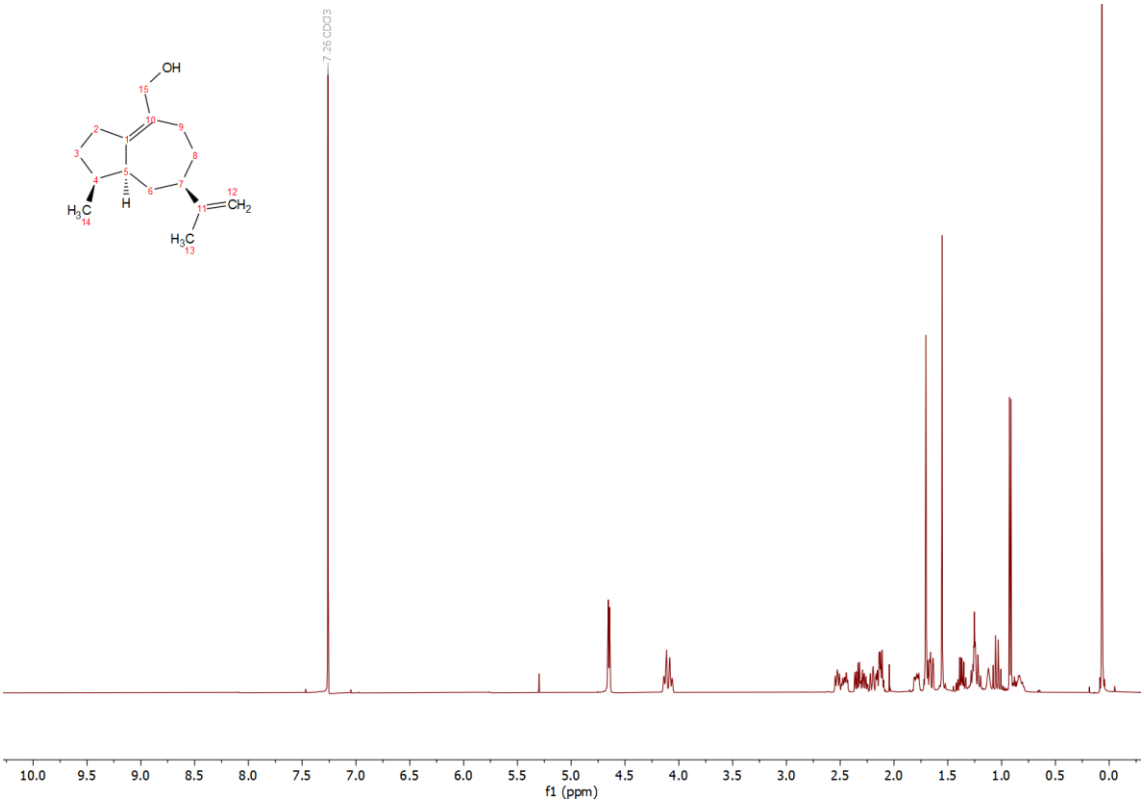


Figure 2.1 The ^{1}H NMR spectrum of **1a** in $CDCl_3$.

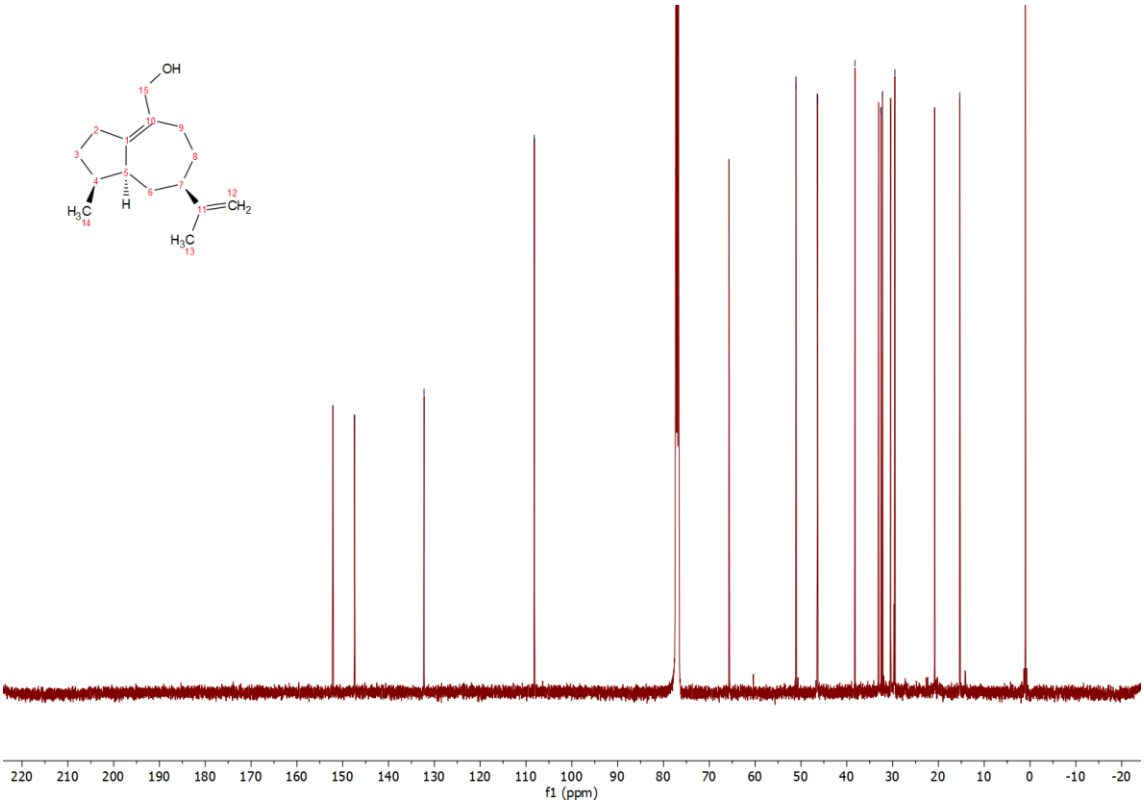


Figure 2.2 The ^{13}C NMR spectrum of **1a** in $CDCl_3$.

Bulnesene-15-al, **1b**

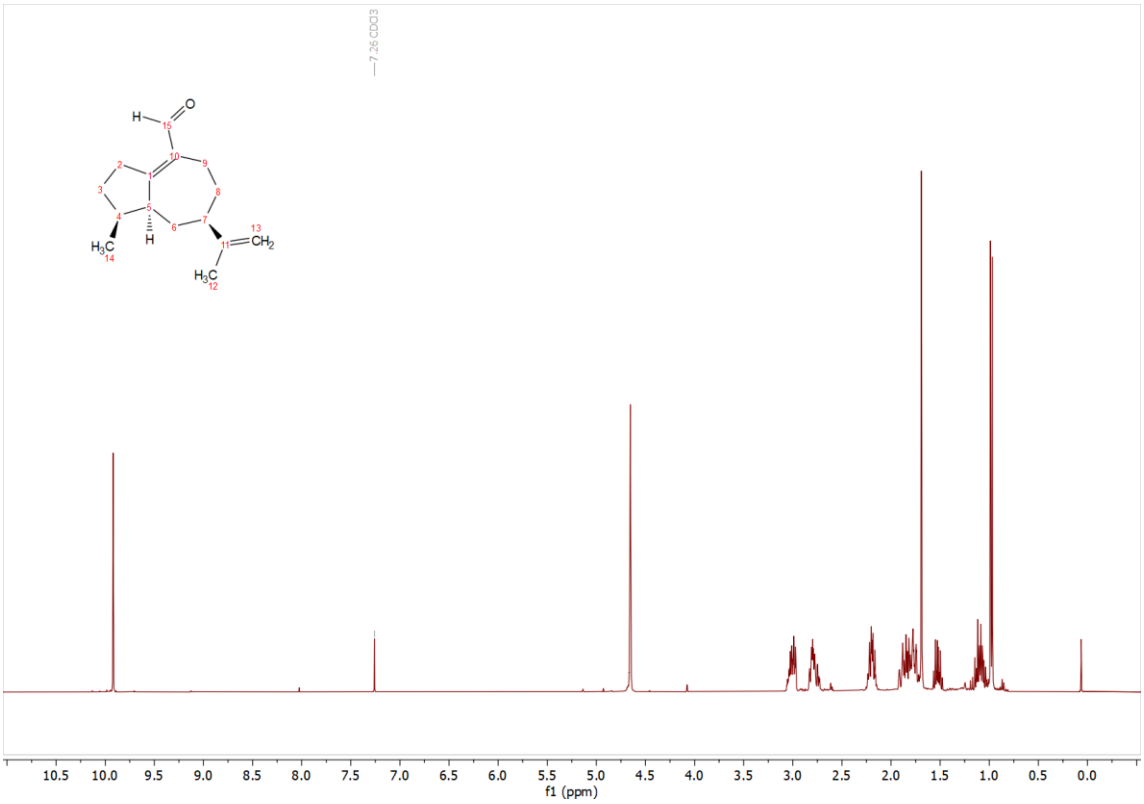


Figure 2.3 The ^{1}H NMR spectrum of **1b** in $CDCl_3$.

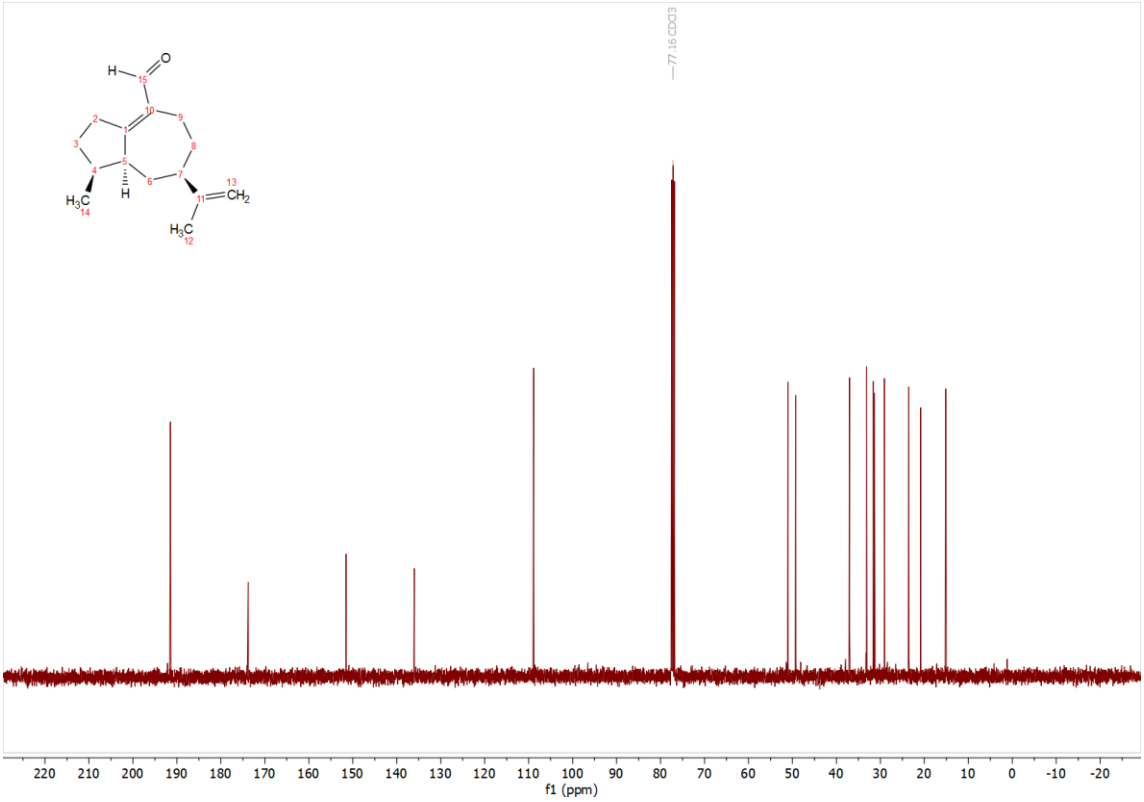


Figure 2.4 The ^{13}C NMR spectrum of **1b** in $CDCl_3$.

(*2S*)-Hydroxy-α-bulnesene, **1c**

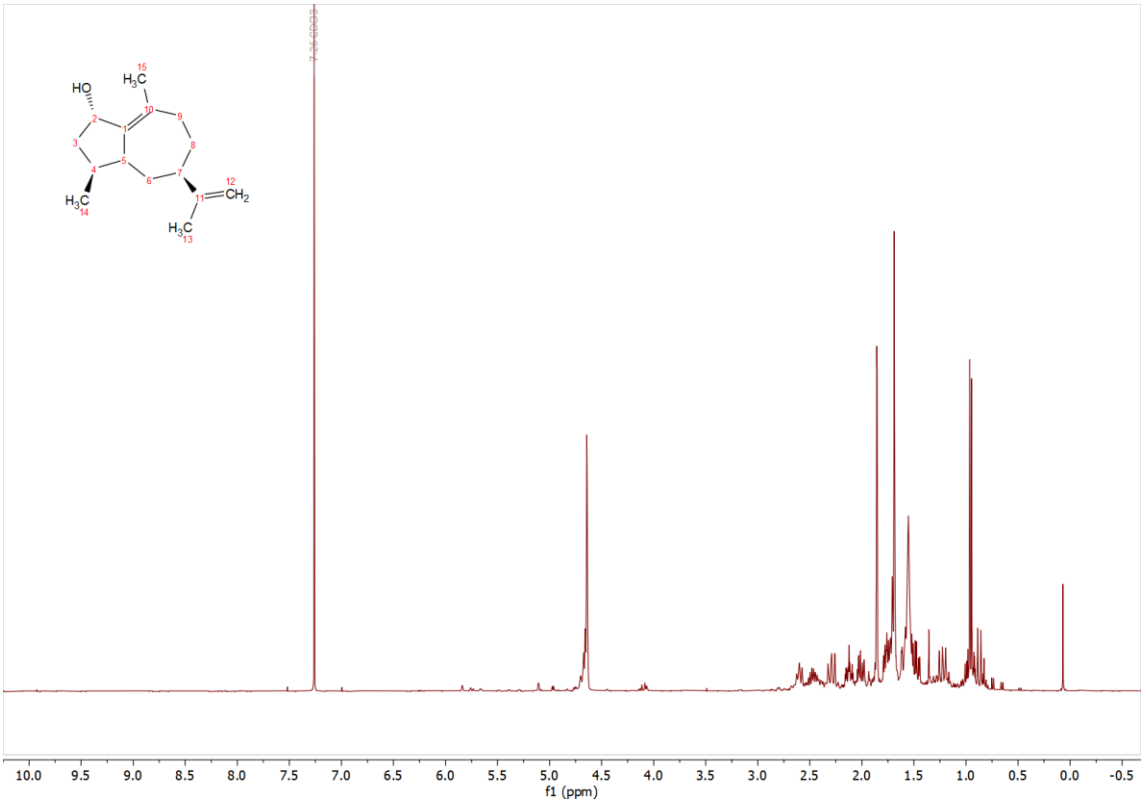


Figure 2.5 The ^{1}H NMR spectrum of **1c** in $CDCl_3$.

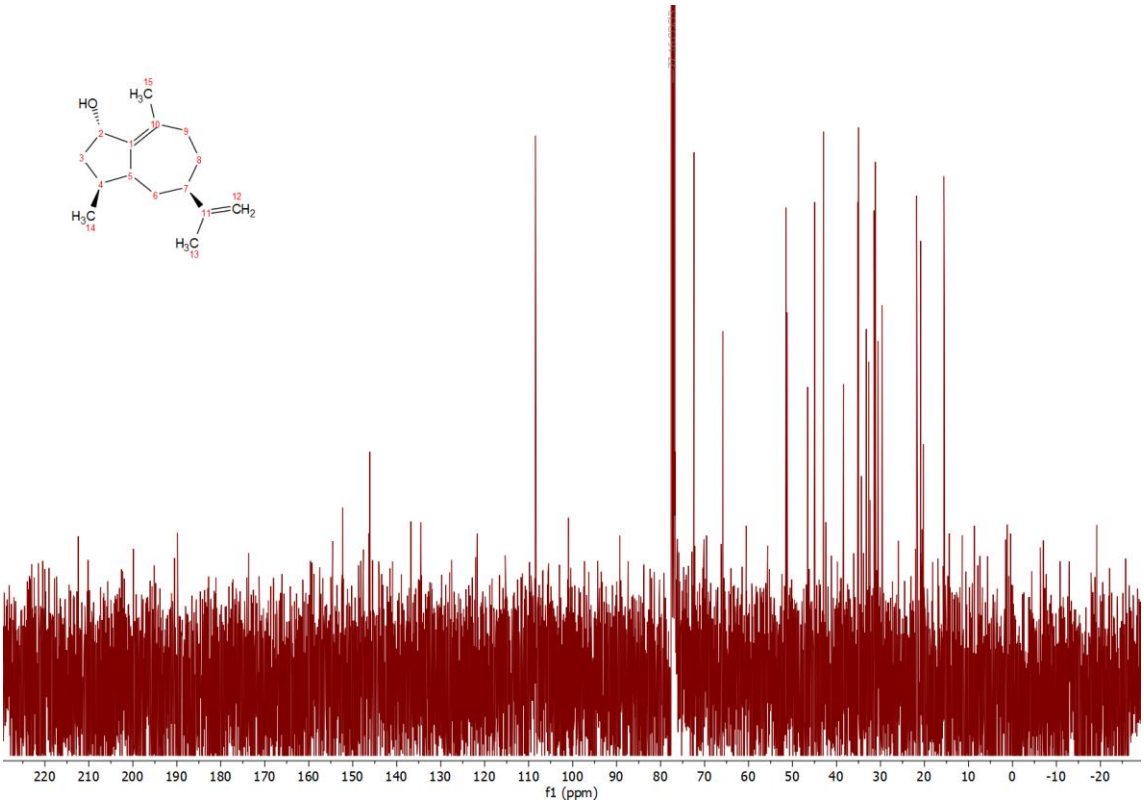


Figure 2.6 The ^{13}C NMR spectrum of **1c** in $CDCl_3$.

(*2R*)-Hydroxy-α-bulnesene, **1d**

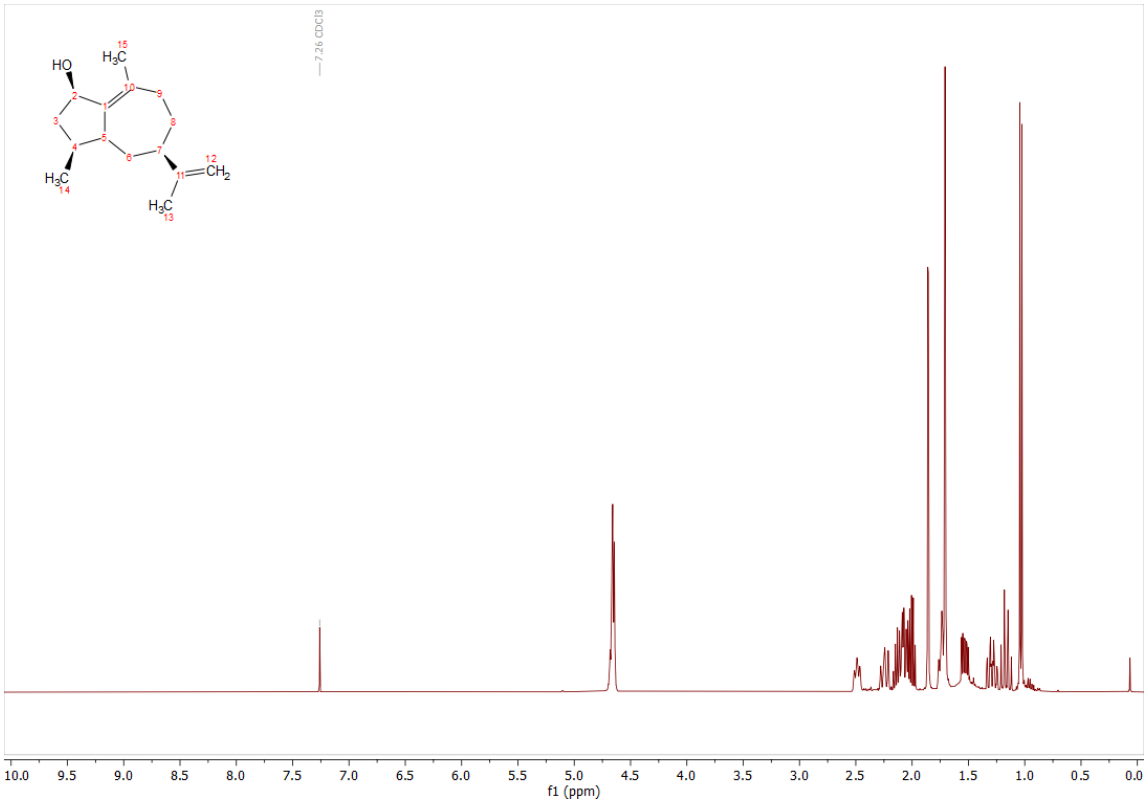


Figure 2.7 The ^{1}H NMR spectrum of **1d** in $CDCl_3$.

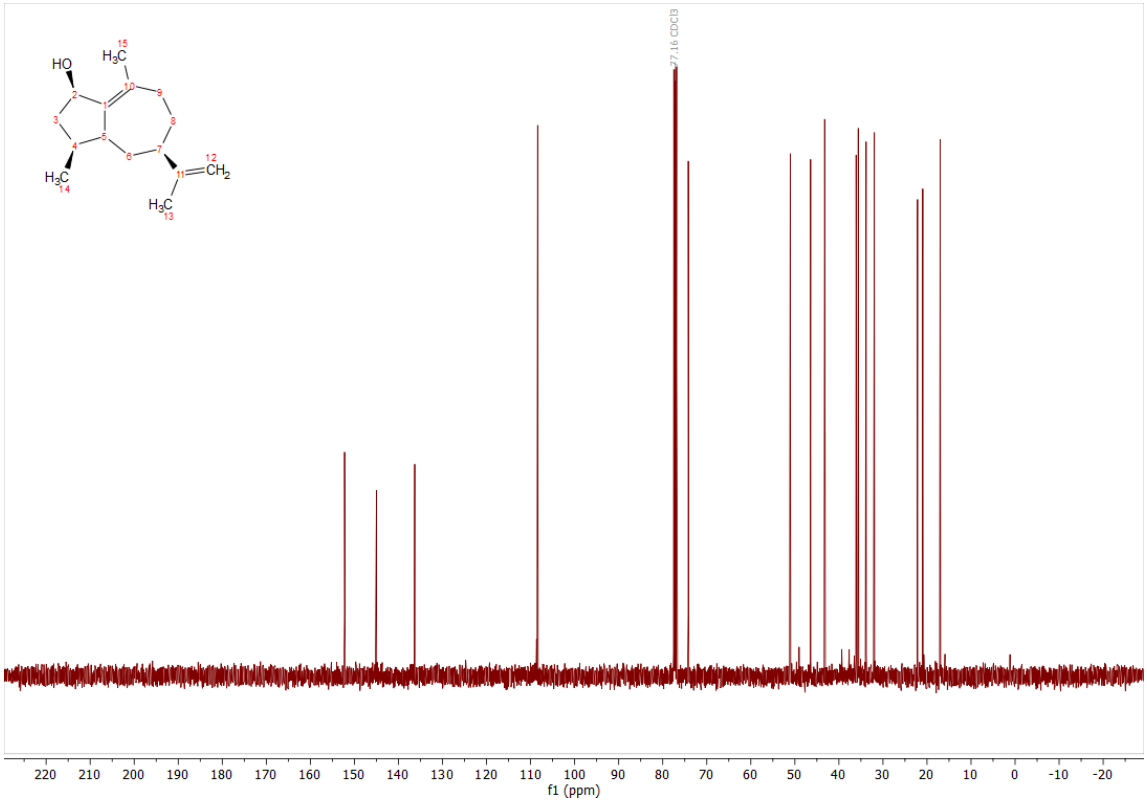


Figure 2.8 The ^{13}C NMR spectrum of **1d** in $CDCl_3$.

11,12-Epoxy-15-hydroxy-α-bulnesene, **1e**

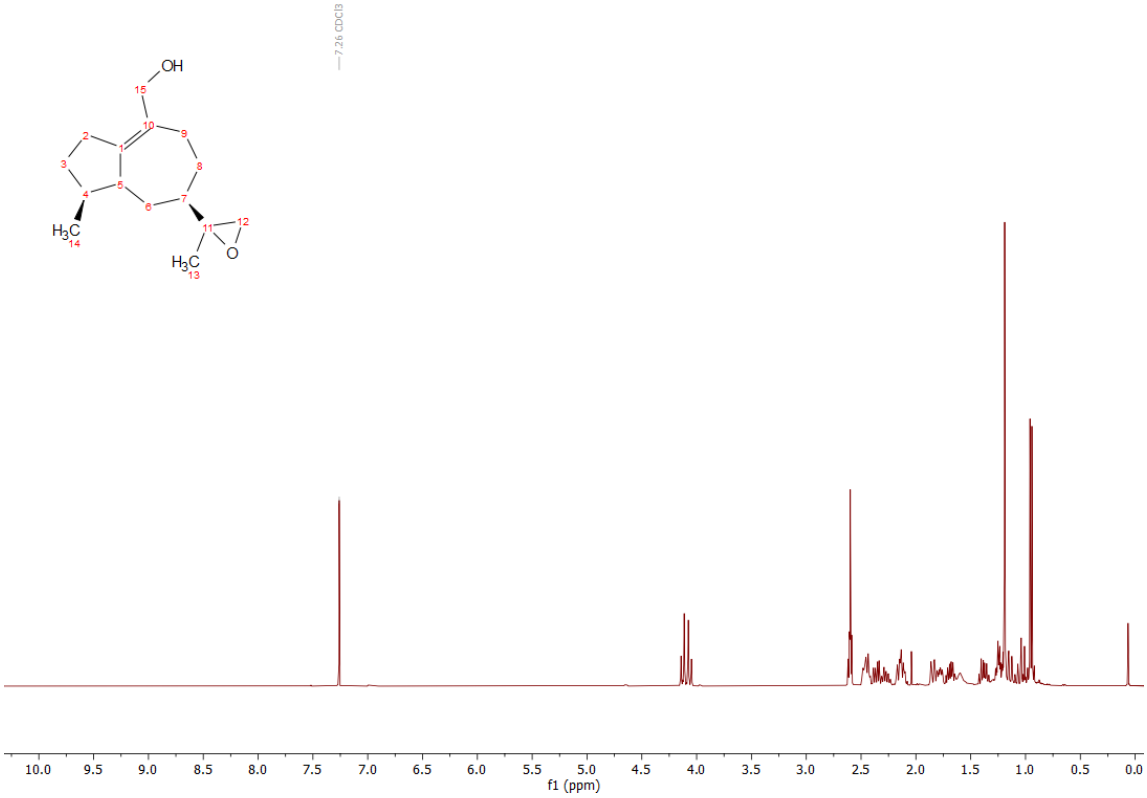


Figure 2.9 The ^{1}H NMR spectrum of **1e** in $CDCl_3$.

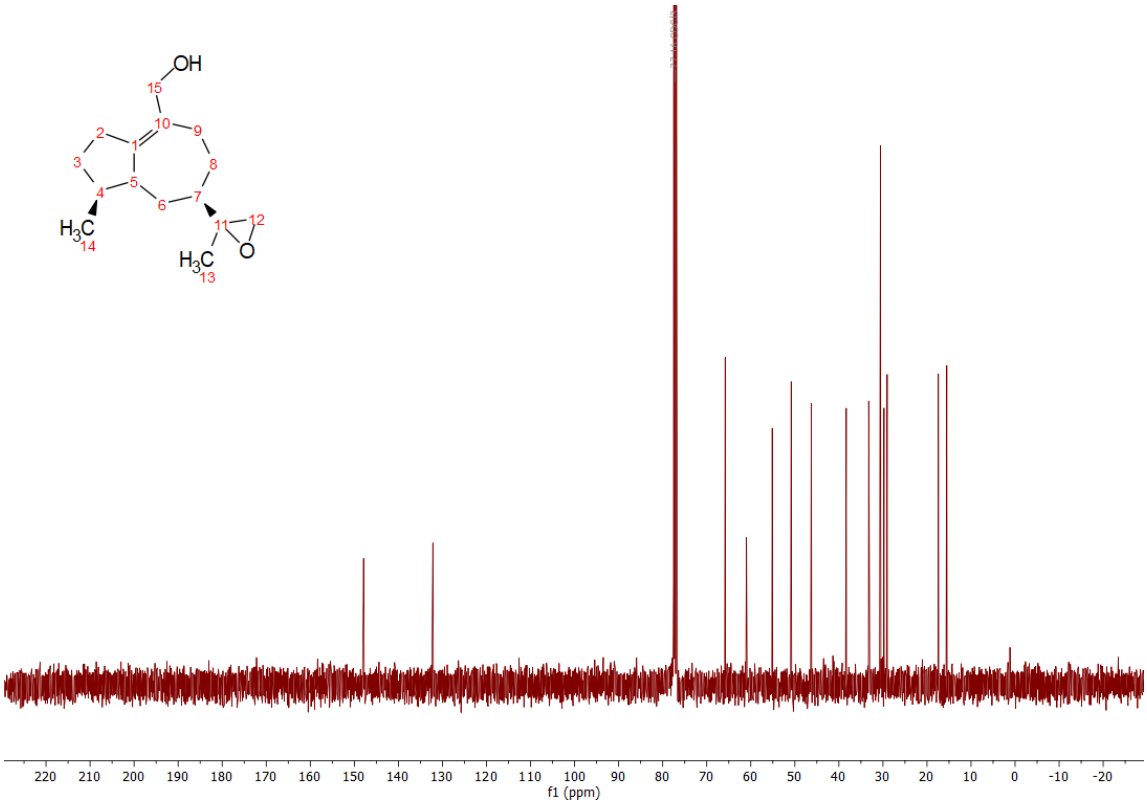


Figure 2.10 The ^{13}C NMR spectrum of **1e** in $CDCl_3$.

9-Hydroxy-α-bulnesene, **1f**

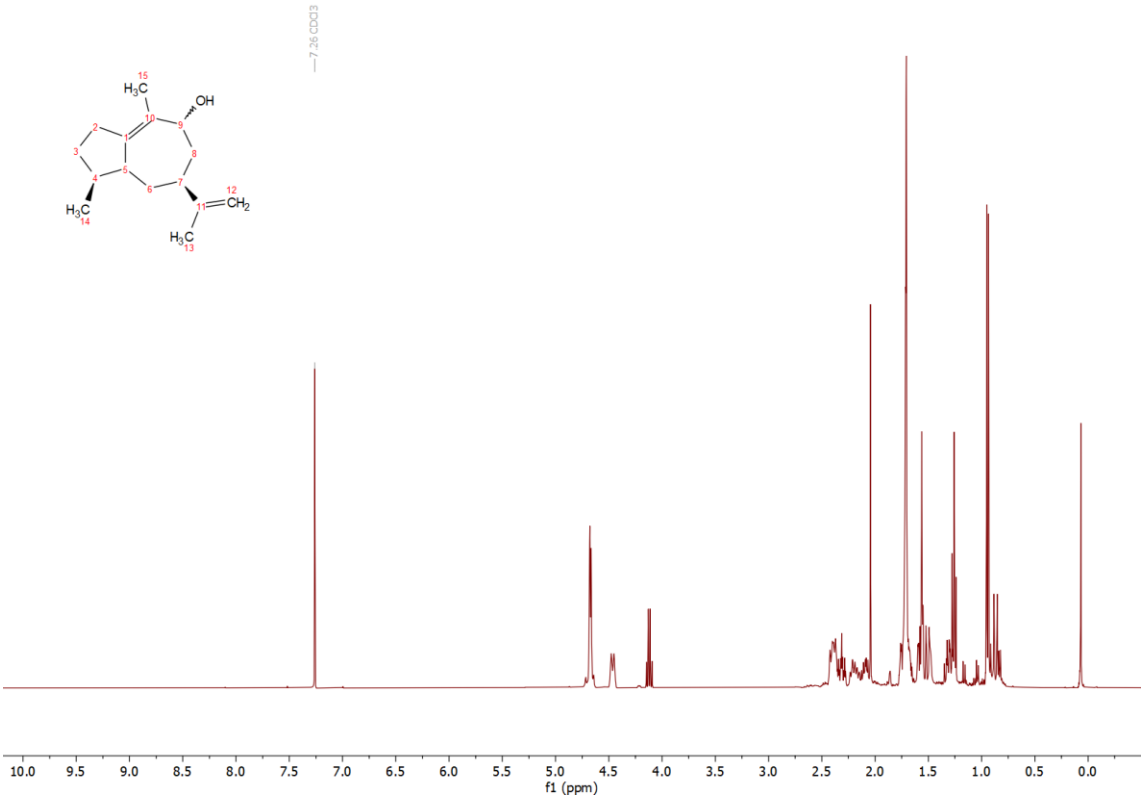


Figure 2.11 The ^{1}H NMR spectrum of **1f** in $CDCl_3$.

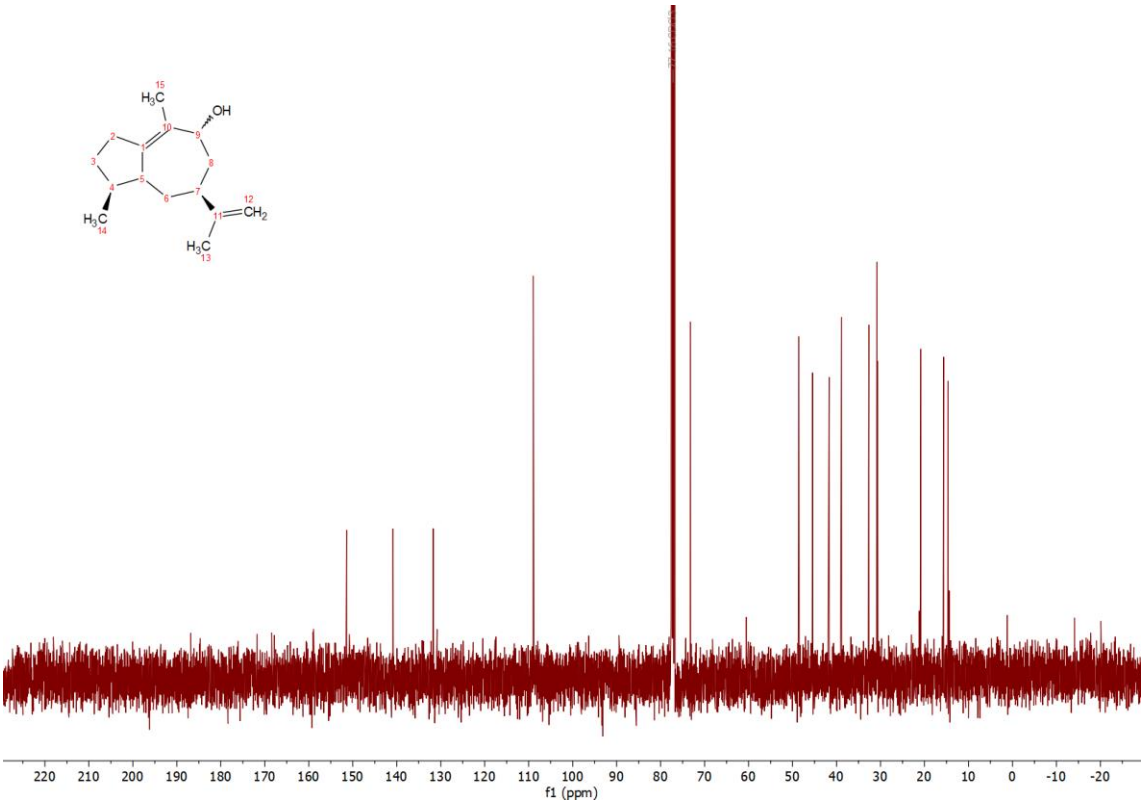


Figure 2.12 The ^{13}C NMR spectrum of **1f** in $CDCl_3$.

Bulnesene-15-COOH, **1g**

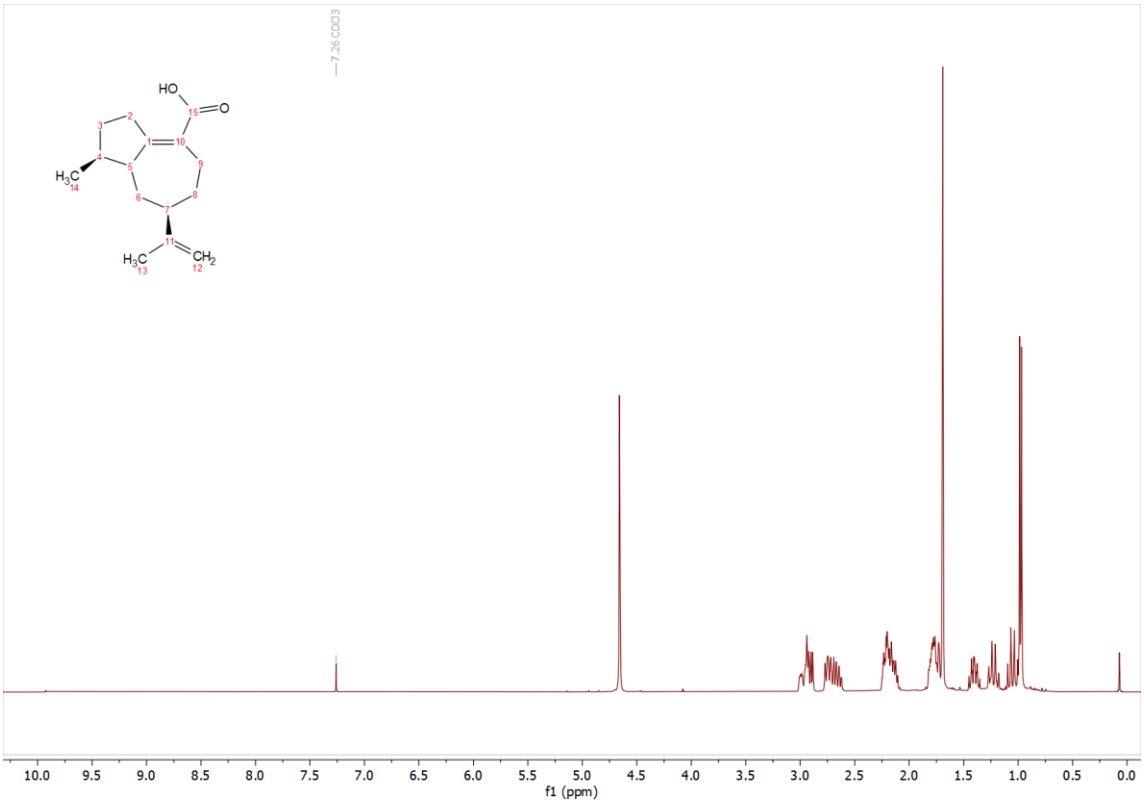


Figure 2.13 The ^{1}H NMR spectrum of **1g** in $CDCl_3$.

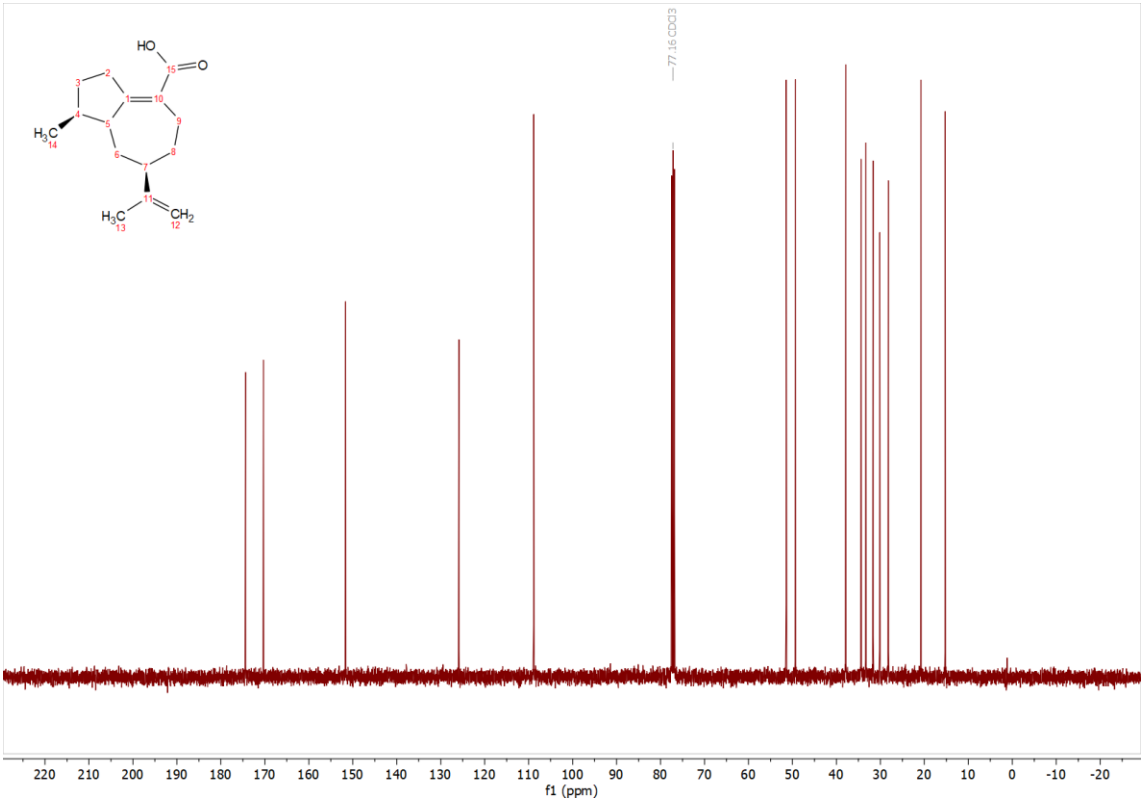


Figure 2.14 The ^{13}C NMR spectrum of **1g** in $CDCl_3$.

11,12-Epoxy-bulnesene-15-COOH, **1h**

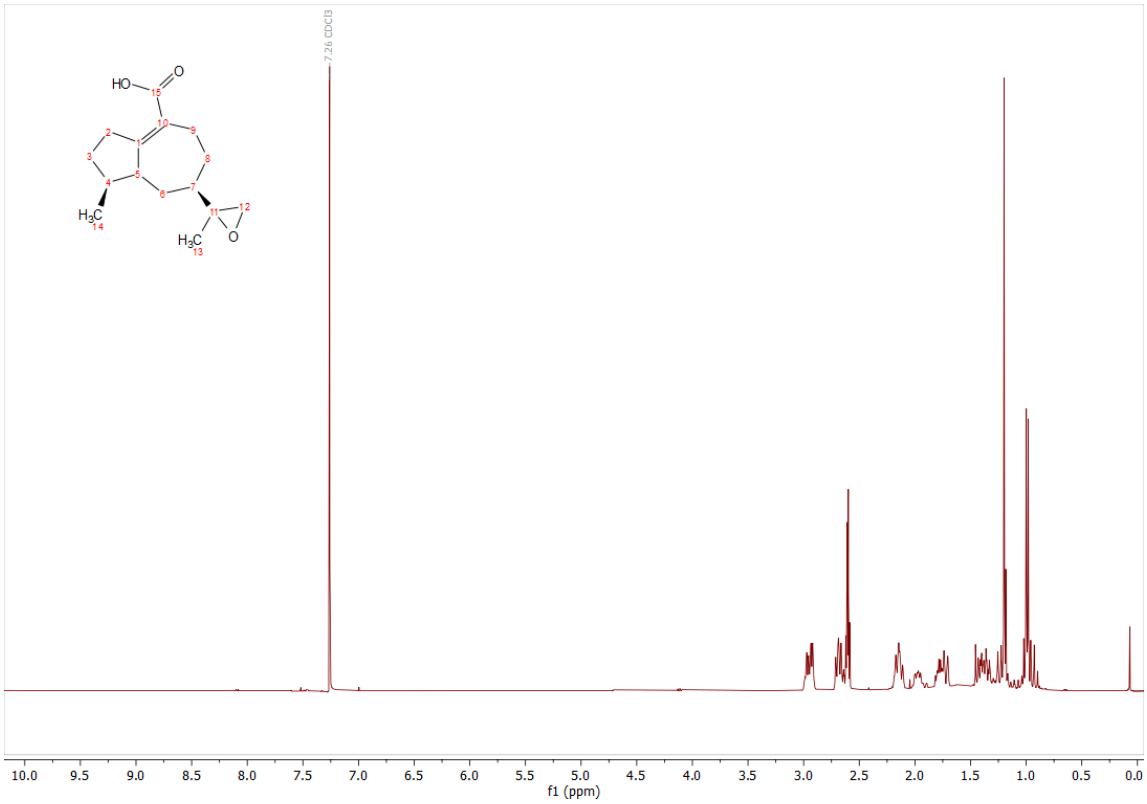


Figure 2.15 The ^{1}H NMR spectrum of **1h** in $CDCl_3$.

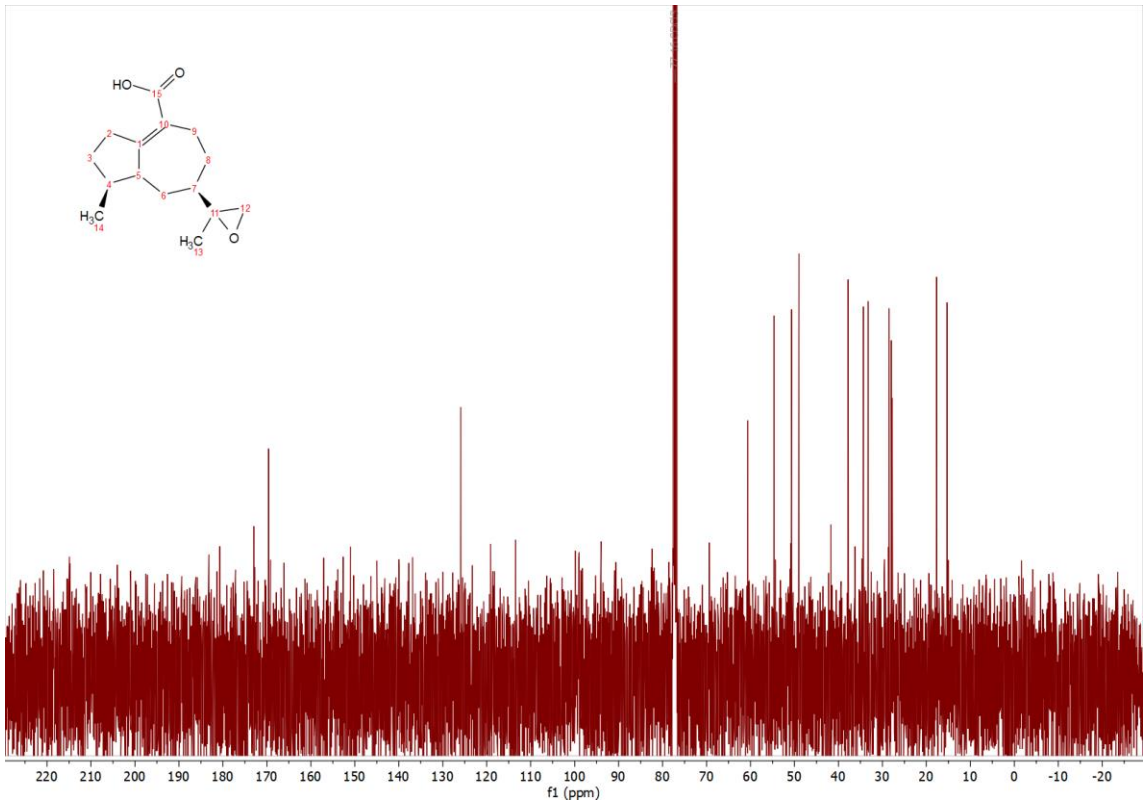


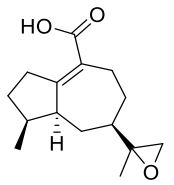
Figure 2.16 The ^{13}C NMR spectrum of **1h** in $CDCl_3$.

Appendix 3: Synthesis of bulnesene-15,2-lactone, 1j

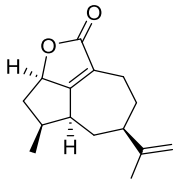
Table 3.1 Variants and mutations in the new generation library.

	Variants	Mutations
1	F87A/A184I/A330P	F87A/A184I/A330P
2	F87A/A184I/I263G/A264G /A328G	F87A/A184I/I263G/A264G/A328G
3	F87A/A184I/I263G/A264G /A330P	F87A/A184I/I263G/A264G/A330P
4	F87A/A184I/I263G/A330P	F87A/A184I/I263G/A330P
5	F87A/A330P	F87A/A330P
6	F87A/I263G	F87A/I263G
7	F87A/I263G/A264G	F87A/I263G/A264G
8	F87A/I263G/A330P	F87A/I263G/A330P
9	R19/F87A/A184I/I263G/A264G /A330P	R47L/Y51F/F87A/H171L/A184I/I263G/A264G/Q307H/N319Y /A330P
10	R19/F87A/A184I/I263G/A330P	R47L/Y51F/F87A/H171L/A184I/I263G/Q307H/N319Y/A330P
11	R19/F87I/A82M/A184I/A330I/ L71F	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ L71F
12	R19/F87I/A82M/A184I/A330I/ L71M	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ L71M
13	R19/F87I/A82M/A184I/A330I/ S72V	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ S72V
14	R19/F87I/A82M/A184I/A330I/ S72W	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ S72W
15	R19/F87I/A82M/A184I/A330I/ V178F	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ V178F
16	R19/F87I/A82M/A184I/A330I/ L188F	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ L188F
17	R19/F87I/A82M/A184I/A330I/ S332M	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ S332M
18	R19/F87I/A82M/A184I/A330I/ S332I	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ S332I
19	R19/F87I/A82M/A184I/A330I/ S332L	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ S332L
20	R19/F87I/A82M/A184I/A330I/ L437V	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ L437V
21	R19/F87I/A82M/A184I/A330I/ L437F	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ L437F
22	R19/F87I/A82M/A184I/A330I/ L437M	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ L437M
23	R19/F87I/A82M/A184I/A330I/ T438F	R47L/Y51F/F87I/H171L/Q307H/N319Y/A82M/A184I/A330I/ T438F

Table 3.2 Conversion (%) and selectivity (%) of each bulnesne-15-COOH metabolite by variants from the progressive screening library. Variants with <10% conversion are not included.



1h/1h(a)



1j

Initial Screening_V2					
	Variants	**Conv.**	**1j**	**1h(a)**	**1h**
1	GV/A184I/I263G/A328G	16	–	2	81
2	GVQ/I263G/A328L	38	–	40	36
3	K19/F87A/I263G	29	–	15	69
4	K19/F87V/E267V	100	–	–	96
5	K19/F87V/I263G	29	–	–	93
6	K19/F87V/Q403P	41	–	7	93
7	K19/F87V/V78I/E267V	67	–	21	79
8	KT2/L188G/I263G	54	–	6	87
9	KU3/S72W/A330P	41	–	2	89
10	R19/F87A	28	–	10	76
11	R19/F87A/A184I	52	5	22	52
12	R19/F87A/A328G	51	–	2	90
13	R19/F87A/A328L	13	–	17	25
14	R19/F87A/I263G	83	6	20	47
15	R19/F87A/V78I/A328I	14	–	3	56
16	RP/H171L/I263G	21	–	5	95
R19/F87A panel_V2					
	Variants	**Conv.**	**1j**	**1h(a)**	**1h**
1	R19/F87A/A264G	20	–	–	100
2	R19/F87A/T260G/I263G/A264G /A328G	57	–	–	7
3	R19/F87A/L262G/A328G	31	–	14	81
4	R19/F87A/Q403P	22	–	16	72
5	R19/F87A/T269G	56	–	3	80
R19/F87A/A184I (panel)					
	Variants	**Conv.**	**1j**	**1h(a)**	**1h**
1	R19/F87A/A184I/A264G	32	–	–	64
2	R19/F87A/A184I/A330P	97	–	9	21
3	R19/F87A/A184I/A82M/L181F/T260G/A328G	23	–	–	13
4	R19/F87A/A184I/A82M/T260G	19	–	–	–

5	R19/F87A/A184I/A82M/T260G/A328G/A330I	17	–	–	55
6	R19/F87A/A184I/A82M/T260G/A328G/A330L	21	–	–	55
7	R19/F87A/A184I/A82M/T260G/A328G/S332M	19	–	–	–
8	R19/F87A/A184I/I263G/A264G	53	14	1	41
9	R19/F87A/A184I/I263G/A264G/A328G	98	55	–	9
10	R19/F87A/A184I/I263G/A328G	42	15	5	19
11	R19/F87A/A184I/S72F/A82M/T260G/A328G	32	–	–	41
12	R19/F87A/A184I/S72F/T260G/A264G/A328G	15	–	–	–
13	R19/F87A/A184I/T260G/A264G	29	–	–	–
14	R19/F87A/A184I/T260G/I263G	23	–	–	–
15	R19/F87A/A184I/T260G/I263G/A264G	10	–	–	–
16	R19/F87A/A184I/T260G/I263G/A328G	12	–	–	–
	K19/F87A/A82M panel				
	Variants	**Conv.**	**1j**	**1h(a)**	**1h**
1	K19/F87A/A82M/A264G	97	–	–	95
2	K19/F87A/A82M/A264G/A328G	67	–	–	86
3	K19/F87A/A82M/A330W	30	–	–	57
4	K19/F87A/A82M/I263G/A264G	14	–	–	35
5	K19/F87A/A82M/I263G/A264G/A330W	21	–	–	29
6	K19/F87A/A82M/I263G/A328G	8	–	–	–
7	K19/F87A/A82M/I263G/A330W	7	–	–	–
8	K19/F87A/A82M/I263W	13	–	–	–

Diastereoisomer of 11,12-epoxy-bulnesene-15-COOH, **1h(a)**

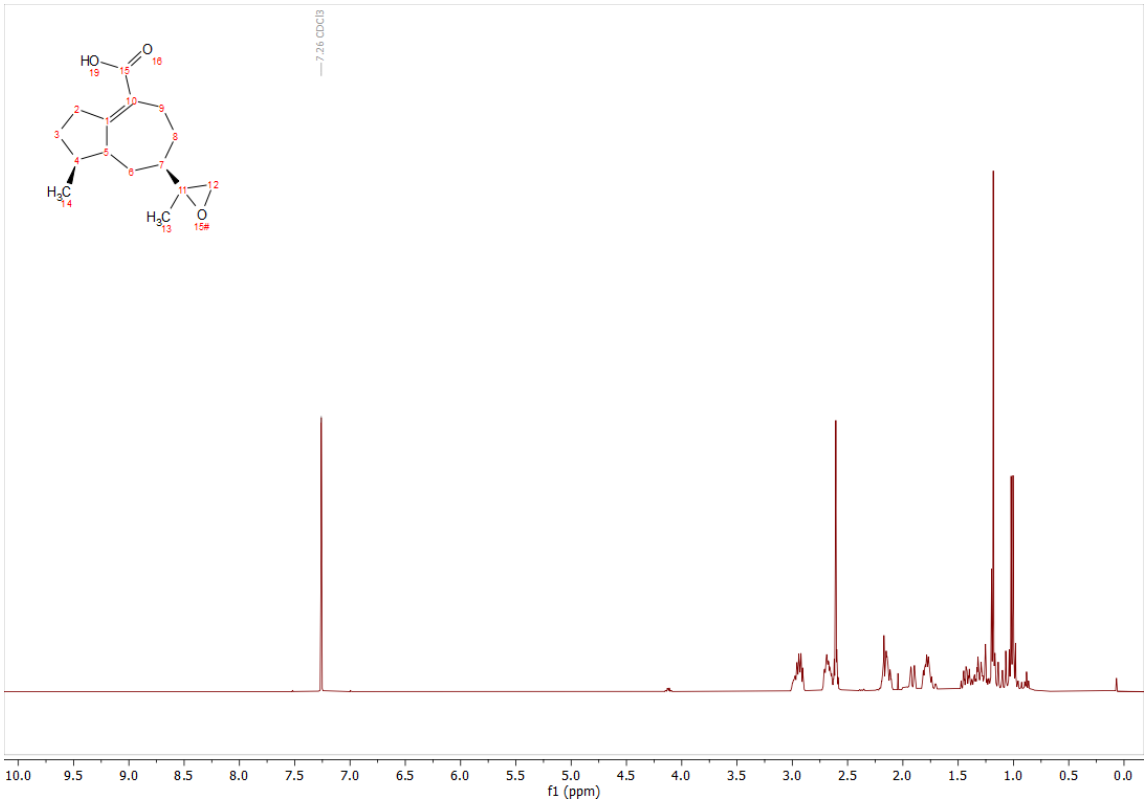


Figure 3.1 The ^{1}H NMR spectrum of **1h(a)** in $CDCl_3$.

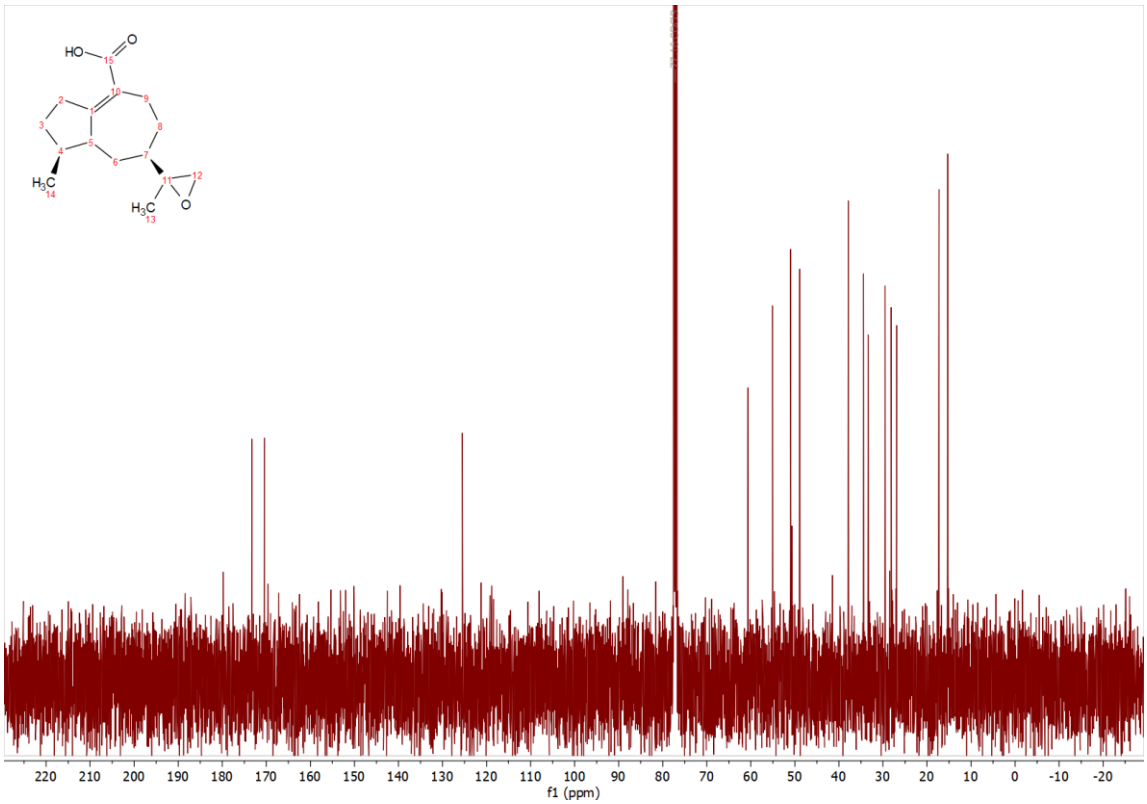


Figure 3.2 The ^{13}C NMR spectrum of **1h(a)** in $CDCl_3$.

2-Hydroxy-bulnsene-15-COOH, **1i**

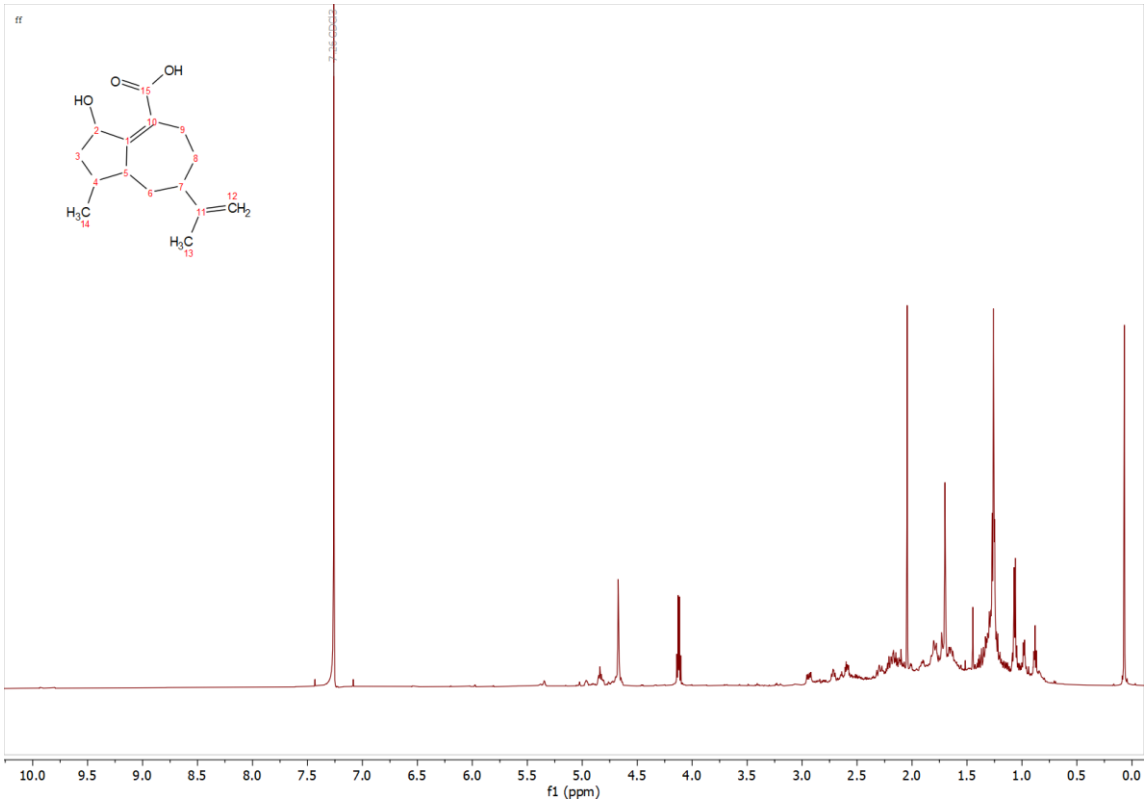


Figure 3.3 The ^{1}H NMR spectrum of **1i** in $CDCl_3$.

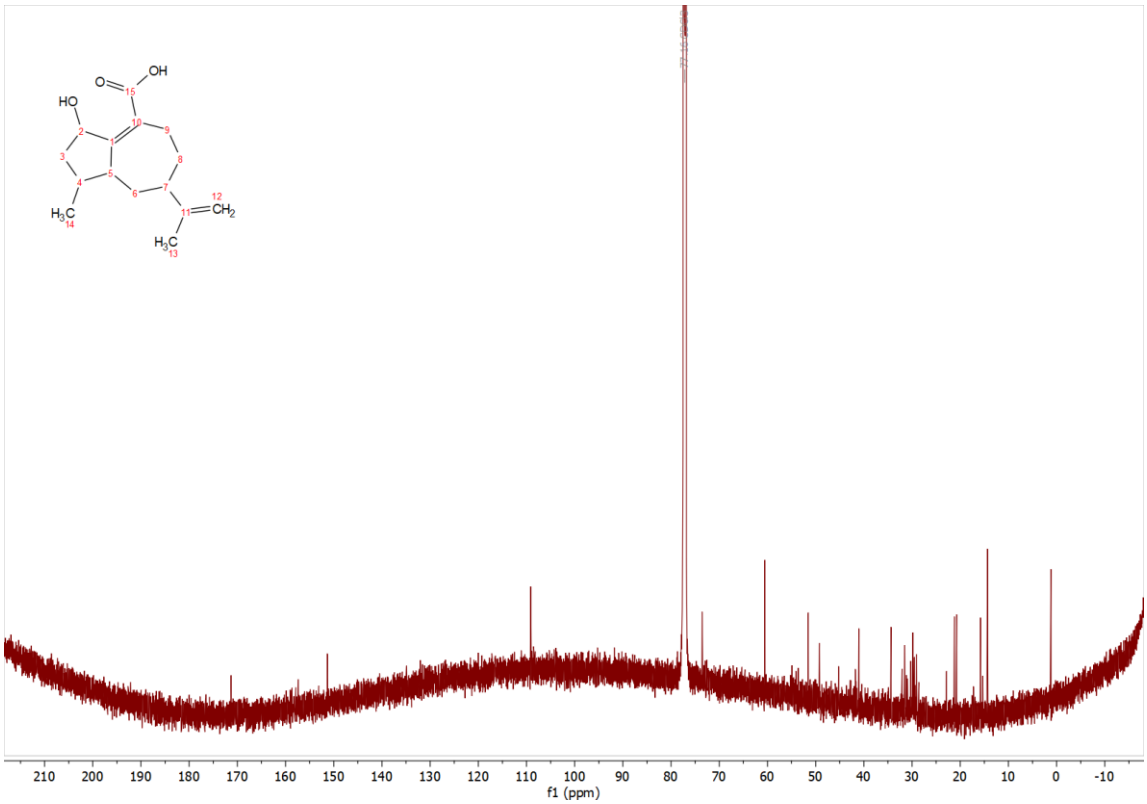


Figure 3.4 The ^{13}C NMR spectrum of **1i** in $CDCl_3$.

Bulnesene-15,2-lactone, **1j**

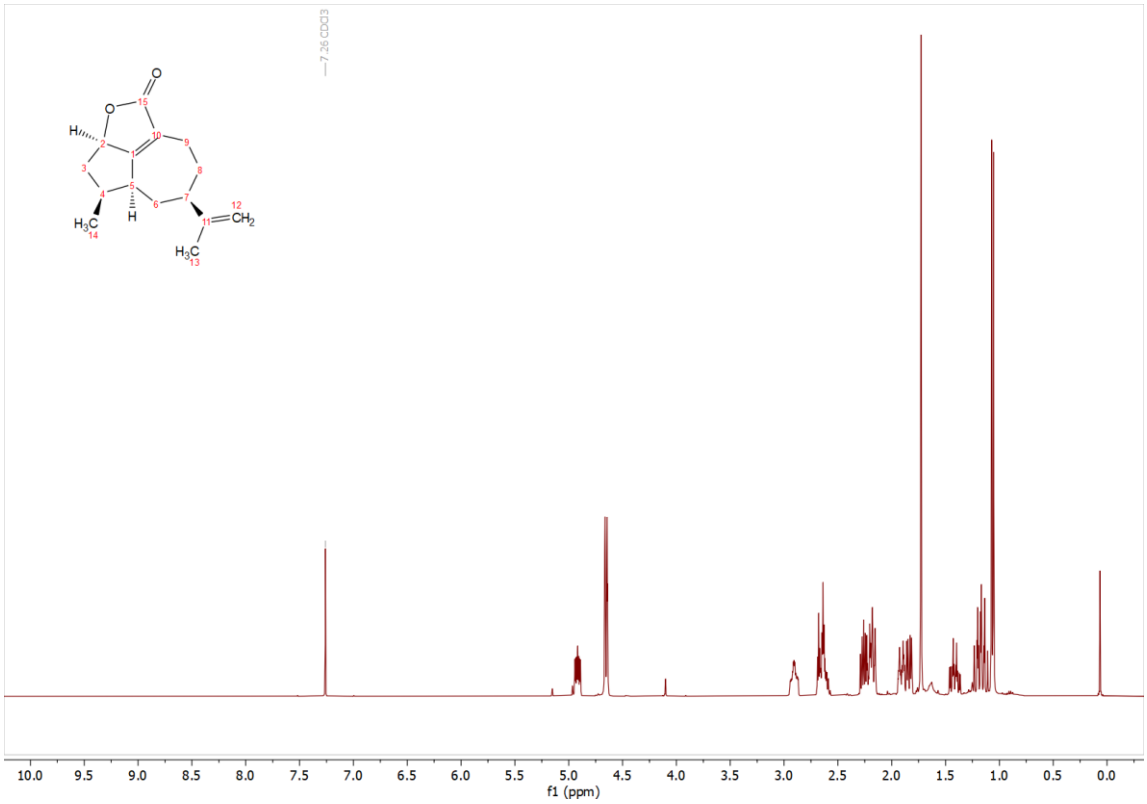


Figure 3.5 The ^{1}H NMR spectrum of **1j** in $CDCl_3$.

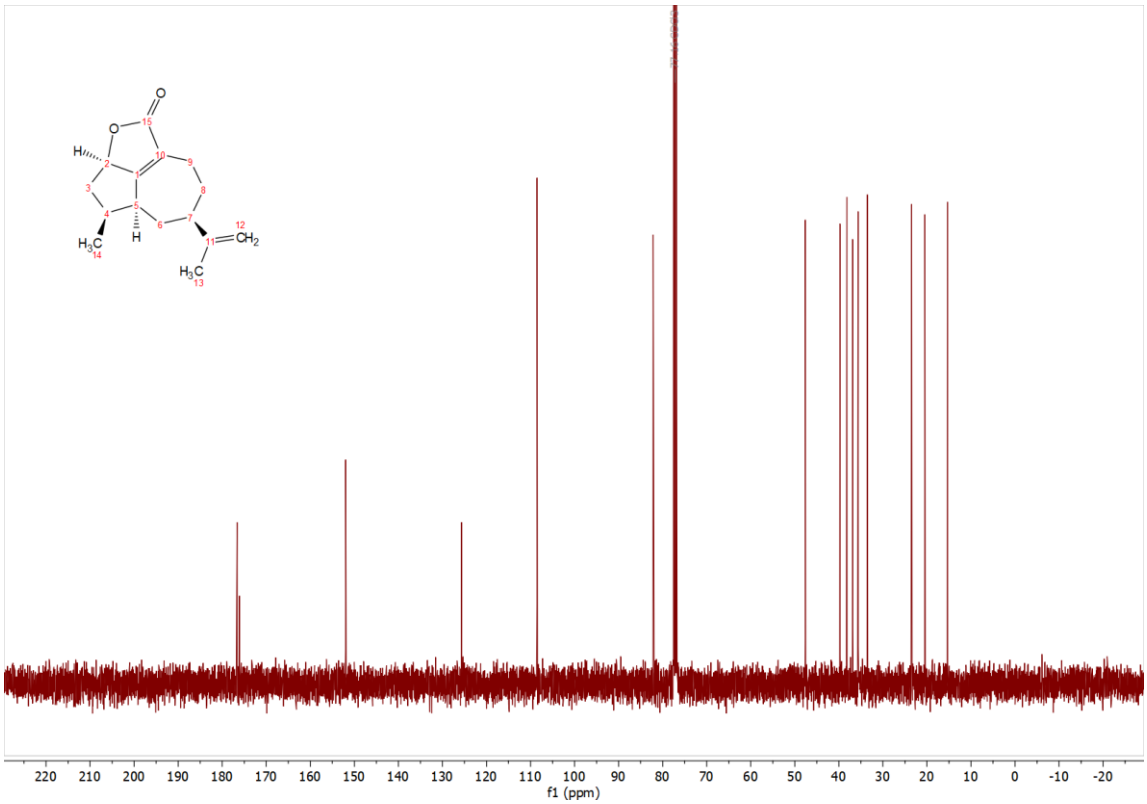


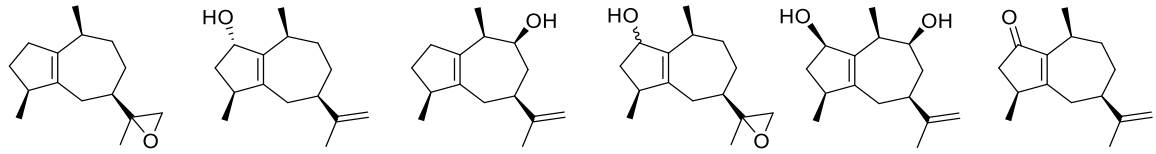
Figure 3.6 The ^{13}C NMR spectrum of **1j** in $CDCl_3$.

Appendix 4: Oxidation of guaiene, 2

Table 4.1 Glycine library assembled for the guaiene oxidation.

	Variants	Mutations
1	GV/A184I/I263G/A264G	A74G/F87V/A184I/I263G/A264G
2	GV/A184I/I263G/A328G	A74G/F87V/A184I/I263G/A328G
3	GV/A184I/A264G/A328G	A74G/F87V/A184I/A264G/A328G
4	GV/A184I/I263G/A264G/A328G	A74G/F87V/A184I/I264G/A264G/A328G
5	R19/F87A/A184I/I263G/A264G	R47L/Y51F/F87A/H171L/A184I/I263G/A264G/Q307H/N319Y
6	R19/F87A/A184I/I263G/A328G	R47L/Y51F/F87A/H171L/A184I/I263G/Q307H/N319Y/A328G
7	R19/F87A/A184I/A264G/A328G	R47L/Y51F/F87A/H171L/A184I/A264G/Q307H/N319Y/A328G
8	R19/F87A/A184I/I263G/A264G/ A328G	R47L/Y51F/F87A/H171L/A184I/I263G/A264G/Q307H/N319Y/ A328G
9	GVQ/I263G/A264G	A74G/F87V/L188Q/I263G/A264G
10	GVQ/I263G/A328G	A74G/F87V/L188Q/I263G/A328G
11	GVQ/I263G/A264G/A328G	A74G/F87V/L188Q/I263G/A264G/A328G

Table 4.2 Conversion (%) and selectivity of each guaiene metabolite (%) by variants from the progressive screening library. Variants with <10% conversion are not included.



2a **2b** **2c** **2d** **2e** **2f**

Initial screening_V1								
		Conv.	**2a**	**2b**	**2c**	**2d**	**2e**	**2f**
1	GV/A184I	97	–	–	5	11	27	9
2	GVQ	74	9	6	–	15	–	–
3	GVQ/A264G	96	5	7	22	33	5	–
4	GVQ/I263G	94	12	7	–	45	1	–
5	GVQ/I263G/A328L	90	78	6	–	7	–	4
6	K19/F87A/A82M/E267F	45	4	68	–	–	–	6
7	K19/F87A/A82M/I263G	42	26	27	21	6	8	6
8	K19/F87A/F81W	92	6	18	18	9	4	6
9	K19/F87A/I263A	76	7	13	10	20	8	2
10	K19/F87A/I263G	27	42	42	–	–	–	–
11	K19/F87V	86	21	35	8	6	14	4
12	K19/F87V/A264G	96	5	9	13	38	5	–
13	K19/F87V/A328V	51	5	50	4	–	–	5
14	K19/F87V/E267V	94	–	3	7	9	22	5
15	K19/F87V/V78I/E267V	31	31	26	22	5	5	–
16	K19/F87V/Q403P	68	27	22	–	3	–	–
17	KU3/A330P	90	30	35	–	18	–	–

18	KU3/A328I/A330P	93	–	5	18	11	21	6
		Conv.	**2a**	**2b**	**2c**	**2d**	**2e**	**2f**
19	R19	82	39	2	–	29	10	–
20	R19/A184I/A328L	60	25	56	–	–	–	5
21	R19/F87A	100	–	17	2	13	17	8
22	R19/F87A/A184I	53	42	41	–	–	–	–
23	R19/F87A/A328F	84	50	21	–	11	–	–
24	R19/F87A/A328I	95	–	3	–	12	2	–
25	R19/F87A/S72W/A328I	95	–	37	3	7	4	7
26	R19/F87A/V78I/A328I	95	2	41	16	4	–	6
27	R19/F87A/A328L	97	11	39	4	21	–	–
28	R19/F87A/I263G	95	3	2	–	10	–	–
29	R19/F87I	75	17	31	21	6	–	–
30	R47L/Y51F/I263R	89	38	28	–	13	–	–
31	RP/A82M/I263A	28	7	58	–	–	–	–
32	RP/F87V	91	5	7	12	11	16	3
33	RP/F87V/V78I	49	44	28	–	4	–	–
34	RP/H171L/I263G	51	22	34	–	–	–	–
35	RP/H171L/A184I/I263G	72	14	9	–	27	–	–
36	RP/F87V/V78I/H171L/I263G	81	7	5	3	20	5	–
37	RP/I263A/E267V	79	19	20	–	29	3	–
38	RP/V78I/E267V	89	9	11	18	11	7	–
39	RP/A82W/I263A	45	11	37	–	–	–	–
40	RT2	100	–	–	5	11	28	5
41	RT2/V78I/A184I/A330P	13	36	32	–	–	–	–
42	RT2/S72G/A330W	98	5	50	–	14	–	–
43	RT2/S72W/A330W	90	10	57	–	23	–	–
	GVQ panel							
		Conv.	**2a**	**2b**	**2c**	**2d**	**2e**	**2f**
1	GVQ/A328G	90	2	16	31	7	11	–
2	GVQ/I263G/A328F	100	–	–	–	23	11	–
3	GVQ/I263G/A328I	100	10	6	–	76	2	–
4	GAQ/I263G	94	5	4	–	14	–	–
5	GAQ/I263G/A328G	95	4	5	–	18	–	–
6	GAQ/I263G/A328L	100	–	–	–	3	–	–
7	GLQ/I263G	96	28	52	–	17	–	–
8	GLQ/I263G/A328I	100	8	45	–	43	–	–
9	GLQ/I263G/A328G	100	7	48	–	16	7	–
10	GLQ/I263G/A328L	100	19	2	–	69	5	–
11	GLQ/I263G/A328F	96	71	5	–	21	–	–
12	GGQ/I263G	88	3	24	–	22	–	–
13	GGQ/I263G/A328F	97	1	4	–	8	–	–
14	GGQ/I263G/A328G	88	–	44	–	11	–	–
15	GGQ/I263G/A328L	98	–	–	–	3	–	–

RP panel								
		Conv.	**2a**	**2b**	**2c**	**2d**	**2e**	**2f**
1	RP/A82M/A330W	35	5	66	–	–	–	–
2	RP/A82M/H171L/I263G	58	5	73	–	5	–	–
3	RP/A82M/I263A/A330W	42	5	68	–	–	–	–
4	RP/A82M/I263A/E267F	30	–	23	–	–	–	–
5	RP/A82M/I263G	36	6	80	–	–	–	14
6	RP/H171L/E267F	29	12	35	–	–	–	–
7	RP/I263A/A330W	27	–	32	–	–	–	–
8	RP/S72G/H171L/I263	53	18	36	–	25	–	–
9	RP/S72W/H171L/I263G	86	8	16	–	46	–	–
RT2 panel								
		Conv.	**2a**	**2b**	**2c**	**2d**	**2e**	**2f**
1	RT2 /S72V/A330W	94	11	63	–	16	–	–
2	RT2 /S72Y/A330W	94	15	59	–	18	–	–
3	RT2 /S72H/A330W	97	7	59	–	9	–	–
4	RT2/S72R/A330W	23	–	22	–	–	–	–
5	RT2/S72G/A330W/L437LA	81	19	64	–	12	–	–
6	RT2/S72G/A330W/L437LM	14	34	56	–	–	–	–
7	RT2/S72G/A330W/I263G	70	21	37	–	36	–	–
8	RT2/A330W	70	21	48	–	4	–	–
9	RT2/L437F	100	2	57	–	37	–	–
10	RT2/I263G	100	1	4	–	59	–	–
11	RT2 /I263G/A330W	100	–	3	–	35	–	–
12	RT2/I263A/A330W	36	15	44	–	–	–	–
K19/F87A/A82M panel								
		Conv.	**2a**	**2b**	**2c**	**2d**	**2e**	**2f**
1	K19/F87A/A82M	13	–	78	–	–	–	21
2	K19/F87A/A82M/A264G	80	2	31	–	7	–	13
3	K19/F87A/A82M/A264G/A328G	90	–	23	–	5	–	18
4	K19/F87A/A82M/AW	31	–	40	–	9	–	–
5	K19/F87A/A82M/F261G	75	–	45	–	5	4	12
6	K19/F87A/A82M/I263G/A264G	89	6	35	–	6	–	11
7	K19/F87A/A82M/I263G/A264G/A328G	79	5	32	–	5	–	16
8	K19/F87A/A82M/I263G/A264G/A330W	39	11	39	–	14	–	5
9	K19/F87A/A82M/I263G/A328G	88	–	31	–	–	–	19
10	K19/F87A/A82M/I263G/A330W	21	–	42	–	–	–	–
11	K19/F87A/A82M/T260G	75	–	62	–	3	–	7
Glycine panel								
		Conv.	**2a**	**2b**	**2c**	**2d**	**2e**	**2f**
1	GV/A184I/I263G/A264G	92	16	23	–	15	–	6
2	GV/A184I/I263G/A328G	90	9	20	–	7	–	11
3	GV/A184I/A264G/A328G	73	4	24	7	8	–	6
4	GV/A184I/I263G/A264G/A328G	73	7	21	–	4	4	–
5	R19/F87A/A184I/I263G/A264G	91	5	18	–	7	–	12

		Conv.	2a	2b	2c	2d	2e	2f
6	R19/F87A/A184I/I263G/A328G	95	4	15	–	16	–	3
7	R19/F87A/A184I/A264G/A328G	69	2	23	–	3	–	13
8	R19/F87A/A184I/I263G/A264G/A328G	79	6	18	–	11	–	5
9	GVQ/I263G/A264G	90	22	21	–	11	–	5
10	GVQ/I263G/A328G	78	17	44	15	4	–	–
11	GVQ/I263G/A264G/A328G	80	14	32	–	5	–	14

11,12-Epoxy-α-guaiene, **2a**

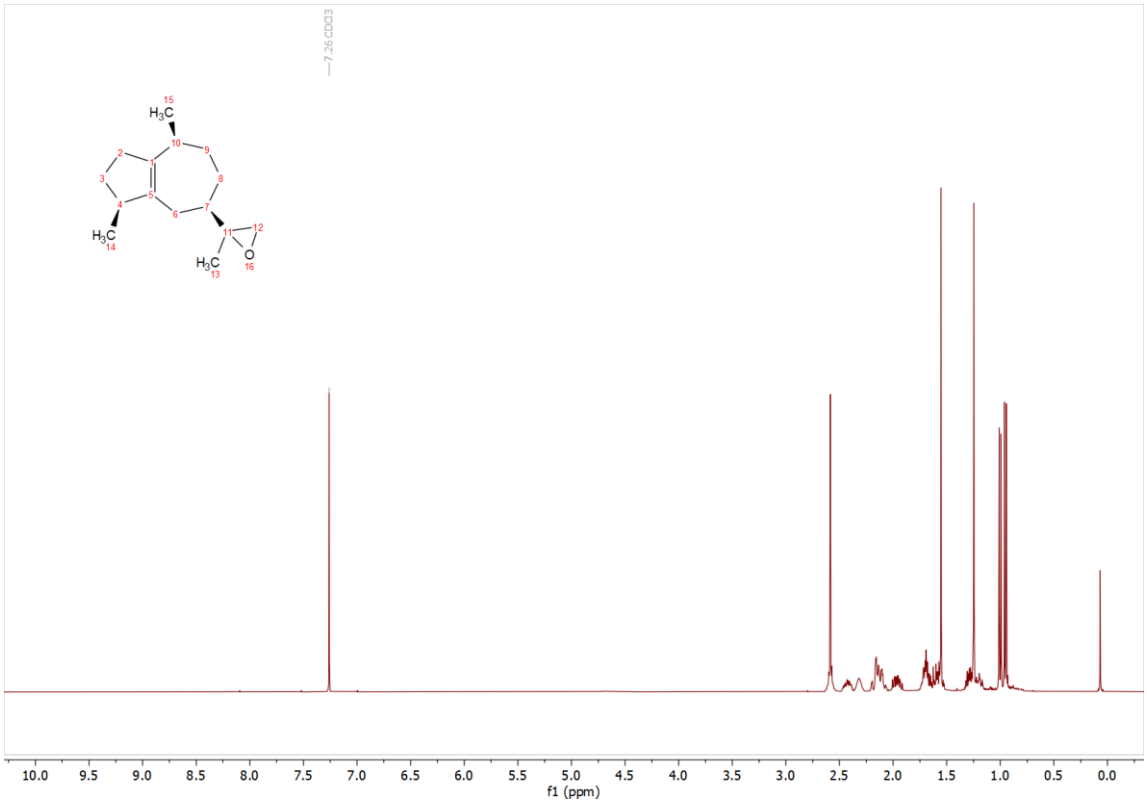


Figure 4.1 The ^{1}H NMR spectrum of **2a** in $CDCl_3$.

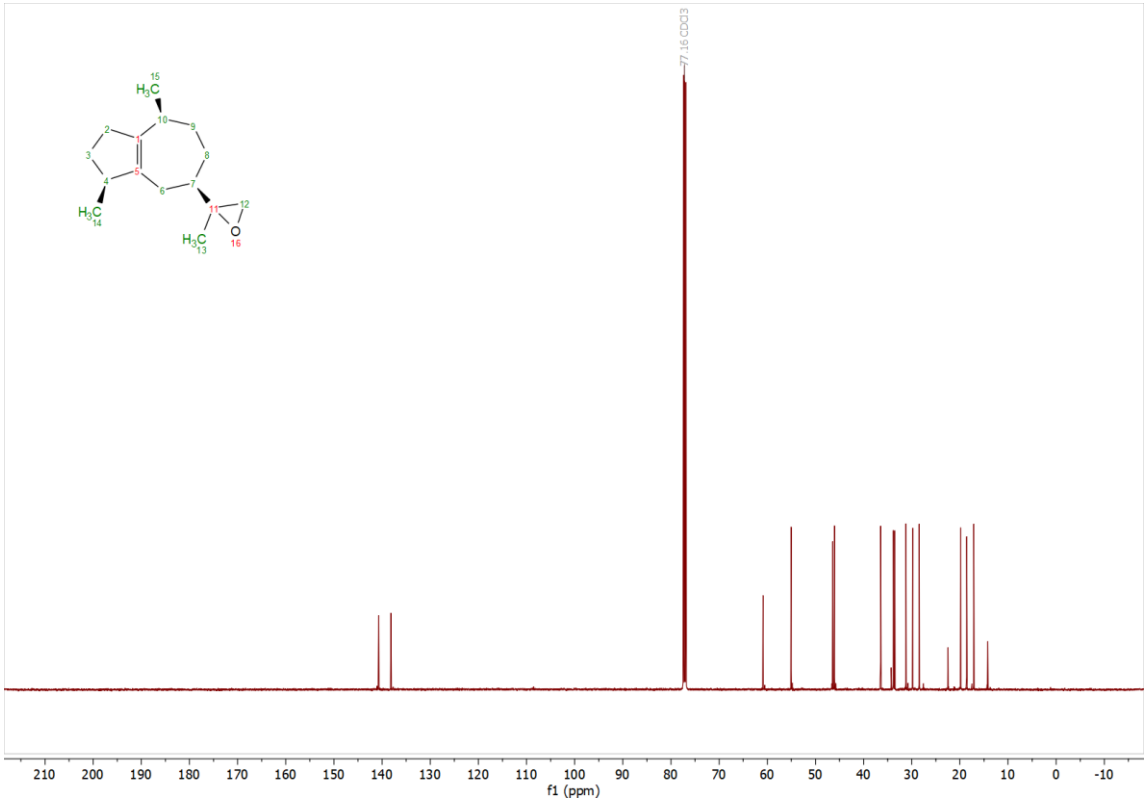


Figure 4.2 The ^{13}C NMR spectrum of **2a** in $CDCl_3$.

(*2S*)-Rotundol, **2b**

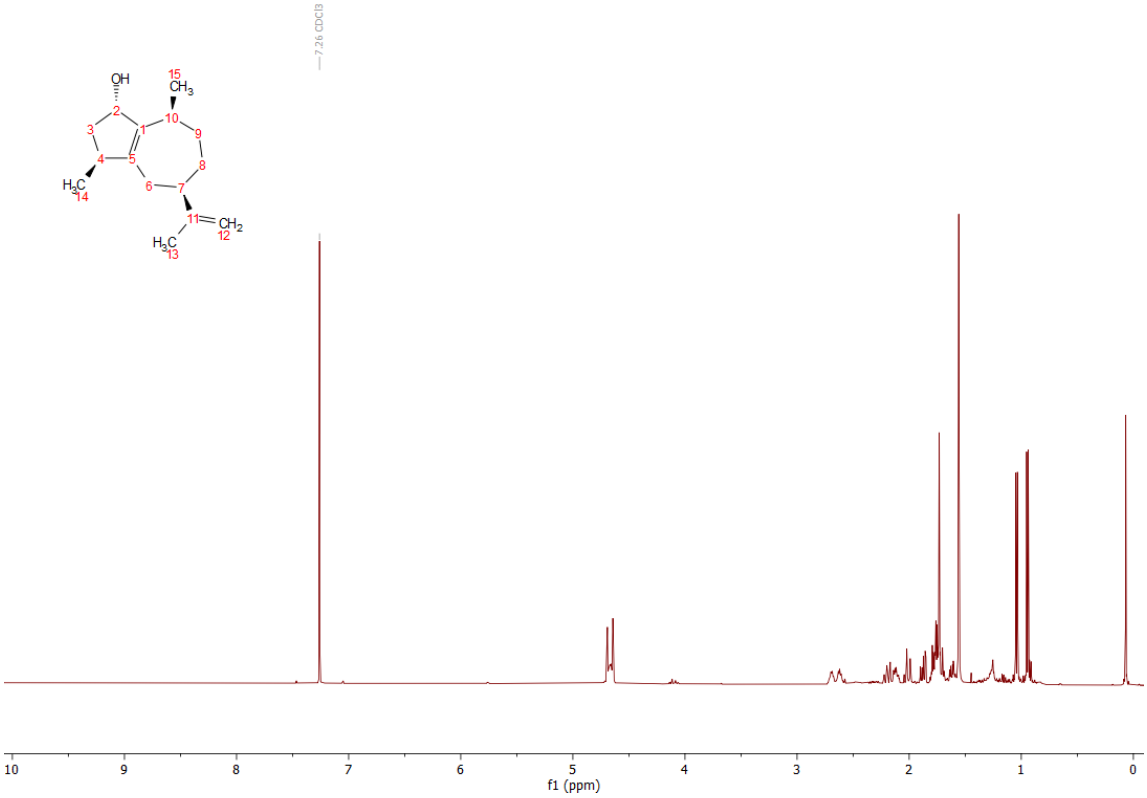


Figure 4.3 The ^{1}H NMR spectrum of **2b** in $CDCl_3$.

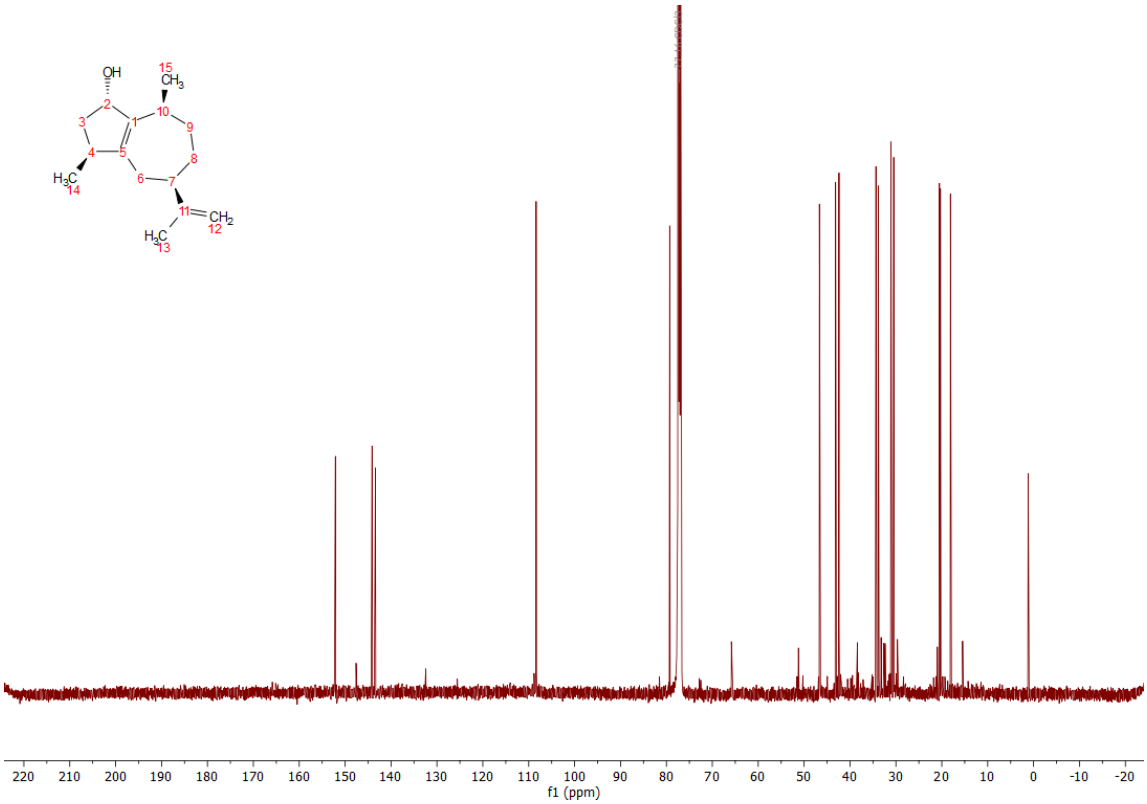


Figure 4.4 The ^{13}C NMR spectrum of **2b** in $CDCl_3$.

(*9S*)-Hydroxy-guaiene, **2c**

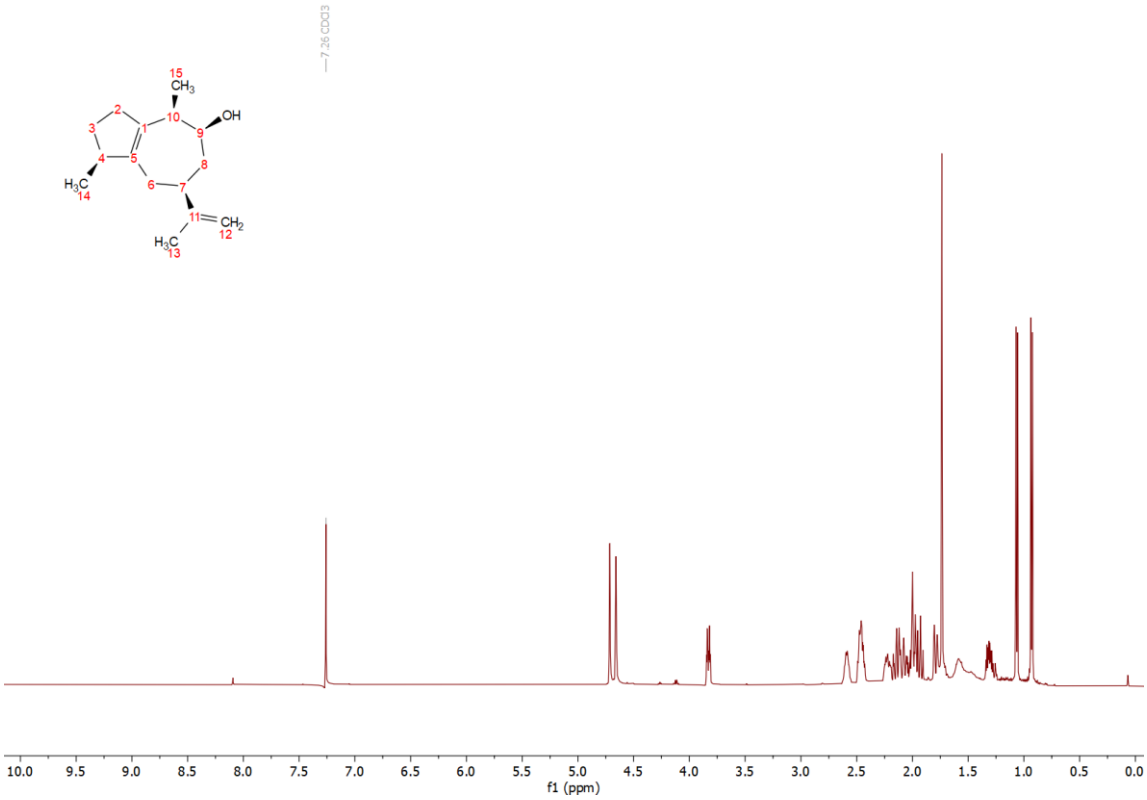


Figure 4.5 The ^{1}H NMR spectrum of **2c** in $CDCl_3$.

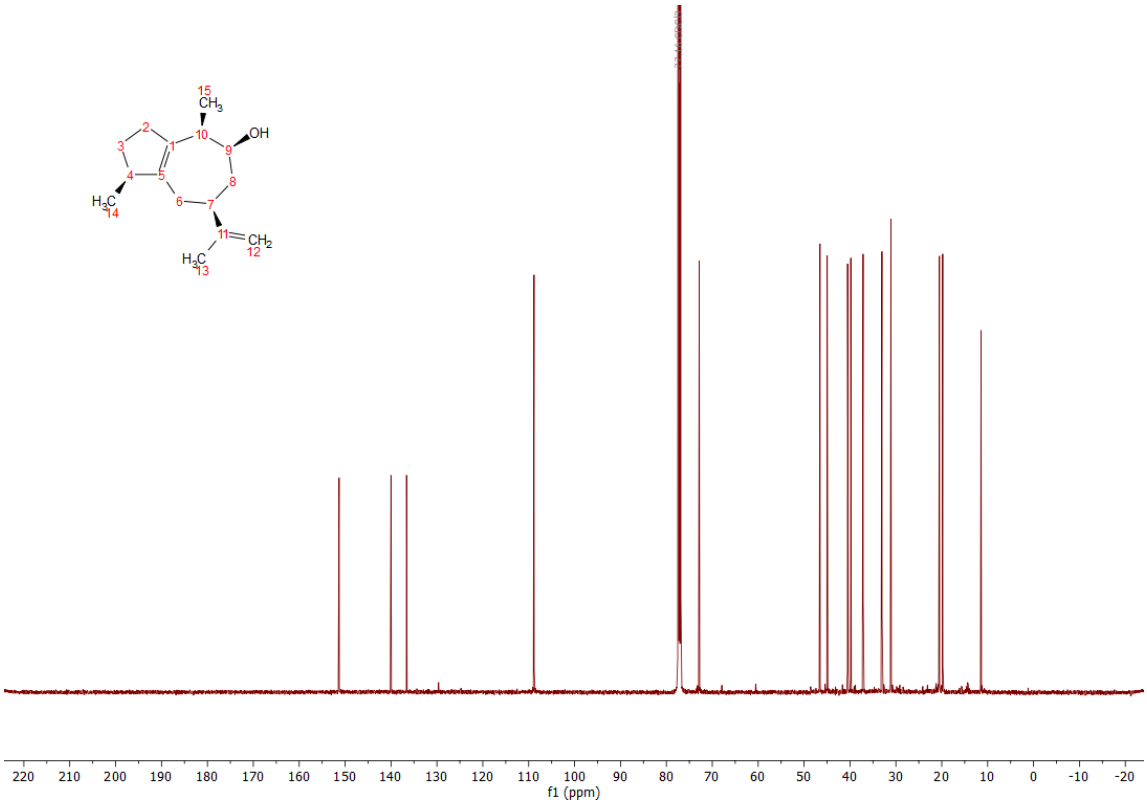


Figure 4.6 The ^{13}C NMR spectrum of **2c** in $CDCl_3$.

2-Hydroxy-11,12-epoxy-α-guaiene, **2d**

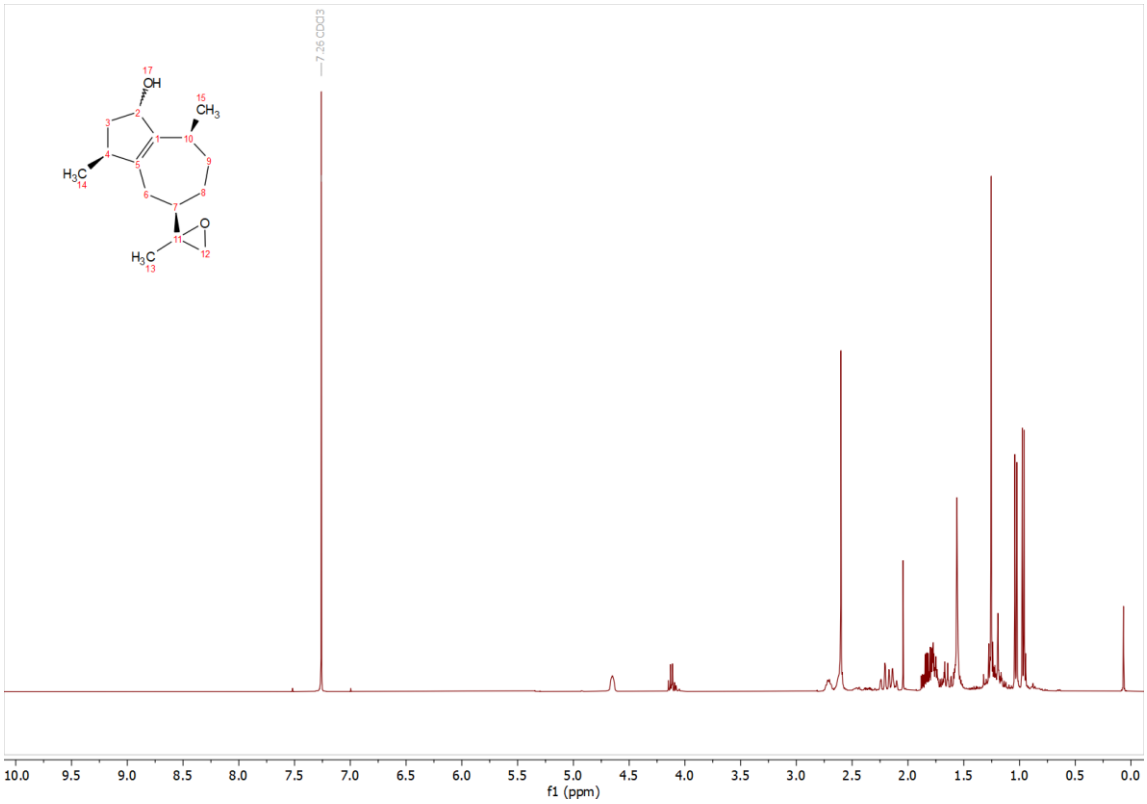


Figure 4.7 The ^{1}H NMR spectrum of **2d** in $CDCl_3$.

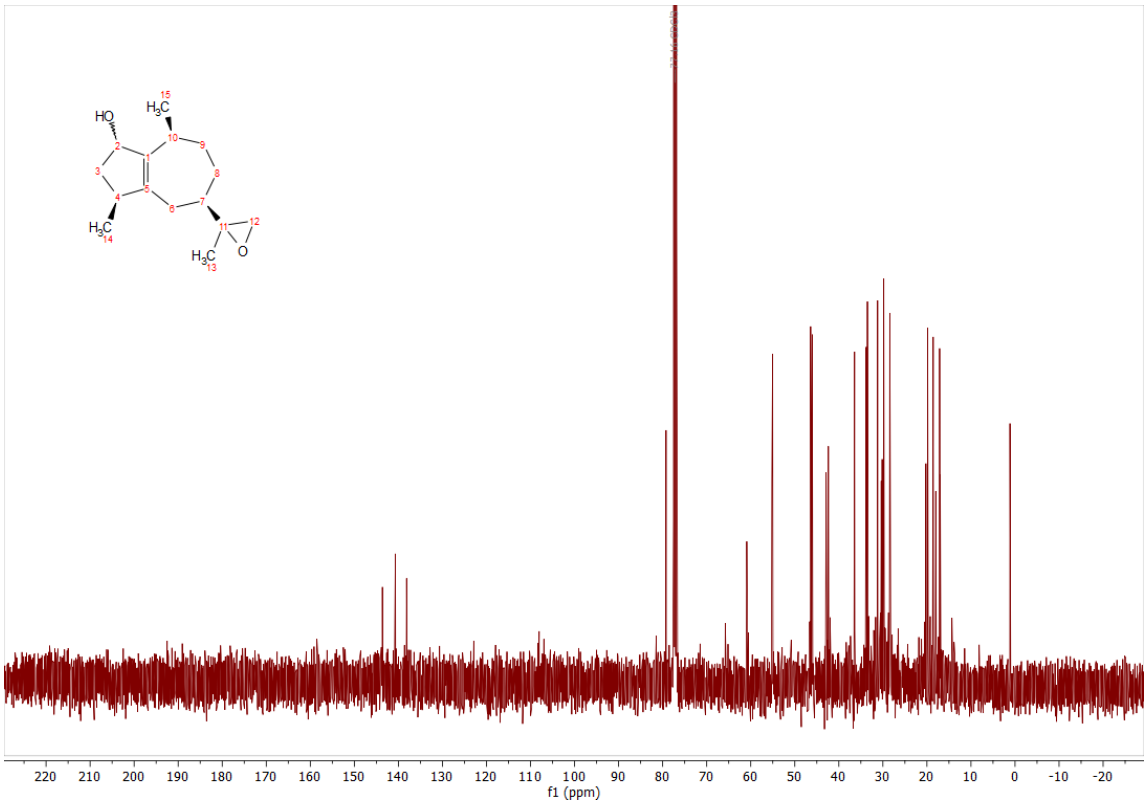


Figure 4.8 The ^{13}C NMR spectrum of **2d** in $CDCl_3$.

(*2R*,*9S*)-Dihydroxy-α-guaiene, **2e**

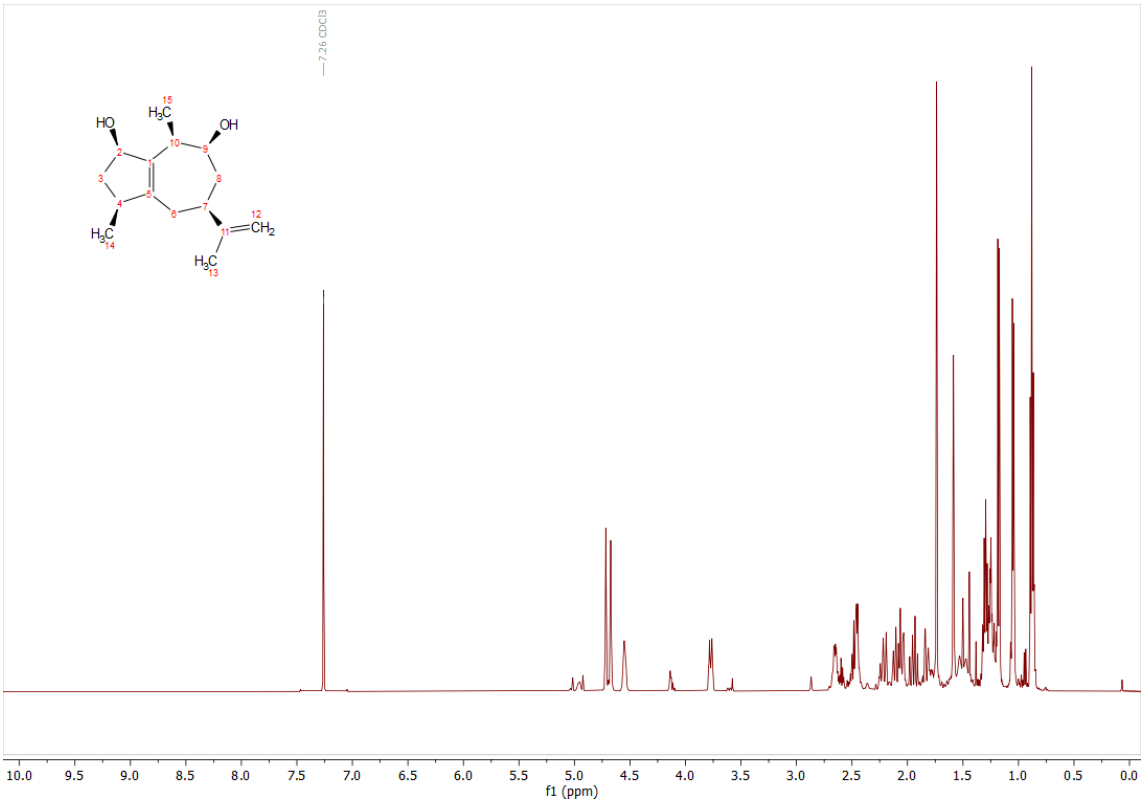


Figure 4.9 The ^{1}H NMR spectrum of **2e** in $CDCl_3$.

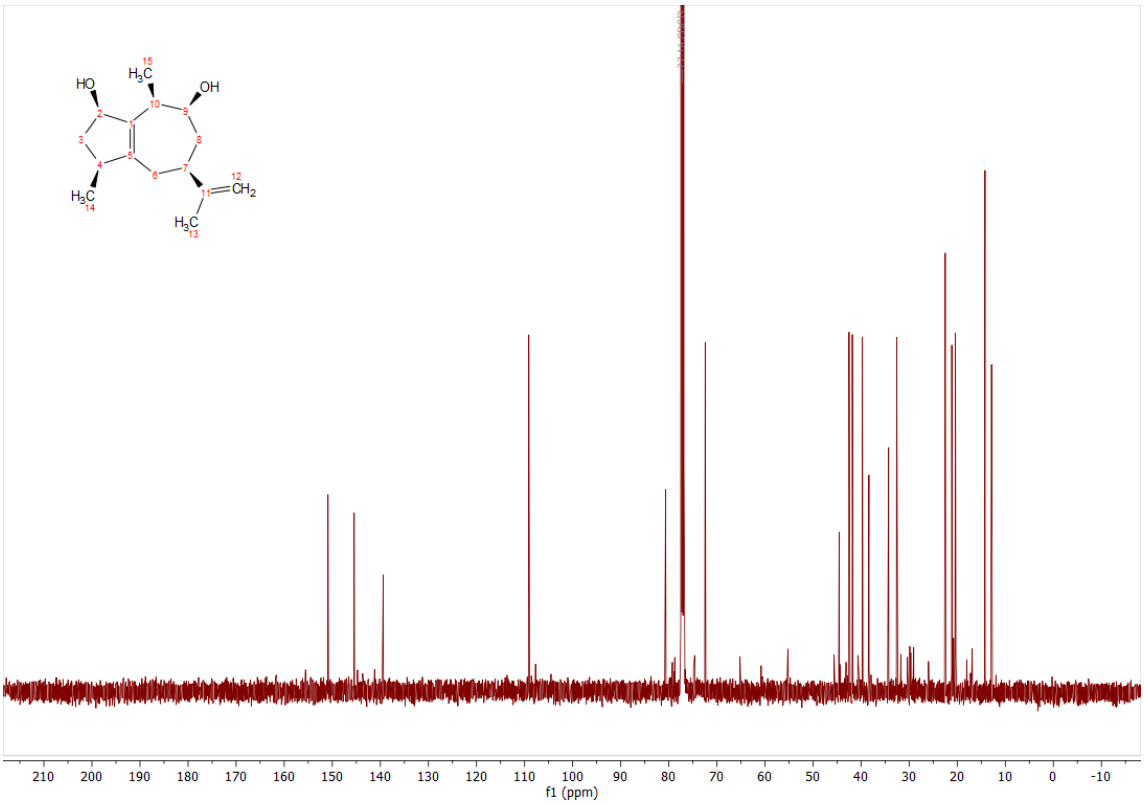


Figure 4.10 The ^{13}C NMR spectrum of **2e** in $CDCl_3$.

Rotundone, **2f**

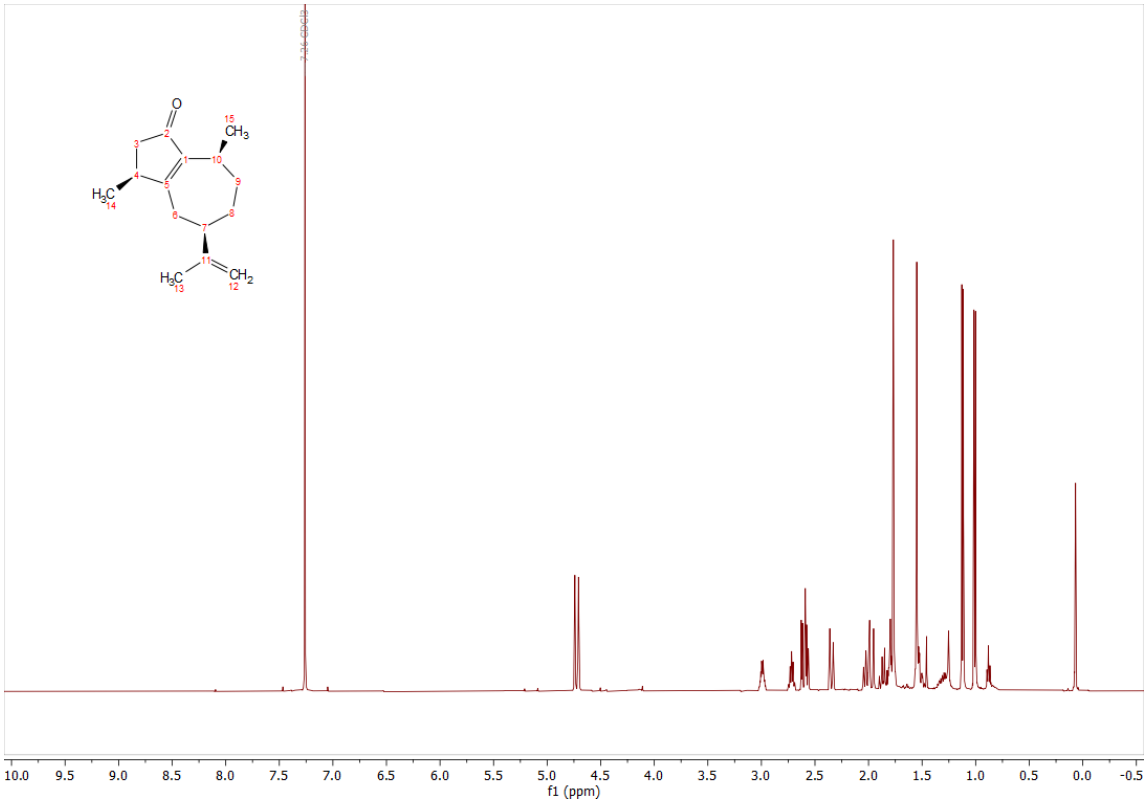


Figure 4.11 The ^{1}H NMR spectrum of **2f** in $CDCl_3$.

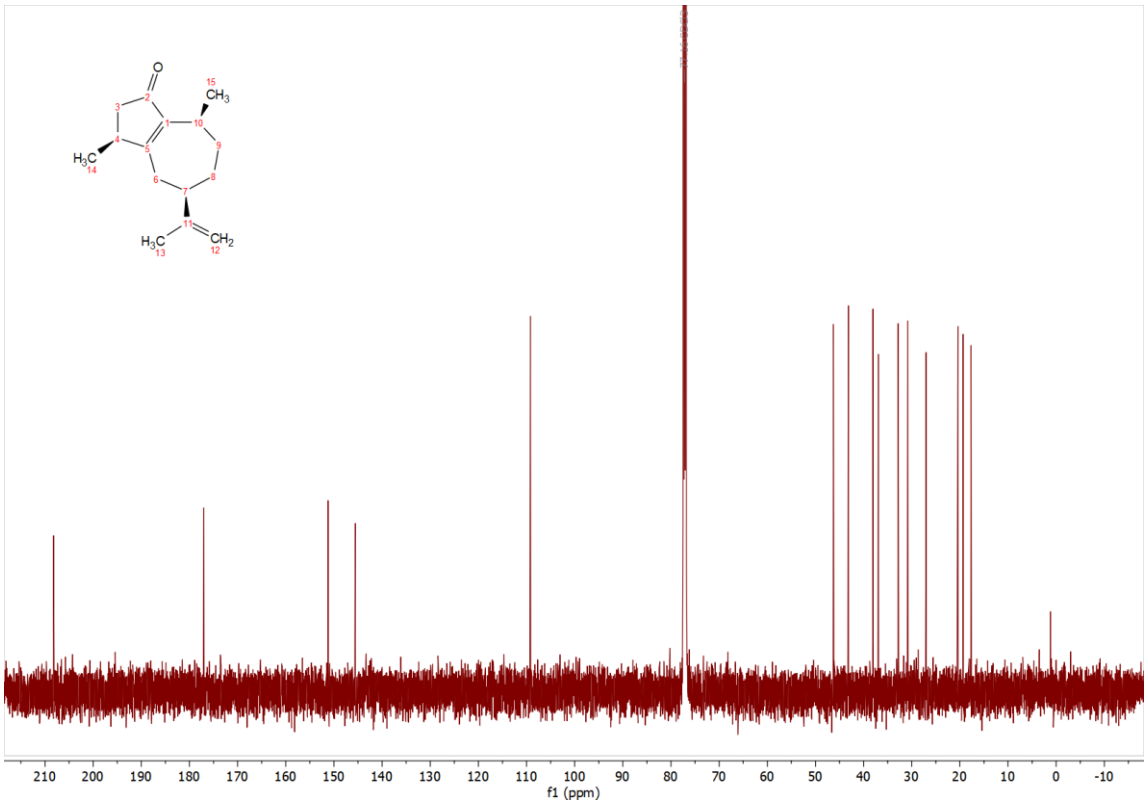


Figure 4.12 The ^{13}C NMR spectrum of **2f** in $CDCl_3$.

Appendix 5: Oligonucleotides used for site-directed mutagenesis

Table 5.1 Sequence of oligonucleotides used for site-directed mutagenesis.

Primers	F/R	Tm	Sequence
F87I/A82M	F	60	CGTGATTTT**ATG**GGAGACGGGTTA**ATT**ACAAGCTGGACGCATG
	R	60	CAGCTTGT**AAT**TAACCCGTCTCC**CAT**AAAATCACGTACAAATTTAAGCGC
A184I	F	60	CTGGATGAA**ATA**ATGAACAAGCTGCAGCGAGCA
	R	60	CTTGTTCAT**TAT**TTCATCCAGTGCACGGACCATACTT
V26L	F	62	CACAGATAAACCG**CTT**CAAGCTTTGATGAAAATTGCGGATG
	R	62	CATCAAAGCTTG**AAG**CGGTTTATCTGTGTTTAATAACGGTAA
V26M	F	61	CACAGATAAACCG**ATG**CAAGCTTTGATGAAAATTGCGGATG
	R	61	TTTCATCAAAGCTTG**CAT**CGGTTTATCTGTGTTTAATAACGGTAAATTT
S72V	F	56	AAAAACTTA**GTT**CAAGCGCTTAAATTTGTACGTGATTTT
	R	56	TAAGCGCTTG**AAC**TAAGTTTTTATCAAAGCGTGATTC
S72I	F	62	GATAAAAACTTA**ATT**CAAGCGCTTAAATTTGTACGTGATTTTGCAGG
	R	62	AAGCGCTTG**AAT**TAAGTTTTTATCAAAGCGTGATTCATC
S72L	F	60	GATAAAAACTTA**CTT**CAAGCGCTTAAATTTGTACGTGATTTTGCAGG
	R	61	CAAATTTAAGCGCTTG**AAG**TAAGTTTTTATCAAAGCGTGATTCATCGC
A74L	F	55	CTTAAGTCAA**CTT**CTTAAATTTGTACGTGATTTTGCAGGAGAC
	R	56	GTACAAATTTAAG**AAG**TTGACTTAAGTTTTTATCAAAGCGTGATTCAT
F87I/T88L	F	59	GAGACGGGTTA**ATTCTA**AGCTGGACGCATGAAAAAAATTGGAAA
	R	60	TGCGTCCAGCT**TAGAAT**TAACCCGTCTCCTGCAAAATC
F87I/T88F	F	59	GAGACGGGTTA**ATTTTT**AGCTGGACGCATGAAAAAAATTGGAAA
	R	60	TGCGTCCAGCTAAAAATTAACCCGTCTCCTGC**AAAAT**CA
F87I/T88M	F	59	GAGACGGGTTA**ATTATG**AGCTGGACGCATGAAAAAAATTGGAAA
	R	59	TGCGTCCAGCT**CATAAT**TAACCCGTCTCCTGCAAAATC
L181F	F	56	GTCCGTGCA**TTC**GATGAAGCAATGAACAAG
	R	58	TGCTTCATC**GAA**TGCACGGACCATACTTGT
I263W	F	60	ACATTCTTA**TGG**GCGGGACACGAAACAACAAGTGGTC
	R	60	GTGTCCCGC**CCA**TAAGAATGTAATAATTTGATAGCGAATGTTCTCG
A264F	F	55	TTCTTAATT**TTC**GGACACGAAACAACAAGTGGTCTTTTATC
	R	55	TTCGTGTCC**GAA**AATTAAGAATGTAATAATTTGATAGCGAATG
A330V	F	64	CCAACTGCTCCT**GTT**TTTTCCCTATATGCAAAAGAAGATACGGTGC
	R	64	CATATAGGGAAAA**AAC**AGGAGCAGTTGGCCATAAGCGCAGC
A330I	F	62	CCAACTGCTCCT**ATT**TTTTCCCTATATGCAAAAGAAGATACGGTGC
	R	62	CATATAGGGAAAA**AAT**AGGAGCAGTTGGCCATAAGCGCAGC
A330F	F	61	CCAACTGCTCCT**TTT**TTTTCCCTATATGCAAAAGAAGATACGGTGC
	R	62	CATATAGGGAAAA**AAA**AGGAGCAGTTGGCCATAAGCGCAGC
S332L	F	62	CCTGCGTTT**CTC**CTATATGCAAAAGAAGATACGGTGCTTG
	R	62	TGCATATAG**GAG**AAACGCAGGAGCAGTTGGCCATAAG
S332I	F	62	CCTGCGTTT**ATC**CTATATGCAAAAGAAGATACGGTGCTTG
	R	62	TGCATATAG**GAT**AAACGCAGGAGCAGTTGGCCATAAG
A330P	F	61	CAACTGCTCCT**CCG**TTTTCCCTATATGCAAAAGAAG
	R	63	GGGAAAA**CGG**AGGAGCAGTTGGCCATAAG

F87A	F	60	GACGGGTTA**GCT**ACAAGCTGGACGCATGAAAAAAATTGG
	R	60	CCAGCTTGT**AGC**TAACCCGTCTCCTGCAAAATCAC
A264G	F	61	CATTCTTAGGT**GGG**GGACACGAAACAACAAGTGG
	R	60	GTTTCGTGTCC**CCC**ACCTAAGAATGTAATAATTTGATAGCGAAT
A328G	F	66	GGCCAACT**GGT**CCTGCGTTTTCCCTATATGC
	R	65	AAACGCAGG**ACC**AGTTGGCCATAAGCGCAGC
L71F	F	57	CGCTTTGATAAAAAC**TTT**AGTCAAGCGCTT
	R	59	GCGCTTGACT**AAA**GTTTTTATCAAAGCGTG
L71M	F	56	CACGCTTTGATAAAAAC**ATG**AGTCAAGCGC
	R	54	GCGCTTGACT**CAT**GTTTTTATCAAAGCGTG
S72W	F	58	GATAAAAACTTA**TGG**CAAGCGCTTAAATTTGTACGTGAT
	R	58	AAGCGCTTG**CCA**TAAGTTTTTATCAAAGCGTGATTCATC
V178F	F	61	TATTACAAGTATG**TTT**CGTGCACTGGATGAAGCAATGAACAAG
	R	62	ATCCAGTGCACG**AAA**CATACTTGTAATAAATGGATGAGGC
A184I/ L188F	F	62	CTGGATGAA**ATA**ATGAACAAG**TTT**CAGCGAGCAAATCCAGACGACCCAG
	R	62	TGCTCGCTG**AAA**CTTGTTCATTATTTCATCCAGTGCACGGACCATAC
S332M	F	62	CCTGCGTTT**ATG**CTATATGCAAAAGAAGATACGGTGCTTGGAGG
	R	62	TGCATATAG**CAT**AAACGCAGGAGCAGTTGGCCATAAGCG
S332I	F	62	CCTGCGTTT**ATC**CTATATGCAAAAGAAGATACGGTGCTTG
	R	62	TGCATATAG**GAT**AAACGCAGGAGCAGTTGGCCATAAG
S332L	F	62	CCTGCGTTT**CTC**CTATATGCAAAAGAAGATACGGTGCTTG
	R	62	TGCATATAG**GAG**AAACGCAGGAGCAGTTGGCCATAAG
L437V	F	61	ATTAAAGAAACT**GTA**ACGTTAAAACCTGAAGGCTTTGTGG
	R	61	CAGGTTTTAACGT**TAC**AGTTTCTTTAATATCCAGCTCGTAGT
L437F	F	58	GAAACT**TTT**ACGTTAAAACCTGAAGGCTTTGTG
	R	58	CAGGTTTTAACGT**AAA**AGTTTCTTTAATATCCAGC
L437M	F	59	ATTAAAGAAACT**ATG**ACGTTAAAACCTGAAGGCTTTGTGGT
	R	59	CAGGTTTTAACGT**CAT**AGTTTCTTTAATATCCAGCTCGTAGTTTGTATG
T438F	F	55	GAAACTTTA**TTC**TTAAAACCTGAAGGCTTTGTGGTAAAAGC
	R	55	AGGTTTTAA**GAA**TAAAGTTTCTTTAATATCCAGCTCGTAGTTTG