Engineering terpenoid oxidation by cytochrome $P450_{BM3}$ for fine chemical synthesis

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

Wenyu Chen

Wolfson College

Trinity 2020

Engineering terpenoid oxidation by cytochrome $P450_{BM3}$ for fine chemical synthesis

Wenyu Chen D Phil. Thesis abstract

Wolfson College Trinity, 2020

$P450_{BM3}$ (CYP102A1) has a unique structure of having the electron-transfer reductase domain fused to the haem monooxygenase domain. The intramolecular electron transfer process makes $P450_{BM3}$ catalytically self-sufficient and promotes high turnover rates. P450 enzymes activate C–H bonds via the insertion of an oxygen atom from atmospheric dioxygen and play key roles in the biosynthesis of endogenous compounds (terpenes, alkaloids and steroids).

Oxygenated terpenoid compounds have desirable properties such as aromas, flavours and medicinal effects. The initial oxygenation step in terpenoid biosynthesis from the terpene precursors is invariably catalysed by a P450 enzyme. Chapter 3 describes the engineering and application of $P450_{BM3}$ variants to develop efficient, two-step chemo-enzymatic synthesis routes to high value the norisoprenoid and isoprenoid aroma compounds β-damascenone, tabanone isomers, and rose oxide by late-stage C–H activation.

The generation of diverse metabolites of common steroids via P450-catalysed direct C–H activation has important applications in the synthesis of steroidal agents for drug screening and development. Chapter 4 describes the engineering of $P450_{BM3}$ for the mono- and di-hydroxylation of androstenedione (AD), dehydroepiandrosterone (DHEA) and testosterone (TST). To design altered steroid binding orientations to broaden the regioselectivity, the structures of wild type $P450_{BM3}$ and the steroid aromatase CYP19A1 are overlaid to identify key regions for scanning glycine mutagenesis which greatly expanded the product range of AD, DHEA and TST. In Chapter 5, more steroidal compounds are screened to show hydroxylation

at multiple sites with high regio- and stereo-selectivity for androsterone (AN), epi-androsterone (EAN), progesterone (PROG) and pregnenolone (PREG).

Acknowledgement

I would like to thank Luet Wong for his guidance and support throughout my whole DPhil study. I am really grateful for having Luet as my supervisor. Everytime when I have questions or failed at experiments, he would be there to help and guide me through. Not only he taught us how to do experiments, the more important thing is he would guide us to think in a wider and bigger picture to have a more thorough insight about the project. I could not thank him enough for his kindness and patience, and sometimes the "where is my short gun" moment.

I would like to thank Oxford Biotrans for studentship support and the generous loan of ketoreductase enzymes for screening.

I am very grateful for all the support I have received from two of my colleages, Yushu Li and Matthew Fisher. I would like to thank my buddy Yushu for her accompany throughout my DPhil study with constantly mumbling, and all the precious memories we have created. And Matthew, for his kindness and support whenever I needed, I will miss you and the eager-to-win-Matt when you play princess card. I really miss the days working with you two.

Also, I would like to thank my buddy Xinxin for her humour and all the support, especially when I am sick or upset, she is always there for me. Alice, for her thoughtfulness for everyone; it is nice to have you around. Alethea for her sport spirit to keep me rolling. Lucy for her great suggestions and every "hung-over" Saturday talk. Vicky for her cute smile. I am also grateful for all the part II's I have worked with (Lydia, Noah, Rory, Sophie and Aaron) for their contribution on the work and the happiness they have brought to the group.

I would like to thank my family, my parents, my sisters and brother for their continous support and love. You will always be there for me.

Finally, I would like to thank Yuan for putting up my bad tempers and giving me all the support and trust and keeping me calm whenever I am upset or panic about works. I could not have done this without you.

Contents

List of Abbreviations

AD	androstenedione
ADH	alcohol dehydrogenase
AdR	adrenoxin reductase
Adx	adrenodoxin
AN	androsterone
BMP	$P450_{BM3}$ haem domain
BMR	$P450_{BM3}$ reductase domian
CAST	combinatorial active-site saturation test
COSY	correlation spectroscopy
CPR	cytochrome P450 reductase
CYP	cytochrome P450
DCIB	1,2-dichloroisobutane
DHEA	dehydroepiandrosterone
DMAPP	dimethyallyl pyrophosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
EAN	epi-androsterone
ET	electron-transfer
ep-PCR	error-prone PCR
F&F	flavour and fragrances
FAD	flavin adenine dinucleotide
FB	flavobacterium
Fdx	ferredoxin
FdR	ferredoxin reductase

FID	flame ionisation detector
FMN	flavin mononucleotide
FPP	farnesyl pyrophosphate
GC	gas chromatography
GDH	glucose dehydrogenase
GGPP	geranylgeranyl pyrophosphate
GPP	geranyl pyrophosphate
HVA	homovanillic acid
HMBC	heteronuclear multiple-bond correlation spectroscopy
HSQC	heteronuclear single quantum correlation spectroscopy
IPP	isopentenyl pyrophosphate
IPTG	isopropyl-β-D-thiogalactopyanoside
ISM	iterative saturation mutagenesis
kan	kanamycin
KRED	ketoreductase
LB	lysogeny broth
MDMA	3,4-methylenedioxymethylamphetamine
MVA	mevalonate
$NAD(P)^+$	nicotinamide adenine dinucleotide (phosphate)
NAD(P)H	nicotinamide adenine dinucleotide (phosphate), reduced form
NMR	nuclear magnetic resonance
NPG	*N*-palmitoylglycine
NOESY	nuclear overhauser effect spectroscopy
nOe	nuclear Overhauser effect
PAF	platelet-activating factor
PAH	polycyclic aromatic hydrocarbon

PCR	polymerase chain reaction
PEO	patchouli essential oil
PFOR	phthalate dioxygenase reductase
PREG	pregnenolone
PROG	progesterone
RMSD	root mean square deviation
RPM	rotations per minute
SB	substrate bound
SF	substrate free
SFC	supercritical fluid chromatography
SOC	super optimal broth with catabolite repression
SRS	substrate recognition site
TAE	tris-acetate-EDTA
TLC	thin layer chromatography
TON	turnover number
TTN	total turnover number
TST	testosterone
v/v	volume by volume
WT	wild type (*i.e.* mutation free)

Chapter 1

1 Introduction

1.1 C–H activation and functionalisation

As the most fundamental building block of organic compounds, C–H bonds are chemically inert due to the high bond dissociation energy. The activation and functionalisation of C–H bonds has been recognised as one of the biggest challenges in organic synthesis that a number of catalysts or synthetic methods have been explored for C–H bond cleavage, such as the use of transitional metal complexes, radical C–H activation, and enzymatic C–H oxidation, etc.[1-6] Iron-catalysed C–H arylation was discovered during the investigation of a cross-coupling reaction. The side product, 2-biphenylpyridine, was formed in 8% yield during the synthesis of 2-phenylpyridine by cross-coupling between 2-bromopyridine and a phenylzinc reagent (Figure 1.1–A). It was found that air leaked into the reaction vessel was the oxidant that triggered the reaction,[7] and 1,2-dichloroisobutane (DCIB) was later discovered to be an optimum oxidant to carry out this reaction.[8] Figure 1.1–B shows an example of iron-catalysed arylation of C–H bonds through directed C–H activation of benzoquinoline with DCIB as oxidant.[8]

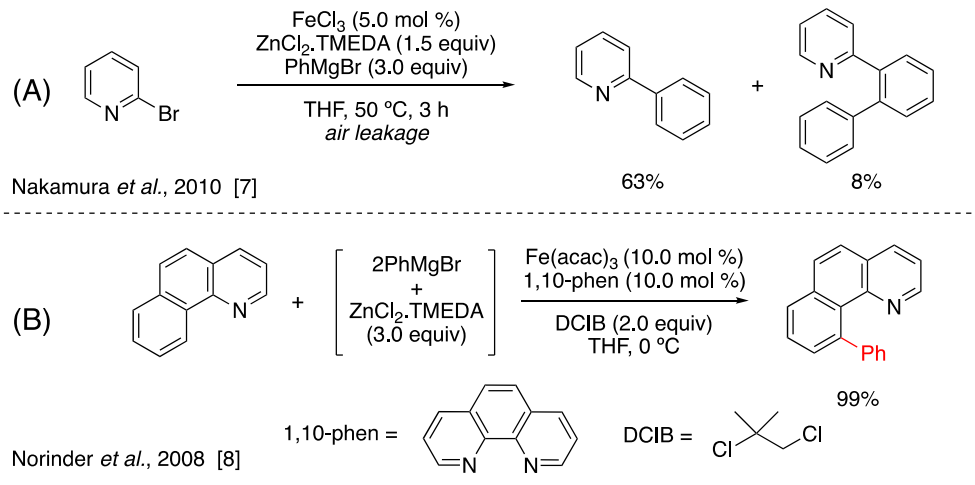


Figure 1.1 Iron-catalysed C–C bond formation through C–H bond activation.[7,8]

Non-haem iron complexes along with H_2O_2 have been investigated and improved for C–H hydroxylation.[9,10] White and co-workers reported an iron-based small molecule catalyst that selectively oxidised a broad range of substrates on the basis of the electronic and steric properties of the C–H bonds (Figure 1.2).[10] It appeared that for insertion pathways the reactivity trend of oxidation was typically: tertiary > secondary > primary.[11] Since most C–H activation reagents are electrophilic, the more electron rich a C–H bond is, the more easily it is oxidised. Apart from the electronic property of C–H bonds, steric hindrance is another factor, which explains the selectivity of oxidation sites in Figure 1.2 by an Fe-(*S*,*S*)-PDP catalyst.[10]

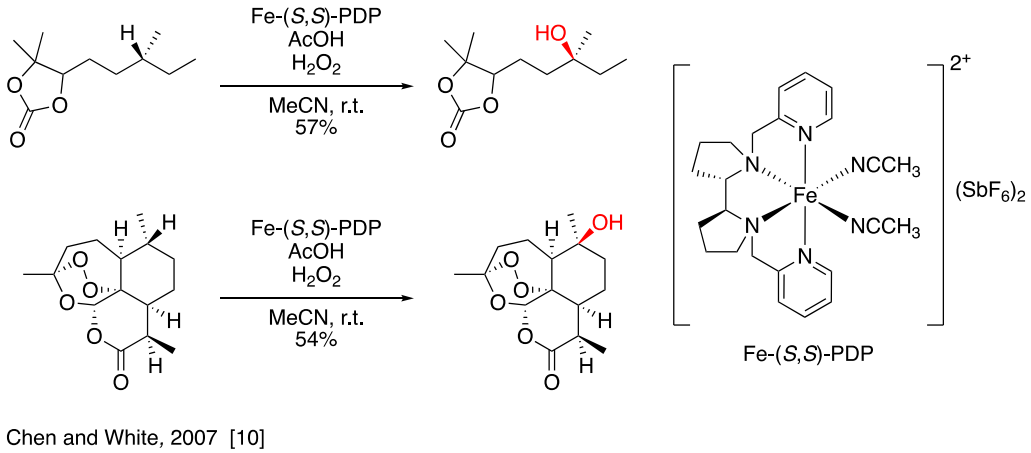


Figure 1.2 Fe-(*S*,*S*)-PDP catalysed C–H oxidation.[10]

In the past decade, palladium-catalysed C–H activation/C–C bond formation has been widely applied in many industrial processes.[6,12-16] Pd(0)/Pd(II) complexes catalyse the formation of C–C and C–O bonds with the aid of an oxidant. Tsukano *et al.* employed Pd(0)-catalysed C(sp^3)–H activation of carbamoyl chlorides for chemoselective lactam formation to synthesise various oxindoles via intramolecular cyclisation (Figure 1.3).[17] The Cramer group also performed the Pd(0)-catalysed C(sp^3)–H activation of trifluoacetimidoyl chlorides to form a range of 2-(trifluoromethyl)indoles (Figure 1.3).[18] The palladium (0) species is reformed by

reductive elimination of the palladium (II) compound by carbonates in the catalytic cycle. These reactions are relatively environmentally friendly and atom efficient.

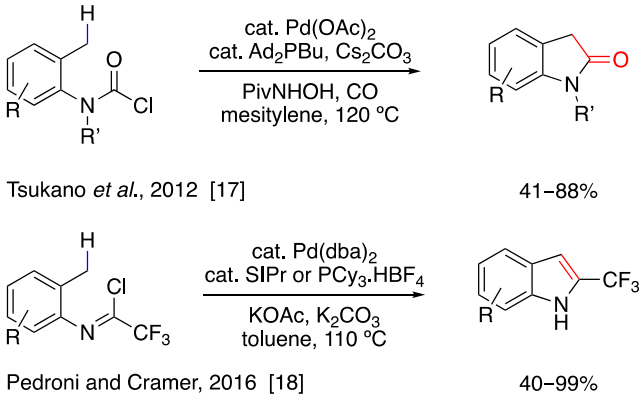


Figure 1.3 Pd(0)/Pd(II)-catalysed intramolecular cyclisation of benzylic C(sp^3)–H bonds.[17,18]

Radical C–H activation/cross-coupling chemistry has also stimulated great interest in industrial synthesis.[19-21] These reactions provide unconventional, efficient methods to connect molecular fragments with the help of photocatalysis and first-row transition metal catalysis together with peroxide added as radical initiators.[4,19] In 2011, McNally and co-workers performed a radical-radical cross-coupling reaction for the synthesis of benzylic amines via the photoredox-catalysed C–H arylation of amines. In this reaction, $Ir(ppy)_3$ was used as the photocatalyst under a light source (Figure 1.4).[22] Later, Hoshikawa and Inoue employed a photochemical 4-pyridination of C(sp^3)–H bonds without the use of transition metal complexes (Figure 1.4).[23] Although the mechanism remained to be clarified, a radical-based pathway was proposed. In this reaction, the key step was the H-atom abstraction of benzophenone by the photochemically generated oxy radical, producing the alkyl radical and the α-hydroxy radical. This electron-rich α-hydroxy radical rejointed with the electron-deficient 4-cyanopyridine substrate and produced the stable arene radical anion which led to the formation of the final product.[23] Other than C–H

activation, Lei and co-workers also demonstrated the radical-radical oxidative coupling for the C(sp^3)–N bond formation with the use of Ni catalyst (Figure 1.4).[19]

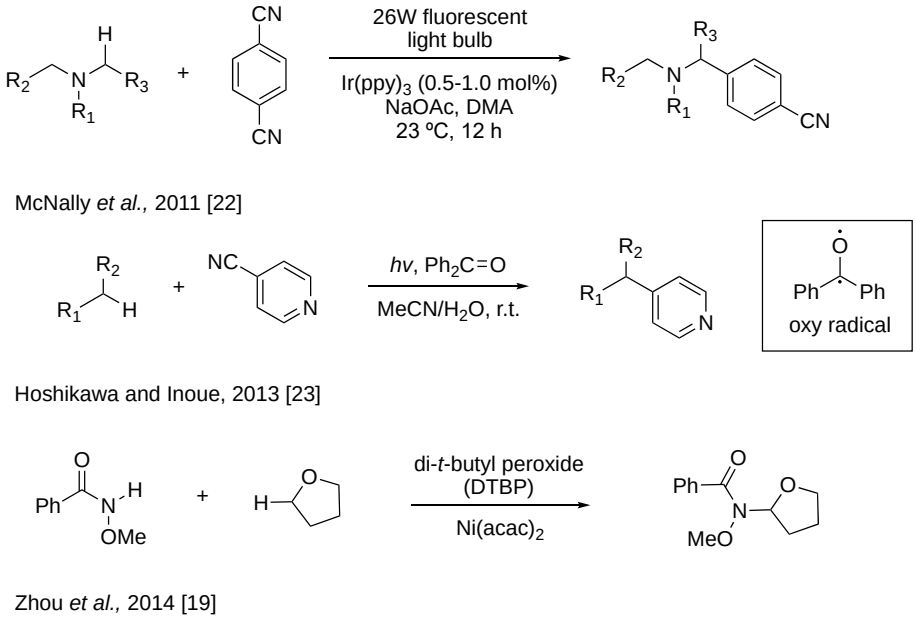


Figure 1.4 Examples of radical C–H and C–N activation/radical cross-coupling reactions using photochemistry and transition metal catalysis.[19,22,23]

Although the radical and transition-metal catalysed C–H bond activation play important roles in synthesis research, these methods suffer from several critical drawbacks, such as low regio- and enantio-selectivity over products, low total turnover number of reactions, narrow substrate scope and functional group compatibility.[24,25] A general pathway for C–H functionalisation is selective oxidation to an alcohol or carbonyl compound which can be converted readily to other functional groups. Therefore, C–H oxy-functionalisation under milder conditions with high regio- and enantio-selectivity along with high turnover number is highly desirable. It is well established that enzymes are capable of C–H functionalisation with high efficiency under ambient conditions.

1.2 Enzymatic reactions

Enzymes are classified in six groups: hydrolases (hydrolysis), oxidoreductases (transfer of electrons), lyases (breaking chemical bonds while forming new double bond), transferases (transfer of functional groups), ligases (bond formation accompanied with hydrolysis of bonds) and isomerases (isomerisation reaction). Of these enzymes, hydrolases such as lipases and amidases are most commonly used in organic synthesis since they often exhibit high levels of enantioselectivity.[26-28] For instance, dynamic kinetic resolution (DKR) of racemic alcohols with *Pseudomonas cepacia* lipase and a Ru catalyst has demonstrated the formation of the corresponding ester with 96% *ee* (Figure 1.5).[29]

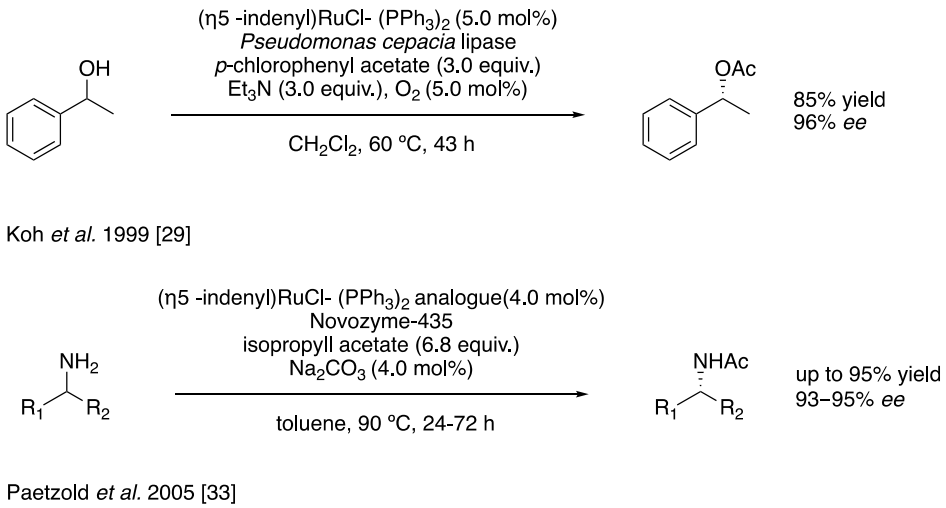


Figure 1.5 Examples of DKR of racemic alcohol and amine catalysed by lipases and transition-metal complexes.[29,33]

As with alcohols, enantiomerically pure amines could also be obtained by the use of lipases.[30,31] The first DKR of an amine was reported by Reetz and Schimossek, employing a lipase and a palladium catalyst to convert the racemic phenylethylamine to enantiomerically pure acetate with 99% *ee*, but with long reaction time and moderate conversion.[32] To improve this reaction, Backvall and co-workers found Ru complexes that can be used as an efficient racemisation

catalyst to form primary amines (Figure 1.5).[33,34] A transferase is a class of enzymes that transfer the specific functional groups, such as methyl or glycosyl groups, from one molecule (the donor) to another (the acceptor). One example of using transferase is in the discovery of the mechanism of the biodegradation of dopamine via the catechol-*O*-methyltransferase to form homovanillic acid (HVA) (Figure 1.6).[35,36]

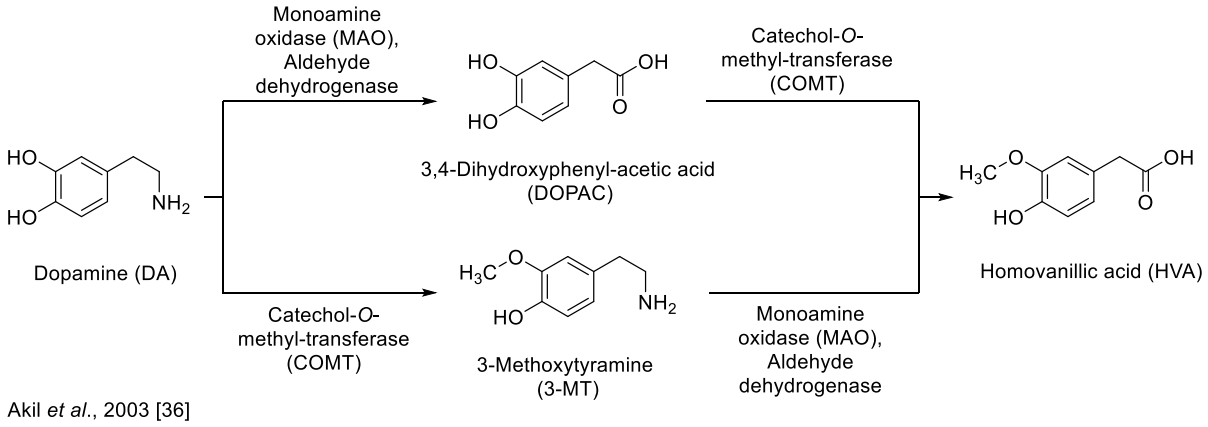


Figure 1.6 Biodegradation of dopamine via catechol-*O*-methyltransferase (along with other enzymes).[36]

Isomerases play important roles in human disease because deficiencies of this enzyme can cause disorders in human.[37,38] 3-Oxosteroid-Δ^5-Δ^4-isomerase catalyses the interconversion of a 3-oxo-Δ^5-steroid to a 3-oxo-Δ^4-steroid (Figure 1.7) in *C. testosteroni* and *P. putida*.[39]

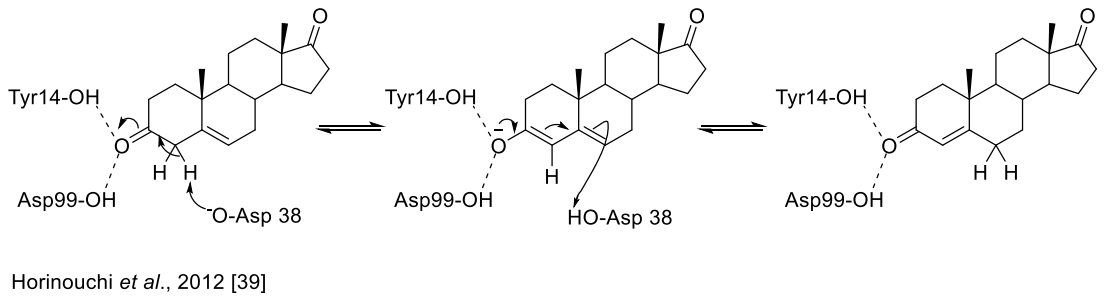


Figure 1.7 A schematic description of the isomerisation catalysed by *C.testosteroni* steroid delta-isomerase.[39]

1.3 Cytochrome P450 enzymes: direct C–H activation

Cytochrome P450 enzymes (CYPs) are a class of haem monooxygenase that catalyse the cleavage of molecular oxygen and introduce one of the oxygen atoms into C–H bonds to form alcohols; the other oxygen atom is reduced to water.[40-42]

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

Cytochrome P450s were discovered from their characteristic spectral properties, displaying a unique absorption peak at 450 nm with the reduced CO-bound form; cytochrome stands for a hemoprotein and P for pigment.[43] Later, this pigment was identified as a hemoprotein, specifically a haem *b*-dependant monooxygenase. The property of the reduced form of this enzyme showing a absorption peak at 450 nm with CO binding is still used for the calculation of P450 content.[44] Individual cytochrome P450 follow the nomenclature based on sequence homology: CYP, followed by a number indicating "family", then a letter indicates "subfamily", and another number indicates "specific protein"; for instance, CYP102A1 is the first protein identified in family 102, subfamily A.[45] Proteins sharing >40% identity in sequence are classified as a family, e.g. CYP3, CYP260, while subfamily members share >55% identity in sequence, e.g. CYP3A, CYP11B.

More than 300,000 CYP genes have been identified across the kingdoms of life. CYP enzymes catalyse essential reactions such as carbon source assimilation, synthesis of hormones and secondary metabolites, and degradation of xenobiotics. Due to their ability of selective hydroxylation of a broad range of substrates, there is great potential to use these enzymes in synthesis. The expression and purification of these enzymes at large scale are difficult, especially the eukaryotic P450s; thus, the heterologous production of these enzymes is

intensively explored.[46,47] The bacteria *Escherichia coli*, devoid of P450 enzymes, can be grown and cultured easily and inexpensively in a laboratory setting, which could be used as microbial cell factories for the production of various P450s.[46,48,49] P450 catalyses a wide variety of oxidation reactions including olefin epoxidation,[50-53] aromatic hydroxylation,[54-56] *N*-oxidation,[57] deamination[58,59] and dehalogenation,[60] as well as *N*-[57,61], *O*- and *S*-dealkylation.[57] Other activity types include C–C and C–O phenol coupling,[62] C–C bonds cleavage,[63] Baeyer-Villiger oxidation[64,65] and rearrangement reactions (Figure 1.8).[41,64,66,67]

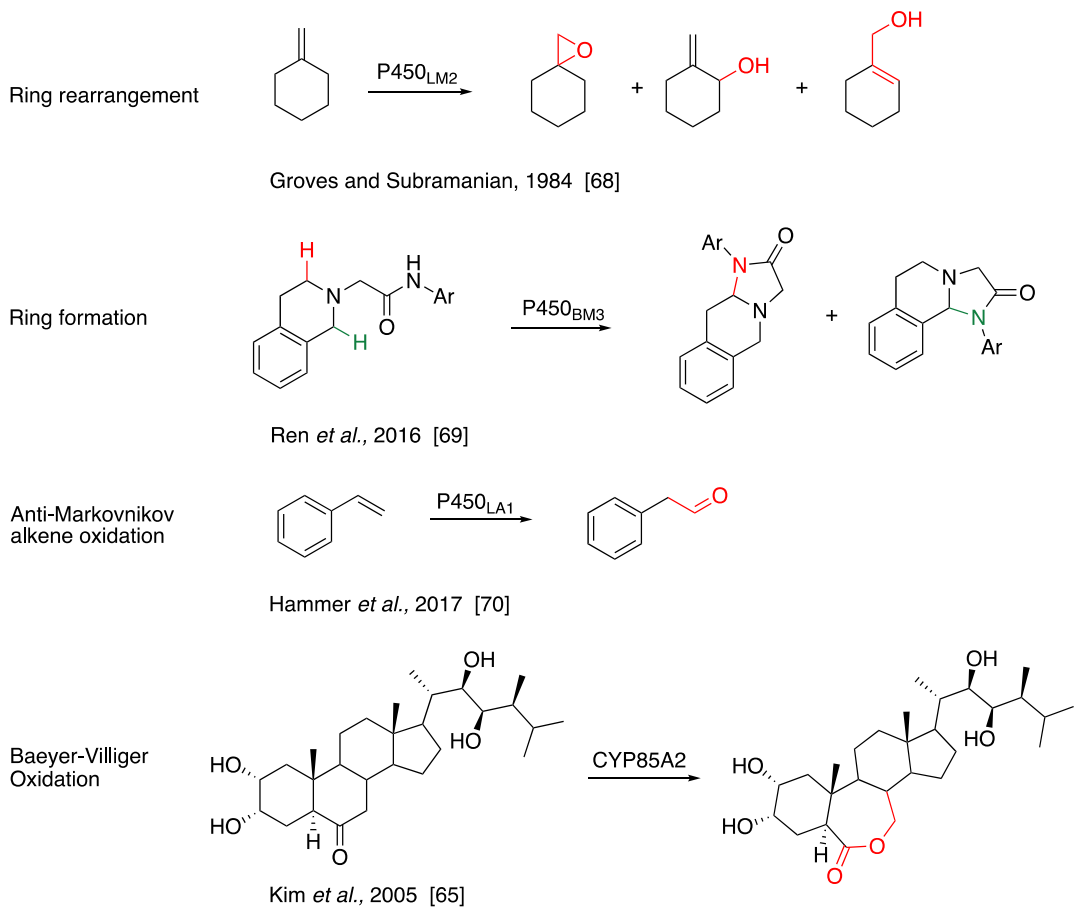


Figure 1.8 Examples of unusual oxidations catalysed by various P450s.[65,68-70]

1.3.1 Electron transfer in P450 oxidation

P450s catalyse region- and stereo-specific oxidations of non-activated C–H bonds under mild conditions. Most P450s are external monooxygenases that utilise electrons derived from the pyridine cofactors NADH or NADPH to activate dioxygen. Therefore, P450s must be associated with redox partner proteins or domains that transfer electrons from NAD (P) H to the P450 haem domain. Based on types of redox proteins, P450s can be classified as two major classes (class I and class II) (Figure 1.9). Class I enzymes comprise a redox 2Fe-2S iron-sulphur ferredoxin (Fdx) and a ferredoxin reductase (FdR). FdR contains flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) dependant enzymes that accept the hydride from NAD(P)H and transfer the two electrons to the ferredoxin domain one-at-a-time. The prototype bacterial class I P450s system is P450cam from *Pseudomonas putida*, which catalyses stereoselective 5-*exo* hydroxylation of *d*-camphor.[71-73] In human mitochondrial P450s, adrenoxin reductase (AdR)/adrenodoxin (Adx) combination is the most well-known FdR/Fdx electron transfer chain that catalyse several steps in the biosynthesis of steroid hormones.[74] The classic example is CYP11A1, also known as P450scc, which catalyses the cleavage of the cholesterol side-chain leading to the formation of pregnenolone.[74,75] Another subfamily member, CYP11B1, is responsible for the biosynthesis of corticosterone and cortisol via the 11β-hydroxylation of their corresponding precursors.[76,77]

Class II cytochromes P450 adopt a different system that electrons are delivered via the cytochrome P450 reductase (CPR) which contains separate FAD and FMN domains (Figure 1.9).[78-81] Class II P450s are the most common enzymes in eukaryotes and are responsible for the oxidative metabolism of both endogenous and exogenous compounds such as fatty acids, steroids and therapeutic drugs.[43,82] As the number of characterised P450s increase, other novel

electron transfer chains have been identified that belong neither to class I nor class II systems (Figure 1.9).[72,83-86]

Class III shares the characteristic three-component P450 system as class I, and the electrons are delivered via the redox partners FAD and FMN (Flavodoxin) which belong to separate proteins. CYP176A1 (P450cin) is the prototype of class III bacterial systems.[87,88]

Class IV is the first discovered thermophilic cytochrome P450 that contains a ferrodoxin and its corresponding partner, 2-oxo-acid:ferredoxin oxidoreductase to constitute redox chain.[89,90] The soluble CYP119 has been identified in this category and intensely studied due to its extreme stability to temperature (T_M = 91 °C) and pressure (up to 200 MPa).[91,92]

Class V consists of two separate protein components: a NAD(P)H-dependant reductase and a cytochrome P450-ferredoxin-fusion protein. The sterol 14α-demethylase CYP51 is an example of this system, with a [3Fe–4S] type ferredoxin domain fused to the P450 haem domain.[93]

Similar to class V systems, the class VI P450s also contain two separate proteins: a NAD(P)H-dependant reductase and a different P450-flavodoxin-fusion protein. FAD, FMN and haem are used as redox centres in this structure, which somewhat resembles the $P450_{BM3}$ and P450cin systems.[94 95]

Class VII P450 systems are fused enzymes. CYP116B2 ($P450_{RhF}$) is the first reported class VII P450.[96] A P450 domain is C-terminally fused to a reductase domain and a phthalate dioxygenase reductase domain (PFOR), which shows three distinct functional parts: a FMN-binding domain (accepts electrons from NAD(P)H), a NADH-binding domain and a [2Fe–2S] ferredoxin domain.[97]

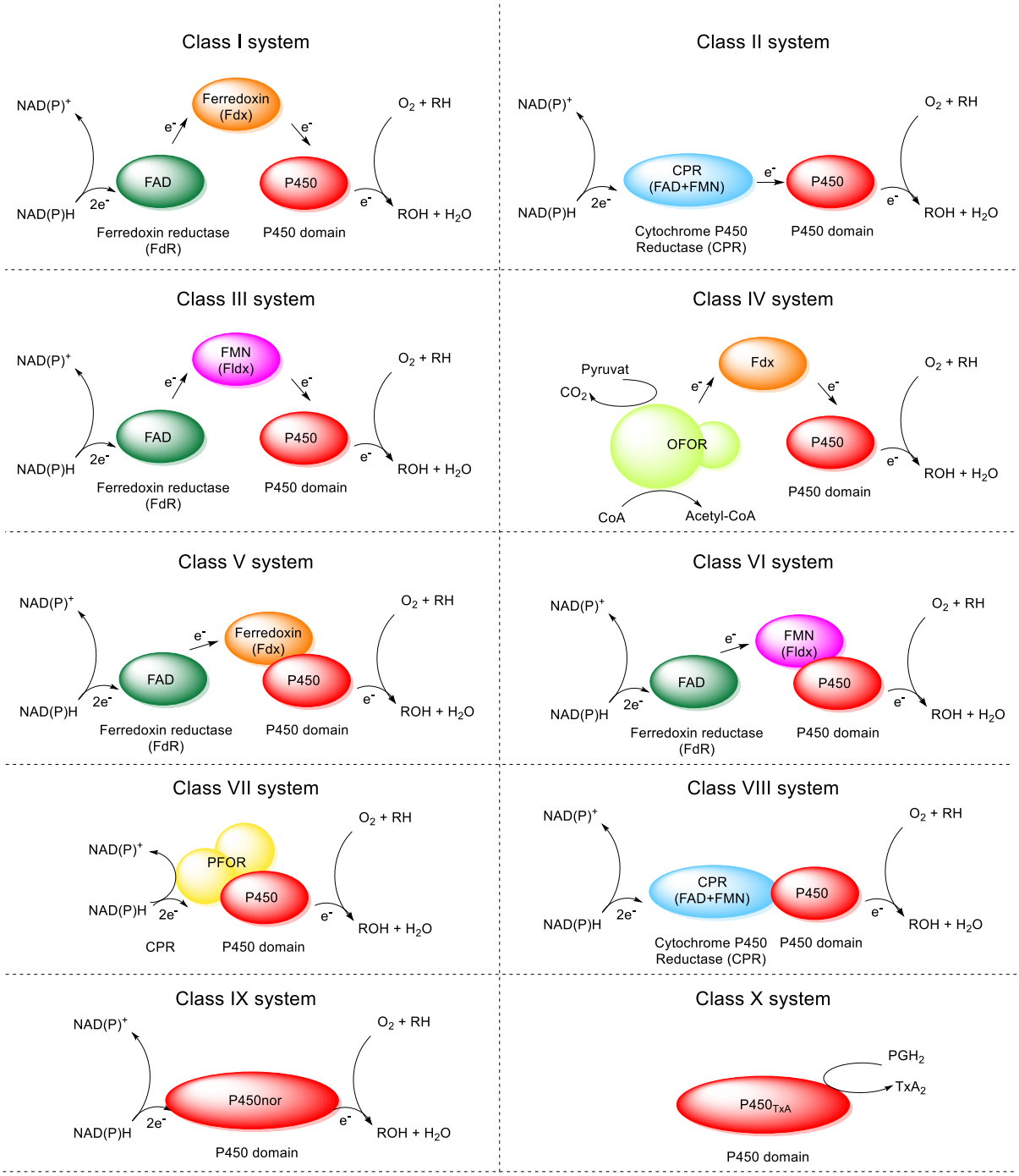


Figure 1.9 The ten classes of redox proteins in P450s.[71-104]

Class VIII contains a eukaryotic-like diflavin reductase, cytochrome P450 reductase (CPR), that is fused to the P450 domain in a single polypeptide chain, which make it catalytically self-sufficient as monooxygenases. The most well-studied member in this class is the CYP102A1 ($P450_{BM3}$) which is composed of a P450 domain connected via a short protein linker to the

reductase domain that contains 1 equivalent each of the cofactors FAD and FMN.[98,99] The electrons are transferred one-at-a-time from FAD to FMN and further on to the P450 haem domain.[100,101]

Class IX and X are special cases of a cytochrome P450. CYP55 (P450nor), an example of a class IX system, uses NADH as electron donor and catalyses reaction independent of other electron transfer proteins.[102,103] P450s in class X catalyse substrate conversion using an independent intramolecular transfer system, such as the allene oxide synthase and fatty acid hydroperoxide synthase.[104]

1.3.1 The P450 catalytic cycle

The catalytic cycle of P450-mediated reaction has been well studied (Figure 1.10).[105] In the resting state (**I**), the ferric iron atom is six-coordinated, bonded with the four equatorial nitrogen ligands in the porphyrin ring; a water molecule and a cysteine thiolate group are axially *trans*-coordinated to the ferric iron atom. The arrival of a substrate causes the loss of the water molecule from the ferric iron which becomes penta-coordinated (**II**) and the ferric spin state changes from low spin (S = 1/2) in **I** to high spin (S = 5/2) in **II**. These changes could be observed spectrophotometrically that the absorption maximum peak shifted from *ca.* 418 nm to *ca.* 390 nm. The first electron provided by NAD(P)H is shuttled to the ferric iron centre and reduces it to the Fe^{2+} form (**III**). Binding of dioxygen yields the oxy complex **IV** which is converted to the ferric peroxy complex **V** with the transfer of the second electron. Protonation of the distal oxygen gives a hydroperoxy complex **VI**, also called "Compound 0". A second protonation at the distal oxygen causes the loss of a water molecule via the heterolytic cleave of the O–O bond and gives Compound I (**VII**), a ferryl oxo species which has been captured and characterised.[106,107] **VII** is the active intermediate responsible for C–H bond activation. In

hydroxylation reactions, a hydrogen atom is abstracted from the substrate by the ferryl oxo species **VII** to give a substrate radical and the Fe^{IV}–OH complex **VIII** (Compound II).[108] Finally, the "radical rebound" of the substrate radical with the OH• gives an alcohol product R–OH loosely bound to a ferric haem which returns to **I** or **II** depending on whether a water molecule re-coordinates with the ferric iron to generate the resting state **I** or another substrate displaces the product and starts another cycle of reaction from complex **II**. This oxygen or radical rebound mechanism of the C–H activation reaction gives the impression that an oxygen atom has been directly inserted into the C–H bond of substrates.[42]

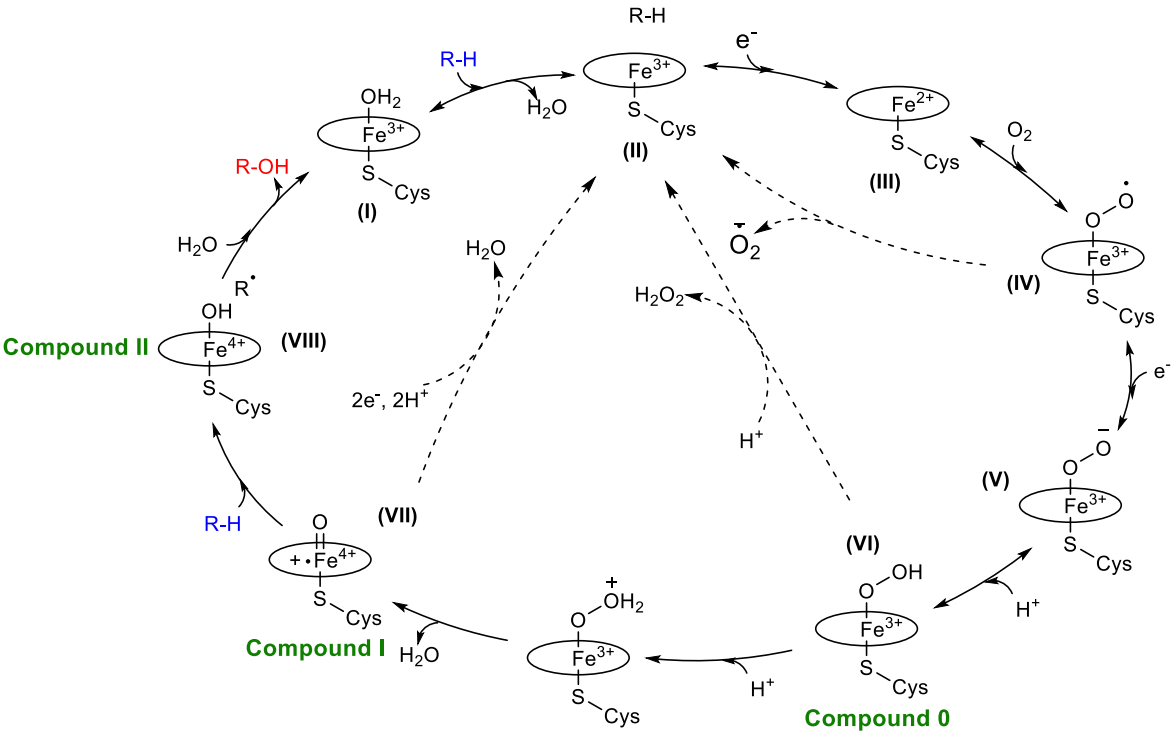


Figure 1.10 The catalytic cycle of P450-mediated reaction.[107,110]

In the catalytic cycle, NAD(P)H consumption is not always fully coupled to product formation due to three major uncoupling reactions.[109,110] In the first reaction, if dioxygen binding is hindered, slow transfer of the second electron to complex **IV** may result in the loss of superoxide and return of **IV** to the resting state (superoxide uncoupling). This may occur if

oxygen binding could not be stabilised by the correct conformation of enzymes or the binding is sterically interfered by substrates. In the second reaction, hydrogen peroxide may be formed if **VI** is protonated at the iron-bound oxygen instead of the distal oxygen due to the loosely fitting substrates. In this reaction the P450 enzyme displays peroxygenase activity. Finally, the ferryl-oxo species may be reduced back to the ferric form (**II**) if the substrate is bound too far away or no C–H is available to be abstracted from the substrate. Thus, the binding between enzymes and substrates may have to be improved to promote the correct substrate binding conformations for higher monooxygenase activity.

1.4 $P450_{BM3}$

CYP102A1, the first catalytically self-sufficient P450 enzyme, was discovered by Fulco *et al.* to act on medium chain fatty acids.[99,111-113] As the third P450 to be isolated from *Bacillus megaterium*, CYP102A1 was originally designated as $P450_{BM3}$; it has 1046 amino acid residues with a molecular weight of 120 kDa and comprises a 55 kDa P450 haem domain fused at its C-terminus to a 65 kDa reductase domain.[114,115] This reductase domain contains two flavin prosthetic groups, FAD and FMN, in an equimolar ratio.[116] This fused structure provides for intramolecular electron transfer from the flavin domain to the haem domain without the complications that arise from the protein-protein recognition and binding steps required for the eukaryotic P450s.[117,118]

Another unusual feature of $P450_{BM3}$ is that it dimerises in solution (Figure 1.11).[117,119] The dimer is the predominant form at higher concentration under most conditions and it is the only form with significant activity.[120] Detailed investigation revealed that the electron-transfer (ET) pathway in $P450_{BM3}$ passes through both protein molecules; from the FAD domain of one molecule to the FMN domain of the other molecule (intermolecular ET) before passing to either

haem domain.[120] It was observed that the activity dropped when the dimer irreversibly dissociates to inactive monomers at low ionic strengths, suggesting at least one of the electron transfer steps, i.e. FAD-FMN or FMN-to-haem, occurs across the dimer interface.

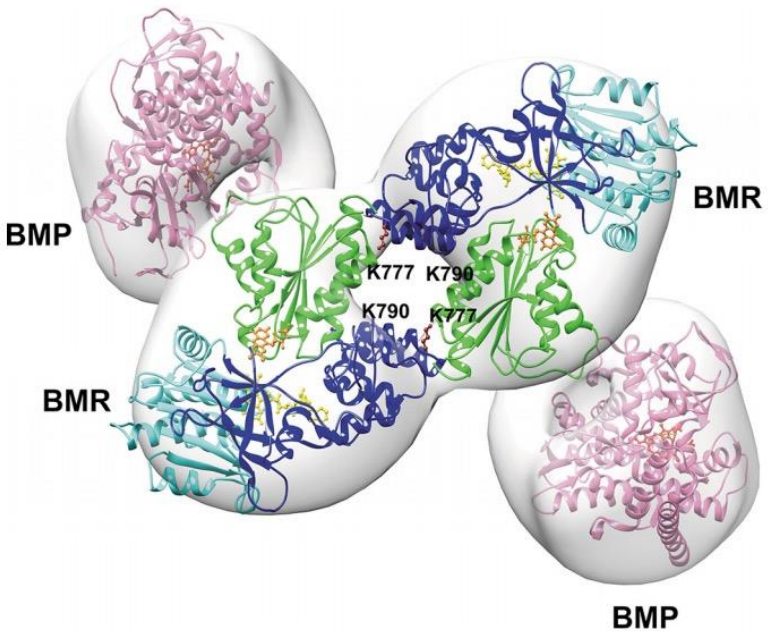


Figure 1.11 3D model of dimeric $P450_{BM3}$ showing the close view of the dimerization.[121] The crystal structure of the BMP (PDB: 4KEW) and the BMR homology model were constructed using Chimera. BMP molecule (haem domain) is shown in magenta, and the FMN, FAD, and nucleotide-binding domains of the BMR are shown in green, blue and cyan, respectively. The haem and cofactors FMN, FAD is shown in red, orange and yellow, respectively.

This unusual fused structure of FAD-FMN domains makes $P450_{BM3}$ catalytically self-sufficient and facilitates the intramolecular electron transfer process in the reaction, hence promoting high turnover rate (up to 4,000 min^{-1}) for the oxidation of medium-chain (C12–C20) fatty acids–the natural substrates for $P450_{BM3}$.[41,122] For example, the oxidation of lauric acid by $P450_{BM3}$ gives oxidation at sub-terminal positions, with ω–1, ω–2 and ω–3 hydroxylation in approximately equal proportions (Figure 1.12).[41]

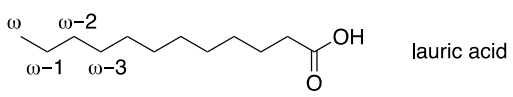


Figure 1.12 Product distributions in the oxidation of lauric acid by $P450_{BM3}$.[41]

1.5 Crystal structure of $P450_{BM3}$

1.5.1 Substrate-bound and substrate-free forms of $P450_{BM3}$

The crystal structure of the functional dimer form of $P450_{BM3}$ has not been reported but the separately expressed haem domains of both the substrate-free (SF) and substrate-bound (SB) have been crystallised. The SF-form of $P450_{BM3}$ structure is shown in Figure 1.13–A with helices labelled in different colours. The haem iron sits nearly in the centre of the structure, with the I-helix (yellow) above the haem centre and the L-helix (red) underneath the haem porphyrin ring. The F-helix is positioned over and is at right angles to the I-helix.

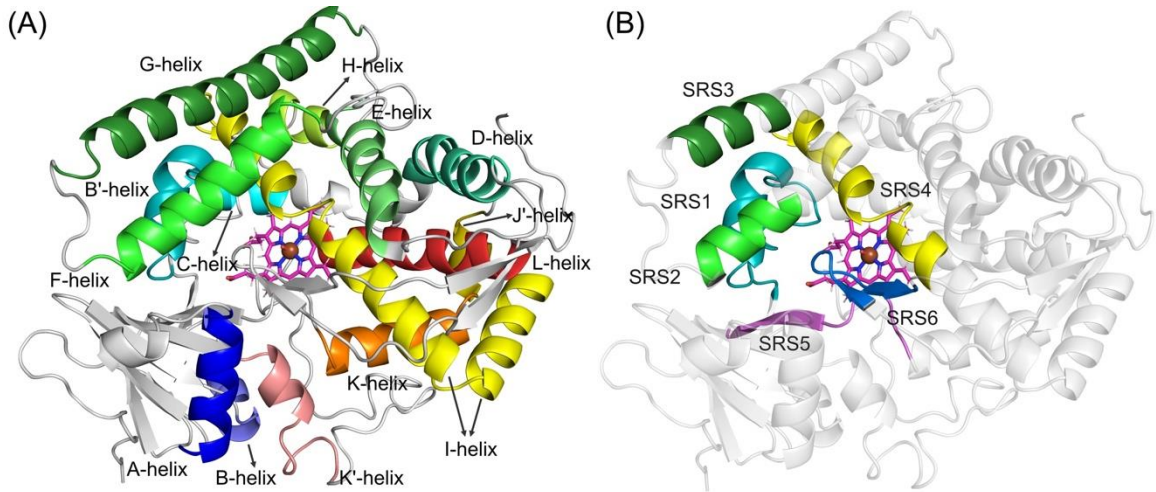


Figure 1.13 (A) 3D structure of $P450_{BM3}$ with helices shown in different colours (PDB: 1BU7);[118] A-helix: Lys24–Gly37 in blue; B-helix: Ser 54–Cys62 in slate; B'-helix: Ser72–Gly83 in cyan; C-helix: Glu93–Ile102 in sky-blue; D-helix: Tyr115–Arg132 in split-tea; E-helix: Pro142–Asn159 in pale green; F-helix: His171–Arg190 in green; G-helix: Ala197–Ser226 in forest; H-helix: Asp232–Asn239 in limo green; I-helix: Asp250–Leu298 in yellow; J'-helix: Ser304–Gln310 in dark yellow; K-helix: Leu311–Trp325 in orange; K'-helix: Leu356–Gly368 in salmon; L-helix: Gly402–Lys419 in red. **(B)** Six SRSs regions; SRS1: Lys69–His92 in teal; SRS2: Leu181–Leu188 in green; SRS3: Glu200–Asp208 in forest; SRS4: Asn253–Gly271 in yellow; SRS5: Trp325–Ala335 in purple; SRS6: Lys434–Pro441 in blue.

The alignment of $P450_{BM3}$ with 14 other sub-family members of A1 suggests a few most variable regions–the flexible substrate recognition regions (SRS)–which enable P450s to be the most versatile biological catalyst.[41,123] All the reported point mutations that significantly affect

the substrate specificities of the parental CYP102 enzymes fell within or overlapped some of these SRSs.[123,124] There are six regions identified as substrate recognition sites (SRS1–6) across the P450 superfamily (Figure 1.13–B).[41,125,126] The alignment shows that there is relatively small deviation in SRS6 (residues 434–441) while SRS4 (residues 253–271) is almost fully conserved. SRS1 (69–92) shows half conserved/half abnormal property, especially at residues 72–74. SRS2 (181–188) and SRS5 (325–335) each present interesting deviations. SRS3 (200–208), on the other hand, is almost completely different across the 15 sub-family enzymes.[127]

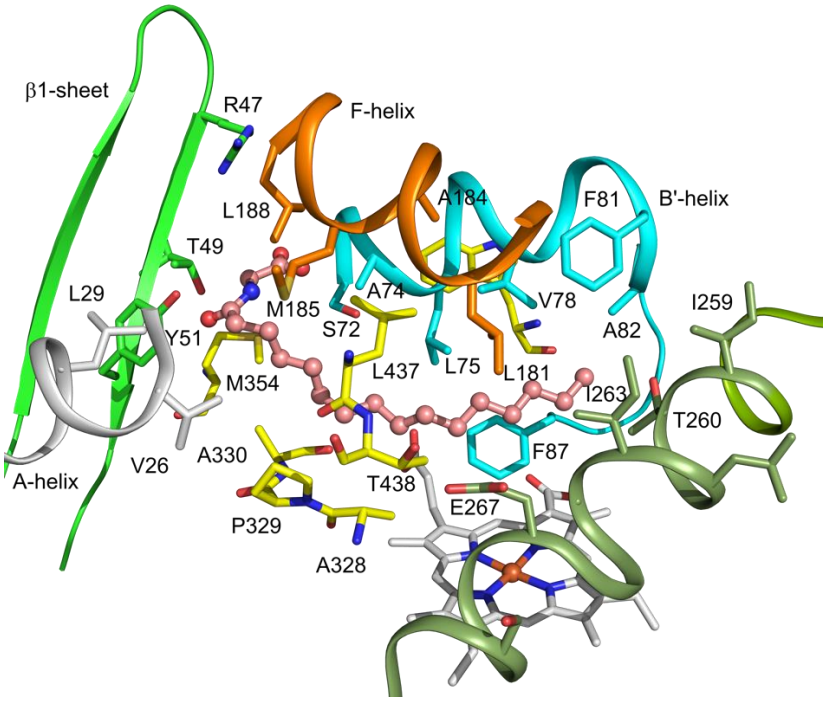


Figure 1.14 The active site structure of *N*-palmitoylglycine-bound $P450_{BM3}$ (PDB: 1JPZ) and key amino acid residues.

The active site structure of $P450_{BM3}$ and some key residues that contact the bound *N*-palmitoylglycine (NPG) substrate are shown in Figure 1.14.[128] Fatty acid substrates are bound in an extended conformation with the carboxylate group located at the entrance to the substrate access channel where the carboxylate oxygens interact with Arg47 and Tyr 51.[41] The aliphatic

side of the molecule reaches over and beyond the haem iron, containing residues V78 and A82 deep into the substrate pocket (Figure 1.14).

The oxidation of *N*-palmitoylglycine by $P450_{BM3}$ showed ω–1, ω–2, and ω–3 hydroxylation. Surprisingly, in the SB crystal structure, the bound *N*-palmitoylglycine was far away from the haem iron centre that none of the carbon atoms were within 7 Å of the haem iron. This suggested that conformational changes or significant substrate motion must occur during the catalytic cycle for the oxidation of *N*-palmitoylglycine. Even though the SB structure of the haem domain is not representative, this structure has revealed residues that are in contact with the substrate and provided basis for mutagenesis and molecular dynamics simulation studies.

Other conformational changes are observed between the SF and SB structures such as a displacement of the F and G helices to make space for the fatty acid aliphatic chain.[129] In the SF form, the axial water molecule bound to the haem iron forms a hydrogen bond with the carbonyl oxygen of A264 in the I helix. With the introduction of substrate (SB form), this water molecule is removed and the I helix moves away from the haem iron compared with the SF form, notably the I263 and A264 residues.[128] The extended aliphatic chain of *N*-palmitoylglycine also induces the rotation of the aromatic ring of F87 by ~90°.[130] In general, the SB structure is much more compact compared with the SF form, therefore, if the enzyme could be engineered towards a SB-like conformation, the oxidation of unnatural substrates may be improved.

1.5.2 Key residues of $P450_{BM3}$

The entrance of substrate access channel has two key residues, Arg47 and Tyr51, which are proposed to be involved in fatty acid recognition; hence, the R47/Y51 couplet was an early target for mutagenesis. It was found that the R47A mutation[131,132] reduced fatty acid oxidation

activity while the R47E mutant accepted alkyl trimethyl ammonium as a new substrate.[133] The hydrophobic substitutions R47L and Y51F were combined to promote the entry and binding of hydrophobic compounds such as polycyclic aromatic hydrocarbon (PAH).[134] The R47L/Y51F combination was then used as the base mutant for further engineering of $P450_{BM3}$.[69,134,135]

Phe87 is close to the haem iron and strongly influences $P450_{BM3}$ substrate specificity and product selectivity.[136,137] It has been proposed that the rotation of the benzene ring in the SB form is responsible for the displacement of the distal axial water ligand.[128] It is the most commonly mutated residue; F87V and F87A being the most useful substitutions as they create space for sterically demanding substrates to approach the haem.[137-139] Ala264 is hydrogen-bonded to the axial water ligand. The combination of F87A and A264G mutations greatly improved the oxidation of PAHs such as pyrene.[140]

In the SB form of WT $P450_{BM3}$ (Figure 1.14), the end of the aliphatic chain reaches over the haem iron and contacts Ala82. It was hypothesised that the side chain of A82 created a hole that may isolate the substrate from the haem iron, therefore, a bulky substitution at A82 would fill in this hole and increase the oxidation activity.[141] The smaller active site generated by A82F and A82W mutations showed increased oxidation activity for *N*-palmitoylglycine and omeprazole.[124] Other than Ala82, Peterson also identified another residue that should influence substrate binding, Ala328, which resides directly above the haem cofactor and is closest to the proximal side of the haem iron.[142] The substitution of a hydrophobic valine onto A328 showed dramatic shift in the regioselectivity of hydroxylation on longer alkanes such as octane and nonane.[143]

1.6 Protein engineering

1.6.1 Rational protein design and directed evolution

In order to develop $P450_{BM3}$ variants suitable for oxidation of diverse range of substrates, a variety of techniques of enzyme engineering are currently applied to $P450_{BM3}$ to explore its catalytic potential. There are two general strategies for protein engineering: rational protein design and directed evolution. Rational design of protein requires detailed knowledge of the structure and function of a protein to mark targeted regions and residues for mutagenesis. With clear and defined targets, rational design mutagenesis is relatively inexpensive and technically easy to perform. However, the major drawback is that the detailed structural knowledge of a protein is not always available, even with the proteins that have been purified and crystallised, the actual transition state of the reaction remains impossible to be captured. Therefore, it would be difficult to predict the effects of various mutations and combinations. Directed evolution requires no prior knowledge of a protein and little prediction of effects of mutations. In directed evolution, random mutagenesis methods such as error-prone PCR (ep-PCR) and DNA shuffling could generate a large number of mutants with random mutations at random residues. For instance, the introduction of 20 amino acid substitutions in a protein of 500 amino acids would generate a maximum of 20^{500} mutants. Site-saturation mutagenesis is another common technique in directed evolution where all possible substitutions are introduced at amino acids residues.[144] Even saturation only at 2 sites would generate 20^2 mutants for screening. This large number of variants often require expensive robotic equipment to automate the process and high-throughput screening for data analysis.

1.6.2 Combination of rational protein design and directed evolution

These two methods are not mutually exclusive, the combination of these methods has been widely applied in protein engineering. A method called rational evolution, is applied to engineer $P450_{BM3}$ variants for fatty acid oxidation.[145] Eight residues were selected from computational simulation for site-saturation mutagenesis. Some variants were found to produce the blue pigment indigo from indole oxidation.[145,146] With further site-saturation and random mutagenesis, a triple mutant A74G/F87V/L188Q (GVQ) was identified to hydroxylate indole, leading to the formation of indigo and indirubin.[145,147] The GVQ variant also showed activity on a broader range of substrates than the WT, including alkanes, alicyclic, aromatic, and heterocylic compounds.[147] Thus, it validates the use of indole oxidation as an indicator for active mutants in the screening.

Similarly, Wong and co-workers used indigo formation from indole oxidation as a preliminary activity screen and then sub-screening *in vivo* for selectivity as well as activity.[148] ep-PCR was carried out on the haem domain-coding regions of the WT $P450_{BM3}$ gene and mutant F87A. Variants showing the most intensive indigo formation were selected for DNA shuffling and a second round of indigo screening. Of ~800 colonies, ~130 were picked for whole-cell oxidation of two substrates, naphthalene and propylbenzene. Four of the most promising variants were identified with much enhanced activity: A330P; A191T/N239H/I259V/A276T/L353I (KT2); F87A/H171L/Q307H/N319Y (KSK19); F87A/A330P/E377A/D425N (KT5), which were used as base variants for further mutagenesis.[69,135,149,150] WT, the F87A variant, and the highly active GVQ variant developed by Schmid and co-workers, were tested together with these four new variants. GVQ showed higher product formation rate (PFR) comparing with WT (1015 min^{-1} *vs.* 606 min^{-1}), but these new variants showed even higher activity by at least an order of

magnitude over WT; the highest activity for propylbenzene was 2688 min^{-1} catalysed by R47L/Y51F/KT2 compared to 606 min^{-1} for WT.

Of these four variants, only F87A is close to the haem centre. Ala330 is in a loop between two β-strands, His171 is in the E/F loop, Ala191 is in the F/G loop, Asn239 is in the H/I loop, Ile259 and Ala276 are in the I helix and Leu353 is near the substrate channel. Other residues such as Gln307, Asn319, Glu377 and Asp425 are on the surface of the haem protein. It is unusual to have a single mutation to have such high activity, variants A330P and I401P were crystallised for detailed analysis (Figures 1.15 and 1.16).[151] In Figure 1.15, compared with the WT structure, the proline mutation on Ala330 leads to the relocation of the adjacent P329 residue towards the haem centre which was closer to the F87 side chain (5.9 Å instead of 9 Å). The substrate pocket was reshaped dramatically, and the more constrained active site abolished fatty acid oxidation while enhancing activity for smaller molecules, such as naphthalene and propylbenzene.

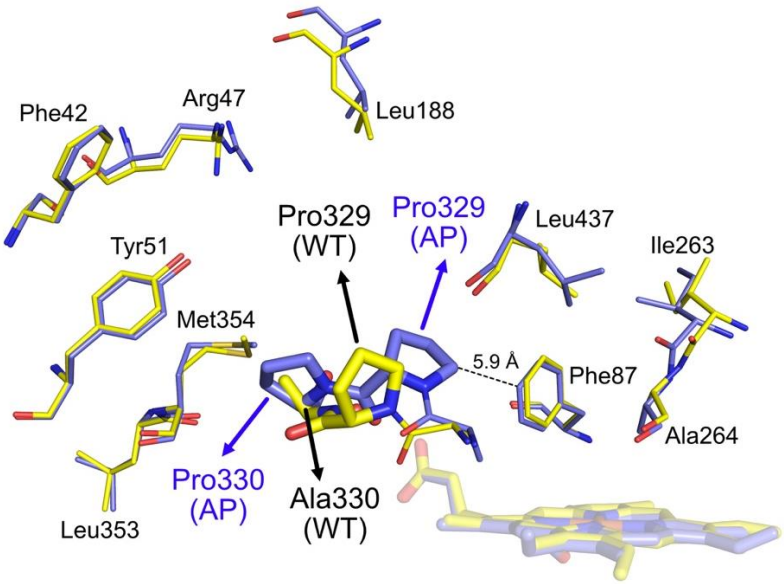


Figure 1.15 Overlay structure of the active-site residues of variant A330P (in blue, PDB: 3M4V)[151] and WT (in yellow, PDB: 1BU7)[118] showing the relocation of the Pro329 residue in A330P into the substrate pocket above the haem centre, and the distance between this Pro329 residue and the side chain of Phe87 is only 5.9 Å compared to 9 Å in WT.

The I401P mutant showed unexpectedly high activity for non-natural substrates.[152] The overlaid crystal structures of I401P and SF WT showed a number of unexpected features (Figure 1.16). The I401P mutation, adjacent to the proximal Cys400 residue, caused the proximal loop to drop, taking the haem iron down slightly from the substrate pocket, resulting in lengthening of the iron-axial water distance from 2.6 Å to 3.6 Å. In this respect, the structure resembled that of the SB form of the WT. Two I-helix residues, Gly265 and His266, were reorientated, and prevented the formation of several intrahelical hydrogen bonds. Overall, the I401P mutation induces conformational changes that facilitates the initiation of the catalytic cycle, hence, increases the oxidation activity for non-natural substrates.

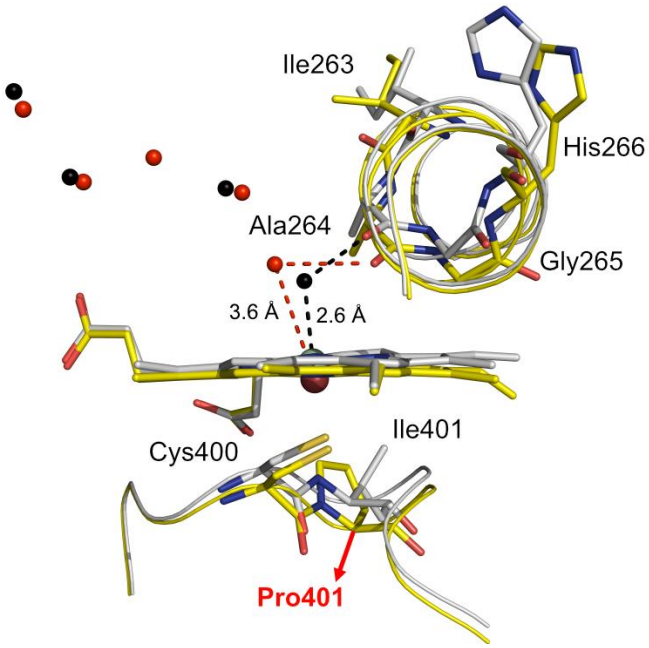


Figure 1.16 The overlay structure of active site in I401P variant (in yellow, PDB: 3HF2)[151] with SF WT (in grey, PDB: 1BU7)[118] showing the iron atom, the porphyrin and the proximal loop residues of I401P dropping away from the substrate pocket comparing with the WT.

The KT2 variant also showed unexpected high activity for non-natural substrates. Because the mutations in KT2 (A191T/N239H/I259V/A276T/L353I) are at residues of no obvious structural importance, it was not immediately clear how these mutations effected the increased activity. The crystal structure of SF KT2 showed that the I helix adopted a different structure

from SF WT (Figure 1.17–A).[153] A number of interhelix hydrogen bonds and salt bridges were broken (black dashes), and the hydrogen bond between the haem iron-bound water and the Ala264 carbonyl was broken (blue spheres). These conformational changes resulted in the retreat of the I helix of KT2 which surprisingly resembled the I helix structure of SB WT[118] (Figure 1.17–B). The overlay between SF KT2 and SB WT was good, and a newly formed hydrogen bond between the carbonyl of I263 and amide of E267 stabilised the structure. The KT2 mutant was thus described as being in a "catalytic ready" conformation.

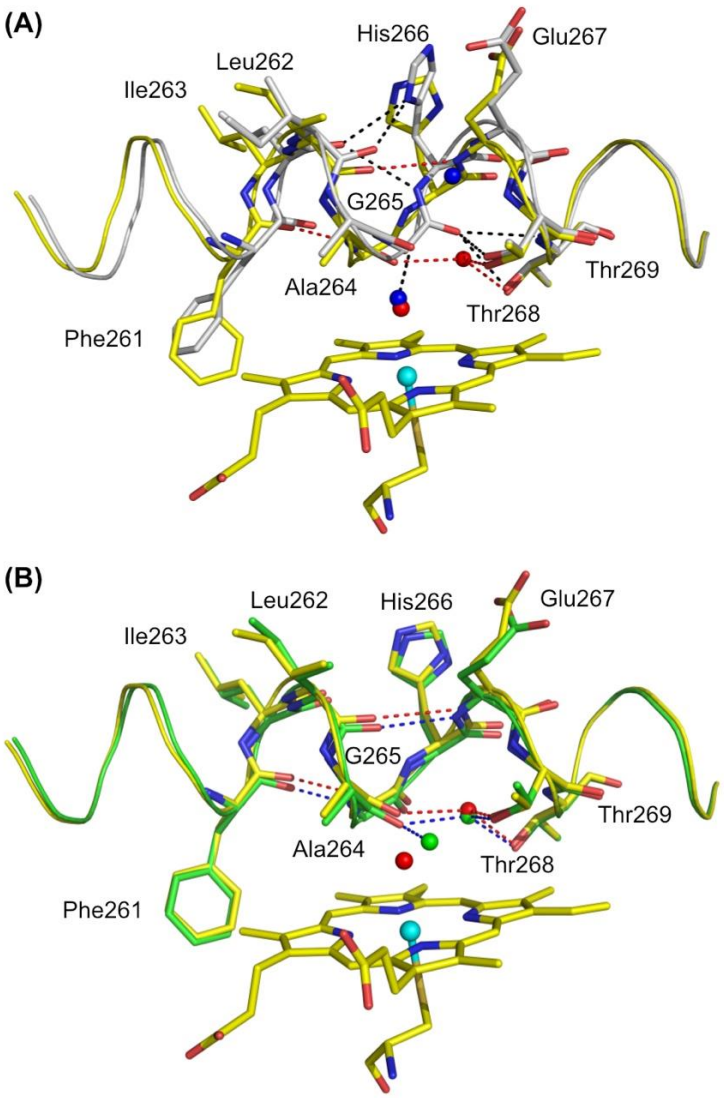


Figure 1.17 Crystal structure of I-helix of SF KT2 (PDB: 3PSX);[153] **(A)** Overlay structure of I-helix of SF KT2 (yellow, with hydrogen bonds in red and water molecules as red spheres) with SF WT (in grey, with hydrogen bonds in black and water molecules as blue spheres). **(B)** Overlay structure of I-helix of SF KT2 (yellow, with hydrogen bonds in red and water molecules as red spheres) with SB WT[118] (in green, with hydrogen bonds in blue and water molecules as green spheres).

Arnold and co-workers used combinational mutagenesis strategies to improve the hydroxylation of alkanes by $P450_{BM3}$.[154] Five rounds of random mutagenesis targeted to the haem domain, together with recombination of beneficial mutations, led to the variant 139-3 with greatly improved octane oxidation activity (TTN~1000).[155] This variant, with 11 mutations (V78A/H138Y/T175I/V178I/A184I/H236Q/E252G/R255S/A295T/L353V), not only improved octane oxidation, but also enhanced propane oxidation (TTN~500).[143,155] After further rounds of random mutagenesis followed by site-saturation mutagenesis, thousands of variants were generated and screened using dimethyl ether as surrogate and detecting the formaldehyde formed via demethylation. Variant 35E11 was found to have greatly increased propane oxidation activity to 2-propanol (TTN~6000).[156] Out of the 17 mutations accumulated in variant 35E11, only two were at active site residues, highlighting the power of directed evolution in tracking beneficial mutations through the great effort of high-throughput screening. Finally, in order to further evolve variant 35E11, a "domain engineering strategy" was proposed. Based on the previously identified beneficial mutations within the P450 domain, random and site-saturation mutagenesis were also performed separately in the FMN and FAD reductase domains. The combination of these beneficial mutations on three domains evolved a proficient propane monooxygenase, name $P450_{PMO}$, possessing high catalytic efficiency and coupling efficiency for this reaction (up to 45,000 TTN).[157-159]

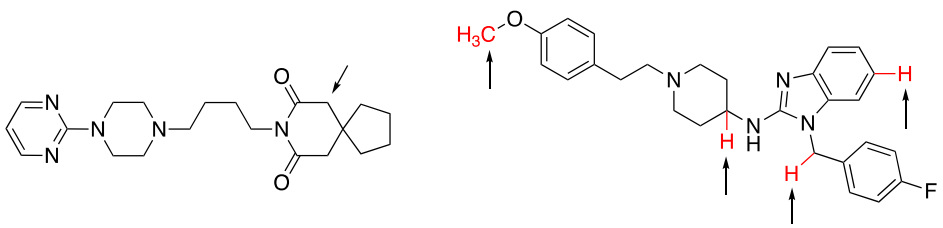


Figure 1.18 Examples of human drug metabolites catalysed by the engineered $P45_{BM3}$ variants; oxidation sites are indicated by arrows.[160,161]

Other than propane oxidation, these mutants were also found to oxidise numerous drug molecules,[160] ranging from smaller molecules such as ibuprofen to larger ones such as buspirone[160] and verapamil (Figure 1.18).[161]

Vermeulen and co-workers rationally designed a triple mutant R47L/F87V/L188Q that was found to have enhanced activity for the oxidation of drug molecules like testosterone (TST), amodiaquine, dextromethorphan, and 3,4-methylenedioxymethylamphetamine (MDMA).[162,163] Based on this variant, further random mutagenesis was conducted to improve its activity. Three generations of ep-PCR gave more than 1,000 mutants, from which R47L/F87V/L188Q/E267V/G415S & R47L/E64G/F81I/F87V/E143G/L188Q/E267V/G415S (M01 & M11) were identified as the most active mutants that gave the highest activity for MDMA and dextromethorphan.[164] It was found that increasing the size of the active site residue at position 82 (A82W and A82F) in M01 and M11 variants improved the oxidation of testosterone at the 16β position,[165] while the extra S72I mutation inverted the selectivity towards 16α-hydroxylation, both with moderate selectivity and conversion (Figure 1.19).[166] However, multiple rounds of random and site-saturation mutagenesis may be required to improve the selectivity for targeted products.

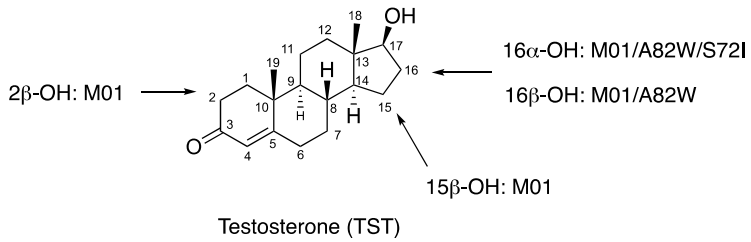


Figure 1.19 Oxidation sites of testosterone by engineered $P450_{BM3}$ variants developed by Vermeulen and co-workers.[162,163,165,166]

1.6.3 Combinatorial active-site saturation test (CAST) and iterative saturation mutagenesis (ISM)

In order to narrow down the target list of mutants and reduce the size of mutant library, Reetz and co-workers developed two methods, Combinatorial Active-site Saturation Test (CAST)[167] and Iterative Saturation Mutagenesis (ISM),[168] to generate a more active-site-focused variant library guided by structural information on the enzyme. CAST involves randomisation of amino acids at several sets of spatially close positions around the active site. For instance, on the basis of the secondary structure of an enzyme, if one member of the pair occurs in a protein sequence at position *n*, then the second one is found at a sequence of (*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. Randomisation on these spatially close pairs of residues will generate a number of sub-libraries, also called focus libraries, for activity screening. This method is always combined with ISM for further rounds of mutagenesis. In ISM, the beneficial mutant gene from a given sub-library is used as the template for another round of saturation mutagenesis at other sites; multiple rounds may be required until reaching a desired level of activity. These two approaches significantly reduce the time required to obtain hits, especially when a high throughput screen is not available. Reetz and co-workers have demonstrated directed evolution of $P450_{BM3}$ for testosterone oxidation at 2β and 15β positions using iterative CAST.[169] The F87A mutant was selected as template, it gave a roughly 1:1 ratio of 2β and 15β alcohols of testosterone. Twenty active site residues were grouped into nine sites, A (A47, T49, Y51), B (V78, A82), C (M185, L188), D (S72, A74, L75), E (L181, A184), F (T260, I263); G (A328, A330), H (L437, T438) and I (V26, T29). Three residues, S72, A328 and A330, were selected for saturation mutagenesis while site B residues (V78, A82) were randomised using NNK codon degeneracy and sites A and C (A47, T49, Y51, M185, L188) were subjected to NDC mutagenesis. The NDC codon degeneracy encodes only 12 amino acids

which are mostly hydrophobic amino acids substitutions. Up to 18,000 variants were generated for testosterone oxidation screening and high regio- and stereo-selectivity for 2β- and 15β-hydroxylation were observed, with F87A/A330W giving 97% of 2β-hydroxy-TST at 79% conversion and R47Y/T49F/V78L/A82M/F87A giving 96% of 15β-hydroxy-TST at 85% conversion (Figure 1.20).[169]

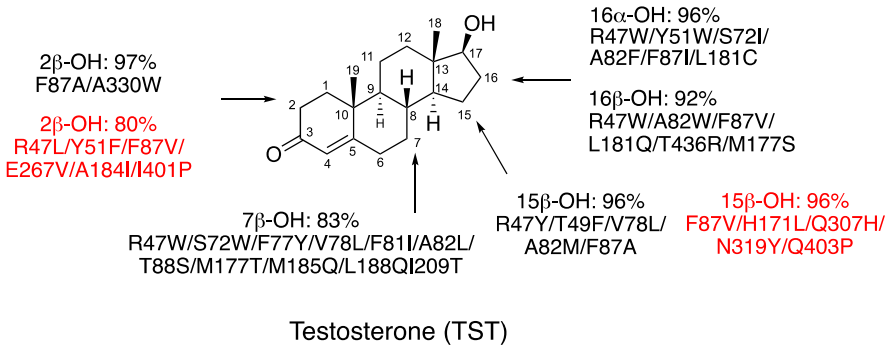


Figure 1.20 Oxidation sites of testosterone by engineered $P450_{BM3}$ variants developed by Reetz and co-workers (2β, 7β, 15β, 16α, and 16β-hydroxylation in black) and Wong and co-workers (2β and 15β-hydroxylation in red).[135,169-172]

Later, using similar ISM-CAST approach but with a broader coverage of residues and more site-saturation mutagenesis, also taking into account beneficial mutations discovered by Vermeulen and co-workers (A82W and S72I), Reetz and co-workers developed $P450_{BM3}$ mutants for 16α- and 16β-hydroxylation for testosterone and 4 other steroidal substrates with high selectivity and conversion (Figure 1.20).[170] More recently, a larger library (more than 30,000) was generated using similar approach by randomisation of 2 amino acids across 15 residues with the aim of 7β-hydroxylation of testosterone and 5 other steroids, which were obtained at medium to high selectivity and conversion (Figure 1.20).[171] However, testosterone oxidation showed low diversity of products even screening >50,000 mutants across different efforts.

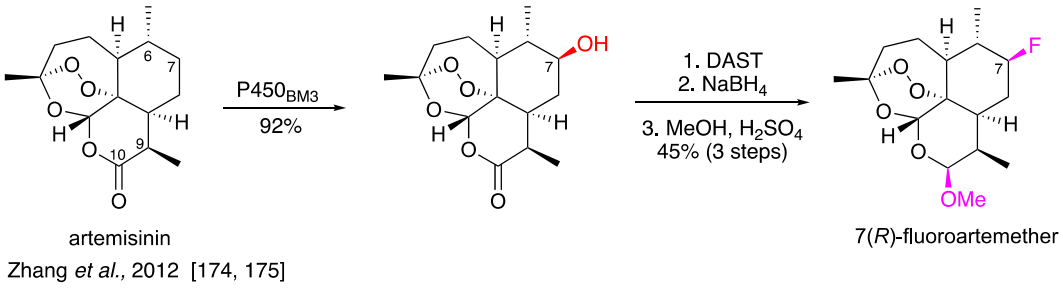
1.6.4 Rational designed focused libraries

To obtain high regio- and stereo-selectivity of testosterone metabolites, even though with the development of ISM and CAST approaches to reduce the size of variant library, huge efforts are required to mutate large collection of recombinant DNA and high-throughput screening. Wong and co-workers developed a panel of ~100 $P450_{BM3}$ variants by rational protein design. This library was constructed from five base variants, K19 (H171L/Q307H/N319Y), R19 (R47L/Y51F/K19), RP (R47L/Y51F/I401P), RT2 (R47L/Y51F/KT2), and GVQ (A74G/F87V/L188Q).[148,152] Additional mutations at two or more active site residues (R47, Y51, S72, V78, F81, A82, F87, A184, I263, A264, E267, A328, A330 and L437) were introduced to these base variants to generate diverse substrate pocket shape and size for accepting unnatural substrate.[69,135,149,150] Variants in this library were broadly divided into three groups. One group contained F87A or F87V mutations to create space close to the haem, which were crucial for large molecule oxidation such as sesquiterpenes and steroids. Another group included A330P and A330W mutations which restricted and reshaped the substrate pocket; the third contained wild type-like variants. Screening with this small library of engineered mutants, diverse range of substrates have shown high selectivity and TTN. For example, testosterone oxidation with this focused library achieved high regioselectivity for 15β-hydroxylation (96%) with excellent substrate conversion (83%); C2 oxidation of testosterone was also observed with 80% selectivity and 20% conversion (Figure 1.20).[135] The screening of drug molecules such as diclofenac, naproxen, chlorzoxazone, amitriptyline and lidocaine; smaller drug core motifs quinolines and tetrahydroquinoline; and diterpene eleuthoside not only showed high selectivity and TTN, but also gave corresponding alcohol precursors for valuable drugs synthesis.

Seifert *et al.* also constructed a $P450_{BM3}$ library by combining five hydrophobic amino acids (A, V, F, L, and I) in two positions, F87 in SRS1 and A328 in SRS5.[173] F87 has been shown to

be significant in creating space when mutated to smaller amino acids for oxidation reaction while mutations at A328 has been predicted to systematically alter the substrate pocket shape by changing the size of side chains. Twenty-four variants were generated and tested for oxidation of geranylacetone, nerylacetone, (4*R*)-limonene and (+)-valencene. The oxidation of geranylacetone showed improved selectivity for the 9,10-epoxide (97%) comparing with WT (54%); new oxidation sites at C11 and C12 were also observed with 37% and 80% selectivity. Oxidation of nerylacetone and limonene largely improved the selectivity for their corresponding epoxides (up to 94%). Valencene was oxidised preferentially at allylic C2 position to produce nootkatol and small amounts of the grapefruit nootkatone.

Fasan and co-workers developed a larger library of ~500 mutants from the base variants in Arnold's work and a variety of mutants derived from random mutagenesis studies. This library was also called "training set" that was used to screen for residues with high selectivity. For example, 20 out of the 500 variants showed activity for artemisinin oxidation. Variant V78A/F81S/A82V/F87A/F142S/T175I/A180T/A184V/A197V/F205C/S226R/H236G/R255S /A290V/L353V (FL#62) gave 83% selectivity towards 7β hydroxylation with 10% of 7α-hydroxy-artemisinin. C7 oxidation of artemisinin remains inaccessible by chemical oxidation via iron-based catalysts, which favour the more electronically activated C–H bond in C6. Therefore, the preferential C7 oxidation of artemisinin by engineered $P450_{BM3}$ provided an ideal precursor for the synthesis of 7(*R*)-fluoroartemether, a fluorinated derivative of the antimalarial drug artemether, which is the major metabolite in human (Figure 1.21).[174,175]



artemisinin

Zhang *et al.*, 2012 [174, 175]

7(*R*)-fluoroartemether

Figure 1.21 Chemoenzymatic synthesis of 7(*R*)-fluoroartemether from artemisinin via the catalyst of $P450_{BM3}$.[174,175]

To improve the activity of this variant, a series of active site libraries were constructed via site-saturation (NNK) mutagenesis across 7 residues (74, 78, 81, 82, 87, 181, 184). After screening of 12,500 variants, selectivity for the 7β-alcohol was improved to 100% with a TTN of 360. This synthetic route via late-stage functionalisation has been applied to a range of drug-like molecules.[176]

1.7 Application of mutant $P450_{BM3}$

Over the past decade, $P450_{BM3}$ has been engineered for the oxidation of alkanes, terpenes, heteroaromatics, alkaloids, steroids and other classes of compounds.[41,42,154,177-179] Engineering biocatalysts for the production of fine chemicals through biosynthetic pathways is becoming attractive as an alternative to synthetic chemistry. Keasling and co-workers developed a novel semi-biosynthetic route for the production of artemisinin using engineered $P450_{BM3}$ variants, with the aid of computational study.[180] Both WT and mutated active sites of $P450_{BM3}$ were constructed and cross-compared using Rosetta algorithm. It was found that the F87A[181] mutation greatly relieved the steric hindrance imposed upon substrate amorphadiene and allowed more substrate-haem interaction in the active site. Mutation A328L appeared to restrict the mobility of amorphadiene in the active site and promote more epoxidation. Incorporating

these mutations into $P450_{BM3}$ together with R47L/Y51F couplet enabled the *in vivo* selective oxidation of amorphadiene to the epoxide at titres of 250 mg/L in *E. coli* (Figure 1.22).

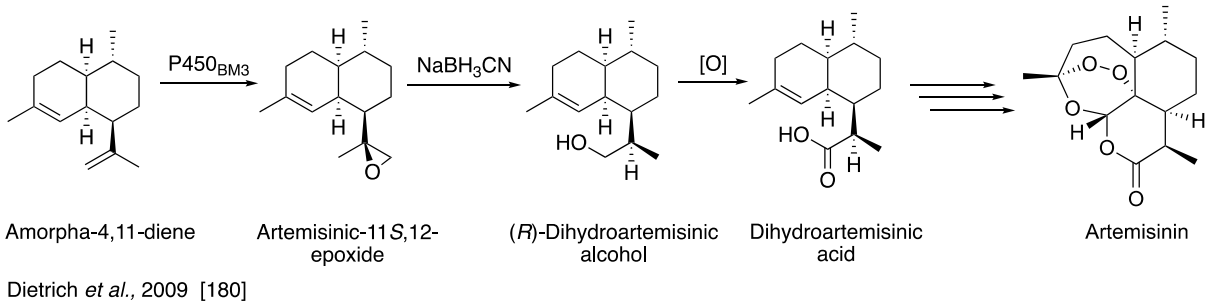


Figure 1.22 Semi-synthesis strategies for production of artemisinin.[180]

The sesquiterpenoids are a large diverse class of naturally occurring compounds with biological or medicinal properties or found as flavourings and aroma compounds. The oxidation sites of oxygenated sesquiterpenes are often chemically inert, which make these compounds interesting and challenging targets for enzymatic reaction.[175,182,183] (+)-Nootkatone, a sesquiterpenoid isolated from grapefruit, is a highly valued aroma compound. The extraction of (+)-nootkatone from nature suffers from low yields and chemical synthesis requires environmentally harmful reagents and solvents. (+)-Valencene, found in orange oil, is the biological precursor to (+)-nootkatone. Chemical oxidation of (+)-valencene to (+)-nootkatone has been studied, ranging from hydroperoxide oxidation to transition metal catalysis.[184-188]

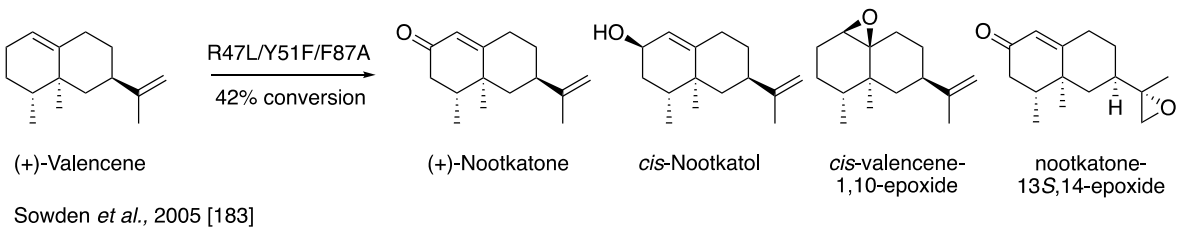


Figure 1.23 Oxidation products of valencene with $P450_{BM3}$ variant.[183]

Biological oxidation is an alternative possibility, but microbial oxidation to (+)-nootkatone commonly suffers from product inhibition and low yields. Wong and co-workers designed $P450_{BM3}$ mutants that exert higher activity for (+)-valencene oxidation (Figure 1.23).[183] The most successful mutant gave 42% conversion of (+)-valencene.

1.8 Thesis objectives

1.8.1 Chemoenzymatic semi-synthesis of aroma compounds

Terpenes and terpenoids have huge potential in the fine chemical industry and in drug discovery. Norisoprenoids are classified as one class of terpenes and possess biological activity and are valuable products in the flavour, fragrance and pharmaceutical industries. Fine chemicals such as α-irone, β-damascenone, tabanone isomers and rose oxide possess very different aromas which have found varied applications. Due to their low concentration in nature, chemical or enzymatic synthesis would be attractive alternatives for large scale production. α-Ionone, β-ionone, β-damascone, and (–)-β-citronellol are readily available and inexpensive, and thus, would be ideal starting materials for the synthesis of α-irone, β-damascenone, tabanone isomers and rose oxide (Figure 1.24).

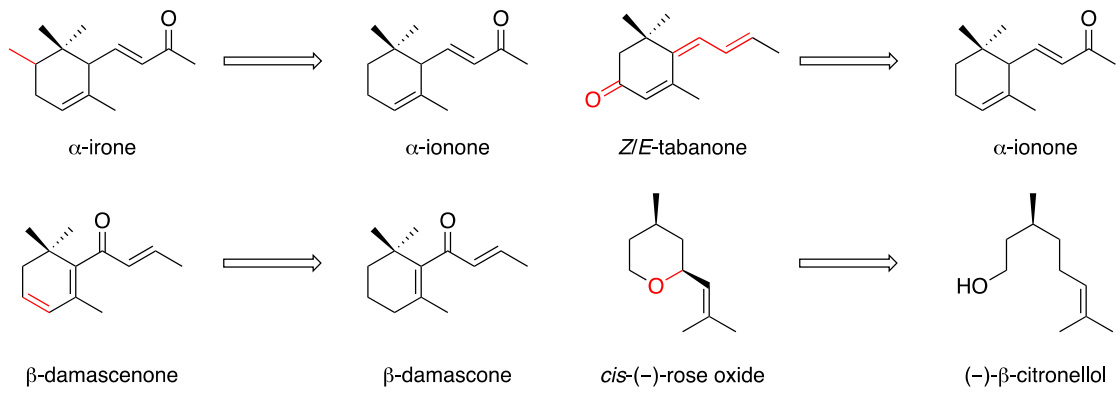


Figure 1.24 The proposed semi-synthesis of valuable norisoprenoids and isoprenoids compounds in this project.

Chapter 3 describes the development of potential synthesis routes for these compounds via late-stage C–H bond oxidation of the precursors. The initial screening and variant selection for the production of precursor alcohols with high selectivity and turnover number, followed by chemical functionalisation and optimisation to achieve high product yield are reported. Moreover, the engineered mutants shifted the oxidation regioselectivity away from allylic alcohol and epoxide to diverse oxidation products, such as tertiary alcohols, and aldehydes. The roles of selected key residues on the activity and selectivity were also studied in this work.

1.8.2 Oxidation of steroids

Previous work in the Wong group focused on generating a small but active library of mutants in order to functionalise diverse range of natural and non-natural substrates. After screening with the initial library, effective mutations were combined to generate sub-libraries (rational design) for specific classes of substrates, such as steroid molecules. Previously, high regio- and stereo-selectivity for 2β- and 15β-hydroxy-testosterone were obtained with the initial library; however, oxidation at other sites was not observed with any significant selectivity. Oxidation of other steroid molecules such as androstenedione and dehydroepiandrosterone, especially at remote, non-activated sites remains challenging for all P450s. The second aim in this thesis was to demonstrate rational protein design of $P450_{BM3}$ variants for the C–H functionalisation of these steroidal compounds at previously inaccessible regions. The roles of the key residues were explored with the help of molecular dynamic simulation and docking studies. Chapter 4 describes the initial screening of androstenedione, dehydroepiandrosterone, and testosterone followed by a nature-inspired design of second-generation library that altered steroid binding orientation and gave a broader range of products (Figure 1.25).

Progesterone oxidation has been thoroughly studied by various recombinant P450s, with high regio- and stereo-selectivity found whereas pregnenolone oxidation remains challenging. Also, it would be interesting to compare the activity for progesterone and pregnenolone oxidation due to the different structures of the A and B ring, likewise for androstenedione and dehydroepiandrosterone. Similarly, androsterone and epi-androsterone are a pair of diastereomers that differ by the C3 stereochemistry. The oxidation of these molecules is more challenging due to the saturated C–H bonds and lack of directing functional groups. Therefore, these molecules are interesting and challenging candidates for testing the activity of $P450_{BM3}$ variants.

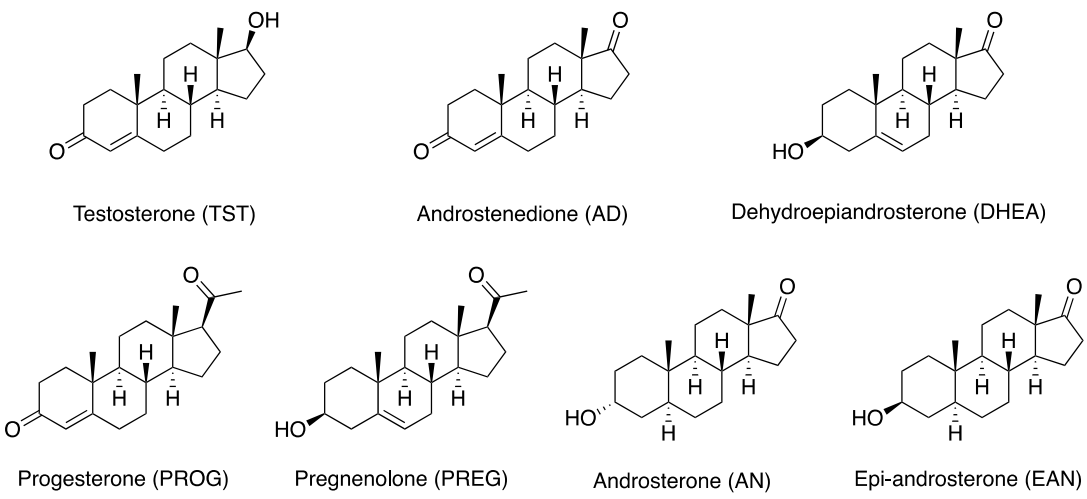


Figure 1.25 Steroidal molecules to be studied in this project.

Chapter 5 describes the oxidation of progesterone, pregnenolone, androsterone, and epi-androsterone by engineered $P450_{BM3}$ variants (Figure 1.25). Oxidation products favoured by different effective mutations were compared and molecular docking study was performed to provide insights on substrate binding.

Chapter 2

2 Materials and methods

2.1 Materials, chemicals and equipment

General reagents, chemicals and HPLC grade solvents were purchased from Alfa-Aesar, Fisher Scientific, Fluorochem, and Sigma-Aldrich, UK. Norisoprenoids and steroidal substrates were purchased from Sigma-Aldrich, Acros, and Fluka, UK. Media components, kanamycin and IPTG were supplied by Melford Laboratories, UK. Hen egg white lysozyme was from Sigma-Aldrich. $NADP^+$ was from Prozomix, UK; glucose dehydrogenase (GDH) was from Codexis, California, USA. Oligonucleotides were supplied by Eurofins Genetic Service, UK.

Sonication was carried out using a Misonix 3000 sonicator and Fisher Scientific FB-505 ultrasonic processor. Mutagenesis and PCR reactions were carried out in a Techne thermocycler or a Cetus thermocycler from Perkin-Elmer. UV-vis spectra were acquired on a Varian CARY50 spectrophotometer at 30°C using 1 cm pathlength quartz cuvettes. Oxidation products were purified by flash silica gel column chromatography using Geduran Silica 60, 40–63 μm (Fisher Scientific, UK) or preparative scale supercritical fluid chromatography (SFC) on a Waters Prep-15 instrument using a 2-ethylpyridine (2-EP) column (5 μm, 10 mm × 150 mm) and eluting with methanol as a modifier of the liquid CO_2 mobile phase. Analytical thin-layer chromatography was performed on petroleum ether/ethyl acetate solvent system and bands were either visualised under ultraviolet light or with permanganate or phosphomolybdic acid staining reagents. ^1H, ^{13}C, COSY, HSQC, HMBC, NOESY, and 1D nOe NMR spectra were acquired on Bruker AVIII-500 (500/125 MHz), Bruker AVIII-400 (400/100 MHz), or Bruker AV-400 (400/100 MHz) spectrometers. High resolution mass spectra (HRMS) were obtained on a Bruker microTOF spectrometer.

Table 2.1 Growth media and buffer components.

Media/buffer	Components, per litre water
LB	10 g tryptone, 5 g yeast extract, 10 g NaCl
LB_{agar}	10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g bacto-agar
2YT	16 g tryptone, 10 g yeast extract, 5 g NaCl
$FB_{glycerol}$	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 (after autoclaving, add 2.0 mL 1 M $MgSO_4.7H_2O$, 3.0 mL trace elements solution, 1.0 mL 10% *w*/*v* thiamine.HCl)
Trace elements	20.1 g Na_2EDTA, 0.74 g $CaCl_2.H_2O$, 0.25 g $CoCl_2.6H_2O$, 0.10 g $CuSO_4.5H_2O$, 16.7 g $FeCl_3.6H_2O$, 0.132 g $MnSO_4.4H_2O$, 0.18 g $ZnSO_4.7H_2O$
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)
Innoculation solution	15 mg kanamycin, 250 mg $MgSO_4 \cdot 7H_2O$, 50 mg thiamine·HCl, add trace elements soluion to 1.5 mL
Induction solution/40 mL	380 mg IPTG, 672 mg δ-ALA·HCl, 112 mg $FeSO_4 \cdot 7H_2O$, 600 mg kanamycin in 40 ml H_2O
Buffer G	4.9 g K_2HPO_4, 1.6 g KH_2PO_4, 0.15 g DTT
50mM Tris-HCl	50 mL 1 M Tris-HCl, 950 mL H_2O
200 mM KPi	2.4 g KH_2PO_4, 31.7 g K_2HPO_4, 1000 mL H_2O

Water used for bacterial growth and buffers was from a Milipore ELIX-5 system (resistivity ≥8 MΩ cm^{-1}). The components of media are shown in Table 2.1. All media were sterilised prior to use. Kanamycin (30 mg L^{-1}) was added after the media had cooled to room temperature. The buffer pH was adjusted using either concentrated HCl (Tris buffer) or orthophosphoric acid (phosphate buffer).

2.2 DNA manipulations

2.2.1 PCR site-directed mutagenesis

Oligos used in this project are listed in Appendix 2, Table 1.4. Solutions of both the forward and reverse primers, each at 10 μM concentration, were prepared in ultrapure water. To a 200-μL autoclaved PCR microcentrifuge tube the following components were added: 29 μL water, 5 μL dNTP, 5 μL 10× KOD polymerase buffer, 1−2 μL parent DNA, 4 μL $MgSO_4$, 2 μL DMSO, 1.5 μL forward primer, 1.5 μL reverse primer and 1 μL polymerase to make a total volume of 50 μL. After being gently pipette-mixed and centrifuged for 5 s at 2,900 *g*, the microcentrifuge tube was placed in the thermocycler and the following temperature program was run: 94 °C for 20 s for strand separation; 55−65 °C for 20 s for annealing; 72 °C for 160 s for extension; cycle repeated 30 times. The temperature was then kept at 72 °C for 10 min before being reduced to and maintained at 4 °C. The PCR reactions were then analysed by agarose gel electrophoresis.

2.2.2 Agarose gel electrophoresis

A suspension of agarose powder (0.5 g) in 50 mL 1× TAE (Tris-acetate-EDTA) buffer in a conical flask was heated to form a clear solution in a microwave (2 × 30 s). RedSafe™ staining solution (5 μL) was added into the solution which was poured into the gel tray with a comb to form sample wells and left to set. The gel was then placed in the electrophoresis apparatus and extra 1× TAE buffer was poured into the tray until the gel was immersed in the buffer. 2 μL of loading dye was mixed with 10 μL of the PCR mixture and then carefully pipetted into a well in the gel; 5 μL of a DNA size ladder was added to an adjacent well in the gel to provide a size scale. A voltage difference of 100 V was applied across the gel for 20 min and a long wavelength UV lamp was used to irradiate the gel and the bands were compared with the DNA

ladder standard. A strong band with the same mobility as a linearised 8500 bp fragment suggested a successful PCR step. 1 μL of DpnI, was then added to digest the template DNA for 2 h at 37 °C.

2.2.3 Post-mutagenesis transformation

Escherichia coli (*E. coli*) DH5α competent cells were stored as 120-μL aliquots in microcentrifuge tubes at −80 °C. One aliquot of cells was defrosted on ice, 10 μL of the DpnI–digested PCR mixture was added into 40 μL of competent cells which were incubated for 45 min on ice. The cells were heat-shocked at 42 °C for 45 s and then placed back on ice for 2 min. 600 μL of filter-sterilised SOC media was added to the mixture which was then incubated for 2 h in a shaker at 180 rpm, 37 °C. The cells were spread onto a LB_{KAN} agar plate and placed in a 37 °C oven to grow overnight.

2.2.4 DNA purification and transformation

One single colony was picked from the plate and inoculated into 5 mL LB_{KAN} media in a 15-mL centrifuge tube at 180 rpm, 37 °C for 16 h. The culture was centrifuged at 9,000 *g* and 4 °C for 10 min. Plasmid DNA was then purified using a GeneJET plasmid miniprep kit from ThermoFisher Scientific. The purified DNA was recovered in 50 μL of deionised H_2O. The DNA concentration (μg/μL) was determined from 5× A_{260} of a 1:100 diluted sample (5 μL solution in 495 μL H_2O) and the purity was assessed by the ratio A_{260}/A_{280}. If the purity ratio was between 1.5 and 1.8 and concentration greater than 0.1 μg μL^{-1}, then 15 μL of the DNA stock was sent for DNA sequencing by Geneservice, UK using two primers, T7F and BM3FP2 (see Table 2.2), to ensure the presence of the desired mutation(s). If the correct sequence was obtained, the plasmid DNA (0.25 μL) was transformed into 40 μL of competent *E. coli* BL21 (DE3) as described in section 2.2.3.

Table 2.2 Nucleotide sequences of the T7F and BM3FP2 primers.

Primers	Sequence
T7F	TAATACGACTCACTATAGG
BM3FP2	CTGCAGCGAGCAAATCCAGACGAC

2.3 Protein expression and purification

One colony of transformed *E. coli* BL21 (DE3) from a LB_{KAN} plate was inoculated into 500 mL FB_{gly} media with addition of 1.5 mL of inoculation solution (Table 2.1) containing 30 mg/L kanamycin in a 2-litre flask and shaken at 160 rpm for 24 h at 37 °C. A 5 mL aliquot of the culture was mixed with 5 mL sterile 50% glycerol in a 15-mL centrifuge tube to prepare a cell bank which was then stored as 1-mL aliquots at −80 °C. The incubator temperature for the overnight culture was then lowered to 20 °C and the culture left to shake for 1 h. Protein production was induced by adding 1 mL of induction solution (Table 2.1) into each 2 L flask. The culture was grown for a further 48−72 h, shaking at 120 rpm, 20 °C. Cells were harvested by centrifugation at 9000 *g*, 4 °C for 10 min.

The cell pellet was re-suspended in 30 mL buffer P per 500 mL culture, transferred to a beaker on ice and the cells were lysed by sonication with a 5-min programme of 15-s pulses and 30-s cooling, at 45% amplitude. The lysate was centrifuged for 15 min at 9,000 *g*, 4 °C to pellet the cell debris. The P450 enzyme was recovered from the supernatant by ammonium sulfate fractionation. The amount of ammonium sulphate required to take the concentration to 40% saturation was calculated (http://www.encorbio.com/protocols/AM-SO4.htm) and added as the solid to the solution at 4 °C and the mixture stirred until all the solids had dissolved. The solution was centrifuged at 9,000 *g*, 4 °C for 20 min. To the retained supernatant was added the amount of ammonium sulfate to take the solution from 40% to 60% saturation to precipitate the

P450 enzyme. The pellet was retained after centrifugation at 9,000 *g* for 30 min at 4 °C and then dissolved in 5 ml buffer P. The protein was aliquoted into 1.5-mL tubes and stored at −20 °C.

2.4 Quantification of cytochrome P450

Before the P450 enzyme was stored at −20 °C, 50 μL of enzyme was added into 950 μL of buffer P and the UV-Vis spectrum (200–500 nm) was recorded. A few crystals of sodium dithionite were added to reduce the protein and the UV-Vis spectrum was recorded of the reduced protein which was then used to obtain the baseline. 10~15 bubbles of CO were gently bubbled through the reduced protein; and the difference spectrum was recorded. More CO was bubbled through the solution and the spectrum recorded or until the characteristic 450 nm peak height was maximised. The concentration of the protein was determined from:

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

where $\varepsilon_{450–490}$ = 91,000 M^{-1} cm^{-1} is the extinction coefficient.[44]

2.5 Compound screening and characterisation

2.5.1 Activity screening in 24-well plates

β-Damascone (**1**), α-ionone (**5**), α-ionol (**7**), ms-citronellol (**14**), ts-citronellol (**15**), and citronellol (**16**) were screened against the $P450_{BM3}$ library at a ratio of 2500:1 of substrate to enzyme (Table 2.3). Norisoprenoid substrates were dissolved in MeOH and added as a stock, typically at 200 mM concentration. 100 μL of each enzyme, previously diluted to 10 μM, was added into the 24-well plates. The screening was carried out in 200 mM phosphate buffer (pH

7.9) in a volume of 0.5 mL in these 24-well plates. In this reaction system, the final concentration of $P450_{BM3}$ variants was 2 μM and the substrates was at 5 mM. GDH (20 U/mL) and glucose (100 mM) was used to regenerate the NADPH cofactor. $NADP^+$ (40 μM) was added to initiate the reaction. Methyl-β-cyclodextrin in 200 mM stock was added to solubilise substrates at such high concentration. Screening plates were shaken at 120 rpm for 16 h at 20 ºC and then extracted with 300 μL of ethyl acetate. After centrifugation at 14300 *g* to separate the phases, the organic extracts were analysed by GC; the oven temperature was held at 140 ºC for 1 min, then raised at 15 ºC/min to 240 ºC and held at this temperature for 1 min.

Table 2.3 Reaction components for screening of β-damascone (**1**), α-ionone (**5**), α-ionol (**7**), ms-citronellol (**14**), ts-citronellol (**15**) and citronellol (**16**).

Substance	**Stock concentration**	**Volume/μL**	**Final concentration**
Glucose	1 M	50	100 mM
Substrate	200 mM	12.5	5 mM
GDH	200 U/mL	10	2 U/mL
$NADP^+$	1 M	10	80 μM
Enzyme	10 μM	100	2 μM
Methyl-cyclodextrin	200 mM	12.5	5 mM
200 mM KPi buffer	200 mM	305	N/A

Androstenedione (AD, **19**), dehydroepiandrosterone (DHEA, **20**), testosterone (TST, **21**) were screened against the $P450_{BM3}$ library at a ratio of 1000:1 of substrate to enzyme (Table 2.4). These steroidal substrates were dissolved in MeOH and added as a stock, typically at 100 mM concentration in the screening reaction. GDH (20 U/mL) and glucose (100 mM) was used to regenerate the NADPH cofactor. $NADP^+$ (40 μM) was added to initiate the reaction. Screening plates were shaken at 120 rpm for 16 h at 20 ºC and then extracted with 300 μL of ethyl acetate. After centrifugation at 14300 *g* to separate the phases, the organic extracts were analysed by GC: the oven temperature was held at 220 ºC for 1 min, then raised at 20 ºC/min to 300 ºC and held at this temperature for 3.5 min.

Table 2.4 Reaction components for screening of androstenedione (AD, **19**), dehydroepiandrosterone (DHEA, **20**), testosterone (TST, **21**).

Substance	Stock concentration	Volume/μL	Final concentration
Glucose	1 M	50	100 mM
Substrate	100 mM	10	2 mM
GDH	200 U/mL	10	2 U/mL
$NADP^+$	1 M	10	80 μM
Enzyme	10 μM	100	2 μM
200 mM KPi buffer	200 mM	320	N/A

Androsterone (AN, **22**), epi-androsterone (EAN, **23**), progesterone (PROG, **24**) and pregnenolone (PREG, **25**) were screened against the $P450_{BM3}$ library at a ratio of 500:1 of substrate to enzyme due to their lower solubility in aqueous reactions (Table 2.5). 200 mM of methyl-β-cyclodextrin was added in the reaction to solubilise these steroidal substrates. AN, EAN and PROG were dissolved in MeOH while PREG, the least soluble substrate, was dissolved in DMSO; these steroids were added as a stock at 100 mM concentration in the screening reaction.

Table 2.5 Reaction components for screening of androsterone (AN, **22**), epi-androsterone (EAN, **23**), progesterone (PROG, **24**) and pregnenolone (PREG, **25**)

Substance	Stock concentration	Volume/μL	Final concentration
Glucose	1 M	50	100 mM
Substrate	100 mM	5	1 mM
GDH	200 U/mL	10	2 U/mL
$NADP^+$	1 M	10	80 μM
Enzyme	10 μM	100	2 μM
Methyl-β-cyclodextrin	200 mM	5	5 mM
200 mM KPi buffer	200 mM	320	N/A

After shaking for 16 h, the reaction mixtures were transferred to a microcentrifuge tube and extracted with 300 μL of ethyl acetate by vortexing for 1 min, and centrifuged at 14,800 *g* for 1 min, which were analysed by GC with program stated above.

2.5.2 Preparative scale reactions

Preparative scale reactions (20 mL − 200 mL) were carried out overnight with the same concentrations as in the initial screening. Progress of the reactions was monitored by GC analysis of organic extracts of 0.5 mL aliquots. Once no further product formation was detected, the reaction mixture was extracted with 30 mL of ethyl acetate at least three times. The combined extracts were dried with Na_2SO_4 before being filtered under suction with a sinter glass funnel and finally evaporated on a rotary evaporator.

2.5.3 Product purification and characterisation

The residue after solvent evaporation was re-dissolved in the ethyl acetate and transferred to a glass vial. The mixture was then tested with thin-layer chromatography (TLC) to select the solvent or combination of solvents that could separate the substrate and products. Flash column chromatography purification was carried out with Geduran Silica 60, 40−63 μm. Fractions were analysed by TLC, staining with a solution of potassium permanganate or phosphomolybdic acid; selected fractions were then analysed by GC. Fractions containing the same compounds were combined and solvents were removed with a rotary evaporator. The purified compound was characterised by NMR and MS.

2.6 Molecular dynamics simulations and steroid docking

MD simulations were performed on wild type CYP102A1 and variants that showed high selectivity for the observed steroid oxidation products. All structures were relaxed via MD simulations performed in GROMACS 2018.6.[189] Structures of variants were prepared by addition of mutations to the WT CYP102A1 crystal structure (PDB code: 1jpz)[128] via Pymol. *N*-Palmitoylglycine, which was crystallised within the structure, was removed. The protein was

prepared using the Amber 99SB*-ILDN force field[190] with TIP3P water.[191] The haem moiety was simulated in the Compound I state, using parameters from Shahrokh, *et al.*[192] The protein was placed into the centre of a truncated octahedral box, with a minimum distance of 10 nm to any box edge. The protein was then solvated with *ca.* 17500 water molecules and the net charge neutralised by addition of Na^+ ions. Steepest-descent energy minimisation was performed until the maximum force was <500 kJ mol^{-1} nm^{-2}. Periodic boundary conditions were then used to model the system, with electrostatic interactions treated by the particle mesh Ewald method[193] and bond lengths involving H-atoms constrained with the LINCS algorithm.[194] Short-range non-bonded interactions were calculated within a 1 nm cut-off and a timestep of 2 fs was used. For equilibration steps, all protein backbone Cα atoms were restrained with a positional restraint force constant of 1000 kJ mol^{-1} nm^{-2}. A NVT equilibration step was performed at 298 K for 100 ps, using the modified Berendsen thermostat.[195] This was followed by an NPT equilibration step for 1250 ps utilising the Berendsen barostat with a time constant of 1 ps.[196] All positional restraints were then removed and 100 ns of production MD was performed in triplicate using the Parrinello-Rahman barostat[197] with a time constant of 5 ps. Structures were recorded every 10 ps. Stability of the simulations was confirmed by monitoring the RMSD of the backbone Cα atoms before the individual trajectories were clustered using the Daura algorithm[198] with a cut-off of 1.2 nm. The three most populated clusters were used as receptor structures for docking studies, which were performed in Autodock Vina.[199] All water molecules were removed from the structure prior to the docking calculation and both receptor and substrate were treated as rigid entities. The docking site was defined as a 30 × 30 × 30 Å box, centred on the ferryl oxygen, and poses were ranked using the Autodock Vina scoring function.

Chapter 3

3 Chemoenzymatic routes to norisoprenoid and isoprenoid aroma compounds

3.1 Introduction

Terpenes are a diverse group of organic compounds that are hydrocarbons mainly produced by plants and insects. Terpenoids are modified terpenes with different functional groups that mostly contain oxygen. Terpenes and terpenoids with strong aromas and flavour are constituents of plant essential oils that have been widely used in perfumery and food industries. Norisoprenoids are classified as one class of terpenes and possess desirable fragrance and medicinal properties (Figure 3.1). For instance, besides boosting the fruity aroma, β-damascone also shows a variety of biological activities such as cancer chemopreventive potential and repellent properties towards specific species of mosquitoes. α-Irone, found in iris oil, has a sweet floral, iris and woody odor and is one of the major odorants in Chanel No.19, the last fragrance to be formulated by Coco Chanel. β-Damascenone, extracted from Bulgarian roses, is a minor component of rose oil but a major contributor to the aroma with a very low odour threshold of 0.002 ppb, and is widely used in food and wines. Tabanone is found in trace quantities in nature and imparts a tobacco aroma that has found application in electronic tobacco liquids.

These products are valuable but difficult to extract from nature due to the low concentration and high expense. The most common norisoprenoids are the ionones and damascones that are readily available and inexpensive, therefore the usage of these compounds as starting material to synthesise valuable products would be plausible. The subject of this chapter is to explore and develop chemoenzymatic routes to β-damascenone, tabanone and rose oxide from different precursors with the use of P450s to shorten the synthesis pathways.

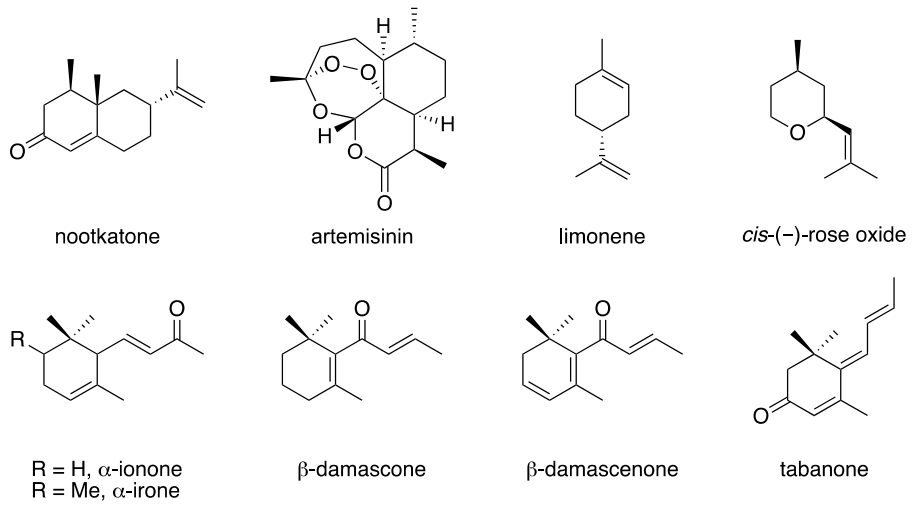


Figure 3.1 A selection of terpenes, terpenoids and norisoprenoids compounds.

3.2 β-Damascenone synthesis

β-Damascenone, first isolated from the Bulgarian rose, has an extremely low odour threshold of 0.002 ppb. It belongs to a family of chemicals known as the rose ketones which also include damascones and ionones, the components of variety of essential oils. Rose ketones have a flowery and sweet aroma, especially β-damascenone which has been widely used in perfumery and foods. The demands for β-damascenone is growing yearly due to the fast-developing cosmetics and perfumery industries. In nature, β-damascenone is biosynthesised *via* the degradation of carotenoids which are converted to neoxanthin, the proposed precursor to β-damascenone, by various oxidases.[200] The synthesis of β-damascenone from 9′-*cis*-neoxanthin in a model system by peroxyacid oxidation and two-phase thermal degradation without the involvement of enzymatic activity is shown in Figure 3.2.[201,202] Oxidation of neoxanthin by hydroperoxide leads to the formation of an epoxide **a1**, which is hydrolysed to the vicinal diol **a2**. Heating **a2** above 90 °C leads to thermolysis of the C9–C10 bond with subsequent disproportionation to yield the allenic triol **a3**, which is converted to β-damascenone *via* a

Meyer-Schuster rearrangement under mildly acidic conditions. Skouroumounis *et al.* suggested that compound **a3** is a key intermediate in the formation of β-damascenone and the 3-hydroxy-β-damascone is the major precursor of β-damascenone (5% yield).[203]

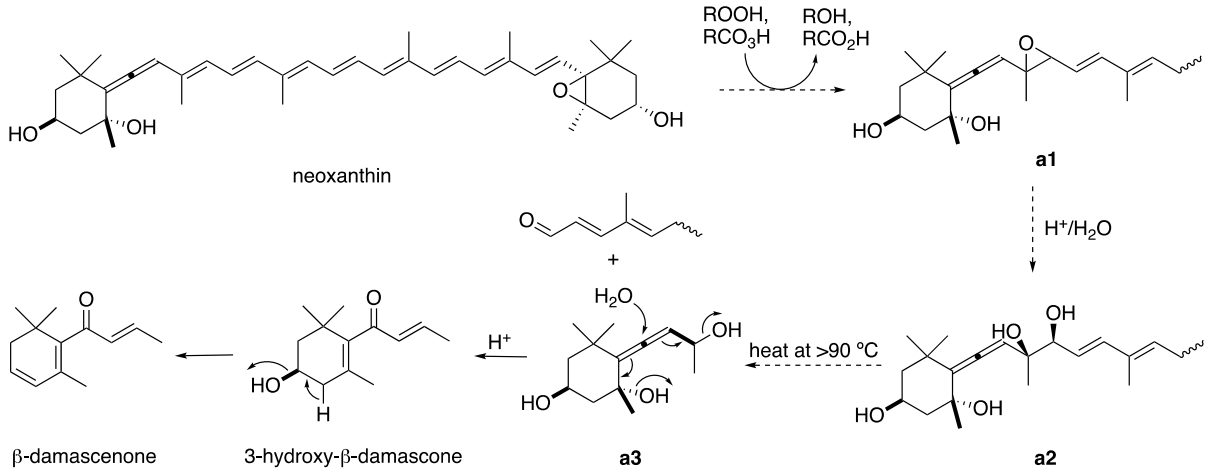


Figure 3.2 Hypothetical biosynthesis of β-damascenone from neoxanthin.

Because of its limited availability from natural sources, numerous chemical synthesis of β-damascenone has been developed. Two examples are shown in Figure 3.3. Route A shows the final step of β-damascenone synthesis starting from β-ionone, which undergoes a ketone shift via an allylsilane derivative to give β-damascone. Allylic bromination at C4 followed by dehydrohalogenation under vacuum at 80 °C in the presence of DABCO/DMAP gives β-damascenone.[204] Route B utilises cobalt-catalysed oxidative esterification of the allylic C–H bond at higher temperature, which once again indicates the difficulty of C–H bond activation, followed by acid-catalysed elimination of benzoic acid.[205]

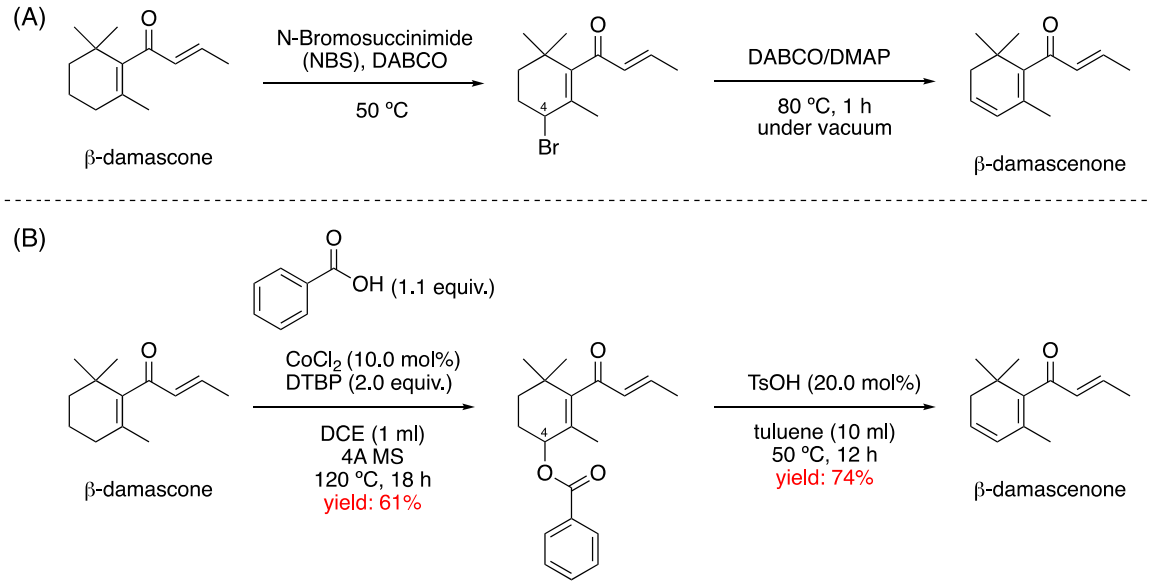


Figure 3.3 Two synthetic routes of β-damascenone from β-damascone.

Biotransformation routes with P450s provides a greener method for late-stage C–H activation under milder conditions to generate "natural grade" compounds. The proposed route involves formation of the precursor, 4-hydroxy-β-damascone, by utilising $P450_{BM3}$ variants followed by acid-catalysed dehydration (Figure 3.4).

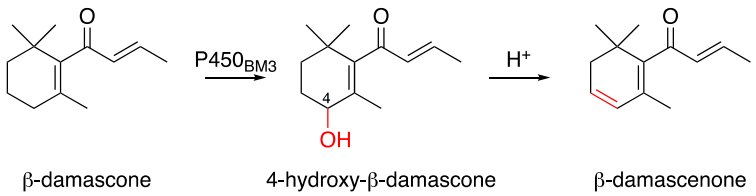


Figure 3.4 Proposed chemoenzymatic route to β-damascenone from β-damascone by $P450_{BM3}$.

3.2.1 Oxidation of β-damascone by $P450_{BM3}$

A library of 96 CYP102A1 variants was screened for the oxidation of β-damascone (5 mM substrate and 2 μM enzyme). Full activity and selectivity data are given in Appendix 3, Table 1.5. β-Damascone (**1**) was a good substrate for the variant library, as 66 out of the 96 variants

showed >50% conversion (total turnover number, TTN >1250) to a collection of 5 major products. These were purified and characterised as 2-hydroxy-β-damascone (**1a**), 3-hydroxy-β-damascone (**1b**),[206] 4-hydroxy-β-damascone (**1c**),[206] 4-oxo-β-damascone (**1d**),[207] and 10-hydroxy-β-damascone (**1e**) (Figure 3.5). The NMR data for products **1b**, **1c** and **1d** (Appendix 3) are consistent with literature. Products **1a** and **1e** are new metabolites of β-damascone; their full characterisation data are collected in Appendix 3.

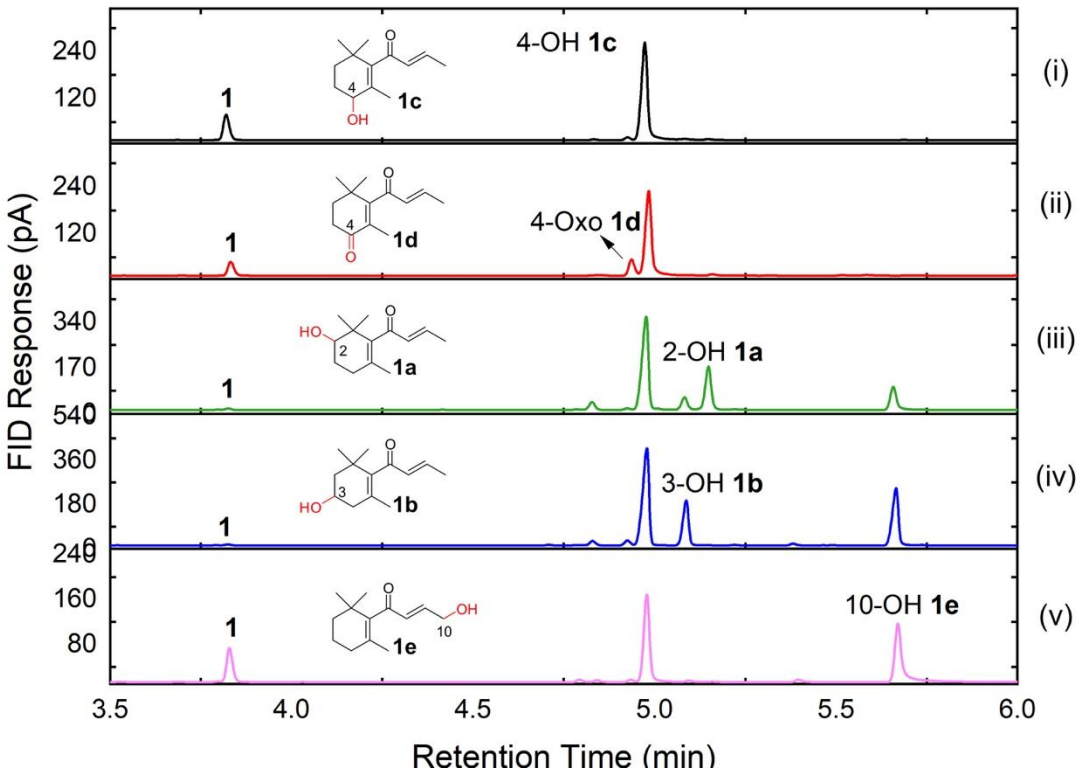


Figure 3.5 Gas chromatography analysis of selected variants giving β-damascone oxidation products **1a**–**1e**: (i) 97% of 4-hydroxy-β-damascone **1c** by GV/A184I/A264G; (ii) 13% of 4-oxo-β-damascone **1d** by R19/F87A/A184I/A264G; (iii) 21% of 2-hydroxy-β-damascone **1a** by R19/F87A/A328F; (iv) 34% of 3-hydroxy-β-damascone **1b** by R19/F87A/A328L; (v) 46% of 10-hydroxy-β-damascone **1e** by RT2/S72G/A330W/L437LA.

The MS data of 10-hydroxy-β-damascone (**1e**) showed that it was a mono-oxygenation product (M+16). The allylic C10 methyl resonance of β-damascone was missing from the ^{1}H NMR spectrum of **1e**, indicating that this methyl group had been oxidised. This was consistent with the two-proton allylic alcohol resonance at 4.38 ppm. Both the resonances at 6.78 and 6.37 ppm

that corresponded to H9 and H8 changed from doublets of quartets in β-damascone to doublets of triplets, which was entirely consistent with C10 hydroxylation (Figure 3.6).

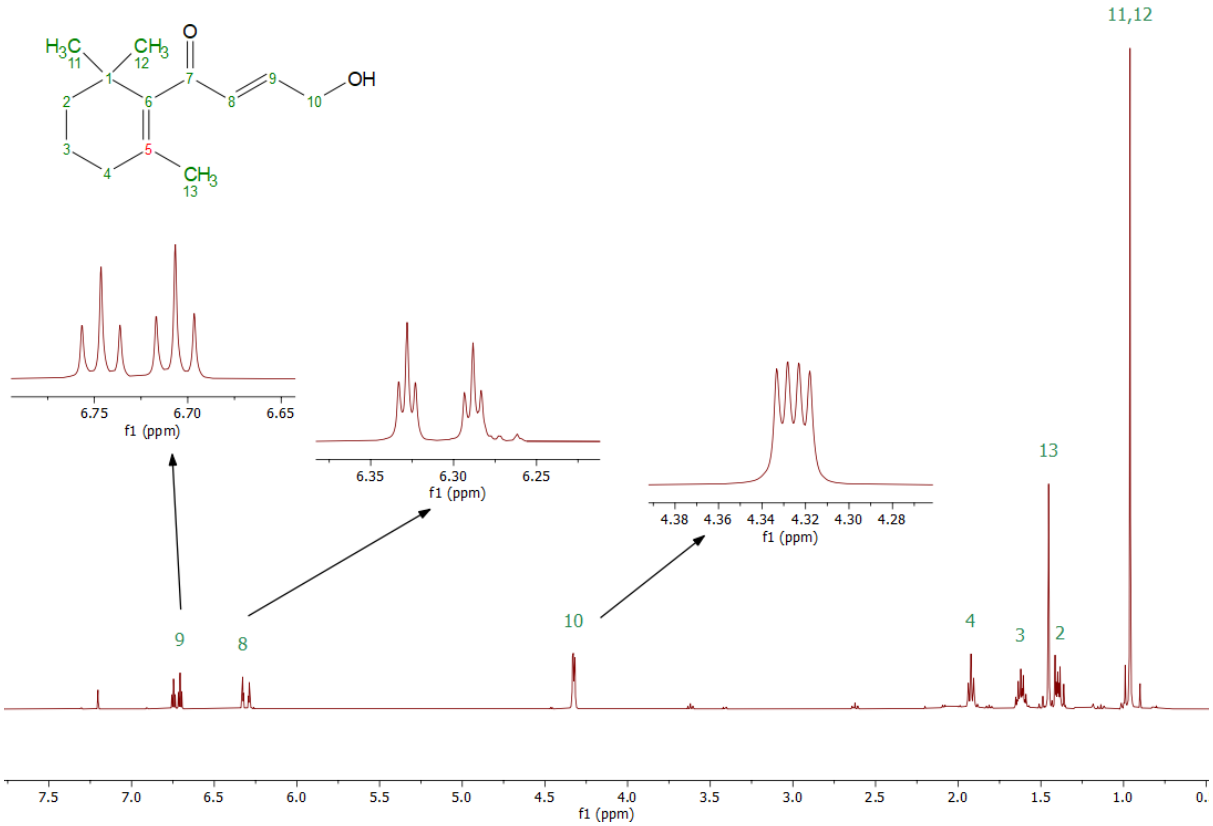


Figure 3.6 The ^{1}H NMR spectrum of a new metabolite of β-damascone, 10-hydroxy-β-damascone (**1e**) in $CDCl_3$.

2-Hydroxy-β-damascone (**1a**) could not be isolated free of **1e** by flash silica chromatography (Figure 3.7). The MS data (M+16) of the mixture of **1a** and **1e** showed that both were mono-hydroxylation products. In the COSY and HSQC spectra, the doublet of doublet resonance at 3.57 ppm only showed coupling to one pair of CH_2 group, indicating the oxidation was at C2.

Of the library of 96 variants, 82 gave 4-hydroxy-β-damascone (**1c**) as the major product (selectivity >50%), probably due to the allylic nature of C4. Variant R47L/Y51F/I263R (95% **1c**, 91% conversion) showed the highest turnover number (TON = 2160, Entry 7).

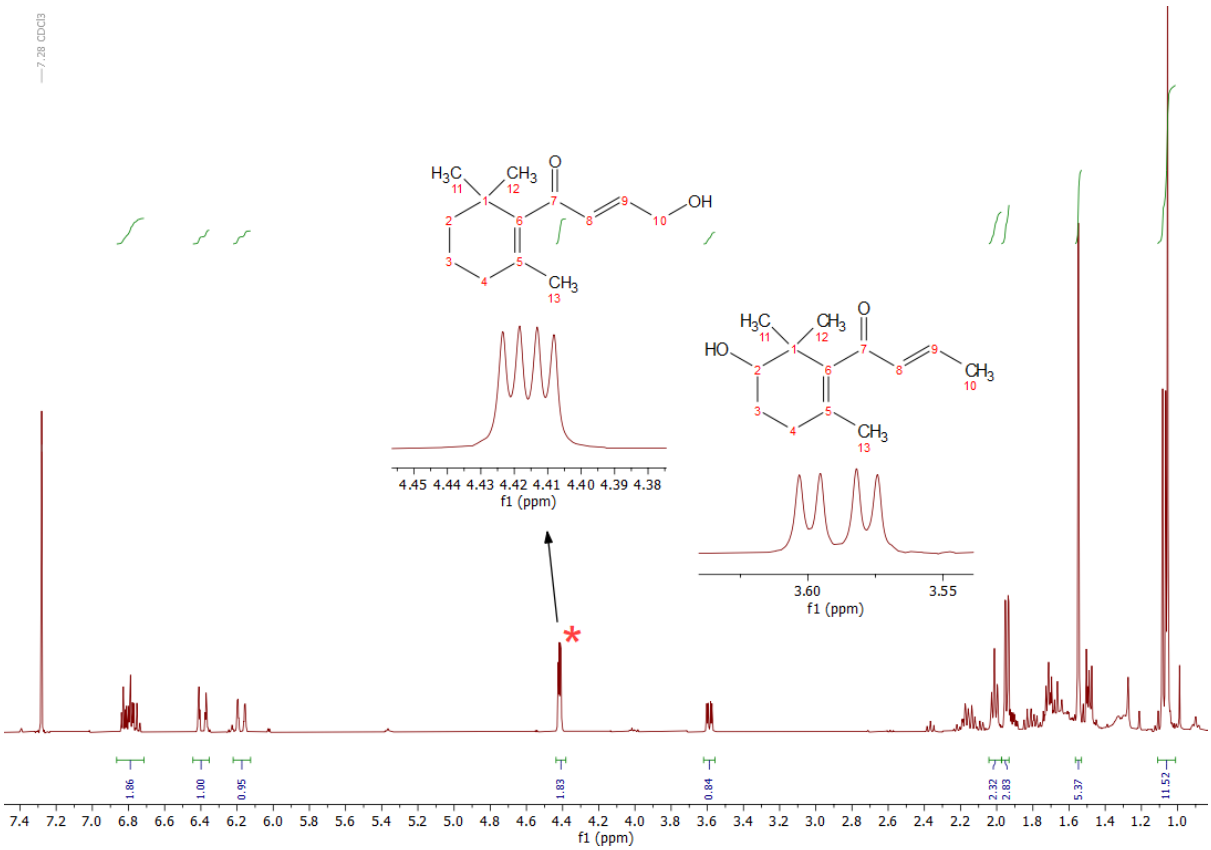
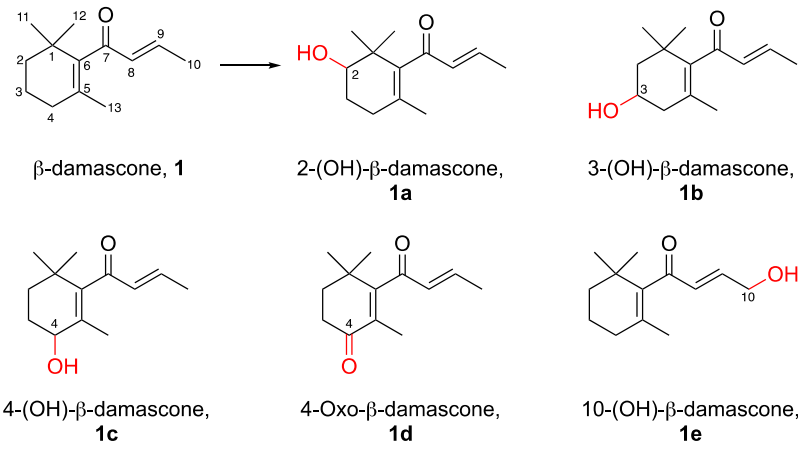


Figure 3.7 ^{1}H NMR spectrum of mixture of new metabolites, 2-hydroxy- **1a** and 10-hydroxy-β-damascone **1e** (red star).

Further oxidation of **1c** to the ketone **1d** was promoted by variants containing the F87A mutation. For instance, R19/F87A/A184I gave 66% of **1c** and 15% of **1d**, and R19/F87A/A184I/A264G gave 87% of **1c** and 13% of **1d** (Entries 11 & 12), although full conversion to **1d** was not observed.

Addition of bulky substitutions at A328 to F87A-based variants shifted oxidation from C4 towards C2 to form 2-hydroxy-β-damascone, **1a** (R19/F87A/A328I: 7%, R19/F87A/A328F: 21%, Entries 1 & 4). Introduction of V78I or S72H mutations to R19/F87A/A328I also slightly increased the proportion of **1a** to 13% (Entries 2 & 3). In sharp contrast, the A328L mutation promoted a different product, 3-hydroxy-β-damascone, **1b** (R19/F87A/A328L: 34%, Entry 5). The extra mutation E267F on variant R19/F87A/A328I also gave 32% **1b** (Entry 6).

Table 3.1 Substrate conversion and product selectivity of selected variants for the oxidation of β-damascone, **1**.[a]



Entry	Variants	1a	1b	1c	1d	1e	Others	Conv.	TON[b]
1	R19/F87A/A328I	**7%**	3%	77%			13%	85%	149
2	R19/F87A/A328I/S72H	**13%**	3%	79%	3%		2%	89%	290
3	R19/F87A/A328I/V78I	**14%**		67%			19%	36%	126
4	R19/F87A/A328F	**21%**	5%	50%		14%	10%	79%	415
5	R19/F87A/A328L	17%	**34%**	38%			10%	100%	850
6	R19/F87A/A328I/E267F	11%	**32%**	40%			17%	97%	776
7	R47L/Y51F/I263R			**95%**			5%	91%	2161
8	GV/A184I/A264G			**97%**			3%	81%	1965
9	K19/F87V/A264G			**98%**			2%	80%	1960
10	RP/HL/I263G/S72G			**98%**			2%	84%	2058
11	R19/F87A/A184I			66%	**15%**		19%	87%	327
12	R19/F87A/A184I/A264G			87%	**13%**			88%	286
13	RT2/S72H/A330W			88%		**10%**	2%	85%	213
14	RT2/S72G/A330W			87%		**12%**	1%	88%	264
15	RT2/S72Y/A330W			83%		**15%**	2%	90%	338
16	KU3/A330P			53%		**31%**	16%	40%	310
17	RT2/S72W/A330W			50%		**35%**	15%	45%	394
18	KU3/A330P/S72W			39%		**45%**	16%	43%	484
19	RT2/S72G/A330W/L437LA[c]			52%		**46%**	2%	83%	955

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 5 mM β-damascone, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold in each reaction. [c] L437LA denotes the insertion of an alanine residue after Leu437.

10-Hydroxy-β-damascone (**1e**) was favoured by retention of F87 and bulky substitutions at A330 (A330P and A330W). Comparison between variants RT2/S72(H,Y,G)/A330W and RT2/S72W/A330W showed that only the S72W mutation favoured the formation of **1e** (Entries

13–15 and 17). The importance of S72W was also demonstrated by A330P-based variants (Entries 16 & 18). The L437LA insertion mutation (insertion of an alanine residue after Leu437) improved the selectivity for **1e** (46%, TON = 960, Entry 19) when added to the RT2/S72G/A330W variant (12% **1e**).

3.2.2 Scalability of 4-hydroxy-β-damascone formation – Step 1

For large scale production of β-damascenone (**2**), the generation of the precursor, 4-hydroxy-β-damascone (**1c**), with high selectivity and turnover number was the crucial first step. The screening reactions were at 0.5 mL scale which contained β-damascone (**1**) and enzymes in a ratio of 2500:1 (5 mM of **1**). In order to demonstrate the scalability of β-damascone oxidation by $P450_{BM3}$, preparative scale reactions of **1** with R19/F87A/A328I were conducted at the higher ratios of 5000:1, 10000:1 and 20000:1 (Figure 3.8).

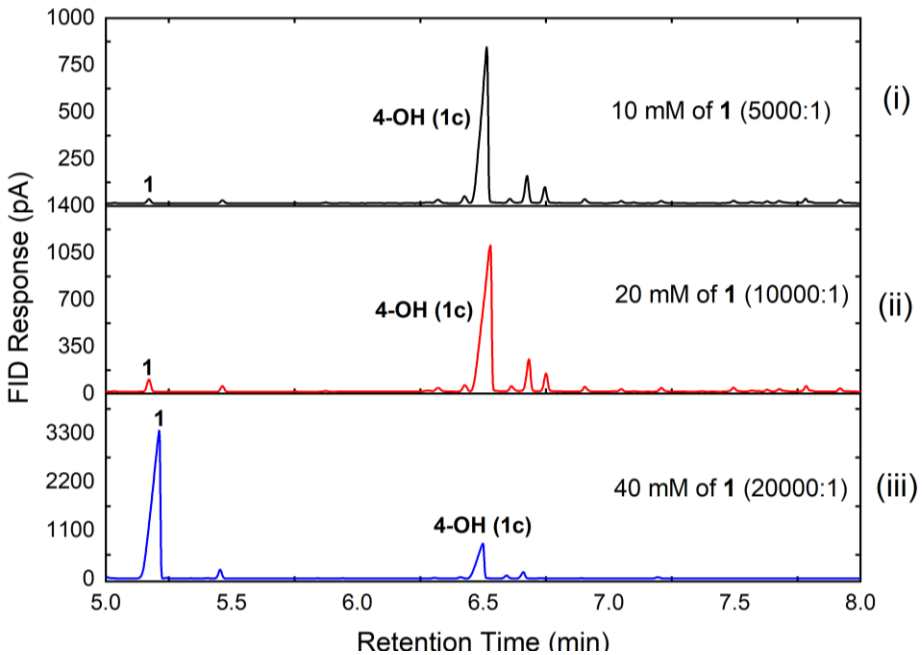


Figure 3.8 Preparative scale reactions of β-damascone (**1**) with R19/F87A/A328I in various conditions. (i) Reaction (0.96 g) was conducted at 500 mL scale in 200 mM phosphate, pH 8.0, containing 2 μM R19/F87A/A328I, 10 mM of **1**, 40 μM NADP+, 100 mM glucose, and 20 U/mL GDH for 16 h. (ii) Reaction (1.92 g) was conducted at 500 mL scale containing 2 μM R19/F87A/A328I and 20 mM of **1** for 16 h. (iii) Reaction (3.84 g) was conducted at 500 mL scale containing 2 μM R19/F87A/A328I and 40 mM of **1** for 16 h.

Gram-scale reactions with 10 mM (5000:1 ratio, 0.96 g) and 20 mM (10000:1 ratio, 1.92 g) of **1** achieved >95% conversion after 16 h, showing 77% and 75% selectivity for 4-hydroxy-β-damascone (**1c**). The reaction with 40 mM β-damascone (20000:1 ratio, 3.84 g) only gave 21% conversion after 16 h, suggesting the substrate was inhibitory to the enzyme at higher concentration, reducing both the rate of reaction and the TON. An alternative way to achieve 40 mM scale up reaction would be adding substrate **1** as 20 mM aliquot in the reaction for 8-16 h and then a second aliquot. Removal of products using selective resins before adding more aliquot of substrate may also help to increase the total reaction titre.[208]

3.2.3 Synthesis of β-damascenone from 4-hydroxy-β-damascone – Step 2

Following the gram-scale generation of 4-hydroxy-β-damascone (**1c**), the second step would be acid/base-catalysed elimination reaction of **1c**. Five synthesis routes for β-damascenone **2** with various agents and conditions were tested (Figure 3.9–3.13).

Route i: The first route was to convert the allylic alcohol in **1c** to a good leaving group such as TsO^- or MsO^-, followed by base treatment (Figure 3.9). However, the secondary alcohol on the ring was too hindered to be protected, therefore, direct protonation of **1c** was then attempted, on the expectation that the stabilised allyl cation would readily generate β-damascenone (**2**), driven by conjugation.

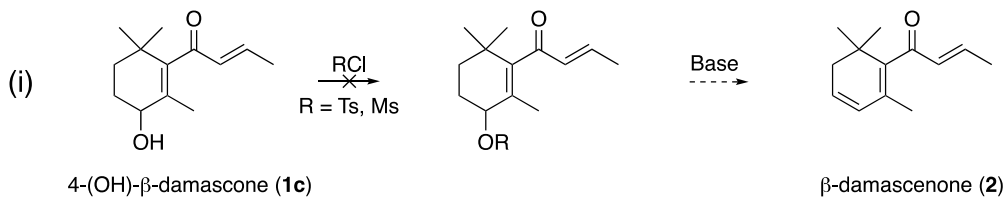


Figure 3.9 Protection of **1c** followed by base treatment.

Route ii: Heating **1c** with dilute H_2SO_4 (pH = 1) at 45 °C did not trigger any reaction; when the temperature was raised to 70 °C for 16 h, **1c** was fully converted to β-damascenone (**2**) as well as three other uncharacterised products (Figure 3.10). Despite the total conversion, not only was the selectivity for **2** low but the use of mineral acid is unacceptable for "natural" grade product labelling. The use of organic acids was therefore explored.

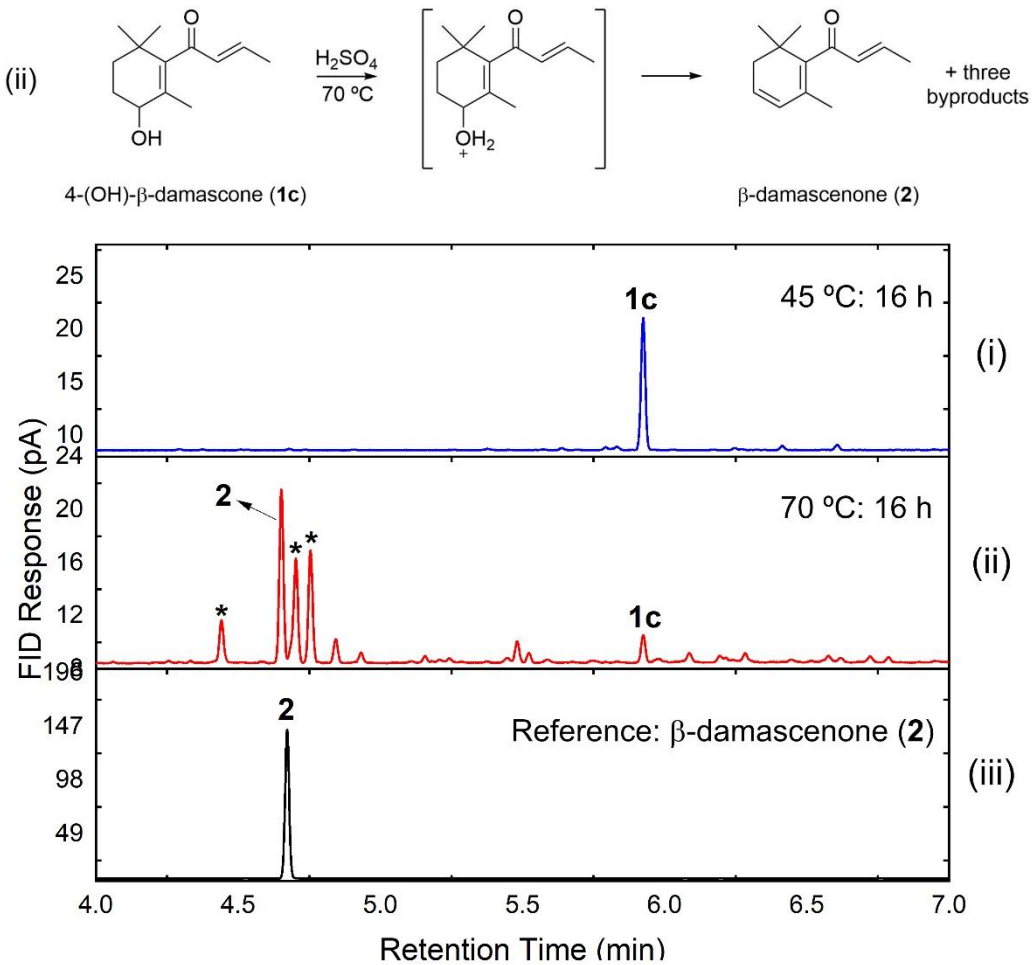


Figure 3.10 Time-course of treatment of 4-hydroxy-β-damascone (**1c**) with dilute H_2SO_4. (i) 20 mg of **1c** in 2 mL of chloroform was stirred at 45 °C for 16 h. (ii) Reaction was stirred at 70 °C for 16 h, giving a mixture of 4 products which contained 22% of β-damascenone **2**. (iii) Commercial sample of **2** as reference. * denotes uncharacterised products.

Route iii: With organic acids the reaction proceeds via the ester intermediate which eliminates the free acid on heating. In route iii, acetic acid (pK_a = 4.75) and acetic anhydride were firstly studied with 4-hydroxy-β-damascone (**1c**) at room temperature (RT); no reaction was observed.

The temperature was raised to 65 °C for 24 h and 40% of **1c** was converted to the ester **3**. Total conversion to **3** was achieved at 85 °C; however, no formation of β-damascenone (**10**) from the ester was observed (Figure 3.11).

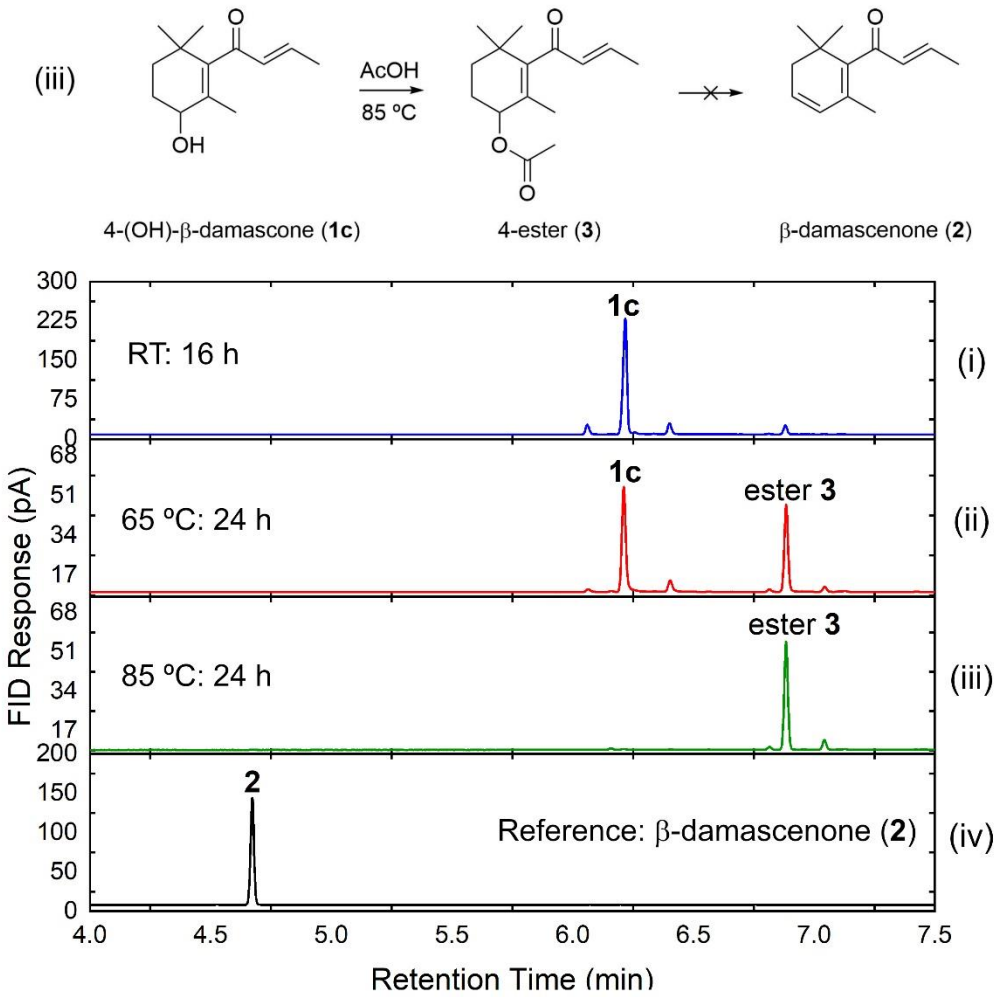


Figure 3.11 Time-course reaction of 4-hydroxy-β-damascone (**1c**) with acetic acid. (i) 20 mg of **1c** in 2 mL of toluene was stirred at RT for 16 h. (ii) Reaction was stirred at 65 °C for 24 h, 40% of **1c** was converted to the ester **3**. (iii) Reaction was stirred at 85 °C for 24 h, **1c** was fully converted to the ester **3**. (iv) Commercial sample of β-damascenone **2** as reference.

Route iv: A much stronger acid, trifluoroacetic acid (pK_a = 0.23) partially converted 4-hydroxy-β-damascone (**1c**) to the corresponding ester **4** at 45 °C in 48 h but no elimination was observed after 5 d. When the reaction temperature was raised to 90 °C, fast elimination occurred to give **2** (Figure 3.12). Activated molecular sieves were added to shift the equilibrium further by

removing water. Although this demonstrated the synthesis of **2** from **1c** with the use of trifluoroacetic acid, the process was less desirable due to the harsh reaction conditions and long reaction time.

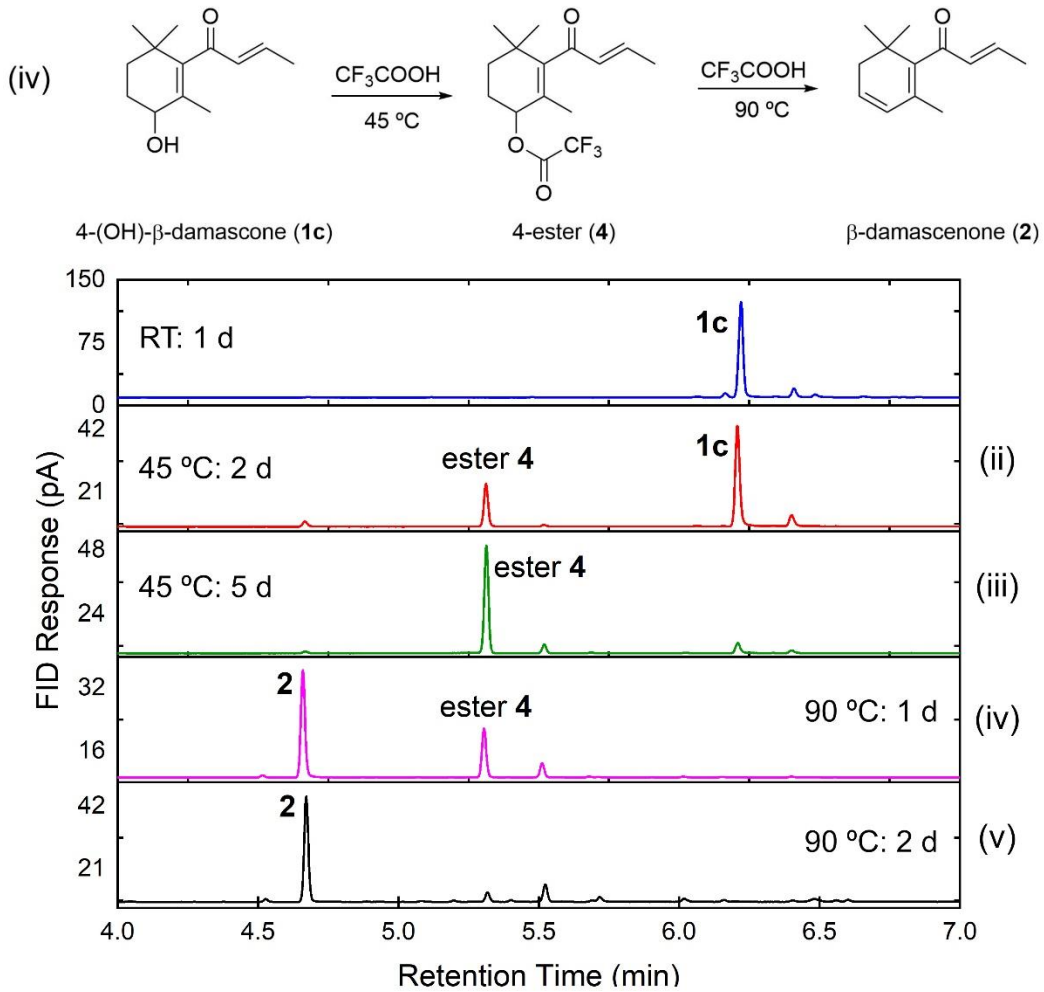


Figure 3.12 Time course reaction of 4-hydroxy-β-damascone (**1c**) with trifluroacetic acid. (i) 20 mg of **1c** in 2 mL of toluene was stirred at RT for 1 d. (ii) Reaction was stirred at 45 °C for 1 d, 25% of **1c** was slowly converted to the ester **4**. (iii) Reaction was stirred at 45 °C for another 3 d, **1c** was fully converted to the ester **4**. (iv) Reaction was stirred at 90 °C for 1 d, 68% of **1c** was converted to β-damascenone **2**. (v) **1c** was totally converted to β-damascenone **2**.

Route v: The pK_a of oxalic acid (1.25) is between those of acetic acid and trifluoracetic acid. The reaction of 20 mg of **1c** with 10 mg of oxalic acid (**1c**: oxalic acid = 1:1) was carried out at 65 °C for 16 h in the presence of activated molecular sieves; **1c** was fully converted to β-damascenone (**2**) with no evidence of any intermediate (Figure 3.13). Oxalic acid converted the

allylic alcohol into a better leaving group, leading to accelerated elimination to give **2**, which was isolated and fully characterised by NMR and MS (Appendix 3).

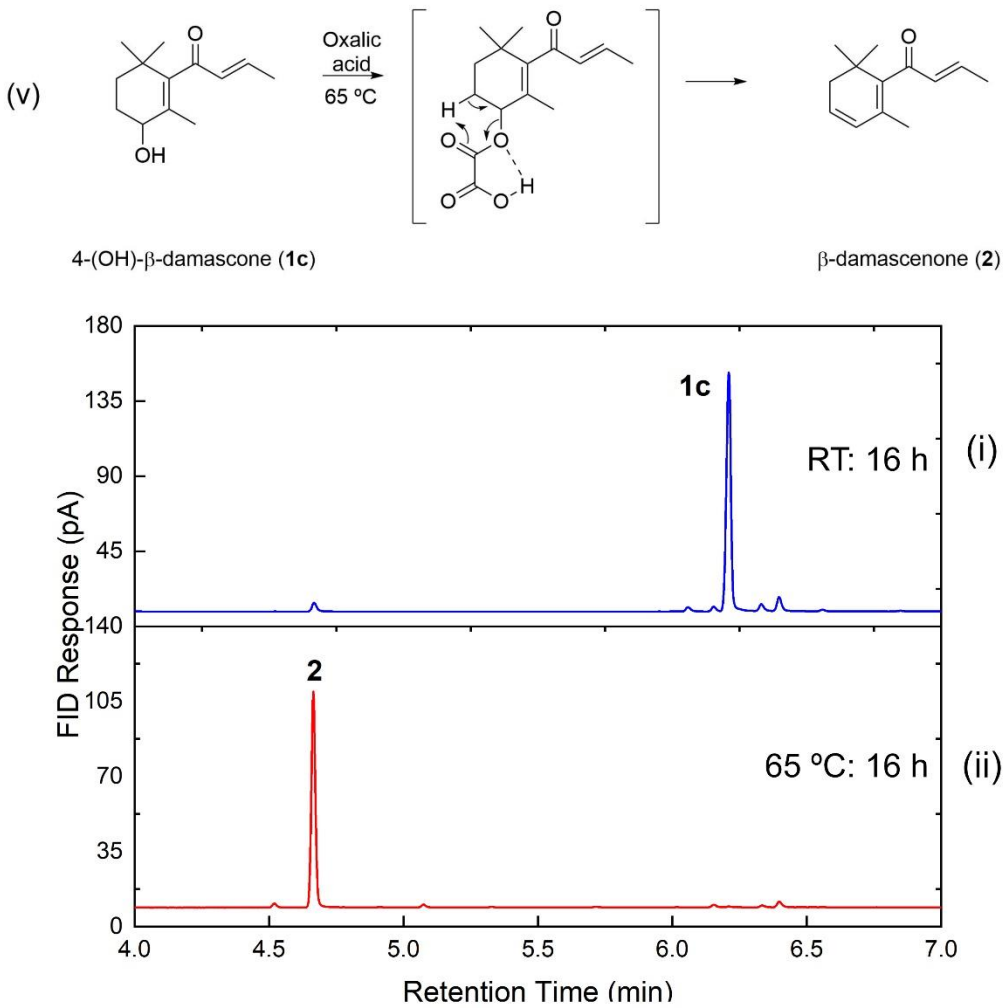


Figure 3.13 Time course reaction of 4-hydroxy-β-damascone (**1c**) with oxalic acid. (i) 20 mg of **1c** in 2 mL of toluene was stirred at RT for 16 h. (ii) The reaction was stirred at 65 °C for 16 h; **1c** was fully converted to β-damascenone **2**.

In order to test the scalability of the acid treatment, a 265-mg reaction of **1c** with 200 mg of oxalic acid was performed, heating at 65 °C in toluene (10 mL). The reaction was completed in 16 h with full conversion and the isolated yield was 66%.

3.2.4 β-Damascenone synthesis summary

The chemoenzymatic synthesis of β-damascenone (**2**) from β-damascone (**1**) has been established using a $P450_{BM3}$ variant for the generation of the 4-hydroxy precursor **1c** and oxalic acid for the dehydration of **1c** to **2** (Figure 3.14). The purified β-damascenone has a very powerful floral fruity note that was rose-like with hints of ripe plum. The ^{1}H NMR spectrum of **2** is shown in Figure 3.15.

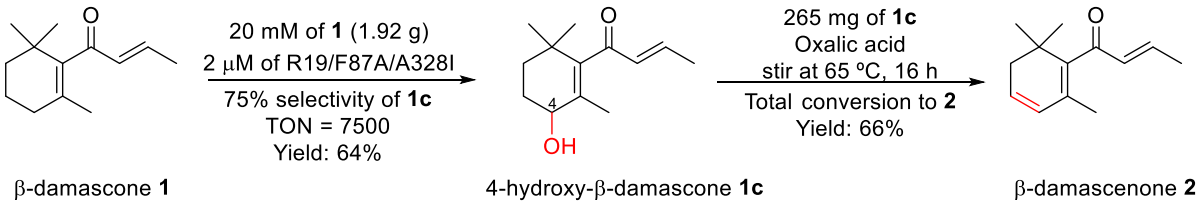


Figure 3.14 Scalability of β-damascenone synthesis.

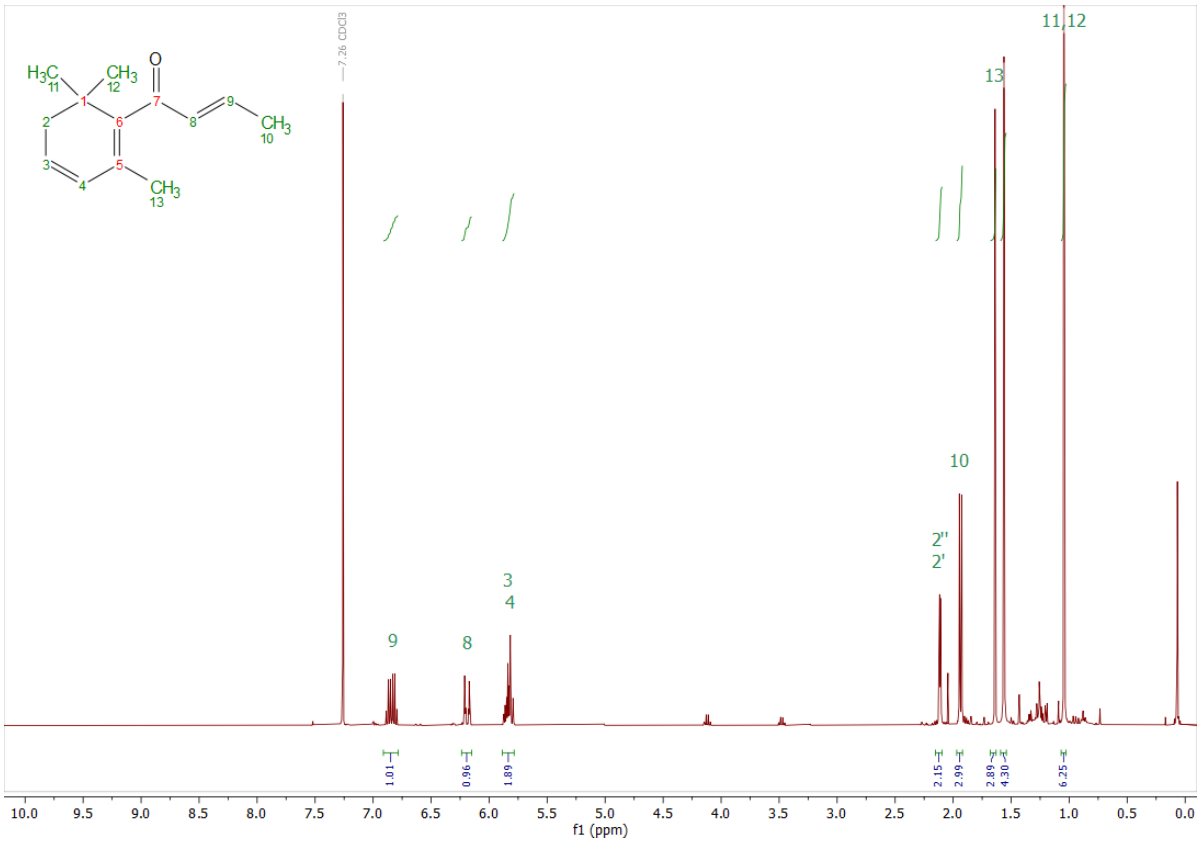


Figure 3.15 The ^{1}H NMR spectrum of chemoenzymatically synthesised β-damascenone, **2** in $CDCl_3$.

3.3 Tabanone synthesis

Tobacco is one of the richest sources for degraded carotenoids with almost 100 norisoprenoids and derivatives having been identified. Four major carotenoids are found in green tabacco that are transformed to a large number of metabolites including ionones, damascones and tabanones. Tabanones, also known as megastigmatrienones (Figure 3.16), are the most important components of electronic cigarette cartridges that impart the aroma of tabacoo notes.

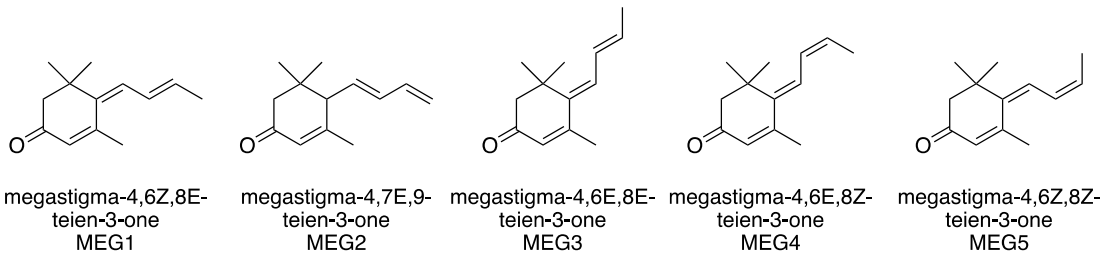


Figure 3.16 Five isomers of the megastigmatrienones.

The 95% pure synthetic tabanone used in aromas and flavourings retails for approximately £1500 per kilo; the price of the pure samples of specific isomers marketed by specialists is up to 60 times higher. Due to difficulties in controlling the selectivity, only a few synthetic methods for these megastigmane derivatives have been reported. In one synthesis, tabanone isomers with a ratio of *Z*/*Z* : *Z*/*E* : *E*/*Z*: *E*/*E* = 3 : 10 : 1 : 6, were obtained in 50% yield from 3-oxo-α-damascol which was synthesised in three steps from isophorone with an overall yield of 27% (Figure 3.17).[209] In spite of the successful synthesis of tabanone, the long pathways resulted in low yield and required the use of hazardous substances. On the other hand, biocatalytic routes could provide alternative pathways to tabanone which not only minimise the amount of organic solvents and hazardous chemicals but also reduce the energy input in the reaction.

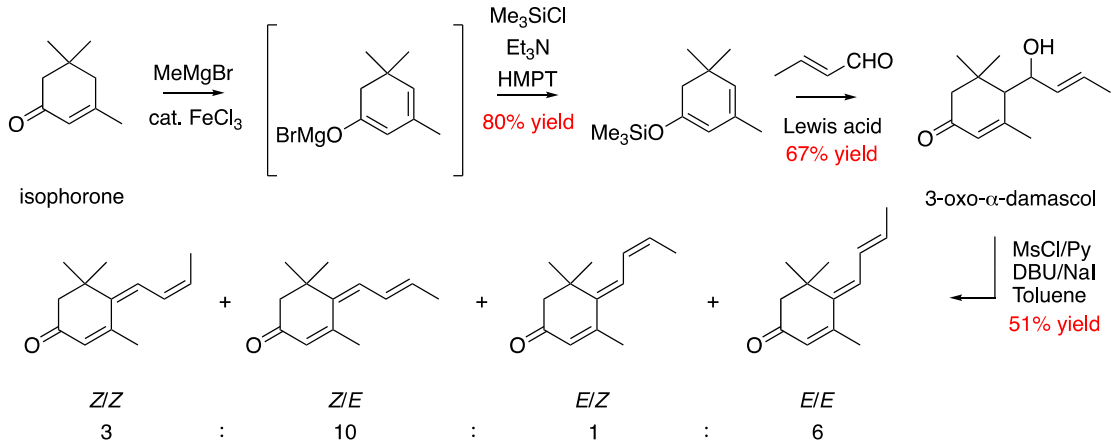


Figure 3.17 Synthesis of a mixture of tabanone isomers from isophorone.

Three biocatalytic routes to tabanone via the oxidation of α-ionone, β-ionone and α-ionol were proposed (Figure 3.18). Both ionones are cheap and commercially available, α-ionol could be easily obtained by reduction of α-ionone with sodium borohydride.

The first pathway involves the oxidation of α-ionone to 3-hydroxy-α-ionone by $P450_{BM3}$ variants, followed by further oxidation to 3-oxo-α-ionone either by $P450_{BM3}$ or an alcohol dehydrogenase. CYP101B1, from *Novosphingobium aromaticivorans* DSM12444, catalyses C3 oxidation of α-ionol and α-ionone as well as C3/C4 oxidation of β-ionol and β-ionone.[210] Further oxidation of 3-hydroxy-α-ionone to the ketone product, however, showed lower conversion. Classical chemical synthesis routes could also give this ketone product from α-ionone, e.g. by using *t*-butyl hydroperoxide and catalytic amount of $Co(acac)_2$ in acetone at room temperature for 15 h in 75% yield,[211] and oxidation using chromium trioxide in acetic acid for 6 h.[212]

In the second route, the reduction of β-ionone to the *E*/*E* megastigmatriene by sodium dimethylaminoborohydride followed by toluene-4-sulfonic acid has been reported.[213] Oxidation of megastigmatriene to the alcohol and then ketone would lead directly to tabanone.

The third pathway involves the oxidation of α-ionol to 3-hydroxy-α-ionol followed by the selective oxidation of the C3-alcohol to the ketone. Chemical oxidising agents would preferentially attack the less hindered C9-alcohol; therefore, the use of CYP102A1 or an alcohol dehydrogenase would be required.

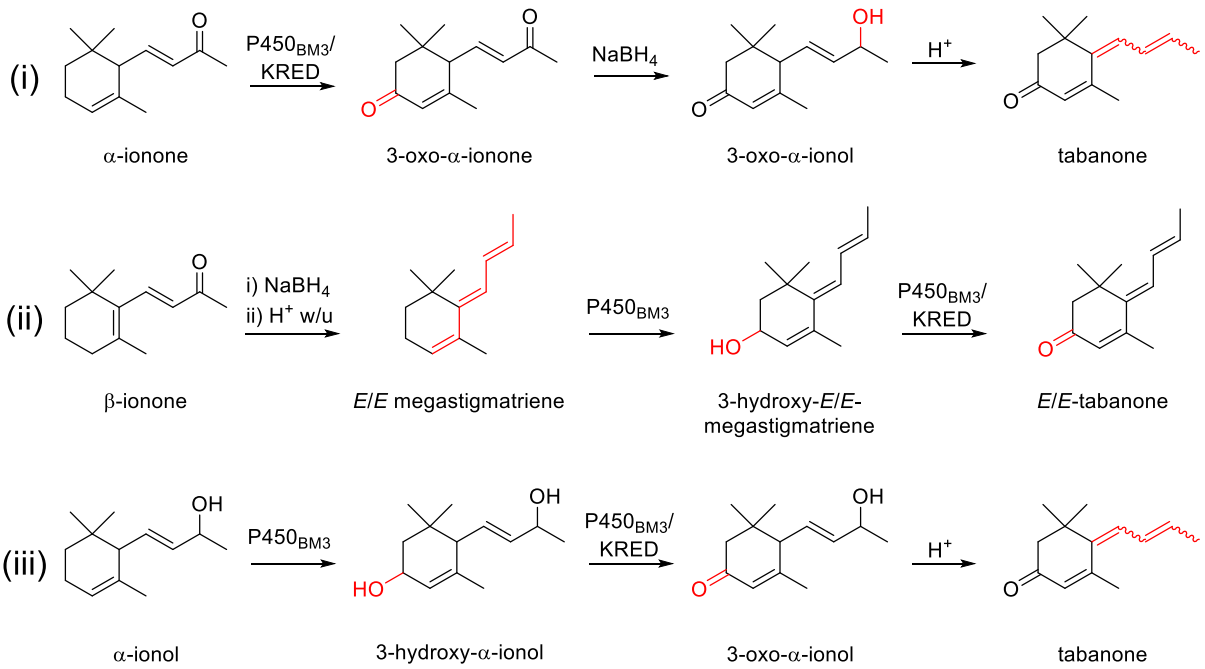


Figure 3.18 Three proposed synthetic routes of tabanone from i) α-ionone; ii) β-ionone; iii) α-ionol.

3.3.1 Route i – α-ionone oxidation by P450$_{BM3}$

A library of 144 P450$_{BM3}$ variants was screened against α-ionone **5** (5 mM substrate and 2 μM enzyme). Full activity and selectivity data are given in the Appendix 4. 71 out of 144 variants showed >50% conversion and led to five major products which were purified and characterised (Figure 3.19) as the 4,5-epoxide (**5a**),[214] *cis*-3-hydroxy-α-ionone (**5b**),[210] *trans*-3-hydroxy-α-ionone (**5c**),[210] 3-oxo-α-ionone (**5d**)[210] and a ring cyclisation product (**5e**). The NMR data of compounds **5a**, **5b**, **5c** and **5d** are in agreement with literature data and shown in Appendix 4.

Compound **5e** is a new metabolite of α-ionone. The full characterisation data for **5e** is collected in Appendix 4.

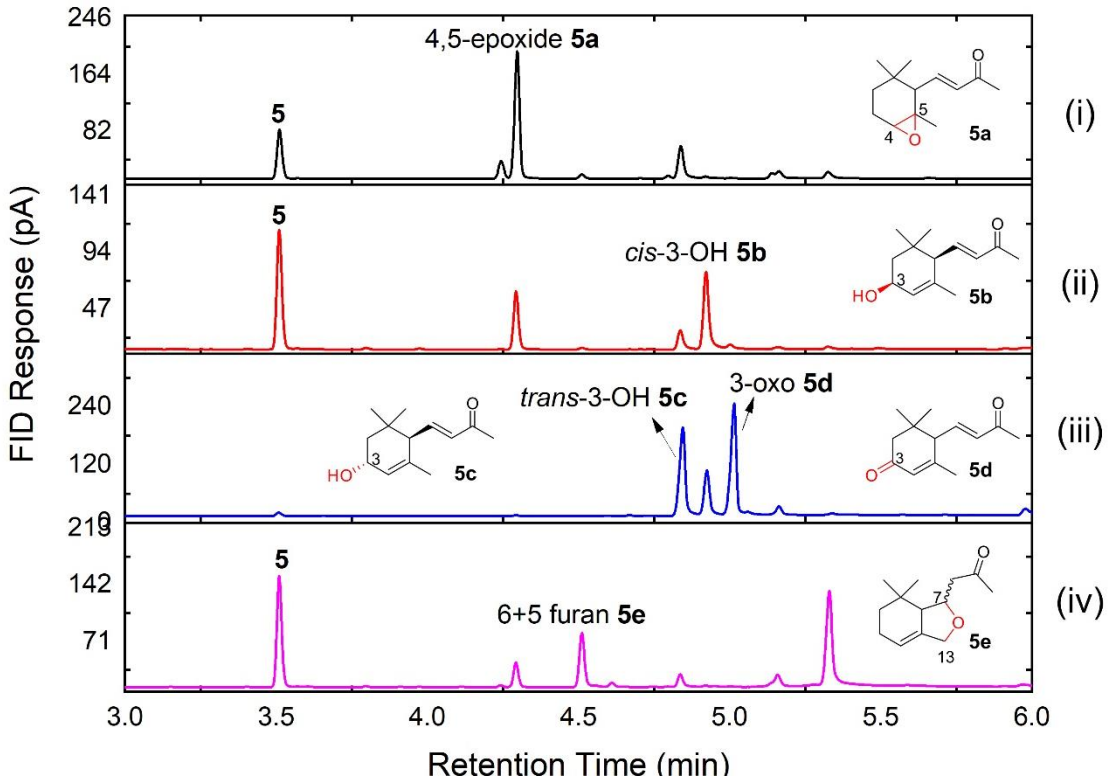


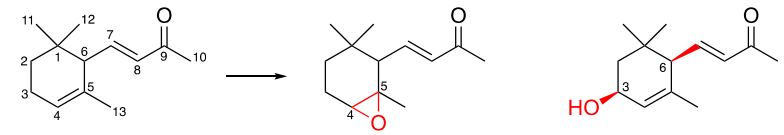
Figure 3.19 GC profile of α-ionone (**5**) oxidation products. (i) 62% of 4,5-epoxide **5a** by KU3/A330P/I263G; (ii) 50% of the *cis*-3-hydroxy-α-ionone **5b** by R19/A82M/T260G; (iii) 33% of the *trans*-3-hydroxy-α-ionone **5c** and 39% of 3-oxo-α-ionone **5d** by R19/F87A/A328I/S72H; (iv) 24% of the furan product **5e** by R19/A82M/I263G/A264G/A328G.

The epoxide **5a** was the predominant product for variants without a mutation of F87; 29 variants based on RP, RT2 and KU3 gave >50% of **5a**. Mutations of the I-helix residue I263 increased this further (Table 3.2, Entries 1–3).

Variants based on F87A/V/I favoured allylic oxidation, with the *cis*-3-alcohol **5b** being the major product. R19/F87A gave 44% of **5b**; addition of A328 mutations increased this to 70% for R19/F87A/A328L (Entry 10). The F87A/I263A combination also preferred the *cis* alcohol; K19/F87A/I263A and K19/F87A/A328I/I263A showed 62% and 56% of **5b** (Entries 8 & 7). The *trans*-3-alcohol **5c**, was favoured by variants without F87 mutations, with

R19/A184I/T260G giving the highest selectivity (Entry 13). The A328I mutation also favoured the formation of **5c** in variants K19/F87A/A184I/A328I and RT2/F81W/A328I (Entries 11 & 12).

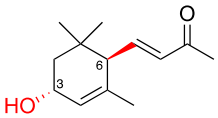
Table 3.2 Substrate conversion and product selectivity of selected variants for the oxidation of α-ionone, **5**.[a]



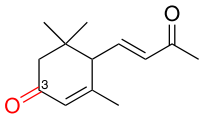
α-ionone, **5**

4,5-epoxy-α-ionone, **5a**

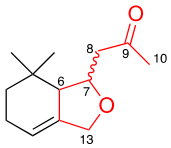
cis-3-(OH)-α-ionone, **5b**



trans-3-(OH)-α-ionone, **5c**



3-oxo-α-ionone, **5d**



6+5 furan **5e**

Entry	Variants	5a	5b	5c	5d	5e	Others	Conv.	TON[b]
1	KU3/A330P/I263G	**62%**		17%			21%	81%	1256
2	RP/H171L/I263G	**65%**		27%			8%	44%	715
3	RT2/I263A/A330W	**76%**	2%	9%		3%	10%	50%	950
4	R19/F87A		**44%**	22%	27%		7%	80%	880
5	R19/F87A/A328I		**42%**	38%	13%		6%	73%	767
6	R19/F87A/A328F		**53%**	32%	4%		11%	43%	570
7	K19/F87A/I263A		**62%**	24%	6%		8%	34%	527
8	K19/F87A/A328I/I263A		**56%**	29%	8%		7%	85%	1190
9	R19/F87A/A328I/E267F		**69%**	15%	8%		8%	92%	1587
10	R19/F87A/A328L	14%	**70%**	14%			2%	91%	1593
11	K19/F87A/A328I/A184I		25%	**46%**	26%		3%	92%	1058
12	RT2/F81W/A328I	9%	22%	**58%**			11%	37%	537
13	R19/A184I/T260G	3%	19%	**58%**	13%		7%	46%	667
14	F87A		35%	21%	**37%**		7%	90%	833
15	R19/F87A/A328I/S72H		17%	33%	**39%**		11%	99%	966
16	R19/F87A/A328I/S72A		4%	23%	**39%**		34%	100%	975
17	R19/I263G/A264G/A328G	26%	2%	8%		**16%**	48%	77%	308
18	R19/A82M/I263G/A263G/A328G	10%		6%		**24%**	60%	70%	420
19	R19/A82M/I263G/A328G	4%		8%		**28%**	60%	45%	315

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 5 mM α-ionone **5**, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold.

Further oxidation of **5b** and **5c** to the 3-oxo product **5d** was favoured by F87A-based variants. The single mutation variant F87A gave 37% of **5d** with 90% conversion, and the variants R19/F87A/A328I/S72H and R19/F87A/A328I/S72A both gave 39% selectivity for **5d** with full conversion (Entries 14–16). **5d** was the main target product in the oxidation of α-ionone; therefore, variants giving high selectivity for C3 oxidation (*cis*- **5b** and *trans*- **5c**) were selected for the subsequent oxidation to **5d** by KREDs.

The furan product **5e**, was a minor metabolite that was likely formed via the proposed mechanism in Figure 3.20 whereby the allylic C13-alcohol undergoes a Michael addition on the α,β-unsaturated ketone. The proposed precursor, 13-hydroxy-α-ionone, was not observed in the reaction, indicating the rapid ring closure. Combinations of space-creating mutations, especially to Gly mutation, gave highest selectivity for **5e** (Entries 17–19).

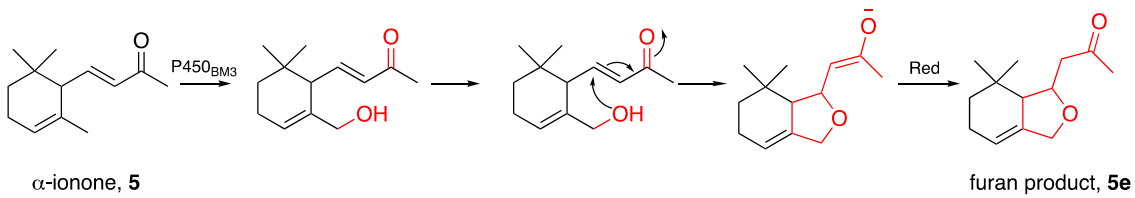


Figure 3.20 Proposed mechanism of formation of product **5e**.

3.3.2 Route i – Ketoreductase (KRED)

The proposed precursor of tabanone in route i, 3-oxo-α-ionone (**5d**), was a minor product in the reaction with the highest selectivity of 39%. This could be improved for the subsequent selective reduction and acidification steps by further engineering of the CYP102A1 enzymes such as K19/F87A/A328I/A184I which gave 97% C3 oxidation, of which 26% was **5d**. In an attemptive approach, three different ketoreductase (KRED) 231, 641 and 771 from Almac, were tested for the oxidation of alcohols **5b** and **5c** to this ketone.

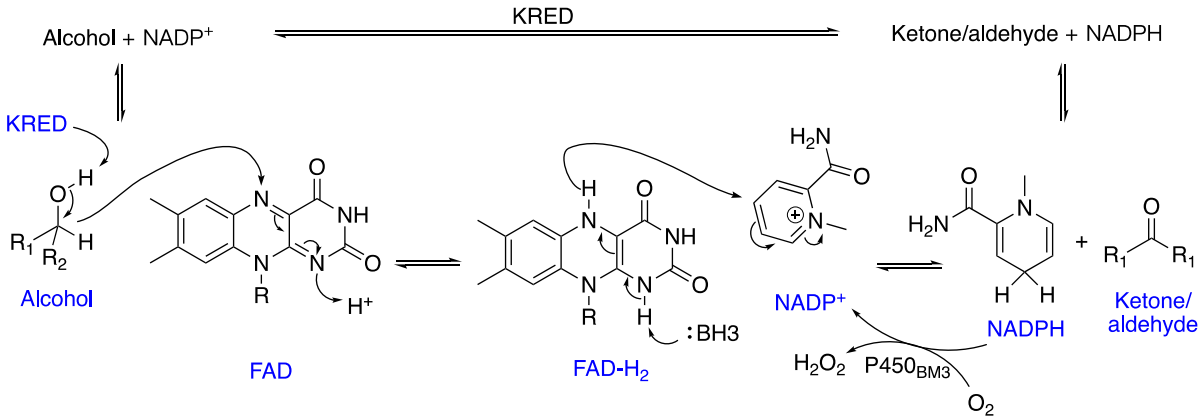


Figure 3.21 Mechanism of the oxidation of alcohol to ketone/aldehyde by the use of FAD-dependent KRED with the aid of $P450_{BM3}$.

Ketoreductases (KRED) catalyse the oxidation of an alcohol on a target substrate by using $NAD(P)^+$ as a hydride acceptor. The mechanism of FAD-dependent KREDs is shown in Figure 3.21. The hydride is transferred to the FAD prosthetic group to give the FAD-H_2 form, which passes the hydride onto $NAD(P)^+$ to give NADPH. This reaction is reversible, and the equilibrium position depends on the relative binding strengths of substrate and product, temperature and the pH of the reaction. For alcohol oxidation the equilibrium can be shifted to the ketone by recycling of the $NADP^+$ cofactor. Moreover, this can be achieved by using the reductase domain of $P450_{BM3}$ which oxidises NADPH and transfers the two electrons to reduce oxygen to hydrogen peroxide.

The three keto-reductases were screened using 5 mM of product mixture (*trans*-3-ol **5c**, *cis*-3-ol **5b** and 3-oxo-α-ionone **5d**) with the R19/F87A variant in 1 mL of phosphate buffer (200 mM, pH 8.5). The reaction containing 5 mM $NADP^+$ and 1 mg of KRED was stirred at 37 °C for 2 h. GC analysis showed that KRED 231 selectively oxidised the *trans*-3-ol **5c** to the ketone **5d** while KRED 641 selectively oxidised the *cis*-3-ol **5b** to the ketone, and KRED 771 did not show any activity (Figure 3.22). Therefore, the substrates were tested with a mixture of KRED 231 and 641 on a larger scale; a 125 mL reaction (30 mM product mixture, 10 mg KRED 231,

10 mg KRED 641, 5 mg of $NADP^+$, 0.4 μM $P450_{BM3}$) was performed at 37 °C for 16 h to give >90% conversion of the two alcohols to the ketone which constituted 75% of the products in the final reaction mixture.

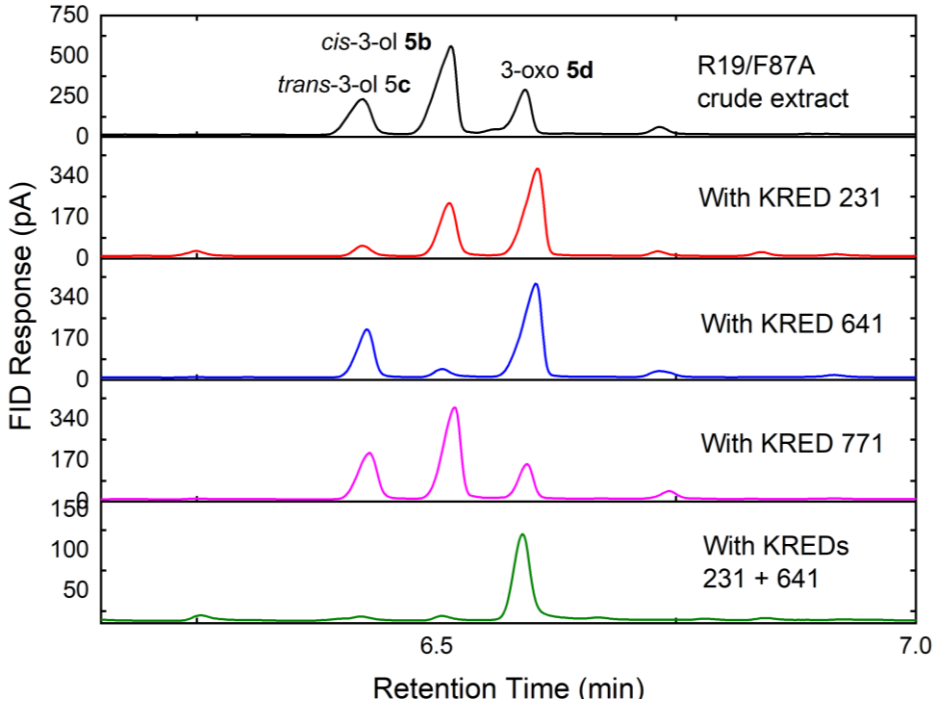


Figure 3.22 The test trials of starting material, including mixture of *trans*-3-hydroxy-α-ionone **5c**, *cis*-3-hydroxy-α-ionone **5b** and 3-oxo-α-ionone **5d**, with different kinds of KRED performed at 37°C.

3.3.3 Route ii - (*E/E*)-megastigmatriene oxidation by $P450_{BM3}$

These experiments were carried out by another member of the group, Rory Woodhouse, and the results are simply noted here. 250 mg of β-ionone was dissolved in 20 mL THF with 20 mg of $NaBH_4$ and stirred for 30 min; the preparation of (*E/E*)-megastigmatriene (**6**) was completed after the addition of 5 mL of 1 M HCl and stirring for 3 h (99% conversion, 84% yield). The structure of **6** was assigned by comparison with literature NMR data, and the *E/E* geometry was established by the NOESY spectrum (Figure 3.23), which showed coupling of H8 with H11/12 and of H7 with H13.[215]

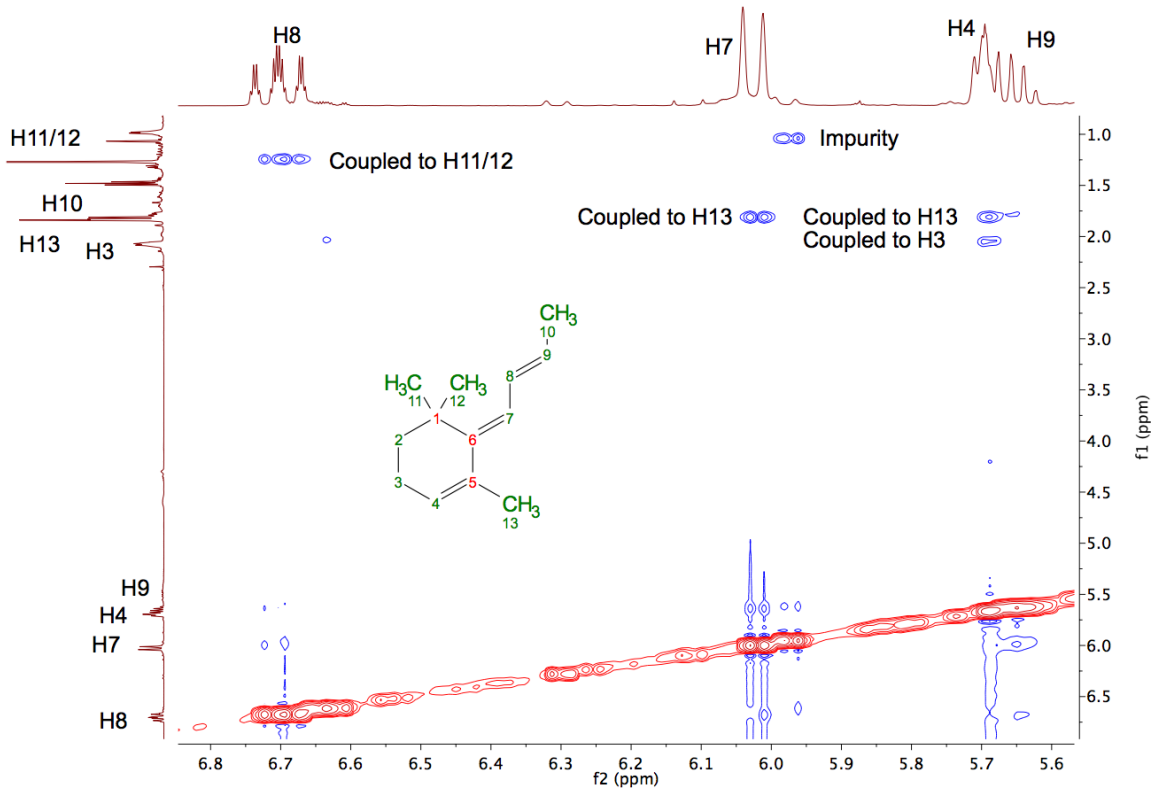


Figure 3.23 The NOESY spectrum of **6** establishing the *E*/*E* geometry.

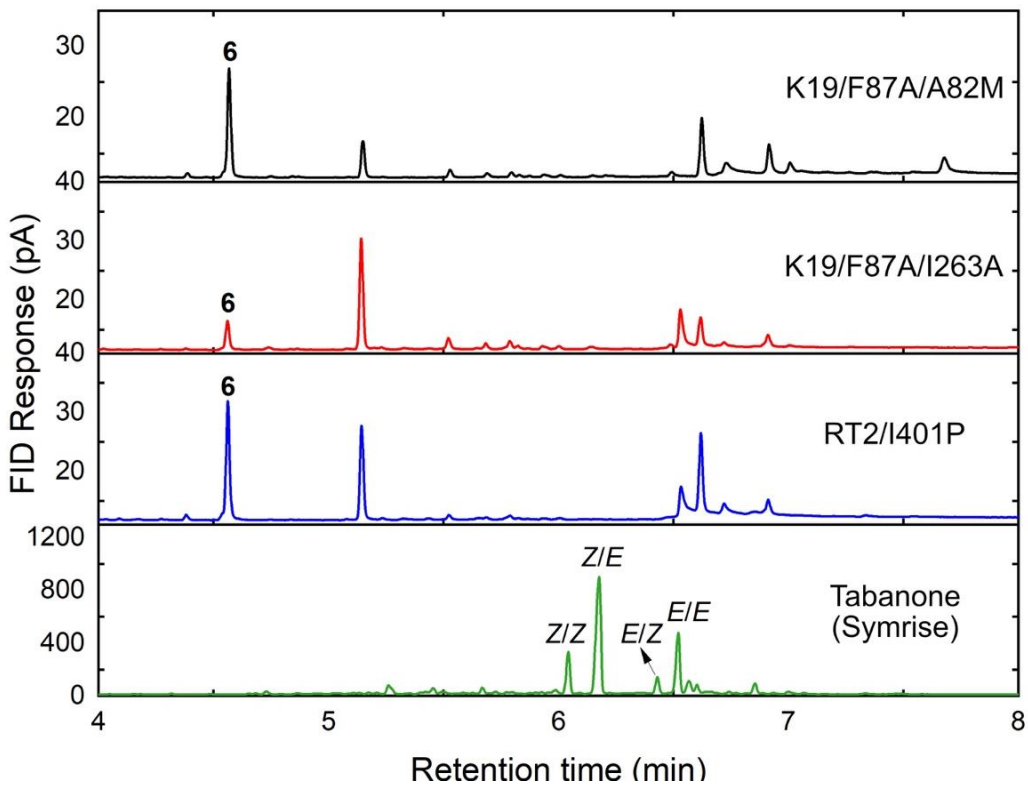


Figure 3.24 GC profile of (*E*/*E*)-megastigmatriene **6** oxidation products and the commercial tabanone from Symrise.

Due to its low aqueous solubility, (*E/E*)-megastigmatriene (**6**) was screened with 24 variants at 1 mM substrate and 2 μM enzyme (500:1 ratio), with acetone used as a co-solvent (2.5% *v/v*) and 14 variants showed >20% conversion (Figure 3.24).

The product profiles were compared with a commercial sample of tabanone from Symrise which contained 17% *E*/*E*-, 43% *Z*/*E*-, 5% *E*/*Z*-, and 12% *Z*/*Z*-tabanone (Figure 3.24).[216] No product peak matched any of the tabanone isomers. Preparative scale reactions (40 mg of **6**) were also conducted but no product could be eluted by silica gel chromatography. Therefore, this route was not pursued further.

3.3.4 Route iii – α-ionol oxidation by $P450_{BM3}$

120 $P450_{BM3}$ variants were screened for oxidation of racemic α-ionol (5 mM substrate and 2 μM enzyme). The enzymes were selected from a large library (around 600 mutants) based on expression levels (>60 mg protein per L of culture) and turnover numbers (>20% conversion) towards norisoprenoid substrates. Full activity and selectivity data are given in the Appendix 5. α-Ionol (**7**) was a good substrate with 61 enzymes showing >50% conversion; the variant library gave 8 major products including the 4,5-epoxide (**7a**),[217] *cis*-3-hydroxy-α-ionol (**7b**),[218] *trans*-3-hydroxy-α-ionol (**7c**),[218,219] 13-hydroxy-α-ionol (**7d**), 3-oxo-α-ionol (**7e**),[219] 3-oxo-13-hydroxy-α-ionol (**7f**),[220] 3-oxo-α-ionone (**7g**)[210] and a cyclisation product (**7h**) (Figure 3.25). The NMR data of compounds **7a**, **7b**, **7c**, **7e**, **7f** and **7g** are in agreement with literature and shown in Appendix 6. Compounds **7d** and **7h** are new metabolites of α-ionol.

The screening results showed that the mutations on F87V and I263G favoured the epoxide **7a** (Table 3.3, Entries 1–7). Few variants showed high selectivity at high conversion for the *cis*- and *trans*-3-hydroxy-α-ionols **7b** and **7c**. Those showing moderate selectivity (Table 3.3,

Entries 8–16), tended to be less active, with low conversion. However, the main target product was 3-oxo-α-ionol (**7e**), the precursor to tabanone. Given that the 3-hydroxy-α-ionol isomers **7b** and **7c** could be converted to the ketone **7e** by the action of KREDs, the most relevant selectivity trait was the total C3 selectivity, i.e. the total percentage of the C3 alcohols and the 3-oxo compound **7e**. However, this compound co-eluted with its oxidation product 3-oxo-α-ionone (**5d**/**7g**) on GC which made it difficult to establish the selectivity for its formation. Variants with the highest C3 selectivity (>93%) contained mutations at I-helix residues (T260G and S270G) (Entries 23–25), also showing high conversion.

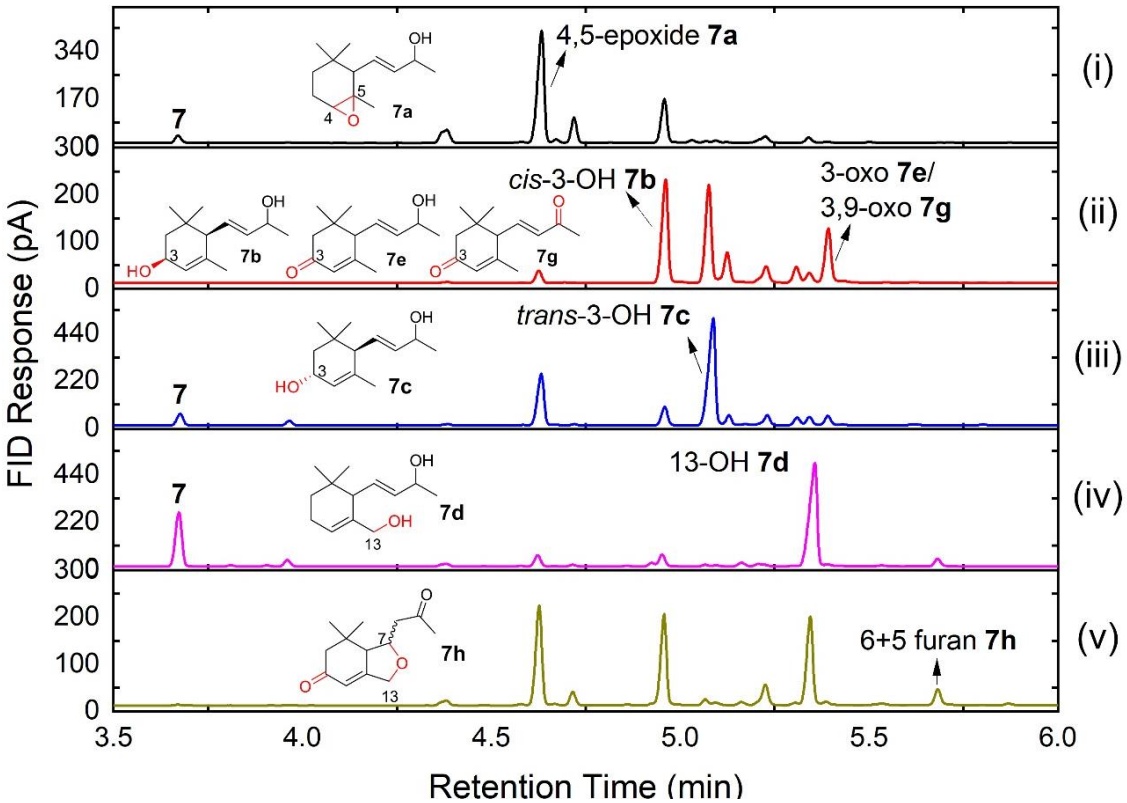
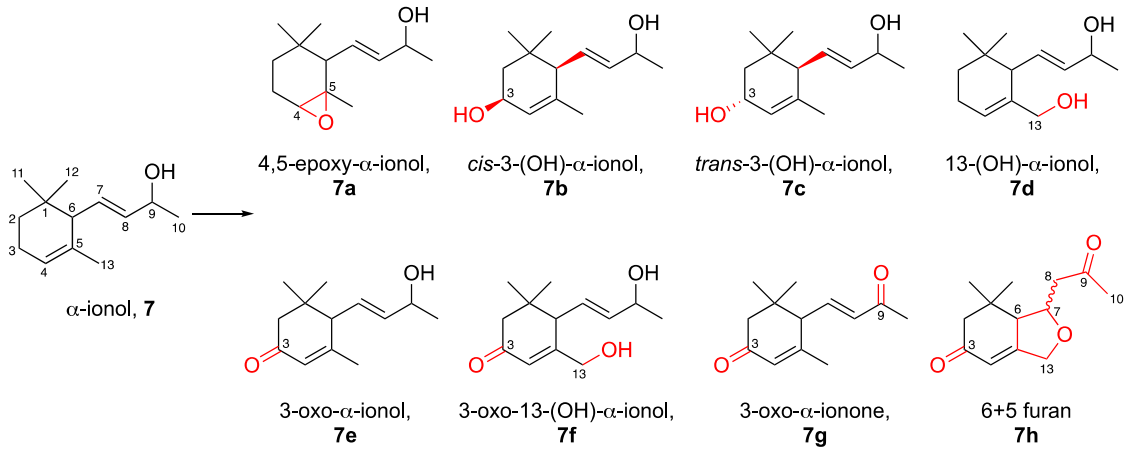


Figure 3.25 GC profile of α-ionol **7** oxidation products. (i) 54% of 4,5-epoxide **7a** by RP/H171L/I263G; (ii) 31% of the *cis*-3-hydroxy-α-ionol **7b**, and 17% of 3-oxo-α-ionol **7e**/3-oxo-α-ionone **7g** by R19/F87A/A328I/S72A; (iii) 62% of the *trans*-3-hydroxy-α-ionol **7c** by R19/F87I; (iv) 70% of 13-hydroxy-α-ionol **7d** by variant R19/A82M/I263G/A264G/A328G; (v) 30% of **7a**, 27% of **7b**, 26% of **7d** and 5% of the furan product **7h** by RP/H171L/I263G/A82M.

Table 3.3 Substrate conversion and product selectivity of selected variants for the oxidation of α-ionol, **7**.[a]



Entry	Variants	7a	7b	7c	7d	7f	7e/7g	*A*	*B*	*C*
1	GV/A184I	11%		6%		20%		56%	100%	
2	K19/F87V	15%	4%	26%		16%	14%	25%	100%	
3	GV/A184I/A264G	**47%**	26%	16%			5%	6%	71%	834
4	GV/A184I/I263G/A328G/A264G	**47%**	14%		6%			33%	75%	881
5	K19/F87V/A264G	**51%**	9%	16%			10%	14%	88%	1122
6	RP/HL/I263G	**54%**	19%					27%	97%	1310
7	RP/HL/I263G/S72G	**54%**	20%					26%	99%	1337
8	R19/F87A/A328I/S72A	4%	**31%**	28%	3%	17%	8%	9%	100%	775
9	K19/F87A/A82M/I263G/A264G	7%	**45%**	33%	7%		5%	3%	29%	326
10	K19/F87A/A82M/I263G/A328G		**32%**	32%	7%		7%	22%	48%	384
11	K19/F87A/A82M/I263G/A264G/A328G		**41%**	25%	7%		7%	20%	26%	267
12	R19/F87A/A328I/S72W		**45%**	24%		4%	16%	11%	47%	529
13	GV/A184I/A328G/T260G	14%	19%	**52%**				25%	22%	286
14	R19/F87A/A328G/T260G		33%	**55%**	6%		6%		29%	399
15	K19/F87I	17%	10%	**57%**		6%		10%	44%	627
16	R19/F87I	13%		**62%**	3%			22%	40%	620
17	R19/I263G/A264G/A328G	18%	9%		**45%**			28%	94%	1058
18	R19/I263G/A328G	7%	7%		**52%**		5%	29%	100%	1300
19	R19/A82M/I263G/A328G	5%	9%		**63%**		8%	15%	47%	740
20	R19/A82M/I263G/A264G/A328G	6%	7%		**70%**		2%	15%	76%	1330
21	K19/F87A/A82M/A330W		20%	45%		**23%**		12%	96%	552
22	R19/F87A/A184I		17%	35%	3%	**34%**	4%	7%	100%	850
23	R19/F87A/A328G/S270G		28%	28%			**38%**	6%	89%	949
24	R19/F87A/A184I/T260G		26%	28%			**39%**	7%	95%	926
25	R19/F87A/A184I/S270G		5%	3%			**48%**	44%	100%	1200

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 5 mM α-ionol **7**, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. *A* denotes "Others" that consists of uncharacterised products. *B* denotes "Conversion" that refers to the % of substrate converted to products. *C* denotes "TON" that refers to the turnover number for the formation of the product in bold.

13-Hydroxy-α-ionol, **7d**, from oxidation of an allylic methyl group, was favoured by variants without F87 mutations. For example, the variant R19/A82M/I263G/A264G/A328G gave 70%

of **7d** (Entry 20) while the F87A-containing K19/A82M/F87A/I263G/A264G/A328G showed 66% of **7b** and **7c**, with only 7% of **7d** (Entry 11); K19/A82M/F87A/I263G/A328G (7% of **7d**, Entry 10) and R19/A82M/I263G/A328G (63%, Entry 19). Comparison of Entries 17 and 20 also illustrated the importance of A82M for giving **7d**. The cyclisation product **7h** was formed from a similar Michael addition mechanism for the formation of **5e** (Figure 3.26). Either the 13-hydroxy compound **7f** or 3-oxo-α-ionone **7g** could be the precursor; **7f** could be oxidised at C9 to the ketone which rapidly cyclised, or **7g** could be hydroxylated at C13 followed by ring closure. The 13-hydroxy-9-ketone intermediate for both pathways was not observed.

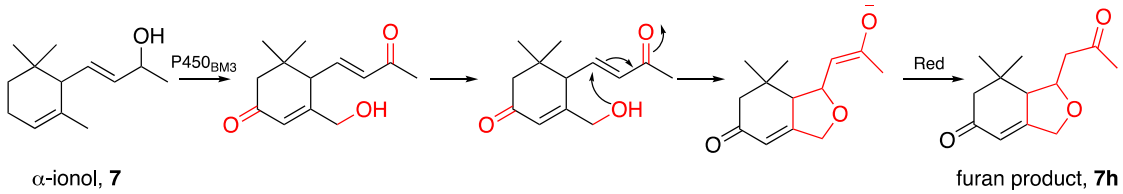


Figure 3.26 Proposed mechanism of formation of the furan product, **7h**, from α-ionol, **7**.

3.3.5 Route iii – Tabanone synthesis from α-ionol

The second step for the tabanone synthesis trial was the dehydration of 3-oxo-α-ionol **7e** catalysed either by acids or bases (Figure 3.27–3.30).

Base-catalysed elimination: 3-Oxo-α-ionol **7e** was heated with TsCl in pyridine at 55 °C. No reaction was observed until the addition of DMAP which led to full conversion of **7e** to the chloride **8** after 4 h (Figure 3.27 & Figure 3.28-iii). Product **8** was then treated with 20% NaOH in MeOH for 10 min at RT to give a mixture of 2 tabanone isomers, *Z*/*E* (MEG1) **9** and *E*/*E* (MEG3) **10** (Figure 3.28-iv) which were identified by NMR data. Under alkaline conditions, elimination of Cl^- from **8** occurred by abstraction of the most acidic proton H6, leading to a mixture of the **9** and **10** isomers via an E2 mechanism.

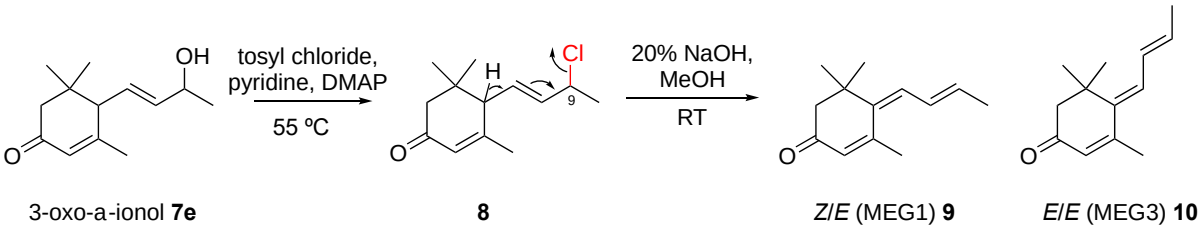


Figure 3.27 The base treatment of 3-oxo-α-ionol **7e** giving a mixture of tabanone isomers **9** and **10**.

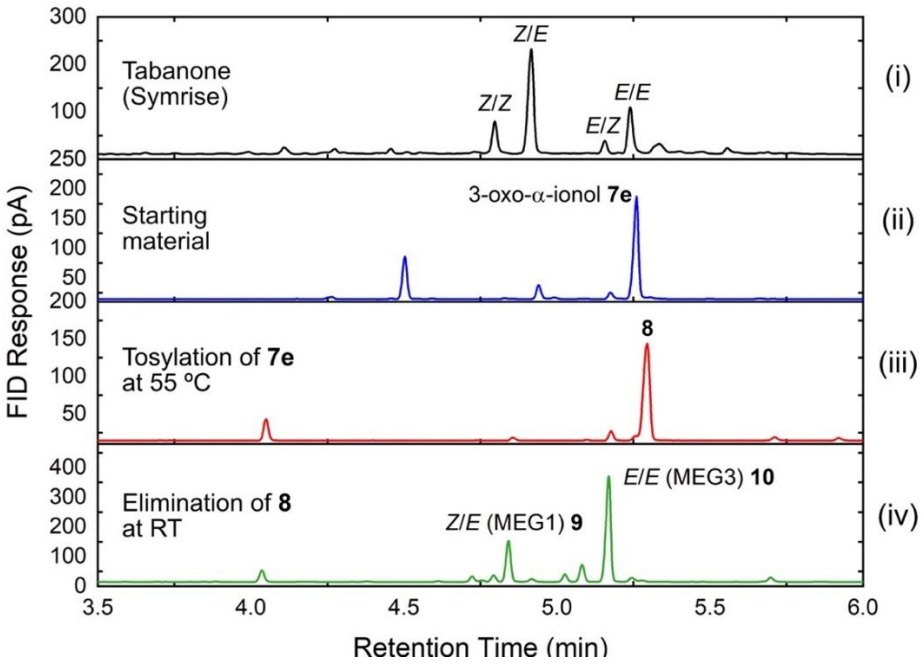


Figure 3.28 The base treatment of 3-oxo-α-ionol **7e** generated a mixture of tabanone isomers **9** and **10**.

Acid-catalysed elimination*:* The reaction of **7e** with oxalic acid did not show any conversion until the temperature was raised to 80 °C for 1 h (Figure 3.29-A); **7e** was fully converted to the terminal dehydration product **11**, the tabanone isomer MEG2 (Figure 3.30-iii). However, the other MEG isomers like **9** or **13** (Figure 3.29-C), with a C6=C7 double bond and a terminal C10 methyl group, were not formed. The reaction of **7e** with *p*-toluenesulfonic acid (*p*-TsOH) gave a ring cyclisation product **12** when heated at 110 °C for 10 min (Figure 3.30-iv), most likely via the terminal alkene **11** followed by acid-catalysed cyclisation (Figure 3.29-B). Other

acids such as trifluoracetic acid and sulfuric acid were also tested at different temperatures but only the terminal alkene product MEG2 (**11**) was formed.

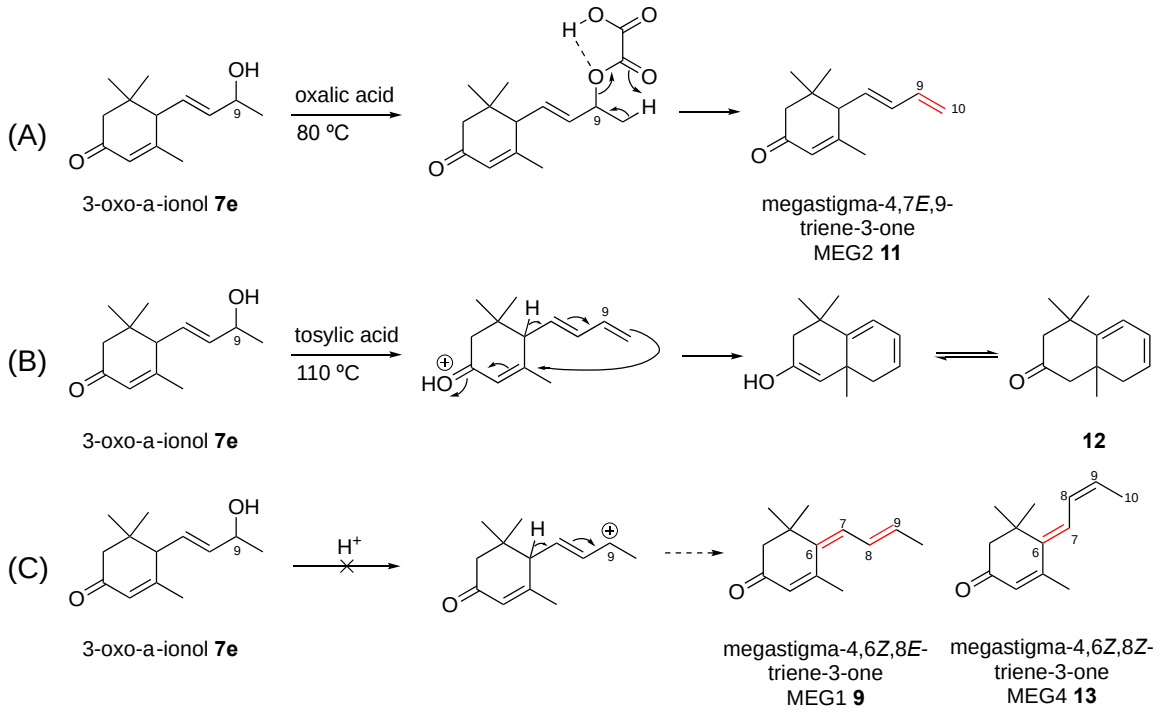


Figure 3.29 The acid treatment of 3-oxo-α-ionol **7e** to isomers **11** and **12**.

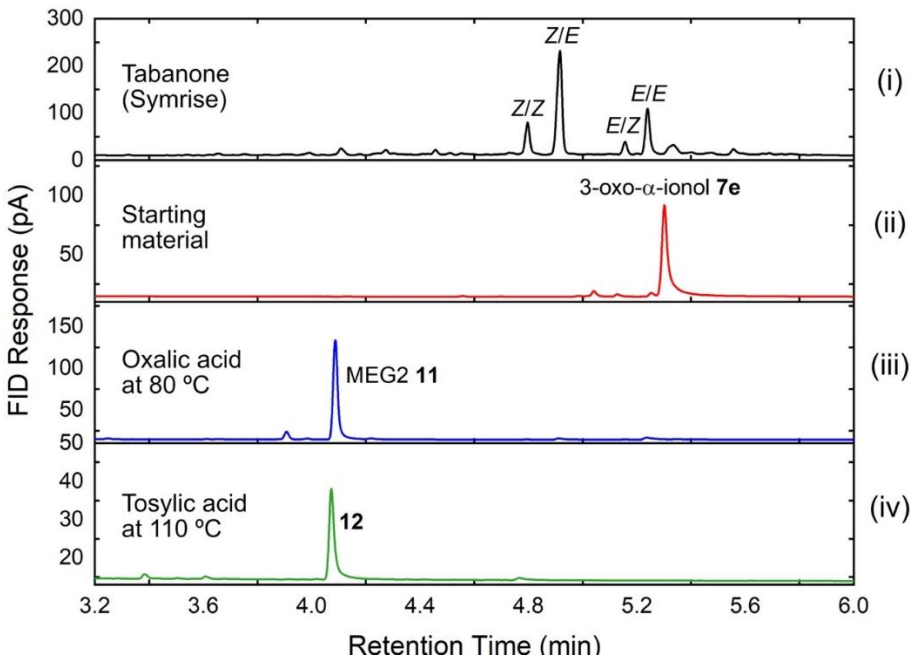


Figure 3.30 Acid treatment of 3-oxo-α-ionol **7e** giving products **11** or **12** at different conditions.

3.3.6 Combinatorial treatments for the synthesis of tabanone

In order to assess the best way to synthesise tabanone, two routes were provided in Figure 3.31 starting from α-ionone **5**.

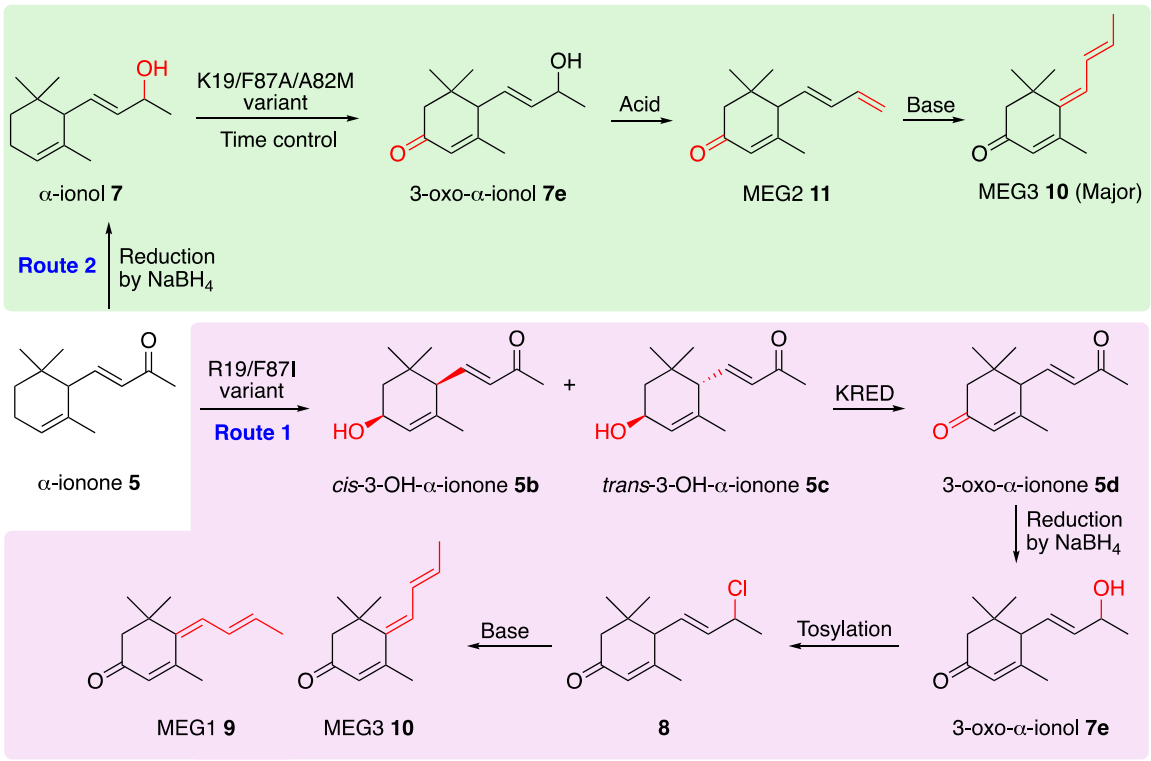


Figure 3.31 The established synthesis route of tabanone isomers **9** and **10** from α-ionone **5**.

In route 1 (pink panel in Figure 3.31), a preparative scale oxidation of α-ionone **5** (500 mg, ratio 5000:1) with R19/F87I was performed for 16 h and gave a mixture of 3-hydroxy-α-ionones **5b** and **5c** as major products. The pH of the reaction mixture was adjusted to 8.5 and the mixture was stirred at 37 ºC for 2 h with the addition of 100 mg of a mixture of KRED 231 and KRED 641. **5b** and **5c** were fully converted to 3-oxo-α-ionone **5d** (61% in the crude extract) (Figure 3.32-i). Compounds **5d** and **7e** co-eluted on the GC but could be separated on TLC, with R_f values of 0.88 and 0.75, respectively. Reduction of **5d** using $NaBH_4$ in MeOH was tested with various conditions. With an excess amount of $NaBH_4$, **5d** was fully converted to a

mixture of 3-hydroxy-α-ionol **7b** and **7c** within 1 min (Figure 3.32-ii). When 1:1 ratio of $NaBH_4$ was added, the crude extract **5d** was fully converted to **7e** together with small amounts of **7c** formed (Figure 3.32-iii). Finally, the most appropriate concentration of $NaBH_4$ and **5d** was found to be 1:0.5, which gave total conversion to **7e** as well as suppressed the formation of **7c** (Figure 3.32-iv).

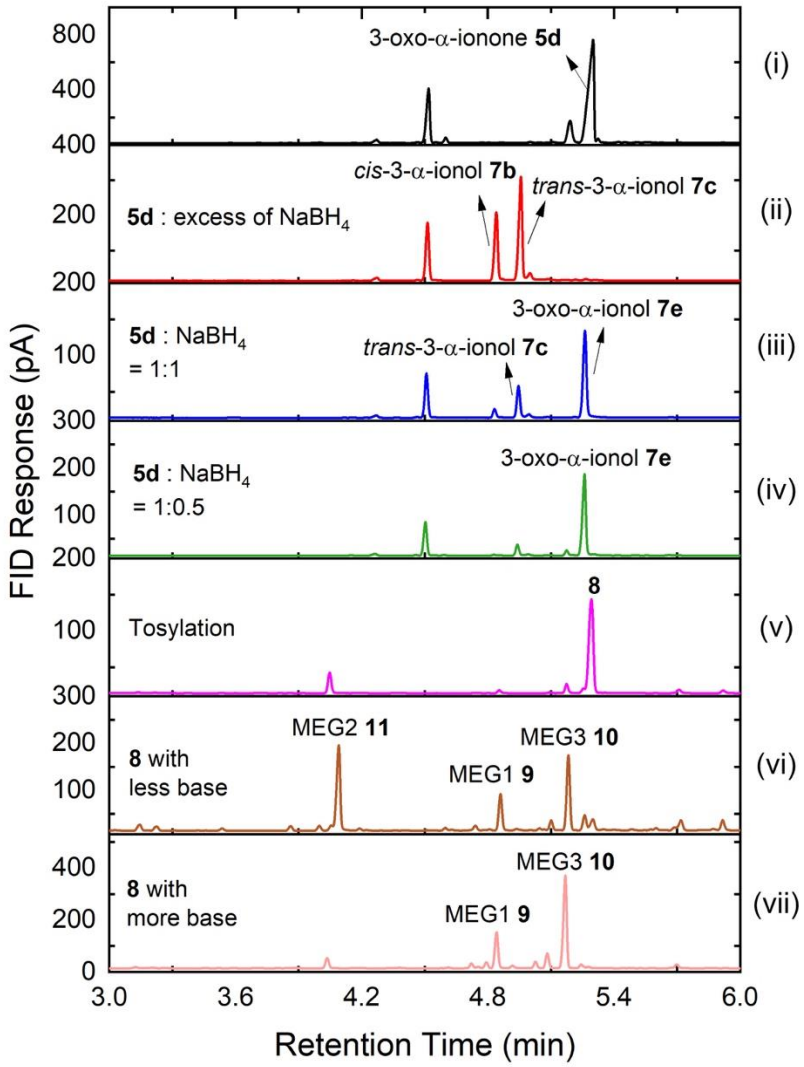


Figure 3.32 The GC profile of synthesis of tabanone from α-ionone in route 1.

The subsequent tosylation of **7e** was conducted in pyridine at 55 ºC (Figure 3.32-v) followed by base treatment of **8** with 20% NaOH, which gave 34% of the terminal alkene MEG2 **11**,

13% of MEG1 **9** and 27% of MEG3 **10** (Figure 3.32-vi). Higher concentration of NaOH (50%) was then added to the mixture; **11** was fully converted to **9** and **10**, which constituted 28% and 45% (by GC) of the crude mixture (Figure 3.32-vii). This suggested that the terminal alkene **11** could generate **9** and **10** under alkaline conditions.

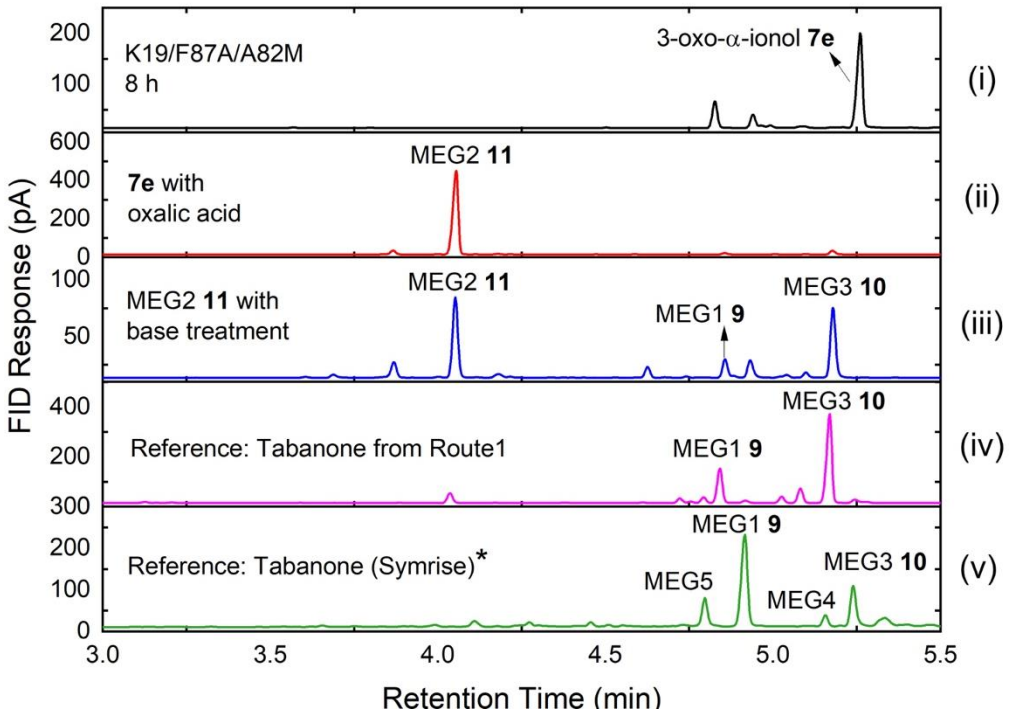


Figure 3.33 The GC profile of synthesis of tabanone isomers **9** and **10** from α-ionone **5** in route 2. The * indicates a retention time shift due to the change of column.

In route 2 (green panel in Figure 3.31), α-ionone **5** was reduced to α-ionol **7** by $NaBH_4$. A preparative scale oxidation of α-ionol **7** (194 mg, 3000:1) with K19/F87A/A82M was performed. **7** was fully converted to a mixture of 3-hydroxy-α-ionol **7b** and **7c** after 4 h, which were then further oxidised to the 3-oxo-α-ionol **7e** for another 4 h, which was monitored using GC and TLC (Figure 3.33-i). The subsequent acid treatment of **7e** was conducted with oxalic acid for 6 h stirring at 75 ºC with the addition of molecular sieves, leading to a single product MEG2 **11** (Figure 3.33-ii). After work up, compound **11** was treated with 50% NaOH and 55% of **11** was converted to a mixture of tabanone isomers with MEG3 **10** being the major product

(55% selectivity, Figure 3.33-iii). This acid treatment only promoted one single product **11** with relatively clean background but gave lower conversion to tabanone isomers compared with route 1 (Figure 3.33-iv).

3.3.7 Tabanone synthesis summary

The chemoenzymatic synthesis of tabanone isomers **9** (MEG1-*Z*/*E*), **10** (MEG3-*E*/*E*) and **11** (terminal alkene MEG2) from α-ionone and α-ionol has been established either using a $P450_{BM3}$ variant or an alcohol dehydrogenase to produce the desired precursor ketone, followed by dehydration by acids or bases. The reaction conditions remain to be optimised for large scale production of tabanone with high yield.

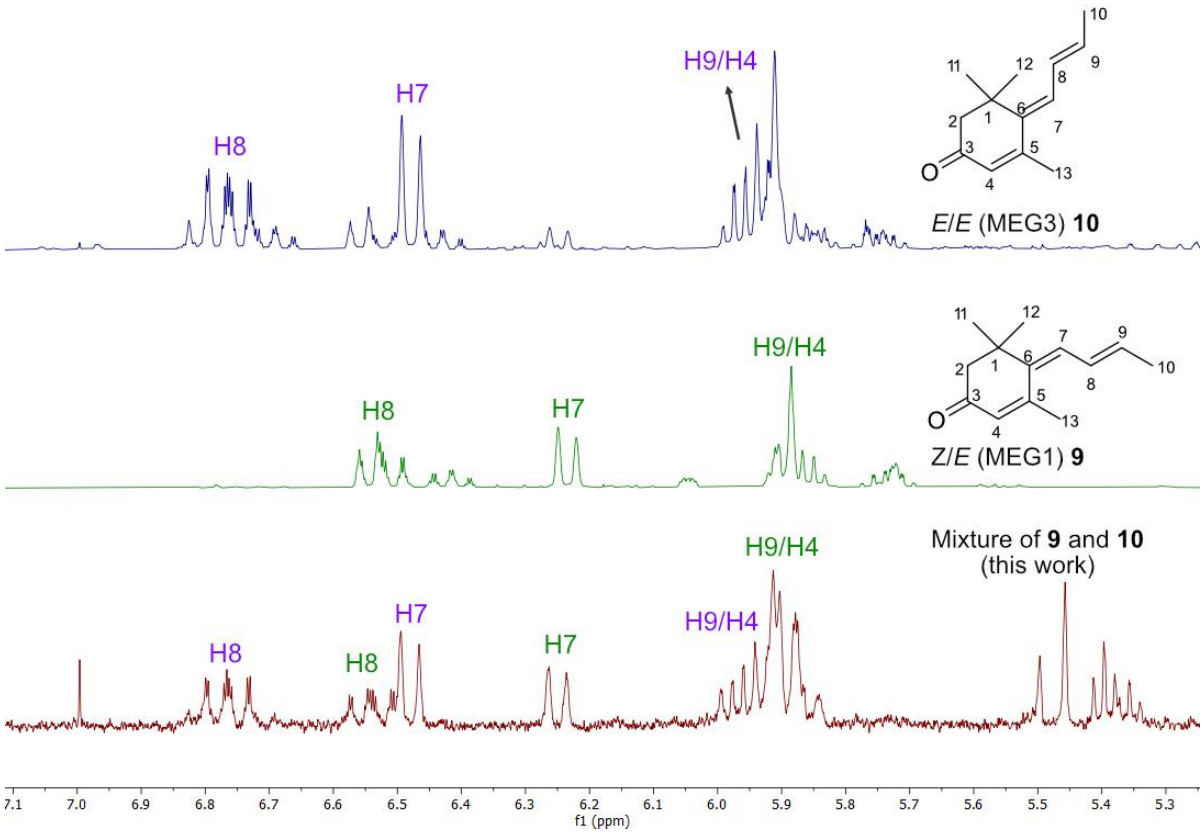


Figure 3.34 Comparison between the ^{1}H NMR spectra of the commercial and chemoenzymatic synthesised tabanone isomers **9** and **10**; the *E*/*E* (MEG3) **10** is in purple and the *Z*/*E* (MEG1) **9** is in green.

The pure isomer **11** possesses a sweet and herbal note while the mixture of **9** and **10** isomers has a herbal, earthy and tabacco like aroma. The aroma of the synthesised tabanone mixture **9** and **10** is different from the commercial sample from Symrise, probably due to different percentage of ingredients. Purer fractions of **9** and **10** were purified from a column of Symrise tabanone sample and were compared with the mixture of isomers **9** and **10** obtained from this work (Figure 3.34).

3.4 Rose oxide synthesis

Rose oxide is found in small amounts in some essential oils, such as Bulgarian rose oil and geranium oil. It contributes to the unique rose aroma and is one of the most important fragrance materials in perfumery. Rose oxide has four stereo isomers, the *cis*- and *trans*-rose oxide each with a (+)- and (–)-enantiomer (Figure 3.35). The (–)-*cis* isomer is responsible for the floral green and strong diffusive rose scent with an odor threshold of 0.5 ppb. Its enantiomer, (+)-*cis*-rose oxide has a herbal, green floral and earthy odor with a threshold of 50 ppb. The *trans* isomers are less rose-like; (+)-*trans*-rose oxide is described as having a herbal green, citrus scent with an odor threshold of 80 ppb, while (–)-*trans*-rose oxide displays floral green and minty notes with an odor threshold of 160 ppb.

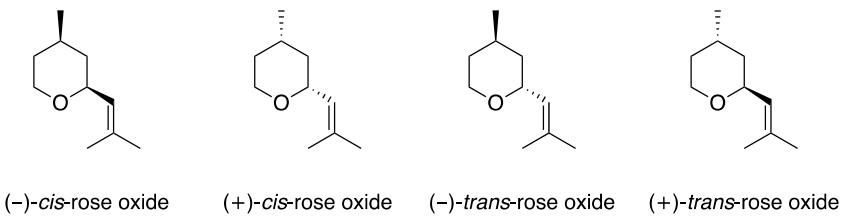


Figure 3.35 The four isomers of rose oxide.

Rose oil was originally obtained by careful steam distillation of freshly picked flowers of *Rosa damascene*. Due to the low concentrations, it required 3–4 tonnes of flowers to produce 1 kg of

rose oil. Therefore, it is one of the most expensive essential oil that is widely used in perfumes and household products. Synthetic racemic rose oxide is often used to impart a diffusive effect for its characteristic scent of roses. Considering its low concentration in nature, rose oxide is almost never isolated from the essential oils. Instead, pure rose oxide is only produced by chemical synthesis. A four-step synthesis of rose oxide from crotonaldehyde by a Grignard reaction followed by Claisen rearrangement, reduction and cyclisation was reported (Figure 3.36).[221] When the precursor alcohol **a3** was exposed to iodine, a mixture of products containing 42% racemic rose oxide and 25% of the iso-propenyl isomer was formed; the latter may be equilibrated to provide more rose oxide by stronger acids at higher temperature.

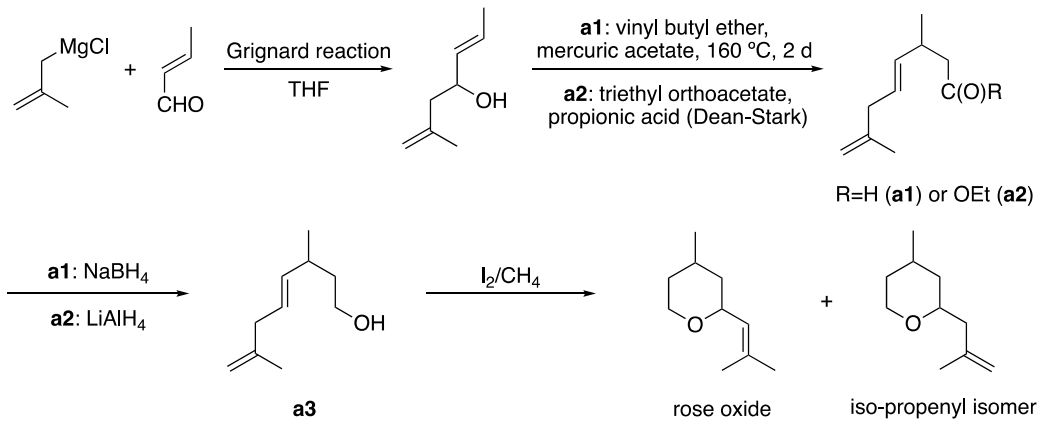


Figure 3.36 Total synthesis of rose oxide from crotonaldehyde by Grignard reaction.

The high cost and high impact of rose oxide has stimulated interest in its synthesis from citronellol, the direct biological precursor to rose oxide. The Dragoco (now Symrise) process utilised bengal rose-photosensitised oxidation of citronellol in methanol followed by sodium sulfite reduction, which gave a mixture of two allyl alcohols **p1** and **p2** in a ratio of 40:60 (Figure 3.37).[222] Alcohol **p2** underwent acid catalysed cyclisation to a mixture of *cis*- and *trans*-rose oxide in a ratio of 9:1.

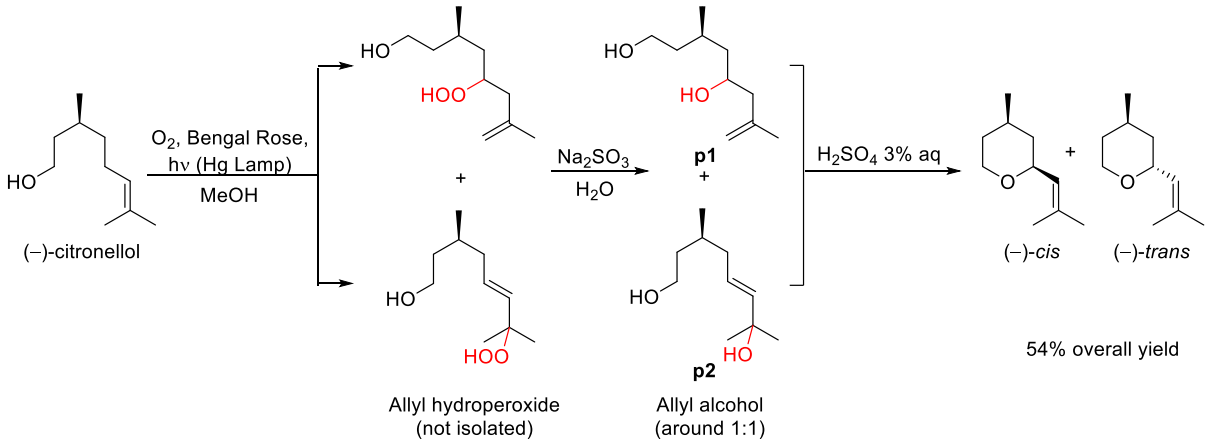


Figure 3.37 Synthesis of (–)-rose oxide (*cis*- and *trans*-) through photooxygenation of citronellol (Dragoco Patent).

This process depended on the selection of appropriate citronellols, (–)-citronellol gave the (–)-rose oxides, while racemic citronellol gave racemic products. Compound **p1** is a by-product that formed the iso-propenyl isomer in Figure 3.36, leading to low overall yield.

In order to optimise the Dragoco protocol, different conditions and reagents were used in the oxidation step to improve the percentage of **p2** in the synthesis (Figure 3.38).[223] In route 2, the Schenck ene oxygenation step was carried out in a microreactor equipped with LEDs.[224-226] The application of microreactors for photooxygenation reactions diminished the hazard related to the use of highly explosive chemicals and provided wider range of emission wavelength light sources for various reactants.[227] However, the long average irradiation time of 65 h gave the same ratio of **p1** and **p2** in the second step.

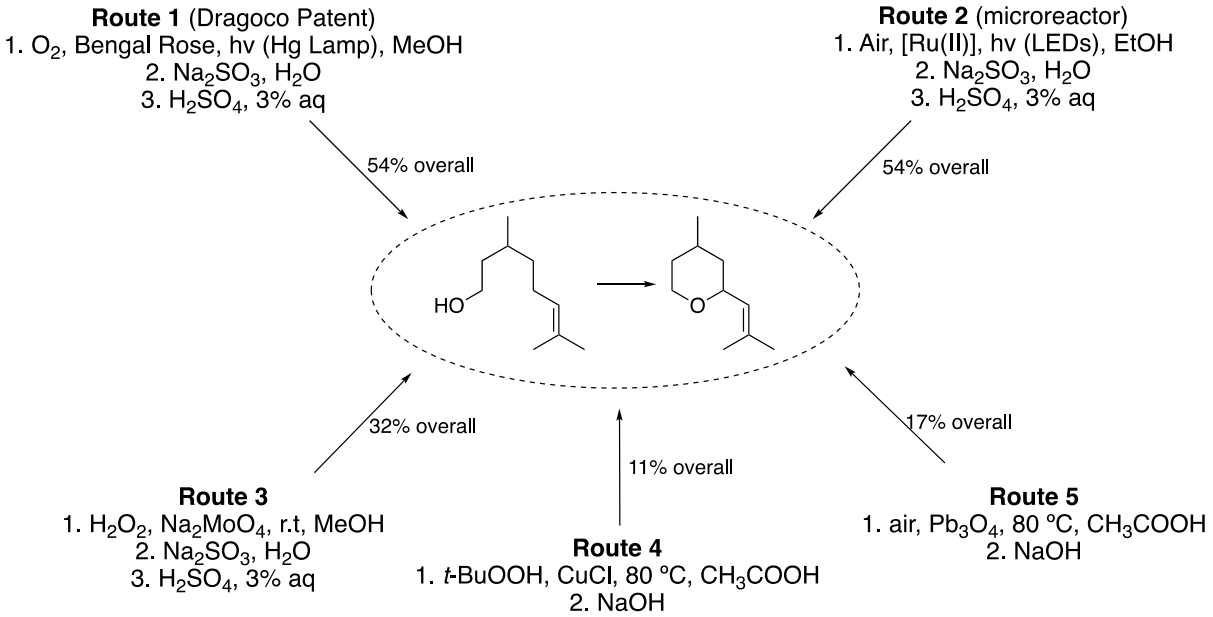


Figure 3.38 Selected routes for the synthesis of rose oxides.

Other than photooxygenation, the alternative approach (route 3) to generate the 'dark' singlet oxygen was via molybdate-catalysed disproportionation of hydrogen peroxide.[228] Singlet oxygen was generated in 71% yield and used for the oxygenation of citronellol, giving an equimolar mixture of **p1** and **p2**. Routes 4 and 5 provided alternative reagents to generate singlet oxygen involving *t*-butyl hydroperoxide and Pb_3O_4 with acetic acid as solvent.[229] The overall yield for these routes were lower due to the lack of selectivity for **p2**.

A biocatalyst such as a P450 enzyme offers a green alternative to these chemical synthesis routes. Enzymes also have the potential advantages of high selectivity via protein engineering. Two proposed routes for the synthesis of rose oxide are shown in Figure 3.39. In route A, the primary alcohol in (−)-β-citronellol is protected followed by screening with $P450_{BM3}$ variants for the C5 hydroxylation product, which would be treated with base to form rose oxide. In route B, (−)-β-citronellol is directly screened against $P450_{BM3}$ variants for C5 hydroxylation, followed by acid treatment for the cyclisation to (−)-rose oxide.

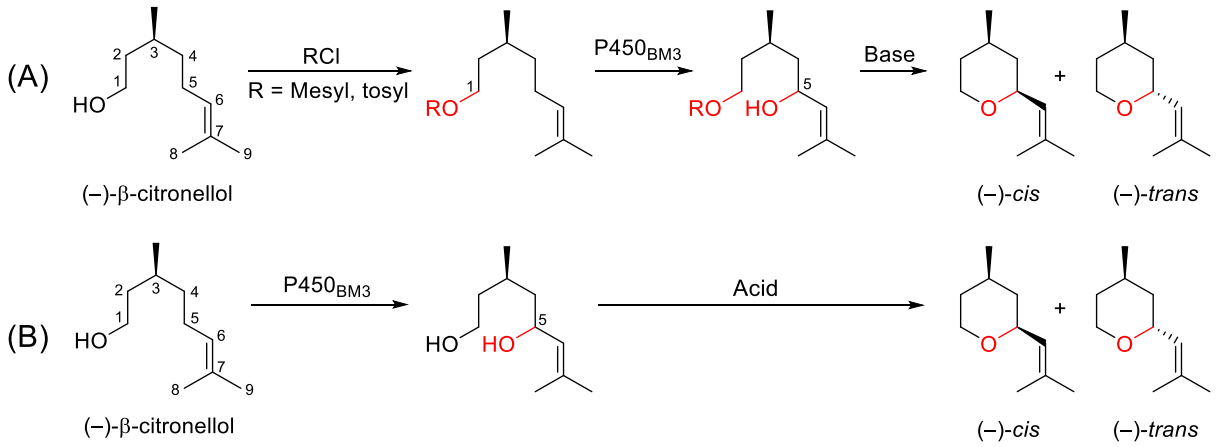


Figure 3.39 Proposed synthesis routes of *cis*- and *trans*- (–)-rose oxide from (–)-β-citronellol.

3.4.1 Route A – Protection of citronellol and oxidation by P450$_{BM3}$

***Mesylation*:** The protection of citronellol by methanesulfonyl chloride (MsCl) was conducted following the protocol in Appendix 7 and the NMR data of product **14** was in agreement with literature data.[230] A panel of 48 CYP102A1 variants (Appendix 7) were screened against the mesylated citronellol (5 mM substrate and 2 μM enzyme); 37 variants showed >50% conversion. The variants gave 6 major products, 4 of which were purified and characterised as the 6,7-epoxide (**14a**), 3-hydroxy-mesylated citronellol (**14b**), a ring product (**14c**), and 3,9-dihydroxy-mesylated citronellol (**14d**) (Figure 3.40). Products **14a**–**14d** are new metabolites for mesylated citronellol **14** and the full characterisation data is collected in Appendix 7. The longer retention times of the two uncharacterised products **u1** and **u2** suggested these to be di-hydroxylation products of **14**.

Epoxidation of citronellol was the predominant reaction in the screening, with 42 variants giving >50% selectivity for **14a**. A few variants giving high selectivity and turnover number for **14a** are shown in Table 3.4 (Entries 1–3). The C3-hydroxylation product **14b** was another major product, which was also the precursor of products **14c** and **14d**. Comparing with the

template variant RP/H171L which mainly gave the epoxide **14a**, the additional I263G mutation shifted the selectivity towards **14b** (Entry 5). K19/F87A/A328I/A184I also gave small amounts of **14b**, but more of **14c**, the cyclisation product originating from **14b** in a proposed mechanism shown in Figure 3.41. Product **14c** was favoured by variants which contained the F87A/A328I combination (Entries 6–8).

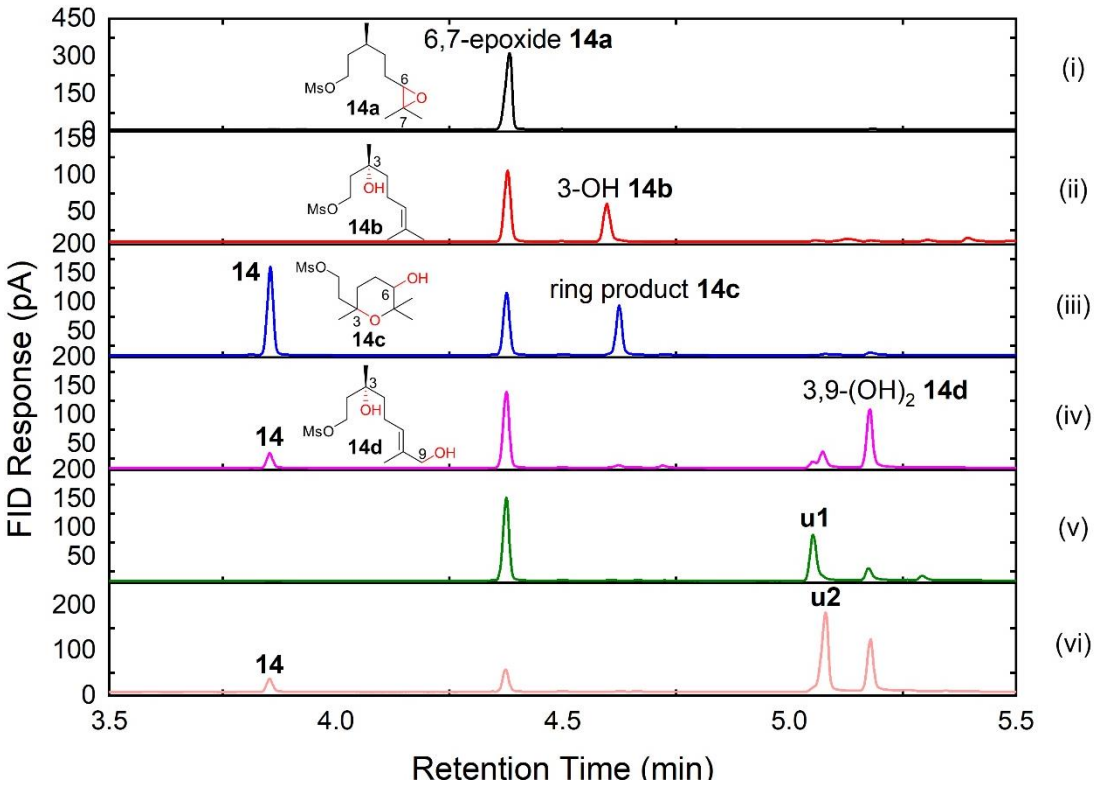


Figure 3.40 Gas chromatographic analysis of selected variants giving mesylated citronellol oxidation products. (i) 97% of the 6,7-epoxide **14a** by RP; (ii) 30% of 3-hydroxy-mesylated citronellol **14b** by RP/H171L/I263G; (iii) 43% of the cyclisation product **14c** by K19/F87A/A328I/V78I; (iv)) 36% of 3,9-dihydroxy-mesylated citronellol **14d** by RT2/S72W/A330W; (v) 34% of **u1** by RP/F87V/E267V/V78F; (vi) 51% of **u2** by RT2/F81W/A328I.

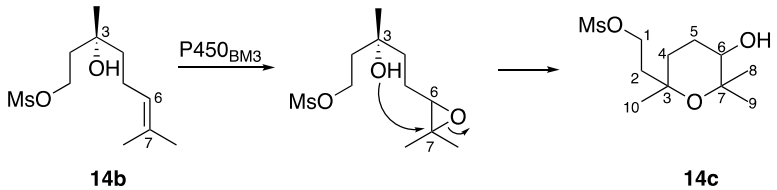
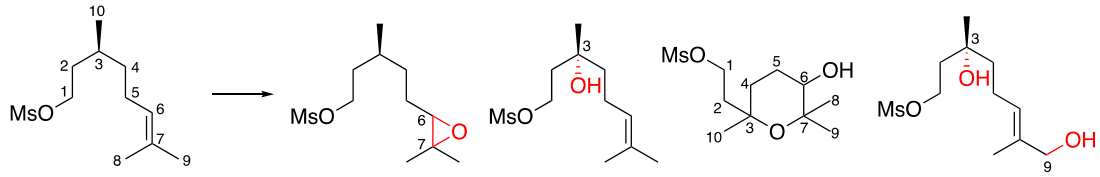


Figure 3.41 Proposed mechanism for the formation of product **14c**.

Table 3.4 Substrate conversion and product selectivity of selected variants for the oxidation of the mesylated citronellol, **14**.[a]



Mesylated citronellol, **14** **14a** **14b** **14c** **14d**

Entry	Variants	14a	14b	14c	14d	u1	u2	Others	Conv.	TON[b]
1	RP	**97%**						3%	100%	2425
2	GVQ	**96%**						4%	100%	2400
3	RP/H171L	**96%**						4%	100%	2400
4	K19/F87A/A184I/A328I	45%	**11%**	25%	8%	5%	5%	1%	100%	275
5	RP/H171L/I263G	53%	**30%**				5%	12%	96%	720
6	R19/F87A/A328I	51%		**33%**	9%			7%	94%	776
7	K19/F87A/A328I/Y51L	51%		**35%**	8%	4%		2%	78%	683
8	K19/F87A/A328I/V78I	54%		**43%**	3%				58%	624
9	RT2/A330W	34%			**30%**	12%	18%	6%	99%	743
10	RT2/S72W/A330W	45%			**36%**		15%	4%	92%	828
11	K19/F87V/A32I	56%			14%	**25%**		5%	100%	625
12	RP/FV/EV/V78F	53%			9%	**34%**		4%	100%	850
13	RT2/A330P/A82M	43%			23%	6%	**24%**	4%	89%	534
14	RT2/F81W/A328I	12%			32%		**51%**	5%	93%	1186

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 5 mM mesylated citronellol **14**, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold.

The 3,9-dihydroxy-mesylated citronellol **14d** was another further oxidation product of **14b**. RT2/A330W gave 30% of **14d** while the template variant RT2 showed 94% selectivity for **14a**, with no indication of C3-hydroxylation (Entry 9). The additional S72W mutation slightly improved the selectivity and turnover number for **14d** (Entry 10). Mesylated citronellol was a good substrate for $P450_{BM3}$ variants; C3, C6, C7 and C9 oxidation occurred with high turnover number. However, the C5-alcohol, the precursor to rose oxide, was not observed.

***Tosylation*:** Protection of citronellol by 4-toluenesulfonyl chloride was conducted following the protocol in Appendix 8.[231] Tosylated citronellol (**15**) was screened with the same set of 48 $P450_{BM3}$ variants (5 mM of substrate and 2 μM enzyme) and only 17 showed >50% conversion.

The full activity and selectivity data are given in Appendix 8. The lower reactivity of **15** may be due to the aromatic ring that resulted in lower aqueous solubility. The screening showed 5 major products which were characterised as the 6,7-epoxide (**15a**), 9-hydroxy-tosylated citronellol (**15b**), 9-oxo-tosylated citronellol (**15c**), 8/9-hydroxy-dihydro-tosylated citronellol (**15d**) and the dihydro aldehyde **15e** (Figure 3.42). Products **15a**–**15e** are new metabolites of tosylated citronellol **15** and the full characterisation data are collected in Appendix 8.

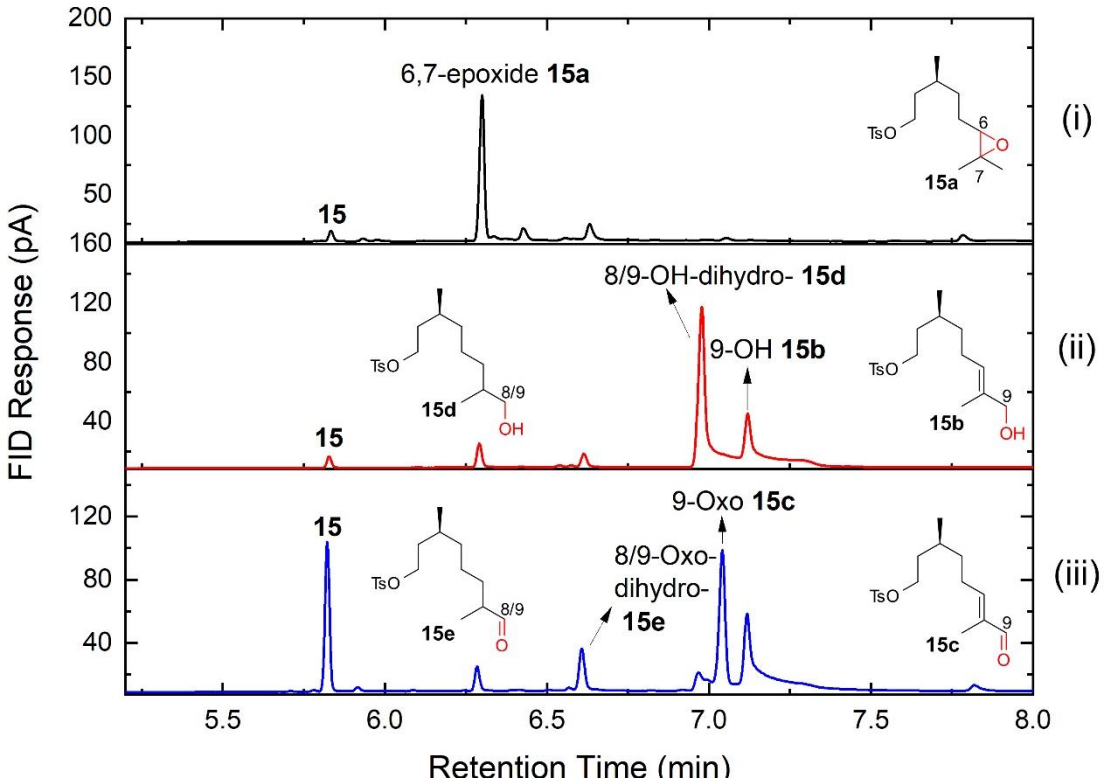
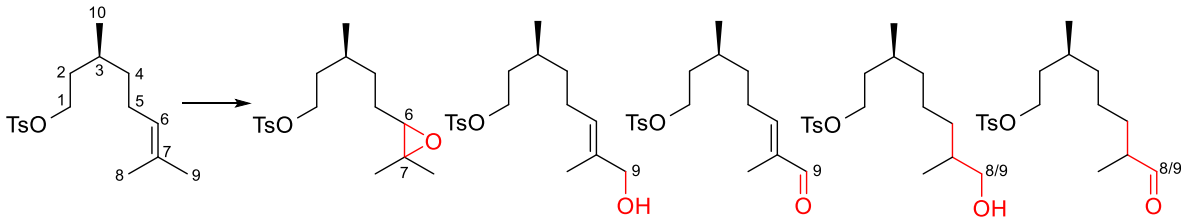


Figure 3.42 Gas Chromatographic analysis of selected variants giving tosylated citronellol oxidation products **15a-15e**. (i) 72% of the 6,7-epoxide **15a** by RT2/P329V/A330P; (ii) preparative scale reaction with R19/F87A/A328I gave 22% of 9-hydroxy-tosylated citronellol **15b** and 51% of the 8/9-hydroxy-dihydro tosylated citronellol **15d**; (iii) preparative scale reaction with K19/F87V/A328I gave 36% of the 9-oxo-tosylated citronellol **15c** and 9% of the 8/9-oxo-dihydro tosylated citronellol **15e**.

As observed with mesylated citronellol, the epoxide of tosylated citronellol **15a** was one of the predominant products, with 23 variants showing >50% selectivity. Variants giving high selectivity and turnover number of **15a** are shown in Table 3.5 (Entries 1–4). Another major product was the C9 allylic alcohol **15b**, which was favoured by the A328I mutation (Entries

5–9). The further oxidation product of **15b**, the C9 aldehyde **15c**, was a minor product (Entries 10 & 11). The reduction of the double bond in **15b** and **15c** gave products **15d** and **15e**, most likely by the reductase domain of $P450_{BM3}$. Since non-conjugated double bonds are difficult to reduce enzymatically, the α,β-unsaturated aldehyde **15c** could be the precursor to **15e**, likewise for **15b** and **15d**. The mechanism by which these compounds is formed is not known.

Table 3.5 Substrate conversion and product selectivity of selected variants for the oxidation of the tosylated citronellol, **15**.[a]



Tosylated citronellol, **15** **15a** **15b** **15c** **15d** **15e**

Entry	Variants	15a	15b	15c	15d	15e	Others	Conv.	TON[b]
1	R47L/Y51F	**72%**	**5%**			11%	12%	79%	1422
2	RT2/P329V/A330P	**72%**	6%			10%	12%	91%	1638
3	K19/F87V/Q403P	**73%**	12%				15%	82%	1497
4	R19/F87A	**77%**					23%	70%	1348
5	K19/F87A/I263A/A328I	21%	**57%**			5%	17%	26%	371
6	K19/F87A/A184I/A328I		**65%**	5%		8%	22%	77%	1252
7	R19/F87A/A328I		**68%**				32%	86%	1462
8	K19/F87V/A328I	4%	**72%**	13%			11%	93%	1674
9	K19/F87A/Y51L/A328I	5%	**84%**				11%	78%	1638
10	RT2/F81W/A328I	5%	69%	**15%**	7%		4%	26%	98
11	RP/F87V/E267V	28%	36%	**15%**			21%	68%	850
12	RT2/F81W	47%	15%	5%		**13%**	20%	81%	264
13	RT2	70%	5%			**12%**	13%	94%	282

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 5 mM tosylated citronellol **15**, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold.

The 48-variant library oxidised mesylated citronellol to the epoxide and C3-alcohol as major products while tosylated citronellol gave the C9-alcohol and epoxide as the predominant

products. Neither of these two protected citronellol gave the C5 oxidation product, the precursor to rose oxide.

3.4.2 Route B – Direct oxidation of citronellol by $P450_{BM3}$

A library of 48 variants was screened against (–)-β-citronellol (**16**, 5 mM substrate and 2 μM enzyme). Full data are given in Appendix 9. Citronellol was oxidised to a similar range of products to those observed with the protected derivatives, including the epoxide **16a**, the cyclisation product **16b**, and the C9 and C8 oxidation products **16d** and **16e**.[232] However, the C5 alcohol **16c**, the precursor to rose oxide, was formed (Figure 3.43). **16a**, **16b** and **16c** are new metabolites of citronellol and the full characterisation data are provided in Appendix 9. The NMR data for **16d** and **16e** are congruent with literature data.

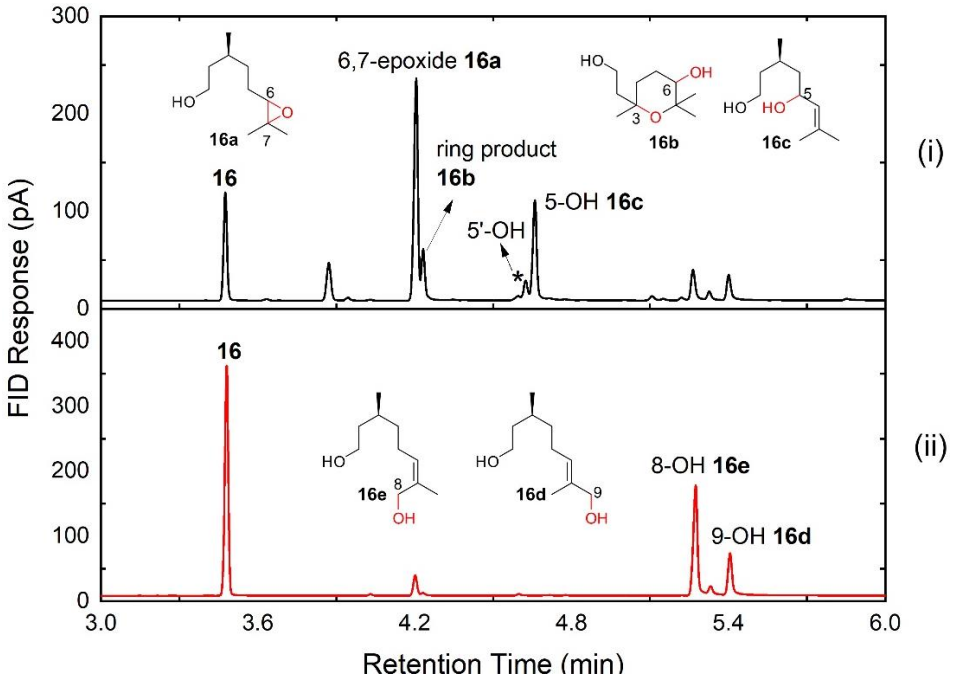
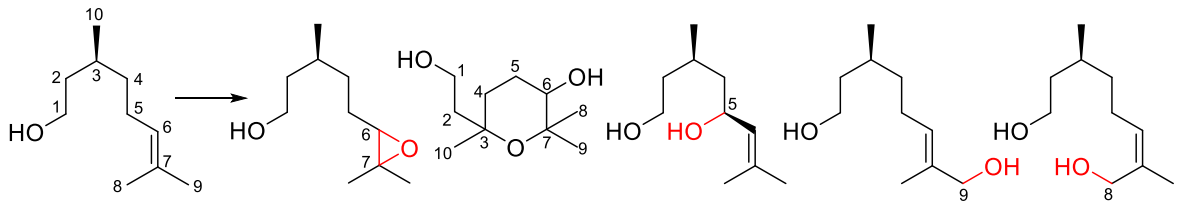


Figure 3.43 Gas chromatographic analysis of selected variants giving citronellol oxidation products **16a**–**16e**. (i) RP/A82M/I263A gave 44% of the 6,7-epoxide **16a**, 10% of the ring cyclisation product **16b**, and 20% of 5-hydroxy-citronellol **16c**; (ii) RT2/F81W/A328I gave 25% of 9-hydroxy-citronellol **16d** and 57% of the 8-hydroxy-citronellol **16e**. The * indicates product that has not been fully characterised, the crude NMR suggested it is the diastereoisomer of 5-hydroxy-citronellol.

Of the enzyme panel, 36 out of 48 variants showed >50% selectivity for the epoxide **16a** (Table 3.6). Even though the C3 alcohol was not observed, the cyclisation product **16b** was likely

formed *via* the mechanism in Figure 3.41. The I263G mutation promoted the ring closure product **16b** (up to 57%) (Entries 7 & 8). Oxidation at the allylic positions C8 and C9 were observed, with C8 oxidation being more favoured (Entries 13–16). The C5 alcohol **16c** was formed in low selectivity. Among the few variants that gave **16c**, the A82M mutation was crucial (Entries 9–12), with the A82M/I263A and A82M/T260G combinations giving 20% and 21% selectivity for **16c**, respectively (Entries 10 & 11). These two enzymes would be good templates for further engineering to increase selectivity for **16c**.

Table 3.6 Substrate conversion and product selectivity of selected variants for the oxidation of citronellol, **16**.[a]



Entry	Variants	16a	16b	16c	16d	16e	Others	Conv.	TON[b]
1	R19/F87A	**91%**	**5%**				4%	88%	2002
2	K19/F87V/Q403P	**92%**	3%				5%	84%	1932
3	GVQ	**92%**	3%				5%	88%	2024
4	RP/F87V	**93%**	3%				4%	98%	2279
5	RT2/A184I/A330P	21%	**17%**	6%	26%	19%	11%	44%	187
6	RP/I263A/E267V	18%	**40%**	4%	7%	6%	25%	26%	260
7	RP/HL/I263G	3%	**42%**		12%	5%	38%	63%	662
8	R19/I263G		**57%**				43%	96%	1368
9	R19/A82M/A184I/T260G	70%	2%	**16%**			12%	99%	396
10	RP/A82M/I263A	44%	10%	**20%**	6%	7%	13%	84%	420
11	R19/A82M/T260G	67%		**21%**			12%	97%	510
12	R19/A82M/I263G/A264G/A328G	56%		**25%**			19%	21%	132
13	RT2/A330P/A184I	21%	17%	6%	**26%**	19%	11%	44%	286
14	RT2/A330P/A82M	22%		6%	**29%**	27%	16%	33%	240
15	RT2/A330P/V78I/A184I	34%	11%	5%	12%	**31%**	7%	56%	434
16	RT2/F81W/A328I	11%			25%	**57%**	7%	51%	727

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 5 mM citronellol **16**, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold.

3.4.3 Route B – Synthesis of rose oxide from citronellol

The second step of the synthesis was the cyclisation of 5-hydroxy-citronellol **16c**, leading to the formation of rose oxide. The first approach was to convert the primary alcohol group at C1 to a better leaving group using methanesulfonyl chloride and 4-toluenesulfonyl chloride followed by base treatment (Figure 3.44–A). However, selective protection of the primary alcohol was difficult to control, therefore, **16c** was treated directly with acid (Figure 3.44–B).

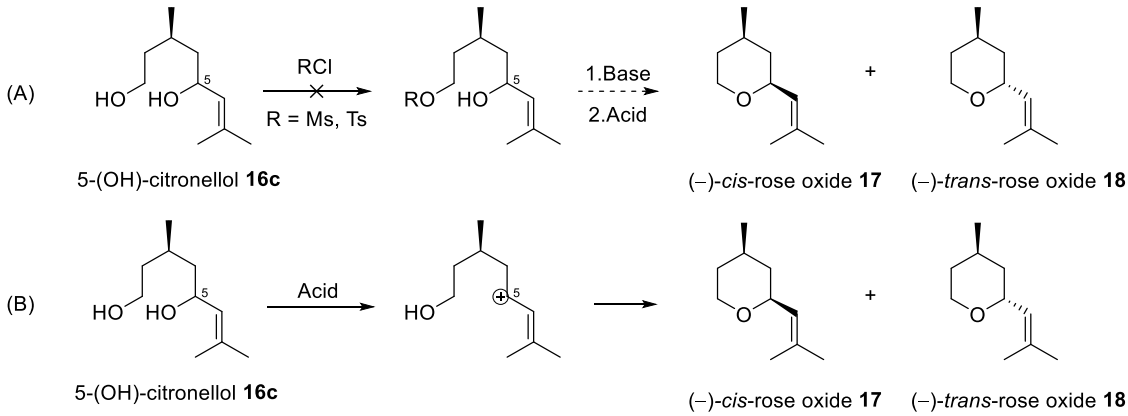


Figure 3.44 Two synthetic routes from the 5-hydroxy-citronellol (**16c**) to the mixture of rose oxide **17** and **18**.

A preparative scale oxidation of **16** (192 mg, 15000:1) with RP/A82M/I263A was performed. From 157 mg of crude extract, 5-hydroxy-citronellol (**16c**) was isolated and purified by the flash silica column chromatography (Figure 3.46–ii). 7 mg of **16c** was treated with 10% H_2SO_4 at room temperature for 1 h to give a mixture of (–)-*cis*-rose oxide **17** and (–)-*trans*-rose oxide **18** in a ratio of 7.5:2.5 (54% yield, Figure 3.46–iii).

In order to test the possibility of one-pot synthesis of rose oxide without work up, 50 mL of the reaction mixture was treated directly with 10% H_2SO_4 at room temperature for 3 h. The reaction crude contained 44% of **16a**, 10% of **16b**, and 20% of **16c** (Figure 3.46–iv). Acid treatment gave a mixture of **17** and **18** (3.4 mg, 20% yield) in a ratio of **17** to **18** of 9:1 (Figure 3.46–v).

The acidification of epoxide **16a** firstly gave two products **16a.1** and **16a.2** by epoxide ring opening via nucleophilic attack by the C1−OH (Figure 3.45). It appeared that these were turned into water soluble products after 3 h, leaving the mixture of rose oxides which were extracted into ethyl acetate. Further work will be required to understand why different ratios of *cis*- and *trans*-rose oxide **17** and **18** (by GC) were obtained (Figure 3.46–iii & v) under the different conditions.

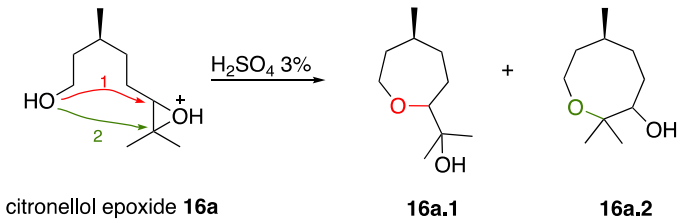


Figure 3.45 Acid treatment of the citronellol epoxide **16a**.

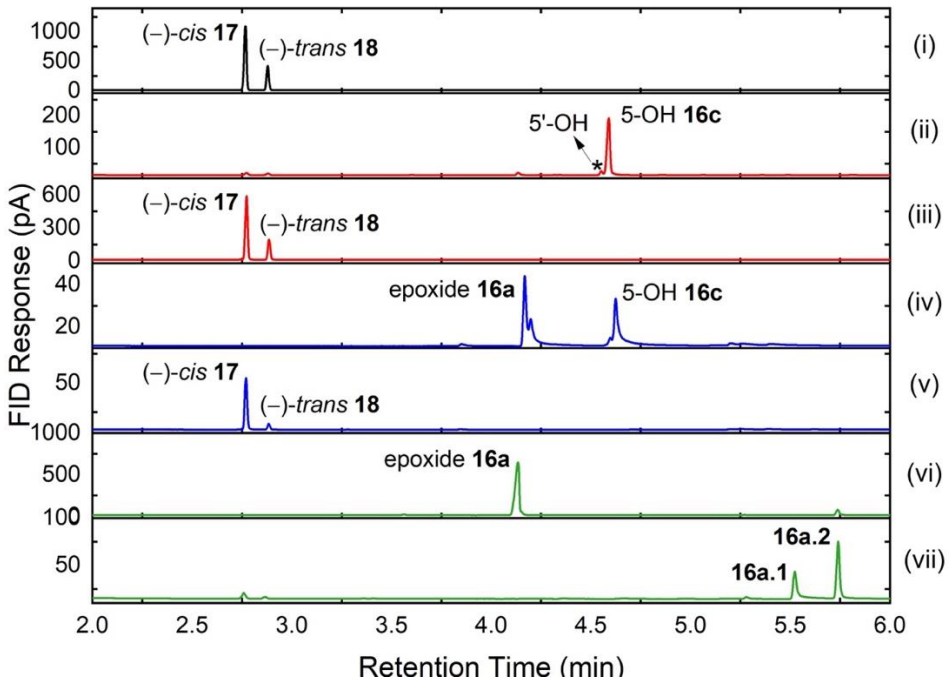


Figure 3.46 Gas chromatographic analysis of rose oxide formation. (i) Commercial sample of rose oxide, a mixture of the (–)-*cis*-rose oxide **17** and (–)-*trans*-rose oxide **18** in 72:28 ratio; (ii) 95% pure 5-hydroxy-citronellol **16c** separated from flash silica column chromatography. The * indicates product that has not been fully characterised due to insufficient material, the crude NMR spectrum suggested it is the diastereoisomer of **16c**. (iii) Pure **16c** was treated with 10% H_2SO_4 and gave a mixture of **17** and **18** in 75: 25 ratio. (iv) 17.2 mg of a mixture of the epoxide **16a** and 5-hydroxy-citronellol **16c**. (v) The reaction mixture in (iv) was treated with 10% H_2SO_4 for 2 h at RT, giving a mixture of **17** and **18** in 9:1 ratio; (vi) pure epoxide **16a**; (vii) **16a** was treated with 3% H_2SO_4 for 0.5 h and gave products **16a.1** and **16a.2**.

3.4.4 Rose oxide synthesis summary

The synthesis of (–)-*cis*-rose oxide from citronellol has been demonstrated on a small scale (< 200 mg). The low yield (17%) for the first step was due to the low selectivity for the production of the precursor 5-hydroxy-citronellol **16c** (20%). Further protein engineering will be needed to improve the selectivity for **16c** for larger scale generation of rose oxide. Acid treatment of **16c** was well established and the isolated yield was 54% from purified **16c**, together with some unknown products. Further investigation is needed to establish the reasons for the difference *cis*/*trans* ratio of the two methods.

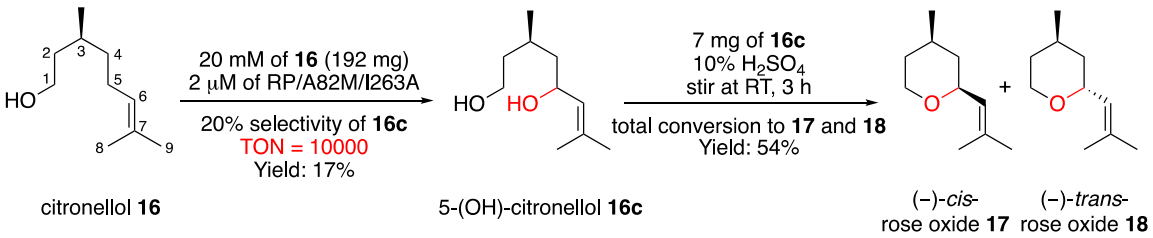


Figure 3.47 Scalability of rose oxide synthesis.

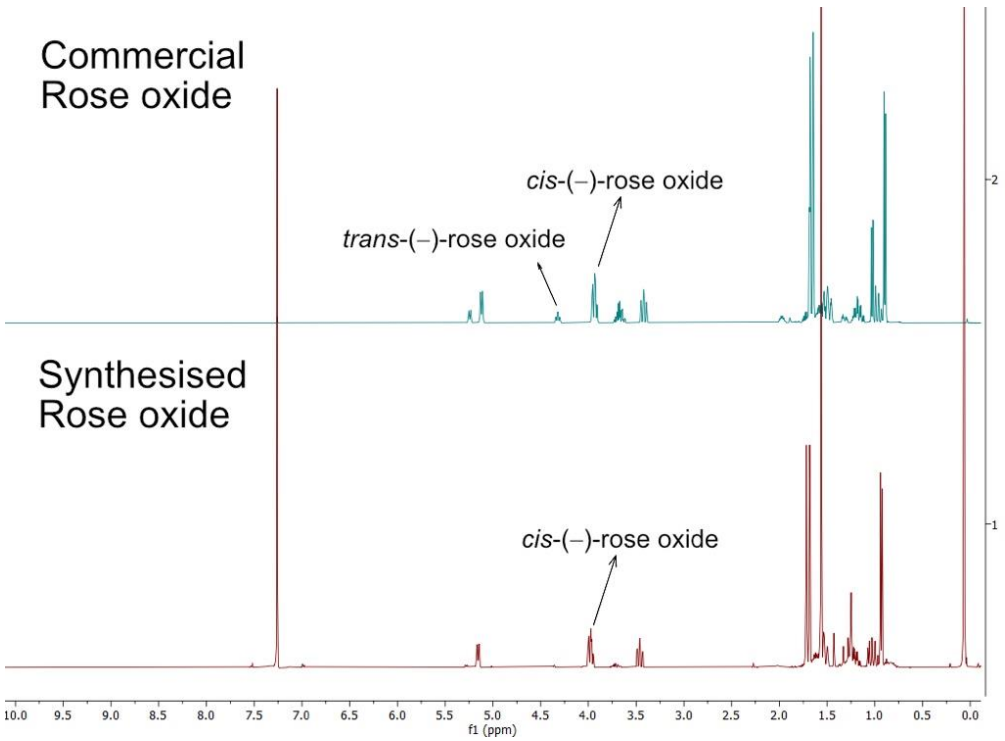


Figure 3.48 Comparison of the ^{1}H NMR spectra of commercial and chemoenzymatically synthesised rose oxide.

Chapter 4

4 Oxidation of androgens

4.1 Introduction of steroids

Steroids are important biological molecules which play important roles in human physiology as signalling molecules. Steroidal drugs are some of the most prescribed medicines that more than 100 FDA-approved steroidal drugs are used for treatment of a variety of conditions including inflammation, heart disease and cancer (Figure 4.1). The biological activity of steroids is differentiated by the oxygenation level and substituents of the four rings of the steroidal core and the nature of the side chain at C17.[233,234] Hydroxylation and general oxy-functionalisation of steroids increase their solubility and introduce diversity to their biological effects.

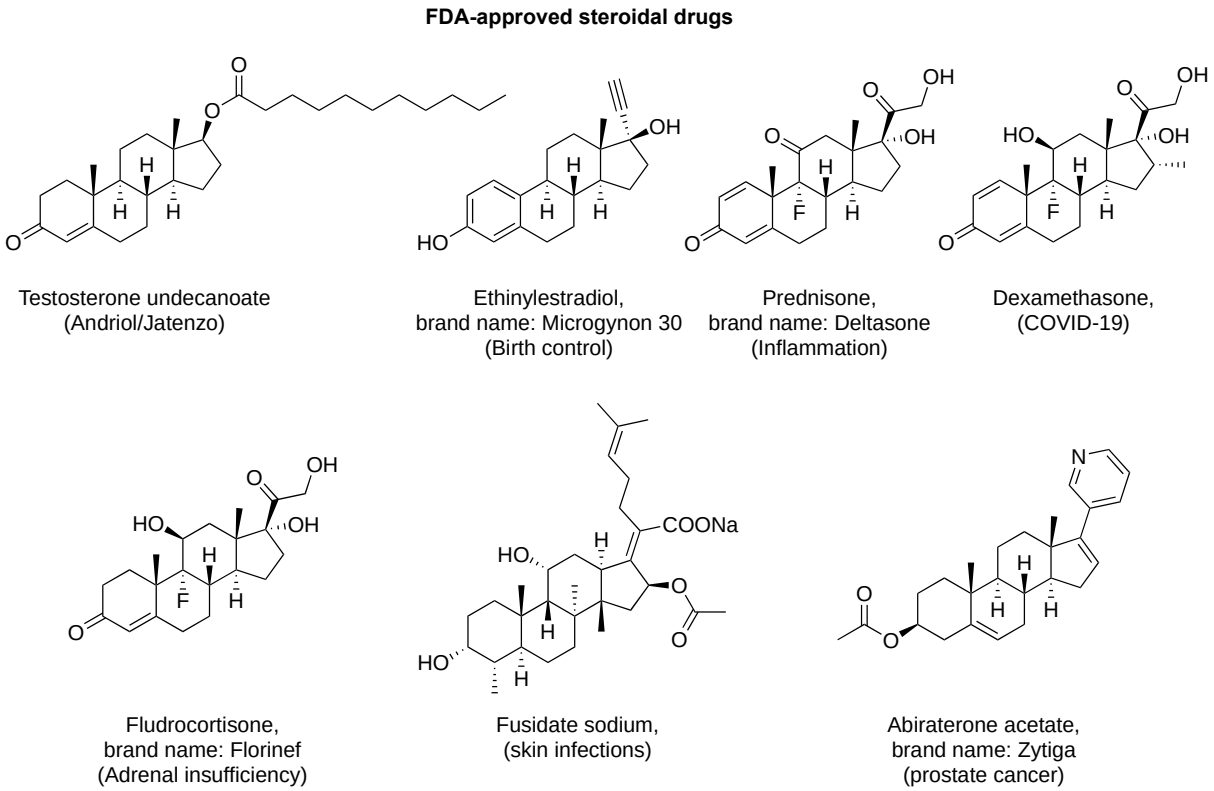


Figure 4.1 Examples of FDA-approved steroidal drugs.

For example, testosterone undecanoate is an anabolic, androgenic steroid that is used mainly in the treatment of low testosterone levels in men. Ethinylestradiol is an estrogen medication that is widely used in birth control pills in combination with progestin. Prednisone is a glucocorticoid medication that is mostly used to suppress the immune system and decrease inflammation. Dexamethasone and hydrocortisone are both glucocorticoids that have been hailed as a ground-breaking treatment for patients seriously ill with Covid-19.

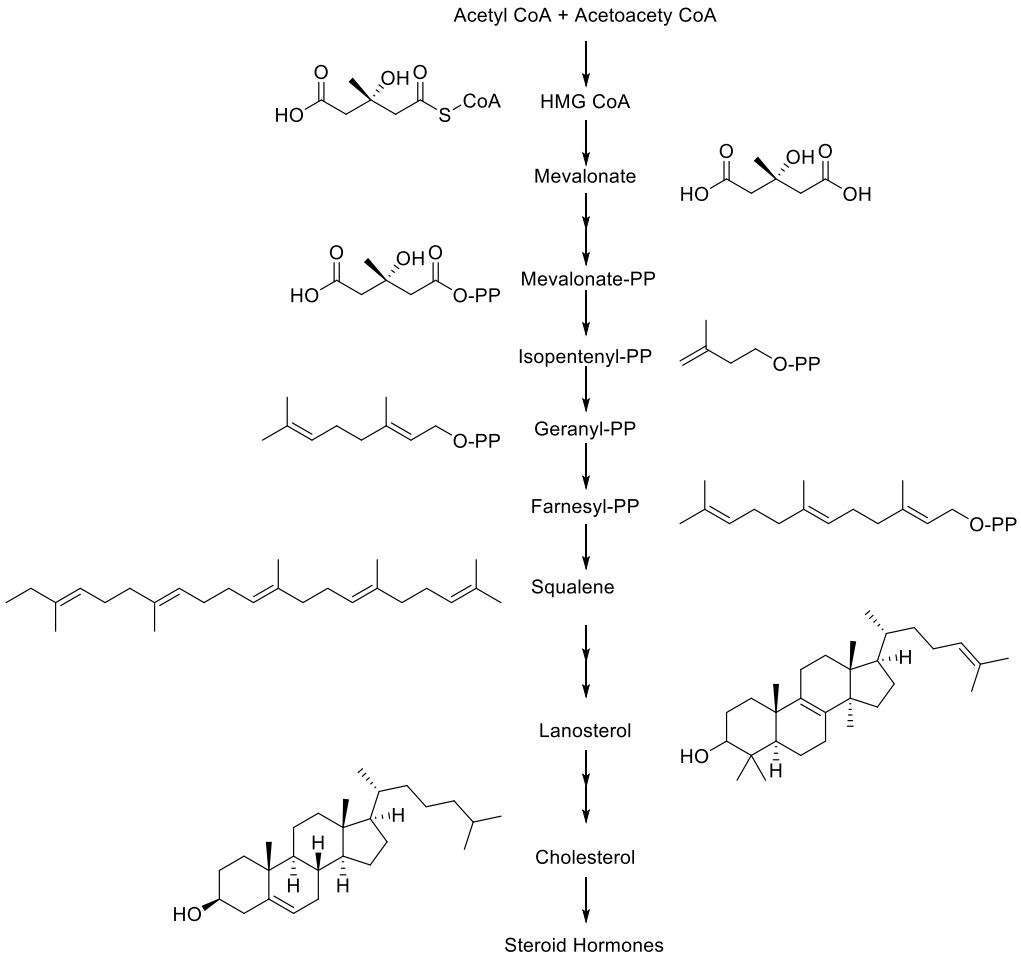


Figure 4.2 The mevalonate pathway and biosynthesis of cholesterol. CoA = Coenzyme A; HMG = 3-hydroxy-3-methylglutaryl; PP = pyrophosphate.

There has been a huge amount of research in the biosynthesis and function of steroids due to their crucial roles in human health. In humans, the biosynthesis of steroids follows the mevalonate pathway that uses acetyl CoA as building blocks to make dimethylallyl

pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP) in Figure 4.2.[235] DMAPP and IPP join to form geranyl pyrophosphate (GPP) which gives farnesyl pyrophosphate (FPP) with an additional unit of IPP. The coupling of two units of FPP gives squalene which is the precursor to lanosterol. Lanosterol is converted to cholesterol which is then modified into other steroids by steroidogenesis transformations (Figure 4.3).[236]

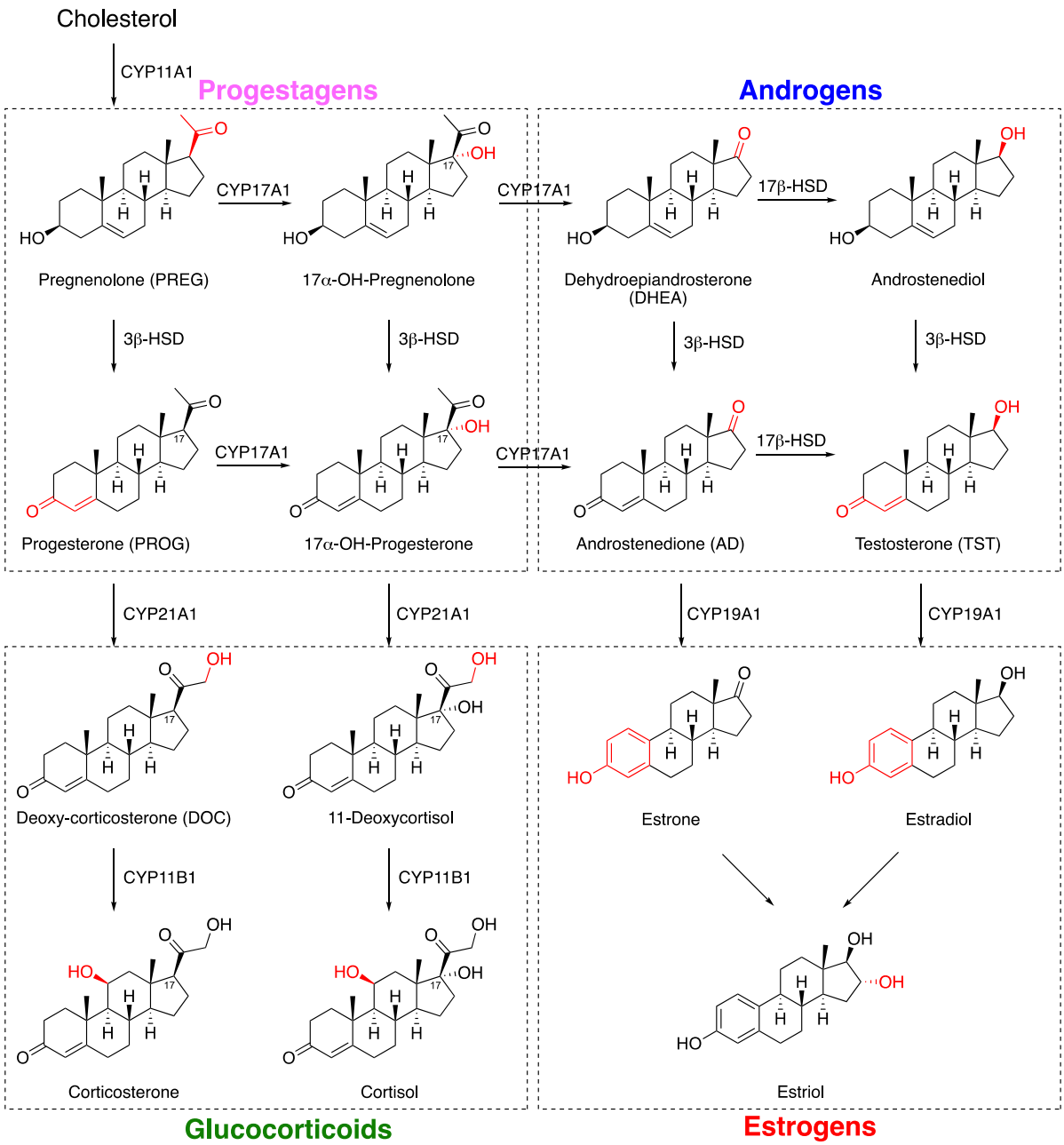


Figure 4.3 The steroidogenesis transformations catalysed by steroidogenic P450s and chemical structures of endogenous steroid hormones. 3β/17β-HSD: 3β/17β-hydroxysteroid dehydrogenase/isomerase.

In steroidogenesis, cholesterol is converted to 4 major classes of steroid hormones (Figure 4.3); the progestagens, androgens, estrogens and glucocorticoids.[237] Androgens such as androstenedione and testosterone regulate the male sexual function and behaviour,[238] while progestagens and estrogens are female sex hormones.[239] Glucocorticoids and mineralocorticoids are involved in a wide range of physiological processes including stress response, immune response, regulation of inflammation, sodium and water levels.[240] Difficiency in the production of sex hormones caused various diseases, so it is important to regulate these steroids hormones.[241] For example, congenital adrenal hyperplasia (CAH) results from mutations in the CYP11B1 gene which encodes a steroidogenic enzyme, 11β-hydroxylase.[242-244] Deficiency of 11β-hydroxylase prevents the conversion of 11-deoxycortisol to cortisol and deoxy-corticosterone to corticosterone (Figure 4.3), respectively, which results in high levels of androgenic steroid.[243,245] This inherited disorder due to low levels of cortisol and high levels of male hormones caused development of male characteristics in new-born females. Synthetic corticosteroids can mimic the actions of naturally occurring corticosteroids and may be used to replace steroids in people that are unable to produce sufficient amount of corticosteroids.[246] Therefore, synthetic steroidal drugs that have anti-inflammatory and immune-modulating properties are widely used to treat diseases such as asthma, arthritis and allergic or inflammatory conditions.

Total synthesis of steroids such as cholesterol and cortisone were reported by Woodward *et al.* using hydroquinone in 1951.[247,248] Later, Johnson and co-workers developed a total synthesis route for progesterone with a key step of stereoselective Wittig condensation of the aldehyde **a1** and ylide **a2** (Figure 4.4–A).[249] Pattenden also demonstrated a new approach to construct the steroid ring based on a novel radical cascade sequence to form a 4-ring system with 7 chiral centres (Figure 4.4–B).[250-252] However, the huge demand for steroidal drugs made the total

synthesis of steroids impractical due to the long synthetic pathways and low overall yield. Instead, all steroids are obtained by semisynthetic methods whereby natural steroids are used as starting materials to produce other novel compounds with distinct chemical and medicinal properties.[253-255] Marker and Rohrmann discovered a semisynthesis of progesterone in a better yield (~50%) via the degradation of diosgenin (Marker's degradation) which was extracted from Mexican yams (Figure 4.4–C).[256]

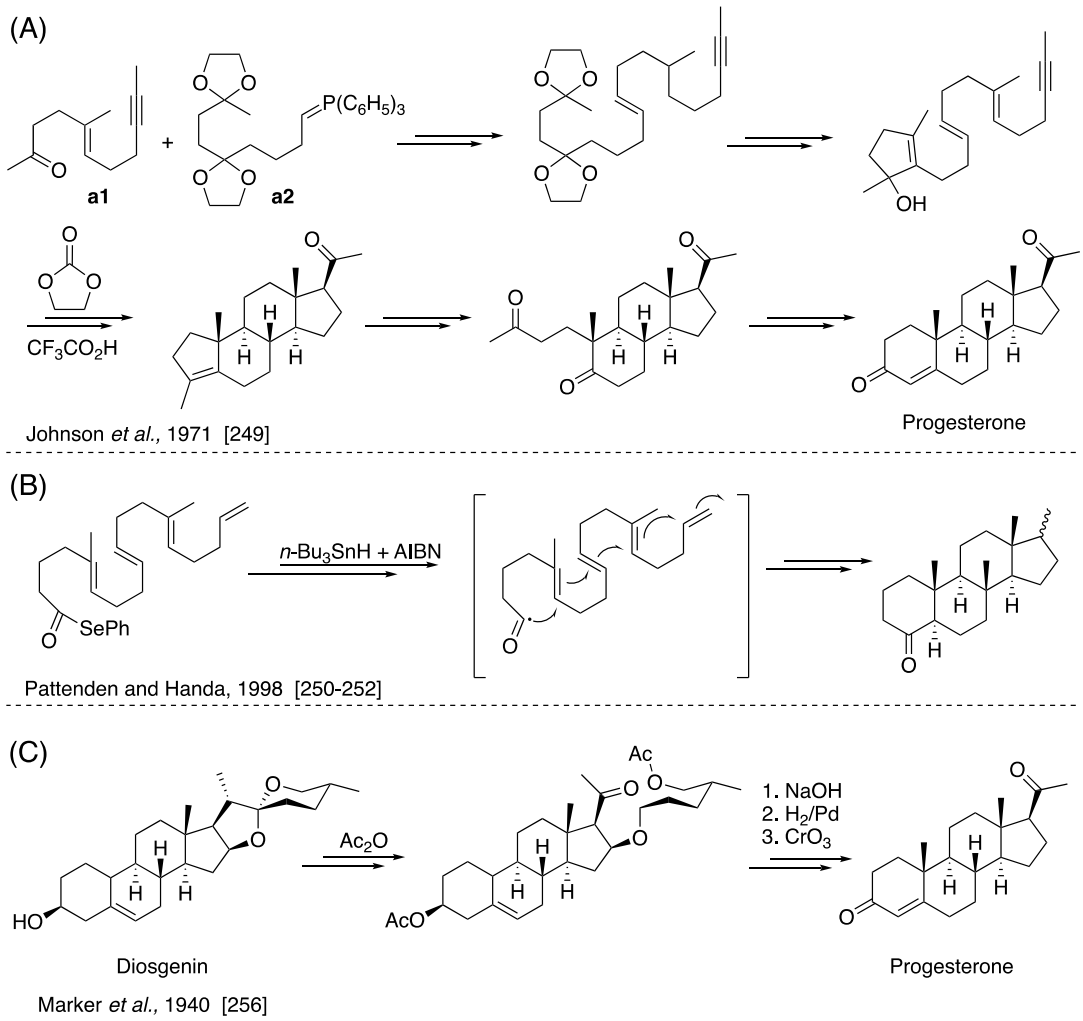


Figure 4.4 (A) Johnson's biomimetic total synthesis route of progesterone; **(B)** Pattenden's synthesis route of steroid skeleton; **(C)** Marker's degradation route from diosgenin to progesterone.

Given the medicinal value of the cardenolides in the treatment of congestive heart failure, a recent report of a scalable synthesis of ouabagenin, a bioactive cardenolide, from cortisone

would be another good example of semisynthesis.[257,258] More than 20 steps are developed in this semisynthetic route which contains numerous functional group conversions to achieve oxidation at specific sites with stereochemistry control (Figure 4.5). It appeared that it required 4 steps to insert an alcohol at C19 methyl group, 4 steps for the C1 and C5 hydroxylation, and then another 7 steps for the C14 oxidation with an inversion of the stereochemistry at C14.

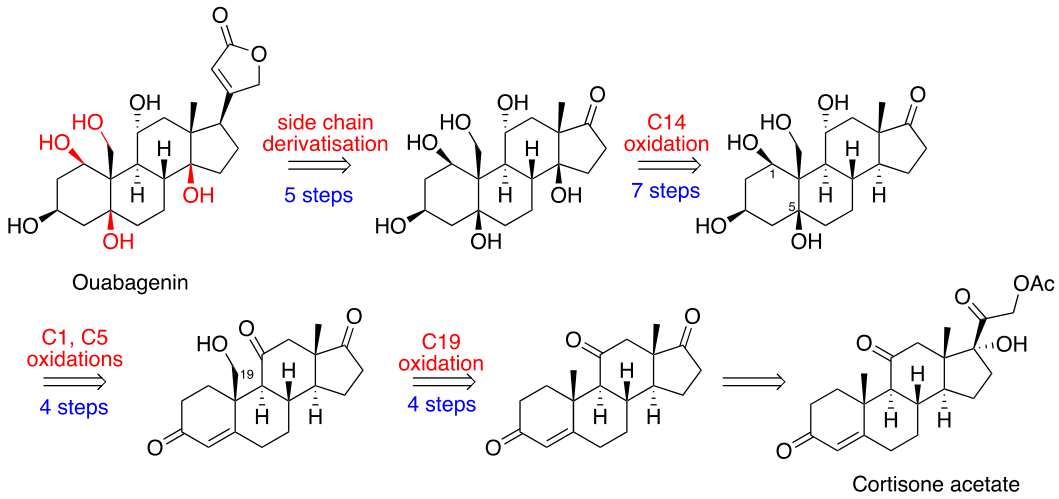


Figure 4.5: Retrosynthetic route of ouabagenin.[257,258]

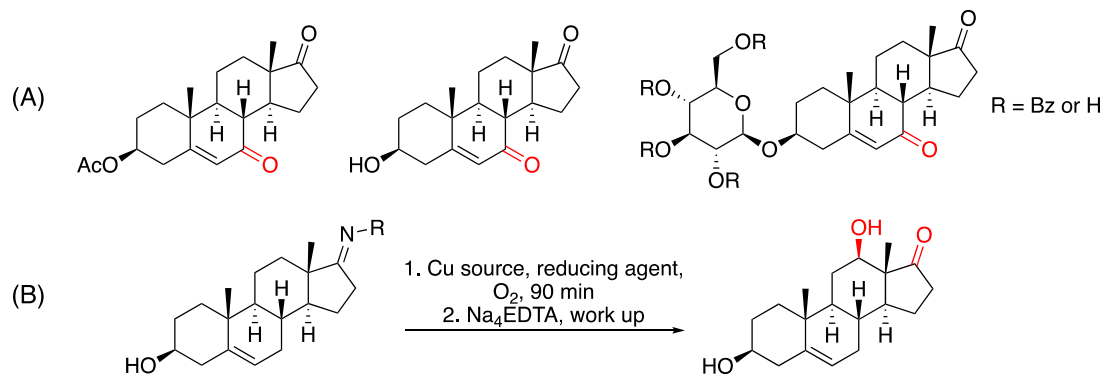


Figure 4.6: **(A)** Examples of allylic C–H oxidation of steroids by electrochemical methods.[259] **(B)** C12 oxidation of steroids with copper via Schönecker oxidation.[260]

Despite the successful demonstration of semisynthesis of ouabagenin, it also shows the difficulties in activation of C–H bonds in steroids due to the predominance of chemically inert sp^3 C–H bonds in steroids. Baran and co-workers have reported scalable electrochemical

oxidation of allylic C–H bonds of steroids (Figure 4.6–A).[259] Another example is the Cu-mediated oxidation at a carbon at the β-position from an alcohol directing group via the optimisation of Schönecker oxidation (Figure 4.6–B).[260]

Even though these approaches provided C–H bond functionalisation of steroids, the oxidation sites are limited. Cytochrome P450 enzymes, on the other hand, catalyse hydroxylation in steroid biosynthesis in microorganisms, giving a wide range of products. Steroid metabolism by human cytochromes P450 has been studied, both with recombinant P450s and in human liver microsomes and mitochondria.

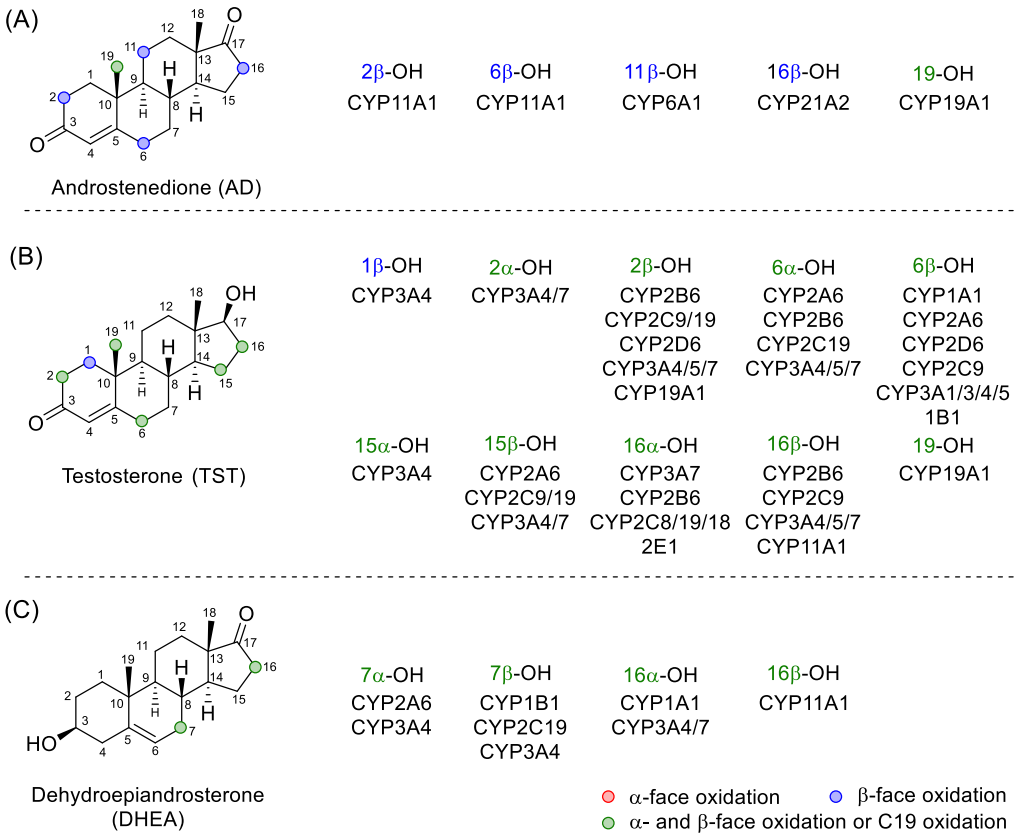


Figure 4.7 Summary of oxidation products of AD, TST and DHEA by various human P450s.

It has been shown that CYP1A2, 2A6, 2B6, 2C8, C29, 2C19, 2D6, 2E1, 3A4, and 3A5 are the major P450 enzymes involved in the oxidation of steroids in human livers.[261] Among these

P450 enzymes examined, CYP3A4 is the most important for steroid metabolism, catalysing, for example, testosterone (TST) 6β-, 15β- and 2β-hydroxylations,[262,263] and dehydroepiandrosterone (DHEA) 7α-, 7β- and 16α-hydroxylations.[261] Androstenedione (AD) is metabolised by other P450s at the C2, C6, C11 and C16 positions on the β-face (Figure 4.7).[261,264] Another essential enzyme, CYP19A1, an estrogen synthase, plays vital roles in steroidogenesis as it catalyses the aromatisation of AD and TST to estrone and estradiol, respectively.[265] Metabolism of testosterone by human P450s in liver microsomes has been the most extensively studied; hydroxylation at most positions is observed but in most cases with little product selectivity.[261,266]

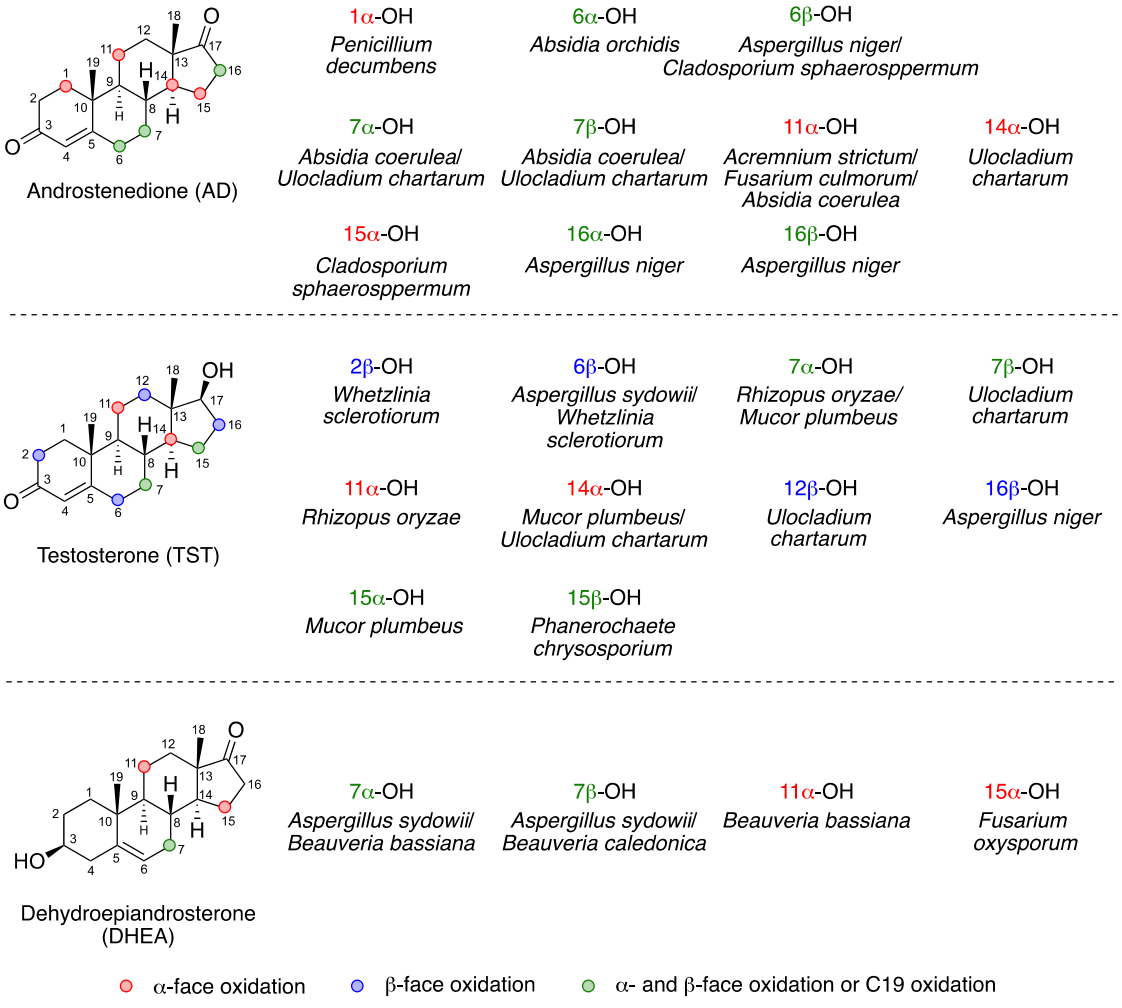


Figure 4.8 Summary of oxidation products of AD, TST and DHEA by selected fugal strains.

Other than human P450s, fungal steroid biotransformation has also been widely used to convert and synthesise steroidal compounds. Biotransformation by fungus gives a wide range of products of AD, TST and DHEA (Figure 4.8). AD is hydroxylated at almost every position except C8, C9 and C12 while TST shows all oxidation products except at C1, C8, C9.[266-270] No C19 oxidation of AD and TST has been observed in the reaction catalysed by fungi. DHEA oxidation is limited to C7, C11 and C15, with C7 oxidation products being predominant.[266,267,271] Although fungal biotransformation of steroids generates a wide range of metabolites, the product selectivity is too low for application in the synthesis of steroidal compounds.

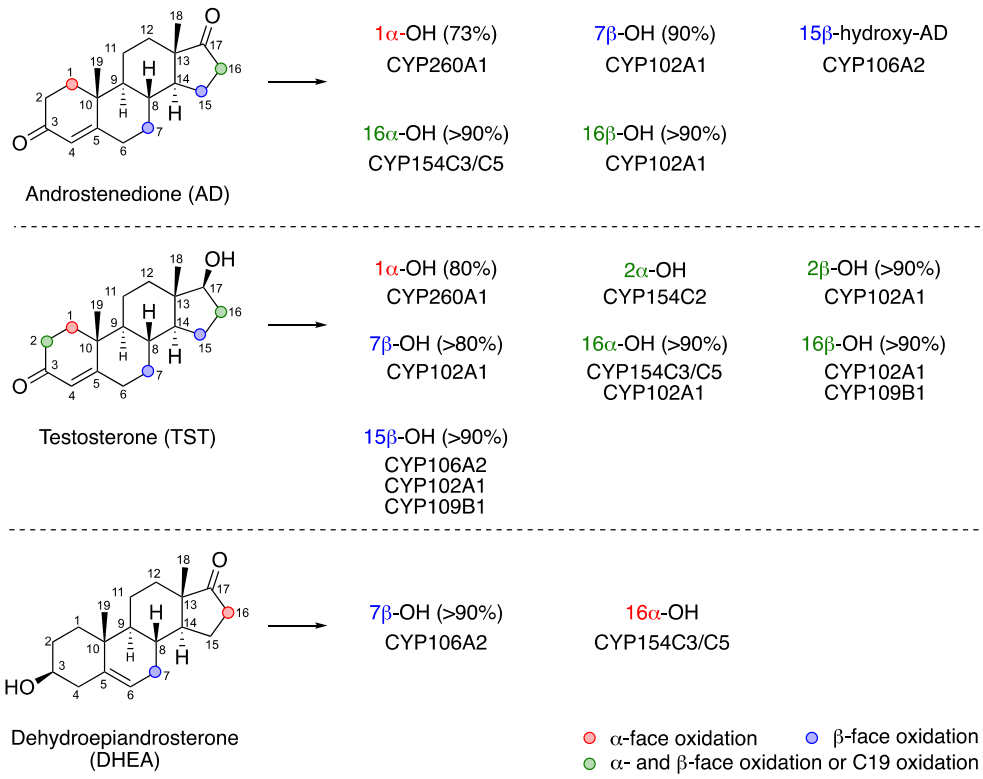


Figure 4.9: Summary of oxidation products of AD, TST and DHEA by various recombinant bacterial P450s.

Recombinant bacterial P450 enzymes offer well-defined systems that could be further engineered to target specific products, potentially providing a broader spectrum of product with high activity and selectivity.[234] Hydroxylation sites of steroids by different bacterial P450s are

collected in Figure 4.9. CYP106A2 from *Bacillus megaterium* catalyses 15β hydroxylation of 3-oxo-Δ^4-steroids including AD and TST,[272-275] and 7β hydroxylation of 3-hydroxy-Δ^5-steroids such as DHEA.[276] CYP260A1 from *Sorangium cellulosum* So ce56 selectively oxidises C-19 steroids such as AD and TST at the 1α-position.[277,278] CYP154C3 and CYP154C5 are highly selective for 16α oxidation of AD, TST and DHEA.[279,280]

CYP102A1 ($P450_{BM3}$) from *Bacillus megaterium* has been engineered by Reetz and co-workers to oxidise TST at 2β and 15β positions by variants F87A/A330W and R47Y/T49F/V78L/A82M/F87A, as well as PROG at 2β and 16β sites with high regio- and stereoselectivity.[135,162,165,166,169,281] Vermeulen and co-workers identified a template variant M01 (R47L/F87V/L188Q/E267V/G415S) that showed high activity for TST oxidation at the 2β, 15β and 16β positions, with 15β oxidation being the dominant pathway.[281] Addition of the A82W mutation to M01 shifted the oxidation site from 15β to 16β for TST, and further addition of the S72I mutation inverted the stereoselectivity of TST towards 16α-oxidation.[165,166] CYP102A1 was further engineered by a combination of ISM and CAST to give high (>90%) selectivity for 7β-, 16α- and 16β- oxidation of AD and TST.[166,171,172]

As shown in Figure 4.9, AD showed 1α-, 7β-, 15β-, 16α-, and 16β-hydroxylation, mostly with high selectivity, but other oxidation products are formed in trace quantities. For TST, CYP102A1 provided almost total selectivity for 2β-, 7β-, 15β-, 16α-, and 16β-hydroxylation. DHEA has only been reported to be oxidised at 7β and 16α positions by two bacterial P450 enzymes. Greater diversity of mono- and polyhydroxylated products will facilitate applications of P450 enzymes in the synthesis of steroid derivatives. In addition to the relatively narrow range of products, it was surprising how few bacterial P450s possessed steroid oxidation activity. The aim of this chapter is to generate a more diverse range of oxidation products of

AD, DHEA and TST than hitherto reported, by a combination of enzyme library screening and rational design of CYP102A1 variants. Initial screening of AD, DHEA and TST with a library of ~100 CYP102A1 variants showed a narrow range of products with relatively low selectivity and turnover number, suggesting that the active sites across the variant library allowed a limited number of binding modes for steroids. In order to access different binding orientations, the structure of WT CYP102A1 was overlaid with the structure of AD-bound human CYP19A1 to identify regions and residues in CYP102A1 that blocked this AD binding mode. Based on these observations, scanning glycine mutagenesis across 12 residues was conducted iteratively on 6 template variants found active in the initial library; these first and the second-generation variants showed increased product diversity in the oxidation of AD, DHEA and TST.[172]

4.2 Initial screening

4.2.1 Initial $P450_{BM3}$ library and screening conditions

The initial library of 48 enzymes for steroid oxidation screening was listed in Table 4.1. AD, DHEA and TST were screened for oxidation by the initial library to find active mutants, the screening data was analysed for the effect of mutations, further engineering was then performed based on these mutations. The full screening results are given in Tables 1.11–1.13 in the Appendices 10–12.

Screening reactions were carried out in 24-well plates in 200 mM phosphate buffer (pH = 8.0) in a total volume of 0.5 mL containing 2 μM $P450_{BM3}$ enzyme, 1 mM substrate, 20 U/mL glucose dehydrogenase (GDH) and 100 mM glucose. $NADP^+$ (40 μM) was added to initiate the reaction. Plates were shaken at 120 rpm for 6 h at 20 °C. Due to the lower substrate

concentration, the reaction time was shortened from the more typical 16 h to 6 h in order to limit further oxidation.[149]

Table 4.1: The list of variants in the initial library.

No.	Variants	No.	Variants
1	A330P	25	A82M/A330P
2	A82T/I263R	26	I263A/A330P
3	GVQ	27	F81W/T260A
4	K19	28	GV/A184I
5	K19/F87V	29	GVQ/A264G
6	K19/F87V/E267V	30	K19/F87V/A264G
7	K19/F87A/A82M/I263G	31	K19/F87V/E267V V78I
8	K19/F87V/A328I	32	K19/F87A/A82M/E267F
9	K19/F87A/F81W	33	R19/F87A/A328I/E267F
10	K19/F87A/I263A	34	R19/F87A/A328L
11	K19/F87V/Q403P	35	R19/F87A/A328I/S72W
12	KU3/A330P/A328I	36	R19/F87A/A328I/V78I
13	KU3/A330P/S72W	37	RLYF/IR
14	KU3/A330P/V78I	38	R19/F87A/A328F
15	R19	39	R19/F87A/I263G
16	R19/F87A/A184I	40	RP/A82W/I263A
17	R19/F87A/A328I	41	RP/A82M/I263A
18	R19/F87A	42	RP/F87V
19	RP/H171L	43	RP/F87V/V78I
20	RT2	44	RP/H171L/I263G A184I
21	RT2/S72G/A330W	45	RT2/H171L/I263G/A330W
22	RT2/A330P/V78I/A184I	46	RP/H171L/I263G/F87V/V78I
23	KU3/A330P	47	RP/V78I/E267V
24	RP/H171L/I263G	48	RT2/S72W/A330W

Reaction mixtures were extracted with ethyl acetate and analysed by GC (Chapter 2.5.1). Variants showing high activity and product selectivity were chosen for preparative scale reactions whereby the reaction volume was increased to 50–100 mL using the same concentration of components as in screening scale reactions. The reactions were monitored by extraction and GC analysis of small aliquots of reaction mixtures. When no further conversion occurred, organics were extracted with ethyl acetate. Products were purified by flash silica chromatography and characterised by NMR and MS data which are provided in the Appendices.

4.2.2 Oxidation of androstenedione (AD)

The enzyme library showed reasonable androstenedione (AD, **19**) oxidation activity; 25 out of 48 variants showed >20% conversion. Full activity and selectivity data are provided in Appendix 10. The five major products were purified and characterised as 6β-hydroxy-AD (**19a**),[282] 16α-hydroxy-AD (**19b**),[282] 16β-hydroxy-AD (**19c**),[282] 16-oxo-testosterone (**19d**),[283] and 2β,16β-dihydroxy-AD (**19e**)[172] (Figure 4.10 and Table 4.2).

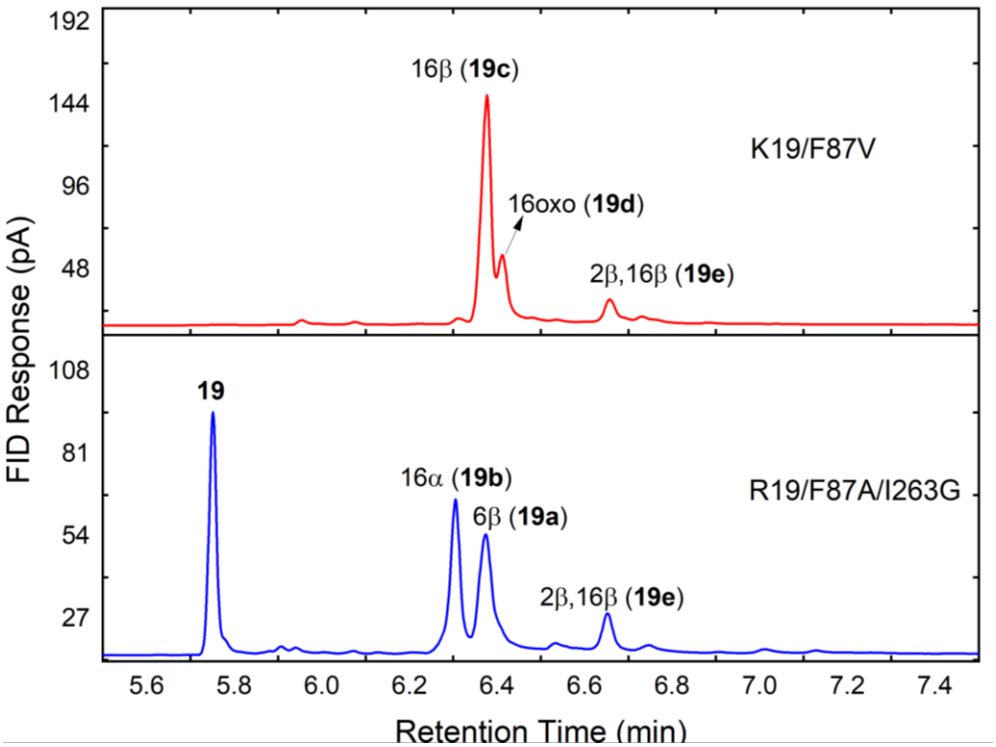
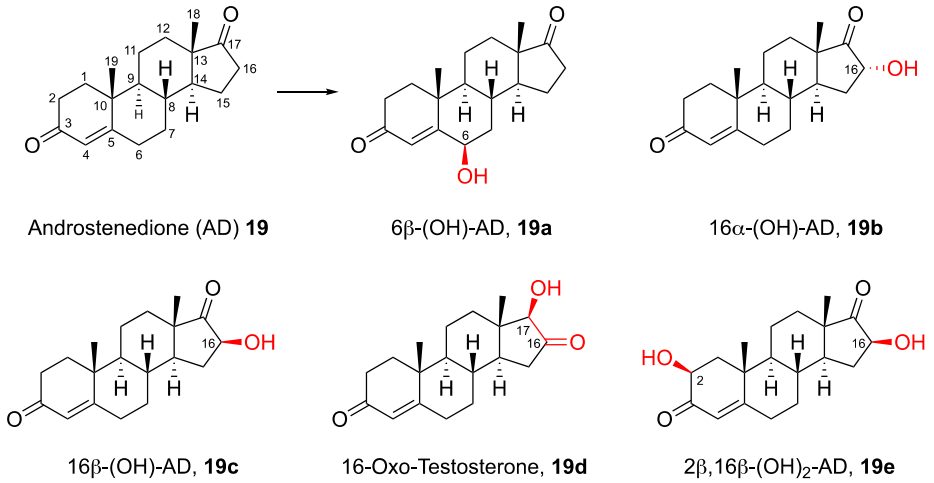


Figure 4.10: GC analysis of AD (**19**) oxidation by two variants in initial screening.

The conversion rates showed that the F87A or F87V mutations were important in promoting AD oxidation by CYP102A1, indicating their important role in creating space close to the haem for steroid binding.[135,169,170] 16β-Hydroxy-AD (**19c**) was a prominent product across the library; it was favoured by the F87V mutation, at up to 67% for the K19/F87V/Q403P variant (Table 4.2, Entries 5–7). The F87A mutation also biased the selectivity towards C16 oxidation but it

promoted further oxidation to 2β,16β-dihydroxy-AD, **19e** (Entries 8 & 9); R19/F87A/A184I gave 68% of this dihydroxylation product. Since the 2β-alcohol was not observed, the 16β-alcohol **19c** was the most likely precursor to **19e**.

Table 4.2: Substrate conversion rates and product selectivity of selected variants for the oxidation of AD, **19**.[a]



Entry	CYP102A1 variant	19a	19b	19c	19d	19e	others	Conv.	TON [b]
1	K19/F87A/I263G	44%	**48%**	–	–	–	8%	50%	120
2	R19/F87A/I263G	**36%**	23%	–	13%	–	28%	46%	83
3	RP/H171L/I263G/A184I	**47%**	31%	–	–	–	26%	57%	134
4	RP/H171L/I263G/F87V/V78I	**46%**	14%	–	23%	–	17%	57%	131
5	K19/F87V/Q403P	–	–	**67%**	22%	–	11%	84%	281
6	RP/F87V	–	–	**62%**	27%	–	11%	60%	186
7	K19/F87V	–	–	**54%**	21%	15%	10%	98%	265
8	R19/F87A	–	–	16%	2%	**66%**	16%	96%	317
9	R19/F87A/A184I	–	–	7%	2%	**68%**	23%	95%	323

[a] The selectivity for the two most selective variants for each product are highlighted in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 1 mM AD (500:1), 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 6 h. [b] TON refers to the turnover number for the formation of the product in bold.

The minor product 16-oxo-testosterone, **19d**, an isomer of 16β-hydroxy-AD (**19c**), was found with F87V-containing variants (Entries 5–7). RP/F87V showed the highest extent of isomerisation, giving 27% of **19d** together with 62% of **19c** (Entry 6). When the enzyme loading was lowered to 0.1 mol% (2 mM AD, 2 μM enzyme), both the isomerisation of **19c** to **19d** and

further oxidation of **19c** to **19e** were significantly suppressed. Under these conditions, the RP/F87V variant gave 91% of **19c** at 66% conversion (TON = 600) while K19/F87V/Q403P gave 80% of **19c** at 94% conversion (TON = 750), with no **19d** and **19e** formed. Reduced isomerisation of **19c** to **19d** at lower enzyme concentration suggested involvement of the enzyme in the reaction which likely involved two keto-enol tautomerisation steps (Fig. 4.11).

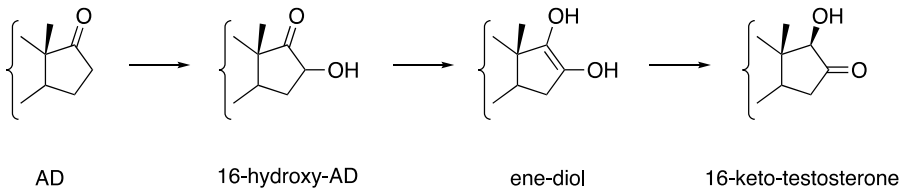


Figure 4.11: Proposed mechanism of the isomerisation of the C16 alcohols **19b** and **19c** to 16-oxo-testosterone, **19d**.

Few other mutations had significant effects on selectivity except for the I263G which abolished 16β oxidation and promoted the formation of 6β-hydroxy-AD (**19a**) and 16α-hydroxy-AD (**19b**) (Entries 1–4). Reetz and co-workers reported CYP102A1 variants from directed evolution and high-throughput screening that possessed >90% selectivity for 16α- and 16β-hydroxylation of AD (Table 4.3).[170] Comparison between Entries 2 and 4 shows the importance of the F87V mutation for 16β-oxidation. Vermeulen and co-workers had reported that mutations of A82 led to a switch between 16α- and 16β-oxidation of testosterone;[166] this was the basis for the further engineering by Reetz and co-workers. There is limited data on 16α-oxidation of AD; in addition to A82F, the S72I mutation had a pivotal effect.[170] Although neither product is a target in this thesis, since high selectivity had already been achieved, addition of the I263G mutation to the Entry 1 variant in Table 4.3 could increase the selectivity for **19b** above 95%.

Table 4.3: Comparison of CYP102A1 variants promoting 16α-hydroxy-AD (**19b**) and 16β-hydroxy-AD (**19c**).

Comparison	Entry	Mutations	Conversion	16α-OH	16β-OH
Literature	1	R47W/Y51W/S72I/A82F/F87I/L181C	85%	95%	–
	2	R47W/Y51H/A82W/F87V/L181Q/A330M	93%	–	100%
Initial library	3	F87A/H171L/I263G/Q307H/N319Y	50%	48%	–
	4	R47L/Y51F/F87V/I401P	66%	–	91%

4.2.3 Oxidation of dehydroepiandrosterone (DHEA)

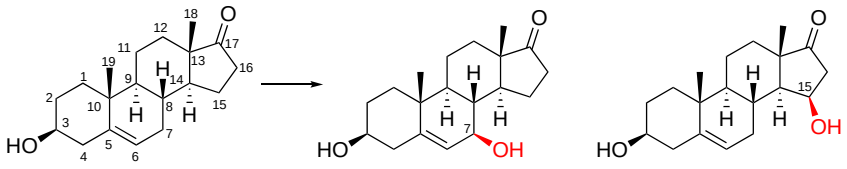
Dehydroepiandrosterone (DHEA, **20**) was a good substrate for the CYP102A1 variant library, as 27 out of 48 variants showed >20% conversion to a collection of 5 major products; 7β-hydroxy-DHEA (**20a**),[272,276] 15β-hydroxy-DHEA (**20b**),[284] 16β-hydroxy-DHEA (**20c**),[285] 16-oxo-androstenediol (**20d**),[286] and 7β,15β-dihydroxy-DHEA (**20e**) (Figure 4.12 and Table 4.4).

As observed for AD oxidation, the F87A and F87V mutations altered the product selectivity of DHEA oxidation. F87V-based variants gave mainly 16β-hydroxy-DHEA, **20c**, and its isomerisation product, 16-oxo-androstenediol, **20d** (Table 4.4, Entries 1 & 2), presumably via the same mechanism as proposed for **19d** (Figure 4.11). With a higher catalyst loading (0.2 mol%, 500:1), the variant A74G/F87V/A184I formed more of the isomerisation product **20d** (50%) than **20c** (38%, Entry 1), whereas a lower loading (0.1 mol%, 1000:1) suppressed **20d** and favoured **20c** formation (90% of **20c** and 4% of **20d**). The K19/F87V variant also gave 26% of 15β-hydroxy-DHEA, **20b**.

The F87A mutation, on the other hand, re-oriented DHEA binding to give 7β-hydroxy-DHEA (**20a**), 15β-hydroxy-DHEA (**20b**) and 7β,15β-dihydroxy-DHEA (**20e**). Addition of I263G/A or A264G mutations shifted the selectivity away from 16β-alcohol **20c** toward the 7β-alcohol (Entries 3–5), with K19/F87A/A82M/I263G giving 53% of **20a**. The F87A/A328I combination

promoted 15β-oxidation to give 48% of **20b** (Entries 6 & 7). Addition of the A184I and F81W mutations led to greater extent of further oxidation to the dihydroxylation product **20e** (Entries 8 & 9). Screening of the 7β and 15β alcohols with variants that formed **20e** showed that only the 15β alcohol was further oxidised to **20e**.

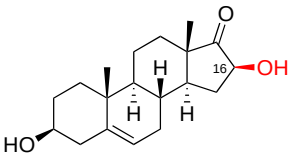
Table 4.4: Substrate conversion and product selectivity of selected variants for the oxidation of DHEA, **20**.[a]



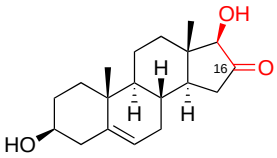
Dehydroepiandrosterone (DHEA) **20**

7b-(OH)-DHEA, **20a**

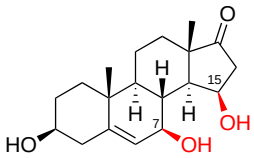
15b-(OH)-DHEA, **20b**



16b-(OH)-DHEA, **20c**



16-Oxo-Androstenediol, **20d**



7b,15b-$(OH)_2$-DHEA, **20e**

Entry	CYP102A1 variant	20a	20b	20c	20d	20e	others	Conv.	TON [b]
1	A74G/F87V/A184I (GV/A184I)	–	5%	38%	**50%**	–	7%	98%	245
2	K19/F87V	4%	26%	**34%**	17%	–	10%	99%	168
3	K19/F87A/A82M/I263G	**53%**	–	20%	–	–	26%	77%	204
4	K19/F87A/I263A	**49%**	25%	–	–	–	24%	65%	160
5	K19/F87V/A264G	**36%**	–	23%	11%	–	23%	98%	177
6	R19/F87A/A328I	4%	**48%**	27%	–	–	22%	62%	149
7	R19/F87A/V78I/A328I	5%	**34%**	23%	11%	–	27%	85%	145
8	R19/F87A/A184I	13%	8%	10%	–	**57%**	12%	98%	280
9	K19/F87A/F81W	15%	3%	8%	–	**57%**	17%	97%	277
10	R19/F87A	–	37%	10%	–	**42%**	11%	83%	175

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 1 mM DHEA (500:1), 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 6 h. [b] TON refers to the turnover number for the formation of the product in bold.

Overall, the variant library displayed low to modest selectivity for DHEA oxidation. Compared to AD, the new oxidation sites were C7 and C15, whereas AD showed C6 oxidation. Both sets of data pointed to significant effects of the I helix mutations I263A/G and A264G.

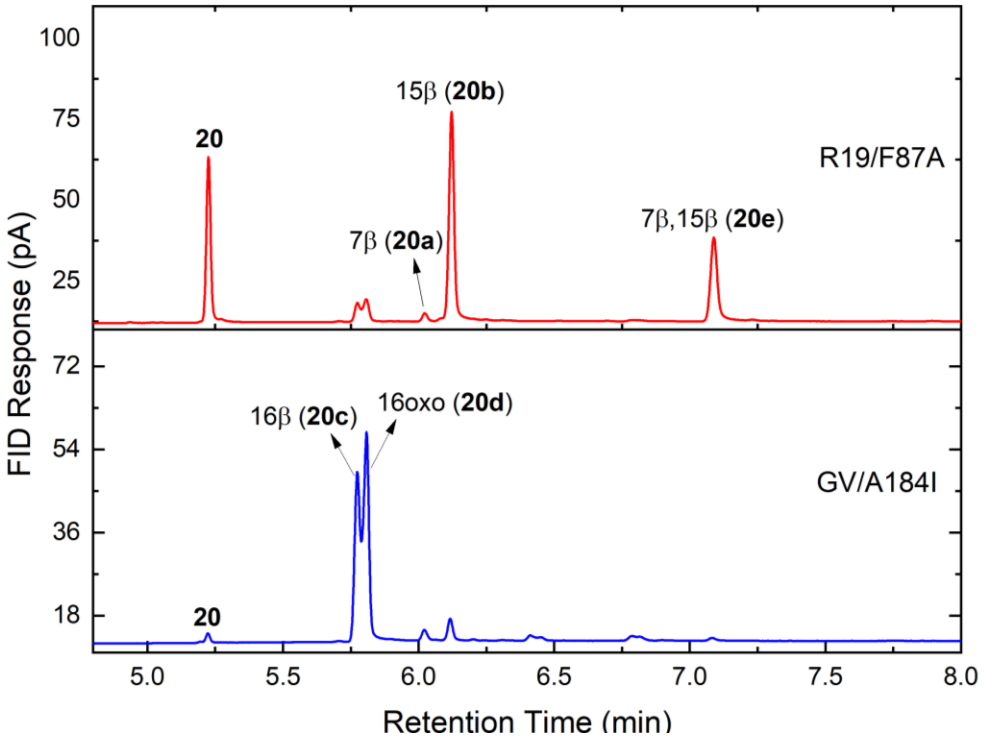


Figure 4.12: GC analysis of the product profile of DHEA (**20**) oxidation by two variants in initial screening.

4.2.4 Oxidation of testosterone (TST)

Testosterone (TST, **21**) oxidation by CYP102A1 has been extensively studied. High conversion and selectivity for 2β-, 7β-, 15β-, 16α- and 16β-hydroxylation (Figure 4.13) have been achieved by CAST and site-saturation mutagenesis and screening of thousands of mutants.[169,170,287] A small library of about 100 CYP102A1 variants created by rational design also showed high conversion and selectivity for the 2β- and 15β-alcohols.[135] Since 9 of the 48 variants in the initial screening library used for the work in this thesis were from this previous library, 2β-hydroxy-TST (**21a**) and 15β-hydroxy-TST (**21b**) were the major products. Variants with the F87V mutation showed strong preference for the 15β-alcohol whereas F87A-based variants gave more of the 2β-alcohol (Figure. 4.13). There was no evidence for C16 oxidation, presumably because the mutations A82F/W were absent; the combinations F87I/A82F and

F87V/A82W had been reported to bias the oxidation of AD and TST towards the 16α and 16β alcohols, respectively.[166,170]

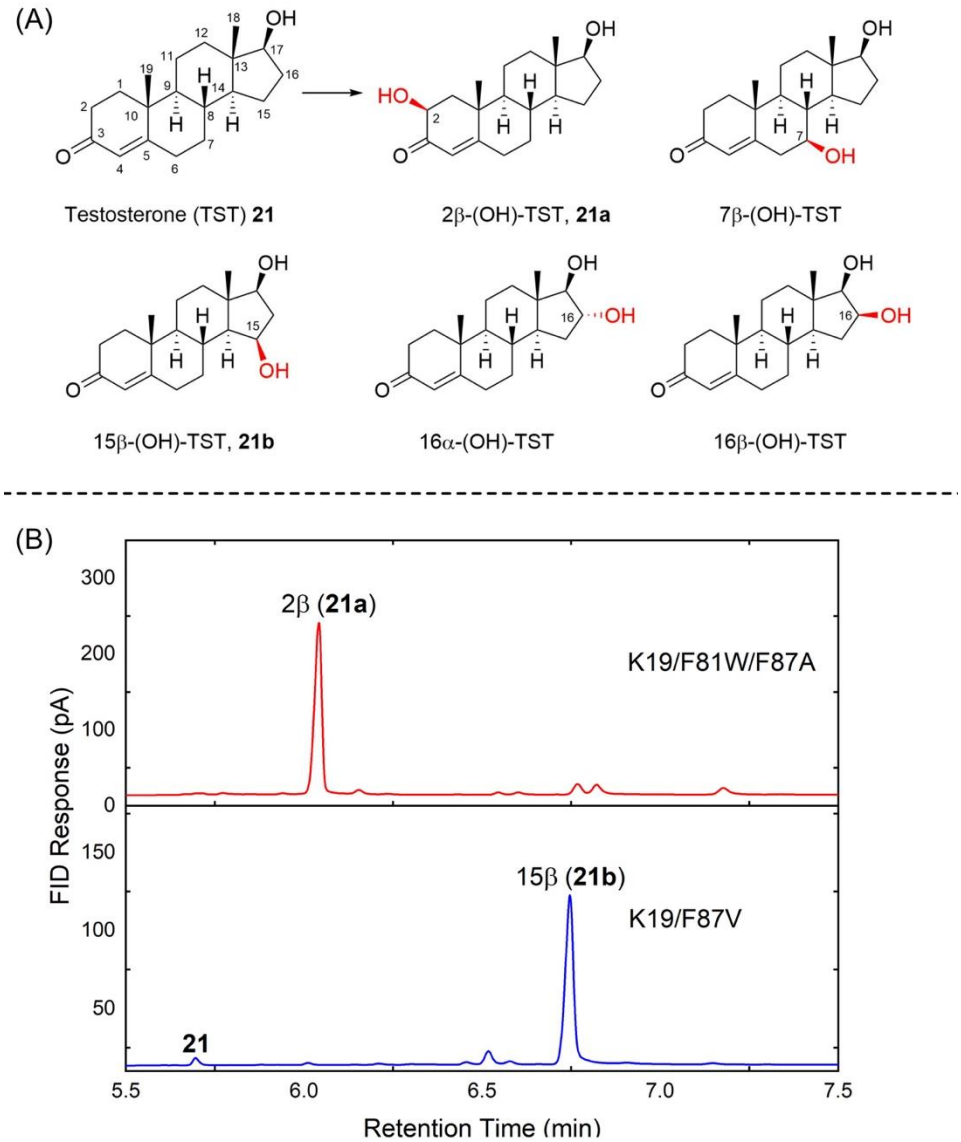


Figure 4.13: (A) Previously reported oxidation products of TST (**21**) by engineered variants of CYP102A1.[135,166,170] (B) GC analysis of TST oxidation by selected variants in the initial library to form **21a** and **21b**.

4.3 Nature-inspired design of scanning glycine mutagenesis

The screening library of variants oxidised AD and DHEA mainly at activated positions and on the lower half of the steroid B and D rings (C6, C7, C15 and C16), in most cases with low

selectivity. As for TST oxidation, no new product other than the 2β- and 15β- alcohols was found. The results showed that F87A or F87V mutations were crucial for steroid oxidation, and the I263G and A264G mutations provided access to oxidation sites other than C16 of AD and DHEA. The substitutions I263A and I263G had been shown to alter the oxidation selectivity of terpenes[183] and lauric acid,[288] in the latter case shifting the oxidation from sub-terminal ω–1, ω–2 and ω–3 position internally to ω–7, ω–8 and ω–9 hydroxylation.[288] Molecular dynamics simulation results suggested that the I helix, especially the region between T260 and T268, limited the regioselectivity shift of lauric acid; substitution of I263 with smaller side chains created space for the fatty acid to bind differently.

Steroids are rigid and sterically more demanding than fatty acids with their aliphatic chains. The restriction of TST oxidation products to the 2β- and 15β-alcohols in the screening library, and the need for specific combinations of mutations at A82 and F87 to access 16α and 16β oxidation, indicated that the CYP102A1 substrate pocket constrained steroid binding orientations such that only a limited number of C–H bonds were exposed to the ferryl oxygen. It is fascinating that 2α- and 15α-oxidation was not observed with TST in any previous study. This suggested that the TST binding orientations in CYP102A1 variants only exposed the β-face to the haem, and that the α-face was not accessible without dramatic changes in the binding orientation which required specific combinations of mutations. The challenge in identifying the required mutations is demonstrated by the A82/F87 combinations for TST C16 oxidation,[166] and the 15-residue site-saturation mutagenesis with high-throughput screening followed by further mutagenesis that provided 7β oxidation of AD and TST.[287]

In order to re-orientate steroid binding and expose other C–H bonds of the steroidal core for oxidation, the crystal structure of WT CYP102A1 was overlaid with that of CYP19A1 bound

with AD (Figure 4.14) to explore what changes in the CYP102A1 substrate pocket might enable AD to bind in a similar orientation to that in CYP19A1. The CYP19A1 aromatase enzyme catalyses C19-demethylation of AD to estrone, and of TST to estradiol.[265] Steric clashes are readily identified between a bound AD molecule with the side chain of F87, the main chain and side chain of I263 and A264 in the I helix, and the side chain of A328 in the β4 strand, so that AD could not adopt a similar orientation in CYP102A1 (Figure 4.14). The clashes are consistent with the effect of the mutations F87A and F87V, but especially for the I263G and A264G mutations, in the screening results.

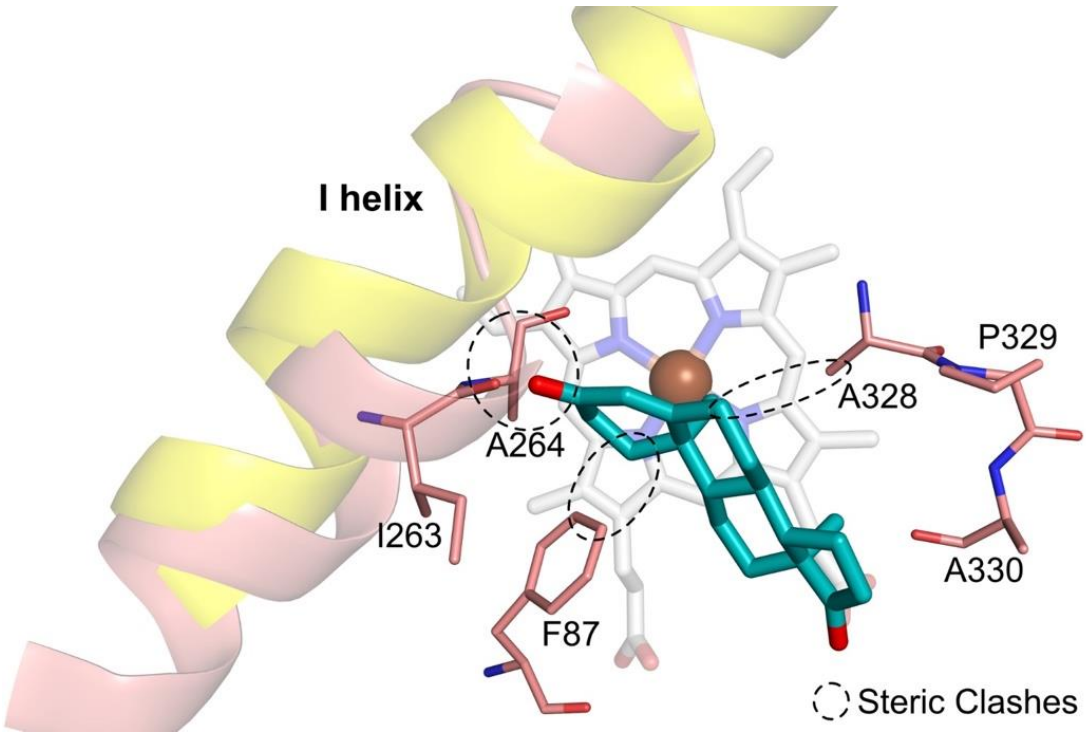


Figure 4.14: Overlay of a portion of the I helix of the aromatase CYP19A1 (yellow, haem group in grey, PDB code: 5jl6) bound with AD (cyan) with the structure of *N*-palmitoylglycine (NPG)-bound wild-type CYP102A1 (salmon, PDB code: 1jpz). NPG was omitted for clarity. Key CYP102A1 residues are shown in sticks (salmon).

The obvious effect of a glycine mutation is to create space in the active site, potentially to accommodate some part of a substrate. In addition, glycine residues are commonly found at the end of helices, and they can cause local distortions and partial unfolding of α-helices. Such

local distortions in the middle of a helix increase flexibility of the active site for substrate binding. Crucially, such distortions can be introduced (although not always predictably) by glycine mutations a few residues away from a contact residue. Therefore, although the structural overlay only indicated clashes of AD with I263, A264 and A328, glycine mutations were introduced at other residues, by scanning along the I helix (I259, T260, T261, L262, E267, T269, S270) and along the β4 strand (A328, P329 and A330), and combinations thereof. The active variants K19/F87A, K19/F87A/A82M, R19/F87A, R19/F87A/A184I, GV/A184I (A74G/F87V/A184I) and K19/F87V were used as templates for this scanning glycine mutagenesis approach.

4.4 Second-generation variants screening and computational studies

Scanning glycine mutagenesis across 11 residues and combinations thereof on 6 template variants led to the preparation of 74 new variants in the second-generation library (Table 4.5). The screening reactions of AD, DHEA and TST with this new library were conducted under the same conditions as in section 2.5.1, except the substrate and enzyme ratio was raised from 500:1 to 1000:1 (2 mM substrate and 2 μM enzyme) to avoid further oxidation and product isomerisation.[172] In order to gain insight into enzyme-steroid binding interactions, molecular dynamics (MD) simulations were also performed, in collaboration with Matthew Fisher, on a number of variants that oxidized **19**–**21** with high selectivity. Firstly, structures of variants were prepared by adding mutations to the NPG-bound WT CYP102A1 crystal structure (PDB code: 1jpz)[128] via Pymol. The bound NPG substrate was removed, and the haem moiety was simulated in the Compound I state. Docking of a substrate to this state of the enzyme simulated the ferryl oxygen atom abstracting a substrate hydrogen atom at a certain distance, providing a visual aid to rationalise product selectivity based on the enzyme-substrate interaction thus revealed.

Table 4.5: Variants in the second-generation library generated from scanning glycine mutagenesis across 11 residues.

No.	Variants	No.	Variants
1	GV/A184I/A264G	38	R19/F87A/A184I/I263G/A264G
2	GV/A184I/A264G/A328G	39	R19/F87A/A184I/I263G/A264G/A328G
3	GV/A184I/A328G	40	R19/F87A/A184I/I263G/A328G
4	GV/A184I/A328G/T260G	41	R19/F87A/A184I/I263W
5	GV/A184I/I263G	42	R19/F87A/A184I/L262G
6	GV/A184I/I263G/A264G	43	R19/F87A/A184I/S270G
7	GV/A184I/I263G/A264G/A328G	44	R19/F87A/A184I/T260G
8	GV/A184I/I263G/A328G	45	R19/F87A/A184I/T269G
9	GV/A184I/I263W	46	R19/F87A/A264G
10	GV/A184I/T260G	47	R19/F87A/A328G
11	GVQ/A264G/P329G/A330P	48	R19/F87A/A328G/A264G
12	GVQ/A264G/P329G/A330W	49	R19/F87A/A328G/I259G/I263G
13	GVQ/A328G	50	R19/F87A/A328G/I263G/A264G
14	GVQ/I263G/A264G	51	R19/F87A/A328G/I263W
15	GVQ/I263G/A264G/A328G	52	R19/F87A/A328G/L262G
16	GVQ/T260G	53	R19/F87A/A328G/P329G/A330G
17	K19/F87V/A264G	54	R19/F87A/A328G/P329G/A330G/T260G
18	K19/F87V/A184I	55	R19/F87A/A328G/S270G
19	K19/F87V/I263G	56	R19/F87A/A328G/T260G
20	K19/F87A/A82M/A264G	57	R19/F87A/A328G/T269G
21	K19/F87A/A82M/A264G/A328G	58	R19/F87A/F81W/A328G/T260G
22	K19/F87A/A82M/A328G/T260G	59	R19/F87A/G265GG
23	K19/F87A/A82M/A330W	60	R19/F87A/I259G/I263G
24	K19/F87A/A82M/F261G	61	R19/F87A/I263G/A264G
25	K19/F87A/A82M/I263G/A264G	62	R19/F87A/I263G/A328G
26	K19/F87A/A82M/I263G/A264G/A328G	63	R19/F87A/I263G/P329G/A330P
27	K19/F87A/A82M/I263G/A264G/A330W	64	R19/F87A/I263G/P329G/A330W
28	K19/F87A/A82M/I263G/A328G	65	R19/F87A/I263W
29	K19/F87A/A82M/I263G/A330W	66	R19/F87A/L262G
30	K19/F87A/A82M/I263W	67	R19/F87A/P329G/A330P
31	K19/F87A/A82M/L262G	68	R19/F87A/P329G/A330W
32	K19/F87A/A82M/S270G	69	R19/F87A/P329G/A330W/T260G
33	K19/F87A/A82M/T260G	70	R19/F87A/Q403P
34	R19/F87A/A184I/A264G	71	R19/F87A/S270G
35	R19/F87A/A184I/A264G/A328G	72	R19/F87A/T260G
36	R19/F87A/A184I/A328G/T260G	73	R19/F87A/T269G
37	R19/F87A/A184I/I259G/I263G	74	R19/F87A/W90V

The protein was prepared using the Amber 99SB*-ILDN force field with TIP3P water.[191] The heme moiety was simulated in the Compound I state, using parameters from Shahrokh *et al.*[192] The protein was placed into the centre of truncated octahedral box and solvated with water molecules followed by neutralisation with addition of Na^+ ions in GROMACS.[289] After

equilibration, triplicate, 100 ns of production MD were performed. The stability of individual trajectories was monitored by the RMSD of the backbone Cα atoms before they were clustered using the Daura algorithm with a cutoff of 0.12 nm.[166,169,170,290] The three most populated clusters usually covered >80% of the 10,000 conformations for each replica (sampling every 10 ps for 100 ns). These three most populated clusters were used as receptor structures for docking studies, which were performed in Autodock Vina.[291] Each substrate-enzyme docking normally generated 9 docking poses that were then either classified as "productive" or "non-productive". The productive poses referred to those with a carbon atom within 4.2 Å of the ferryl oxygen while the non-productive poses had the substrate docked too far away, or the C3-alcohol in DHEA, the C3-ketone in AD or C17-alcohol in TST were closest to the ferryl oxygen.

4.4.1 AD oxidation with the second-generation library

The new variants not only showed increased product selectivity and activity for products **19a**–**19e**, but also generated 3 new metabolites of AD in reasonable selectivity and turnover number. One metabolite was oxidation at the A ring, 2β-hydroxy-AD (**19f**),[292] and two others were at non-activated positions, 7β-hydroxy-AD (**19g**)[293] and 15β-hydroxy-AD (**19h**) which were labled in blue in Table 4.6.[292]

In the initial screening, the I263G mutation showed preference for 6β- and 16α-hydroxy-AD (**19a** and **19b**) formation. From the second-generation library, the combination of mutations I263G/A328G further optimised the selectivity for **19a** (Table 4.6, Entries 6–9). The highest selectivity of 78% was promoted by variant K19/F87V/I263G although at low conversion (Entry 10). The variants in the second-generation library had less preference towards 16α-hydroxy-AD (**19b**) that R19/F87A/I263G/P329G/A330W (48% of **19b**, Entry 11) only slightly improved the conversion and turnover number compared to R19/F87A/I263G (45% of **19b**,

Entry 5). The T260G mutation led to a new product, 7β-hydroxy-AD, **19g**, when combined with F87A (Entries 14 & 15). With the extra introduction of A328G, selectivity of **19g** was increased from 50% to 81% with almost full conversion (Entries 16 & 18, Figure 4.15-vi). Another example was R19/F87A/F81W/T260G/A328G (76% of **19g**, Entry 17) compared with the template K19/F87A/F81W which favoured 16β oxidation. 7β-Oxidation of AD was reported recently by Reetz and co-workers with 90% selectivity and 90% conversion for 1 mM substrate loading with the CYP102A1 variant (Entry 21).[287]

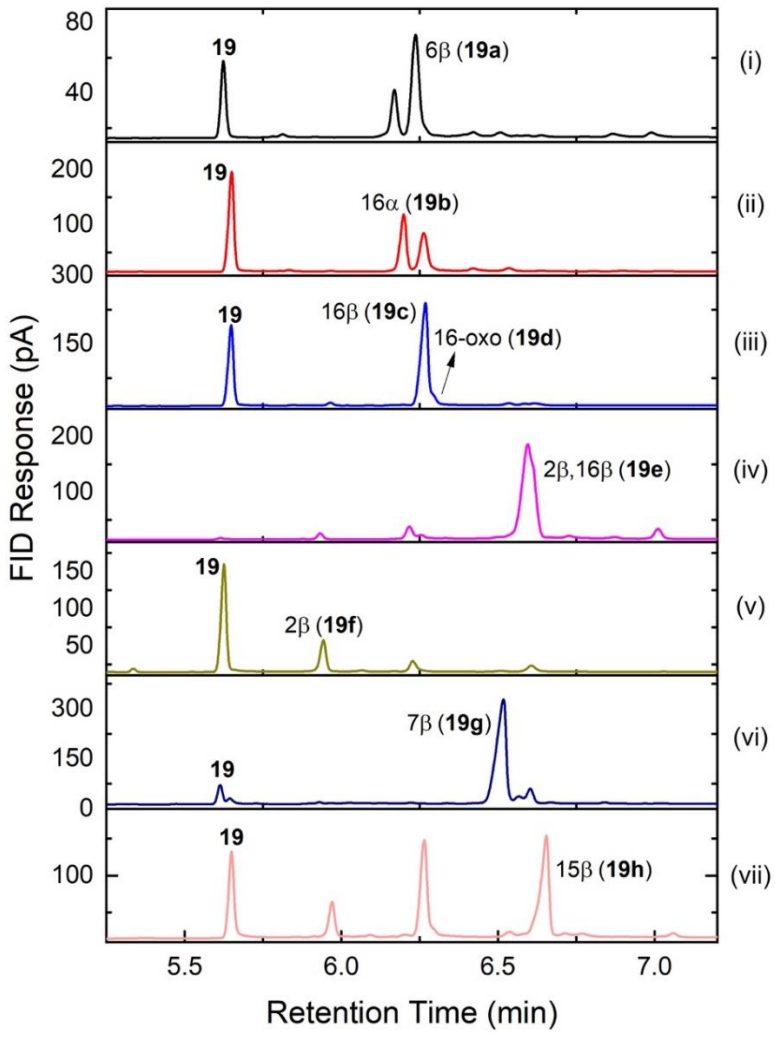
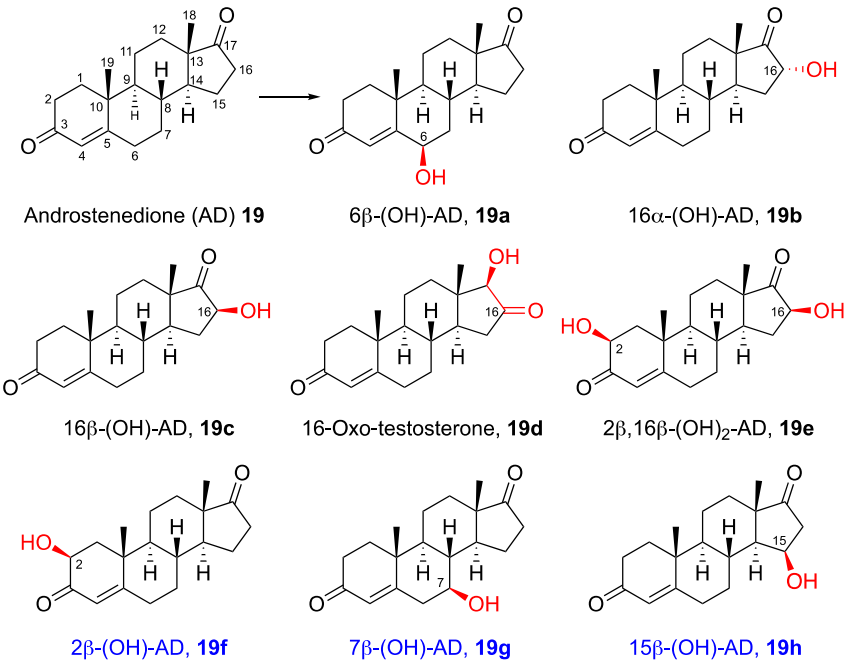


Figure 4.15: Gas chromatographic analysis of the oxidation of AD (**19**) catalysed by CYP102A1 variants: (i) 6β-hydroxy-AD **19a** (63%) by R19/F87A/A184I/I263G/A328G; (ii) 16α-hydroxy-AD **19b** (48%) by R19/F87A/I263G/P329G/A330W; (iii) 16β-hydroxy-AD **19c** (91%) and 16-oxo-testosterone **19d** (2%) by RP/F87V; (iv) 2β,16β-dihydroxy-AD **19e** (66%) by R19/F87A; (v) 2β-hydroxy-AD **19f** (60%) by K19/A82M/F87A/I263G/A264G/A328G; (vi) 7β-hydroxy-AD **19g** (81%) by R19/F87A/A184I/T260G/A328G; (vii) 15β-hydroxy-AD **19h** (50%) by R19/F87A/A264G.

Table 4.6: Selectivity and turnover number of $P450_{BM3}$ variants in the secondary library for the oxidation of AD.[a]



Entry	Variants	19a	19b	19f	19g	19h	Others	TON	Con.
1	K19/A82M/F87A*						100%[b]		88%
2	R19/F87A/A184I*						100%[c]		99%
3	GV/A184I*						100%[d]		65%
4	K19/F87A/I263G*	**46%**	45%				9%	133	29%
5	R19/F87A/I263G*	**48%**	45%				7%	168	35%
6	K19/A82M/F87A/I263G/A328G	**57%**	6%	20%	7%	10%		300	52%
7	GV/A184I/I263G/A328G	**63%**	20%			9%	8%	271	43%
8	R19/F87A/A184I/I263G/A328G	**63%**	24%				13%	490	77%
9	R19/F87A/I263G/A328G	**69%**	27%				4%	428	62%
10	K19/F87V/I263G	**78%**	22%					172	22%
11	R19/F87A/I263G/P329G/A330W	43%	**48%**				9%	274	57%
12	K19/A82M/F87A/I263G/A264G			**59%**			41%	212	36%
13	K19/A82M/F87A/I263G/A264G/A328G			**60%**	17%	13%		222	37%
14	K19/A82M/F87A/T260G				7%		93%[e]		98%
15	R19/F87A/A184I/T260G				12%	31%	57%[f]		69%
16	K19/A82M/F87A/T260G/A328G				**50%**	6%	44%[g]	460	92%
17	R19/F81W/F87A/T260G/A328G			5%	**76%**	8%	11%	623	82%
18	R19/F87A/A184I/T260G/A328G				**81%**		19%	761	94%
19	R19/ F87A/A184I/A264G			15%		**42%**	43%	315	75%
20	R19/ F87A/A264G			12%		**50%**	38%	400	80%
21	R47W/S72W/F77Y/V78L/F81I/A82L/ T88S/M177T/M185Q/L188Q/I209T[171]				90%				90%

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 2 mM AD, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 6 h. [b] denotes 77% of **19c** and 19% of **19e**. [c] denotes 71% of **19e**. [d] denotes 75% of **19c**. [e] denotes 80% of **19c**. [f] denotes 20% of **19c**. [g] denotes 37% of **19c**. "Others" contains mainly **19c**, **19d**, and

19e. "TON" refers to the turnover number for the formation of the product in bold. "Conversion" refers to the % of substrate converted to products. *Denotes variants in the initial screening library.

The additional A264G mutation on base variants R19/F87A and R19/F87A/A184I supressed oxidation at C16 and gave a new product, 15β-hydroxy-AD (**19h**) (Entries 19 & 20, Figure 4.15-vii). Other than **19h**, these variants with the A264G mutation also gave another new product, 2β-hydroxy-AD (**19f**), in small amounts, likewise for variant R19/F87A/A328G (15% of **19f**). Since both the A264G and A328G mutations led to some 2β oxidation, more glycine mutations on I helix and β4 strand residues, i.e., combinations of I263G, A264G, A328G, P329G and A330G, were added to the template variants. It was found that only combination I263G/A264G added to the K19/A82M/F87A template increased the 2β oxidation, to 59% (Entry 12), while the template favoured 16β oxidation (Entry 1). The additional A328G (Entry 13, Figure 4.15-v) gave similar selectivity of 2β oxidation (60%), suggesting that the glycine mutations on I helix residues were the dominant factor in promoting A ring oxidation of AD.

4.4.2 Computational studies of AD oxidation products

The K19/F87A/A82M/I263G/A264G variant was one of the most selective for 2β oxidation of AD (59% of **19f**), with **19c** (16β-OH) and **19d** (16-oxo-TST) as minor products. Six clustered structures were generated from the three replica MD simulations of this variant. AD was docked into these six clusters; of the 54 poses, 18 were productive poses that indicated C1, C2, C15, C16 and C19 oxidation. C2β and C15β oxidation was the most populated among these productive poses, with 8 and 6 poses, respectively, even though the 15β alcohol (**19h**) was not observed for this variant experimentally. The 2β binding poses of AD were within the five lowest energy poses for all of these clusters, which are overlaid in Figure 4.16–A showing the similar AD binding orientation.

In order to gain more information on substrate-enzyme interaction, the structure of this variant with the lowest energy 2β pose was overlaid with WT P450$_{BM3}$ in Figure 4.16–B. In this pose, C2 of AD was 3.2 Å from the ferryl oxygen, with C1 at 3.74 Å and C19 at 3.29 Å. The C3 carbonyl oxygen of this 2β pose of AD would clash with the side chains of I263 and A264 in WT P450$_{BM3}$, indicating the need for glycine mutation for AD to adopt this 2β pose. The C3 carbonyl oxygen contacted the side chains of T260 and the mutated residues M82 and A87, while the C17 carbonyl oxygen was hydrogen-bonded to the side chain of S72. The M82 side chain was in an extended conformation above A87. It had extensive contacts with AD, consistent with its importance in promoting 2β oxidation. Although C2 and C19 were at similar distances from the ferryl oxygen, the more activated secondary C–H bond of C2 was expected to be preferentially oxidised over the primary C–H bond of the C19 methyl.

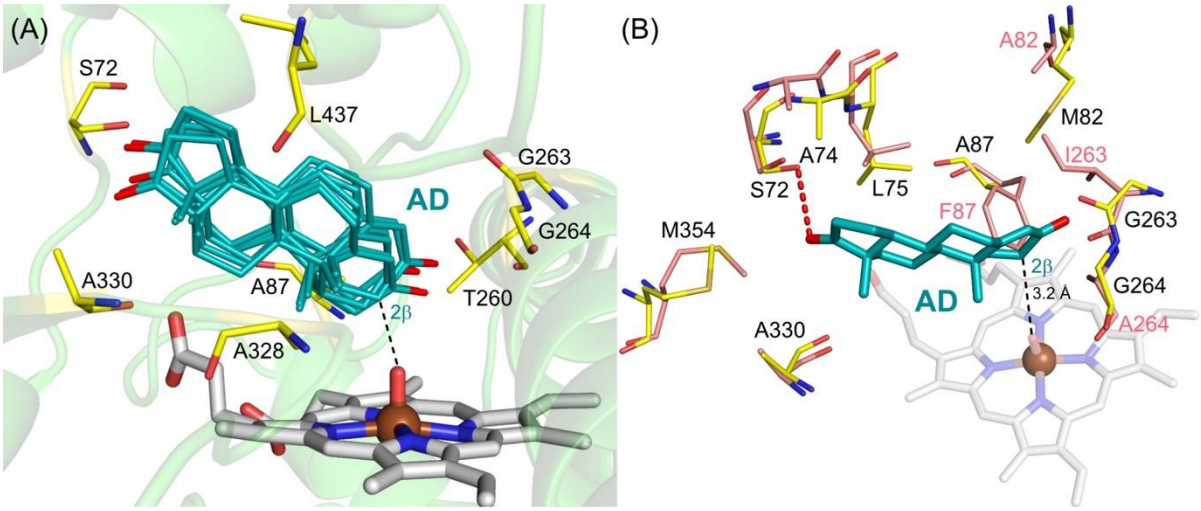


Figure 4.16: **(A)** The overlaid structure of six 2β binding poses from six clusters of variant K19/A82M/F87A/I263G/A264G showing the similarity between binding poses, key contact residues are shown in yellow. **(B)** K19/A82M/F87A/I263G/A264G variant with docked AD (teal) bound in a 2β pose overlaid with WT P450$_{BM3}$ (salmon) showing clashes of AD with F87, I263 and A264. Hydrogen-bonding between the C17 carbonyl oxygen and the side chain of S72 is shown in red dashes. Other key contacting residues within van der Waals contact distance of AD are M354 and A330 (to C17 carbonyl), M82 and T260 (C3 carbonyl), A87, L75 and A74 (C2, C14 and C16).

The simulated structure of the variant was then overlaid with the CYP19A1 structure with AD bound (Figure 4.17–A) and also compared with the structure prepared by Pymol before MD

simulation (Figure 4.17–B). It was surprising to see how close the docked 2β pose in the MD simulated structure of the variant was to that of AD in CYP19A1 (Figure 4.17–A). Arguably the more important observation is the structure of the I helix around these glycine mutations. As shown in Figure 4.17–B, even with the I263G and A264G mutations, the AD binding pose in CYP19A1 would still clash with the peptide backbone of residues 263 and 264 if the structure of the helix was intact, i.e. this binding pose would not be possible. The I helix in CYP19A1 adopts a "stepped back" structure at the point when the bound AD molecule approaches most closely, to create the space required for the observed binding pose. The fact that the CYP102A1 variant was able to bind AD in almost exactly the same pose was due to partial unwinding of the helix as a result of the two glycine mutations (Figure 4.17–A). This "stepping back" effect mimicked the I helix conformation in CYP19A1, thus creating space that allowed AD to bind in a similar position in the CYP102A1 variant to that in CYP19A1. This was unexpected but could be rationalised by the known tendency of glycine to disrupt the structure of α helices.

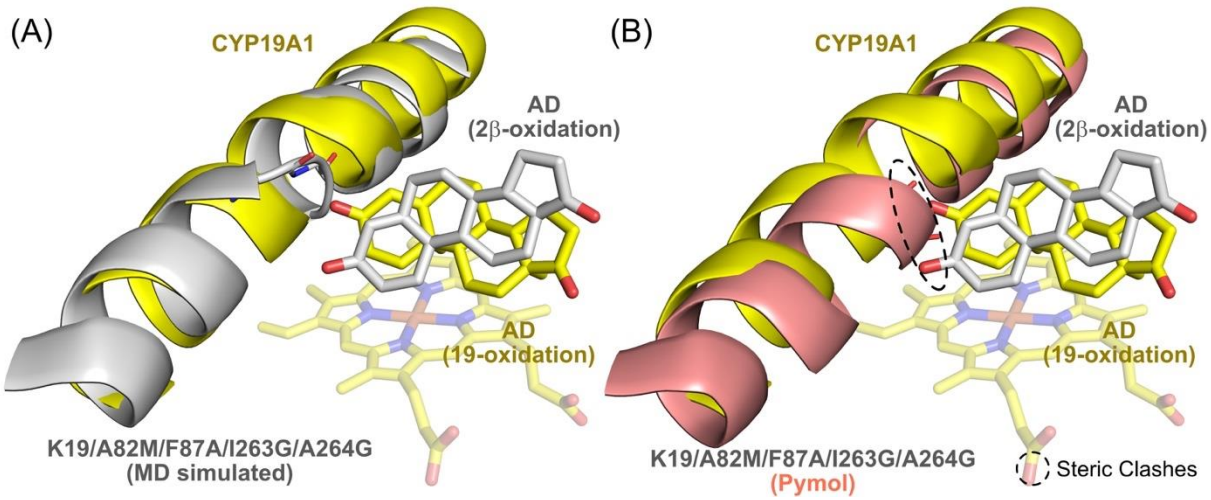


Figure 4.17: **(A)** Haem domain of CYP19A1 (yellow), bound with AD (yellow), was overlaid with the same cluster of variant K19/A82M/F87A/I263G/A264G (grey), with AD (grey) docked in 2β pose. Two of the key residues are shown as lines, I263G and A264G. **(B)** CYP19A1 (yellow), bound with AD (yellow), was overlaid with the same variant generated from WT $P450_{BM3}$ (salmon) using Pymol without MD simulation and the mutations are shown as lines. AD (grey) docked in 2β pose from Figure A is included to show the clashes with the I helix.

This nature-inspired mutagenesis of CYP102A1 was aimed at steroid A ring oxidation. If the AD binding mode in CYP19A1 was exactly reproduced, C19 oxidation should be observed. No C19 oxidation was observed with AD but the docking poses suggested that the more activated C–H bond of C2 was closer to the ferryl oxygen, and its more activated nature led to C2 oxidation. The approach therefore gave A ring oxidation of AD, and also opened up the possibility of forming products not observed in the initial screening library, e.g. 7β- and 15β-hydroxy-AD, in reasonable regio- and stereoselectivity.

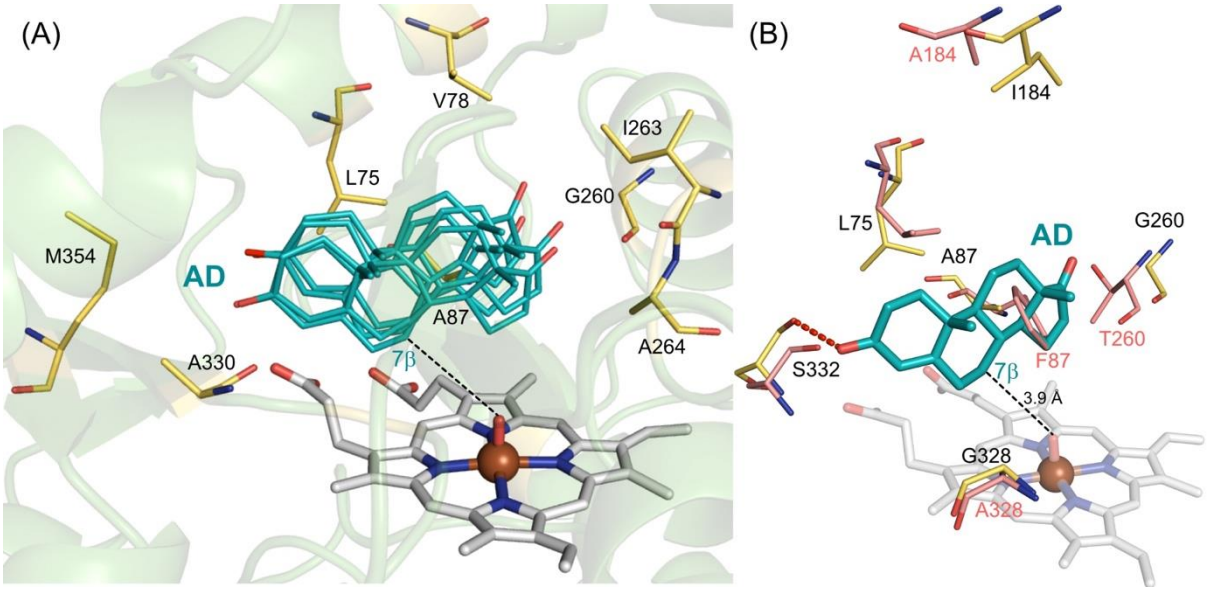


Figure 4.18 (A) The overlaid structure of four 7β binding poses from four clusters of variant R19/F87A/A184I/T260G/A328G shows the similarity between poses. Key contacting residues are shown in yellow. **(B)** R19/F87A/A184I/T260G/A328G variant with docked AD (teal) bound in a 7β pose overlaid with WT $P450_{BM3}$ (salmon) showing clashes of AD with the side chain of F87. Hydrogen-bond between the C3 carbonyl oxygen and the side chain of S332 is shown in red dashes. Other key contacting residues within van der Waals contact distance of AD are L75 and T260.

R19/F87A/A184I/T260G/A328G (V18) was the most selective variant for 7β-hydroxy-AD, **1g**, (81%). MD simulation was performed on this variant and 8 clustered structures were generated for docking. Out of the 72 poses, only 15 were productive poses, 6 of which placed the 7β hydrogen closest to the ferryl oxygen. The other poses included 1 pose for C1β, and 4 poses each for C2β and C15β oxidation. Most of the 7β binding poses of AD, with similar binding

orientations, were within the top five poses for all of these clusters. The four lowest energy poses are overlaid in Figure 4.18–A and the lowest energy pose is overlaid with WT $P450_{BM3}$ in Figure 4.18–B. C7 of AD was 3.9 Å from the ferryl oxygen, with C15 at the same distance; however, C15 oxidation was not observed in the reaction. The clash between AD and the side chain of F87 in the WT was evident in Figure 4.18–B. In the variant, the C17 carbonyl oxygen of AD contacted the side chain of G260, while the C3 carbonyl oxygen was hydrogen-bonded to the side chain of S332. The screening data in Table 4.6 suggested that A328G mutations was essential for 7β oxidation of AD. However, the rigid docking analysis did not show any contact between these residues and AD.

Proteins were prepared using the Amber 99SB*-ILDN force field with TIP3P water, which estimated the forces between atoms within molecules and between molecules. The interaction between residues in the active site was calculated using this force field and a pocket was generated above the porphyrin ring for substrate to interact with the haem centre. Three residues (S72, T260 and A328) were selected in Pymol to show this substrate pocket (in mesh) for various mutants in Figure 4.19. The mesh view is a solvent accessible surface which is traced by rolling a "water molecule" (at 1.4 A sphere) on the surface and it will only show up in the interface if there are holes large enough to accommodate a water molecule. Different variants showed various pocket sizes and shapes that led to different binding poses.

The substrate pocket of this clustered MD simulation structure of V18 variant bound with AD in a 7β pose is shown in mesh in Figure 4.19–B. In order to investigate how MD simulations might affect docking results, AD was docked into the structure of the variant generated by *in-silico* mutagenesis in Pymol, without MD simulation. The docking generated nine poses, none of which was productive, i.e. two poses with C3 exposed to the ferryl oxygen and the others bound too far away. In Figure 4.19–A, the substrate pocket of this Pymol structure is shown in

mesh, together with the 7β pose from docking into the MD simulated structure. It is clear that the *in-silico* structure has a narrower pocket that was unable to accommodate the 7β pose. The clash between the side chain of I263 and C17 of AD prevents it from adopting this 7β pose in the *in-silico* structure. However, the clash is absent in the MD simulated structure because of a slight rotation of the I helix so that the side chain of I263 avoided clashes with AD, therefore allowing the adoption of 7β pose (Figure 4.19–B). This comparison suggests that MD simulation may generate a better representation of mutant structures for substrate docking, even though the simulated structures may not be accurate in all structural details.

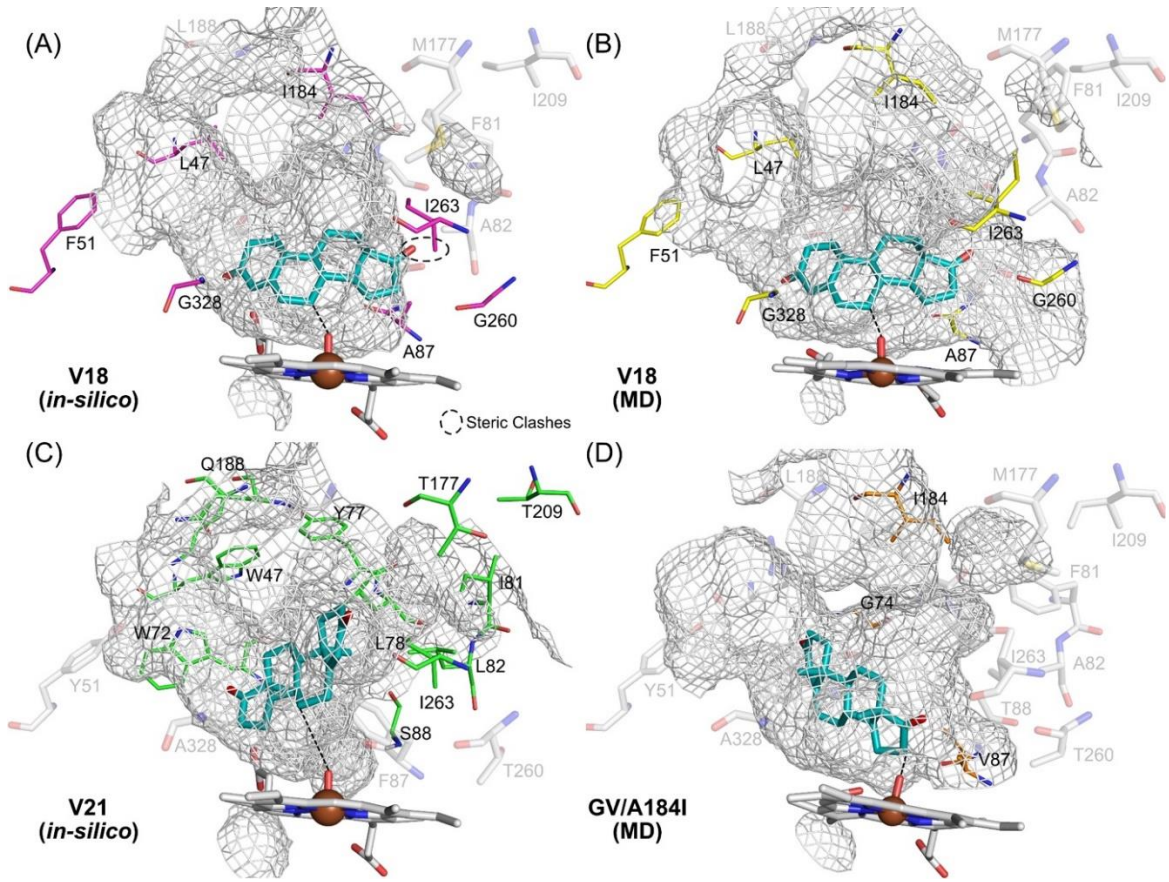


Figure 4.19: Selected CYP102A1 variants with residues S72, T260 and A328 selected to show mesh surface around the substrate pocket. The mutations of each variants are labelled in different colours. Other key residues are labelled in light grey for comparison (transparent sticks). **(A)** Variant R19/F87A/A184I/T260G/A328G (V18) from the second-generation library promoting 7β-hydroxy-AD, generated by mutagenesis in Pymol. **(B)** Variant V18 generated after MD simulation with AD docked in a 7β pose. **(C)** Variant R47W/S72W/F77Y/V78L/F81I/A82L/T88S/M177T/M185Q/L188Q/I209T (V21) from literature promoting 7β alcohol,[287] generated by mutagenesis in Pymol with AD docked in 7β pose. **(D)** Variant A74G/F87V/A184I from the second-generation library promoted 16β-hydroxy-AD, is generated after MD simulation with AD bound in 16β pose.

Reetz and co-workers reported recently a CYP102A1 variant giving 90% of 7β-hydroxy-AD; the R47W/S72W/F77Y/V78L/F81I/A82L/T88S/M177T/M185Q/L188Q/I209T (V21). In order to compare the docked 7β poses adopted by these two completely different variants (V18 and V21), AD was docked into the structure of V21 generated by *in-silico* mutagenesis in Pymol. Out of the nine poses generated from docking, only one was productive which showed 7β oxidation. The substrate pocket of variant V21 was similar to that of variant V18 but there is less space in the vicinity of I263 and A264, for obvious reasons. The 7β pose of AD in this structure is shown in Figure 4.19–C. The AD substrate is in a horizontal orientation, almost parallel to the haem. The major difference between the 7β pose in V21 compared to V18 was the AD molecule being "flipped over", with the C3 carbonyl contacting the I helix, instead of the C17 carbonyl in V18 (Figure 4.19–B). It may be speculated that the structure of the mutant would contain local conformation changes that promoted the adoption of more 7β poses.

The A74G/F87V/A184I variant favoured 16β-hydroxy-AD, **1c** (75%). The MD simulated structure showed a relatively vertical pocket shape extending from the haem for AD to dock in 16β poses (Figure 4.19–D). This avoided steric clashes with the I helix residues T260, I263 and A264. By contrast, the glycine mutations, T260G and A328G in variant R19/F87A/A184I/T260G/A328G altered the pocket shape for AD binding, enabling 7β docking poses such as that in Figure 4.19–B. The close approach of the D-ring to the V87 side chain is consistent with the need for mutation of F87 for steroid oxidation activity.

4.4.3 DHEA oxidation with the second-generation library

The second-generation variants not only improved selectivity for products **20a**–**20e** but also gave 6 new oxidation products of DHEA, which included the A ring oxidation product, 1β-hydroxy-DHEA (**20f**),[294] the isomer of the 7β-alcohol **20a**, 7α-hydroxy-DHEA (**20g**),[276] the

isomer of the 15β-alcohol, 15α-hydroxy-DHEA (**20h**),[271] and three further oxidation products of **20g**, namely 7-oxo-DHEA (**20i**),[276] 7α,15α-dihydroxy-DHEA (**20j**),[295] and 7α,16β-dihydroxy-DHEA (**20k**), which were labelled in blue in Table 4.8.

7α-Hydroxy-DHEA (**20g**) and 7β-hydroxy-DHEA (**20a**) co-eluted on the GC under all conditions. Therefore, 16 of the most selective (50%–97%) variants for C7 oxidation were selected for preparative scale reactions (Table 4.7). The crude extract of each was then analysed by the ^{1}H NMR spectrum. The 7α and 7β alcohols gave characteristic C$_7$*H*(OH) resonances in ^{1}H NMR spectra, which were then used to determine the isomer ratio by peak integration.

Table 4.7: Selectivity and turnover number of P450$_{BM3}$ variants in the second-generation library for the oxidation of 7α-hydroxy-DHEA (**20g**) and 7β-hydroxy-DHEA (**20a**).[a]

Entry	CYP102A1 variant	Selectivity[a]		de[b]	Conv.	Others	TON
		7β (20a)	7α (20g)				
1	R19/F87A/A264G	34%		100%	92%	66%	313
2	R19/F87A/A328G	35%		100%	82%	65%	287
3	R19/F87A/A328G/T260G	70%		100%	29%	30%	203
4	R19/F87A/A264G/A184I	74%		100%	93%	26%	688
5	R19/F87A/A264G/A328G	89%		100%	100%	11%	890
6	R19/F87A/T260G/A184I	93%		100%	100%	7%	930
7	R19/F87A/A184I/A264G/A328G	94%		100%	100%	6%	940
8	R19/F87A/A184I/T260G/A328G	98%		100%	100%	2%	980
9	R19/F87A/A184I/I263G/A264G	83%	6%	86%	33%	11%	20
10	R19/F87A/I263G/A264G	66%	14%	65%	39%	20%	55
11	K19/AM/F87A/I263G/A264G	52%	34%	21%	38%	14%	129
12	R19/F87A/A328G/P329G/A330G	30%	57%	–31%	95%	13%	542
13	K19/F87A/I263G/A330W	8%	56%	–75%	42%	36%	235
14	K19/F87V/I263G	10%	71%	–75%	39%	19%	277
15	GV/A184I/I263G/A328G	5%	54%	–83%	83%	41%	448
16	K19/AM/F87A/I263G/A328G		96%	–100%	48%	4%	461

[a] The selectivity for the two most selective variants for each product are in red. Prep-scale reactions at 10 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 2 mM DHEA, 40 μM NADP$^+$, 100 mM glucose, and 20 U/mL GDH. Flasks were shaken at 120 rpm at 20 °C for 6 h. [b] de denotes the diastereoselectivity of C7 oxidation (% 7β–% 7α), positive value refers to more 7β alcohol formation and negative value refers to more 7α alcohol formation. "Others" contains products **20b**–**20f**, **20h**–**20j**. "TON" for Entries 1-8 refers to the turnover number for the 7β-hydroxy-DHEA (**20a**), Entries 9-16 refers to the turnover number for the 7α-hydroxy-DHEA (**20g**). "Conversion" refers to the % of substrate converted to products.

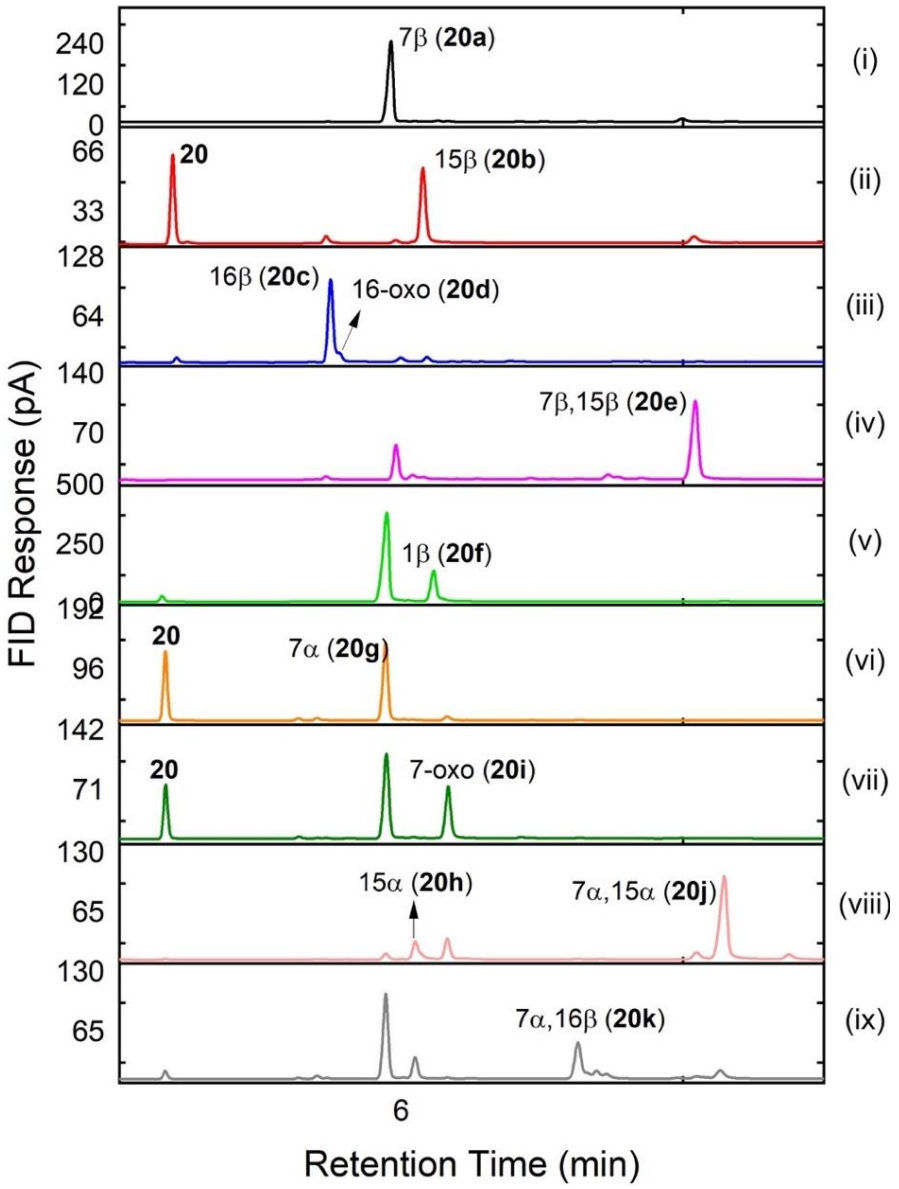


Figure 4.20: Gas chromatographic analysis of the oxidation of DHEA (**20**) catalysed by CYP102A1 variants: (i) 7β-hydroxy-DHEA **20a** (97%) by R19/F87A/A184I/T260G/A328G; (ii) 15β-hydroxy-DHEA **20b** (91%) by R19/F87A/W90V; (iii) 16β-hydroxy-DHEA **20c** (90%) and 16-oxo-androstenediol **20d** (5%) by GV/A184I; (iv) 7β,15β-dihydroxy-DHEA **20e** (74%) by R19/F87A/A184I/S270G; (v) 1β-hydroxy-DHEA **20f** (23%) by R19/F87A/F81W/T260G/A328G; (vi) 7α-hydroxy-DHEA **20g** (96%) by K19/F87A/A82M/I263G/A328G; (vii) 7-oxo-DHEA **20i** (35%) by K19/F87A/A82M/I263G/A264G/A328G; (viii) 15α-hydroxy-DHEA **20h** (22%) and 7α,15α-dihydroxy-DHEA **20j** (62%) by GV/A184I/I263G/A264G/A328G; (ix) 7α,16β-dihydroxy-DHEA **20k** (33%) by K19/F87A/A82M/A330W.

Introduction of the glycine mutations A264G and A328G to R19/F87A led to slight improvement for the 7β alcohol **20a** (34% and 35%, Entries 1 & 2) compared to the template, which favoured 15β oxidation. The combination of A264G/A328G, however, drastically increased the selectivity for **20a** to 89% (Entry 5). Another I helix glycine mutation, T260G,

also increased the selectivity for **20a** to 70% when added to variant R19/F87A/A328G (Entry 3). With the extra introduction of the A184I mutation to A264G and T260G, selectivity of **20a** was further increased to 74% and 93% (Entries 4 & 6). Finally, combining these positive residues together yielded almost full selectivity of **20a**, 94% and 98% with high turnover number (Entries 7 & 8, Figure 4.20-i).

In contrast to the 7β alcohol **20a**, 7α-hydroxy-DHEA (**20g**) was favoured by the I263G mutation (Table 4.7). The introduction of I263G into 7β alcohol-favouring variants started to give the 7α alcohol. For example, variant K19/A82M/F87A/I263G/A264G showed 86% of C7 oxidation with a 7β:7α ratio of 60:40 (Entry 11). Based on I263G, the addition of the A328G mutation further improved the selectivity for 7α oxidation; K19/A82M/F87A/I263G/A328G variant gave the highest selectivity for **20g**, 96% in 100% de (Entry 16, Figure 4.20-vi). The triple mutation combination of A328G/P329G/A330G improved the turnover number (TON = 542) for **20g**, but at the same time with a lower selectivity (Entry 12).

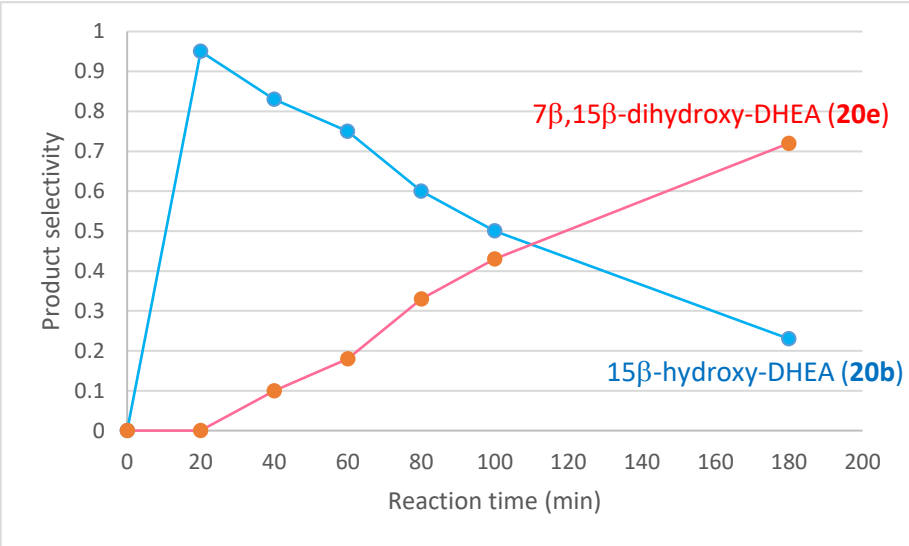


Figure 4.21: Time course reaction of DHEA with variant R19/F87A, monitoring the formation of 15β-hydroxy-DHEA (**20b**) and its further oxidation product, 7β,15β-dihydroxy-DHEA (**20e**). 50 mL reaction was in 200 mM phosphate buffer (pH 8.0) containing 2 μM R19/F87A variant, 2mM DHEA, GDH (20 U/mL), glucose (100 mM), and $NADP^+$ (40 μM) was added to initiate the reaction which was shaken at 120 rpm for 3 h at 20 ºC.

In the initial library, active variants such as R19/F87A and R19/F87A/A184I produced the diol 7β,15β-dihydroxy-DHEA (**20e**). A 50 mL scale oxidation of 2 mM of DHEA with 2 μM R19/F87A was monitored over time for the formation of **20e** (Figure 4.21). An aliquot (0.5 mL) of reaction mixture was taken every 20 min, extracted with 0.3 mL of ethyl acetate and analysed by GC. During the first 20 min, 61% of DHEA was converted to 15β-hydroxy-DHEA (**20b**) with 95% selectivity (TON = 580), with no evidence of **20e** formed. After another 20 min, **20b** was further oxidised, forming 10% of **20e**. After 180 min, 72% of **20e** was formed, with 23% of **20b** left in the reaction (Figure 4.20-iv). This suggested that high conversion and selectivity of **20b** and **20e** could be both achieved with accurate time control in small scale reactions. However, due to its high activity, R19/F87A would not be suitable for large scale production of **20b** since it requires accurate time control and repeated removal of product to prevent further oxidation. Therefore, new variants were explored to stop the further oxidation.

Based on variant R19/F87A, the additional I helix glycine mutations improved the selectivity for the 15β alcohol (**20b**), with less **20e** formed. The R19/F87A/L262G and R19/F87A/T269G variants gave 55% and 62% of **20b** with 10% and 13% of **20e**, respectively (Table 4.8, Entries 1 & 2). However, the highest selectivity for **20b** of 91% was found with the R19/F87A/W90V variant identified from screening with the larger library in the collection (Entry 3, Figure 4.20-ii). The proximity of F87 and W90 might lead to conformational changes that altered the DHEA binding.

The glycine mutation combination T260G/A328G not only increased the selectivity for 7β alcohol **20a**, but also led to A ring oxidation of DHEA to give 1β-hydroxy-DHEA (**20f**). The K19/F87A/A82M/T260G/A328G variant gave 18% of **20f** (Table 4.8, Entry 6) and R19/F87A/F81W/T260G/A328G increased it to 23% at 97% conversion (Entry 7, Figure 4.20-

v). The selectivity was slightly higher with other combinations, such as variant R19/F87A/T260G/A328G/P329G/A330G giving 30% but conversion was lower (Entry 8).

Apart from 7α-hydroxy-DHEA (**20g**), I263G-containing variants also gave another new product, 7-oxo-DHEA (**20i**) by further oxidation of the 7α alcohol. Based on the F87A mutation, the double glycine mutation on I helix residues, I263G/A264G, gave small amount of **2i** (Table 4.8, Entries 9 & 10). With the additional A328G, the selectivity of **20i** was increased to 27% and 35% (Entries 11 & 12, Figure 4.20-vii).

On the other hand, the F87V mutation favoured another further oxidation product from 7α alcohol, the diol 7α,15α-dihydroxy-DHEA (**20j**). Based on GV/A184I, double mutations I263G/A264G and I263G/A328G gave 12% and 40% of **20j** (Entries 13 & 14). The combination of these variants, I263G/A264G/A328G, further increased **20j** to 42% and 64% (Entries 15 & 16, Figure 4.20-viii).

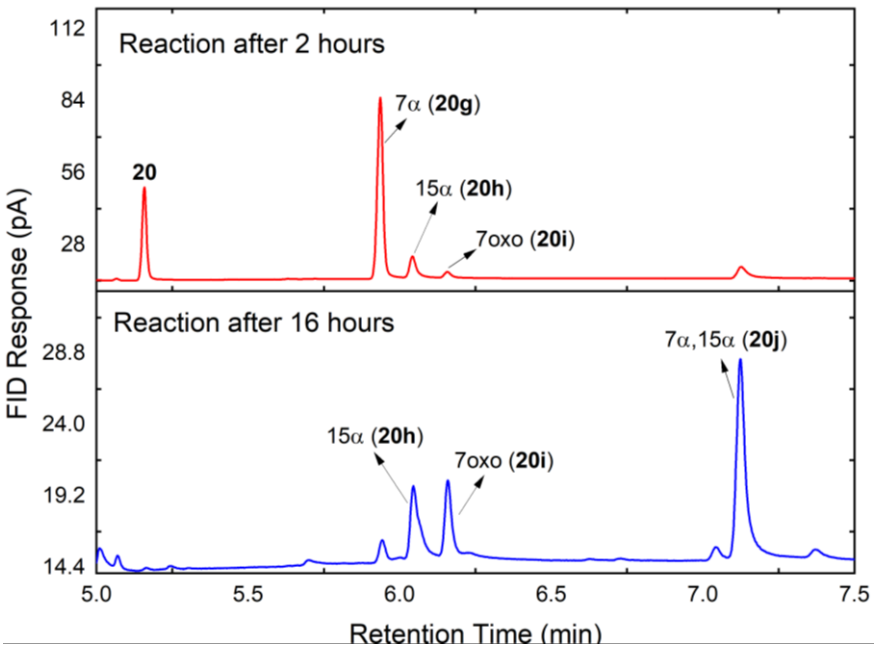
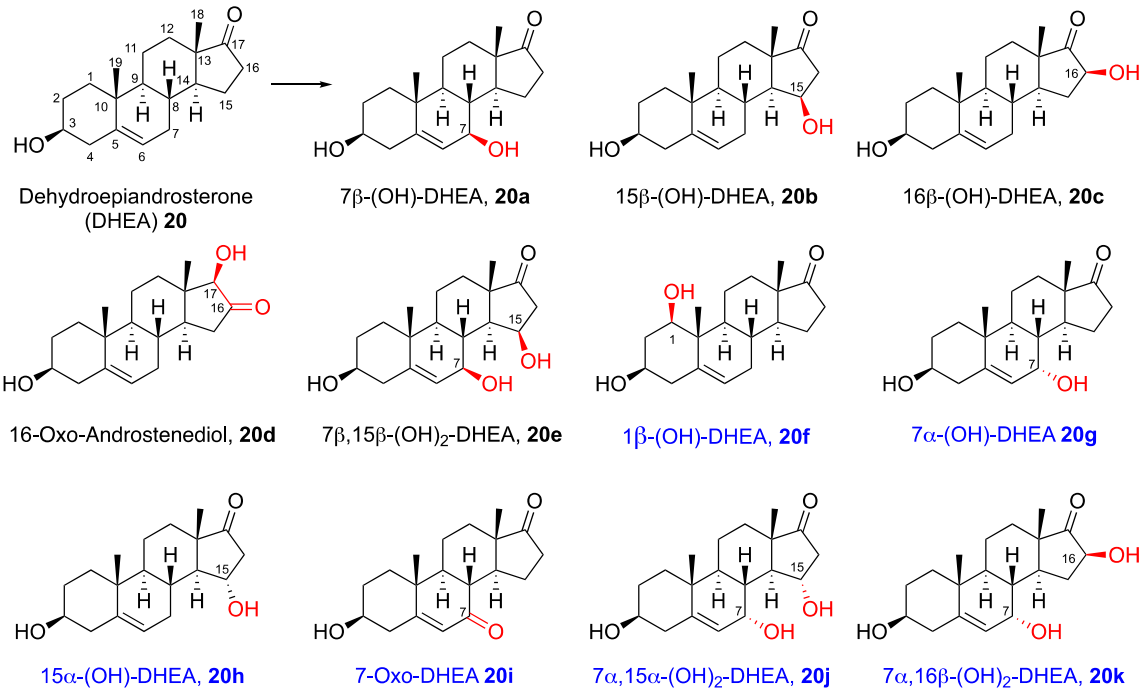


Figure 4.22: Time course reaction of DHEA with GV/A184I/I263G/A264G/A328G was conducted in 200 mM phosphate buffer (pH 7.9) containing 2 μM variant, 2mM DHEA, GDH (20 U/mL), glucose (100 mM), and $NADP^+$ (40 μM) was added to initiate the reaction which was shaken at 120 rpm for 16 h at 20 ºC.

A time-course reaction of DHEA oxidation with GV/A184I/I263G/A264G/A328G was performed to monitor the further oxidation of the 7α alcohol **20g** (Figure 4.22). After 2 h, 76% of DHEA was converted to 78% of **20g**, together with 11% of a new product **20h**, and 3% of the 7-oxo product **20i**. After 16 h, **20j** was the major product (62%) with a trace amount of **20g** left, confirming the 7α alcohol as the precursor to this diol. This observation was in sharp contrast to the formation of the 7β,15β-diol **20e** where the 15β-alcohol **20b** was the precursor. The new peak **20h** was purified and characterised as 15α-hydroxy-DHEA,[271] that appeared as a small shoulder peak close to the 15β alcohol in the second-generation screening.

Dihydroxylation was a common feature of DHEA oxidation by CYP102A1. Apart from 7α,15α-dihydroxy-DHEA (**20j**) and 7β,15β-dihydroxy-DHEA (**20e**), another diol was found in the screening, 7α,16β-dihydroxy-DHEA (**20k**). Addition of the I helix mutations F261G and L262G to K19/F87A/A82M led to small amount of **20k** with 10% and 18%. GV/A184I/I263G gave 24% of **20k** at 97% conversion (Entry 17). Interestingly, the F87A/A330W combination also promoted this reaction; with the K19/F87A/A82M/A330W variant forming 33% of **20k** (Entry 18, Figure 4.20-ix).

Table 4.8: Selectivity and turnover number of $P450_{BM3}$variants in the second-generation library for the oxidation of DHEA.[a]



Entry	Variants	20b	20c	20e	20f	20i	20j	20k	*A*	*B*	*C*
1	R19/F87A/L262G	**55%**	33%	10%					2%	44%	242
2	R19/F87A/T269G	**62%**	14%	13%					11%	44%	273
3	R19/F87A/W90V	**91%**	9%							61%	555
4	K19/F87A/A82M*	31%	**54%**	3%					12%	92%	497
5	GV/A184I*		**90%**						10%	84%	756
6	K19/F87A/A82M/T260G/A328G				**18%**				82%[b]	95%	171
7	R19/F87A/F81W/T260G/A328G				**23%**				77%[c]	97%	223
8	R19/F87A/A328G/A329G/A330G/T260G				**30%**				70%[d]	37%	111
9	R19/F87A/I263G/A264G	7%				**6%**			87%[e]	67%	40
10	K19/F87A/A82M/I263G/A264G	7%				**7%**			86%[f]	63%	44
11	R19/F87A/I263G/A264G/A328G	13%				**27%**			60%[g]	50%	135
12	K19/F87A/A82M/I263G/A264G/A328G					**35%**			65%[h]	78%	273
13	GV/A184I/I263G/A264G	19%		6%			**12%**		63%[i]	82%	98
14	GV/A184I/I263G/A328G	8%					**40%**		52%[j]	98%	392
15	GVQ/I263G/A264G/A328G	5%				7%	**42%**		46%[k]	98%	412
16	GV/A184I/I263G/A264G/A328G	12%				12%	**64%**		12%	100%	640
17	GV/A184I/I263G	11%					7%	**24%**	58%[l]	97%	233
18	K19/F87A/A82M/A330W			27%		6%		**33%**	34%	96%	317

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 2 mM DHEA, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 6 h. [b] denotes 80% of **20a**. [c] denotes 77% of **20a**. [d] denotes 67% of **20a**/**20g**. [e] denotes 60% of **20a**/**20g**. [f] denotes 64% of **20a**/**20g**. [g] denotes 37% of **20a**/**20g** and 13% of **20b**. [h] denotes 56% of **20a**/**20g**. [i] denotes 52% of **20a**/**20g**. [j] denotes 39% of **20a**/**20g**. [k] denotes 43% of **20a**/**20g**. [l] denotes 42% of **20a**/**20g**. *A* denotes "Others" that consists of mainly of the C7 oxidation products. *B* denotes "Conversion" that refers to the % of substrate converted to products. *C* denotes "TON" that refers to the turnover number for the formation of the product in bold. *Denotes variants in the initial screening library.

4.4.4 Computational studies of DHEA oxidation products

GV/A184I/I263G/A264G/A328G was one of the most reactive and selective variants for 7α-hydroxy-DHEA **20g** (78%), with 15α-hydroxy-DHEA **20h** and 7-oxo-DHEA **20i** as minor products (Figure 4.22). Five clustered structures were generated from the three replicas of this variant for DHEA docking, which showed 16 productive poses that indicated C1α/β, C2β, C7α/β, C11α, C15α/β and C19 oxidation. C7α and C15α were the most dominant positions, with 3 and 4 poses, respectively. The overlaid structure of the 3 poses of C7α oxidation is shown in Figure 4.23–A, 2 of which were the lowest energy poses. The C17 carbonyl contacted the side chains of P329 and A330 while the C3 alcohol contacted the side chains of F81 and M177, all within van der Waals contact distance. Curiously, there was no hydrogen–bond to the oxygen functionalities at either end of DHEA.

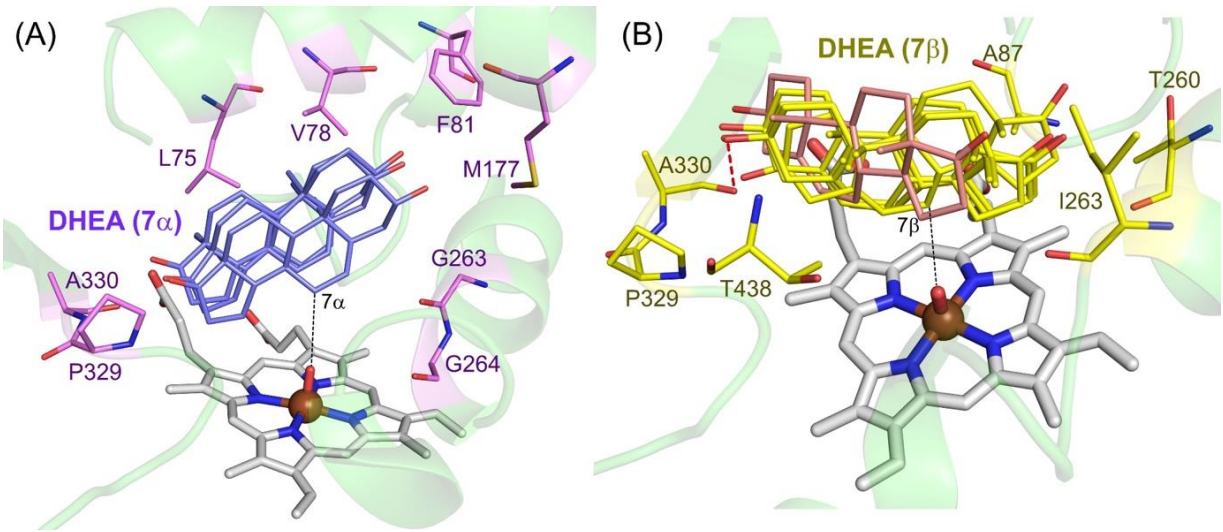


Figure 4.23: **(A)** GV/A184I/I263G/A264G/A328G variant with docked DHEA (purple) bound in 7α pose. Key contacting residues (magenta) within van der Waals contact distance of DHEA are L75, V78, F81, M177, G263, G264, P329 and A330 residues. **(B)** R19/F87A/A264G/A328G variant with docked DHEA bound in 7β pose. 4 DHEA poses were well aligned (in yellow) and the other one was at a different angle (in salmon). A H–bond was observed between the C3–OH in DHEA molecules and the A330 carbonyl. Key contacting residues (yellow) within van der Waals contact distance of DHEA are A87, T260, I263, P329, A330 and T438 residues.

The R19/F87A/A264G/A328G variant favoured the formation of 7β-hydroxy-DHEA, **20a,** (89%). The docking showed 14 productive poses from 4 clusters that consisted of C1β, C2β, C4β, C7β and C15β oxidation. C7β and C2β were the dominant poses, with 5 and 4 poses, respectively, even though C2β oxidation was not observed in the screening. Of the five 7β poses, four were well aligned and shown in yellow in Figure 4.23–B while the other pose was at a different angle (shown in salmon). A hydrogen bond was observed between the C3–OH in DHEA and the A330 carbonyl. The side chains of A87, I263 and T260 were in van der Waals contact with the C17 carbonyl while the side chains of I263 and T438 contacted the C18 and C19 methyl groups, respectively.

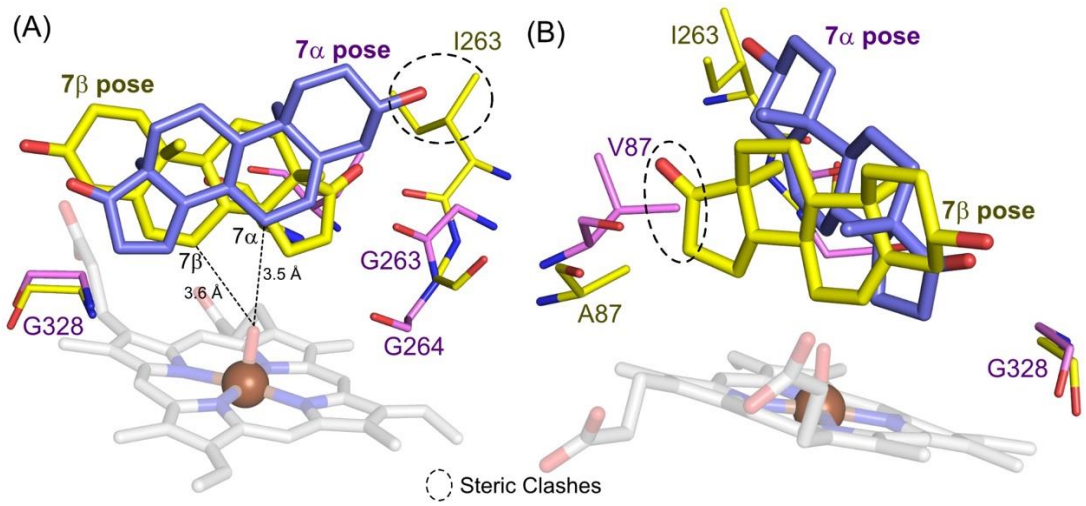


Figure 4.24: **(A)** Variant GV/A184I/I263G/A264G/A328G (magenta) docked with DHEA in a 7α pose (blue) is overlaid with variant R19/F87A/A264G/A328G (yellow) docked with DHEA in a 7β pose (yellow). **(B)** Back view of (A) showing the flip of DHEA between the 7α (blue) and 7β (yellow) poses.

The lowest energy poses of these two variants giving 7α and 7β alcohols are overlaid in Figure 4.24. As shown in Figure 4.24–A, in the 7β poses in the F87A-based variant (in yellow), the β-side of DHEA is facing the haem with the side chain of A87 contacting the DHEA substrate. DHEA is flipped over in the 7α poses in the F87V-based variant (in blue). The α-side of DHEA is facing the haem, such that the C18 and C19 methyl groups contacted the V87 side chain. The

7α pose would not be stable in the F87A-based variant which had steric clashes between the C3–OH and the I263 side chain (Figure 4.24–A). Likewise, the F87V-based variant would not accommodate the 7β pose due to the clashes with the side chain of V87 (Figure 4.24–B).

The R19/F87A/T260G/A328G variant afforded 26% of 1β-hydroxy-DHEA (**20f**) together with 70% of the 7β alcohol **20a**. DHEA was docked into the 4 clusters of this variant which generated 36 poses in total. These poses included 2 poses of C1β, 1 pose of C7β, 3 poses of C15β, 2 poses of C11β and 1 pose of C16β, although only 1β and 7β oxidation were observed in the reaction. Two of the lowest energy DHEA binding poses corresponded to 1β oxidation, while the third lowest energy pose had 7β and 15β C–H bonds closest to the ferryl oxygen. The allylic nature of C7 would account for the 7β alcohol being the major product.

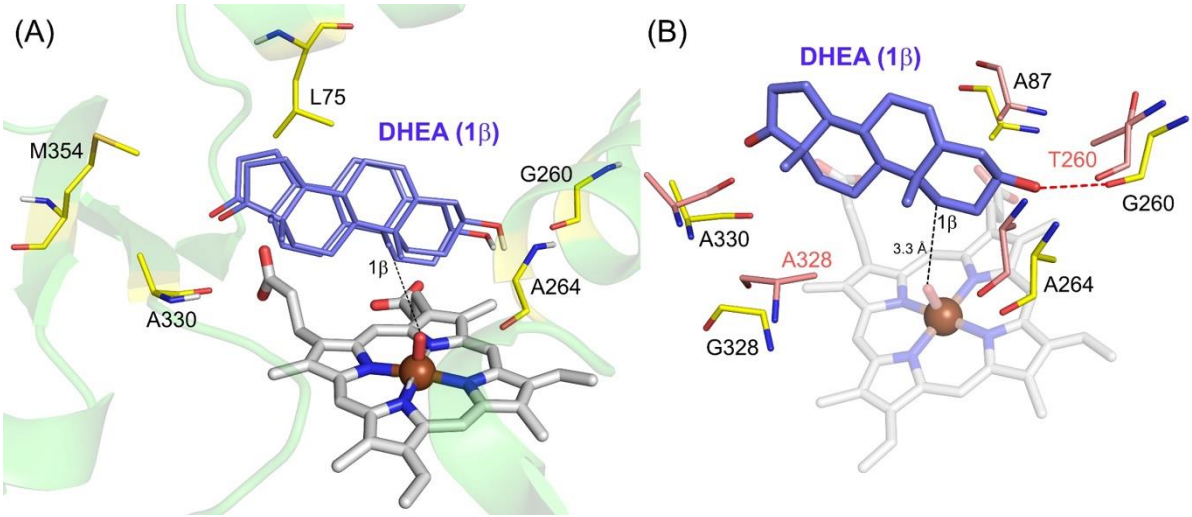


Figure 4.25: **(A)** The overlaid structure of two 1β binding poses (blue) from 2 clusters of variant R19/F87A/T260G/A328G. Key contacting residues L75, G260, A264, A330 and M354 are shown in yellow. **(B)** R19/F87A/T260G/A328G variant docked with DHEA in 1β pose (blue) overlaid with R19/F87A precursor (salmon). Hydrogen bond between the C3-OH and the G260 backbone carbonyl group is shown in red.

DHEA was bound in similar orientations in the two 1β poses (Figure 4.25–A). A hydrogen bond was found between the C3–OH and the G260 backbone carbonyl oxygen. C1 of DHEA was 3.3 Å from the ferryl oxygen, and C19 was at 3.2 Å. The absence of C19 oxidation

suggested that the more activated C–H bond of C1 was preferentially oxidised. When this binding pose was overlaid with the structure of the R19/F87A precursor (Figure 4.25–B), the C3–OH group of DHEA clashed with the T260 side chain of R19/F87A. Therefore, the I helix glycine mutation T260G was crucial for 1β hydroxylation.

4.4.5 TST oxidation with the second-generation library

Formation of the A ring oxidation products, 2β-hydroxy-AD (**19f**) and 1β-hydroxy-DHEA (**20f**), encouraged screening of the second-generation variants for the oxidation of TST, **21**. The second-generation variants gave 3 new products, one was from A ring oxidation, 1β-hydroxy-TST (**21d**),[283] another was the allylic oxidation product 6β-hydroxy-TST (**21e**),[264] and the third was the isomer of 15β alcohol, 15α-hydroxy-TST (**21f**), which were labelled in blue in Table 4.9.[271]

The I helix glycine mutation T260G promoted TST 1β-oxidation. R19/F87A/A184I/T206G gave 28% of 1β-hydroxy-TST (**21d**) whereas its precursor R19/F87A/A184I formed 56% of the previously reported 2β-hydroxy-TST (**21a**), with no evidence for **21d** (Table 4.9, Entries 1 & 2). The additional β4 strand mutations further increased the selectivity for **21d**, with R19/A184I/T260G/A328G and R19/F87A/T260G/P329G/A330W variants giving 51% and 71% of **21d**, respectively (Entries 3 & 4, Figure 4.26-iv).

In DHEA oxidation, the addition of I263G mutation shifted its oxidation site from 7β to 7α position. Likewise, in TST oxidation, the I263G mutation promoted the isomer of the 15β alcohol (**21b**), 15α-hydroxy-TST **21f**. The GV/A184I/I263G variant gave 23% of **21f** whereas its precursor GV/A184I only formed **21b**, at 76% (Entries 6 & 5). Based on GV/A184I/I263G, the introduction of A264G and A328G gave 41% and 40% of **21f**, respectively (Entries 8 & 7).

Inclusion of these two mutations, the GV/A184I/I263G/A264G/A328G variant further raised the proportion of **21f** to 73% (Entry 9, Figure 4.26-vi). Other than the I263G/A264G/A328G combination, the mutation I263W also gave high selectivity of **21f** (75%) but at lower conversion (Entry 10). The GV/A184I/I263G variant not only gave 23% of 15α-hydroxy-TST (**21f**), but also formed 18% of a new oxidation product, 6β-hydroxy-TST, **21e** (Entry 6). The addition of A328G further increased the selectivity for **21e** to 27% (Entry 7, Figure 4.26-v).

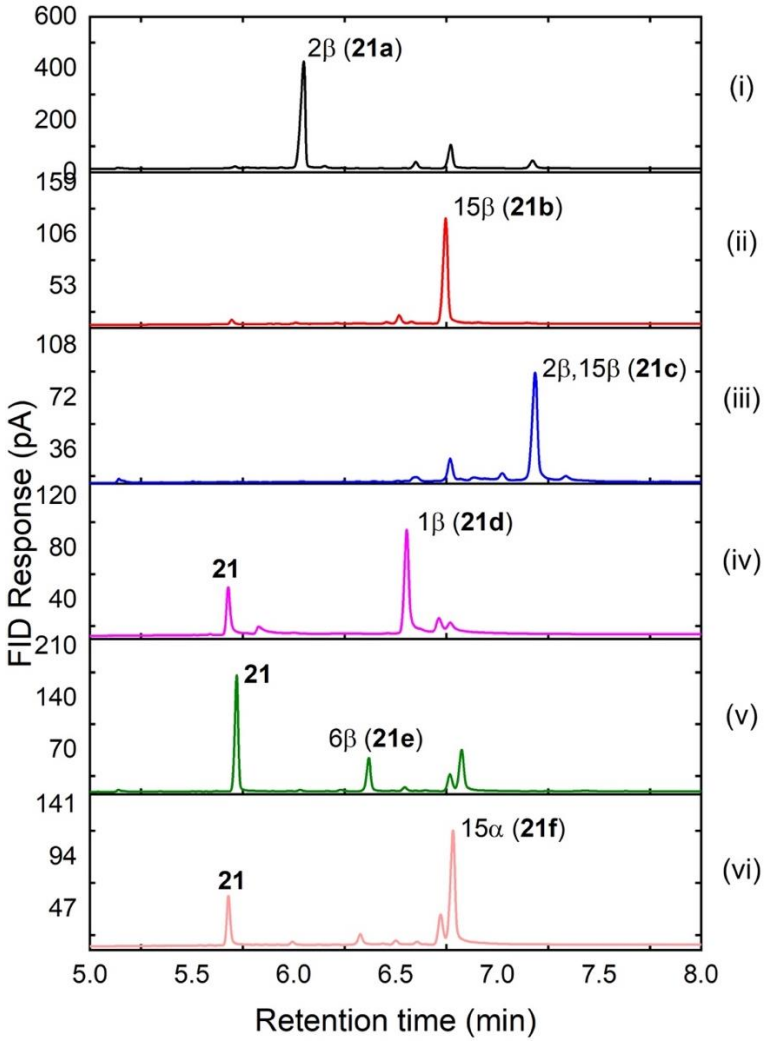
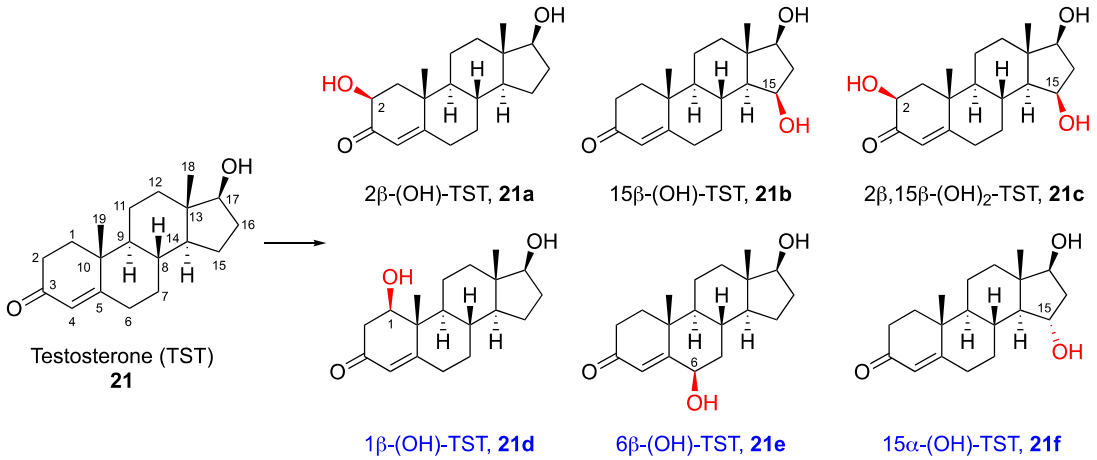


Figure 4.26: Gas chromatographic analysis of the oxidation of TST (**21**) catalysed by CYP102A1 variants: (i) 2β-hydroxy-TST **21a** (66%) by R19/F87A/A184I/L262G; (ii) 15β-hydroxy-TST **21b** (83%) by K19/F87A/A82M/I263A/A330W; (iii) 2β,15β-dihydroxy-TST **21c** (72%) by K19/F87A/A82M/A264G; (iv) 1β-hydroxy-TST **21d** (71%) by R19/F87A/T260G/P329G/A330W; (v) 6β-hydroxy-TST **21e** (27%) by GV/A184I/I263G/A328G; (vi) 15α-hydroxy-DHEA **21f** by GV/A184I/IG/A264G/A328G (73%).

Table 4.9 Selectivity and turnover number of $P450_{BM3}$ variants in the second-generation library for the oxidation of TST. [a]



Entry	Variants	21a	21b	21d	21e	21f	Others	Con.	TON
1	R19/F87A/A184I	56%	13%	**4%**			27%	100%	40
2	R19/F87A/A184I/T260G	18%	52%	**28%**			2%	90%	252
3	R19/F87A/A184I/T260G/A328G	4%	13%	**51%**		31%	1%	100%	510
4	R19/F87A/T260G/P329G/A330W[a]		11%	**71%**		10%	8%	76%	270[b]
5	GV/A184I*	7%	**76%**				17%	96%	730
6	GV/A184I/I263G		37%		**18%**	23%	22%	63%	113
7	GV/A184I/I263G/A328G		15%		**27%**	40%	18%	57%	154
8	GV/A184I/I263G/A264G	6%	36%			**41%**	17%	45%	185
9	GV/A184I/I263G/A264G/A328G		15%		5%	**73%**	7%	82%	600
10	GV/A184I/I263W		7%			**75%**	18%	45%	338

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 2 mM TST, 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 6 h. [b] TST: enzyme ratio = 500: 1. "Conversion" refers to the % of substrate converted to products. "TON" refers to the turnover number for the formation of the product in bold. *Denotes variants in the initial screening library.

4.4.6 Computational studies of TST oxidation products

15α-Hydroxy-TST (**21f**) was a new TST oxidation product for CYP102A1. Variant GV/A184I/I263G/A264G/A328G gave 73% of **21f** and 7% of the isomer 15β-hydroxy-TST **21b**. The three replicas of this variant generated five clusters for TST docking which only gave 11 productive poses, including C1β, C7α, C11α, C15α and C15β oxidation. C15α and C7α were the dominant positions, with 3 and 4 poses, respectively. These 3 poses showing the similar TST binding orientations are overlaid in Figure 4.27–A. The A330 carbonyl oxygen

was hydrogen-bonded to the C17-OH in TST, while the side chain of M177 contacted the C3 carbonyl oxygen on the other side. Residues L75 and V78 contacted the C18 and C19 methyl groups of TST, presenting the α-face to the ferryl oxygen.

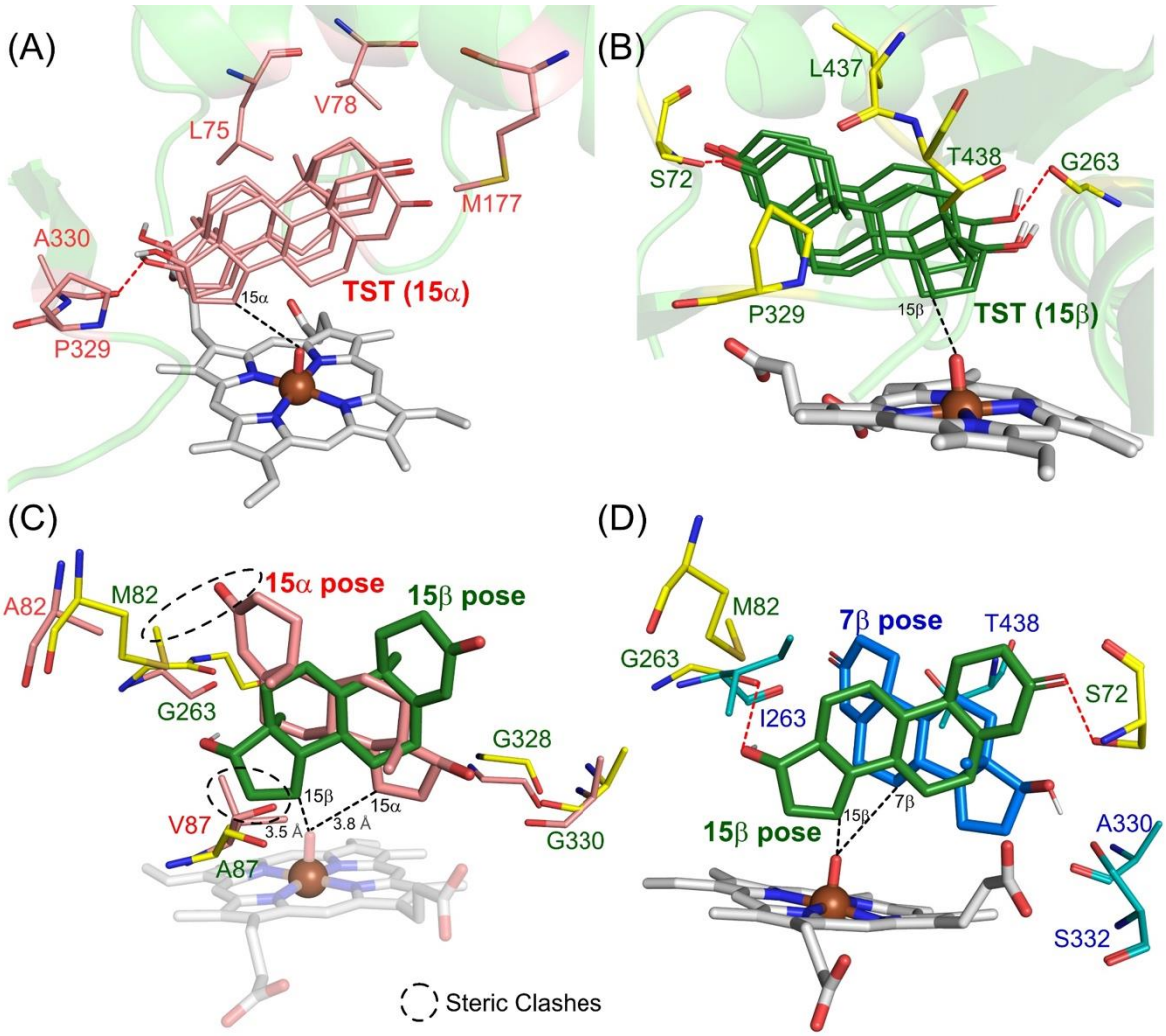


Figure 4.27: **(A)** The overlaid structure of three 15α binding poses (salmon) from clusters of GV/A184I/I263G/A264G/A328G; key contact residues are shown in salmon. **(B)** The overlaid structure of three 15β binding poses from 5 clusters of variant K19/F87A/A82M/I263G/A264G/A328G; key contact residues are shown in forest. **(C)** The overlaid structure of variants in (A) and (B) showing 15α and 15β docking poses. **(D)** V21 (sky blue) from literature promoting 7β alcohol,[287] generated by mutagenesis in Pymol with TST docked in 7β pose (blue), which is overlaid with the 15β promoting variant (forest) in (B).

Variant K19/F87A/A82M/I263G/A264G/A328G, on the other hand, mainly promoted the 15β alcohol **21b** (56%), with 2β-hydroxy-TST **21a** as a minor product (14%). Five clusters were generated from the three replicas of this variant, and the docking with TST showed 11

productive poses. These poses contained 7 poses of C15β, 3 poses of C2β and 1 pose of C1β oxidation. The three lowest energy poses of 15β oxidation were overlaid in Figure 4.27–B. Hydrogen-bonds were formed at both ends of the TST molecule, which helped stabilise the 15β docking pose. One was between the side chain of S72 and the C3 carbonyl oxygen, the other was between the G263 carbonyl and C17-OH. Residues P329, L437 and T438 contacted the TST molecule within van der Waals contact distance.

The two lowest energy poses of these two variants are overlaid in Figure 4.27–C, showing the flipped over docking poses of 15α (salmon) and 15β (forest) alcohols. The comparison shows that the 15β pose would be unstable in the F87V-based variant, due to the steric clash from the side chain of V87. Similarly, the F87A-based variant would not adopt the 15α pose due to the clash from the M82 side chain.

Both 7β and 15β oxidation were observed with AD and DHEA with reasonable selectivity and turnover number. The docking results also suggested that C7 and C15 were both close to the ferryl oxygen. However, only C15 oxidation was observed for TST oxidation in the second-generation screening. 7β-Hydroxy-TST was reported recently by Reetz and co-workers, which was provided by the 15-residue site-saturation mutagenesis with high-throughput screening followed by further mutagenesis of CYP102A1 variants.[171] The reported variant (V21) was generated by mutagenesis in Pymol into which TST was docked. The docking showed nine poses, only one was productive which indicated C7β oxidation. In Figure 4.27–D, this pose is overlaid with variant K19/F87A/A82M/I263G/A264G/A328G from the second-generation library, which adopted 15β pose with hydrogen-bond formed on both end of TST. In V21 variant, residues A330 and S332 contacted the C17-OH of TST while I263 and T438 contacted the C3 carbonyl oxygen, all within van der Waals contact distance. Interestingly, there was no

hydrogen bond to either end of the TST molecule. This suggested that the absence of strong interaction with TST may be the reason of forming 7β oxidation in V21 or disfavouring the existing hydrogen-bond in the F87A-based variant may promote more B ring oxidation.

4.5 Summary

The initial screening library of variants oxidised AD at the 6β, 16α and 16β positions with medium to low selectivity. The new approach of scanning glycine mutagenesis in the I helix and the β4 strand led to new oxidation sites for CYP102A1 at C2, C7 and C15 on the β face (Figure 4.28).[172] 7β-Oxidation of AD was also reported recently by Reetz and co-workers, at 90% selectivity (*vs.* 81% by R19/F87A/A184I/T260G/A328G).[171] On the other hand, 2β and 6β-hydroxylation of AD were only known in fungi.[271,283,292]

The initial library gave DHEA oxidation at C7, C15 and C16 positions on the β face, with 16β-hydroxy-DHEA formed in high selectivity by F87V-based variants. The second-generation CYP102A1 variants not only broadened the range of DHEA products, giving 1β-, 7α- and 15α-oxidation, but also significantly improved their selectivity and turnover number (Figure 4.29).[172] CYP106A2 from *Bacillus megaterium* gives 90% 7β-oxidation of DHEA (*vs.* 98% by R19/F87A/A184I/T260G/A328G), but no 7α oxidation (*vs.* 96% by K19/F87A/A82M/I263G/A328G).[276] The 15β- and 16β-alcohols of DHEA have only been reported in fungi,[271,285] while there is only one report of 1β-hydroxy-DHEA by chemical synthesis.[294]

Previous work on TST oxidation has demonstrated 2β- and 15β-hydroxylation with high selectivity and turnover number. With the screening of the second-generation library, other

oxidation sites have also been discovered, including 1β-, 6β- and 15α-hydroxylation of TST (Figure 4.30).[172] Both 1β- and 15α-hydroxylation of TST have only been reported in fungi.[283]

Steroid A-ring oxidation has been observed on all three substrates. C2 oxidation, given by F87A-based variants, is common for TST as it is activated position but not for AD. 2β-Hydroxy-AD was not observed until the screening with the second-generation variants. The T260G/A328G mutation combination shifted from the 2β alcohol of TST to 1β alcohol, same as for DHEA on the promotion of 1β-hydroxylation. However, the same combination of mutations gave B-ring oxidation for AD to give the 7β alcohol. Interestingly, variants promoting 7β-hydroxy-AD also favoured the formation of 7β-hydroxy-DHEA, with the 1β-hydroxy-DHEA as minor product. The I263G mutation gave 6β hydroxylation of AD and TST, whereas in DHEA it promoted the 7α alcohol stereoselectively. 16β-Oxidation of AD and DHEA were mainly promoted by F87V-containing variants, which instead gave the 15β alcohol for TST, with no sign of C16 oxidation. Of the I helix glycine mutations, the F261G and E267G had little effect on product selectivity in the reaction. Variants with the I259G mutation showed poor expression level, hence lower enzyme activity in the screening.

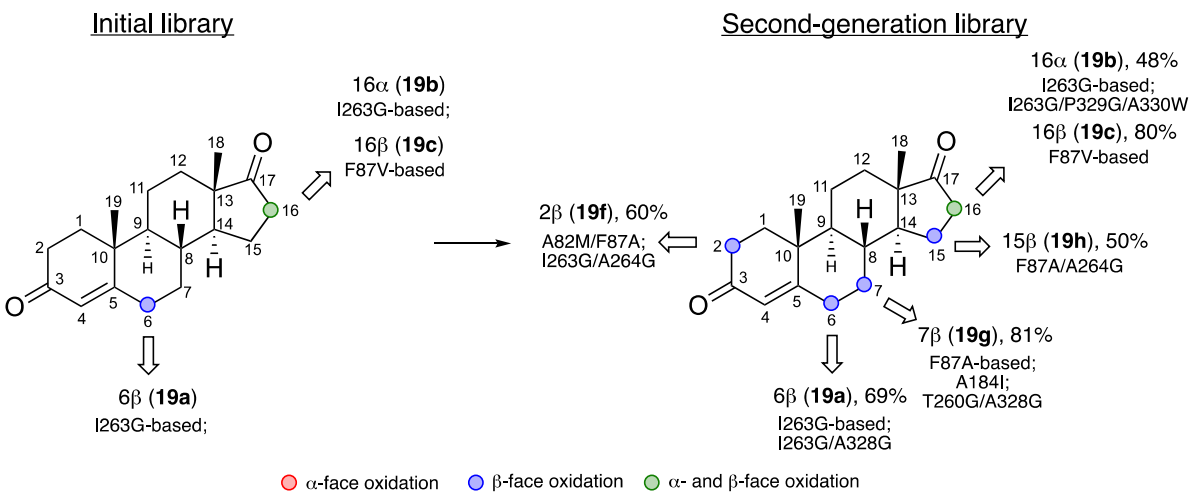


Figure 4.28: Summary of mutations for the promotion of different AD oxidation products.

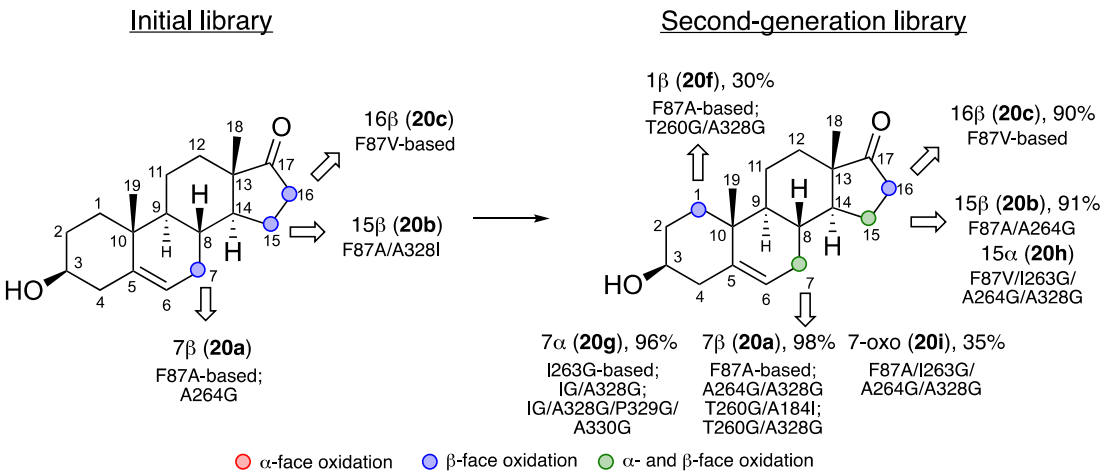


Figure 4.29: Summary of mutations for the promotion of different DHEA oxidation products.

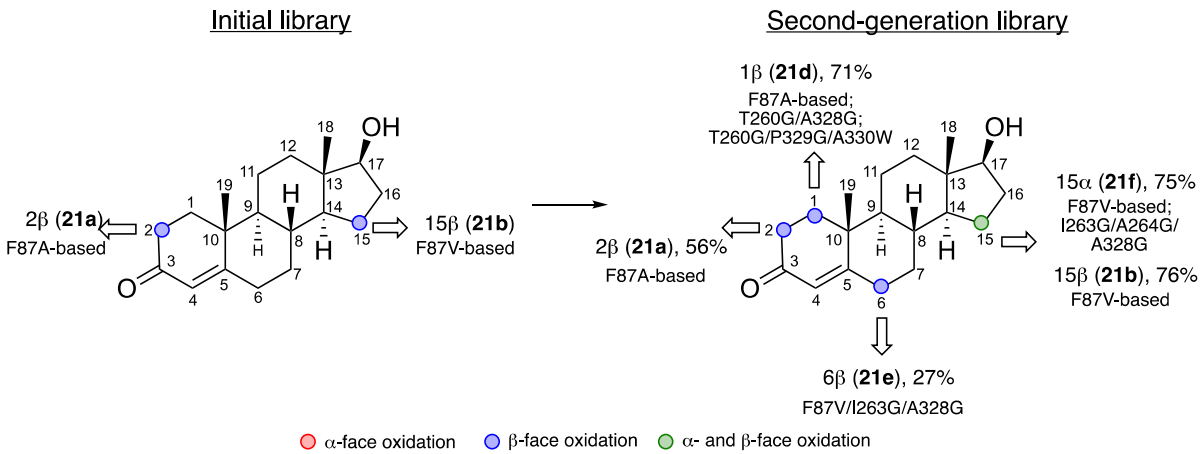


Figure 4.30: Summary of mutations for the promotion of different TST oxidation products.

Other non-active-site residues, L262G and T269G, however, were effective in giving high selectivity of 15β-hydroxy-DHEA. The S270G significantly increased the substrate conversion and generated more further oxidation products including 2β,16β-dihydroxy-AD and 7β,15β-dihydroxy-DHEA. This suggested that these glycine mutations had caused conformational changes on the I helix that enhanced steroid oxidation. The most important mutations in the I helix are T260G, I263G and A264G. T260G was particularly crucial for the A-ring oxidation of DHEA and TST. The docking results suggested that the glycine mutation on T260 greatly reduced the clash between the side chain and the substrates, therefore, allowing them to adopt

a 1β pose. I263G played important roles in promoting oxidation products on the α-face, such as 16α-hydroxy-AD, 7α-hydroxy-DHEA and 15α-hydroxy-TST. The combination of T260G, I263G and A264G with F87A and F87V increased the product diversity. Especially, with the addition of the β4 strand glycine mutations, the product range has been greatly expanded.

In conclusion, scanning and combination of glycine mutagenesis, inspired by the overlaid structures of CYP19A1 and CYP102A1, has provided access to mono- and dihydroxylation of steroids via direct C–H bond activation. In order to further increase the selectivity for these monooxygenation products, site-saturation mutagenesis on effective residues selected from computational simulation will be conducted, such as S72, A82, F87, A184, A328, P329, A330, and L437. Although some of these sites may have been targeted for saturation mutagenesis and gave thousands of variants,[169,170] the reported product range was still limited. Therefore, it seems that directed evolution, rational design, MD-simulation and docking study will make a better combination that enables easier construction of P450 variants with improved selectivity and high activities.

Chapter 5

5 Oxidation of progestogens and synthetic steroids

5.1 Introduction

Progestogens are a class of steroid hormones that regulate and maintain the female reproductive system. Pregnenolone (PREG) is one of the most important progestogens that is produced from cholesterol in human steroidogenesis. PREG is described as a neurosteroid and anti-inflammatory drug that was used in the treatment of rheumatoid arthritis and soft-tissue rheumatism in 1950s. It is also the biosynthetic precursor to most of the steroid hormones including other progestogens, androgens, estrogens, glucocorticoids, and mineralocorticoids. PREG is converted to 17α-hydroxy-PREG by the microsomal steroid 17α-hydroxylase (CYP17A1) for the subsequent formation of DHEA and 17α-hydroxy-progesterone.

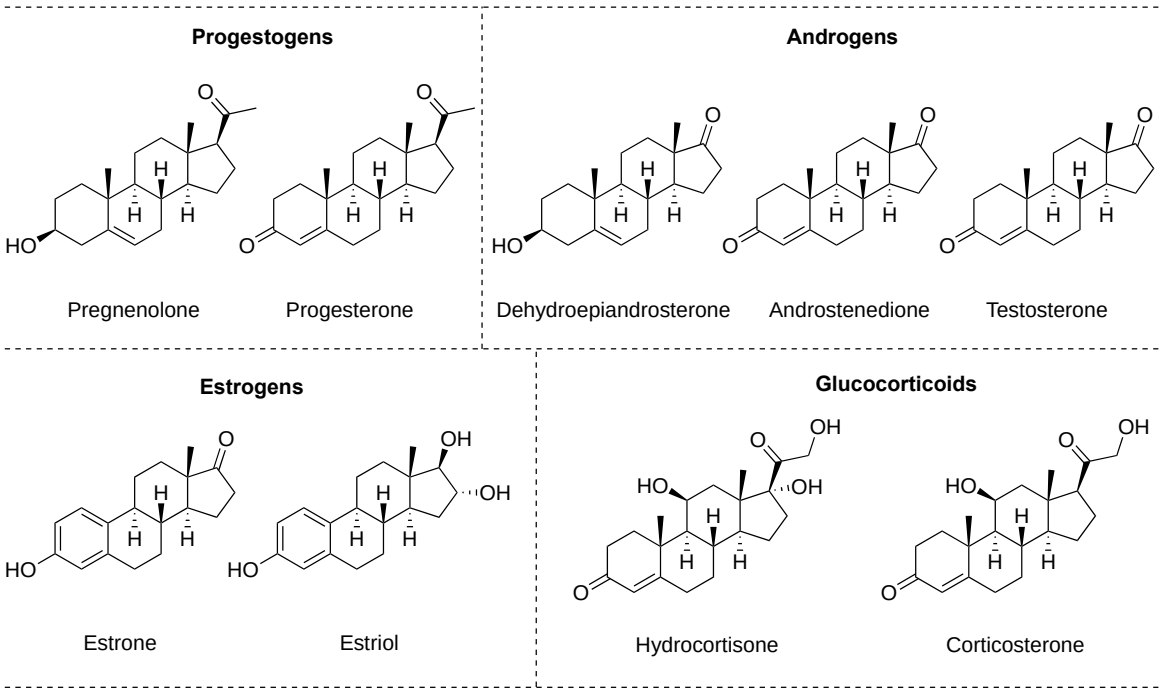


Figure 5.1 The four major classes of steroids generated in human steroidogenesis.

Progesterone (PROG) is also generated from PREG by a 3β-hydroxysteroid dehydrogenase (3β-HSD). Progestins are synthetic progestogens that are used in medicine, such as the 17α-hydroxy-progesterone derivative medrosyprogesterone acetate and the 19-nortestosterone derivative norethisterone. These progestins are structural analogues of progesterone and possess progestogenic activity. In additional to their roles as natural hormones, progestogens and their derivatives are of great interest for the potential biological and pharmaceutical values.

Androsterone is the main metabolite of testosterone and possesses weak androgenic activity. There are three isomers of androsterone, each with different pharmacological properties – epi-androsterone (3β-isomer), etiocholanolone (5β-isomer), and epi-etiocholanolone (3β,5β-isomer) (Figure 5.2–A).

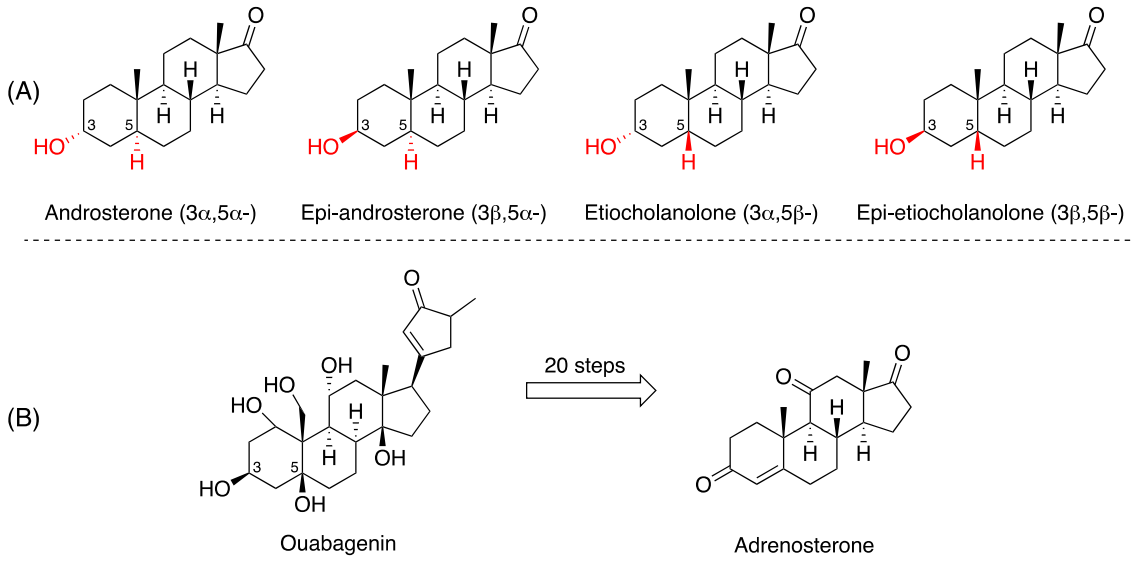


Figure 5.2 (A) The structure of four isomers of androsterone; **(B)** Scalable synthesis of ouabagenin from adrenosterone.[257]

Androsterone and etiocholanolone are the main metabolites of testosterone and dihydrotestosterone (DHT) that possess anticonvulsant effects.[296,297] Epi-androsterone is naturally produced from DHEA by the enzyme 5α-reductase. Although these steroids are less well known in human metabolism pathways, there is great potential of these steroids and their

derivatives in the synthesis of drug-type molecules, e.g. the semisynthesis of ouabagenin from adrenosterone via innovative methods for C–H and C–C activation and C–O bond homolysis (Figure 5.2–B).[257] Given that the epi-etiocholanolone possesses the same 3β,5β-stereochemistry as the target product ouabagenin, using epi-etiocholanolone as the starting material might shorten a few steps and provide a more efficient semisynthesis route for ouabagenin.

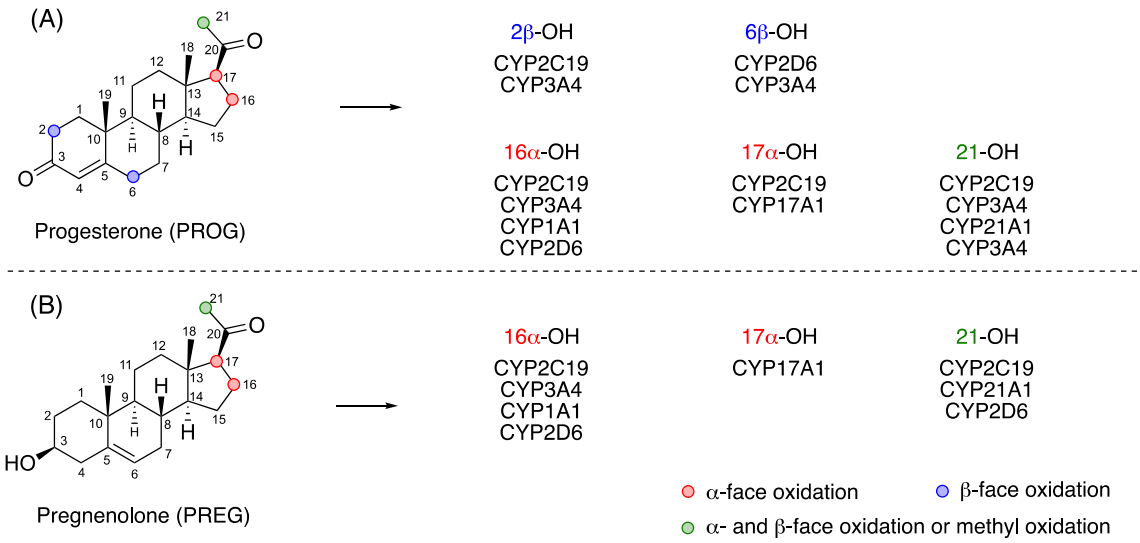


Figure 5.3 Summary of oxidation products of PROG and PREG by various human P450s.

Since PROG plays vital roles in human steroidogenesis, the metabolism of PROG by various human P450s has been thoroughly studied (Figure 5.3).[261] It was demonstrated that CYP2C19 mainly oxidised PROG to form 21-hydroxy-PROG, with 16α- and 17α-hydroxy-PROG as minor products.[298] CYP3A4 oxidised PROG to produce 16α-, 6β- and 2β-hydroxy-PROG as major products with a small amount of the 21-hydroxylation product (Figure 5.3–A).[261] As for pregnenolone oxidation, 16α-hydroxylation was catalysed efficiently by CYP1A1 and CYP3A4.[299] CYP17A1 and CYP2C9 catalysed the less common 17α- and 21-hydroxylation of

PREG (Figure 5.3–B).[261] No report of oxidation of androsterone (AN) and epi-androsterone (EAN) by human P450s was found.

Other than human P450s, fungal biotransformation of EAN, PROG and PREG has also been studied (Figure 5.4). It is interesting to observe that the oxidation by fungal strains was mainly on the α-face of the molecules. No oxidation of AN has been reported for fungal strains. The oxidation sites for EAN were limited to the 7α-,[300] 7β-,[301] 9α-[300] and 11α-positions;[300] PREG was oxidised at the 7α-,[261] 11α-[302] and 15α-sites[284] (Figure 5.4–A & C) but no 17- or 21-oxidation was observed. PROG was mainly oxidised at the C- and D-ring positions 11α-,[303] 11β-,[304] and 15α-[261] (Figure 5.4–B).[305] No A- and B-ring oxidation or 21-hydroxylation of PROG was observed with fungus.

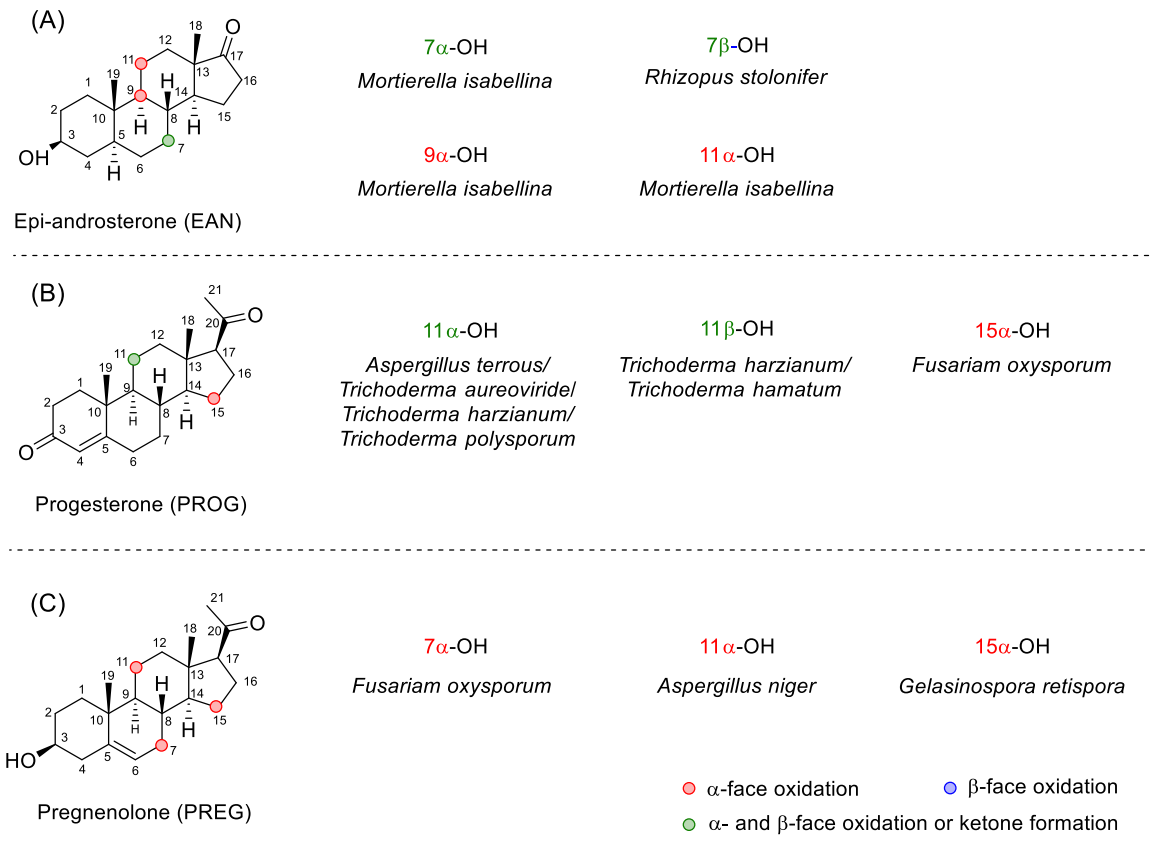


Figure 5.4 Summary of oxidation products of EAN, PROG and PREG by selected fungal strains.

The oxidation of PROG by recombinant bacterial P450s has been extensively studied; oxidation at every position except C7, C8, C12, C14 and C21 has been reported for various P450s. CYP106A2 from *Bacillus megaterium* catalysed 15β-hydroxylation of PROG.[273-275,306] Further engineering shifted the selectivity away from the 15β-alcohol to 9α-, 6β-, and 11α-hydroxylation.[306] CYP106A2 also catalysed the C7β oxidation of PREG.[276] CYP260A1 from *Sorangium cellulosum* So ce56 selectively oxidised C-19 steroids such as AD and TST at the 1α-position, but the selectivity for PROG was low.[277] With semi-rational protein engineering, two highly regio- and stereo-selective variants, S276N and S276I, were identified for the formation of 1α- and 17α-hydroxylation of PROG.[307] CYP154C3 and CYP154C5 are highly selective for 16α oxidation of PROG and PREG.[279,280] The engineering of CYP102A1 ($P450_{BM3}$) from *Bacillus megaterium* led to high regio- and stereo-selectivity for 2β- and 16β-oxidation of PROG.[169] To the author's knowledge, the oxidation of AN and EAN by any bacterial P450s has not been reported.

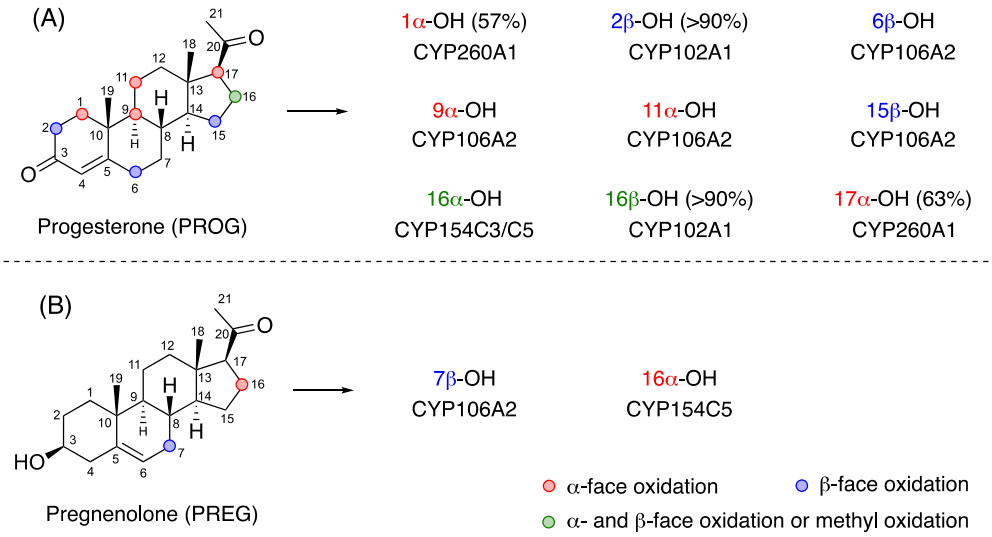


Figure 5.5 Summary of oxidation products of PROG and PREG by various recombinant bacterial P450s.

As shown in Figures 5.3–5.5, PROG oxidation by various fungal strains and P450s have been studied extensively; oxidation at every position except for C7 and C12 has been found. PREG has only been reported to be oxidised at the 7β- and 16α-positions by two bacterial P450 enzymes. The oxidation of AN and EAN, however, was very limited, especially for AN; no P450s or fungal oxidation has been reported.

There are two aims for the work described in this chapter. The first is to generate a diverse range of oxidation products of AN, EAN, PROG and PREG. The second is to compare the oxidation sites of these substrates and correlate these with the effective mutations and gain insights for further engineering with the aid of MD simulations and substrate docking studies.

5.2 Oxidation of androsterone (AN)

Androsterone (AN, **22**) was a good substrate for the CYP102A1 variant library; 46 out of 105 variants showed >20% conversion (Appendix 13). Six major products were purified and characterised as androstanedione (**22a**),[308] 1β-hydroxy-AN (**22b**),[309] 7β-hydroxy-AN (**22c**),[309] 15β-hydroxy-AN (**22d**),[310] 16β-hydroxy-AN (**22e**),[311] and 4α,16β-dihydroxy-AN (**22f**) (Figure 5.6 and Table 5.1). The NMR data for **22a**–**22e** (Appendix 13) agreed with literature data. Product **22f** was new metabolite of AN; full characterisation data are collected in Appendix 13.

The conversion rates showed that variants without F87A or F87V mutations were inactive for AN oxidation except for the variant R19/A330W which gave 90% of **22a**, the C3 alcohol oxidation product at 32% conversion (Table 5.1, Entry 1). As observed for AD and DHEA oxidation, the F87A mutation favoured the formation of 7β-hydroxy-AN **22c**. Glycine mutations in the I helix (A264G) and β4 strand (A328G) improved the selectivity for **22c** from 24% for R19/F87A to 58% and 49% respectively (Table 5.1, Entries 4, 5 and 15). Adding both

glycine mutations increased the selectivity to 86% with R19/F87A/A264G/A328G (Entry 9). The combination T260G/A328G also improved the selectivity for **22c**, to 78% in variant R19/F87A/T260G/A328G but with lower conversion (Entry 8). Addition of A184I to these two glycine combinations increased the selectivity for **22c** to 90% and 96% with full conversion for the variants R19/F87A/A184I/A264G/A328G and R19/F87A/A184I/T260G/A328G (Entries 10 & 12, Fig. 5.6–i).

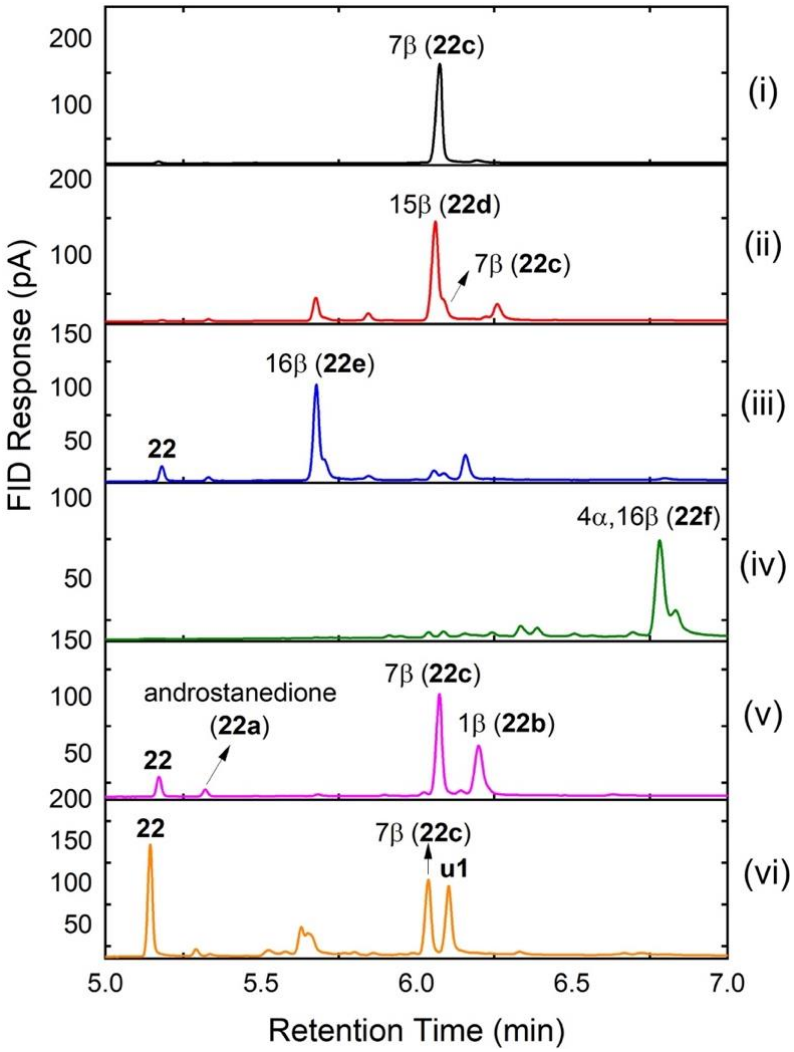
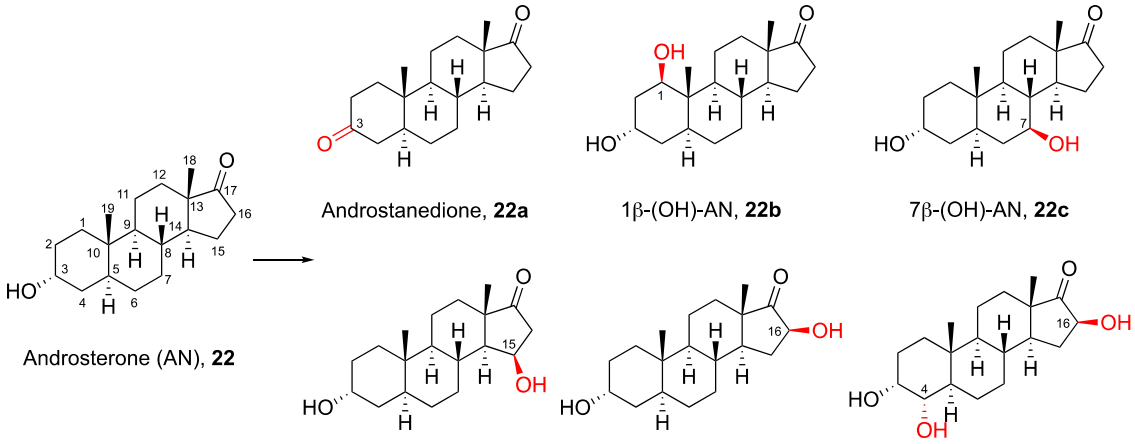


Figure 5.6 Gas chromatography analysis of selected variants giving AN oxidation products **22a**–**22e**: (i) 96% of 7β-hydroxy-AN **22c** by R19/F87A/A184I/T260G/A328G; (ii) 65% of 15β-hydroxy-AN **22d** by GV/A184I/I263G/A264G/A328G; (iii) 57% of 16β-hydroxy-AN **22e** by R19/F87A/A184I/I263G/A264G; (iv) 54% of 4α,16β-dihydroxy-AN **22f** by K19/F87A/A82M/S270G; (v) 53% of **22c** and 38% 1β-hydroxy-AN **22b** by K19/F87A/A82M/T260G/A328G. (vi) 34% of **22c** and 38% of an unidentified but likely AN mono-hydroxylation product **u1** by K19/F87A/A82M/I263G/A330W.

Table 5.1 Selectivity and turnover number of 2nd generation $P450_{BM3}$ variants for the oxidation of AN, **22**.[a]



Entry	Variants	22a	22b	22c	22d	22e	22f	Others	Conv.	TON[b]
1	R19/A330W	**90%**						10%	32%	144
2	R19/F87A/F81W/T260G/A328G		**27%**	67%				6%	56%	76
3	K19/F87A/A82M/T260G/A328G		**38%**	53%				9%	93%	177
4	R19/F87A			24%	29%	17%	5%	25%	60%	
5	R19/F87A/A264G			**58%**	18%			24%	67%	195
6	R19/F87A/A184I			**44%**	24%			32%	59%	130
7	R19/F87A/A184I/T260G		**32%**	55%				13%	56%	90
8	R19/F87A/T260G/A328G			**78%**				22%	20%	78
9	R19/F87A/A264G/A328G			**86%**	7%			7%	100%	430
10	R19/F87A/A184I/A264G/A328G			**90%**				10%	100%	450
11	R19/F87A/A184I/A264G			**93%**				7%	100%	465
12	R19/F87A/A184I/T260G/A328G			**96%**				4%	99%	480
13	GV/A184I/I263G	15%			21%	38%		32%	72%	
14	GV/A184I/A328G	34%		5%	6%	20%		35%	43%	
15	R19/F87A/A328G			49%	**40%**			11%	90%	180
16	GV/A184I/I263G/A328G	5%			**52%**	29%		14%	85%	221
17	GV/184I/I263G/A264G/A328G			5%	**65%**	13%		17%	100%	325
18	GVQ	30%				**35%**		45%	26%	46
19	K19/F87V	7%				**40%**		53%	59%	118
20	R19/F87A/A184I/I263G/A328G					**49%**		51%[c]	80%	196
21	R19/F87A/A184I/I263G/A264G			4%	6%	**57%**		33%	93%	266
22	R19/F87A/A184I/I263G/A264G/A328G			16%		**57%**	7%	20%	100%	285
23	K19/F87V/I263G	21%			9%	**62%**		8%	36%	112
24	K19/F87A/A82M			23%	12%	18%	14%	33%	79%	
25	K19/F87A/A82M/F261G			26%	11%	11%	**28%**	24%	79%	111
26	K19/F87A/A82M/L262G						**53%**	47%	100%	265
27	K19/F87A/A82M/S270G						**54%**	46%	100%	270

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 1 mM AN (500:1), 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold. [c] denotes 42% of uncharacterised product **u1**.

The F87V mutation favoured the formation of 16β-hydroxy-AN (**22e**), with variants GVQ and K19/F87V giving 35% and 40% of **22e** (Entries 18 & 19). Addition of I263G to 7β-favouring variants shifted the selectivity towards 16β-hydroxylation, with variants R19/F87A/A184I/I263G/A264G and R19/F87A/A184I/I263GA/264G/A328G both giving 57% of 16β-hydroxy-AN, **22e** (Entries 21 & 22, Fig. 5.6–iii).

As observed for DHEA and TST 1β-hydroxylation, the T260G mutation played important roles in promoting the 1β-oxidation of AN; R19/F87A/F81W/T260G/A328G gave 27% and R19/F87A/A184I/T260G gave 32% of 1β-hydroxy-AN **22b**, with the 7β-alcohol remaining the major product (Entries 2 & 7). K19/F87A/A82M/T260G/A328G gave the highest selectivity and turnover number for **22b** (38%, TON = 177, Entry 3, Fig. 5.6–v).

Compared with the R19/F87A base variant which gave a wide range of products (Entry 4), the extra A328G mutation significantly suppressed the formation of other products while promoting the 7β- (**22c**, 49%) and 15β- (**22d**, 40%) alcohols (Entry 15). The 15β-alcohol **22d** was also favoured by F87V-based variants; GV/A184I/I263G gave 21% and the addition of A328G led to 52% **22d** (Entries 13 & 16). Variant GV/A184I/I263G/A264G/A328G, which combined these mutations, gave the highest selectivity and turnover number for **22d** (65%, TON = 325, Entry 17, Fig. 5.6–ii).

K19/F87A/A82M was one of the active variants that gave multiple products including 18% of 16β-hydroxy-AN **22e**, which likely to be the precursor of the diol 4α,16β-dihydroxy-AN (**22f**). The addition of the I helix mutations F261G, L262G and S270G to K19/F87A/A82M led to 28%, 53% and 54% of **22f**, respectively (Entries 25–27, Fig. 5.6–iv). Finally, the A330W mutation in the variant K19/F87A/A82M/I263G/A330W led to the formation of 38% of an unidentified product (**u1**, Fig. 5.6–vi) together with 34% of 7β-hydroxy-AN (**22c**); the retention

time of **u1** suggested that it was a mono-hydroxylated AN but the moderate conversion (60%) did not allow its purification and characterisation.

5.3 Oxidation of epi-androsterone (EAN)

Epi-androsterone (EAN, **23**) was a poor substrate for CYP102A1 variants in the initial library, with only 6 out of 48 variants showing >20% conversion. The second-generation variants were more active, and 20 out of 67 variants showed >20% conversion. Five products were purified and characterised (data in Appendix 14) as androstanedione (**23a**)[308], 1β-hydroxy-EAN (**23b**), 7β-hydroxy-EAN (**23c**),[309] 15β-hydroxy-EAN (**23d**), and 16β-hydroxy-EAN (**23e**).[312] The NMR data for **23a**, **23c**, and **23e** (Appendix 14) agreed with literature. Products **23b** and **23d** are new metabolites of EAN; full characterisation data are collected in Appendix 14.

The K19/F87A/A82M variant gave 70% of 16β-hydroxy-EAN (**23e**) with 55% conversion. The extra A330W mutation raised the selectivity to 79% while maintaining the conversion level (Table 5.2, Entry 13). Variant GV/A184I gave 88% selectivity for **23e** but with a lower conversion of 37% (Entry 14). Addition of either A264G or A328G improved the conversion for **23e** with a slight trade-off on selectivity (Entries 15 & 16, Fig. 5.7–i). Interestingly, the A264G/A328G combination suppressed the formation of **23e** but gave a new product, 7β-hydroxy-EAN (**23c**), with 32% selectivity (Entry 2). GV/A184I/I263G/A264G/A328G (Entry 4) further increased the selectivity for **23c** to 51%, as well as promoting 44% of another new product, 15β-hydroxy-EAN (**23d**). In contrast, addition of A264G/A328G to R19/F87A/A184I gave 69% of **23c** with high conversion (Entry 5, Fig. 5.7–iii). Other than the A264G/A328G combination, the T260G mutation also favoured the formation of the 7β-alcohol. The addition of T260G to the 16β-alcohol-favouring variant K19/F87A/A82M gave 38% of the 7β-alcohol

23c while the extra A328G increased it to 76% in variant K19/F87A/A82M/T260G/A328G (Entries 3 & 6).

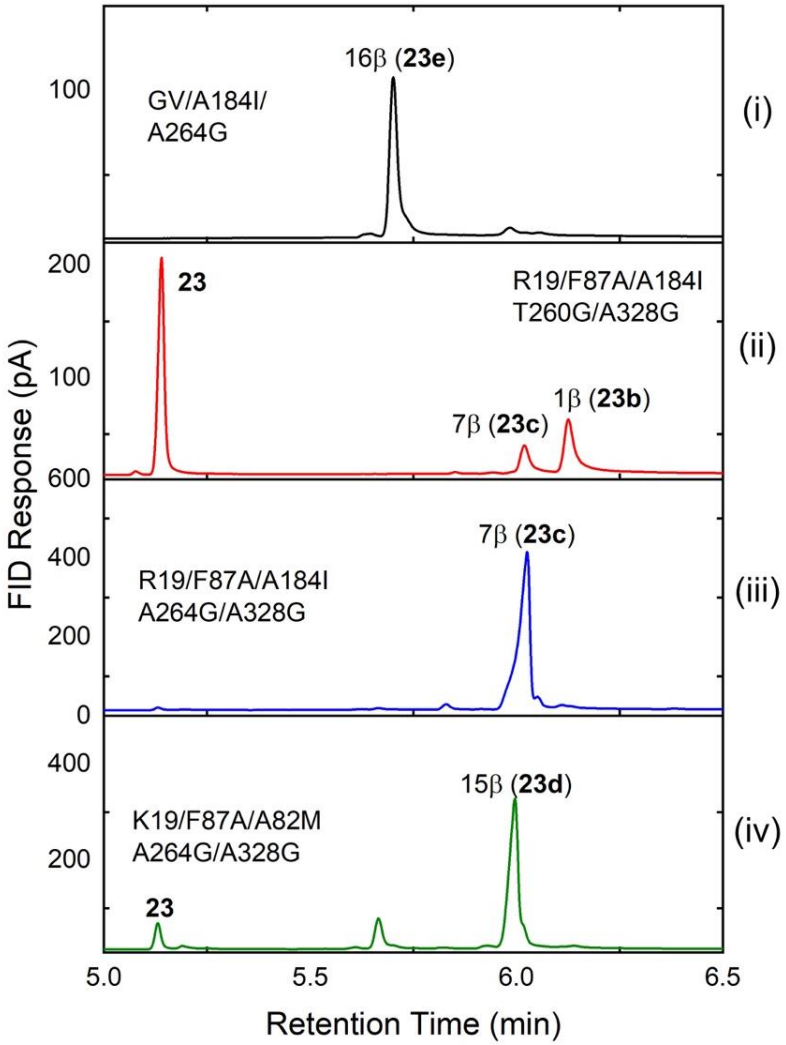
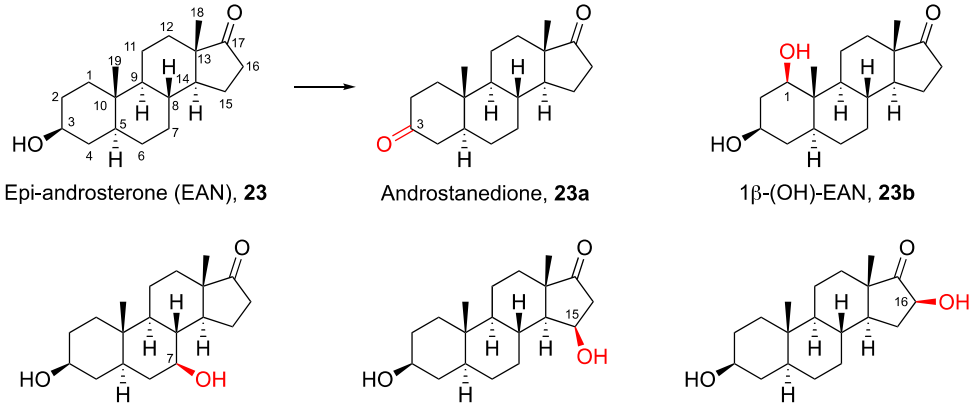


Figure 5.7 Gas chromatography analysis of selected variants giving EAN oxidation products **23a**–**23e**: (i) 77% of 16β-hydroxy-EAN **23e** by GV/A184I/A264G; (ii) 64% of 1β-hydroxy-EAN **23b** by R19/F87A/A184I/T260G/A328G; (iii) 69% of 7β-hydroxy-EAN **23c** by R19/F87A/A184I/A264G/A328G; (iv) 76% of 15β-dihydroxy-EAN **23d** by K19/F87A/A82M/A264G/A328G.

The T260G/A328G combination in variant R19/F87A/A184I/T260G/A328G not only gave 7β-hydroxylation but also promoted 64% of the A-ring oxidation product 1β-hydroxy-EAN, **23b**, (Entry 1, Fig. 5.7–ii). Although the combination A264G/A328G on R19/F87A/A184I and GV/A184I favoured C7 oxidation, the addition of this combination to K19/F87A/A82M shifted the selectivity significantly to 76% of 15β-hydroxy-EAN (**23d**) (Entry 11). GV/A184I/A328G

favoured the 16β-alcohol (81% of **23e**), but the addition of I263G to A328G shifted EAN binding and gave 73% of the 15β-alcohol **23d** (Entry 10, Fig. 5.7–iv). The combination of other glycine mutations on I helix to A328G also favoured the formation of **23d**, but with lower conversion (Entries 7–9).

Table 5.2 Selectivity and turnover number of second-generation $P450_{BM3}$ variants for the oxidation of EAN, **23**.[a]



Entry	Variants	23a	23b	23c	23d	23e	Others	Conv.	TON[b]
1	R19/F87A/A184I/T260G/A328G		**64%**	28%			8%	38%	122
2	GV/A184I/A264G/A328G			**32%**	11%	53%	4%	78%	125
3	K19/F87A/A82M/T260G			**38%**		38%	24%	33%	63
4	GV/A184I/I263G/A264G/A328G			**51%**	44%		5%	58%	148
5	R19/F87A/A184I/A264G/A328G		6%	**69%**	15%		10%	99%	342
6	K19/F87A/A82M/T260G/A328G		**5%**	**76%**		5%	14%	32%	122
7	R19/F87A/T269G/A328G	5%		20%	**50%**	6%	19%	28%	70
8	R19/F87A/S270G/A328G			17%	**55%**	9%	19%	50%	138
9	R19/F87A/L262G/A328G	8%			**59%**	10%	23%	18%	53
10	GV/A184I/I263G/A328G	6%			**73%**	10%	11%	31%	113
11	K19/F87A/A82M/A264G/A328G				**76%**	16%	8%	99%	377
12	K19/F87A/A82M	3%			14%	**70%**	13%	55%	193
13	K19/F87A/A82M/A330W			7%		**79%**	14%	65%	257
14	GV/A184I				5%	**88%**	7%	37%	162
15	GV/A184I/A328G				10%	**81%**	9%	67%	272
16	GV/A184I/A264G			7%		**77%**	16%	100%	385

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 1 mM EAN (500:1), 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold.

5.4 Computational studies of AN and EAN oxidation

Variant R19/F87A/A184I/A264G/A328G was highly selective for 7β-oxidation of DHEA (94% 7β-hydroxy-DHEA, **20a**) and AN (90% 7β-hydroxy-AN, **22b**) but less so for EAN (76% 7β-hydroxy-EAN, **23d**). Docking was performed on both AN and EAN in order to gain more information on the substrates-enzyme interactions. Seven clustered structures were generated from the three replica MD simulations of this variant. AN was docked into these seven clusters and generated 63 poses; 18 of these were productive poses that indicated C1β, C2β, C7β, C11β, C15α/β oxidation. C7β and C15β oxidation were the most populated poses, with 4 and 7 poses, respectively. Three of the four C7β poses were the lowest energy poses for all of these clusters and were stabilised by a hydrogen bond between the side chain of S332 and the AN C3-alcohol (Figure 5.8).

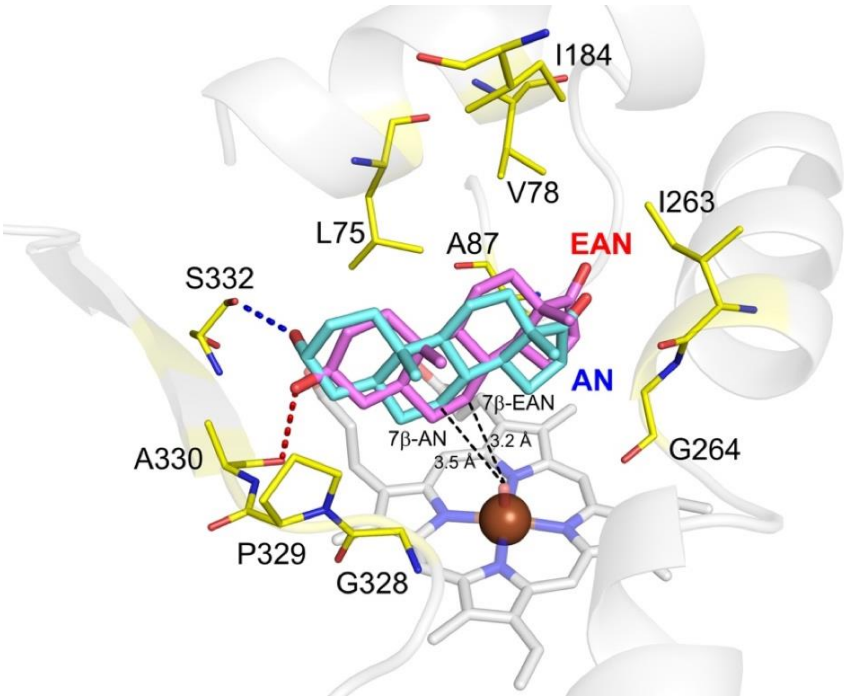


Figure 5.8 The docking poses of 7β-hydroxy-AN (cyan) and 7β-hydroxy-EAN (red) in R19/F87A/A81I/A264G/A328G are overlaid. Hydrogen-bonding between the C3-OH in AN and the side chain of S332 is shown in blue dashes. Hydrogen-bonding between the C3-OH in EAN and the backbone of A330 is shown in red dashes. Other key contacting residues within van der Waals contact distance of AN and EAN ARE L75, V78, A87, G264 and P329.

The docking of EAN into these seven clustered structures gave 15 productive poses, where C1β, C2β, C7α/β, C15β and C16β oxidation were indicated. C1β and C15β were the most populated poses with 4 and 7 poses, respectively and only 1 pose of 7β was observed. This 7β pose was the lowest energy pose for this cluster and was stabilised by a weak hydrogen bond formed between the C3-OH of EAN with the backbone carbonyl oxygen of A330. One of the 7β-hydroxy-AN poses and the 7β-hydroxy-EAN pose are overlaid in Figure 5.8. It is evident that, due to the opposite stereochemistry at C3, the C3-OH in AN (in cyan) appeared to be closer to the side chain of S332 for hydrogen bonding (blue dashes) in this 7β-oxidation pose, while the C3-OH in EAN (purple) was far away from this residue and could only hydrogen-bond to the A330 backbone (red dashes), leading to a less stable 7β-pose. This weakened hydrogen bonding and the lower number of 7β poses, albeit that it was one of the lowest energy poses, are qualitatively consistent with the lower 7β selectivity of EAN oxidation. However, since the affinity energies from the docking algorithm are not sufficiently reliable for firm conclusions to be drawn; further studies at a higher level of theory and MD simulations of the various poses will be required.

5.5 The oxidation of AN and EAN with a third-generation library

In the second-generation variants, the most important mutations were T260G, I263G, A264G and A328G. The combination of these mutations with A184I and A82M increased the product diversity and selectivity. MD simulations and docking studies indicated that these I helix mutations led to local distortions of the I helix to create space for steroid binding. Therefore, further combinations of I-helix glycine mutations, namely T260G/I263G, T260G/A264G, and T260G/I263G/A264G were introduced to R19/F87A, K19/F87A/A82M, R19/F87A/A184I, GV/A184I, and R19/F87A/A328G base variants to generate a small library of 21 variants.

Table 5.3: List of variants in the 3rd generation library.

No.	Variants	No.	Variants
1	R19/F87A/T260G/I263G	12	R19/F87A/A184I/T260G/I263G/A264G
2	K19/F87A/A82M/T260G/I263G	13	R19/F87A/T260G/I263G/A264G
3	R19/F87A/A328G/T260G/I263G	14	K19/F87A/A82M/T260G/I263G/A264G
4	GV/A184I/T260G/I263G	15	R19/F87A/A328G/T260G/I263G/A264G
5	R19/F87A/A184I/T260G/A328G/I263G	16	GV/A184I/T260G/I263G/A264G
6	R19/F87A/A184I/T260G/A264G	17	R19/F87A/A184I/T260G/A328G/I263G/A264G
7	R19/F87A/T260G/A264G	18	R19/F87A/A184I/T260G/A82M
8	K19/F87A/AM/T260G/A264G	19	R19/F87A/A328G/T260G/A82M
9	R19/F87A/A328G/T260G/A264G	20	R19/F87A/A184I/T260G/A328G/A82M
10	GV/A184I/T260G/A264G	21	R19/F87A/T260G/A82M
11	R19/F87A/A184I/T260G/A328G/A264G		

The third-generation variants gave the same range of products as with the second-generation library for AN, with 7β-hydroxy-AN (**22c**) as the major product. The new variants R19/F87A/A184I/T260G/A82M and R19/F87A/A184I/T260G/A264G/A328G (Table 5.4, Entries 7 & 8) gave 88% and 96% of **22c** with almost total conversion. The new variants also improved the selectivity for the A-ring oxidation product 1β-hydroxy-AN (**22b**); combining the

A82, A184I, T260G and A328G mutations that promoted 1β-oxidation led to the R19/F87A/A184I/T260G/A328GA82M variant which gave 48% of **22b** (Entry 1, Fig. 5.9).

Table 5.4 The comparison of variants giving high selectivity of 7β-hydroxy-AN (**22c**) and 1β-hydroxy-AN (**22b**) from the second and third-generation library. The entry number in red indicates variants from the third-generation library.

Entry	K19/F87A	R19/F87A	A82M	A184I	T260G	A264G	A328G	7β	1β	TON
1		✓	✓	✓	✓		✓	49%	**48%**	240
2	✓		✓		✓		✓	53%	**38%**	177
3		✓		✓	✓			55%	**32%**	90
4		✓	✓		✓		✓	66%	**26%**	88
5		✓	✓		✓			**54%**		100
6		✓			✓		✓	**78%**		78
7		✓	✓	✓	✓			**88%**	10%	440
8		✓		✓	✓	✓	✓	**96%**		452
9		✓		✓	✓		✓	**96%**		480

Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 1 mM AN (500:1), 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold.

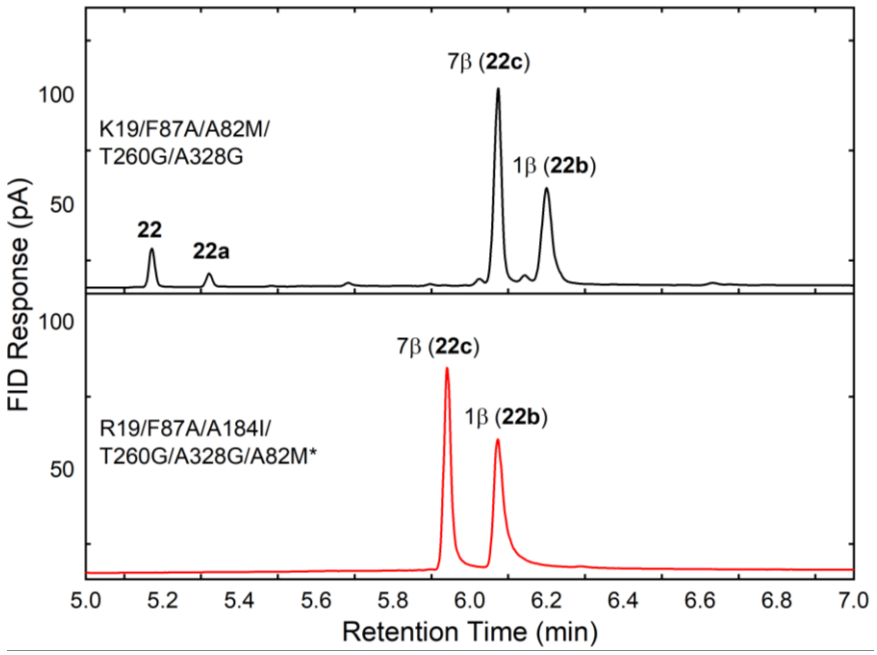


Figure 5.9 GC analysis of selected variants from the second and third-generation library. * indicates retention time shift due to the change of column. Variant K19/F87A/A82M/T260G/A328G (second-generation) gave 53% of **22c** and 38% of **22b**. Variant R19/F87A/A184I/T260G/A328G/A82M (third generation) gave 49% of **22c** and 48% of **22b**.

Only 3 out of the 21 third-generation variants showed >20% conversion of EAN (Table 5.5, Entries 2 and 5–6). Results from the second-generation variants showed the importance of the

A264G/A328G combination when added to R19/F87A/A184I (Entry 3), as well as the significance of the A82M/T260G combination (Entry 4). These mutations were combined in the R19/F87A/A184I/T260G/A328G/A82M and R19/F87A/A184I/T260G/A264G/A327G variants that showed 89% and 95% selectivity for 7β-hydroxy-EAN **23c** (Entries 5 & 6, Fig. 5.10).

Table 5.5 The comparison of variants giving high selectivity of 7β-hydroxy-EAN (**23d**) from the second and third-generation library. The entry number in red indicates variants from the third-generation library.

Entry	K19/F87A	R19/F87A	A82M	A184I	T260G	A264G	A328G	7β	15β	1β	TON
1	✓		✓		✓			**38%**			63
2		✓	✓	✓	✓			**68%**	13%		327
3		✓		✓		✓	✓	**69%**	15%		342
4	✓		✓		✓		✓	**76%**			122
5		✓	✓	✓	✓		✓	**89%**		11%	445
6		✓		✓	✓	✓	✓	**95%**		5%	238

Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 1 mM EAN (500:1), 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold.

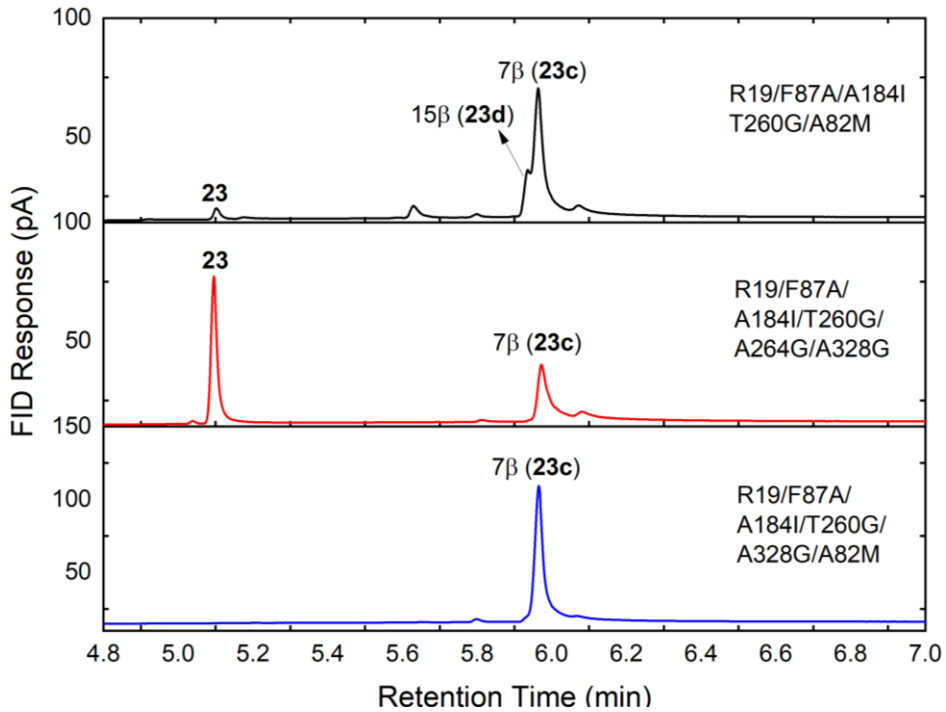


Figure 5.10 GC analysis of selected variants from the third-generation library. Variant R19/F87A/A184I/T260G/A82M gave 68% of **23c** and 13% of **23d**; R19/F87A/A184I/T260G/A264G/A328G: 95% of **23c**; R19/F87A/A184I/T260G/A328G/A82M: 89% of **23c** with full conversion.

5.6 Summary of AN and EAN oxidation

The initial screening library of variants oxidised AN and EAN at C3 and C16β. C16β is the most common position for steroid oxidation by CYP102A1. AN also showed C7β oxidation as one of the minor products. Scanning glycine mutagenesis in the I helix and the β4 strand led to oxidation at new non-activated aliphatic C–H bonds of C1 and C15 for both substrates. AN showed 4α-hydroxylation of the 16β-alcohol precursor to form the 4α,16β-diol. This product is of interest because it represents the only example of C4 oxidation amongst all the steroids studied with over 100 CYP102A1 variants in this work.

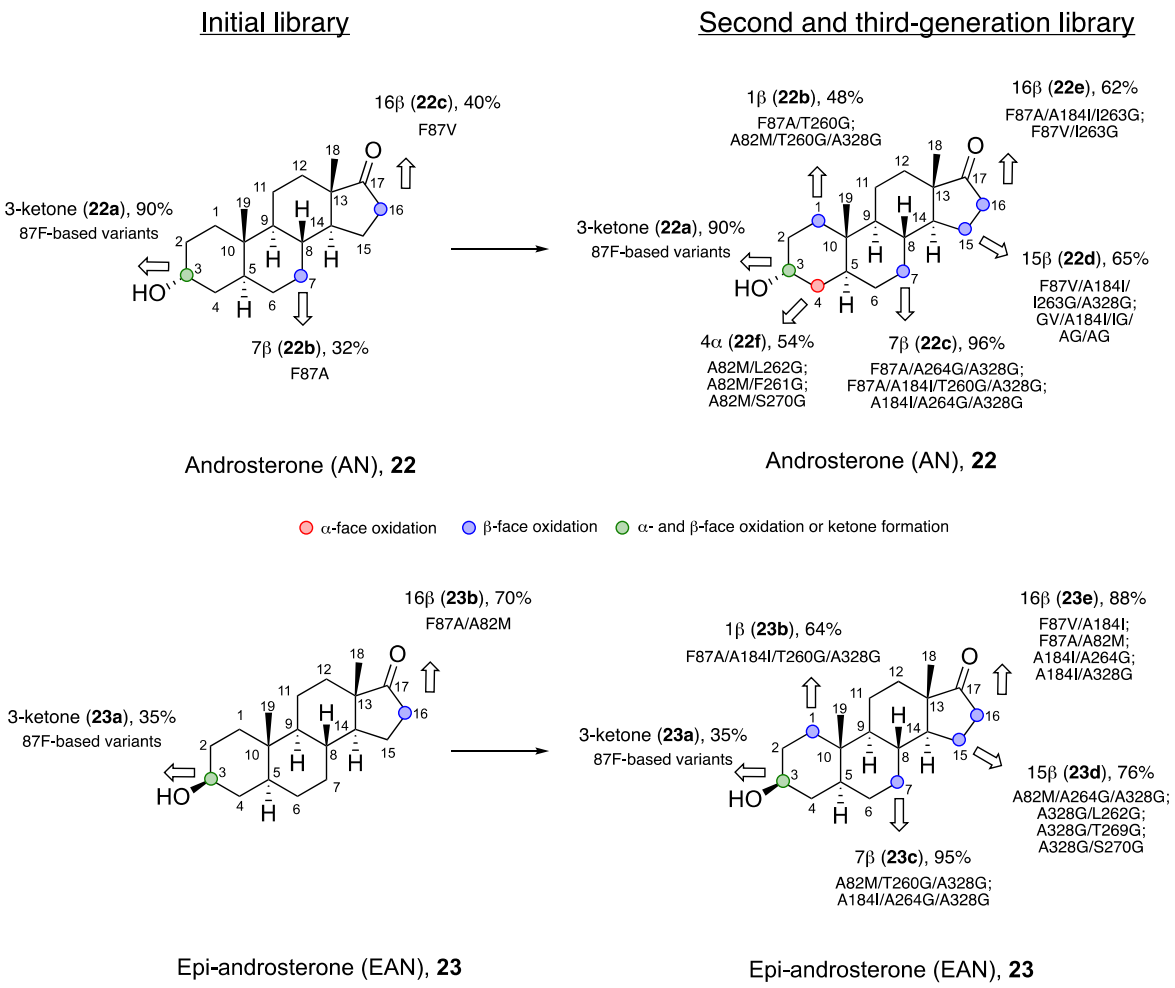


Figure 5.11 Summary of mutations promoting different AN and EAN oxidation products in the initial, second and third-generation library.

AN and EAN are less soluble in aqueous buffer compared to DHEA, leading to lower oxidation activity. As observed with DHEA, the second-generation variants oxidised AN and EAN at C7β, C15β and C16β positions with high selectivity and turnover number. Combinations of I263G, A264G and A328G added to F87A- and F87V-based variants diversified the CYP102A1 substrate pocket topology for steroid binding. The addition of A82M and A184I further increased the product selectivity and led to the formation of new products. The T260G mutation was particularly important for A-ring oxidation of AN and EAN. With the third-generation variants, the selectivity for 1β-hydroxy-AN was further increased with the combination of A82M, T260G and A328G. Likewise, the selectivity for 7β-hydroxy-EAN was improved by the same variant. Of the other I-helix glycine mutations, L262G, T269G and S270G promoted 15β-hydroxy-EAN formation when combined with the A328G mutation.

5.7 Oxidation of progesterone (PROG)

The screening library showed reasonable progesterone (PROG, **24**) oxidation activity, with 35 out of 72 variants giving >20% conversion. Six products were purified and characterised (data in Appendix 15) as 2β-hydroxy-PROG (**24a**),[313] 6β-hydroxy-PROG (**24b**),[313] 15α-hydroxy-PROG (**24c**),[313] 16β-hydroxy-PROG (**24d**),[169] 2β,16β-dihydroxy-PROG (**24e**)[169], and 2β,17α-dihydroxy-PROG (**24f**) (Figure 5.12). The NMR data of **24a**–**24e** (Appendix 15) agreed with literature; **24f** is a new metabolite of PROG and its full characterisation data are collected in Appendix 15. There were other, minor products in the reaction for most of the variants, but these were not formed in sufficient quantities for characterisation.

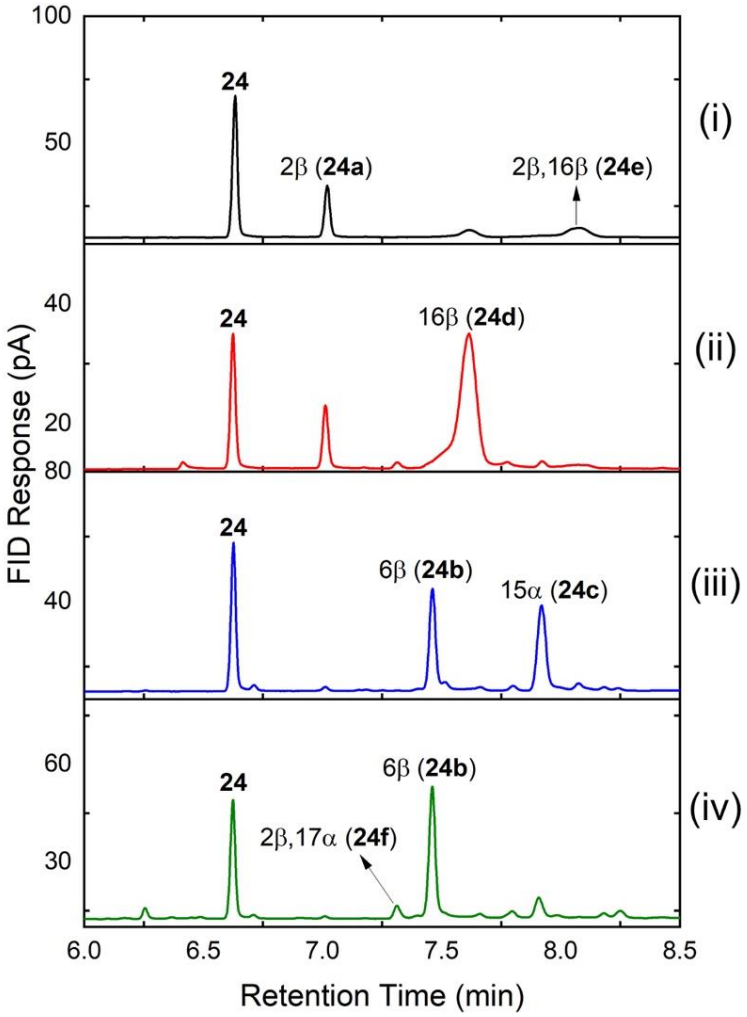


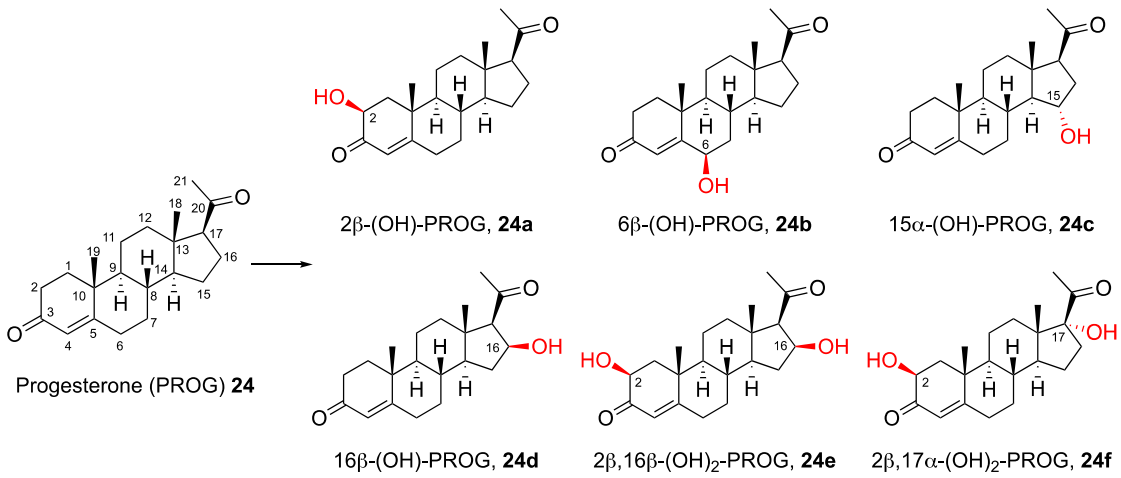
Figure 5.12 GC analysis of selected variants giving PROG oxidation products **24a**–**24f**: (i) 46% of 2β-hydroxy-PROG **24a** and 37% of 2β,16β-dihydroxy-PROG **24e** by GVQ/A264G; (ii) 82% of 16β-hydroxy-PROG **24d** by R19/F87A/A264G; (iii) 36% of 6β-hydroxy-PROG **24b** and 44% of 15α-hydroxy-PROG **24c** by R19/F87A/I263G/A264G; (iv) 61% of 6β-hydroxy-PROG **24b** and 6% of 2β,17α-dihydroxy-PROG **24f** by R19/F87A/I263G/A264G/A328G.

The I-helix mutations I263G and A264G had contrasting effects when added to variants based on the F87A or F87V mutations. The base variant R19/F87A gave 88% of the 16β-hydroxy-PROG **24d** but at low conversion of 18% (Table 5.6, Entry 11). Addition of the A264G mutation to R19/F87A increased its conversion for the 16β-alcohol **24d** from 18% to 93% with 81% selectivity (Entry 13, Figure 5.12-ii). With the F87V-based variant GVQ, however, the A264G mutation led to the 2β-alcohol **24a** (46%) and the 2β,16β-diol **24e** (37%, Entry 1, Figure 5.12-i). Interestingly, the addition of this mutation to F87A/A82M also favoured the formation

of 2β-alcohol **24a** (Entries 3–6), with K19/F87A/A82M/A264G giving 82% of **24a** in 86% conversion (Entry 6). Reetz and co-workers have reported 2β- and 16β-hydroxylation of PROG by CAST and site-saturation mutagenesis and screening of ~9000 CYP102A1 mutants.[169] The starting variant F87A showed 82% of 16β-alcohol and 18% of 2β-alcohol with 11% conversion (Entry 15). Although addition of the A330W mutation (KSA-1) increased the selectivity for the 16β-alcohol to 91%, the conversion remained low (12%, Entry 16). Following the screening of hundreds of mutants, several improved variants showed higher conversion but with a trade off on selectivity of the 16β-alcohol (Entries 17 & 18). The optimal 2β-selective mutant, A82N/F87A, displayed 100% regioselectivity but with only 3% conversion (Entry 19). The improved variant R47Y/T49F/V78T/A82F/F87A slightly improved the conversion to 11% but with a drop of selectivity for the 2β-alcohol to 57% (Entry 20).

The I263G mutation did not affect the selectivity of F87V-base variants but it abolished 16β-oxidation when added to the R19/F87A variant, forming 55% of 6β-hydroxy-PROG (**24b**) and 15% of 15α-hydroxy-PROG (**24c**) (Entry 9). The I263G/A264G combination led to further variations in PROG oxidation selectivity. R19/F87A/I263G/A264G gave 36% of 6β-hydroxy-PROG (**24b**) and 44% of 15α-hydroxy-PROG (**24c**) (Entry 8, Figure 5.12-iii). GVQ/I263G/A264G showed similar selectivity (43%) for the 2β-alcohol **24a** as the GVQ/A264G precursor but 16β oxidation products were eliminated and replaced with 12% of **24b** and 28% of **24c** (Entry 2). Replacing the L188Q mutation with A184I increased the selectivity for the 15α-alcohol **24c** to 43% (Entry 7).

Table 5.6 Selectivity and turnover number CYP102A1 variants in the second-generation library for the oxidation of PROG, **24**.[a]



Entry	Variants	24a	24b	24c	24d	24e	24f	*A*	*B*	*C*
1	GVQ/A264G	**46%**			17%	37%			46%	106
2	GVQ/I263G/A264G	**43%**	12%	28%				17%	14%	30
3	K19/F87A/A82M/I263G/A264G	**51%**						49%	28%	71
4	K19/F87A/A82M/I263G/A264G/A328G	**66%**						34%	36%	119
5	K19/F87A/A82M/A264G/A328G	**69%**						31%	87%	301
6	K19/F87A/A82M/A264G	**82%**						18%	86%	353
7	GV/A184I/I263G/A264G	22%		**43%**				35%	34%	73
8	R19/F87A/I263G/A264G		36%	**44%**				20%	70%	154
9	R19/F87A/I263G		**55%**	15%						
10	R19/F87A/I263G/A264G/A328G		**61%**	11%			6%	22%	71%	217
11	R19/F87A	7%			**88%**	4%		1%	18%	79
12	R19/F87A/Q403P	13%			**80%**		3%	4%	36%	144
13	R19/F87A/A264G				**81%**		5%	14%	93%	377
14	R19/F87A/I263G/P329G/A330W		27%	17%			**19%**	37%	22%	21
15	F87A *	18%			82%				11%	
16	F87A/A330W *	9%			91%				12%	
17	R47Y/T49F/V78I/A92M/F87A *	8%			75%				60%	
18	F87A/V78C/A82G *	13%			81%				43%	
19	V78V/A82N/F87A *	100%							3%	
20	R47Y/T49F/V78T/A82F/F87A *	57%			37%				11%	

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 1 mM PROG (500:1), 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. *A* denotes "Others" that consists of other uncharacterised oxidation products. *B* denotes "Conversion" that refers to the % of substrate converted to products. *C* denotes "TON" that refers to the turnover number for the formation of the product in bold. * literature data from Reetz and co-workers.[169]

Mutations of the β4-strand residues A328–A330 also altered the selectivity. The A328G mutation increased the selectivity for the 6β-alcohol **24b** to 61% in variant

R19/F87A/I263G/A264G/A328G, which also gave 6% of the new product 2β,17α-dihydroxy-PROG (**24f**) at 71% conversion (Entry 10, Figure 5.12-iv). Other β4-strand mutations also influenced the formation of **24f**; the P329G/A330W couplet increased the selectivity to 19% in the variant R19/F87A/I263G/P329G/A330W (Entry 14). Since the 2β-alcohol **24a** was not observed, it was rapidly oxidised if formed; similarly, the 17α-alcohol was not a significant product with either mutant, suggesting that it was also rapidly oxidised once formed.

5.8 Oxidation of pregnenolone (PREG)

Pregnenolone (PREG) is much less soluble than PROG in aqueous buffer. Of the initial screening library, only the R19/F87A/A184I variant gave any significant conversion (42%). With combinations of glycine mutations in the I helix and β4-strand, the 2nd generation variants showed higher activity, with 12 out of 67 variants giving >20% conversion. Notably, most of the active variants contained the A184I mutation. PREG was oxidised to 3 main products which were purified and characterised as 7β-hydroxy-PREG (**25a**),[276] 15β-hydroxy-PREG (**25b**),[314] and 17α-hydroxy-PREG (**25c**).[315,316] The NMR data for **25a**–**25c** (Appendix 16) were in agreement with literature.

The A264G mutation was found to be effective in promoting the formation of 17α-hydroxy-PREG (**25c**); the K19/F87A/A82M/A264G and R19/F87A/A264G variants gave 55% and 46%, respectively, of **25c** (Table 5.7, Entries 6 & 4). Addition of the A184I mutation to R19/F87A/A264G increased the selectivity for **25c** to 53%, with total conversion (Entry 5). Introduction of the A328G mutation further improved the selectivity for **25c** to 64% in R19/F87A/A184I/A264G/A328G, with full conversion (Entry 7, Figure 5.13-ii). Interestingly, and as observed with PROG, addition of the I263G mutation to A264G-containing variants dramatically shifted the product profile. For example, R19/F87A/A184I/A264G and

R19/F87A/A184I/A264G/A328G favoured the 17α-alcohol **25c**, but the addition of I263G to these variants gave 75% and 64%, respectively, of the new product 7β-hydroxy-PREG (**25a**), with no 17α-alcohol observed (Entries 1 & 2, Figure 5.13-i).

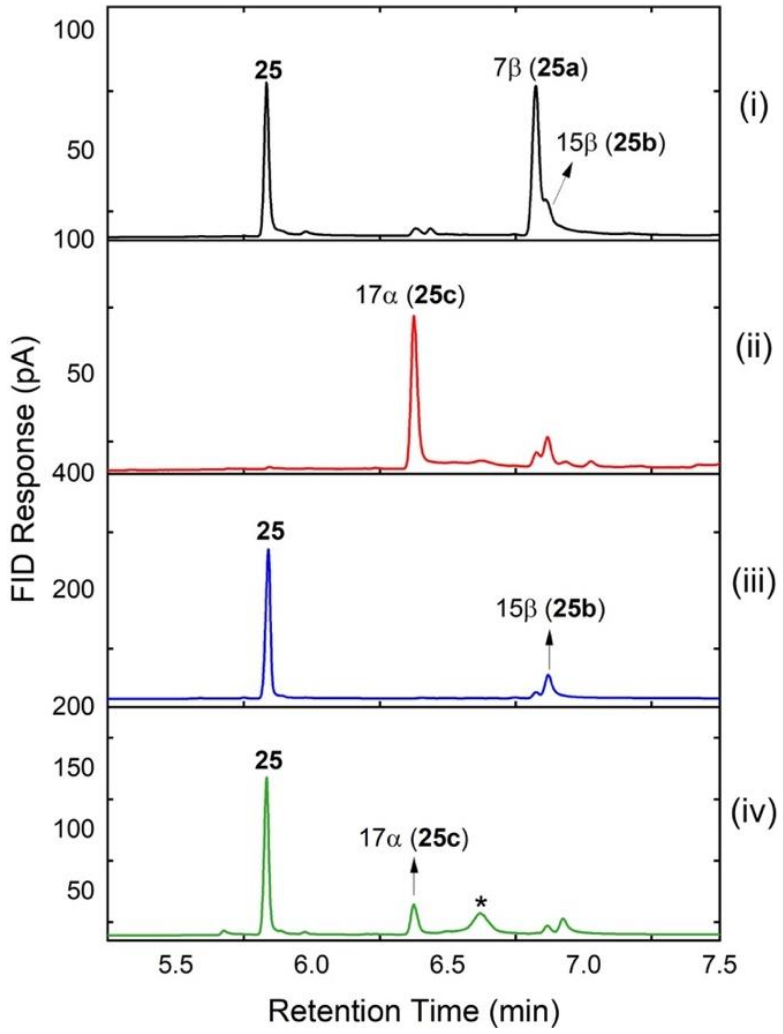
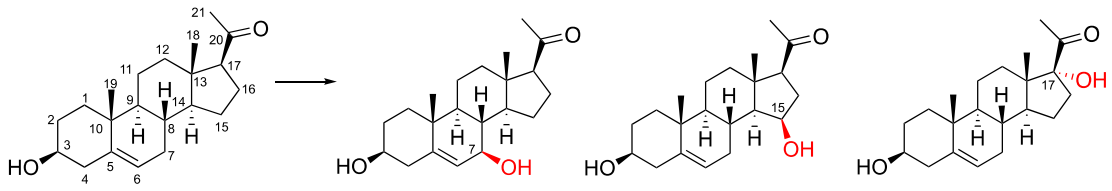


Figure 5.13 GC analysis of selected variants giving PREG oxidation products **25a**–**25c**: (i) 64% of 7β-hydroxy-PREG **25a** and 25% of 15β-hydroxy-PREG **25b** by R19/F87A/A184I/I263G/A264G/A328G; (ii) 64% of 17α-hydroxy-PREG **25c** by R19/F87A/A184/A264G/A328G; (iii) 73% of 15β-hydroxy-PREG **25b** by GV/A184I/I263G/A264G/A328G; (iv) 21% of 17α-hydroxy-PREG **25c** and 44% of an uncharacterised product **u1** (*) by R19/F87A/A184I/S270G.

Other than promoting the 7β-alcohol, these I263G-based variants also gave 15β-hydroxy-PREG (**25b**), e.g. 25% for R19/F87A/A184I/I263G/A264G/A328G (Entry 1). The highest selectivity for **25b** was 73% given by GV/A184I/I263G/A264G/A328G but with lower conversion (Entry 3, Figure 5.13-iii). When added to R19/F87A/A184I, other I-helix glycine

mutations, L262G, T269G, and S270G, favoured the formation of a new product that could not be purified due to insufficient amount of material (Entries 8–10). However, given that the 15β- and 17α-alcohols were amongst the significant products, it may be speculated that this unidentified product could arise from C16 oxidation.

Table 5.7 Selectivity and turnover number CYP102A1 variants in the second-generation library for the oxidation of PREG, **25**.[a]



Pregnenolone (PREG) **25** 7β-(OH)-PREG, **25a** 15β-(OH)-PREG, **25b** 17α-(OH)-PREG, **25c**

Entry	Variants	25a	25b	25c	u1	Others	Conv.	TON[b]
1	R19/F87A/A184I/I263G/A264G/A328G	**64%**	**25%**			11%	68%	218
2	R19/F87A/A184I/I263G/A264G	**75%**				25%	57%	214
3	GV/A184I/I263G/A264G/A328G	11%	**73%**			16%	32%	117
4	R19/F87A/A264G			**46%**	32%	22%	92%	212
5	R19/F87A/A184I/A264G		13%	**53%**	29%	5%	100%	265
6	K19/F87A/A82M/A264G		11%	**55%**	10%	24%	48%	132
7	R19/F87A/A184I/A264G/A328G		16%	**64%**	7%	13%	100%	320
8	R19/F87A/A184I/T269G		12%	33%	**23%**	32%	37%	43
9	R19/F87A/A184I/L262G		16%	31%	**34%**	19%	41%	70
10	R19/F87A/A184I/S270G		7%	21%	**44%**	28%	58%	128

[a] The selectivity for the two most selective variants for each product are in red. Screening scale reactions at 0.5 mL scale were in 200 mM phosphate, pH 8.0, containing 2 μM CYP102A1 enzyme, 1 mM PREG (500:1), 40 μM $NADP^+$, 100 mM glucose, and 20 U/mL GDH. Plates were shaken at 120 rpm at 20 °C for 16 h. [b] TON refers to the turnover number for the formation of the product in bold.

5.9 Computational studies of PROG and PREG oxidation

The R19/F87A/A184I/A264G/A328G variant was one of the most selective for the 17α oxidation of PREG (64%) whereas PROG gave multiple peaks including 16β-, 15α-alcohols and other uncharacterised products. Seven clustered structures were generated from the three replica MD simulations of this variant. Both PROG and PREG were docked into these structures and generated 59 and 63 poses, respectively. The docking of PROG gave 20

productive poses that were equally distributed to 2β, 6β, 7β, 11β, 15β, 18, 19 and 21 oxidations; no 16β- and 15α-poses were observed. PREG gave 24 productive poses that indicated 1β, 2α, 7α/β, 15β, 17α, 19 and 21 oxidations; 15β and 17α hydroxylation were the dominant positions with 4 and 3 poses. Two of the three 17α poses were the lowest energy poses in the clusters and are shown in Figure 5.14. In this pose, 17α-H was 4.0 Å and 3.4 Å from the ferryl oxygen, and two hydrogen-bonds were observed between the C3-alcohol of PREG and the side chains of S72 and S332 which stabilised this 17α-binding mode. The planar structure of the 3-oxo group in PROG, on the other hand, may diminish the possibility of hydrogen-bond formation with these residues, therefore, no formation of 17α-hydroxylation of PROG was observed. Out of the 20 productive poses generated by PROG docking, only two of them showed hydrogen-bonds between the C3 and C20 carbonyl groups with the side chains of S72 and S332, which were docked in C18- and C19-oxidation poses, respectively.

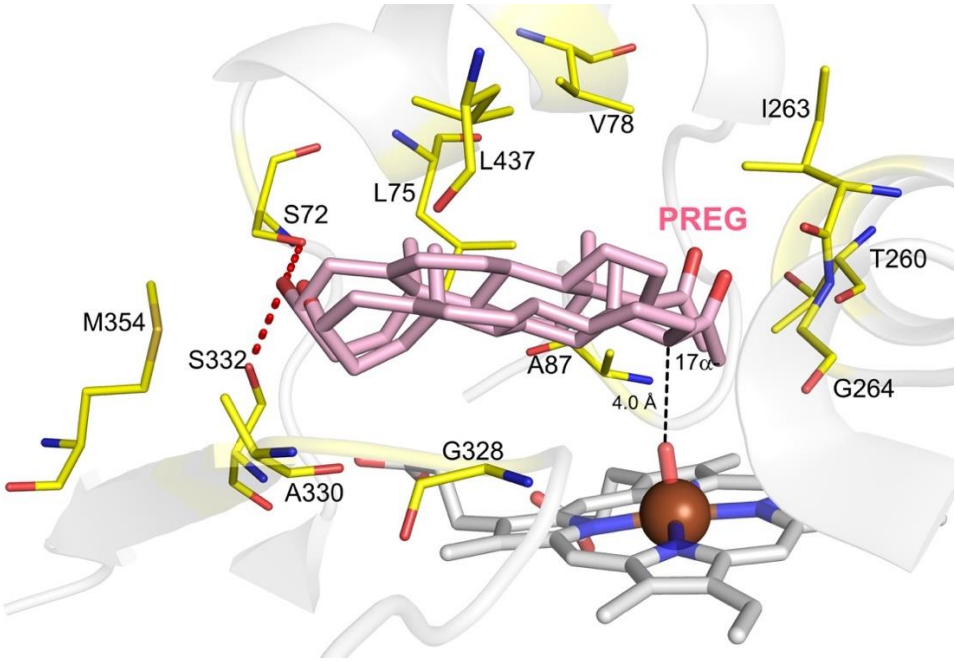


Figure 5.14 R19/F87A/A184I/A264G/A328G variant with docked PREG (pink) bound in 17α pose. Hydrogen-bonding between the C3-OH and the side chains of S72 and S332 are shown in red dashes. Other key contacting residues within van der Waals contact distance of PREG are L75, V78, T260, I263, G264, G328, A330, M354, L437 and T438.

5.10 Comparison of AD, DHEA, PROG and PREG oxidation

The screening library of CYP102A1 variants oxidised PROG at the 2β, 6β, 15α, and 16β positions with medium to high selectivity. The I helix glycine mutation A264G played vital roles in promoting the 2β- and 16β-alcohols of PROG when combined with F87V and F87A mutations, respectively. The addition of I263G abolished the formation of 2β- and 16β-alcohols but gave 6β- and 15α-hydroxylation. PROG also showed 17α-hydroxylation of the 2β-alcohol precursor to form the 2β,17α-diol as a minor product. Reetz and co-workers had reported 2β- and 16β-oxidation of PROG at >90% selectivity but at medium to low conversion, especially for the 2β-alcohol (~10%).

By having the α,β-unsaturated carbonyl group in the A ring, both AD and PROG were oxidised at the activated C2 position by variants with the A82M/A264G combination. The I263G mutation, on the other hand, shifted the selectivity away from the 2β-alcohols towards the C6, the allylic position, for both AD and PROG. Apart from 6β-hydroxylation, the I263G mutation also gave 16α-oxidation of AD but gave 15α-hydroxylation for PROG, possibly due to the steric hindrance of the acetyl group at C17 in PROG. The combination T260G/A328G led to oxidation at the non-activated C7 position for AD but this was not observed for PROG.

PREG has a 3-hydroxy-Δ^5-structure in the A, B rings as for DHEA, and an acetyl group at C17 as for PROG. However, PREG is much less soluble in aqueous buffer compared to DHEA and PROG, leading to much lower oxidation activity. Within the library of CYP102A1 variants, PREG was mainly oxidised by A184I-based variants at the 7β, 15β and 17α positions with medium to high selectivity. Compared to PROG oxidation, no A-ring oxidation was observed with PREG, but 17α-hydroxylation of PREG was one of the dominant activities that was

promoted, in particular, by A264G-based variants. The addition of I263G to F87A-based variants altered the selectivity towards B-ring oxidation at C7β. On the other hand, addition of this mutation to F87V led to 15β-hydroxylation of PREG.

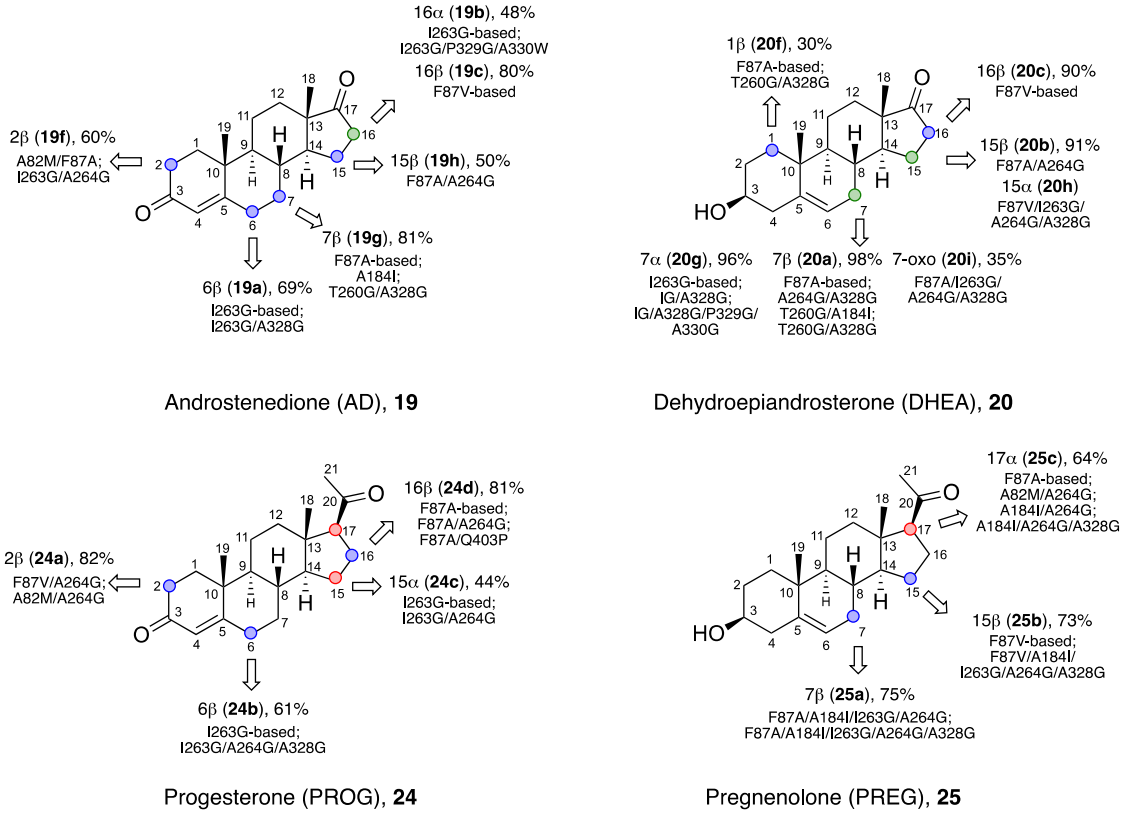


Figure 5.15 Summary of mutations for the promotion of different AD, DHEA, PROG and PREG oxidation products in the screening library.

Chapter 6

6 Conclusion

The main goal of this thesis was to engineer $P450_{BM3}$ and apply the variants to oxidise unnatural substrates to a diverse range of products with high regio- and stereo-selectivity, especially to specific desired precursors for the synthesis of valuable compounds.

Chapter 3 demonstrated the use of $P450_{BM3}$ variants in the chemoenzymatic synthesis of the aroma compounds β-damascenone, tabanone isomers and rose oxide. A number of the engineered variants displayed almost full regioselectivity (up to 98%) at high conversion (up to 91%) for the oxidation of β-damascone to 4-hydroxy-β-damascone, the desired precursor to β-damascenone, probably due to the allylic nature of C4. Larger preparative scale oxidation of β-damascone (up to 3.84 g/L) with R19/F87A/A328I were performed and up to 95% conversion and 77% selectivity for the 4-alcohol was achieved. Acid-catalysed dehydration of the 4-alcohol to form β-damascenone was explored with four different acids (sulphuric acid, acetic acid, trifluoroacetic acid and oxalic acid) under various conditions. The screening scale reaction (~10 mg) of the 4-alcohol with oxalic acid gave total conversion to β-damascenone under mild conditions, which was then scaled to 265 mg and the reaction was completed in 16 h with full conversion and 66% isolated yield. Compared with chemical synthesis routes which require high temperatures and use of toxic reagents to access C4 oxidation of β-damascone, $P450_{BM3}$ provides a greener method for late-stage C–H activation under milder conditions with higher yield. Further optimisation may increase the reaction titre, such as developing an alternative $P450_{BM3}$ variant that gives higher TON and the use of selective resins to remove the 4-alcohol before the addition of more substrate to avoid enzyme inhibition. The reaction conditions of acid treatment and product isolation can be optimised to improve the yield and sustainability of the overall process. This chemoenzymatic synthesis for β-damascenone was a successful

illustration of semisynthesis using a natural feedstock. Although the 4-alcohol was the predominant product in the oxidation of β-damascone, some $P450_{BM3}$ variants exhibited hydroxylation at other positions such as C10 (another allylic position), C2 and C3 (non-activated positions). The C2 and C10 alcohols are new metabolites of β-damascone.

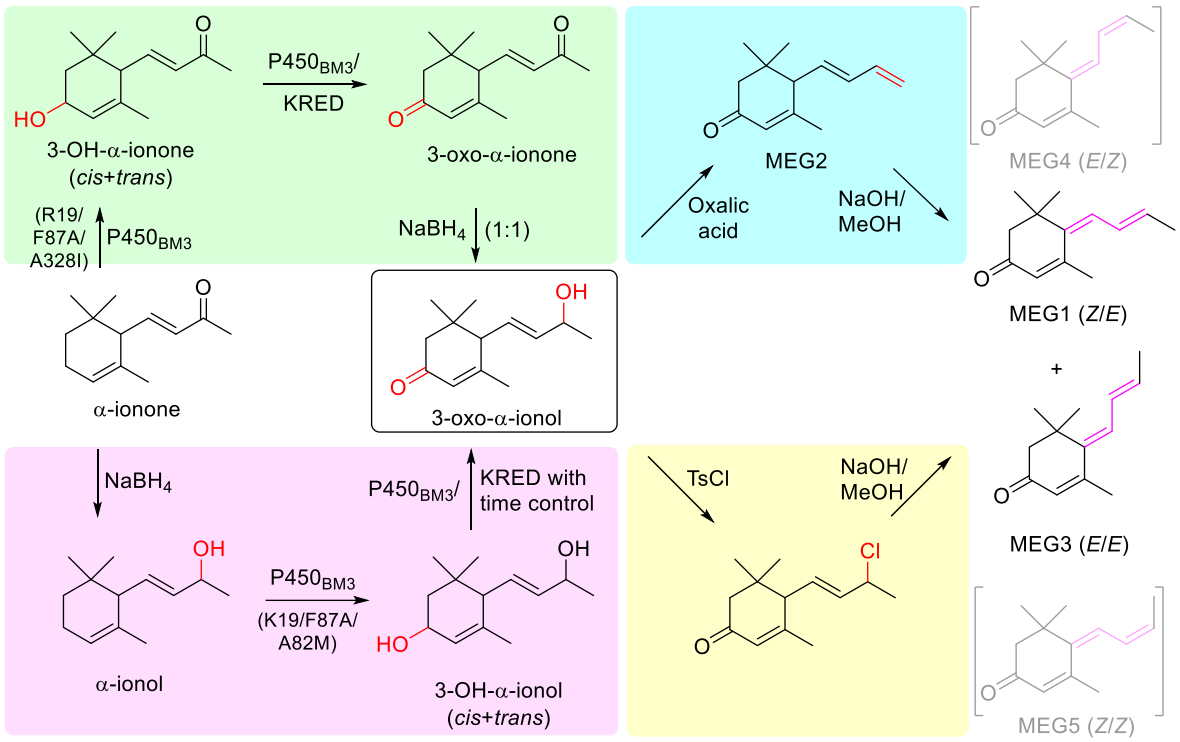


Figure 6.1 Overview of tabanone synthesis (MEG1 & 3) from α-ionone via the elimination of the precursor 3-oxo-α-ionol.

Due to difficulties in controlling the selectivity for the five isomers of tabanone, two synthetic routes for a mixture of tabanone isomers were developed in Chapter 3, via dehydration of 3-oxo-α-ionol. This precursor could be obtained from α-ionone via oxidation to 3-oxo-α-ionone followed by selective reduction of the C9 ketone with careful control of reaction stoichiometry (green panel in Figure 6.1). The R19/F87A variant (66% C3-alcohol and 27% of 3-oxo) was selected for screening for oxidation by KREDs, finally giving 75% selectivity of 3-oxo-α-ionone which was reduced with one equivalent of $NaBH_4$. 3-Oxo-α-ionol could also be obtained from oxidation of α-ionol at C3 catalysed either by $P450_{BM3}$ variants or in combination

with a KRED with reaction time control (pink panel in Figure 6.1). Both pathways have their own advantages and disadvantages compared to the other.

Following the established routes for 3-oxo-α-ionol, two elimination pathways catalysed either by acid or base were explored. Acid treatment with oxalic acid gave full conversion to the terminal alkene product MEG2, which could be converted to a mixture MEG1 and MEG3 under alkaline conditions (blue panel in Figure 6.1). The other route was to convert the alcohol in 3-oxo-α-ionol into a better leaving group by tosylation followed by base treatment to form the mixture of MEG1 and MEG3 (yellow in Figure 6.1). The acid treatment gave clean conversion to MEG2 but only 60% of which was converted to tabanone isomers, whereas base treatment gave full conversion to the tabanone isomers.

There is only one patent (Symrise) for tabanone synthesis; therefore, it would be hard to compare the yield and sustainability between these current methods. However, the biocatalytic synthesis of tabanone using $P450_{BM3}$ minimises the use of organic solvents and hazardous chemicals. This chemoenzymatic synthesis of tabanone isomers has only been explored at small scale (~10 mg) and will require optimisation to pilot larger scale production.

The third example of chemoenzymatic synthesis in Chapter 3 is the synthesis *cis*-(–)-rose oxide from citronellol. The oxidation of citronellol by $P450_{BM3}$ variants gave multiple products including 5-hydroxy-citronellol, the precursor to rose oxide. Although all are allylic positions, the selectivity for C5 hydroxylation was lower than that for C8 and C9. RP/A82M/I263A was one of the variants showing the highest selectivity obtained (21%) for the C5-alcohol with high conversion (84%), and it was selected for the subsequent acid-catalysed cyclisation reaction to form rose oxide. Acid treatment of the pure 5-hydroxy-citronellol gave a 75:25 mixture of *cis*-(–)-rose oxide and *trans*-(–)-rose oxide while the same condition on the reaction crude (mixture

of epoxide and 5-alcohol) gave 90:10 ratio of *cis*- and *trans*- mixture. Further investigations are needed to establish the reasons for the different *cis*/*trans* ratios. Compared with the Dragoco (now Symrise) patent, which showed 54% overeall yield with 60% selectivity for the precursor allyl alcohol, this biocatalytic synthesis of rose oxide using $P450_{BM3}$ showed proof-of-concept for the process. The current yield of this process (17%) could be further improved by protein engineering to improve the selectivity for the 5-alcohol with the aid of MD simulation and docking study.

Chapter 4 and 5 described the synthesis of metabolites of steroidal compounds. The screening of AD, DHEA and TST with the initial library gave few products, mainly from oxidation at activated positions (allylic position or C16) with low to medium selectivity. In order to access different binding orientations, the structure of WT $P450_{BM3}$ was overlaid with the AD-bound CYP19A1 aromatase structure to identify key regions that showed significant steric clashes which blocked this AD binding mode. It appeared that the side chain of F87, the main chain and side chain of I263 and A264 in the I-helix and the side chain of A328 in the β4 strand significantly interfered with the bound-AD. Scanning glycine mutations were introduced along the I helix (I259, T260, T261, L262, I263, A264, E267, T269 and S270) and along the β4 strand (A328, P329 and A330) to create local distortions, allowing the maximum change of substrate pocket shape that would accommodate more binding poses. This nature-inspired mutagenesis of $P450_{BM3}$ was aimed at steroid A ring oxidation. The MD simulated structure of variant K19/F87A/A82M/I263G/A264G, which gave 59% of 2β-hydroxy-AD, was compared with the CYP19A1 aromatase. It showed that with the glycine mutations, the I helix in this new mutant adopted a “stepped back” structure, resembling the I helix conformation in CYP19A1. The screening of DHEA and TST against the second-generation library also showed A ring oxidation at C1β, the new metabolites that have not been reported by any P450 enzyme. Apart

from the promotion of A ring oxidation, oxidation at other non-activated positions was also observed with high selectivity and turnover number, such as the 7β-hydroxylation of AD (81%, TON = 761), 15β-hydroxylation DHEA (91%, TON = 555), and 15α-hydroxylation of TST (75%, TON = 600).

Of the glycine mutations, the most important ones in the I helix are T260G, I263G and A264G. T260G was responsible for the A ring oxidation of DHEA and TST, while I263G was the key mutation for the oxidation products on the α-face such as 16α-hydroxy-AD, 7α-hydroxy-DHEA and 15α-hydroxy-TST. This rationally designed scanning glycine mutagenesis has provided access to broader range of mono- and di-hydroxylated steroids via direct C–H bond activation. Further engineering can be performed to improve the 1β-hydroxylation of DHEA and TST based on the T260G/A328G combination, e.g. by site-saturation mutagenesis at residues identified from the substrate docking studies.

Formation of diverse range of products for AD, DHEA and TST, especially A-ring oxidation, led to screening of the second-generation variants for the oxidation of AN, EAN, PROG and PREG, as reported in Chapter 5. Interestingly, even with almost all inert C–H bonds in AN and EAN, the C7 binding pose was still predominant with 96% and 95% selectivity, respectively. The MD simulation and docking studies suggested that H-bonding with the C3-alcohol of AN and EAN helped to stabilise these binding poses. As observed with DHEA and TST, the T260G/A328G combination promoted 1β-hydroxylation of AN and EAN with up to 48% and 64% selectivity, respectively. Compared with the oxidation of PROG by CYP106A2 variants which gave 6β-, 9α-, 11α-, and 15β-hydroxylation, the screening with $P450_{BM3}$ variants gave hydroxylation mainly at 2β-, 6β-, and 16β-sites. PROG showed 17α-hydroxylation of the 2β-alcohol precursor to form the 2β,17α-diol. PREG has a similar structure to PROG and DHEA

but with a much lower activity due to its low solubility in aqueous solution; only A184I-containing variants showed any activity. In contrast, 17α-hydroxylation of PREG was one of the predominant products in the screening with $P450_{BM3}$ variants.

This approach of designing mutations by taking inspiration from nature can be applied to other substrates and enzymes for the synthesis of natural products and their derivatives. Human mitochondrial CYP11B1 catalyses a one-step regio- and stereo-selective 11β-hydroxylation of different steroids in human steroidogenesis, e.g. 11-deoxycortisol to cortisol,[317] PROG to 11β-hydroxy-PROG,[318] *etc*. Overlay of the CYP11B1 hydroxylase with WT $P450_{BM3}$ could be explored further, in order to identify key regions showing significant differences. Further mutagenesis could be designed based on these observations to generate 11β-hydroxylation of progesterone and derivatives.

7. Reference

1. Crabtree, R. H.; Lei, A., Introduction: CH Activation. *Chem. Rev.* **2017,** *117* (13), 8481-8482.

2. Shang, R.; Ilies, L.; Nakamura, E., Iron-Catalyzed C-H Bond Activation. *Chem. Rev.* **2017,** *117* (13), 9086-9139.

3. Gandeepan, P.; Muller, T.; Zell, D.; Cera, G.; Warratz, S.; Ackermann, L., 3d Transition Metals for C-H Activation. *Chem. Rev.* **2019,** *119* (4), 2192-2452.

4. Yi, H.; Zhang, G.; Wang, H.; Huang, Z.; Wang, J.; Singh, A. K.; Lei, A., Recent Advances in Radical C-H Activation/Radical Cross-Coupling. *Chem. Rev.* **2017,** *117* (13), 9016-9085.

5. He, J.; Wasa, M.; Chan, K. S. L.; Shao, Q.; Yu, J. Q., Palladium-Catalyzed Transformations of Alkyl C-H Bonds. *Chem. Rev.* **2017,** *117* (13), 8754-8786.

6. Newton, C. G.; Wang, S. G.; Oliveira, C. C.; Cramer, N., Catalytic Enantioselective Transformations Involving C-H Bond Cleavage by Transition-Metal Complexes. *Chem. Rev.* **2017,** *117* (13), 8908-8976.

7. Nakamura, E.; Yoshikai, N.; Matsumoto, A.; Norinder, J., Iron-Catalyzed Direct Arylation of Aryl Pyridines and Imines Using Oxygen as an Oxidant. *Synlett* **2010,** *2010* (02), 313-316.

8. Norinder, J.; Matsumoto, A.; Yoshikai, N.; Nakamura, E., Iron-Catalyzed Direct Arylation through Directed C-H Bond Activation. *J. Am. Chem. Soc.* **2008,** *130* (18), 5858-5859.

9. Chen, M. S.; White, M. C., Combined Effects on Selectivity in Fe-Catalyzed Methylene Oxidation. *Science* **2010,** *327* (5965), 566-571.

10. Chen, M. S.; White, M. C., A Predictably Selective Aliphatic C–H Oxidation Reaction for Complex Molecule Synthesis. *Science* **2007,** *318* (5851), 783-787.

11. Newhouse, T.; Baran, P. S., If C-H bonds could talk: selective C-H bond oxidation. *Angew. Chem. Int. Ed. Engl.* **2011,** *50* (15), 3362-74.

12. Basu, D.; Kumar, S.; V, S. S.; Bandichhor, R., Transition metal catalyzed C-H activation for the synthesis of medicinally relevant molecules: A Review. *J. Chem. Sci.* **2018,** *130* (6).

13. Ma, W.; Kaplaneris, N.; Fang, X.; Gu, L.; Mei, R.; Ackermann, L., Chelation-assisted transition metal-catalysed C–H chalcogenylations. *Org. Chem. Front.* **2020,** *7* (8), 1022-1060.

14. Chen, X.; Engle, K. M.; Wang, D. H.; Yu, J. Q., Palladium(II)-catalyzed C-H activation/C-C cross-coupling reactions: versatility and practicality. *Angew. Chem. Int. Ed. Engl.* **2009,** *48* (28), 5094-115.

15. Lyons, T. W.; Sanford, M. S., Palladium-Catalyzed Ligand-Directed C-H Functionalization Reactions. *Chem. Rev.* **2010,** *110* (2), 1147-1169.

16. Kim, D. S.; Park, W. J.; Jun, C. H., Metal-Organic Cooperative Catalysis in C-H and C-C Bond Activation. *Chem. Rev.* **2017,** *117* (13), 8977-9015.

17. Tsukano, C.; Okuno, M.; Takemoto, Y., Palladium-catalyzed amidation by chemoselective C(sp3)-H activation: concise route to oxindoles using a carbamoyl chloride precursor. *Angew. Chem. Int. Ed. Engl.* **2012,** *51* (11), 2763-6.

18. Pedroni, J.; Cramer, N., 2-(Trifluoromethyl)indoles via Pd(0)-Catalyzed C(sp(3))-H Functionalization of Trifluoroacetimidoyl Chlorides. *Org. Lett.* **2016,** *18* (8), 1932-5.

19. Zhou, L.; Tang, S.; Qi, X.; Lin, C.; Liu, K.; Liu, C.; Lan, Y.; Lei, A., Transition-metal-assisted radical/radical cross-coupling: a new strategy to the oxidative C(sp3)-H/N-H cross-coupling. *Org. Lett.* **2014,** *16* (12), 3404-7.

20. Zheng, C.; Lu, F.; Lu, H.; Xin, J.; Deng, Y.; Yang, D.; Wang, S.; Huang, Z.; Gao, M.; Lei, A., Copper-catalyzed selective radical-radical cross-coupling for C-S bond formation: an access to alpha-alkylthionitriles. *Chem. Commun. (Camb)* **2018,** *54* (44), 5574-5577.

21. Leifert, D.; Studer, A., The Persistent Radical Effect in Organic Synthesis. *Angew. Chem. Int. Ed. Engl.* **2020,** *59* (1), 74-108.

22. McNally, A.; Prier, C. K.; MacMillan, D. W. C., Discovery of an α-amino C-H arylation reaction using the strategy of accelerated serendipity. *Science* **2011,** *334* (6059), 1114-1117.

23. Hoshikawa, T.; Inoue, M., Photoinduced direct 4-pyridination of C(sp3)–H Bonds. *Chem. Sci.* **2013,** *4* (8).

24. Herrerias, C. I.; X., Y.; Li, Z.; Li, C., Reactions of C−H Bonds in Water. *Chem. Rev.* **2007,** *107* (6), 2546-2562.

25. Labinger, J. A.; Bercaw, J. E., Understanding and exploiting C–H bond activation. *Nature* **2002,** *417*, 507-514.

26. Verho, O.; Backvall, J. E., Chemoenzymatic dynamic kinetic resolution: a powerful tool for the preparation of enantiomerically pure alcohols and amines. *J. Am. Chem. Soc.* **2015,** *137* (12), 3996-4009.

27. Rachwalski, M.; Vermue, N.; Rutjes, F. P., Recent advances in enzymatic and chemical deracemisation of racemic compounds. *Chem. Soc. Rev.* **2013,** *42* (24), 9268-82.

28. Ghanem, A.; Aboul-Enein, H. Y., Application of lipases in kinetic resolution of racemates. *Chirality* **2005,** *17* (1), 1-15.

29. Koh, J. H.; Jung, H. M.; Kim, M.; Park, J., Enzymatic resolution of secondary alcohols coupled with ruthenium-catalyzed racemization without hydrogen mediator. *Tetrahedron Lett.* **1999,** *40*, 6281-6284.

30. Kamal, A.; Azhar, M. A.; Krishnaji, T.; Malik, M. S.; Azeeza, S., Approaches based on enzyme mediated kinetic to dynamic kinetic resolutions: A versatile route for chiral intermediates. *Coord. Chem. Rev.* **2008,** *252* (5-7), 569-592.

31. Breuer, M.; Ditrich, K.; Habicher, T.; Hauer, B.; Kesseler, M.; Sturmer, R.; Zelinski, T., Industrial methods for the production of optically active intermediates. *Angew. Chem. Int. Ed. Engl.* **2004,** *43* (7), 788-824.

32. Reetz, M. T.; Schimossek, K., Lipase-Catalyzed Dynamic Kinetic Resolution of Chiral Amines- Use of Palladium as the Racemization Catalyst. *Chimia* **1996,** *50* (12), 668-669.

33. Paetzold, J.; Backvall, J. E., Chemoenzymatic Dynamic Kinetic Resolution of Primary Amines. *J. Am. Chem. Soc.* **2005,** *127* (50), 17620-17621.

34. Thalen, L. K.; Zhao, D.; Sortais, J. B.; Paetzold, J.; Hoben, C.; Backvall, J. E., A chemoenzymatic approach to enantiomerically pure amines using dynamic kinetic resolution: application to the synthesis of norsertraline. *Chem. Eur. J.* **2009,** *15* (14), 3403-10.

35. Matsumoto, C.; Shinkai, T.; Hori, H.; Ohmori, O.; Nakamura, J., Polymorphisms of dopamine degradation enzyme (COMT and MAO) genes and tardive dyskinesia in patients with schizophrenia. *Psychiatry Res.* **2004,** *127* (1-2), 1-7.

36. Akil, M.; Kolachana, B. S.; Rothmond, D. A.; Hyde, T. M.; Weinberger, D. R.; Kleinman, J. E., Catechol-O-Methyltransferase Genotype and Dopamine Regulation in the Human Brain. *J. Neurosci.* **2003,** *23* (6), 2008-2013.

37. Krone, W.; Sohneider, G.; Schulz, D., Detection of phosphohexose isomerase deficiency in human fibroblast cultures. *Humangenetik* **1970,** *10*, 224-230.

38. Kugler, W.; Lakomek, M., Glucose-6-phosphate isomerase deficiency. *Baillieres Best Pract. Res. Clin. Haematol.* **2000,** *13* (1), 89-101.

39. Horinouchi, M.; Hayashi, T.; Kudo, T., Steroid degradation in Comamonas testosteroni. *J. Steroid Biochem. Mol. Biol.* **2012,** *129* (1-2), 4-14.

40. Urlacher, V. B.; Girhard, M., Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol.* **2012,** *30* (1), 26-36.

41. Whitehouse, C. J. C.; Bell, S. G.; Wong, L.-L., P450(BM3) (CYP102A1): connecting the dots. *Chem. Soc. Rev.* **2012,** *41* (3), 1218-1260.

42. Bernhardt, R., Cytochromes P450 as versatile biocatalysts. *J. Biotechnol.* **2006,** *124* (1), 128-145.

43. Werck-Reichhart, D.; Feyereisen, R., Cytochromes P450- a success story. *Genome Biol.* **2000,** *1* (6), 3003.1-3003.9.

44. Guengerich, F. P.; Martin, M. V.; Sohl, C. D.; Cheng, Q., Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase. *Nat. Protoc.* **2009,** *4* (9), 1245-51.

45. Nelson, D. R.; Kamataki, T.; Waxman, D. J.; Guengerich, F. P.; Estabrook, R. W.; Feyereisen, R.; Gonzalez, F. J.; Coon, M. J.; Gunsalus, I. C.; Gotoh, O.; Okuda, K.; Nebert, D. W., The P450 Superfamily- Update on New Sequences, Gene Mapping, Accession Numbers, Early Trivial Names of Enzymes, and Nomenclature. *DNA Cell Biol.* **1993,** *12* (1), 1-51.

46. Hausjell, J.; Halbwirth, H.; Spadiut, O., Recombinant production of eukaryotic cytochrome P450s in microbial cell factories. *Biosci. Rep.* **2018,** *38* (2).

47. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **2014,** *5* (172), 1-17.

48. Leonard, E.; Koffas, M. A., Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by Escherichia coli. *Appl. Environ. Microbiol.* **2007,** *73* (22), 7246-51.

49. Quehl, P.; Hollender, J.; Schuurmann, J.; Brossette, T.; Maas, R.; Jose, J., Co-expression of active human cytochrome P450 1A2 and cytochrome P450 reductase on the cell surface of Escherichia coli. *Microb. Cell. Fact.* **2016,** *15* (26), 1-15.

50. Vaz, A. D. N.; McGinnity, D. F.; Coon, M. J., Epoxidation of olefins by cytochrome P450- Evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant. *Proc. Natl. Acad. Sci. USA* **1998,** *95*, 3555-3560.

51. Guengerich, F. P., Mechanisms of Cytochrome P450-Catalyzed Oxidations. *ACS Catal.* **2018,** *8* (12), 10964-10976.

52. Jin, S.; Makris, T. M.; Bryson, T. A.; Sligar, S. G.; Dawson, J., Epoxidation of Olefins by Hydroperoxo-Ferric Cytochrome P450. *J. Am. Chem. Soc.* **2003,** *125* (12), 3406-3407.

53. Newcomb, M.; Shen, R.; Choi, S.; Toy, P. H.; Hollenberg, P. F.; Vaz, A. D. N.; Coon, M. J., Cytochrome P450-Catalyzed Hydroxylation of Mechanistic Probes that Distinguish between Radicals and Cations. Evidence for Cationic but Not for Radical Intermediates. *J. Am. Chem. Soc.* **2000,** *122* (12), 2677-2686.

54. Bathelt, C. M.; Ridder, L.; Mulholland, A. J.; Harvey, J. N., Aromatic Hydroxylation by Cytochrome P450- Model Calculations of Mechanism and Substituent Effects. *J. Am. Chem. Soc.* **2003,** *125* (49), 15004-15005.

55. Shoji, O.; Wiese, C.; Fujishiro, T.; Shirataki, C.; Wunsch, B.; Watanabe, Y., Aromatic C-H bond hydroxylation by P450 peroxygenases: a facile colorimetric assay for monooxygenation activities of enzymes based on Russig's blue formation. *J. Biol. Inorg. Chem.* **2010,** *15* (7), 1109-15.

56. Sarabia, S. F.; Zhu, B. T.; Kurosawa, T.; Tohma, M.; Liehr, J. G., Mechanism of Cytochrome P450-Catalyzed Aromatic Hydroxylation of Estrogens. *Chem. Res. Toxicol.* **1997,** *10* (7), 767-771.

57. Rydberg, P.; Ryde, U.; Olsen, L., Sulfoxide, Sulfur, and Nitrogen Oxidation and Dealkylation by Cytochrome P450. *J. Chem. Theory. Comput.* **2008,** *4* (8), 1369-1377.

58. Giovani, S.; Alwaseem, H.; Fasan, R., Aldehyde and Ketone Synthesis by P450-Catalyzed Oxidative Deamination of Alkyl Azides. *ChemCatChem* **2016,** *8* (16), 2609-2613.

59. Theobald, D. S.; Maurer, H. H., Identification of monoamine oxidase and cytochrome P450 isoenzymes involved in the deamination of phenethylamine-derived designer drugs (2C-series). *Biochem. Pharmacol.* **2007,** *73* (2), 287-97.

60. Guengerich, F. P., Mechanisms of Cytochrome P450 Substrate Oxidation- MiniReview. *J. Biochem. Mol. Toxicol.* **2007,** *21* (4), 163-168.

61. Volz, T. J.; Rock, D. A.; Jones, J. P., Evidence for Two Different Active Oxygen Species in Cytochrome P450 BM3 Mediated Sulfoxidation and N-Dealkylation Reactions. *J. Am. Chem. Soc.* **2002,** *124* (33), 9724-9725.

62. Zerbe, K.; Woithe, K.; Li, D. B.; Vitali, F.; Bigler, L.; Robinson, J. A., An oxidative phenol coupling reaction catalyzed by oxyB, a cytochrome P450 from the vancomycin-producing microorganism. *Angew. Chem. Int. Ed. Engl.* **2004,** *43* (48), 6709-13.

63. Shyadehi, A. Z.; Lamb, D. C.; Kelly, S. L.; Kelly, D. E.; Schunck, W.; Wright, J. N.; Corina, D.; Akhtar, M., The Mechanism of the Acyl-Carbon Bond Cleavage Reaction Catalyzed by Recombinant Sterol 14a-Demethylase of Candida albicans (Other Names Are- Lanosterol 14a-Demethylase, P-45014DM, and CYP51)*. *J. Biol. Chem.* **1996,** *271* (21), 12445-12450.

64. Isin, E. M.; Guengerich, F. P., Complex reactions catalyzed by cytochrome P450 enzymes. *Biochim. Biophys. Acta* **2007,** *1770* (3), 314-29.

65. Kim, T. W.; Hwang, J. Y.; Kim, Y. S.; Joo, S. H.; Chang, S. C.; Lee, J. S.; Takatsuto, S.; Kim, S. K., Arabidopsis CYP85A2, a cytochrome P450, mediates the Baeyer-Villiger oxidation of castasterone to brassinolide in brassinosteroid biosynthesis. *Plant Cell* **2005,** *17* (8), 2397-412.

66. Guengerich, F. P., Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* **2001,** *14* (6), 612-650.

67. Ortiz de Montellano, P. R.; Nelson, S. D., Rearrangement reactions catalyzed by cytochrome P450s. *Arch. Biochem. Biophys.* **2011,** *507* (1), 95-110.

68. Groves, J. T.; Subramanian, D. V., Hydroxylation by Cytochrome P-450 and Metalloporphyrin Models. Evidence for Allylic Rearrangement. *J. Am. Chem. Soc.* **1984,** *106* (7), 2177-2181.

69. Ren, X.; O'Hanlon, J. A.; Morris, M.; Robertson, J.; Wong, L. L., Synthesis of Imidazolidin-4-ones via a Cytochrome P450-Catalyzed Intramolecular C–H Amination. *ACS Catal.* **2016,** *6* (10), 6833-6837.

70. Hammer, S. C.; Kubik, G.; Watkins, E.; Huang, S.; Minges, H.; Arnold, F. H., Anti-Markovnikov alkene oxidation by metal-oxo–mediated enzyme catalysis. *Science* **2017,** *358* (6360), 215-218.

71. Hirakawa, H.; Kamiya, N.; Tanaka, T.; Nagamune, T., Intramolecular electron transfer in a cytochrome P450cam system with a site-specific branched structure. *Protein Eng. Des. Sel.* **2007,** *20* (9), 453-9.

72. Hannemann, F.; Bichet, A.; Ewen, K. M.; Bernhardt, R., Cytochrome P450 systems - biological variations of electron transport chains. *Biochim. Biophys. Acta-Gen. Subj.* **2007,** *1770* (3), 330-344.

73. Kuznetsov, V. Y.; Poulos, T. L.; Sevrioukova, I. F., Putidaredoxin-to-cytochrome P450cam electron transfer: differences between the two reductive steps required for catalysis. *Biochemistry* **2006,** *45* (39), 11934-44.

74. Derouet-Humbert, E.; Roemer, K.; Bureik, M., Adrenodoxin (Adx) and CYP11A1 (P450scc) induce apoptosis by the generation of reactive oxygen species in mitochondria. *Biol. Chem.* **2005,** *386* (5), 453-61.

75. Child, S. A.; Reddish, M. J.; Glass, S. M.; Goldfarb, M. H.; Barckhausen, I. R.; Guengerich, F. P., Functional interactions of adrenodoxin with several human mitochondrial cytochrome P450 enzymes. *Arch. Biochem. Biophys.* **2020,** *694*, 108596.

76. Cao, P.; Bernhardt, R., Modulation of aldosterone biosynthesis by adrenodoxin mutants with different electron transport efficiencies. *Eur. J. Biochem.* **1999,** *265*, 152-159.

77. Zollner, A.; Kagawa, N.; Waterman, M. R.; Nonaka, Y.; Takio, K.; Shiro, Y.; Hannemann, F.; Bernhardt, R., Purification and functional characterization of human 11beta hydroxylase expressed in Escherichia coli. *FEBS J.* **2008,** *275* (4), 799-810.

78. Hanukoglu, I., Electron Transfer Proteins of Cytochrome P450 Systems. In *Physiological Functions of Cytochrome P450 in Relation to Structure and Regulation*, 1996; Vol. 14, pp 29-56.

79. Porter, T. D.; Kasper, C. B., NADPH-Cytochrome P-450 Oxidoreductase- Flavin Mononucleotide and Flavin Adenine Dinucleotide Domains Evolved from Different Flavoproteins1". *Biochemistry* **1986,** *25* (7), 1682-1687.

80. Smith, G. C. M.; Tew, D. G.; Wolf, C. R., Dissection of NADPH-cytochrome P450 oxidoreductase into distinct functional domains. *Proc. Natl. Acad. Sci. USA* **1994,** *91*, 8710-8714.

81. Wang, M.; Roberts, D. L.; Paschke, R.; Shea, T. M.; Masters, B. S. S.; Kim, J. P., Three-dimensional structure of NADPH–cytochrome P450 reductase- Prototype for FMN- and FAD-containing enzymes. *Proc. Natl. Acad. Sci. USA* **1997,** *94*, 8411-8416.

82. Kahn, R.; Durst, F., Chapter Six - Function and Evolution of Plant Cytochrome P450. In *Recent Advances in Phytochemistry*, 2000; Vol. 34, pp 151-189.

83. Fleming, B. D.; Tian, Y.; Bell, S. G.; Wong, L. L.; Urlacher, V.; Hill, H. A., Redox properties of cytochrome p450BM3 measured by direct methods. *Eur. J. Biochem.* **2003,** *270* (20), 4082-8.

84. Wright, R. L.; Harris, K.; Solow, B.; White, R. H.; Kennelly, P. J., Cloning of a potential cytochrome P450 from the Archaeon Sulfolobus solfataricus. *FEBS J.* **1996,** *384*, 235-239.

85. Nishida, C. R.; Ortiz de Montellano, P. R., Thermophilic cytochrome P450 enzymes. *Biochem. Biophys. Res. Commun.* **2005,** *338* (1), 437-45.

86. Oku, Y.; Ohtaki, A.; Kamitori, S.; Nakamura, N.; Yohda, M.; Ohno, H.; Kawarabayasi, Y., Structure and direct electrochemistry of cytochrome P450 from the thermoacidophilic crenarchaeon, Sulfolobus tokodaii strain 7. *J. Inorg. Biochem.* **2004,** *98* (7), 1194-9.

87. Hawkes, D. B.; Adams, G. W.; Burlingame, A. L.; Ortiz de Montellano, P. R.; De Voss, J. J., Cytochrome P450(cin) (CYP176A), isolation, expression, and characterization. *J. Biol. Chem.* **2002,** *277* (31), 27725-32.

88. Meharenna, Y. T.; Li, H.; Hawkes, D. B.; Pearson, A. G.; Voss, J. D.; Poulos, T. L., Crystal Structure of P450cin in a Complex with Its Substrate, 1,8-Cineole, a Close Structural Homologue to D-Camphor, the Substrate for P450cam†,‡. *Biochemistry* **2004,** *43* (29), 9487-9494.

89. Fukuda, E.; Kino, H.; Matsuzawa, H.; Wakagi, T., Role of a highly conserved YPITP motif in 2-oxoacid-ferredoxinoxidoreductase. *Eur. J. Biochem.* **2001,** *268*, 5639-5646.

90. Kojoh, K.; Matsuzawa, H.; Wakagi, T., Zinc and an N-terminal extra stretch of the ferredoxin from athermoacidophilic archaeon stabilize the molecule at high temperature. *Eur. J. Biochem.* **1999,** *264*, 85-91.

91. Puchkaev, A. V.; Ortiz de Montellano, P. R., The Sulfolobus solfataricus electron donor partners of thermophilic CYP119: an unusual non-NAD(P)H-dependent cytochrome P450 system. *Arch. Biochem. Biophys.* **2005,** *434* (1), 169-77.

92. Puchkaev, A. V.; Wakagi, T.; Montellano, P. R. O., CYP119 Plus a Sulfolobus tokodaii Strain 7 Ferredoxin and 2-Oxoacid-Ferredoxin Oxidoreductase Constitute a High-Temperature Cytochrome P450 Catalytic System. *J. Am. Chem. Soc.* **2002,** *124* (43), 12682-12683.

93. Jackson, C. J.; Lamb, D. C.; Marczylo, T. H.; Warrilow, A. G.; Manning, N. J.; Lowe, D. J.; Kelly, D. E.; Kelly, S. L., A novel sterol 14alpha-demethylase/ferredoxin fusion protein (MCCYP51FX) from Methylococcus capsulatus represents a new class of the cytochrome P450 superfamily. *J. Biol. Chem.* **2002,** *277* (49), 46959-65.

94. Rylott, E. L.; Jackson, R. G.; Edwards, J.; Womack, G. L.; Seth-Smith, H. M.; Rathbone, D. A.; Strand, S. E.; Bruce, N. C., An explosive-degrading cytochrome P450 activity and its targeted application for the phytoremediation of RDX. *Nat. Biotechnol.* **2006,** *24* (2), 216-9.

95. Seth-Smith, H. M.; Rosser, S. J.; Basran, A.; Travis, E. R.; Dabbs, E. R.; Nicklin, S.; Bruce, N. C., Cloning, sequencing, and characterization of the hexahydro-1,3,5-Trinitro-1,3,5-triazine degradation gene cluster from Rhodococcus rhodochrous. *Appl. Environ. Microbiol.* **2002,** *68* (10), 4764-71.

96. Roberts, G. A.; Grogan, G.; Greter, A.; Flitsch, S. L.; Turner, N. J., Identification of a new class of cytochrome P450 from a Rhodococcus sp. *J. Bacteriol.* **2002,** *184* (14), 3898-908.

97. Hunter, D. J.; Roberts, G. A.; Ost, T. W.; White, J. H.; Muller, S.; Turner, N. J.; Flitsch, S. L.; Chapman, S. K., Analysis of the domain properties of the novel cytochrome P450 RhF. *FEBS Lett.* **2005,** *579* (10), 2215-20.

98. Warman, A. J.; Roitel, O.; Neeli, R.; Girvan, H. M.; Seward, H. E.; Murray, S. A.; McLean, K. J.; Joyce, M. G.; Toogood, H.; Holt, R. A.; Leys, D.; Scrutton, N. S.; Munro, A. W., Flavocytochrome P450 BM3- an update on structure and mechanism of a biotechnologically important enzyme. *Biochem. Soc. Trans.* **2005,** *33* (4), 747-753.

99. Miura, Y.; Fulco, A. J., (w- 2) Hydroxylation of Fatty Acids by a Soluble System from Bacillus meguterium* *J. Biol. Chem.* **1974,** *249* (6), 1880-1888.

100. Fulco, A. J., P450BM-3 and other inducible bacterial P450 cytochromes- biochemistry and regulation. *Annu. Rev. Pharmacol. Toxicol.* **1991,** *31*, 177-203.

101. Ost, T. W.; Clark, J.; Mowat, C. G.; Miles, C. S.; Walkinshaw, M. D.; Reid, G. A.; Chapman, S. K.; Daff, S., Oxygen Activation and Electron Transfer in Flavocytochrome P450 BM3. *J. Am. Chem. Soc.* **2003,** *125* (49), 15010-15020.

102. Daiber, A.; Shoun, H.; Ullrich, V., Nitric oxide reductase (P450nor) from Fusarium oxysporum. *J. Inorg. Biochem.* **2005,** *99* (1), 185-93.

103. Nakahara, K.; Tanimoto, T.; Hatano, K.; Usuda, K.; Shoun, H., Cytochrome P-450 55A1 (P-450dNIR) Acts as Nitric Oxide Reductase Employing NADH as the Direct Electron Donor. *J. Biol. Chem.* **1992,** *268* (11), 8350-8355.

104. Froehlich, J. E.; Itoh, A.; Howe, G. A., Tomato Allene Oxide Synthase and Fatty Acid Hydroperoxide Lyase, Two Cytochrome P450s Involved in Oxylipin Metabolism, Are Targeted to Different Membranes of Chloroplast Envelope. *Plant Physiol.* **2001,** *125*, 306-317.

105. Meunier, B.; Visser, S. P.; Shaik, S., Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. *Chem. Rev.* **2004,** *104* (9), 3947-3980.

106. Rittle, J.; Green, M. T., Cytochrome P450 Compound I- Capture, Characterization, and C-H Bond Activation Kinetics. *Science* **2010,** *330* (6006), 933-937.

107. Newcomb, M.; Zhang, R.; Chandrasena, R. E.; Halgrimson, J. A.; Horner, J. H.; Makris, T. M.; Sligar, S. G., Cytochrome p450 compound I. *J. Am. Chem. Soc.* **2006,** *128* (14), 4580-1.

108. Li, X. X.; Postils, V.; Sun, W.; Faponle, A. S.; Sola, M.; Wang, Y.; Nam, W.; de Visser, S. P., Reactivity Patterns of (Protonated) Compound II and Compound I of Cytochrome P450: Which is the Better Oxidant? *Chemistry* **2017,** *23* (26), 6406-6418.

109. Loida, P. J.; Sligar, S. G., Molecular Recognition in Cytochrome P-450- Mechanism for the Control of Uncoupling Reactions. *Biochemistry* **1993,** *32* (43), 11530-11538.

110. Grinkova, Y. V.; Denisov, I. G.; McLean, M. A.; Sligar, S. G., Oxidase uncoupling in heme monooxygenases: human cytochrome P450 CYP3A4 in Nanodiscs. *Biochem. Biophys. Res. Commun.* **2013,** *430* (4), 1223-7.

111. Kim, B. H.; Fulco, A. J., INDUCTION BY BARBITURATES OF A CYTOCHROME P-450-DFPENDENT FATTY ACID MONOOXYGENASE IN BACILLUS MEGATERIUM- RELATIONSHIP BETWEEN BARBITURATE STRUCTURE AND INDUCER ACTIVITY. *Biochem. Biophys. Res. Commun.* **1983,** *116* (3), 843-850.

112. P., H. P.; Fulco, A. J., INVOLVEMENT OF A SINGLE HYDROXYLASE SPECIES IN THE HYDROXYLATION OF PALMITATE AT THE o-l, w-2 AND w-3 POSITIONS BY A PREPARATION FROM BA CZLL US MEGA TERZUM. *Biochim. Biophys. Acta-Gen. Subj.* **1976,** *431*, 249-256.

113. Maston, R. S.; Hare, R. S.; Fulco, A. J., CHARACTERISTICS OF A CYTOCHROME P-450-DEPENDENT FATTY ACID w-2 HYDROXYLASE FROM BACILLUS'MEGATERIUM. *Biochim. Biophys. Acta.* **1977,** *487*, 487-494.

114. Narhi, L. O.; Wen, L.; Fulco, A. J., Characterization of the protein expressed in Escherichia coli by a recombinant plasmid containing the Bacillus megaterium cytochrome P-450BM-3 gene. *Mol. Cell. Biochem.* **1988,** *79*, 63-71.

115. Narhi, L. O.; Kim, B. H.; Stevenson, P. M.; Fulco, A. J., PARTIAL CHARACTERIZATION OF A BARBITURATE-INDUCED CYTOCHROME P-450-DEPENDENT FATTY ACID MONOOXYGENASE FROM BACILLUS MEGATERIUM. *Biochem. Biophys. Res. Commun.* **1983,** *116* (3), 851-858.

116. Joyce, M. G.; Ekanem, I. S.; Roitel, O.; Dunford, A. J.; Neeli, R.; Girvan, H. M.; Baker, G. J.; Curtis, R. A.; Munro, A. W.; Leys, D., The crystal structure of the FAD/NADPH-binding domain of flavocytochrome P450 BM3. *FEBS J.* **2012,** *279* (9), 1694-706.

117. Hazzard, J. T.; Govindaraj, S.; Poulos, T. L.; Tollin, G., Electron Transfer between the FMN and Heme Domains of Cytochrome P450BM-3. *J. Biol. Chem.* **1997,** *272* (12), 7922-7926.

118. Sevrioukova, I. F.; Li, H. Y.; Zhang, H.; Peterson, J. A.; Poulos, T. L., Structure of a cytochrome P450–redox partner electron-transfer complex. *Proc. Natl. Acad. Sci. USA* **1999,** *96*, 1863-1868.

119. Black, S. D.; Martin, S. T., Evidence for Conformational Dynamics and Molecular Aggregation in Cytochrome P450 102 (BM-3)* *Biochemistry* **1994,** *33* (40), 12056-12062.

120. Kitazume, T.; Haines, D. C.; Estabrook, R. W.; Chen, B.; Peterson, J. A., Obligatory Intermolecular Electron-Transfer from FAD to FMN in Dimeric P450BM-3. *Biochemistry* **2007,** *46* (42), 11892-11901.

121. Zhang, H.; Yokom, A. L.; Cheng, S.; Su, M.; Hollenberg, P. F.; Southworth, D. R.; Osawa, Y., The full-length cytochrome P450 enzyme CYP102A1 dimerizes at its reductase domains and has flexible heme domains for efficient catalysis. *J. Biol. Chem.* **2018,** *293* (20), 7727-7736.

122. Rowlatt, B.; Yorke, J. A.; Strong, A. J.; Whitehouse, C. J.; Bell, S. G.; Wong, L. L., Chain length-dependent cooperativity in fatty acid binding and oxidation by cytochrome P450BM3 (CYP102A1). *Protein Cell* **2011,** *2* (8), 656-71.

123. Gotoh, O., Substrate Recognition Sites in Cytochrome P450 Family 2 (CYPB) Proteins Inferred from Comparative Analyses of Amino Acid and Coding Nucleotide Sequences. *J. Biol. Chem.* **1992,** *267* (1), 83-90.

124. Butler, C. F.; Peet, C.; Mason, A. E.; Voice, M. W.; Leys, D.; Munro, A. W., Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias. *J. Biol. Chem.* **2013,** *288* (35), 25387-99.

125. Dueholm, B.; Krieger, C.; Drew, D.; Olry, A.; Kamo, T.; Taboureau, O.; Weitzel, C.; Bourgaud, F.; Hehn, A.; Simonsen, H. T., Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. *BMC Evol. Biol.* **2015,** *15* (122), 1-14.

126. Tietz, D. R.; Colthart, A. M.; Sondej Pochapsky, S.; Pochapsky, T. C., Substrate recognition by two different P450s: Evidence for conserved roles in a common fold. *Sci. Rep.* **2017,** *7* (1), 13581.

127. Hasemann, C. A.; Kurumbail, R. G.; Boddupalli, S. S.; Peterson, J. A.; Deisenhofer, J., Structure and function of cytochromes P450-a comparative analysis of three crystal structures. *Structure* **1995,** *2* (1), 41-62.

128. Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A., Pivotal Role of Water in the Mechanism of P450BM-3. *Biochemistry* **2001,** *40* (45), 13456-13465.

129. Li, H.; Poulos, T. L., Fatty acid metabolism, conformational change, and electron transfer incytochromeP-450BM-3. *Biochim. Biophys. Acta* **1999,** *1441*, 141-149.

130. Li, H. Y.; Poulos, T. L., The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat. Struct. Biol.* **1997,** *4* (2), 140-146.

131. Noble, M. A.; Miles, C. S.; Chapman, S. K.; Lysek, D. A.; Mackay, A. C.; Reid, G. A.; Hanzlik, R. P.; Munro, A. W., Roles of key active-site residues in flavocytochrome P450 BM3. *Biochem. J.* **1999,** *339*, 371-379.

132. Graham-Lorence, S.; Truan, G.; Peterson, J. A.; Falck, J. R.; Wei, S.; Helvig, C.; Capdevila, J. H., An Active Site Substitution, F87V, Converts Cytochrome P450 BM-3 into a Regio- and Stereoselective (14S,15R)-Arachidonic Acid Epoxygenase. *J. Biol. Chem.* **1997,** *272* (2), 1127-1135.

133. Oliver, C. F.; Modi, S.; Primrose, W. U.; Lian, L.; Roberts, G. C., Engineering the substrate specificity of Bacillus megaterium cytochrome P450 BM3 hydroxylation of alkyl trimethylammonium compounds. *Biochem. J.* **1997,** *327*, 537-544.

134. Carmichael, A. B.; Wong, L. L., Protein engineering of Bacillus megaterium CYP102 The oxidation of polycyclic aromatic hydrocarbons. *Eur. J. Biochem.* **2001,** *268* (10), 3117-3125.

135. Ren, X.; Yorke, J. A.; Taylor, E.; Zhang, T.; Zhou, W.; Wong, L. L., Drug Oxidation by Cytochrome P450BM3 : Metabolite Synthesis and Discovering New P450 Reaction Types. *Chemistry* **2015,** *21* (42), 15039-47.

136. Urban, P.; Truan, G.; Pompon, D., Access channels to the buried active site control substrate specificity in CYP1A P450 enzymes. *Biochim. Biophys. Acta Gen. Subj.* **2015,** *1850* (4), 696-707.

137. Vottero, E.; Rea, V.; Lastdrager, J.; Honing, M.; Vermeulen, N. P. E.; Commandeur, J. N. M., Role of residue 87 in substrate selectivity and regioselectivity of drug-metabolizing cytochrome P450 CYP102A1 M11. *Journal of Biological Inorganic Chemistry* **2011,** *16* (6), 899-912.

138. Rea, V.; Dragovic, S.; Boerma, J. S.; de Kanter, F. J. J.; Vermeulen, N. P. E.; Commandeur, J. N. M., Role of Residue 87 in the Activity and Regioselectivity of Clozapine Metabolism by Drug-Metabolizing CYP102A1 M11H: Application for Structural Characterization of Clozapine GSH Conjugates. *Drug Metab. Dispos* **2011,** *39* (12), 2411-2420.

139. Graham-Lorence, S.; Truan, G.; Peterson, J. A.; Falck, J. R.; Wei, S.; Helvig, C.; Capdevila, J. H., An Active Site Substitution, F87V, Converts Cytochrome P450 BM-3 into a Regio- and Stereoselective (14S,15R)-Arachidonic Acid Epoxygenase*. *J. Biol. Chem.* **1997,** *272* (2), 1127-1135.

140. Carmichael, A. B.; Wong, L. L., Protein engineering of Bacillus megaterium CYP102 - The oxidation of polycyclic aromatic hydrocarbons. *Eur. J. Biochem.* **2001,** *268* (10), 3117-3125.

141. Huang, W. C.; Westlake, A. C.; Marechal, J. D.; Joyce, M. G.; Moody, P. C.; Roberts, G. C., Filling a hole in cytochrome P450 BM3 improves substrate binding and catalytic efficiency. *J. Mol. Biol.* **2007,** *373* (3), 633-51.

142. Ravichandran, K. G.; Boddupalli, S. S.; Hasemann, C. A.; Peterson, J. A.; Deisenhofer, J., Crystal Structure of Hemoprotein Domain of P45OBM-3, a Prototype for Microsomal P450's. *Science* **1993,** *261* (5122), 731-736.

143. Peters, M. W.; Meinhold, P.; Glieder, A.; Arnold, F. H., Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. *J. Am. Chem. Soc.* **2003,** *125* (44), 13442-13450.

144. Siloto, R. M. P.; Weselake, R. J., Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatal. Agric. Biotechnol.* **2012,** *1* (3), 181-189.

145. Li, Q.-S.; Schwaneberg, U.; Fischer, P.; Schimd, R. D., Directed Evolution of the Fatty acid Hydroxylase P450BM3 into an indole-hydroxylating catalyst. *Chem. Eur. J.* **2000,** *6* (9), 1531-1536.

146. Gillam, E. M. J.; Aguinaldo, A. M. A.; Notley, L. M.; Kim, D.; Mundkowski, R. G.; Volkov, A. A.; Arnold, F. H.; Soucek, P.; De Voss, J. J.; Guengerich, F. P., Formation of Indigo by Recombinant MammalianCytochrome P450. *Biochem. Biophys. Res. Commun.* **1999,** *265* (2), 469-472.

147. Appel, D.; Lutz, S.; Fischer, P.; Schwaneberg, U.; Schmid, R. D., A P450 BM-3 mutant hydroxylates alkanes, cycloalkanes,arenes and heteroarenes. *J. Biotechnol.* **2001,** *88*, 167-171.

148. Whitehouse, C. J. C.; Bell, S. G.; Tufton, H. G.; Kenny, R. J. P.; Ogilvie, L. C. I.; Wong, L. L., Evolved CYP102A1 (P450(BM3)) variants oxidise a range of non-natural substrates and offer new selectivity options. *Chem. Commun.* **2008,** (8), 966-968.

149. Li, Y.; Wong, L. L., Multi-Functional Oxidase Activity of CYP102A1 (P450BM3) in the Oxidation of Quinolines and Tetrahydroquinolines. *Angew. Chem. Int. Ed. Engl.* **2019,** *58* (28), 9551-9555.

150. Syntrivanis, L.-D.; Wong, L. L.; Robertson, J., Hydroxylation of Eleuthoside Synthetic Intermediates by P450BM3 (CYP102A1). *Eur. J. Org. Chem.* **2018,** *2018* (45), 6369-6378.

151. Whitehouse, C. J.; Yang, W.; Yorke, J. A.; Rowlatt, B. C.; Strong, A. J.; Blanford, C. F.; Bell, S. G.; Bartlam, M.; Wong, L. L.; Rao, Z., Structural basis for the properties of two single-site proline mutants of CYP102A1 (P450BM3). *ChemBioChem* **2010,** *11* (18), 2549-56.

152. Whitehouse, C. J. C.; Bell, S. G.; Yang, W.; Yorke, J. A.; Blanford, C. F.; Strong, A. J. F.; Morse, E. J.; Bartlam, M.; Rao, Z. H.; Wong, L. L., A Highly Active Single-Mutation Variant of P450(BM3) (CYP102A1). *ChemBioChem* **2009,** *10* (10), 1654-1656.

153. Whitehouse, C. J.; Yang, W.; Yorke, J. A.; Tufton, H. G.; Ogilvie, L. C. I.; Bell, S. G.; Zhou, W.; Bartlam, M.; Rao, Z.; Wong, L. L., Structure, electronic properties and catalytic behaviour of an activity-enhancing CYP102A1 (P450BM3) variant. *Dalton Trans.* **2011,** *40* (40), 10383-10396.

154. Fasan, R.; Jennifer Kan, S. B.; Zhao, H., A Continuing Career in Biocatalysis: Frances H. Arnold. *ACS Catal.* **2019,** *9* (11), 9775-9788.

155. Glieder, A.; Farinas, E. T.; Arnold, F. H., Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat. Biotechnol.* **2002,** *20* (11), 1135-9.

156. Meinhold, P.; Peters, M. W.; Chen, M. M.; Takahashi, K.; Arnold, F. H., Direct conversion of ethane to ethanol by engineered cytochrome P450 BM3. *ChemBioChem* **2005,** *6* (10), 1765-8.

157. Fasan, R.; Chen, M. M.; Crook, N. C.; Arnold, F. H., Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties. *Angew. Chem. Int. Ed. Engl.* **2007,** *46* (44), 8414-8.

158. Lewis, J. C.; Bastian, S.; Bennett, C. S.; Fu, Y.; Mitsuda, Y.; Chen, M. M.; Greenberg, W. A.; Wong, C. H.; Arnold, F. H., Chemoenzymatic elaboration of monosaccharides using engineered cytochrome P450BM3 demethylases. *Proc. Natl. Acad. Sci. U S A* **2009,** *106* (39), 16550-5.

159. Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H., Evolutionary history of a specialized p450 propane monooxygenase. *J. Mol. Biol.* **2008,** *383* (5), 1069-80.

160. Landwehr, M.; Hochrein, L.; Otey, C. R.; Kasrayan, A.; Backvall, J. E.; Arnold, F. H., Enantioselective r-Hydroxylation of 2-Arylacetic Acid Derivatives and Buspirone Catalyzed by Engineered Cytochrome P450 BM-3. *J. Am. Chem. Soc.* **2006,** *128* (18), 6058-6059.

161. Sawayama, A. M.; Chen, M. M.; Kulanthaivel, P.; Kuo, M. S.; Hemmerle, H.; Arnold, F. H., A panel of cytochrome P450 BM3 variants to produce drug metabolites and diversify lead compounds. *Chemistry* **2009,** *15* (43), 11723-9.

162. van Vugt-Lussenburg, B. M.; Damsten, M. C.; Maasdijk, D. M.; Vermeulen, N. P.; Commandeur, J. N., Heterotropic and homotropic cooperativity by a drug-metabolising mutant of cytochrome P450 BM3. *Biochem. Biophys. Res. Commun.* **2006,** *346* (3), 810-8.

163. Lussenburg, B. M.; Babel, L. C.; Vermeulen, N. P.; Commandeur, J. N., Evaluation of alkoxyresorufins as fluorescent substrates for cytochrome P450 BM3 and site-directed mutants. *Anal. Biochem.* **2005,** *341* (1), 148-55.

164. Vugt-Lussenburg, B. M. A.; Sternschantz, E.; Lastdrager, J.; Oostenbrink, C.; Vermeulen, N. P.; Commandeur, J. N., Identification of Critical Residues in Novel Drug Metabolizing Mutants of Cytochrome P450 BM3 Using Random Mutagenesis. *J. Med. Chem.* **2007,** *50* (3), 455-461.

165. Rea, V.; Kolkman, A. J.; Vottero, E.; Stronks, E. J.; Ampt, K. A.; Honing, M.; Vermeulen, N. P.; Wijmenga, S. S.; Commandeur, J. N., Active site substitution A82W improves the regioselectivity of steroid hydroxylation by cytochrome P450 BM3 mutants as rationalized by spin relaxation nuclear magnetic resonance studies. *Biochemistry* **2012,** *51* (3), 750-60.

166. Venkataraman, H.; Beer, S. B.; Bergen, L. A.; Essen, N.; Geerke, D. P.; Vermeulen, N. P.; Commandeur, J. N., A single active site mutation inverts stereoselectivity of 16-hydroxylation of testosterone catalyzed by engineered cytochrome P450 BM3. *ChemBioChem* **2012,** *13* (4), 520-3.

167. Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A., Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chem. Int. Ed. Engl.* **2005,** *44* (27), 4192-6.

168. Reetz, M. T.; Carballeira, J. D., Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2007,** *2* (4), 891-903.

169. Kille, S.; Zilly, F. E.; Acevedo, J. P.; Reetz, M. T., Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **2011,** *3* (9), 738-743.

170. Acevedo-Rocha, C. G.; Gamble, C. G.; Lonsdale, R.; Li, A.; Nett, N.; Hoebenreich, S.; Lingnau, J. B.; Wirtz, C.; Fares, C.; Hinrichs, H.; Deege, A.; Mulholland, A. J.; Nov, Y.; Leys, D.; McLean, K. J.; Munro, A. W.; Reetz, M. T., P450-Catalyzed Regio- and Diastereoselective Steroid Hydroxylation: Efficient Directed Evolution Enabled by Mutability Landscaping. *ACS Catal.* **2018,** *8* (4), 3395-3410.

171. Li, A.; Acevedo-Rocha, C. G.; D'Amore, L.; Chen, J.; Peng, Y.; Garcia-Borras, M.; Gao, C.; Zhu, J.; Rickerby, H.; Osuna, S.; Zhou, J.; Reetz, M. T., Regio- and Stereoselective Steroid Hydroxylation at C7 by Cytochrome P450 Monooxygenase Mutants. *Angew. Chem. Int. Ed. Engl.* **2020,** *59* (30), 12499-12505.

172. Chen, W. Y.; Fisher, M. J.; Leung, A.; Cao, Y.; Wong, L. L., Oxidative Diversification of Steroids by Nature-Inspired Scanning Glycine Mutagenesis of P450BM3 (CYP102A1). *Acs Catal.* **2020,** *10* (15), 8334-8343.

173. Seifert, A.; Vomund, S.; Grohmann, K.; Kriening, S.; Urlacher, V. B.; Laschat, S.; Pleiss, J., Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *ChemBioChem* **2009,** *10* (5), 853-61.

174. Zhang, K.; Shafer, B. M.; Demars, M. D., 2nd; Stern, H. A.; Fasan, R., Controlled oxidation of remote sp3 C-H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity. *J. Am. Chem. Soc.* **2012,** *134* (45), 18695-704.

175. Zhang, K.; El Damaty, S.; Fasan, R., P450 fingerprinting method for rapid discovery of terpene hydroxylating P450 catalysts with diversified regioselectivity. *J. Am. Chem. Soc.* **2011,** *133* (10), 3242-5.

176. Rentmeister, A.; Arnold, F. H.; Fasan, R., Chemo-enzymatic fluorination of unactivated organic compounds. *Nat. Chem. Biol.* **2009,** *5* (1), 26-8.

177. Roiban, G. D.; Reetz, M. T., Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun. (Camb)* **2015,** *51* (12), 2208-24.

178. Damsten, M. C.; van Vugt-Lussenburg, B. M.; Zeldenthuis, T.; de Vlieger, J. S.; Commandeur, J. N.; Vermeulen, N. P., Application of drug metabolising mutants of cytochrome P450 BM3 (CYP102A1) as biocatalysts for the generation of reactive metabolites. *Chem. Biol. Interact.* **2008,** *171* (1), 96-107.

179. Reinen, J.; Ferman, S.; Vottero, E.; Vermeulen, N. P.; Commandeur, J. N., Application of a Fluorescence-Based Continuous-Flow Bioassay to Screen for Diversity of Cytochrome P450 BM3 Mutant Libraries. *J. Biomol. Screen.* **2011,** *16* (2), 239-250.

180. Dietrich, J. A.; Yoshikuni, Y.; Fisher, K. J.; Woolard, F. X.; Ockey, D.; McPhee, D. J.; Renninger, N. S.; Chang, M. C. Y.; Baker, D.; Keasling, J. D., A Novel Semi-biosynthetic Route for Artemisinin Production Using Engineered Substrate-Promiscuous P450BM3. *ACS Chem. Biol.* **2009,** *4* (4), 261-267.

181. Oliver, C. F.; Modi, S.; Sutcliffe, M. J.; Primrose, W. U.; Lian, L.; Roberts, G. C. K., A Single Mutation in Cytochrome P450 BM3 Changes Substrate Orientation in a Catalytic Intermediate and the Regiospecificity of Hydroxylation. *Biochemistry* **1997,** *36* (7), 1567-1572.

182. Wriessnegger, T.; Augustin, P.; Engleder, M.; Leitner, E.; Muller, M.; Kaluzna, I.; Schurmann, M.; Mink, D.; Zellnig, G.; Schwab, H.; Pichler, H., Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of Pichia pastoris. *Metab. Eng.* **2014,** *24*, 18-29.

183. Sowden, R. J.; Yasmin, S.; Rees, N. H.; Bell, S. G.; Wong, L. L., Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450cam and P450BM-3. *Org. Biomol. Chem.* **2005,** *3* (1), 57-64.

184. Shaffer, G. W.; Eschinasi, E. H.; Purzycki, K. L.; Doerr, A. B., Oxidations of Valencene. *J. Org. Chem.* **1975,** *40* (15), 2181-2185.

185. Davies, A. G.; Davison, I. G. E., The Rearrangements of Allylic Hydroperoxides Derived from (+)-Valencene. *J. Chem. Soc., Perkin Trans.* **1989,** (7), 825-830.

186. Bombarda, I.; Gaydou, E. M.; Smadja, J.; Faure, R., Sesquiterpenic Epoxides and Alcohols Derived from Hydrocarbons of Vetiver Essential Oil. *J. Agric. Food Chem.* **1996,** *44* (1), 217-222.

187. Arantes, S. F.; Farooq, A.; Hanson, J. R., The preparation and microbiological hydroxylation of the sesquiterpenoid nootkatone. *Journal of Chemical Research-S* **1999,** (3), 248-248a.

188. Salvador, J. A. R.; Clark, J. H., The allylic oxidation of unsaturated steroids by tert-butyl hydroperoxide using surface functionalised silica supported metal catalysts. *Green Chem.* **2002,** *4* (4), 352-356.

189. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

190. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78* (8), 1950-1958.

191. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

192. Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 3rd, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **2012,** *33* (2), 119-133.

193. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

194. Hess, B., P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008,** *4* (1), 116-122.

195. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1), 014101.

196. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

197. Parrinello, M.; Rahman, A., Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.

198. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E., Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* **1999,** *38* (1), 236-240.

199. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.

200. Mikami, K.; Hosokawa, M., Biosynthetic pathway and health benefits of fucoxanthin, an algae-specific xanthophyll in brown seaweeds. *Int. J. Mol. Sci.* **2013,** *14* (7), 13763-81.

201. Bezman, Y.; Bilkis, I.; Winterhalter, P.; Fleischmann, P.; Rouseff, R. L.; Baldermann, S.; Naim, M., Thermal Oxidation of 9′-cis-Neoxanthin in a Model System Containing Peroxyacetic Acid Leads to the Potent Odorant beta-Damascenone. *J. Agric. Food Chem* **2005,** *53* (23), 9199-9206.

202. Isoe, S.; Katsumura, S.; Sakan, T., The Synthesis of Damascenone and beta-Damascone and The Possible Mechanism of Their Formation from Carotenoid. *Helv. Chim. Acta* **1973,** *56* (5), 1514-1516.

203. Skouroumounis, G. K.; Massy-Westropp, R. A.; Sefton, M. A.; Williams, P. J., Precursors of Damascenone in Fruit Juices. *Tetrahedron Lett.* **1992,** *33* (24), 3533-3536.

204. Azzari, E.; Faggi, C.; Gelsomini, N.; Taddei, M., Transformation of a- and b-Ionones into a- and b-Damascone and d-Damascenone Using Allylsilane Chemistry. *J. Org. Chem.* **1990,** *55* (3), 1106-1108.

205. Ren, T.-L.; Xu, B.-H.; Mahmood, S.; Sun, M.-X.; Zhang, S.-J., Cobalt-catalyzed oxidative esterification of allylic/benzylic C(sp 3)–H bonds. *Tetrahedron* **2017,** *73* (20), 2943-2948.

206. Ma, M.; Bell, S. G.; Yang, W.; Hao, Y.; Rees, N. H.; Bartlam, M.; Zhou, W.; Wong, L. L.; Rao, Z., Structural Analysis of CYP101C1 from Novosphingobium aromaticivorans DSM12444. *ChemBioChem* **2011,** *12* (1), 88-99.

207. Clemente-Tejeda, D.; López-Moreno, A.; Bermejo, F. A., Non-heme iron catalysis in CC, C–H, and CH2 oxidation reactions. Oxidative transformations on terpenoids catalyzed by Fe(bpmen)(OTf)2. *Tetrahedron* **2013,** *69* (14), 2977-2986.

208. Kiss, F. M.; Lundemo, M. T.; Zapp, J.; Woodley, J. M.; Bernhardt, R., Process development for the production of 15beta-hydroxycyproterone acetate using Bacillus megaterium expressing CYP106A2 as whole-cell biocatalyst. *Microb. Cell. Fact.* **2015,** *14* (28), 1-13.

209. Takazawa, O.; Tamura, H.; Kogami, K.; Hayashi, K., New Synthesis of Megastigma-4,6,8-trien-3-ones, 3-Hydroxy-b-ionol, 3-Hydroxy-b-ionone, 5,6-Epoxy-3-hydroxy-b-ionol, and 3-oxo-a-ionol. *Bull. Chem. Soc. Jpn.* **1982,** *55* (6), 1907-1911.

210. Hall, E. A.; Bell, S. G., The efficient and selective biocatalytic oxidation of norisoprenoid and aromatic substrates by CYP101B1 from Novosphingobium aromaticivorans DSM12444. *RSC Adv.* **2015,** *5* (8), 5762-5773.

211. Qin, Z.; Xiao, Y.; Han, X.; Zhou, Z.; Wan, C., Co(acac)2-Catalyzed Allylic and Benzylic Oxidation by tert-Butyl Hydroperoxide. *Synthesis* **2013,** *45* (05), 615-620.

212. Aasen, A. J.; Kimland, B.; Enzell, C. R., Tabacco Chemistry-18. Absolute Configuration of (9R)-9-Hydroxy-4,7E-megastigmadien-3-one (3-Oxo-a-ionol). *Acta Chem. Scand.* **1973,** *27* (6), 2107-2114.

213. Hutchins, R. O.; Learn, K.; El-Telbany, F.; Stercho, Y., Aminoborohydrides as Reducing Agents. 1. Sodium (Dimethylamino)- and (tert -Butylamino)borohydrides as Selective Reducing Agents18. *J. Org. Chem.* **1984,** *49* (13), 2438-2443.

214. Englert, G., A 13C-NMR. Study of cis-trans Isomeric Vitamins A, Carotenoids and Related Compounds. *Helv. Chim. Acta* **1975,** *58*, 2367-2390.

215. Hutchins, R. O.; Learn, K.; El-Telbany, F.; Stercho, Y., Aminoborohydrides as Reducing Agents. 1. Sodium (Dimethylamino)- and (tert -Butylamino)borohydrides as Selective Reducing Agents. *J. Org. Chem.* **1984,** *49* (13), 2438-2443.

216. Ito, N.; Etoh, T.; Hagiwara, H.; Kato, M., Novel synthesis of degradation products of carotenoids, megastigmatrienone analogues and blumenol-A. *J. Chem. Soc., Perkin Trans.* **1997,** *1* (10), 1571-1579.

217. Yorozu, K.; Takai, T.; Yamada, T.; Mukaiyama, T., A Novel Method for the Preparation of Acid-Sensitive Epoxides from Olefins with the Combined Use of Molecular Oxygen and Aldoacetal Catalyzed by a Cobalt (II) Complex. *Bull. Chem. Soc. Jpn.* **1994,** *67* (8), 2195-2202.

218. Behr, D.; Wahlberg, I.; Nishida, T.; Enzell, C. R., Tabacco Chemistry. 47. (3S, 6R, 7E, 9R)- and (3S*, 6R*, 7E, 9S*)-4,7-Megstimadiene-3,9-diol. Two New Nor-carotenoids of Greek Tobacco. *Acta Chem. Scand.* **1978,** *32* (6), 391-394.

219. D'Abrosca, B.; DellaGreca, M.; Fiorentino, A.; Monaco, P.; Oriano, P.; Temussi, F., Structure elucidation and phytotoxicity of C13 nor-isoprenoids from Cestrum parqui. *Phytochemistry* **2004,** *65* (4), 497-505.

220. Murakami, T.; Kishi, A.; Matsuda, H.; Hattori, M.; Yoshikawa, M., Medicinal Foodstuffs. XXIV.1) Chemical Constituents of the Processed Leaves of Apocynum venetum L.- Absolute Stereostructures of Apocynosides I and II. *Chem. Pharm. Bull.* **2001,** *49* (7), 845-848.

221. Ho, T.-L.; Din, Z. U., A Rose Oxide Synthesis. *Synth. Commun.* **2006,** *12* (14), 1099-1102.

222. Wilhelm Pickenhagen, H.; Dietmar Schatkowski, S. Process for producing rose oxide. 5,892,059, 1999.

223. Ravelli, D.; Protti, S.; Neri, P.; Fagnoni, M.; Albini, A., Photochemical technologies assessed: the case of rose oxide. *Green Chem.* **2011,** *13* (7), 1876-1884.

224. Coyle, E. E.; Oelgemoller, M., Micro-photochemistry: photochemistry in microstructured reactors. The new photochemistry of the future? *Photochem. Photobiol. Sci.* **2008,** *7* (11), 1313-22.

225. Meyer, S.; Tietze, D.; Rau, S.; Schäfer, B.; Kreisel, G., Photosensitized oxidation of citronellol in microreactors. *Journal of Photochemistry and Photobiology A: Chemistry* **2007,** *186* (2-3), 248-253.

226. Ghogare, A. A.; Greer, A., Using Singlet Oxygen to Synthesize Natural Products and Drugs. *Chem. Rev.* **2016,** *116* (17), 9994-10034.

227. Jähnisch, K.; Dingerdissen, U., Photochemical Generation and [4+2]-Cycloaddition of Singlet Oxygen in a Falling-Film Micro Reactor. *Chem. Eng. Technol.* **2005,** *28* (4), 426-427.

228. Alsters, P. L.; Jary, W.; Nardello-Rataj, V.; Aubry, J., "Dark" Singlet Oxygenation of -Citronellol- A Key Step in the Manufacture of Rose Oxide. *Org. Process Res. Dev.* **2010,** *14* (1), 259-262.

229. Sagadevan, A.; Hwang, K. C.; Su, M. D., Singlet oxygen-mediated selective C-H bond hydroperoxidation of ethereal hydrocarbons. *Nat. Commun.* **2017,** *8* (1), 1812.

230. Bruckner, S.; Weise, M.; Schobert, R., Synthesis of the Entomopathogenic Fungus Metabolites Militarinone C and Fumosorinone A. *J. Org. Chem.* **2018,** *83* (18), 10805-10812.

231. Flachsbarth, B.; Fritzsche, M.; Weldon, P.; Schulz, S., Compisiiton of the Cloacal Gland Secretion of Tuatara, Sphenodon punctatus. *Chem. Biodivers.* **2009,** *6*, 1-37.

232. Zhao, X. Z.; Tu, Y. Q.; Peng, L.; Li, X. Q.; Jia, Y. X., Synthetic studies of the HIV-1 protease inhibitive didemnaketals: stereocontrolled synthesis of an ester side chain. *Tetrahedron Lett.* **2004,** *45* (19), 3713-3716.

233. Biellmann, J. F., Enantiomeric Steroids: Synthesis, Physical, and Biological Properties. *Chem. Rev.* **2003,** *103* (5), 2019-2033.

234. Szaleniec, M.; Wojtkiewicz, A. M.; Bernhardt, R.; Borowski, T.; Donova, M., Bacterial steroid hydroxylases: enzyme classes, their functions and comparison of their catalytic mechanisms. *Appl. Microbiol. Biotechnol.* **2018,** *102* (19), 8153-8171.

235. Goldstein, J. L.; Brown, M. S., Regulation of the mevalonate pathway. *Nature* **1990,** *343*, 425-430.

236. Payne, A. H.; Hales, D. B., Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. *Endocrine Reviews* **2004,** *25* (6), 947-970.

237. Miller, W. L.; Auchus, R. J., The Molecular Biology, Biochemistry, and Physiology of Human Steroidogenesis and Its Disorders. *Endocr. Rev.* **2011,** *32* (1), 81-151.

238. Mooradian, A. D.; Morley, J. E.; Korenman, S. G., Biological Actions of Androgens. *Endocr. Rev.* **1987,** *8* (1), 1-28.

239. Nilsson, S.; Gustafsson, J. A., Biological role of estrogen and estrogen receptors. *Crit. Rev. Biochem. Mol. Biol.* **2002,** *37* (1), 1-28.

240. Ericson-Neilen, W.; Kaye, A. D., Steroids- Pharmacology, Complications, and Practice Delivery Issues. *Ochsner J.* **2014,** *14* (2), 203-207.

241. Veiga, S.; Melcangi, R. C.; Doncarlos, L. L.; Garcia-Segura, L. M.; Azcoitia, I., Sex hormones and brain aging. *Exp. Gerontol.* **2004,** *39* (11-12), 1623-31.

242. Khattab, A.; Haider, S.; Kumar, A.; Dhawan, S.; Alam, D.; Romero, R.; Burns, J.; Li, D.; Estatico, J.; Rahi, S.; Fatima, S.; Alzahrani, A.; Hafez, M.; Musa, N.; Razzghy Azar, M.; Khaloul, N.; Gribaa, M.; Saad, A.; Charfeddine, I. B.; Bilharinho de Mendonca, B.; Belgorosky, A.; Dumic, K.; Dumic, M.; Aisenberg, J.; Kandemir, N.; Alikasifoglu, A.; Ozon, A.; Gonc, N.; Cheng, T.; Kuhnle-Krahl, U.; Cappa, M.; Holterhus, P. M.; Nour, M. A.; Pacaud, D.; Holtzman, A.; Li, S.; Zaidi, M.; Yuen, T.; New, M. I., Clinical, genetic, and structural basis of congenital adrenal hyperplasia due to 11beta-hydroxylase deficiency. *Proc. Natl. Acad. Sci. U S A* **2017,** *114* (10), E1933-E1940.

243. Zachmann, M.; Tassinari, D.; Prader, A., Clinical and Biochemical Variability of Congenital Adrenal Hyperplasia Due to 11b-Hydroxylase Deficiency, A Study of 25 Patients*. *JCE & M* **1983,** *56* (2), 222-229.

244. Mancenido, D.; New, M. I., Chapter 3B - The History of Prenatal Diagnosis of Congenital Adrenal Hyperplasia. In Genetic Steroid Disorders, 2014; pp 53-62.

245. White, P. C., Steroid 11β-Hydroxylase Deficiency and Related Disorders. Genetic Steroid Disorders. Elsevier Inc., 2013.

246. Ramamoorthy, S.; Cidlowski, J. A., Corticosteroids: Mechanisms of Action in Health and Disease. *Rheum. Dis. Clin. N. Am.* **2016,** *42* (1), 15-31, vii.

247. Woodward, R. B.; Sondheimer, F.; Taub, D., THE TOTAL SYNTHESIS OF CHOLESTEROL. *J. Am. Chem. Soc.* **1951,** *73*, 3548.

248. Woodward, R. B.; Sondheimer, F.; Taub, D., THE TOTAL SYNTHESIS OF CORTISONE. *J. Am. Chem. Soc.* **1951,** *73*, 4057.

249. Johnson, W. S.; Gravestock, M. B.; McCarry, B. E., Acetylenic Bond Participation in Biogenetic-Like Olefinic Cyclizations. II. Synthesis of dl-Progesterone. *J. Am. Chem. Soc.* **1971,** *93* (17), 4332-4334.

250. Handa, S.; Pattenden, G.; Li, W.-S., A new approach to steroid ring construction based on a novel radical cascade sequence. *Chem. Commun.* **1998,** *3*, 311-312.

251. Cardente, M. A. G.; McCulloch, S.; Pattenden, G., A cascade radical-mediated macrocyclisation–transannulation approach to oestrogen steroids. *C R Acad Sci Paris, Chim/Chem* **2001,** *4*, 571-574.

252. Pattenden, G.; Gonzalez, M. A.; McCulloch, S.; Walter, A.; Woodhead, S. J., A total synthesis of estrone based on a novel cascade of radical cyclizations. *Proc. Natl. Acad. Sci. U S A* **2004,** *101* (33), 12024-9.

253. Pines, S. H., The Merck Bile Acid Cortisone Process: The Next-to-Last Word. *Org. Process Res. Dev.* **2004,** *8* (5), 708-724.

254. Gennari, C.; Carcano, M.; D'onghi, M.; Mongelli, N.; Vanotti, E.; Vulpetti, A., Taxol Semisynthesis: A Highly Enantio- and Diastereoselective Synthesis of the Side Chain and a New Method for Ester Formation at C-13 Using Thioesters. *J. Org. Chem.* **1997,** *62* (14), 4746-4755.

255. Baloglu, E.; Kingston, D. G. I., A New Semisynthesis of Paclitaxel from Baccatin III. *J. Nat. Prod.* **1999,** *62* (7), 1068-1071.

256. Marker, R. E.; Tsukamoto, T.; Turner, D. L., Sterols. C. Diosgenin1. *J. Am. Chem. Soc.* **1940,** *62*, 2525-2532.

257. Renata, H.; Zhou, Q.; Baran, P. S., Strategic Redox Relay Enables A Scalable Synthesis of Ouabagenin, A Bioactive Cardenolide. *Science* **2013,** *339 (6115)*, 59-63.

258. Renata, H.; Zhou, Q.; Dunstl, G.; Felding, J.; Merchant, R. R.; Yeh, C. H.; Baran, P. S., Development of a concise synthesis of ouabagenin and hydroxylated corticosteroid analogues. *J. Am. Chem. Soc.* **2015,** *137* (3), 1330-40.

259. Horn, E. J.; Rosen, B. R.; Chen, Y.; Tang, J.; Chen, K.; Eastgate, M. D.; Baran, P. S., Scalable and sustainable electrochemical allylic C-H oxidation. *Nature* **2016,** *533* (7601), 77-81.

260. See, Y. Y.; Herrmann, A. T.; Aihara, Y.; Baran, P. S., Scalable C-H Oxidation with Copper: Synthesis of Polyoxypregnanes. *J. Am. Chem. Soc.* **2015,** *137* (43), 13776-9.

261. Niwa, T.; Murayama, N.; Imagawa, Y.; Yamazaki, H., Regioselective hydroxylation of steroid hormones by human cytochromes P450. *Drug Metab. Rev.* **2015,** *47* (2), 89-110.

262. Kandel, S. E.; Han, L. W.; Mao, Q.; Lampe, J. N., Digging Deeper into CYP3A Testosterone Metabolism: Kinetic, Regioselectivity, and Stereoselectivity Differences between CYP3A4/5 and CYP3A7. *Drug Metab. Dispos.* **2017,** *45* (12), 1266-1275.

263. Xiao, K.; Gao, J.; Weng, S. J.; Fang, Y.; Gao, N.; Wen, Q.; Jin, H.; Qiao, H. L., CYP3A4/5 Activity Probed with Testosterone and Midazolam: Correlation between Two Substrates at the Microsomal and Enzyme Levels. *Mol. Pharm.* **2019,** *16* (1), 382-392.

264. Fessner, N. D.; Srdič, M.; Weber, H.; Schmid, C.; Schönauer, D.; Schwaneberg, U.; Glieder, A., Preparative-Scale Production of Testosterone Metabolites by Human Liver Cytochrome P450 Enzyme 3A4. *Adv. Synth. Catal.* **2020,** *362* (13), 2725-2738.

265. Ghosh, D.; Egbuta, C.; Lo, J., Testosterone complex and non-steroidal ligands of human aromatase. *J. Steroid Biochem. Mol. Biol.* **2018,** *181*, 11-19.

266. Donova, M. V.; Egorova, O. V., Microbial steroid transformations: current state and prospects. *Appl. Microbiol. Biotechnol.* **2012,** *94* (6), 1423-47.

267. Kozlowska, E.; Urbaniak, M.; Hoc, N.; Grzeszczuk, J.; Dymarska, M.; Stepien, L.; Plaskowska, E.; Kostrzewa-Suslow, E.; Janeczko, T., Cascade biotransformation of dehydroepiandrosterone (DHEA) by Beauveria species. *Sci. Rep.* **2018,** *8* (1), 13449.

268. Lamm, A. S.; Chen, A. R.; Reynolds, W. F.; Reese, P. B., Steroid hydroxylation by Whetzelinia sclerotiorum, Phanerochaete chrysosporium and Mucor plumbeus. *Steroids* **2007,** *72* (9-10), 713-22.

269. Mao, S.; Zhang, L.; Ge, Z.; Wang, X.; Li, Y.; Liu, X.; Liu, F.; Lu, F., Microbial hydroxylation of steroids by Penicillium decumbens. *J. Mol. Catal. B Enzym.* **2016,** *133*, S346-S351.

270. Yildirim, K.; Kuru, A.; Küçükbaşol, E., Microbial transformation of androstenedione by Cladosporium sphaerospermum and Ulocladium chartarum. *Biocatal. Biotransformation* **2019,** *38* (1), 7-14.

271. Peart, P. C.; McCook, K. P.; Russell, F. A.; Reynolds, W. F.; Reese, P. B., Hydroxylation of steroids by Fusarium oxysporum, Exophiala jeanselmei and Ceratocystis paradoxa. *Steroids* **2011,** *76* (12), 1317-30.

272. Schmitz, D.; Janocha, S.; Kiss, F. M.; Bernhardt, R., CYP106A2-A versatile biocatalyst with high potential for biotechnological production of selectively hydroxylated steroid and terpenoid compounds. *Biochim. Biophys. Acta Proteins Proteom.* **2018,** *1866* (1), 11-22.

273. Berg, A.; Gustafsson, J. A.; Ingelman-Sundberg, M., Characterization of a Cytochrome P-450-dependent Steroid Hydroxylase System Present in Bacillus megaterium. *J. Biol. Chem.* **1976,** *251* (9), 2831-2838.

274. Zehentgruber, D.; Hannemann, F.; Bleif, S.; Bernhardt, R.; Lutz, S., Towards preparative scale steroid hydroxylation with cytochrome P450 monooxygenase CYP106A2. *ChemBioChem* **2010,** *11* (5), 713-21.

275. Kiss, F. M.; Schmitz, D.; Zapp, J.; Dier, T. K.; Volmer, D. A.; Bernhardt, R., Comparison of CYP106A1 and CYP106A2 from Bacillus megaterium - identification of a novel 11-oxidase activity. *Appl. Microbiol. Biotechnol.* **2015,** *99* (20), 8495-514.

276. Schmitz, D.; Zapp, J.; Bernhardt, R., Steroid conversion with CYP106A2 - production of pharmaceutically interesting DHEA metabolites. *Microb. Cell. Fact.* **2014,** *13* (81), 1-13.

277. Khatri, Y.; Ringle, M.; Lisurek, M.; von Kries, J. P.; Zapp, J.; Bernhardt, R., Substrate Hunting for the Myxobacterial CYP260A1 Revealed New 1alpha-Hydroxylated Products from C-19 Steroids. *ChemBioChem* **2016,** *17* (1), 90-101.

278. Khatri, Y.; Carius, Y.; Ringle, M.; Lancaster, C. R.; Bernhardt, R., Structural characterization of CYP260A1 from Sorangium cellulosum to investigate the 1alpha-hydroxylation of a mineralocorticoid. *FEBS Lett.* **2016,** *590* (24), 4638-4648.

279. Bracco, P.; Janssen, D. B.; Schallmey, A., Selective steroid oxyfunctionalisation by CYP154C5, a bacterial cytochrome P450. *Microb. Cell. Fact.* **2013,** *12* (95), 1-11.

280. Makino, T.; Katsuyama, Y.; Otomatsu, T.; Misawa, N.; Ohnishi, Y., Regio- and Stereospecific Hydroxylation of Various Steroids at the 16 alpha Position of the D Ring by the Streptomyces griseus Cytochrome P450 CYP154C3. *Appl. Environ. Microbiol.* **2014,** *80* (4), 1371-1379.

281. Vugt-Lussenburg, B. M. A.; Stiernschantz, E.; Lastdrager, J.; Oostenbrink, C.; Vermeulen, N. P. E.; Commandeur, J. N. M., Identification of Critical Residues in Novel Drug Metabolizing Mutants of Cytochrome P450 BM3 Using Random Mutagenesis. *J. Med. Chem.* **2007,** *50* (3), 455-461.

282. Swizdor, A.; Panek, A.; Milecka-Tronina, N., Hydroxylative activity of Aspergillus niger towards androst-4-ene and androst-5-ene steroids. *Steroids* **2017,** *126*, 101-106.

283. Peart, P. C.; Reynolds, W. F.; Reese, P. B., The facile bioconversion of testosterone by alginate-immobilised filamentous fungi. *J. Mol. Catal. B Enzym.* **2013,** *95*, 70-81.

284. Koshimura, M.; Utsukihara, T.; Hara, A.; Mizobuchi, S.; Horiuchi, C. A.; Kuniyoshi, M., Hydroxylation of steroid compounds by Gelasinospora retispora. *J. Mol. Catal. B Enzym.* **2010,** *67* (1-2), 72-77.

285. D'oller, D.; Gros, E. G., Carbon-13 nuclear magnetic resonance spectral data of sterohhl vicinal ketols and related compounds. *Steroids* **1991,** *56*.

286. Marwah, P.; Marwah, A.; Lardy, H. A.; Miyamoto, H.; Chang, C., C19-steroids as androgen receptor modulators: design, discovery, and structure-activity relationship of new steroidal androgen receptor antagonists. *Bioorg. Med. Chem.* **2006,** *14* (17), 5933-47.

287. Li, A.; Acevedo-Rocha, C. G.; D'Amore, L.; Chen, J.; Peng, Y.; Garcia-Borras, M.; Gao, C.; Zhu, J.; Rickerby, H.; Osuna, S.; Zhou, J.; Reetz, M. T., Regio- and Stereoselective Steroid Hydroxylation at C7 by Cytochrome P450 Monooxygenase Mutants. *Angew. Chem. Int. Ed. Engl.* **2020**.

288. Dietrich, M.; Do, T. A.; Schmid, R. D.; Pleiss, J.; Urlacher, V. B., Altering the regioselectivity of the subterminal fatty acid hydroxylase P450 BM-3 towards gamma- and delta-positions. *J. Biotechnol.* **2009,** *139* (1), 115-7.

289. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1* (2), 19-25.

290. Petrovic, D.; Bokel, A.; Allan, M.; Urlacher, V. B.; Strodel, B., Simulation-Guided Design of Cytochrome P450 for Chemo- and Regioselective Macrocyclic Oxidation. *J. Chem. Inf. Model.* **2018,** *58* (4), 848-858.

291. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-61.

292. Jacobsen, N. E.; Kover, K. E.; Murataliev, M. B.; Feyereisen, R.; Walker, F. A., Structure and stereochemistry of products of hydroxylation of human steroid hormones by a housefly cytochrome P450 (CYP6A1). *Magn. Reson. Chem.* **2006,** *44* (4), 467-74.

293. Kollerov, V.; Shutov, A.; Kazantsev, A.; Donova, M., Biotransformation of androstenedione and androstadienedione by selected Ascomycota and Zygomycota fungal strains. *Phytochemistry* **2020,** *169*, 112160.

294. Kobayashi, M.; Mitsuhashi, H., ISOLATION AND SYNTHESIS OF 18,3a,S,68-TETRA- HYDROXY-So-ANDROSTAN-17-ONE FROM THE SOFT CORAL SARCOPHYTON GLAUCUM. *Steroids* **1982,** *40* (6), 673-677.

295. Romano, A.; Romano, D.; Ragg, E.; Costantino, F.; Lenna, R.; Gandolfi, R.; Molinari, F., Steroid hydroxylations with Botryodiplodia malorum and Colletotrichum lini. *Steroids* **2006,** *71* (6), 429-34.

296. Kaminski, R. M.; Marini, H.; Kim, W.; Rogawski, M. A., Anticonvulsant Activity of Androsterone and Etiocholanolone. *Epilepsia.* **2005,** *46* (6), 819-827.

297. Zolkowska, D.; Dhir, A.; Krishnan, K.; Covey, D. F.; Rogawski, M. A., Anticonvulsant potencies of the enantiomers of the neurosteroids androsterone and etiocholanolone exceed those of the natural forms. *Psychopharmacology (Berl)* **2014,** *231* (17), 3325-32.

298. Niwa, T.; Imamura, H.; Katagiri, M., Inhibition of human steroidogenic cytochrome P450 c17 by 21-hydroxypregnenolone and related steroid hormones. *Biol. Pharm. Bull.* **2012,** *35* (9), 1594-1597.

299. Niwa, T.; Yabusaki, Y.; Honma, K.; Matsuo, N.; Tatsuta, K.; Ishibashi, F.; Katagiri, M., Contribution of human hepatic cytochrome P450 isoforms to regioselective hydroxylation of steroid hormones. *Xenobiotica* **1998,** *28* (6), 539-47.

300. Kolek, T.; Milecka, N.; Swizdor, A.; Panek, A.; Bialonska, A., Hydroxylation of DHEA, androstenediol and epiandrosterone by Mortierella isabellina AM212. Evidence indicating that both constitutive and inducible hydroxylases catalyze 7alpha- as well as 7beta-hydroxylations of 5-ene substrates. *Org. Biomol. Chem.* **2011,** *9* (15), 5414-22.

301. Al-Aboudi, A.; Mohammad, M. Y.; Musharraf, S. G.; Choudhary, M. I.; Atta ur, R., Microbial transformation of testosterone by Rhizopus stolonifer and Fusarium lini. *Nat. Prod. Res.* **2008,** *22* (17), 1498-509.

302. Hunter, A. C.; Mills, P. W.; Dedi, C.; Dodd, H. T., Predominant allylic hydroxylation at carbons 6 and 7 of 4 and 5-ene functionalized steroids by the thermophilic fungus Rhizomucor tauricus IMI23312. *J. Steroid Biochem. Mol. Biol.* **2008,** *108* (1-2), 155-63.

303. Swizdor, A.; Kolek, T.; Panek, A.; Bialonska, A., Microbial Baeyer-Villiger oxidation of steroidal ketones using Beauveria bassiana: Presence of an 11alpha-hydroxyl group essential to generation of D-homo lactones. *Biochim. Biophys. Acta* **2011,** *1811* (4), 253-62.

304. El-Kadi, I.; Mostafa, M. E., Hydroxylation of progesterone by some Trichoderma species. *Folia. Microbiol.* **2004,** *49* (3), 285-290.

305. Kolar, N. W.; Swart, A. C.; Mason, J. I.; Swart, P., Functional expression and characterisation of human cytochrome P45017alpha in Pichia pastoris. *J. Biotechnol.* **2007,** *129* (4), 635-44.

306. Sagadin, T.; Riehm, J. L.; Milhim, M.; Hutter, M. C.; Bernhardt, R., Binding modes of CYP106A2 redox partners determine differences in progesterone hydroxylation product patterns. *Commun. Biol.* **2018,** *1* (99), 1-9.

307. Khatri, Y.; Jozwik, I. K.; Ringle, M.; Ionescu, I. A.; Litzenburger, M.; Hutter, M. C.; Thunnissen, A. W. H.; Bernhardt, R., Structure-Based Engineering of Steroidogenic CYP260A1 for Stereo- and Regioselective Hydroxylation of Progesterone. *ACS Chem. Biol.* **2018,** *13* (4), 1021-1028.

308. Mappus, E.; Chambon, C.; Ravel, M. R.; Grenot, C.; Cuilleron, C. Y., Synthesis and characterization by 1H and 13C nuclear magnetic resonance spectroscopy of 17a-cyano, 17a-aminomethyl, and 17a-alkylamidomethyl derivatives of 5a-dihydrotestosterone and testosterone. *Steroids* **1997,** *62*, 603-620.

309. Crabb, T.; Saul, J. A.; Williams, R. O., Microbiological Transformations. Part 4. Microbiological Transformations of 5a-Androstan-17-ones and of 17a-Aza-D-homo-5a-androstan-17-ones with the Fungus Cunnjnghamella elegans. *J. C. S. Perkin I* **1981**, 1041-1045.

310. Yildirim, K.; Kuru, A., Microbial hydroxylation of epiandrosterone by Aspergillus candidus. *Biocatal. Biotransformation* **2017,** *35* (2), 120-126.

311. Matsui, M.; Hakozaki, M., DISULPHATES OF 16-OXYGENATED KETONIC C19 STEROIDS AS BILIARY METABOLITES OF ANDROSTERONE SULPHATE IN FEMALE RATS-16b-AN. *Steroids* **1978,** *31* (2), 219-226.

312. Numazawa, M.; Nagaoka, M.; Mutsumi, A., Stereospecific 1,2-Hydride Shift in the Rearrangement of 16b-Hydroxy-17-oxo Steroids to 17b-Hydroxy-16-ones with Acid and Base. *Chem. Pharm. Bull.* **1987,** *35* (12), 4763-4768.

313. Kirk, D. N.; Toms, H. C.; C., D.; White, K. A.; Smith, K. E.; Latif, S.; Hubbard, R. W. P., A Survey of the High-field 1H NMR Spectra of the Steroid Hormones, their Hydroxylated Derivatives, and Related Compounds. *J. Chem. Soc., Perkin Trans.* **1990**, 1567-1594.

314. Reeder, A. Y.; Joannou, G. E., 15b-Hydroxysteroids (Part I).* Steroids of the human perinatal period: The synthesis of 3b,15b,17a-trihydroxy-5-pregnen-20-one *Steroids* **1996,** *61*, 74-81.

315. Jankowski, K., NMR Spectra of some Cyanosteroids. *Steroids* **1972,** *19* (2), 189-196.

316. Solleder, C.; Schauber, T.; Homoki, J.; Wudy, S. A., Synthesis of deuterium-labelled 17a-hydroxypregenolone for use as internal standard in stable isotope dilution/mass spectrometry. *J. Label. Compd. Radiopharm.* **1998**, 557-565.

317. Gupta, V., Mineralocorticoid hypertension. *Indian J. Endocrinol. Metab.* **2011,** *15* S298-312.

318. Rooyen, D.; Gent, R.; Barnard, L.; Swart, A. C., The in vitro metabolism of 11beta-hydroxyprogesterone and 11-ketoprogesterone to 11-ketodihydrotestosterone in the backdoor pathway. *J. Steroid Biochem. Mol. Biol.* **2018,** *178*, 203-212.