Enantioselective remote functionalisation of cyclic amines by engineered $P450_{BM3}$ variants

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

Yuan Zhang

Merton College

Trinity 2022

Enantioselective remote functionalisation of cyclic amines by engineered P450$_{BM3}$ variants

Yuan Zhang
Merton College

D Phil. Thesis abstract

Trinity, 2022

P450 enzymes activate C–H bonds via the insertion of an oxygen atom from atmospheric dioxygen under ambient conditions in water and accept a broad array of endogenous and exogenous organic substrates. P450$_{BM3}$ (CYP102A1) has a unique structure, with the electron-transfer reductase domain fused to the haem monooxygenase domain. This makes P450$_{BM3}$ catalytically self-sufficient and promotes high turnover rates. Functionalised cyclic amines are versatile intermediates in drug synthesis and important fragment molecules in drug discovery. Selective oxidation of unactivated ring aliphatic C–H bonds is an attractive route to cyclic amine derivatives; however, classical synthetic methods are limited in scope and lack enantioselectivity. The application of engineered P450 enzymes for selective functionalisation of cyclic amines is surprisingly under-explored.

Chapter 2 describes the substrate and enzyme engineering to access remote hydroxylation products of the medium-sized cyclic amines azepane and azocane. Variant library screening is deployed to identify key residues affecting product selectivity, substrate engineering is utilised to alter the selectivity patterns, and rational design is conducted to improve product selectivity. In Chapter 3, the iterative docking-guided mutagenesis (IDGM) approach is explored. Comparative analysis of the correct binding poses and unwanted poses leads to targeted mutation design to disfavour unwanted poses and increase both the activity and product selectivity of amine oxidation. In Chapter 4, the combined enzyme evolution strategies are used for more complex cyclic amines – the bicyclic, 7-azabicyclo[2.2.1]heptane, and spirocyclic, 8-azaspiro[5.4]decane. These approaches led to scalable enantioselective oxidation of all but one of the unactivated C–H bonds in these important cores of drugs and fragment libraries.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Luet Wong for all the support throughout my DPhil study. Your patience, kindness and encyclopedic knowledge have guided me through the hard battles in the research and escorted me to grow up as a scientific researcher. I would also like to thank Professor Jeremy Robertson for all the knowledgeable advice and conversations, especially those about hard-core stereochemistry. I could not possibly imagine accomplishing the thesis without any of you.

I would like to thank the generous financial support from China Scholarship Council and University of Oxford. I could not have been able to grasp this great opportunity to study in Oxford without the help from the CSC-Oxford Scholarship.

More importantly, thank you to Wenyu Chen, my beloved partner in life and in work. Thank you for all your unrequited and selfless support. You have encouraged me and raised me up through all the dark and tough times.

Thank you to the senior DPhil students in the group; Yang Cao, for your great help when I settled down in Oxford. You are like a big brother. Yushu Li, for all the informative ideas about science and life you shared. Matthew Fisher, for all the technical help from computational side, which made everyone's life easier.

Thank you to my fellow students; Xinxin Zhang, for all the cherish memories of World Cup, board games, comedy shows and cooking, playing and laughing together. Yuxuan Lu, for all the great times cooking and chatting together. Hope Lumi grow up soundly and strongly. Alice Wong, for always being considerate to everyone. It has been great to work with you. Vicky Barros, for your energies that always cheer me up. Alethea Lee, for all the great desserts and cakes you made.

Special thanks to all the Part IIs I have worked with. Jamie Lynch, Ruby Smith, Yi Lu, Claudia Bustamante, Alissa Bamford, Sophie Cook, Aaron Leung, Nelson, Ellie Jaques and Noah. You have made our group more like a family. I am so lucky to be able to work with you.

Finally, I would like to thank my family; my father Heguo Zhang, my mother Xiaoli Lian and my sister Yao Zhang. Words have not remotely been able to express my feelings. Thank you for all the love and support.

Table of contents

List of Abbrevations

ADH	alcohol dehydrogenase
AdR	adrenoxin reductase
Adx	adrenodoxin
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
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
ET	electron-transfer
ep-PCR	error-prone PCR
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

HMBC	heteronuclear multiple-bond correlation spectroscopy
HSQC	heteronuclear single quantum correlation spectroscopy
IDGM	iterative docking guided mutagenesis
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
PFOR	phthalate dioxygenase reductase
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
WT	wild type

Chapter 1

1. Introduction

1.1 Cyclic amines and their C–H functionalisation

1.1.1 Cyclic amines and applications

Cyclic amines are among the most important core structures in bioactive natural and synthetic compounds which have numerous applications in the chemical, agricultural, and pharmaceutical industry (Fig. 1.1-A).[1-11] The most notable utilisation of these compounds is in the medical and pharmaceutical sector, with ~60% of U.S. FDA-approved drugs containing nitrogen heterocycles.[7] In most of these cases, as exemplified in Figure 1.1-B, the cyclic amine cores are functionalised at sp^3 C–H bonds, especially at unreactive positions.

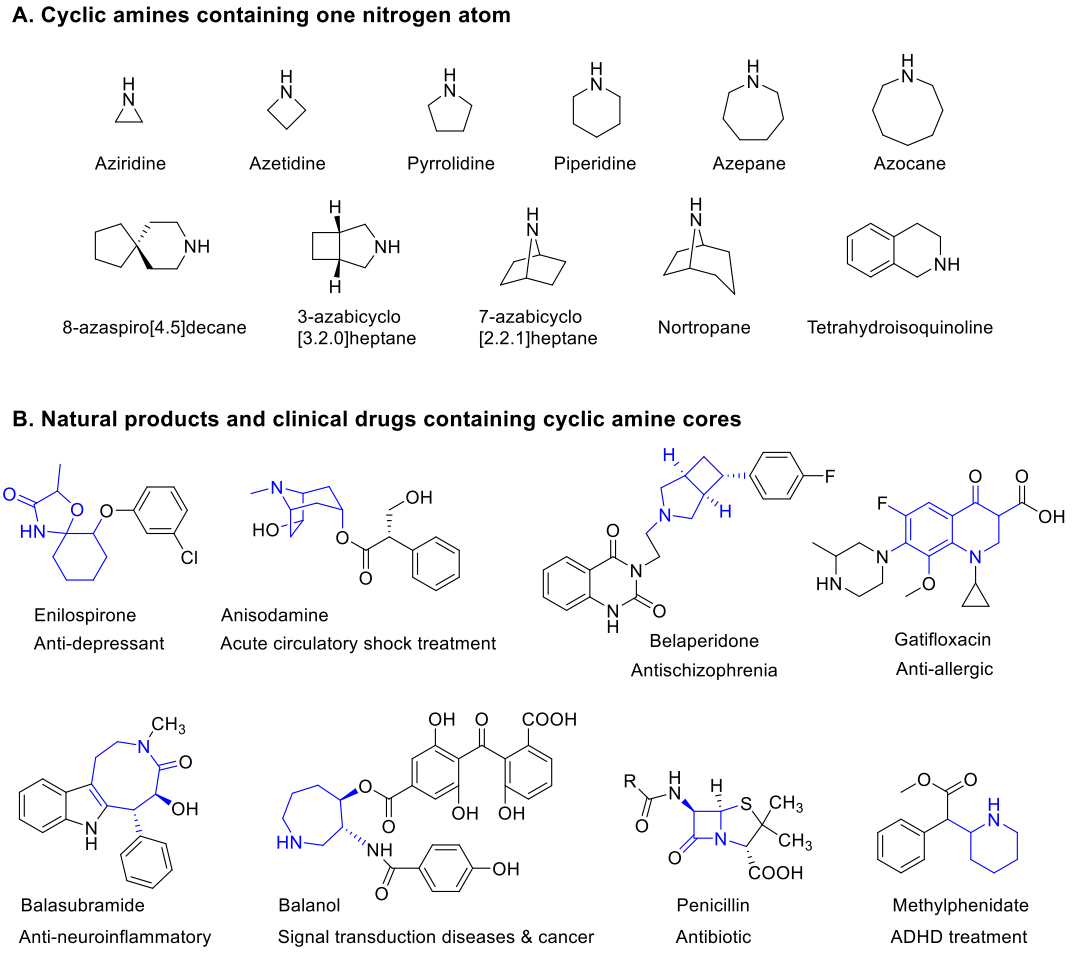


Figure 1.1 Examples of **(A)** saturated cyclic amines, and **(B)** natural products and clinical drugs containing cyclic amine cores.

Nitrogen heterocycle motifs are also heavily used in fragment-based drug discovery (FBDD), a powerful approach that is increasingly applied for identifying ligands for previously intractable drug targets and discovering new biologically active compounds.[2,12-19] The FBDD approach starts with the screening of biological targets against a library (typically <5000 compounds) of small molecules (molecular weight <350 Da), called fragments, to identify potential starting points. These initial hits generally have a weak binding affinity to their target (IC_{50} in mM – μM range). They are then modified further and linked together, usually guided by structural and computational methods, resulting in a "lead" for drugs with significantly improved potency and binding affinity (IC_{50} in the nM range) (Fig. 1.2). Compared with the more traditional high throughput screening (HTS) approach, FBDD requires less screening effort and typically results in drug leads with higher ligand affinity.

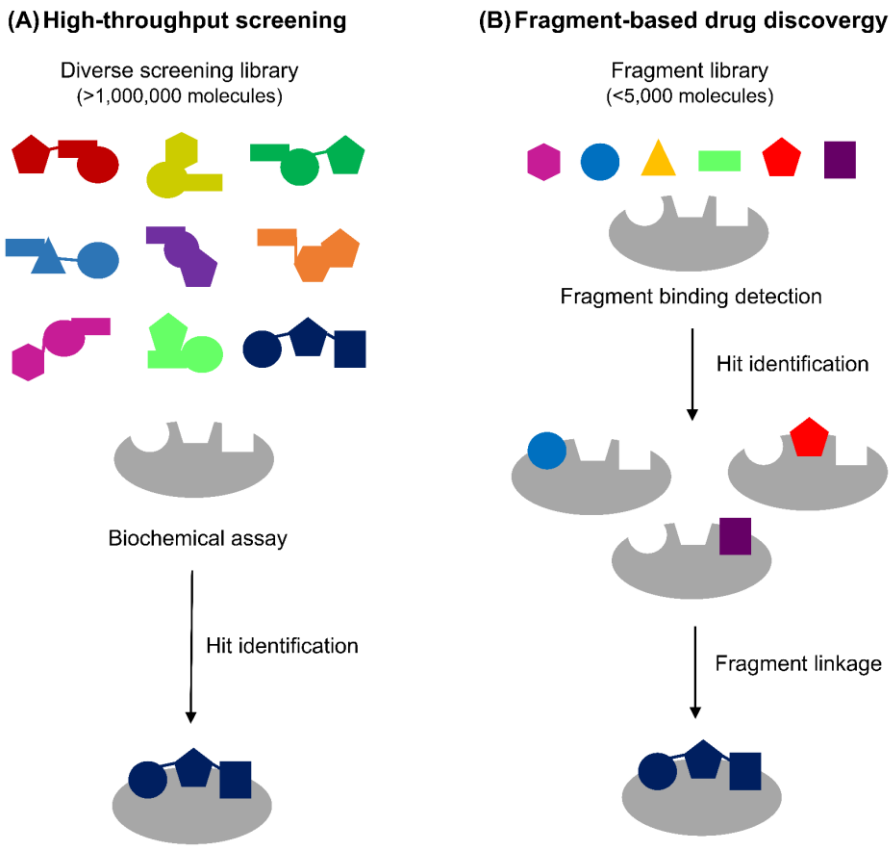


Figure 1.2 Schematic illustrations of processes of **(A)** high-throughput screening (HTS) and **(B)** fragment-based drug discovery (FBDD).[19]

Figure 1.4 shows some examples of drug leads discovered by FBDD. In these cases, fragment hits with low binding affinities (IC_{50} = 1–7100 μM) were identified from fragment libraries containing 3000–5000 compounds. These hits were merged and optimised to afford potent drug leads that displayed significant inhibition for their relevant biological targets.[20] It is worth noting that the cyclic amine cores used in these examples are mostly functionalised on the ring, often at the remote positions, as with those in the drugs listed in Figure 1.3. Therefore, the sp^3 C–H functionalisation of cyclic amines, especially with regio- and enantio-selectivity, is of great importance.

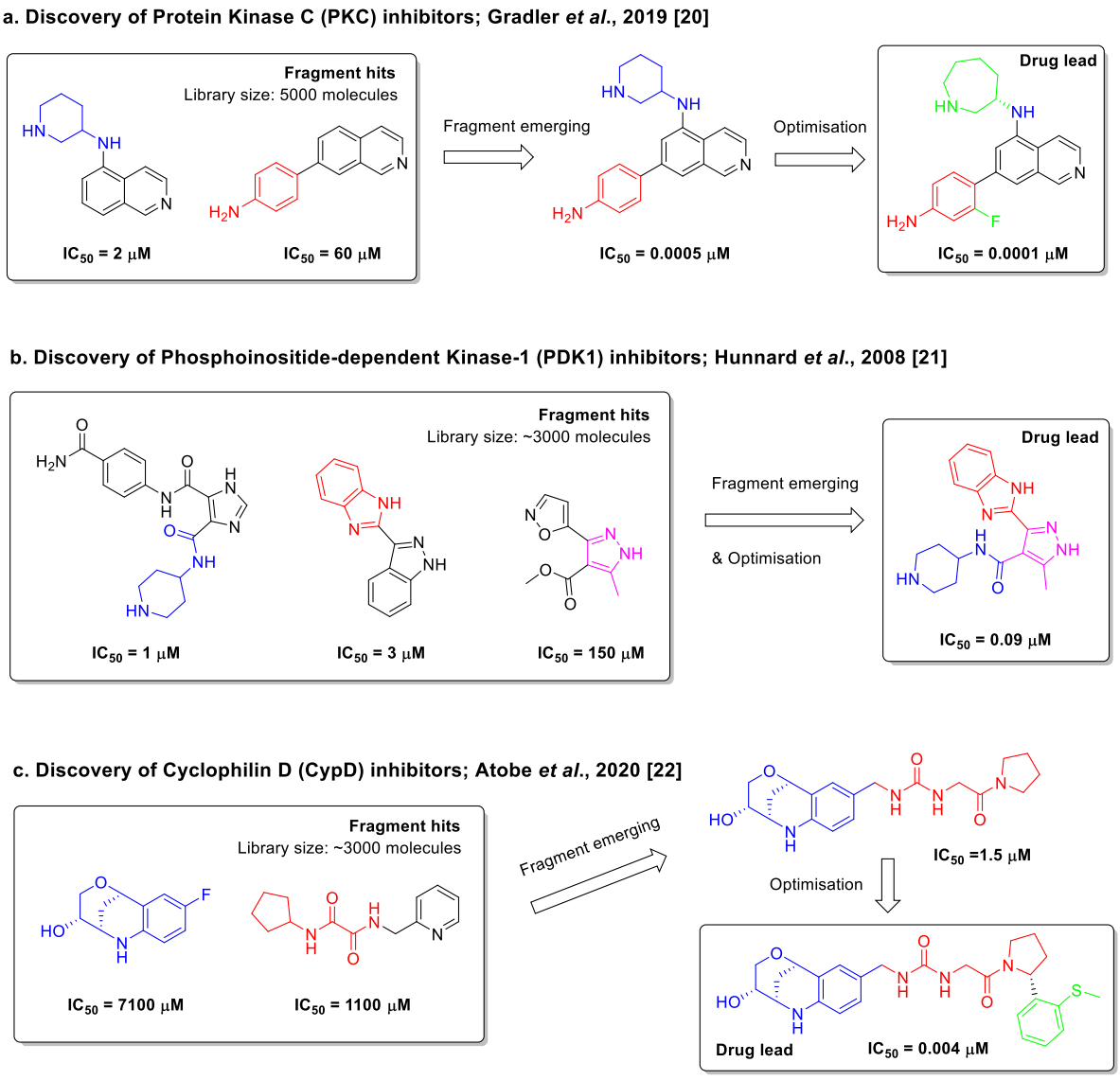


Figure 1.3 Examples of drug lead discovery by FBDD approaches.[20-22]

1.1.2 C–H functionalisation of cyclic amines: synthetic approaches

Unlike their aromatic counterparts, there are still limited strategies for facile sp^3 substitution of cyclic amines and, to date, the vast majority of such methods focus on the functionalisation of the C–H bond adjacent to the nitrogen (the α-C–H bond).[3,23,24] Traditionally, α-C–H functionalisation relies on multi-step synthesis from acyclic precursors.[23-25] Since the pioneering work on lithiation-based α-functionalisation of cyclic amines by Beak and Lee in 1989,[26] direct sp^3 α-C–H functionalisation has drawn significant attention and many more approaches are established. These methodologies can be mechanistically divided into four groups: (a) α-lithiation with alkylithium/diamine complexes which produces a dipole-stabilised carbanion, followed by reaction with electrophiles to provide α-substituted cyclic amines;[26-28] (b) radical-based C–H functionalisation which includes generation of α-amino radicals through 1,5-hydrogen atom transfer (HAT),[29] BEt_3/O_2 or light in the presence of a photosentitiser,[30,31] metal-catalysed C–C bond formation,[32] or photoredox α-amino C–H arylation;[33] (c) redox-neutral C–H functionalisation in which an electrophile/nucleophile pair is generated by chemical ,[34] thermal,[35] electrochemical,[36] photocatalytic,[37] or metal-catalysed oxidation (Cu,[38] Fe,[39] Ru,[40] *etc*.); (d) transition metal-catalysed direct C–H activation with Ta,[41] Ir,[42] Ru,[43] Rh,[44] and with the presence of appropriate directing groups (Fig. 1.4). Most of these approaches require the relevant cyclic amines to be *N*-protected. Chen *et al*.[3] reported a hydride-transfer based method to achieve α-functionalisation of unprotected cyclic amines. These approaches are exemplified by the syntheses of clinical drugs **E1** (a glucokinase activator; Fig. 1.4-a),[45,46] **E2** (kainic acid, an anti-worming agent; Fig. 1.4-b),[31] **E3** (lepadiformine, an anti-arrhythimic; Fig. 1.4-c),[38] and **E4** (mitomycin, used in urothelial cancer treatment; Fig. 1.4-d).[47]

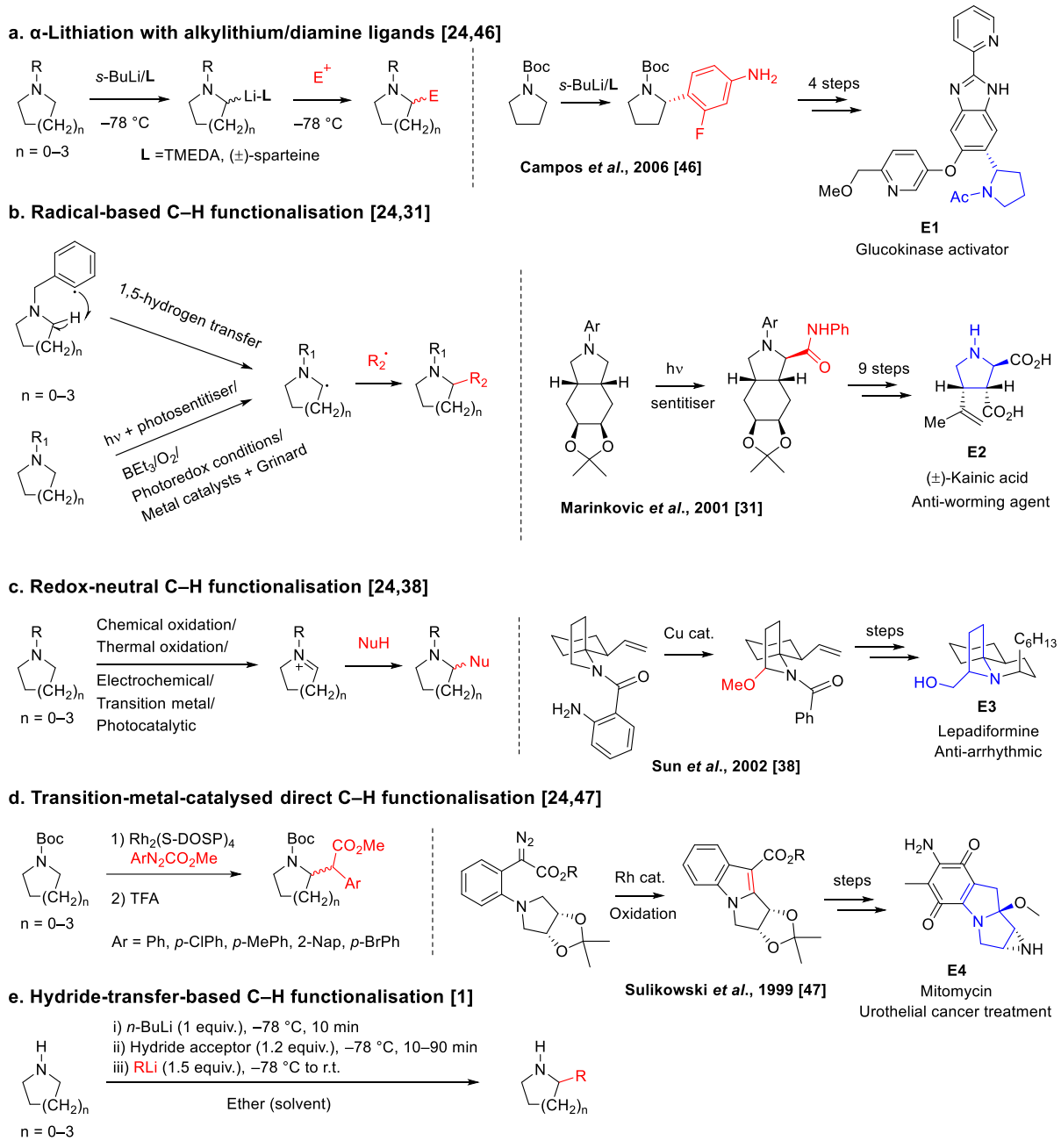


Figure 1.4 Examples of synthetic approaches to prepare α-functionalised cyclic amines and their applications in drug synthesis.[1,24,31,38,46,47]

Direct β- and γ-C–H functionalisation are also reported.[1,4,5,23,48-52] These typically work with protected amines and require transition metals as catalysts, *e.g.*, the undirected borylation approach using an iridium catalyst by Oeschger *et al.*[51] (Fig. 1.5-a) and the palladium-catalysed transannular C–H functionalisation approach by Topczewski *et al.*[4] (Fig. 1.5-b). The applications of these methods are demonstrated by access of drugs **E5** (for asthma treatment),[1]

and **E6** (for schizophrenia treatment).[4] Recently, Chen *et al.*[1] developed a *N*-lithiation based method that furnish β-substituted free cyclic amines and without transition metal catalysts (Fig. 1.5-c). In this method, the *N*-lithiated amines are oxidised by relevant ketones to generate the transient imines, which are subsequently converted to the key intermediate, an endocyclic 1-azaallyl anion. This anionic species is then reacted with an alkylating reagent, followed by reductive work-up to afford the β-functionalised amines.

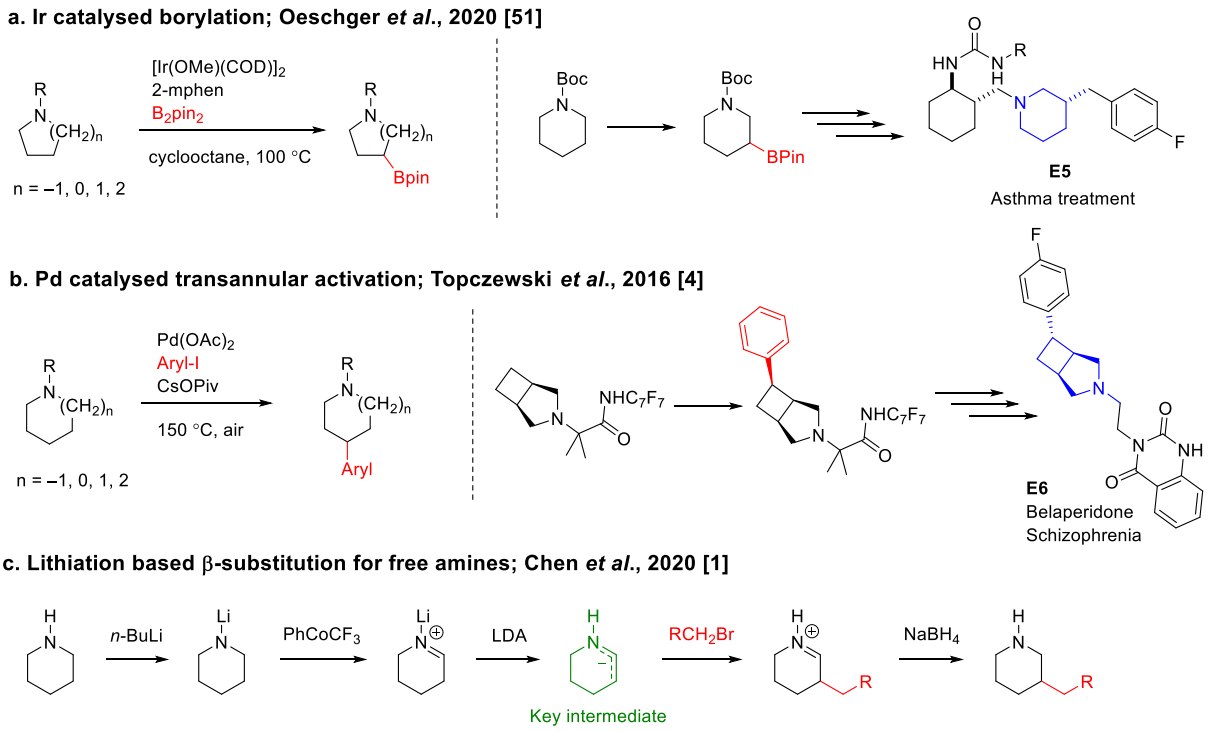


Figure 1.5 Examples of synthetic approaches to prepare β- and γ-functionalised cyclic amines and their applications in drug synthesis.[1,4,51]

Current synthetic β- and γ-C–H activation approaches still face limitations, of which the most challenging is stereochemical control of the substitution.[1,2,4,51] Therefore, the highly desirable selective remote C–H functionalisation of cyclic amines remains a significant challenge to the synthetic and pharmaceutic communities. Biocatalysts, such as cytochrome P450 enzymes, which catalyse C–H oxy-functionalisation under ambient conditions, potentially with high regio- and enantio-selectivity by enzyme engineering, may provide viable alternative approaches.

1.2 Biocatalysis and P450 enzymes

1.2.1 Development of biocatalysis

Biocatalysis is the application and engineering of enzymes or microbes in synthetic chemistry and uses nature's catalysts for new purposes.[53-63] More than a century ago, Rosenthaler's ground-breaking work on synthesis of (*R*)-mandelonitrile from benzaldehyde and hydrogen cyanide using almond emulsin opened the door to biocatalysis as a new research field.[64] Since then, there have been two revolutionary changes in this field, aided with key advances in biochemistry and biotechnology, *e.g.*, the establishments of X-ray crystallography for protein structures, DNA recombination, DNA sequencing, site directed mutagenesis and polymerase chain reaction (PCR) techniques. In pioneering work in the mid-1980s, Estell *et al*.[65] reported improvements in the oxidative stability of subtilisin, a protease, by site directed mutagenesis at position M222. This put forward the application of enzyme engineering in industrial biotechnology. The second revolution emerged in mid-1990s as Pim Stemmer *et al*.[66] and Frances Arnold *et al*.[67] discovered and utilised molecular biology methods that rapidly and extensively modify biocatalysts through an *in vivo* process that mimic Darwinian evolution (Fig. 1.6).

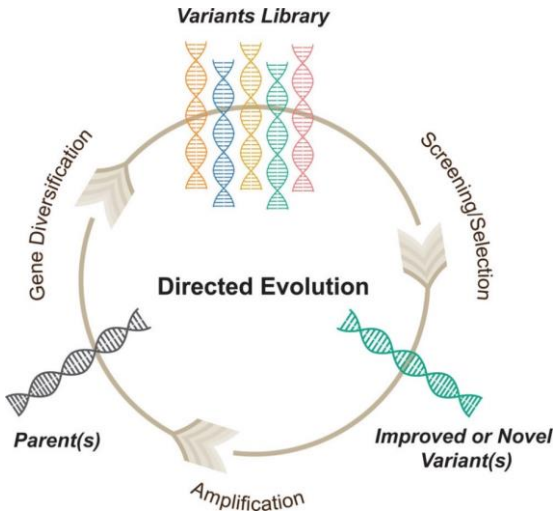


Figure 1.6 Typical process of direct evolution.[55]

Known as “directed evolution”, this approach involves iterative cycles consisting of two main steps: (1) gene diversification by random mutagenesis or DNA shuffling to generate a variant library, and (2) screening and selection to identify variants with improved enzyme stability, substrate specificity and product selectivity, which could also serve as templates for further rounds of evolution (Fig. 1.6).[53-55,59,61] There have been enormous successful applications of enzyme engineering in the synthesis of chiral building blocks, clinical drugs and natural products, and constructing novel synthetic pathways.[53-63,68] Examples include the industrial production of glucose from saccharification of starch by glucose isomerase,[69] the synthesis of β-lactam antibiotics such as amoxicillin using penicillin G amidase,[70] the production of the antimalarial drug precursor artemisinic acid by engineered yeast,[71] and the total synthesis of the Vitamin B12 precursor hydrogenobyrinic acid **E7** using 12 enzymes in one pot,[72] among thousands of other applications (Fig. 1.7).

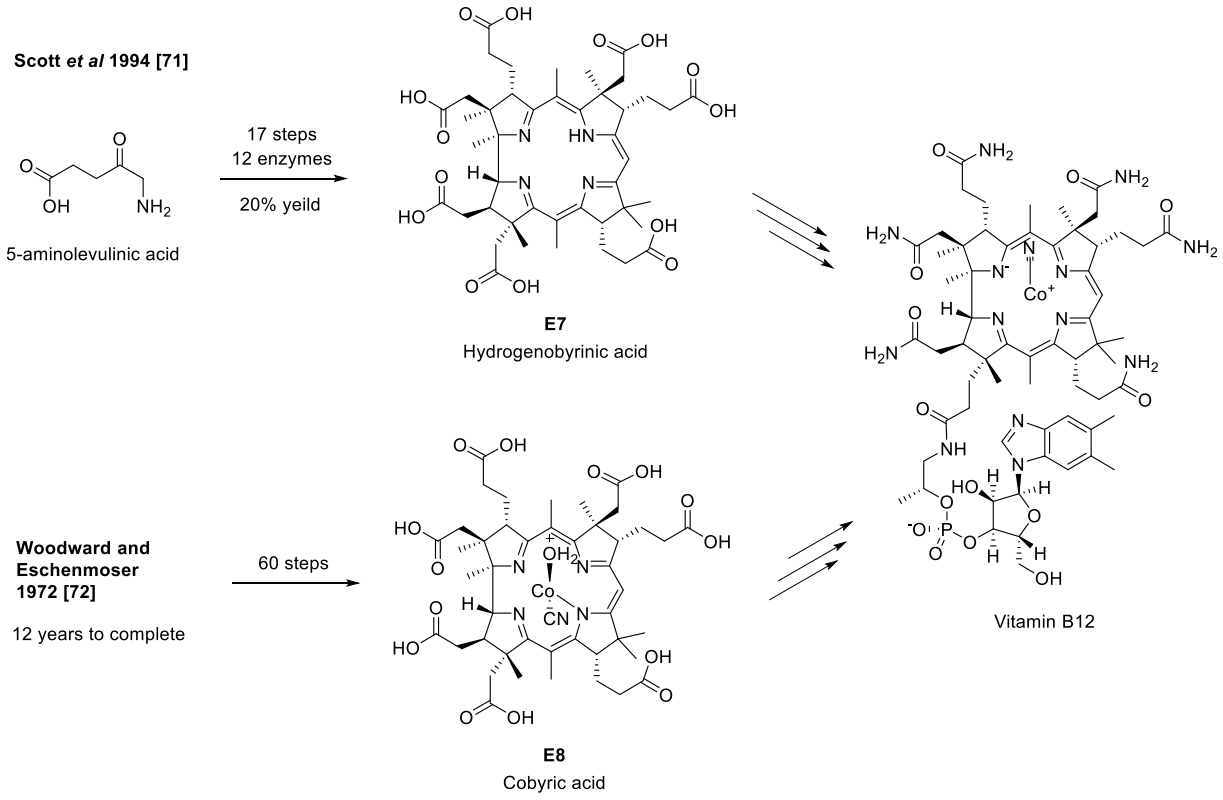


Figure 1.7 Biocatalytic synthesis of Vitamin B12 precursor **E7** and the Woodward-Eschenmoser synthesis of **E8**.

The biosynthesis of hydrogenobyrinic acid **E7** by Scott *et al*.[72] is widely considered one of the most impressive total syntheses using biocatalysts ever achieved (Fig. 1.7). Starting from 5-aminolevulinic acid, the 17 step-route utilising 12 enzymes gave 20% overall yield. In comparison, the entirely organic synthetic pathway of the Vitamin B12 precursor cobyric acid **E8** by Woodward and Eschenmoser took 60 steps and 12 years to complete (Fig. 1.7).[73] Although the direct comparison between these two approaches might be futile as they have different goals, this demonstrated the great potential of biocatalysis.

1.2.2 P450 enzymes

Among the numerous enzymes utilised in biocatalysis research, cytochrome P450 enzymes (P450s) are considered some of the most versatile biocatalysts because of the wide range of substrates accepted and reactions catalysed. P450 enzymes are a superfamily of monooxygenases broadly distributed among living organisms and play crucial roles in xenobiotic degradation, drug metabolism and natural product biosynthesis.[59,74-78] The first P450 enzyme to be discovered was a pigment (hence the "P" in the name) found in rat liver microsomes.[79] Omura *et al*.[80,81] later reported that the CO-bound, reduced form of that pigment exhibited a characteristic absorbance at 450 nm, which donated the "450" in the name P450 and today is still widely used to measure the concentration of P450 enzyme solution. To date, there are >50,000 P450 genes published (data from Uniprot), which are found across the kingdoms of life (human, animals, plants, micro-organisms, and even some viruses), demonstrating their incredible diversity.[62,74,78] P450 enzymes accept a broad array of endogenous and exogenous substrates covering almost all classes of organic molecules in nature: fatty acid, steroids, terpenes, terpenoids, cyclic amines, drugs, pesticides, carcinogens and organic solvents.[74,76] There are more than 20 different types of chemical oxidation reactions P450s can catalyse,

including hydroxylation, epoxidation, decarboxylation, *N*- and *O*-dealkylation, C–C bond coupling and cleavage, to name a few.[59,62,74,76] Many unusual reactions are also discovered to be of P450s' capacity, such as nitration,[82] C–N–C animation,[83] ether ring formation,[84] and Baeyer-Villiger oxidation,[85] as exemplified in Figure 1.9. Some of these reactions have been directly applied to synthesis of drugs of natural products, such as thaxtomin A (**E9**; Fig. 1.8-a), (+)-aureothin (**E10**; Fig. 1.8-c), and brassinolide (**E11**, Fig. 1.8-d).

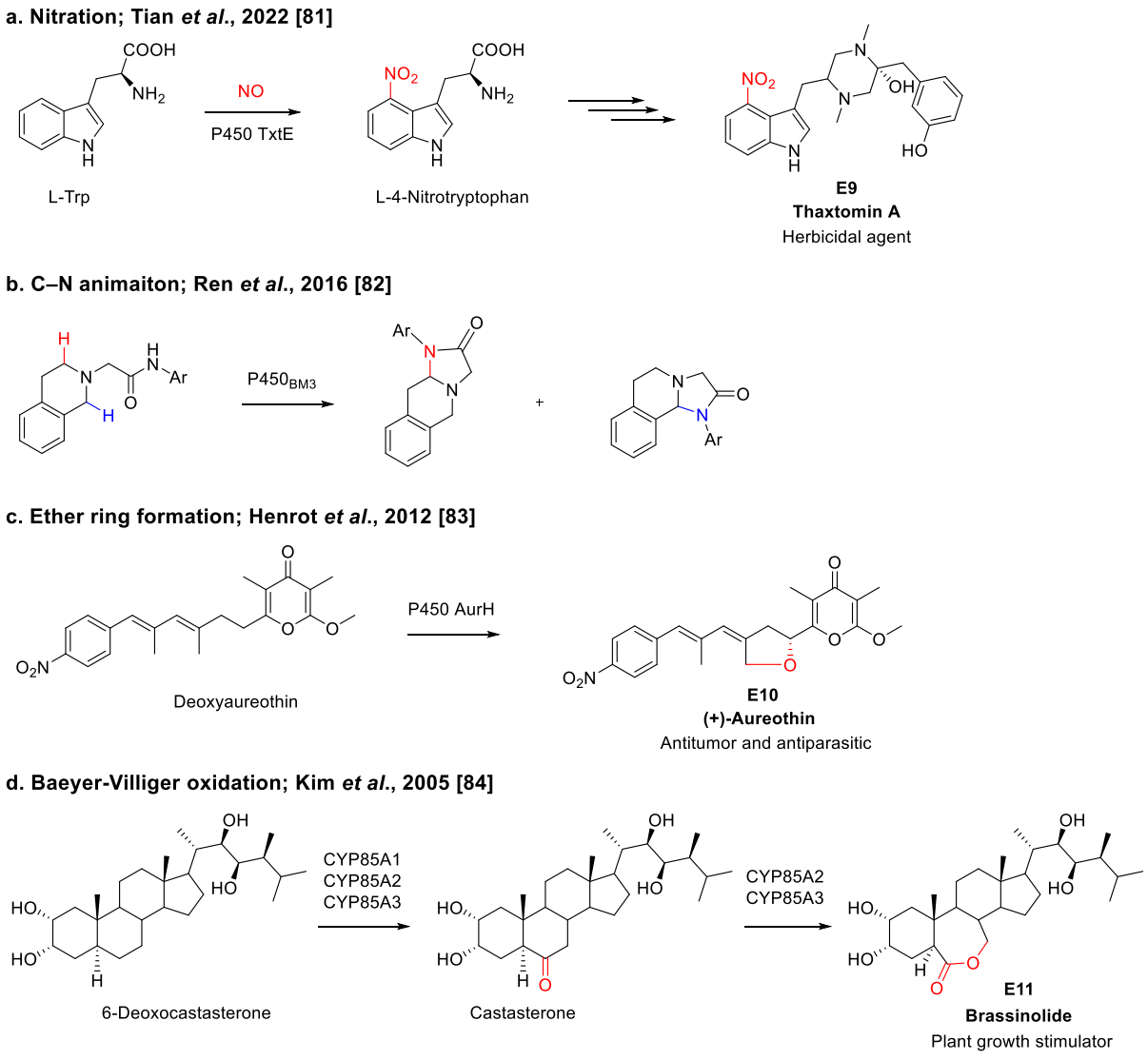


Figure 1.8 Examples of unusual reactions that P450 enzymes can catalyse.

1.2.3 The electron transfer chain of P450 enzymes

Among diverse functionalities, the most important is that P450 enzymes are capable of catalysing enantioselective oxidation of chemically inert C–H bonds using atmospheric oxygen under mild conditions.[62,74-76,86,87] The activation of molecular oxygen requires two electrons, which typically are derived from the pyridine nucleotide cofactors NADH or NADPH, since most P450s are external monooxygenases. Therefore, P450s rely on redox partner proteins or domains to transfer electrons from NAD(P)H to the haem domain.

$$\mathbf{R–H + O_2 + NAD(P)H + H^+ \rightarrow R–OH + H_2O + NAD(P)^+}$$

Depending on the types of redox partners, P450 enzymes were firstly grouped into two major classes (I and II) (Fig. 1.9).[87] Class I systems contain most bacterial as well as mitochondrial P450s from eukaryotes. Both groups are common in the composition of three separate proteins: a redox iron-sulfur ferredoxin (Fdx), a ferredoxin reductase (FdR) and the P450 enzyme. FdR are flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) containing enzymes that accept the hydride from NADH or NADPH and transfer the two electrons, one at a time, to the ferredoxin, which then passes the electron to the P450 enzyme. In bacteria, all three proteins are soluble. The prototype bacterial class I system is P450cam (CYP101) from *Pseudomonas putida*, which catalyses the 5-*exo* hydroxylation step in the breakdown of D-camphor.[87-89] In eukaryotes, the most well-known redox partners for mitochondrial P450s are adrenodoxin reductase (AdR) and adrenodoxin (Adx). The major difference to their bacterial counterparts is that AdR and P450 enzyme are both membrane-associated, whereas Adx is soluble. The classic example of class I mitochondrial P450 is P450scc (CYP11A1), which catalyses the cleavage of the cholesterol side-chain leading to the production of pregnenolone, the rate-limiting step of steroid biosynthesis.[90,91]

Class II P450s adopt a different system where the electrons from NAD(P)H are picked up by the cytochrome P450 reductase (CPR) containing FAD and FMN domains and transferred to the P450 enzyme (Fig. 1.9).[92-94] They are the most common type in eukaryotes and perform extremely diverse catalytic reactions, including oxidative metabolism of fatty acids, steroids, drugs and toxicants, *etc*. in mammals,[74,95] secondary metabolism and hormone biosynthesis in plants,[96,97] and membrane sterol synthesis and phytoalexin detoxification in fungi.[86]

In 2002, a novel class of P450 systems was identified (class III),[87,98] which share the three-component characteristic with class I enzymes but electrons derived from NAD(P)H are transferred from the FAD-containing ferredoxin reductase to an FMN-dependent redox protein, instead of the iron-sulfur cluster Fdx in class I, and then to the P450 enzyme. Class III also resembles class II concerning the use of FAD and FMN redox centres, but the FAD and FMN centres are in separate proteins for class III. P450cin (CYP176A1) from *Citrobacter braakii* is the prototype of class III bacterial systems.[87,98] Since the discovery of P450cin, as the number of characterised P450s increase, several more novel systems with "unusual" electron transfer chains have been found and described that could not be classified as class I nor II.[74,78,87] Hannemann *et al*.[87] summarised eight such systems (class III–X).

The first class IV P450 was also the first thermophilic cytochrome P450 which constitutes the redox chain using a ferredoxin and its partner, 2-oxo-acid:ferredoxin oxidoreductase (OFOR) (Fig. 1.9). This is also the first example that does not obtain electrons from NAD(P)H-dependent flavoprotein; instead, the reducing equivalents are provided by pyruvate and CoA.[99] Being the prototype of the class, CYP119, identified in the thermophilic archaeon *Sulfolobus solfataricus*, has been studied extensively due to its extreme tolerance to high temperature (T_M = 91°C) and pressure (up to 200 MPa).[100,101]

Class V resembles class I in that it principally uses the same redox centres, an FAD-containing NAD(P)H-dependent ferredoxin reductase, an iron-sulfur ferredoxin the P450 enzyme.

However, class V P450s have the ferredoxin and the P450 enzyme fused and are thus a two-component system. The most notable P450 in this class is the sterol 14α-demethylase CYP51 from *Methylococcus capsulatus*.[102]

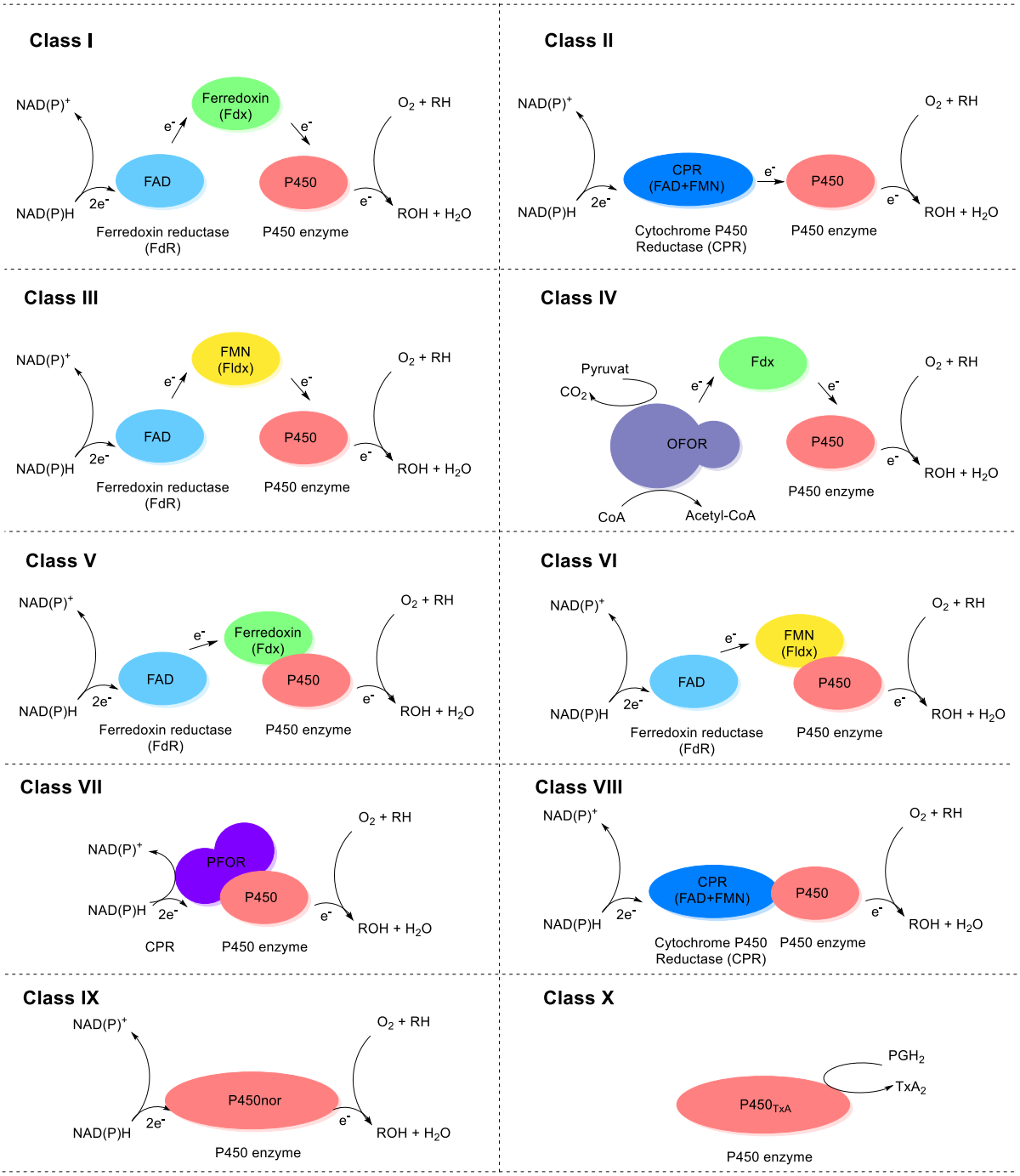


Figure 1.9 The ten classes of P450 redox proteins in electron transfer chain.[87]

Similarly, Class VI P450s are also composed of two components–an FAD-containing reductase domain and an FMN-containing ferredoxin-P450-fusion protein. The most representative example of this class is XplA (CYP177A1) from *Rhodococcus rhodochrous*, which is responsible for the breakdown of the explosive hexahydro-1,3,5-trinitro-1,3,5-triazine.[103]

Classes VII and VIII include fused enzymes in quite unique structural organisations. $P450_{RhF}$ (CYP116B2) from *Rhodococcus* is the first reported Classes VII system where the N-terminal P450 enzyme and the C-terminal reductase are linked by a shorter linker region of 16 amino acids. The reductase part contains three functional parts: an FMN-binding domain, an NADH-binding domain and a [2Fe–2S] ferredoxin.[104,105] Class VIII systems, on the other hand, contain a class II-like diflavin reductase (CPR) which is fused with the P450 enzyme, resulting in high catalytic self-sufficiency. The most well-studied example in this class, as well as among all P450 enzymes, is $P450_{BM3}$ (CYP102A1) from *Bacillus megaterium*, whose reductase domain (CPR) is composed of one equivalent each of the FAD and FMN cofactors.[74,76,106,107] This arrangement of reductase is related to mammalian P450 enzymes.

Class IX, as a special case of P450 enzymes, mainly describes the nitric oxide reductase, with P450nor (CYP55) as a prototype. Independent of other electron transfer proteins, P450nor takes up electrons derived from NADH and catalyses the conversion of two nitric oxide molecules to nitrous oxide.[108,109]

Distinguished from most P450s, which function as monooxygenases and require delivery of two electrons, class X P450 enzymes convert substrates using an independent intramolecular transfer system. Examples of this class include allene oxide synthase AOS (CYP74A), hydroperoxide lyase HPL (CYP74B/C) and divenyl ether synthase DES (CYP74D).[110,111]

1.2.4 The catalytic cycle of P450 enzymes

The P450 catalytic cycle has been extensively studied and long debated. A generally accepted mechanism is shown in Figure 1.10.[74,76,86]

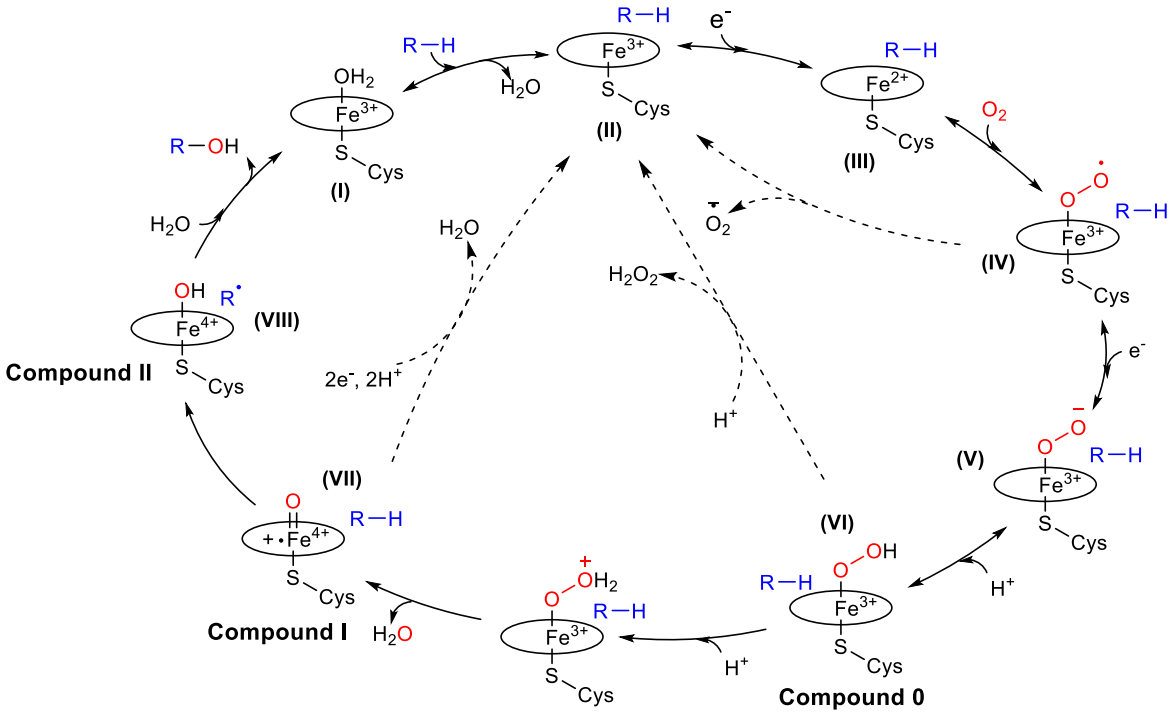


Figure 1.10 The commonly accepted catalytic cycle of P450 enzymes.[74]

In the resting state (**I**), the haem ferric iron atom is six-coordinate with four equatorial nitrogen ligands supplied by the porphyrin system, a proximal cysteine thiolate group, and a water ligand occuping the other axial position, making the ferric iron low spin (S = 1/2). The arrival of a substrate molecule causes the displacement of this water ligand and a high spin (S = 5/2) state of the haem ferric iron (**II**). The ferric haem is reduced to the ferrous form (**III**). The oxy complex (**IV**) is formed upon the binding of dioxygen. A second electron is then transferred to give **V**, which is protonated to give the Fe^{III}-hydroperoxy complex (**VI**), or 'Compound 0'. The distal oxygen is then protonated, which leads to the heterolytic cleavage of the O–O bond and the formation of the ferryl species ($P^{+\bullet}Fe^{IV}$=O) (**VII**), or 'Compound I'. This species then

abstracts a hydrogen atom from the substrate molecule to give a substrate radical and form 'Compound II' (**VIII**). The substrate radical then recombines with OH• from **VIII** to produce the alcohol product and the haem iron returns to the ferric state.

However, it is worth noting that in the catalytic cycle, the consumption of NAD(P)H is not always fully coupled with product formation.[62,74,76,112] If the dioxygen binding to **III** is too hindered, the loss of superoxide from **IV** may be possible due to the slow transfer of the second electron to **IV** (superoxide uncoupling). If the substrate is too loosely fit in the active site and thus fails to prevent the water encroachment, Compound 0 (**VI**) may be protonated at the iron-bound oxygen, leading to loss of hydrogen peroxide (peroxide uncoupling). High concentration of hydrogen peroxide causes the peroxide uncoupling to reverse−this reaction is known as the "peroxide shunt". This reaction allows P450 catalysis to be driven by peroxide instead of NAD(P)H/O_2/H^+, the feature for which the CYP152 family is known. If the substrate is bound in an orientation such that no hydrogen atom is conveniently positioned to be abstracted, the oxygen "atom" in Compound I (**VII**) may be reduced to water (oxidase uncoupling). Therefore, it is necessary to study and understand the binding interactions between P450 enzymes and substrates to promote monooxygenase activity.

1.3 $P450_{BM3}$: a self-sufficient monooxygenase

1.3.1 The electron transfer chain of $P450_{BM3}$

$P450_{BM3}$ (CYP102A1) was first isolated from *Bacillus megaterium* by Fulco *et al.*[113,114] in the 1970s as a medium- to long-chain fatty acid hydroxylase. Characterisation of the full-length enzyme (120 kDa) revealed that the P450 haem domain (BMP, 55 kDa) is fused to the reductase domain (BMR, 65 kDa) containing two flavin prosthetic groups, FAD and FMN, in a 1:1 molecular ratio (Fig. 1.11).[74,76,106,107,114,115]

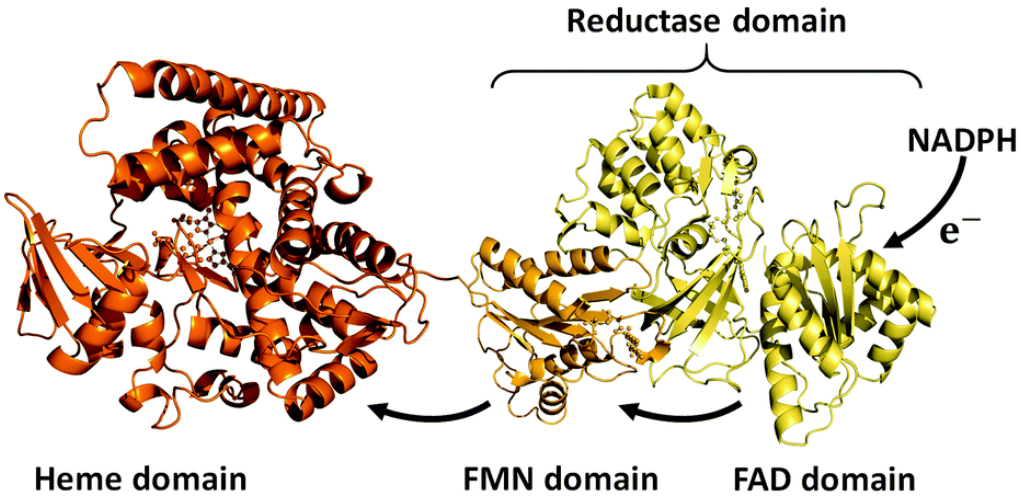


Figure 1.11 Cartoon of the electron transfer pathway of $P450_{BM3}$.[115]

Compared with most other P450 systems which receive electrons from discrete protein partners, this unique structure means the electron transfer to $P450_{BM3}$ haem iron is not contingent on collision encounter and recognition with a redox partner (Fig. 1.11).

Another unique feature of $P450_{BM3}$ is that it is only functional in the dimeric form (Fig. 1.12).[74,116-119] Detailed investigation suggested that electron transfer from FAD to FMN follows a *cis* fashion within the same monomer (monomer A or B), *i.e.*, $FAD_A \rightarrow FMN_A$ or $FAD_B \rightarrow FMN_B$, whereas the subsequent electron transfer to the P450 domain is *trans*, *i.e.*, $FMN_A \rightarrow P450_B$ or $FMN_B \rightarrow P450_A$ (Fig. 1.13).[116,117] This is supported by the variant complementation study to restore activity in heterodimeric $P450_{BM3}$ enzymes by Neeli *et al*. conducted.[118] In this experiment, a mixture of the two non-functional variants, A264H (P450 domain inactivated) and G570D (FMN binding abolished), reconstituted the fatty acid oxidation activity.

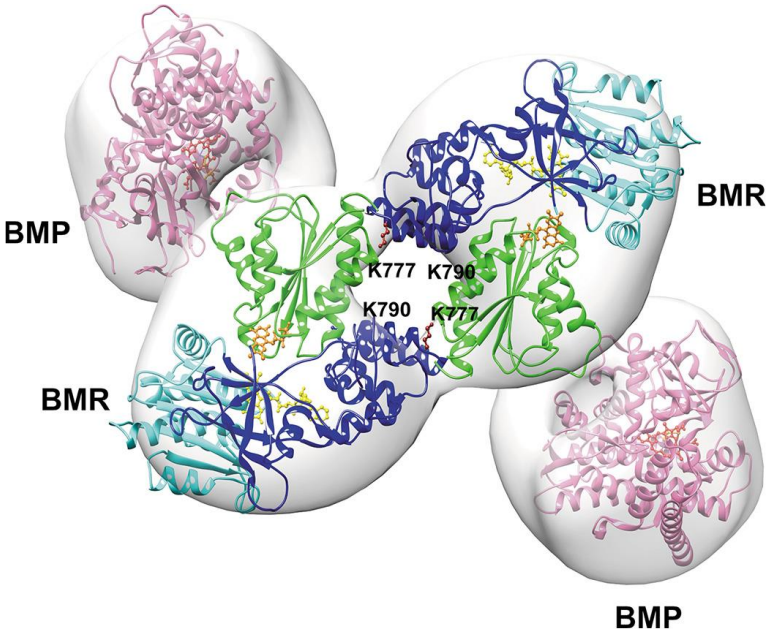


Figure 1.12 3D model of dimeric P450$_{BM3}$ showing the close view of the dimerisation interface.[117] The BMP molecule (haem domain) is shown in magenta, and the FMN, FAD, and nucleotide-binding domains of the BMR are shown in green, blue and cyan, respectively. The haem and cofactors FMN, FAD is shown in red, orange and yellow, respectively.

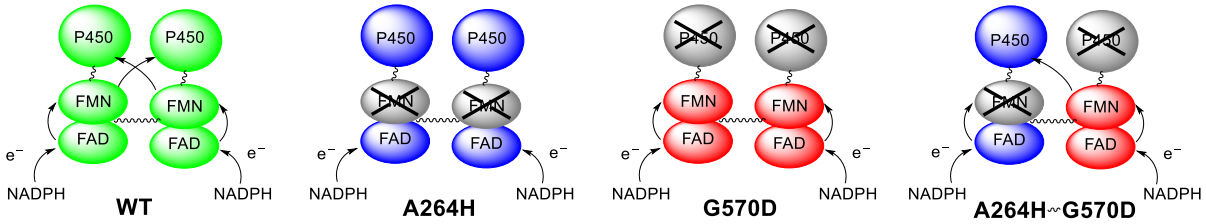


Figure 1.13 Schematic representation of the electron transfer routes in WT P450$_{BM3}$, A264H variant, G570D variant and the mixture of A264H–G570D variants.

These unique structural arrangements make P450$_{BM3}$ self-sufficient and facilitate inter-domain electron transfer, resulting in high activity levels, *e.g.*, >5000 min^{-1} for fatty acids such as arachidonic acid.[74,120-122]

1.3.2 Crystal structure of P450$_{BM3}$

Crystal structures of the separately expressed haem domain of WT P450$_{BM3}$ are available in both substrate free (SF) and substrate bound (SB) forms. As shown in the SF structure (Fig. 1.14-A), there are 15 major helices (A, B, B′, C, D, E, F, G, H, I, J, J′, K, K′ and L) in the

P450$_{BM3}$ haem domain, together with various β-sheets and loop regions. The haem is located roughly in the centre of the structure with the I-helix above and the L-helix underneath the porphyrin ring. The B′- and F-helices are also in the vicinity of the haem. Sequence alignment across the P450 superfamily has revealed six substrate recognition sites (SRS),[123] which in P450$_{BM3}$ are distributed in the B′- (SRS1, K69–H92), F- (SRS2, L181–L188), G- (SRS3, E200–D208), and I- (SRS4, N253–G271) helices and β1- (SRS5, W325–A335), and β4- (SRS6, K434–P441) sheets (Fig. 1.14-B).[74] The alignment of P450$_{BM3}$ with the other 25 members of the CYP102A sub-family shows SRS4 is fully conserved while there are few variations in SRS6. SRS1 is well conserved at the C-terminus but not the N-terminus. SRS2 and SRS5 both reveal anomalies, whereas SRS3 is almost entirely non-conserved.

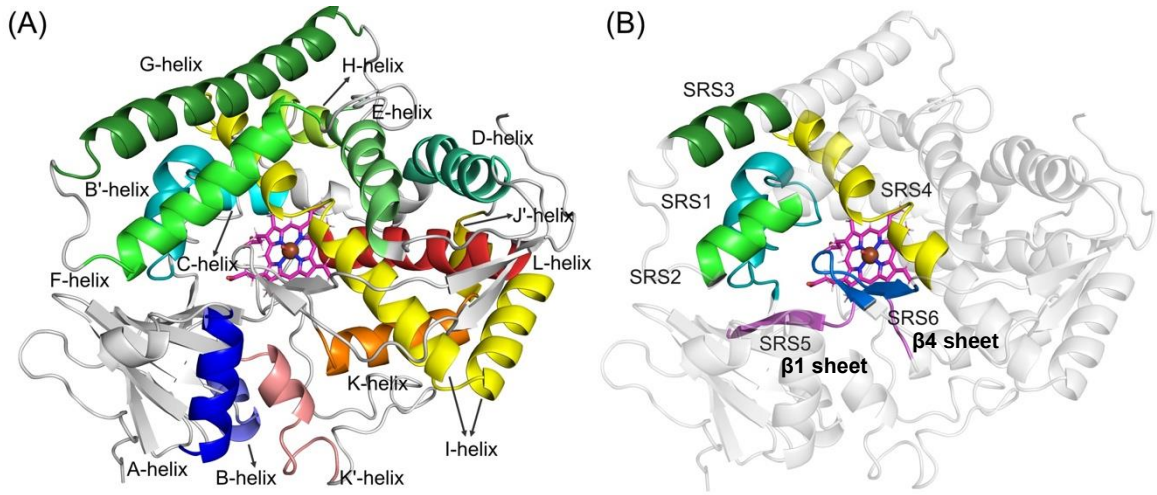


Figure 1.14 (A) The crystal structure of SF WT P450$_{BM3}$ (PDB code: 1BU7)[124] with helices shown in different colours; **(B)** The location of the six substrate recognition sites SRS1–6.

1.3.3 The active site and key residues of P450$_{BM3}$

Similar to certain mammalian P450s,[125] the SB crystal structure of P450$_{BM3}$ [bound with *N*-palmitoylglycine (NPG)] reflects a "pre-catalytic" conformation where no substrate atom lies sufficiently close to the haem iron for hydroxylation to occur (Fig. 1.15). It is thus unclear whether the residues depicted in the crystal structure, especially those lining the access channel,

are catalytically and physiologically relevant. This further complicates structural-based engineering for this enzyme. Nevertheless, analysis of the SB structure reveals important residues for substrate recognition and binding orientation, which in many cases has led to discovery of variants with enhanced activity.[74,122,126] Some of these key residues are discussed here (Fig. 1.15).

Located in the substrate access channel, residues R47 and Y51 are proposed to be involved in the fatty acid substrate recognition. They were therefore early targets for engineering. It was found that mutagenesis of either residue generally lowered fatty acid hydroxylase activity and raised the binding constant, suggesting they act together as the gate-keeper to the active site.[74,122,127] Carmichael and Wong[128] reported the combination of hydrophobic substitutions R47L and Y51F which increased oxidation activity for polycyclic aromatic hydrocarbons (PAHs) by up to 40-fold.

As one of the most mutated residues of $P450_{BM3}$, F87 is directly above and in contact with the haem; its side chain extends into the substrate access channel and strongly influences substrate specificity and product selectivity.[74,122,126] Mutations F87V and F87A expand the $P450_{BM3}$ substrate range and especially play important roles in drug-metabolising activities.[129-131]

Residue A328 sits above the porphyrin ring and opposite to F87. Mutations here affect the binding orientation of substrates and alter the selectivity, *e.g.*, Peters *et al.*[132] reported an A328V-containing variant that dramatically changed the regio- and enantioselectivity of hydroxylation for alkanes such as octane, nonane and decane.

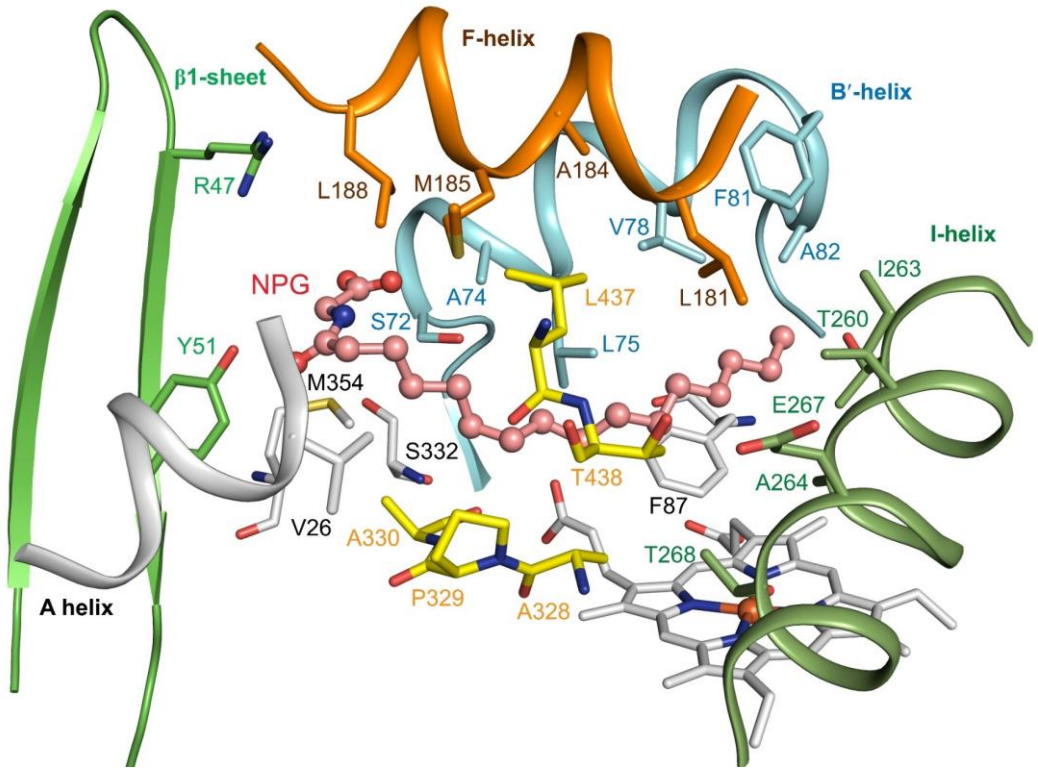


Figure 1.15 The active site and selected residues of the *N*-palmitoylglycine (NPG)-bound WT P450$_{BM3}$ crystal structure.

In the SB crystal structure, the aliphatic terminus of NPG reaches over the haem iron and contacts A82 (Fig. 1.15). Huang *et al.*[133] proposed that A82 might have created a hole above the haem and that a bulky substitution at A82 would fill in this and increase oxidation activity of unnatural substrates. Experimentally, the A82F and A82W variants showed ~800-fold increase in laurate binding affinity, and the formation of indigo in *E. coli* cultures expressing these variants suggested they are also effective catalysts to hydroxylate indole.

The E267-T268 acid-alcohol pair is located close to the kink region of the I-helix and is a highly conserved structural feature across the P450 superfamily. The T268 residue, as one of the few polar groups in the active site, is involved in various roles in the catalytic cycle, such as proton delivery, dioxygen activation, and the stabilisation of the oxy-ferrous or hydroperoxyl intermediate (**IV** and **VI**, Fig. 1.9).[74,134-137] Most of the reported mutations at T268 typically

have adverse impact on the fatty acid oxidation activity,[134,135] while the E267 mutations convert $P450_{BM3}$ to a peroxygenase.[136]

1.4 Enzyme engineering for $P450_{BM3}$

1.4.1 Rational design and directed evolution

Owing to its solubility, self-sufficiency, high expression level in *E. coli*, and tunable regio- and enantioselectivity, $P450_{BM3}$ has been extensively engineered as a biocatalyst.[59,74,75,86] A few of the enzyme's natural substrates are precursors to desirable oxidation products, *e.g.*, linolenic acid is metabolised to leukotoxin whose derivatives are implicated in pathophysiological responses,[138] and 15,16-epoxyoctadeca-9,12-dienoic acid which possesses antifungal properties.[139] However, the vast majority of products of interest have precursors that are not structurally-related to the natural substrate of wild type $P450_{BM3}$ and therefore do not fit the active site well. The primary aim of enzyme engineering is to alter the substrate specificity, with robustness, longevity and stability being important considerations as well.

Early research relied much on rational design aided with analysis of the crystal structure. The first reports of unnatural substate oxidation by $P450_{BM3}$ variants included mutations at F87 and A328 and substrates such as 1,1,2,2,-tetrachloroethane[140] and 2-phenylpropanal.[141] Product formation rate were enhanced by more than two orders of magnitude for the intractable polyaromatic hydrocarbons (PAH) by two space-creating mutations around the haem centre, F87A and A264G, and the R47L/Y51F couplet which neutralise the polar binding site at the access channel.[128] Other examples include F87G and F87V raising activity for propylbenzene,[128] 3-chlorostyrene,[128] β-ionone[142] and monoterpenes,[143] and I263V and A328V improving epoxidation levels for terminal alkenes.[144]

In the directed evolution era, the first randomly generated mutation of $P450_{BM3}$ to be reported was P25Q, which weakened fatty acid binding and changed the product profile.[145] However, such approach drew limited interest from researchers working on $P450_{BM3}$ engineering prior to the discovery that certain active P450 variants led to formation of blue pigments in complex media cultures by oxidising indole from tryptophan degradation to indigo.[146] This provided a viable and cost-effective method to screen many thousands of randomly generated variants for improved oxidase activity. In an early random mutagenesis-based study, error-prone PCR was used to generate a library of random variants. The indigo-producing ones were analysed and these all contained mutations at A74, F87 and L188. Site-saturation mutagenesis at these residues in turn then led to the discovery of a highly active variant, A74G/F87V/L188Q (GVQ),[147] which oxidised a wide range of unnatural substrates including PAHs,[148] chlorinated dioxins,[149] and organophosphorus pesticides.[150,151]

Arnold and co-workers used an activity screening assay similar to that for esterase directed evolution.[152] A library of $P450_{BM3}$ generated by random mutagenesis was screened against *p*-nitrophenyl octyl ether (8-pnpane). Terminal hydroxylation of this compound by active variants would generate the unstable hemiacetal, which decomposed to the aldehyde and *p*-nitrophenolate. The formation of the latter could be monitored at 410 nm and thus be an indication of the activity levels of variants (Fig. 1.16).[152-154]

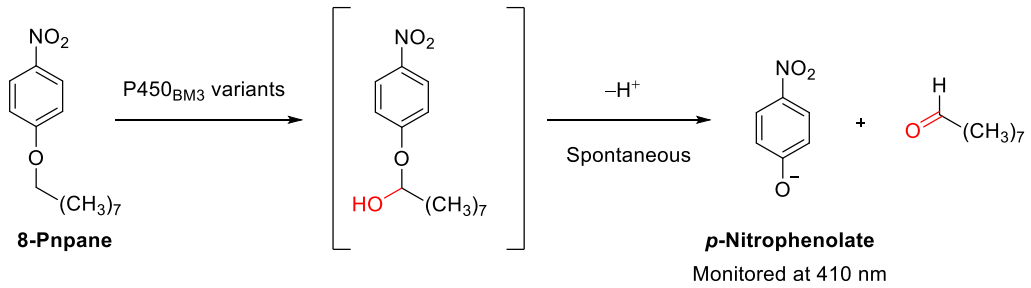


Figure 1.16 Schematic mechanism of the activity screening assay using p-nitrophenyl octyl ether (8-pnpane).

This has been combined with ep-PCR and DNA shuffling to engineer $P450_{BM3}$ for improved oxidation activity for alkanes.[153-157] Five rounds of random mutagenesis, together with recombination of beneficial mutations, led to the discovery of the variant 139-3 (V78A/H138Y/T175I/V178I/A184V/H236Q/E252G/R255S/A295T/L353V) with enhanced octane (TTN~3000) and propane (TTN~1000) oxidation, compared with TTN of 80 and 15 by the WT.[153] Among the 11 mutations in this variant, only V78A is in the active site and five mutations (T175I, V178I, A184V E252G and R255S) are in the highly flexible F-G helix-loop-helix structure and the I helix along which it slides during substrate binding (Fig. 1.17). Other mutations are remote or on the surface. With more rounds of random and site-saturation mutagenesis, variant 35E11 was found, which improved propane oxidation activity further (TTN~6000).[155]

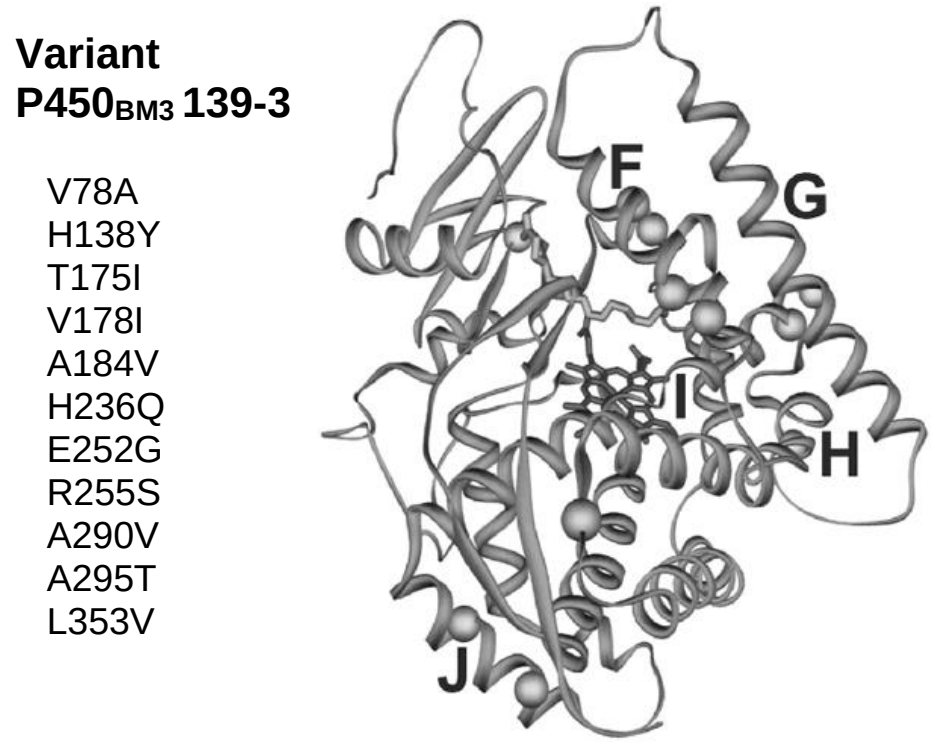


Figure 1.17 Positions of the amino acid substitutions in $P450_{BM3}$ variant 139-3.[153] The variant contains 11 substitutions, represented by spheres on the crystal structure of the SB WT enzyme (PDB: 1FAG).

Then, a "domain engineering strategy", which include site-saturation mutagenesis on not only the P450 domain but also the FAD and FMN reductase domains, was deployed and a proficient propane monooxygenase ($P450_{PMO}$) was produced, with striking catalytic efficiency (TTN~45000).[156] In addition to alkane oxidation, these evolved variants were also found to oxidise numerous drug molecules such as buspirone[158] and verapamil[159] (Fig. 1.18). Further mutagenesis work led to the discovery of the variant P411, with the signature ferrous CO Soret peak shifted to 411 nm from 450 nm by the C400S mutation. The unique serine-haem ligation abolished the monooxygenation activity and enabled efficient carbene insertion activity, such as, olefin cyclopropanation catalysed by P411-CIS[160] and indole alkylation by P411-HF.[161]

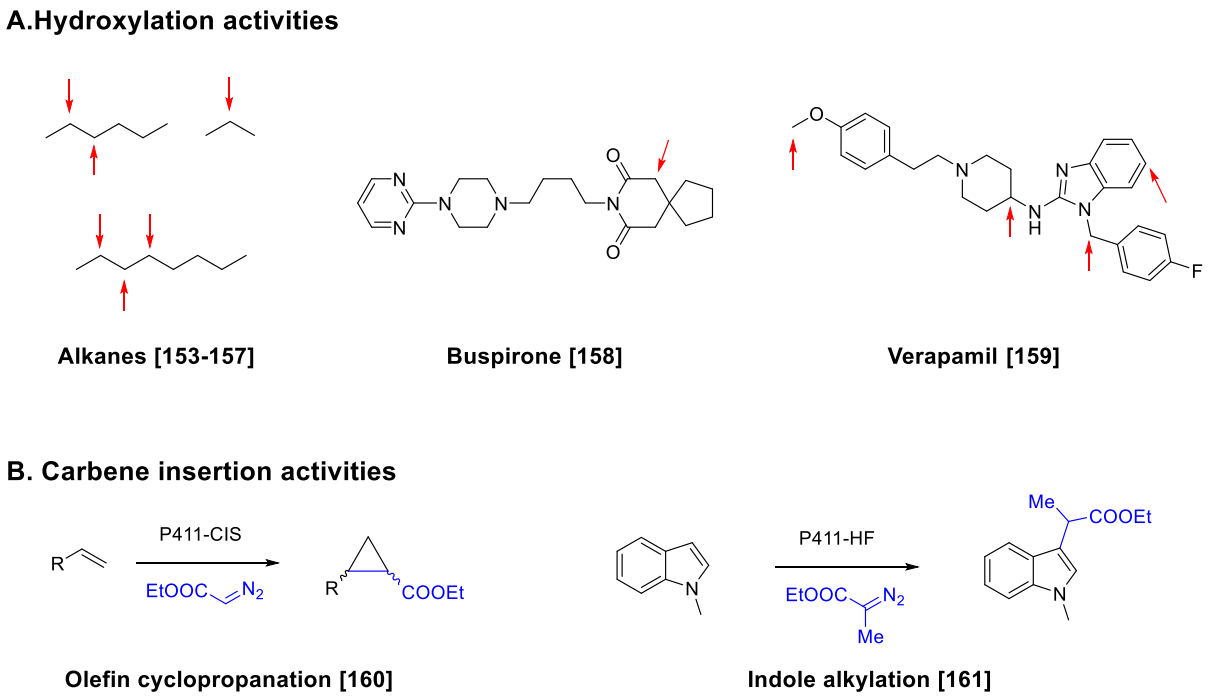


Figure 1.18 Examples of **(A)** alkanes and drug molecules oxidised by $P450_{BM3}$ variants, with oxidation sites indicated by red arrows; and **(B)** carbene insertion activities catalysed by P411 variants of $P450_{BM3}$.

Wong and co-workers utilised the indigo formation characteristic to preliminarily screen for active $P450_{BM3}$ variants from randomly generated library and then perform second-generation mutagenesis to improve enzymatic activity and product selectivity for target compounds.[75,83,162,163] Error-prone PCR was carried out on WT and the F87A variant to generate

random variants, which was transformed into *E. coli* cells and cultured. Variants giving rise to blue colonies were selected for further DNA shuffling and indigo-formation screening. Promising variants were assayed *in vivo* for oxidation activity and product selectivity for two unnatural substrates, naphthalene and propylbenzene. Of these, four most active variants were identified: A330P; A191T/N239H/I259V/A276T/L353I (KT2); F87A/H171L/Q307H/N319Y (KSK19); F87A/A330P/E377A/D425N (KT5), which were then used as template for further mutagenesis.[74,75] All these showed enhanced oxidation activity by at least one order of magnitude over the WT for naphthalene and propylbenzene. The RLYF/A330P variant gave the highest product formation rate (PFR) for naphthalene at 666 min^{-1}, compared with 3.1 min^{-1} by WT. With propyl benzene, the RLYF/KT2 variant gave a PFR of 2688 min^{-1} while WT gave 606 min^{-1}. The unusual high activity of the single variant A330P inspired the introduction of proline mutations at selected residues. This led to the discovery of the I401P variant, which was unexpectedly active, *e.g.*, giving a PFR for propylbenzene of 3578 min^{-1}.[74,75,164]

Crystal structures of the A330P and I401P variants were obtained for detailed analysis (Figures 1.19 and 1.20). Compared with the WT structure, the A330P mutation causes a significant re-location of the adjacent P329 residue towards the haem centre. As a result, the substrate binding pocket was reshaped dramatically (Fig. 1.19). The constrained active site may lead to enhanced activity for smaller molecules such as naphthalene and propylbenzene.

The I401P variant structure showed some unexpected features compared with WT (Fig 1.20).[164,165] The mutated P401 moved towards the proximal direction of haem, causing the adjacent C400 and the haem iron to drop as well. As a result, the distance between iron and the axial water is significantly lengthened to 3.6 Å from 2.6 Å. It is arguable whether the axial water in the I401P variant remains a ligand of the haem iron or not, but in any case, the water molecule would be more easily displaced upon substrate binding compared to WT. Whitehouse *et al.*[165] reported that the haem redox potential and the rate constant for the first electron transfer

step of the I401P variant were almost the same as the palmitic acid-bound WT. Two I-helix residues, G265 and H266, were re-orientated, and thus interrupted several interhelical hydrogen bonds. In summary, the I401P mutation induces conformational changes that facilitate the initiation process of the catalytic cycle and hence increases the oxidation activity for unnatural substrates.

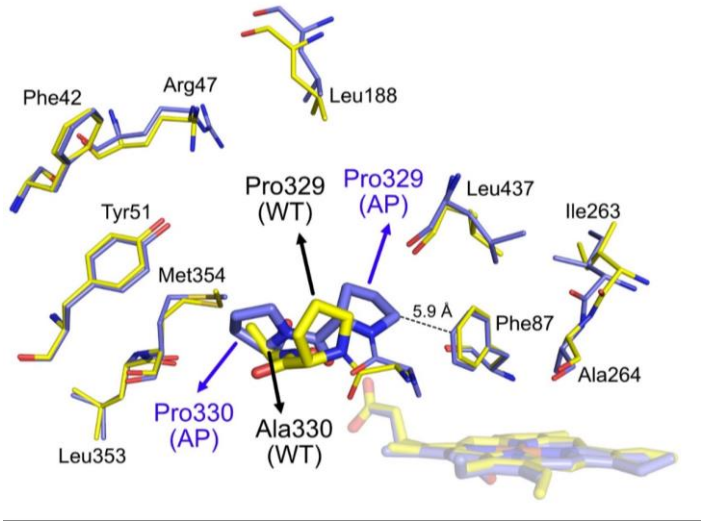


Figure 1.19 Overlaid structure of the SF A330P variant (blue; PDB code: 3M4V)[165] and SF WT (yellow; 1BU7)[124].

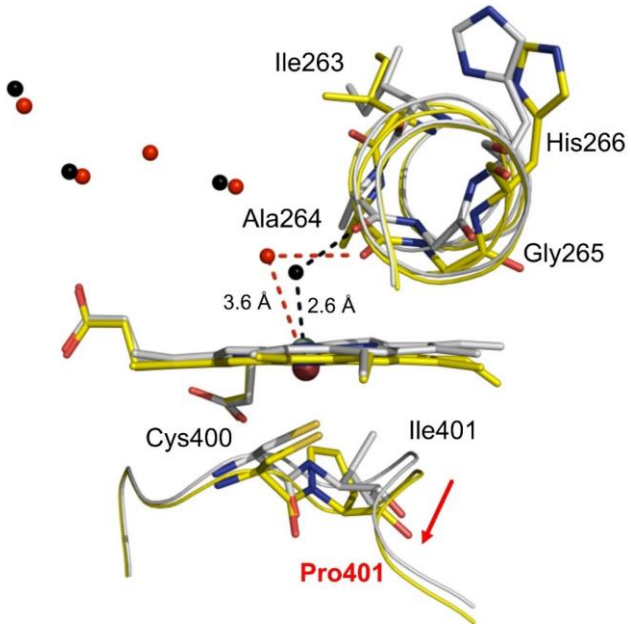


Figure 1.20 Overlaid structure of the I401P variant (yellow, with water molecules as red spheres; PDB code: 3HF2)[165] and SF WT (grey, with water molecules as black spheres; PDB code: 1BU7)[124].

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

One limitation of the directed evolution by ep-PCR is that the size of variant library could become too large to get effective variants for target substrates. In order to narrow down the target list of variants, Reetz and co-workers developed two methods, combinatorial active-site saturation (CAST) and iterative saturation mutagenesis (ISM), both guided by structural information of the enzyme.[166-168] CAST aims to design and generate small and focused initial variant library by randomisation at several sets of two spatially close amino acid positions around the active site. In such randomisation process, the following geometric relation pertains: if one member of the amino acid pair occurs in the 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. Mutagenesis is first conducted at the randomised residues to afford several sub-libraries, also called focused libraries, for activity screening. This approach is always combined with ISM, where beneficial variants from a sub-library serves as template for further round(s) of site-saturation mutagenesis until desired activity and selectivity levels are reached. This combined strategy significantly reduces time and resources needed to obtain hits, especially when high throughput screening methods are not available.

Reetz and co-workers engineered $P450_{BM3}$ for highly selective hydroxylation of testosterone (TST) by deploying the CAST-ISM approach. In their early work, the F87A variant was selected as template and 20 residues were chosen as possible candidates for mutagenesis, using the SB $P450_{BM3}$ crystal structure as a rough guide. These residues were divided into nine sites (A–I), each with 2–3 members: 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). The initial experiments focused on three sites: A, B and C, while site saturation mutagenesis was also conducted at residues S72 (D), A328 (G) and A330 (G). Residues in sites

A and C were randomised using NDC codon degeneracy encoding only 12 amino acids (Gly, Val, Leu, Ile, Phe, Ser, Cys, Tyr, His, Asp, Asn and Arg) to reduce screening efforts, while site B was subjected to NNK degeneracy for all 20 amino acids. Roughly 18,000 variants were generated and tested for oxidation activity against TST and high regio-and stereoselectivity for 2β- and 15β-hydroxylation were observed, with the F87A/A330W variant giving 97% of 2β-hydroxy-TST at 79% conversion and R47Y/T49F/V78L/A82M/F87A giving 96% of 15β-hydroxy-TST at 85% conversion (Fig. 1.21).[168]

Later, Reetz and co-workers further developed the ISM-CAST approach by utilising information derived from mutability landscapes (MLs) and molecular dynamics (MD) simulations. Since first used by Rost and co-workers[169] in developing algorithm to predict the effect of single-amino-acid substitutions in disease-related proteins, the idea of mutability landscape has been increasingly applied in enzyme engineering. For this type of application, data for a large number of enzyme variants are analysed to determine the effect of each single-amino-acid substitution on enzyme activity, product selectivity or stability. Detailed maps of beneficial, neutral and detrimental amino acids for each residue under investigation are thus generated, which serve as basis for further mutagenesis, instead of complete randomisation.[170] This is sometimes aided with molecular dynamics simulations which provide information about substrate-residue interactions. Combination of these tools significantly reduced mutagenesis efforts. In a study targeting TST as model substrate, Reetz and co-workers applied ML analysis on previous data obtained from ISM-CAST. Guided with information from MD-simulations, they chose 10 residues as candidates and divided them into two groups: A (R47, S72, A82, F87 and L437) and B (Y51, V78, L181, L188 and A330). Each residue was to be mutated to 2–4 amino acids only. Using this strategy, ~3000 variants were tested and high selectivities for 16α (highest at 96% by R47W/Y51W/S72I/A82F/F87I/L181C) and 16β (92% by R47W/A82W/F87V/L181Q/T436R/M177S) oxidation were achieved (Fig. 1.21).[171]

More recently, a larger library was generated (~30,000 variants) by randomising 15 residues (R47, S72, K76, F77, V78, R79, D80, F81, A82, T88, M177, M185, L188, F205 and I209) each to 2 amino acids. Such randomisation is achieved by using the uracil-specific excision reagent (USER) method. This library showed medium to high selectivity and conversion for 7β hydroxylation of TST (Fig. 1.21) and products of five other steroids.[172]

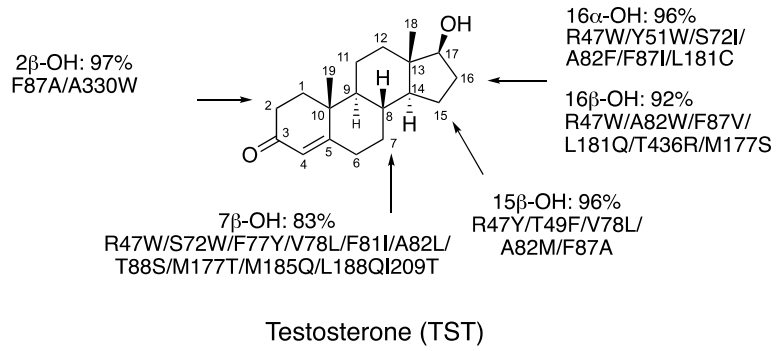


Figure 1.21 Oxidation of testosterone by $P450_{BM3}$ variants developed by Reetz and co-workers.[165-167]

1.4.3 Rationally designed variant library

With the development of the CAST-ISM approach, huge efforts for mutagenesis and screening are still required to obtain high regio- and enantioselectivity for a target product. Therefore, many have adopted library screening approaches in which substrates are firstly tested against a focused or diversified variant library to search for key residues and mutations. Seifert *et al.*[173] developed a focused $P450_{BM3}$ variant library by mutating two residues, F87 and A328, to five hydrophobic amino acids (A, V, F, I and L). Variants in this library showed improved selectivity for oxidation of geranylacetone (9,10-epoxides), nerylacetone (epoxide), limonene (epoxide) and valencene (C2 oxidation to form nootkatol).

Wong and co-workers developed a library containing ~100 $P450_{BM3}$ variants. These are derived from five base variants, A330P, K19 (H171L/Q307H/N319Y), RP (R47L/Y51F/I401P), RT2 (R47L/Y51F/A191T/N239H/I259V/A276T/L353I), and GVQ (A74G/F87V/L188Q), with

additional mutations at 2−4 active site residues (S72, V78, F81, A82, F87, A184, I263, A264, E267, A328, A330, L437 and T438).[75,83,162,163,165,174-176] Screening with this library has shown high activity and selectivity for diverse range of unnatural substrates, such as steroids (TST; Fig. 1.21),[176] diterpenes (eleuthoside),[163] smaller drug core motifs (quinolines and tetrahydroisoquinolines)[175] and drug molecules[162] (diclofenac, naproxen, amitriptyline, lidocaine and chlorzoxazone). These provided corresponding alcohol precursors for synthesis of various valuable drugs.

Fasan and co-workers constructed a larger library of ~500 variants derived from base variants in Arnold's work.[76,177,178] The variants in this library were not specifically selected or designed, instead, they were used as a "training set" to screen for residues that affect activity and selectivity.[177,178] Once hits were identified, further site saturation mutagenesis and analysis could lead to higher selectivity. For example, in targeting artemisinin, the variant V78A/F81S/A82V/F87A/F142S/T175I/A180T/A184V/A197V/F205C/S226R/H236G/R255S/A290V/L353V was identified from the training set, showing selective C7 oxidation (83% 7*R* with 10% 7*S*). Further mutagenesis based on this variant led to the variant II-H10 which showed 100% selectivity for (7*R*)-hydroxy-artemisinin. Selective C7 oxidation of artemisinin is inaccessible by chemical approaches; thus, this work provided a viable route for the synthesis of (7*R*)-fluoroartemether, a fluorinated precursor of the antimalarial drug artemether (Fig. 1.22).[177,178]

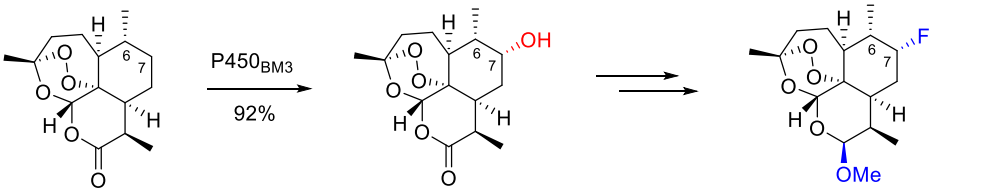


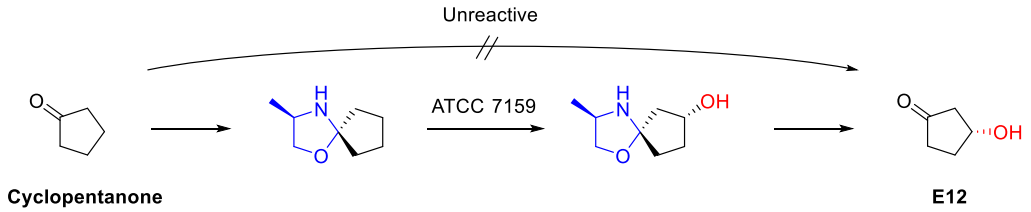
Figure 1.22 Synthesis of (7R)-fluoroartemether enabled by selective C7 oxidation of artemisinin.

1.4.4 Substrate engineering

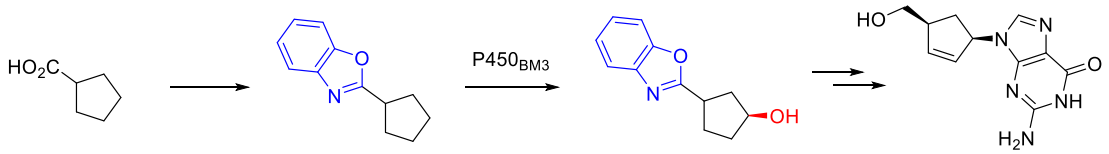
From a totally different angle, modification of substrates themselves (or, substrate engineering) could lead to altered enzymatic activity and product selectivity as well. The idea of substrate engineering for application in bio-hydroxylation was first proposed by Griengl and co-workers,[179] who introduced docking and protecting (d/p) group to unnatural substrates to facilitate reactivity. In an early work they found that cyclopentanone, an unreactive substrate of the fungus *Beauveria bassiana* ATCC 7159, could be effectively hydroxylated after the ketone moiety was derivatised as a spirooxazolidine, to produce **E12** with high ee (Fig. 1.23-A).[179]

Arnold and co-workers[180] derived this concept to apply in $P450_{BM3}$ catalysis. They derivatised the carboxylic acid **E13** to render it reactive for $P450_{BM3}$ oxidation to alcohol precursors for the synthesis of carbovir, **E14** (Fig. 1.23-B). The combination of substrate and enzyme engineering approaches has since led to some successful applications in $P450_{BM3}$ catalysis. Lange and co-workers[181] investigated the use of d/p groups in $P450_{BM3}$-mediated oxidation in a biomimetic fashion. A set of modified nosyl protecting groups were designed and vabicaserin was studied as the model compound. Combined with some mutagenesis work, the F87V variant was identified to give >80% regioselectivity for the aromatic oxidation of a nosyl protected vabicaserin derivative (Fig. 1.23-C). The nosyl groups can also improve substrate solubility and are easy to remove. Urlacher and co-workers[182] improved regioselectivity for $P450_{BM3}$-catalysed oxidation of cembranoids by increasing size, polarity, and ring rigidity of the exocyclic substituents (d/g groups). The dimethyl diester functionality was the most effective. Further mutagenesis led to the V78A/F87A variant, which gave 90% selectivity for C5 oxidation of the dimethyl diester-derivatised cembranoid alcohol **E15** (Fig. 1.23-D).

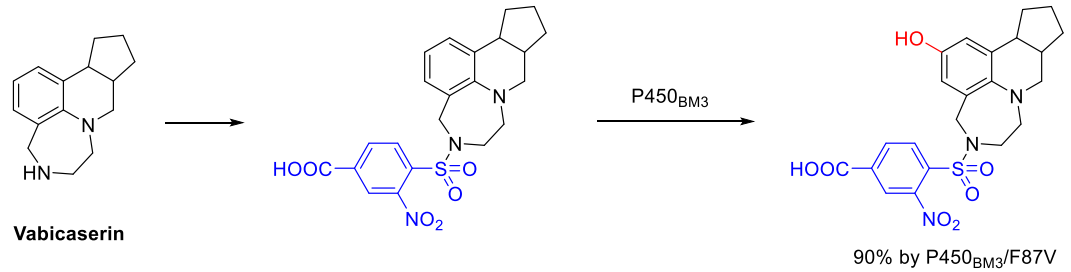
A.Griengl and co-workers, 1999 [179]



B.Arnold and co-workers, 2005 [180]



C.Lange and co-workers, 2018 [181]



D.Urlacher and co-workers, 2018 [182]

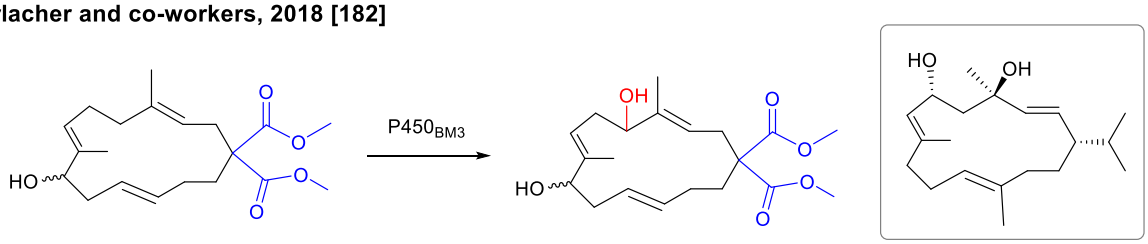


Figure 1.23 Examples of substrate engineering that led to selective hydroxylation of various compounds.

1.5 Molecular dynamics simulations and substrate docking

1.5.1 Molecular dynamic simulations

Although current enzyme engineering methodologies have been successful in improving enzymatic activity and product selectivity for numerous desirable products, most of these approaches still largely rely on random mutagenesis or site-saturation mutagenesis at multiple residues, which requires large screening efforts. Structural analysis of enzyme-substrate binding interactions provided by crystal structures have had significant contributions to mutation design. However, the enzymatic function is not decided only by their rigid structures but often more importantly, by the dynamics of residues and substrates.

With the rapid development of computational tools, molecular dynamics (MD) simulations have emerged as a powerful tool to provide insight into the structural features of enzymes and enzyme-substrate interactions. MD simulations predict how every atom in a protein or other molecules moves (the "internal motions") over a period of time, based on a classical mechanics treatment of interatomic interactions (Fig. 1.24).[183,184,187]

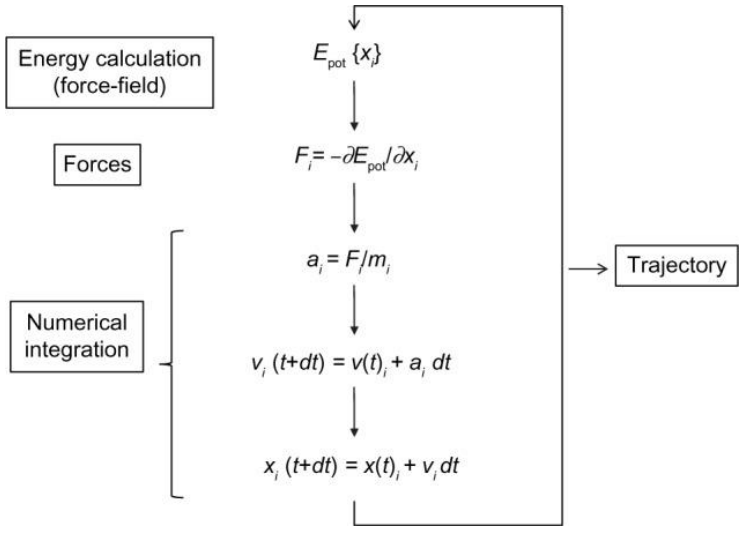


Figure 1.24 Basic algorithm of molecular dynamics (MD) simulations.[187] The output trajectory is an ordered list of 3N atom coordinates for each simulation time. The abbreviations represent: E_{pot}, potential energy; t, simulation time; dt, iteration time; for each spatial coordinate of the N simulated atoms (i): x, atom coordinate; F, force component; a, acceleration; m, atom mass; v, velocity.

Such simulations have been applied to capture a wide range of important biomolecular processes, such as protein folding, ligand binding and conformational change, as well as revealing positions of atoms at fs temporal resolution.[184] The first MD simulations were performed with simple gas molecules in late 1950s by Alder and Wainwright.[185] Then, in late 1970s, Karplus and co-workers ran the first MD simulation of a protein (bovine pancreatic trypsin inhibitor).[186] Since then, this approach has advanced from simulating a few hundreds of atoms to modelling complex biological systems including entire proteins solvated with specific solvents.[187] Simulations of 50,000–100,000 atoms are now routine.[187]

The basic idea of MD simulations is to calculate the forces exerted on each atom by all the other atoms in a biomolecular system, such as a protein solvated by water or a lipid bilayer. An initial model of the system is typically obtained from experimental structures (X-ray crystallography, cryo-EM, NMR, *etc*.) or comparative modelling data. This model is then subject to solvation, the methods of which can be divided to two types: (a) explicit-solvent methods, such as particle mesh Ewald (PME) with TIP3P water, which compute interactions between all pairs of solute and solvent atoms explicitly;[188,189] (b) implicit-solvent methods, such as the generalised Born model, which speed up atomistic simulations by approximating the discrete solvent as a continuum, and are thus used for faster calculations.[190-192] Once the solvated system is built, forces acting on all atoms are obtained through force fields equations, which deduce the potential energy from the molecular structure.[193-197] Although force fields currently used in simulations differ in how they are parameterised, the molecular features are typically represented similarly – spring-like terms for bond length and angles, the Coulomb's law for van der Waals and electrostatic interactions, and periodic functions for the bond rotations and Lennard–Jones potentials (Fig. 1.25).[198] The simplicity of these representations assures fast calculations even for large systems but also means these force fields are inherently approximate. Once the forces acting on individual atoms are calculated, Newton's laws of motions are then

used to predict the spatial position of each atom as a function of time. The resulting trajectory is essentially a three-dimensional movie that describes conformation of the system at every atom during the time interval of the simulations. Since the integration of movement is done numerically, to avoid instability, the time steps in MD simulations must be shorter than the fastest movements in molecules. Typically, each time step is only a few femtoseconds for atomistic simulations. Most of the biochemical processes, *e.g.*, structural changes of enzymes upon substrate binding, take place in ns, μs, or longer. Thus, an MD simulation run requires millions or billions of time steps, each single step evaluating millions of interatomic interactions. These cause simulations to be computationally demanding.

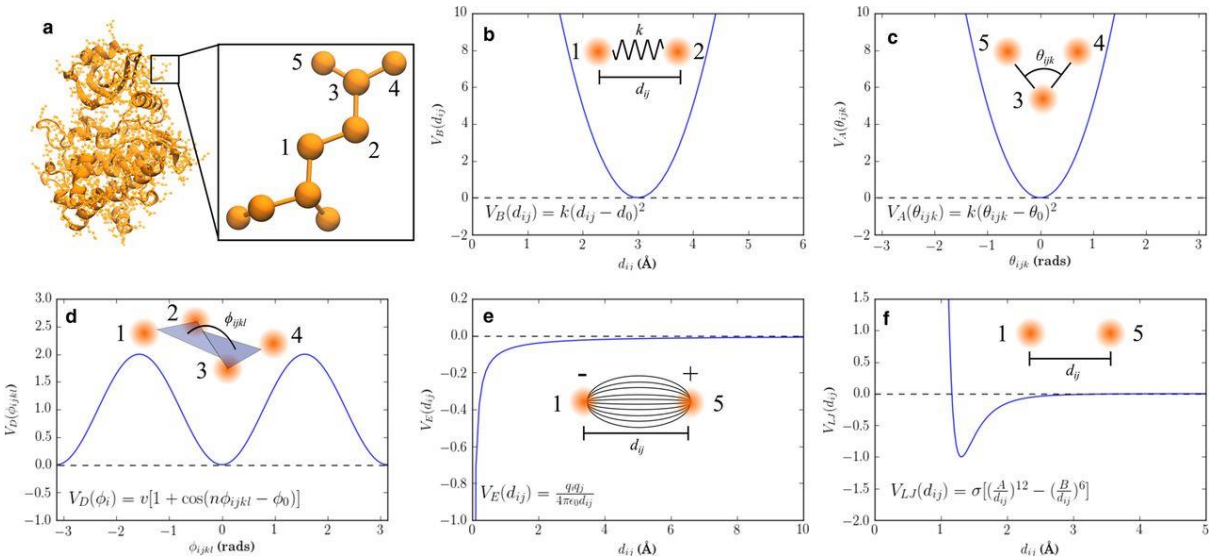


Figure 1.25 Examples of terms used to calculate interactions between atoms in the common force fields of docking algorithms.[198] These are exemplified using atoms 1, 2, 3, 4 and 5 as labelled in **(a)**, and terms are shown for **(b)** bond length; **(c)** bond angle; **(d)** bond rotation; **(e)** electrostatic interactions; and **(f)** Lennard–Jones potentials.

The explosion in developments of supercomputers and graphics processing units (GPUs) over the last few decades have made such simulations more accessible.[184,187] There are plenty of software to use, including GROMACS[199], NAMD[200], AMBER[201], CHARMN[202], and OpenMM.[203] These packages all perform similar computations and support multiple force fields but differ in how efficiently they integrate to different hardware and in the supported features,

such as enhanced sampling methods and temperature/pressure control schemes, *etc*. GROMACS was used in this work since routine analysis within the Wong group has been established with this.

1.5.2 Molecular docking

Molecular docking is used to model the favoured binding interactions between a small molecule (ligand) and a protein (receptor) at atomic level, starting with their unbound structures, MD simulated conformations, or homology modelling, *etc*.[204,205] The docking process involves two basic steps: (a) sampling conformations of the ligand (usually referred to as "poses") within the active site of the protein, using sampling algorithms; and (b) ranking these conformations (poses) by scoring functions. The goal is to reproduce the experimental binding pose and rank it highest among all poses.

Various sampling algorithms have been developed. Matching algorithms (MA) map a ligand into an active site of a protein (both represented as pharmacophores) according to the shape features and chemical information. Each distance between pharmacophores is calculated for a match, and new conformations of ligand are governed by distance matrix between the pharmacophore and the corresponding ligand atoms.[206,207] Incremental construction (IC) methods divide the ligand into several fragments and select one of these (usually the largest fragment) to dock into the active site first as anchor. The remaining fragments are then added incrementally.[208,209] In addition to IC, multiple copy simultaneous search (MCSS)[210] and LUDI[211,212] are also fragment-based approaches. Monte Carlo (MC) methods[213,214] generate poses of ligand through bond rotations or rigid-body translation, and these poses are either accepted or rejected upon test with an energy-based selection criterion. The accepted poses will be recorded and further modification lead to next generation of poses for test. The iteration proceed until a pre-defined quantity of poses are obtained. Genetic algorithms (GA)[215,216] form

a class of widely used stochastic methods. The idea of GA derives from Darwin's theory of evolution. Degrees of freedom of the ligand molecule are encoded as binary strings called "genes". These "genes" make up binding poses of ligand, called "chromosome". Two types of "genetic operators" are introduced: (a) "mutations" which make random changes to the "genes"; and (b) "crossovers" which exchanges "genes" between two "chromosomes". New poses are generated upon effect of these genetic operators and will be assessed by scoring function. Poses that "survived" (*i.e.*, meet the pre-defined criteria) will be recorded and used for next generation. GA has been implemented in many commonly adopted docking software, such as GOLD[217], DIVALI[218] DARWIN[219], and AutoDock series.[216,220,221] Recently, Sanner and co-workers[221] reported a refined GA (Fig. 1.26) which is implemented in the software AutoDock for Flexible Receptors (ADFR).

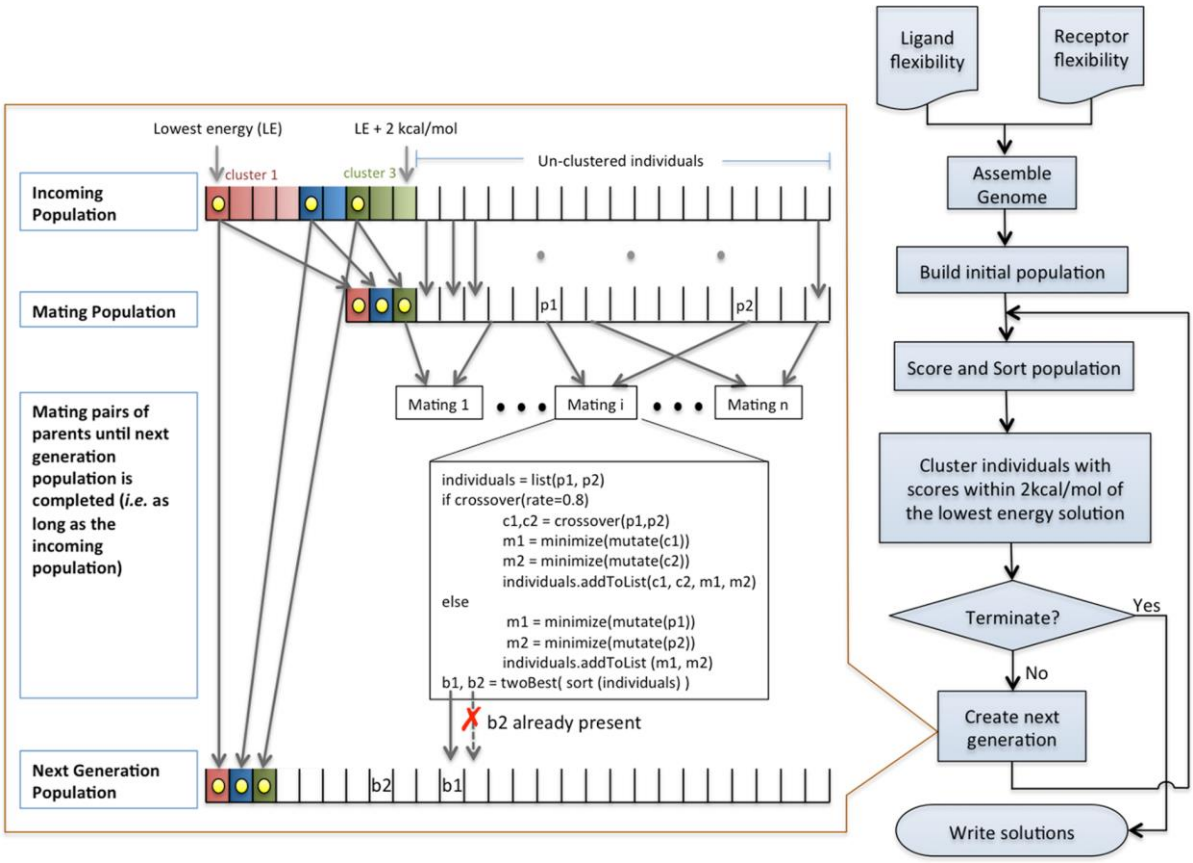


Figure 1.26 Schematic representation of the genetic algorithm (GA) implemented in AutoDock for Flexible Receptors (ADFR).[221]

Scoring functions serve the purpose to delineate the "correct" binding poses, those agreeing with experimental observations, from the "incorrect" ones. This is done by calculating and sorting the binding affinity of poses generated by the sampling algorithm. These calculations usually involve estimations to reduce computational time and depending on the assumptions and simplifications adopted, scoring functions can be divided into three groups: (1) force field-based methods which assess the binding energy by calculating the sum of the non-bonded (electrostatics and van der Waals) interactions.[222,223] The extended versions also consider hydrogen bonds, solvations and entropy contributions and are widely used in common docking software such as GOLD and AutoDock; (2) empirical scoring functions which split the binding energy into several components, such as hydrogen bond, ionic interaction, binding entropy and hydrophobic effect. These components are each multiplied by a coefficient, obtained from regression analysis fitted to test sets of ligand-protein complexes with known binding affinities. The summation of these multiplied components then affords the final score. Examples of empirical scoring functions include LUDI and ChemScore.[211,224,225] (3) knowledge-based methods, which utilise statistical analysis of crystal structures of ligand-protein complexes to obtain frequencies of interatomic contacts. The basic assumption is that the more favourable an interaction is, the greater the occurrence frequency will be. The score is calculated by weighting more the favoured contacts and penalising the repulsive interactions between each atom in the ligand and protein within the pre-defined cut-off. Examples of docking software adopting such methods include PMF, SMoG and Bleep *etc*.[226-229]

The early docking approaches mostly treated both ligand and receptor as rigid bodies, mainly due to the limitation of computation power. In this case, ligand flexibility could be addressed by deploying a set of pre-computed ligand conformations or by allowing a degree of atom-atom overlap between the protein and ligand. This is adopted by early versions of DOCK[230], FLOG[231], and some protein-protein docking software such as FTDOCK.[232] For systems whose behaviour

follows the "induce-fit" theory, it is important to consider the flexibilities of both ligand and receptor as there will be significant structural changes of both upon substrate binding. However, a fully flexible receptor requires too much computational power. Thus, the common approach treats the ligand as flexible while keeping the receptor rigid. Most of docking programs have adopted this methodology, such as AutoDock Vina[220], GOLD[217] and DARWIN.[219] Although incorporating complete receptor flexibility without taking up too much computational time remains a challenge, many docking programs, such as ADFR, have developed functions that allow users to define certain number of residues to be flexible. These residues are usually selected based on their roles in interactions with ligands. In this work, docking calculations are conducted using Autodock Vina[220] or Autodock Flexible Receptors (ADFR),[221] because of their accessibility, fast calculation and friendly user interface.

1.5.3 Applications in enzyme engineering

Although initially designed for and still primarily used as a tool to aid drug discovery, the combined approaches of MD simulations and molecular docking have also revealed great potential in enzyme engineering. In 1995, Ornstein and Paulsen performed the first MD simulations for $P450_{BM3}$ haem domain using the SF crystal structure.[233] The simulations suggested that the entrance of the substrate binding pocket was highly mobile, which indicated that the range of substrates $P450_{BM3}$ could accept was likely broader than what would be deduced from analysis of just the crystal structure. This has inspired some early enzyme engineering of $P450_{BM3}$ towards the oxidation of unnatural substrates.

In early $P450_{BM3}$ engineering, MD simulations and molecular docking were mainly used to investigate the origins of improved enzymatic activity and product selectivity. For example, in an early work to engineer $P450_{BM3}$ for selective steroid oxidation, Reetz and co-workers applied such a strategy for two effective variants, the 2β-selective variant for testosterone (TST), KSA-

2, and the 15β-selective variant KSA-14. MD simulations were performed on these and the output trajectories showed that binding pocket of KSA-2 and KSA-14 had quite different shapes. Docking of TST into the KSA-2 variant revealed only one active binding pose in which solely the 2β hydrogen pointed to the haem iron (Fig. 1.27-A). In contrast, the pose produced by docking with the 15β-selective variant KSA-14 showed the 15β hydrogen in the ideal proximity to the haem (Fig. 1.27-B).[168] Recently, Wong and co-workers used molecular docking to understand the roles of effective mutations that improved product selectivity for androstenedione (AD) oxidation.[176] For example, the K19/F87A/A82M/I263G/A264G variant was one of the most selective for 2β oxidation of AD (59%). Docking of AD with simulated structures of this variant showed the binding poses indicating 2β oxidation to be the most populated. The 2β pose docked structure revealed surprising resemblance to the AD-bound structure of CYP19A1, a C19-demethylase of AD. Since C19 and C2 of AD are very close to each other, the structural similarities suggested that this $P450_{BM3}$ variant mimicked the binding features of CYP19A1.

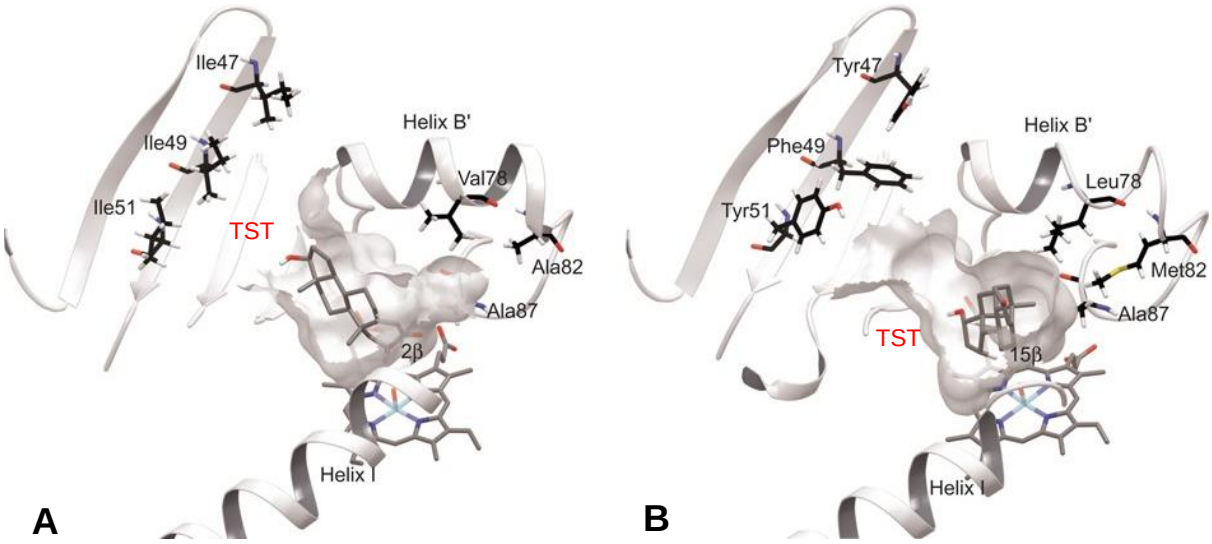


Figure 1.27 Binding poses of TST from the docking with **(A)** the 2β-selective variant, KSA-2; and **(B)** the 15β-selective variant, KSA-14.[168]

In general, applications of utilising the information from MD simulations and docking to guide the enzyme engineering process are greatly under-explored, although there have been some successful examples (Fig. 1.28). In 2014, Pleiss and co-workers reported a hotspot identification approach, which used molecular docking and data mining to identify key residues and mutations for selectivity of target products and thus construct focused libraries to screen for improved selectivity (Fig. 1.28-b).[234] In this work for selective C7 oxidation of (4*R*)-limonene by engineered $P450_{BM3}$, the analysis of the sequence of over 6300 CYP genes and the relevant MD simulations helped identify three key residues A264, A328 and L437. A focused library was generated and tested against (4*R*)-limonene, of which the A328V variant showed 27% regioselectivity for the C7 oxidation product, (*R*)-(+)-perillyl alcohol, while no C7 oxidation was observed for the WT $P450_{BM3}$. After two subsequent rounds of molecular modelling and variant screening, the variant A328V/L437F/A264V was generated, showing 97% of the perillyl alcohol product. This was obtained by screening only 29 variants in total (Fig. 1.28-b).

In 2015, Arnold and co-workers used computational reconstruction in combination with large-scale MD simulations to engineer the nitrating P450 TxtE.[235] To better understand the functional significance of the F/G loop (residues 176–183) of TxtE, which remained unresolved in the crystal structure[236], homology reconstruction was carried out for the F/G loop with the L-troptophan-bound WT TxtE structure, and MD simulations performed for the whole structure. The simulated trajectories revealed that residue H176 was in direct contact with the substrate through a tightly packed edge-to-face interaction. Site-directed mutagenesis was conducted at the H176 residue and a few single variants were generated and tested. Of these, the H176F and H176Y variants increased the binding affinity for L-tryptophan by 15- and 8-fold, respectively, compared with the WT. Surprisingly, both variants also completely switched the product

selectivity, showing >99% of 5-nitro-L-troptophan, while the WT gave 98% of 4-nitro-L-troptophan (Fig. 1.28-c).

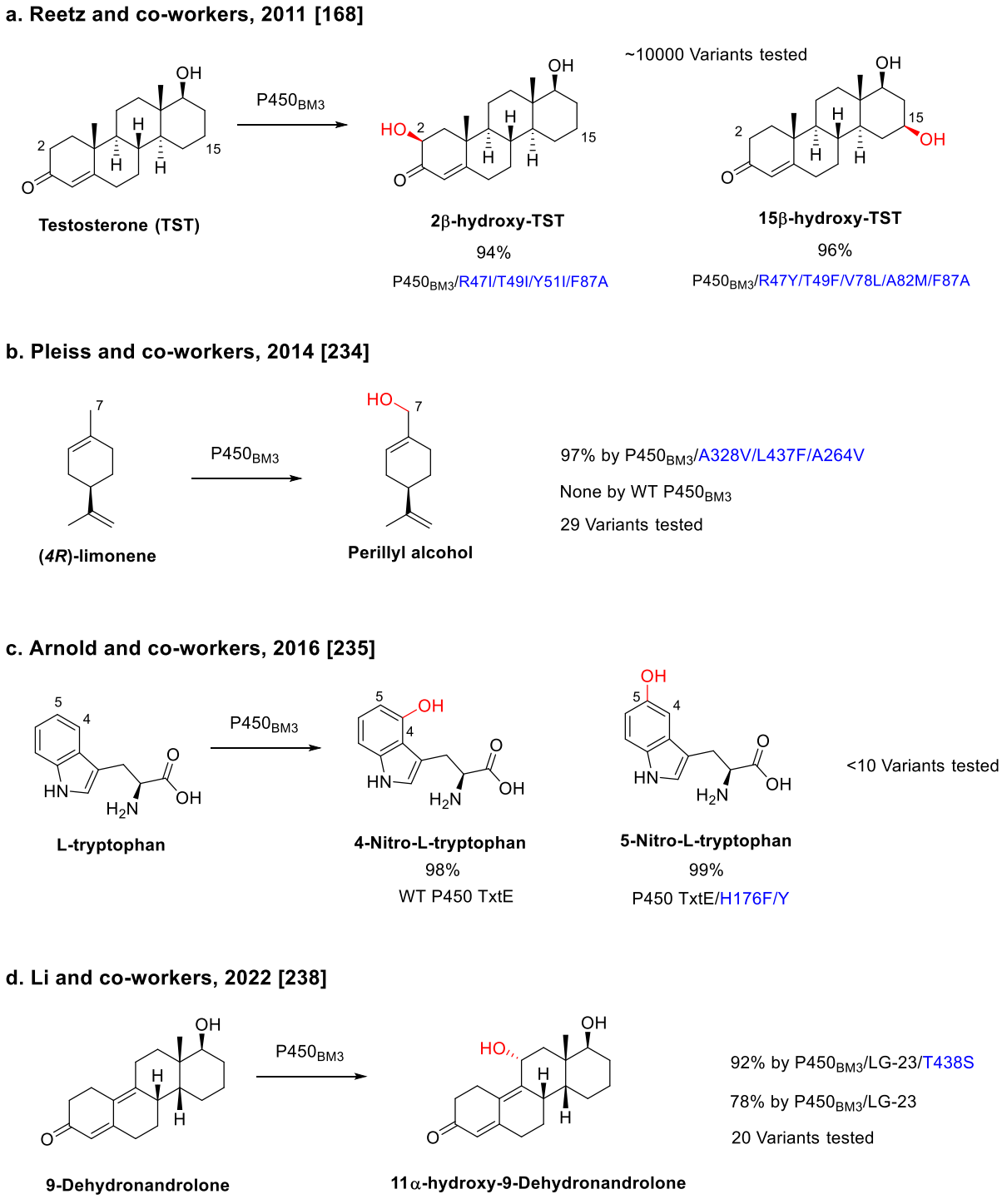


Figure 1.28 Examples of P450 engineering utilising MD simulation and molecular docking as a guide to improve enzymatic activity and product selectivity for target compounds.

Pelletier and co-workers applied the adaptive biasing force (ABF) method to evaluate the substrate binding path of palmitate with $P450_{BM3}$ and therefore identify key residues.[237] The ABF method calculates the mean force along a reaction coordinate describing ligand binding and applies an equal and opposite force at each point of this coordinate. This allows the ligand to escape local energy minima and cross energy barrier in the free-energy landscape of the reaction coordinate. Using this method, a single MD simulation of the palmitate-bound $P450_{BM3}$ identified all the key residues known to be involved in fatty acid binding, demonstrating the accuracy and utility of this approach. In addition, residue Q73, which previously not know to be implicated in binding, was also identified. Molecular docking of a library of 29 palmitate analogues with $P450_{BM3}$ was then conducted. Catalytically relevant binding poses were produced for the entire library. This approach is broadly applicable to predict binding hotspots and has the potential to reduce screening efforts to tailor enzymes for selective functionalisation of desirable products.

In 2022, Li and co-workers applied an *in silico* $P450_{BM3}$ engineering, guided by MD simulation and docking, for selective steroid oxidation.[238] The previously generated variant LG-23 (R47W/S72W/F77Y/V78L/F81I/A82L/F87G/T88S/M177T/M185Q/L188Q/I209T/A328G/A 330W)[172] was a 11α-selective (78%) variant for 9-dehydronandrolone. MD-simulations and molecular docking were then conducted, identifying G87, S88, T260, E267, T327, T438 and L439 as substrate contact residues. Enzyme-substrate contacts in the productive binding mode were compared with those in the non-productive poses. It was noticed that the T438 side chain adopted an inward movement in the non-productive binding mode which abolished the crucial hydrogen bond between the substrate and residues T327 and L439. Site-saturation mutagenesis was the performed at the T438 residue of LG-23 to fine-tune the active pocket. Of these, the LG-23/T438S variant showed greatly improved 11α-selectivity (92% from 78% by LG-23). Only 20 variants were generated to achieved this (Fig. 1.28-d).

To summarise, a few reports have demonstrated the successful applications of MD simulations and molecular docking in P450 engineering, with significantly reduced screening efforts, compared with traditional random mutagenesis or rational design methods. However, most of these approaches remained substrate specific. More generally applicable techniques still await to be established.

1.6 Thesis objective

1.6.1 Selective oxidation of cyclic amines

The overarching aim of this project is the selective remote functionalisation of saturated cyclic amines by engineered $P450_{BM3}$ variants for applications in synthesis, in particular fragment molecule diversification. Library screening with substrate engineering was first deployed to identify key residues and most effective protecting groups for each substrate. Previous work within the Wong and Robertson groups suggested that β-oxidation of mono-ring cyclic amines is difficult to achieve using Boc as the protecting group. Cyclic amine derivatives are therefore prepared and tested with five protecting groups: Boc (-*tert*-Butyloxycarbonyl), Ms (-Methylsulfonyl), Ts (-Toluenesulfonyl), Tf (-Trifluoromethylsulfonyl), and Ips (-Isopropylsulfonyl). These different derivatives were screened against the same panel of $P450_{BM3}$ variants and comparative GC analysis was conducted to identify key mutations and selectivity profiles for each derivative. Docking guided mutagenesis was then explored to target high regio- and enantio-selectivity for remote C–H activation products. The medium size mono-ring amines azepane and azocane, the bicyclic amine 7-azabicyclo[2.2.1]heptane, and the spirocyclic amine 8-azaspiro[5.4]decane are selected as model compounds to carry out these combined strategies. Direct C–H hydroxylation of these molecules using engineered $P450_{BM3}$ has not been reported.

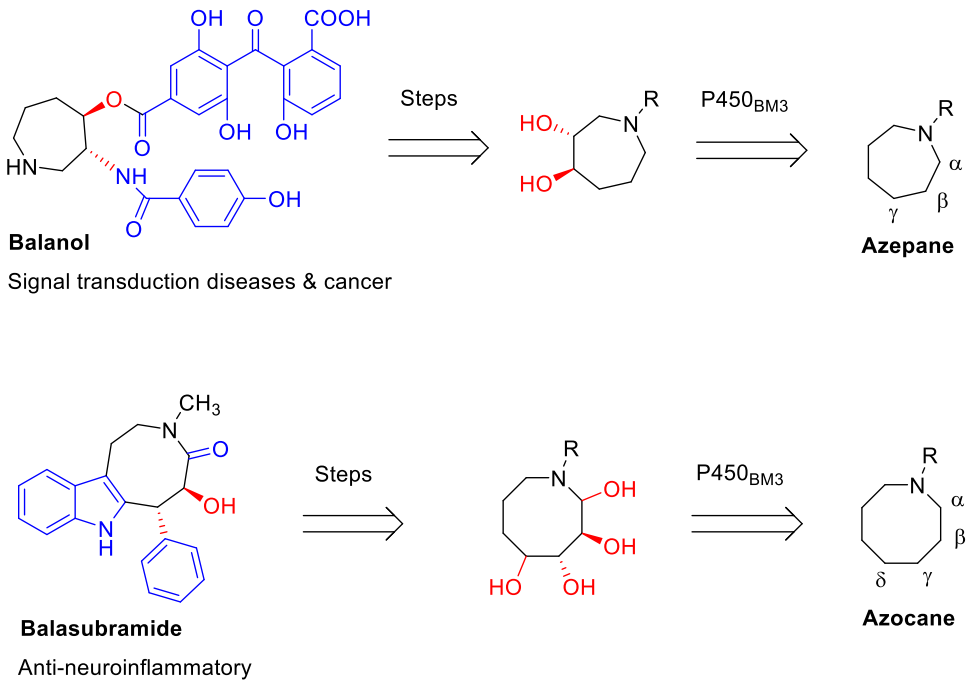
Chapter 2 describes the initial screen, substrate engineering, together with preliminary mutagenesis for azepane and azocane. Medium to high selectivity for the remote functionalisation products were observed. The Ips group is identified to be β oxidation-promoting, while the Boc derivatives show high selectivity for γ alcohols, for azepane and azocane. Chapter 3 describes the systematic analysis of docking results of azepane and azocane with selected variants. The docking-guided mutagenesis approach is explored where the productive binding poses are overlaid with the non-productive ones or the poses for other products to provide information for mutation design. Chapter 4 describes the $P450_{BM3}$ engineering for selective oxidation of the bicyclic and spirocyclic substrates. The combined strategies of library screening, substrate engineering and docking-guided mutagenesis were deployed to achieve high enzymatic activity and product selectivity. For 8-azaspiro[5.4]decane, advanced Mosher ester analysis and X-ray crystallography were explored to determine the stereochemistry of the hydroxylation products.

Chapter 2

2. Azepane and azocane

2.1 Introduction

Azepane and azocane are the core structures of a wide range of natural products and synthesised drugs that are used as insecticides, antibiotics, antifungal, antiviral, anticancer and anti-inflammatory agents.[6,11,239,240] These medium size cyclic amines are increasingly applied in the syntheses of new drugs or drug fragments.[11,239] Selective functionalisation of these amines provides alternative routes to these compounds with fewer steps and under mild conditions. Scheme 2.1 exemplifies the potential applications of P450-mediated selective hydroxylation of azepane and azocane in the synthesis for balanol and balasubramide, two clinically applied drugs.



Scheme 2.1 Potential synthe*tic* routes for balanol and balasubramide through P450-mediated selective hydroxylation of azepane and azocane.

Direct functionalisation at the α positions of azepane and azocane could be accomplished using the various methods described in **1.1.2**. Wu and co-workers[6] applied two stereoselective ketoreductases READH and *Ch*KRED20 for production of both enantiomers of the β- and γ-alcohols of azepane with high ee (up to 99%) from the respective ketones; however, this method requires the ketones to be synthesised. There is currently no report on selective direct hydroxylation at remote positions for azepane and azocane. This chapter explores the $P450_{BM3}$ oxidation of azepane and azocane by library screening and rational design. Substrate engineering is also deployed with the objective of achieving high selectivity for different products while reducing mutagenesis and screening efforts.

The $P450_{BM3}$ oxidation of *N*-Boc-azepane (**1**) has been previously studied within the Wong group. Initial screening showed ≥50% conversion at a substrate-to-enzyme ratio of 2000:1 (TTN ≥1000) to a range of products, some of which were characterised.

2.2 Initial screening with *N*-Boc-azepane

At the beginning of this project, **1** was screened against the 48 variants in the initial library (Table A2.1, Appendix **A2.1**). These variants are derived from five base variants, A330P, K19 (H171L/Q307H/N319Y), RP (R47L/Y51F/I401P), GVQ (A74G/F87V/L188Q), and RT2 (R47L/Y51F/A191T/N239H/I259V/A276T/L353I), which had been shown to have increased activity for the oxidation of a wide range of organic compounds.[75,164,165,174] Additional mutations at 2−4 active site residues (S72, V78, F81, A82, F87, A184, I263, A264, E267, A328, A330, L437 and T438) were introduced (Fig. 2.1).[75,83,162,163,165,174-176] The 48 selected variants in the initial library, as listed in Table A2.1, Appendix **A2.1**, contain different combinations of mutations at these residues to generate diverse binding pocket topologies.

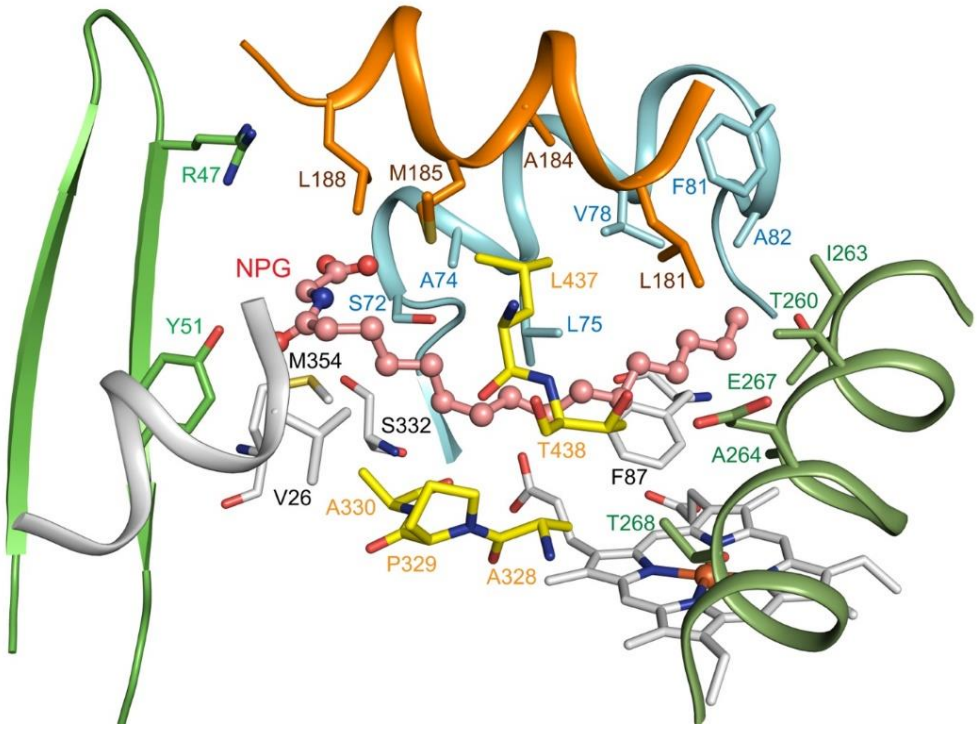


Figure 2.1 The active site structure of *N*-palmitoylglycine (NPG)-bound $P450_{BM3}$ and key amino acid residues.

Screening reactions (0.5 mL, 1 μmol substrate) were carried out at a 2000: 1 substrate to enzyme ratio (2 mM: 2 μM) in 200 mM phosphate buffer (pH 7.9) for 16 hours in 24-well plates. Glucose (100 mM) and glucose dehydrogenase (GDH, 2 U/mL) were added to regenerate the NADPH cofactor. $NADP^+$ (4 μM) was used to initiate the reaction. Reactions were analysed by GC (Appendix **A4.1.1** and **A4.1.2**). Variants showing good activity were used to conduct preparative reactions with the same components as screenings but scaled up by volume (50 mL to 1 L). Products were separated by silica gel column chromatography and characterised.

All major products of **1** were purified and fully characterised by NMR and MS data; these were the γ-alcohol **1a**, the γ-ketone **1b**, the β-alcohol **1c**, the β-ketone **1d**, the α-alcohol **1e** and its ring-opening product **1f** (Fig. 2.2). NMR spectra of these products are in line with literature data and can be found in the appendices (Appendices **A6** and **A7**).

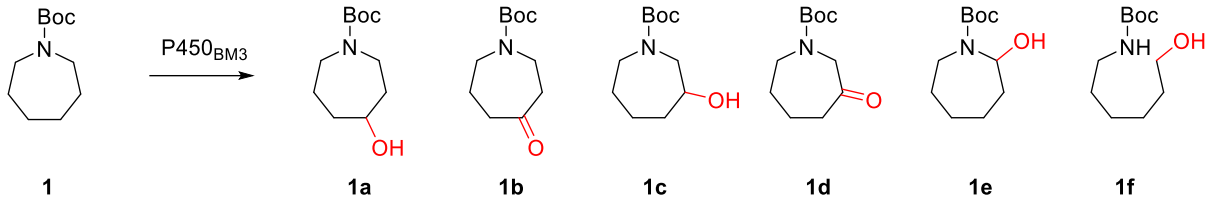


Figure 2.2 Products characterised from the initial screening of *N*-Boc-azepane, **1**.

Chiral GC analysis revealed high regio- and enantioselectivity for the γ-alcohol 4-hydroxy-*N*-Boc-azepane (**1a**) for some variants. The absolute configuration for enantiomers was determined by comparison of the $[\alpha]_D$ with literature data (Table 2.1).[6] The highest (*R*)-selectivity was produced by the variant RP/H171L/I263G/A330W at 70% of **1a** and 75% *ee* for (*R*)-**1a** (Table 2.2, Entry 7). The (*S*)-enantiomer was most favoured by the variant KU3/A330P, giving 92% of **1a** and 75% *ee* for (*S*)-**1a** (Table 2.2, Entry 1). On the other hand, selectivity for the β-alcohol **1b**, remained low, with R19/A330W/S72G being the most selective, affording 48% of **1c** and 90% *ee* for (*R*)-**1c** (Table 2.2, Entry 16). The enantioselectivity for the β-alcohol product by other variants was not further analysed due to low regioselectivity.

Table 2.1 $[\alpha]_D$ of oxidation products and the literature data for pure enantiomers.

Products	Variants	*ee*	$[\alpha]_D^{25}$	Solvent	Literature data
1a γ-alcohol	RLYF/HL/I263G/A330W	90%, ***R***	+5.2°	$CHCl_3$	+5.7° for ***R***[6] –6.7° for ***S***[6]
1a γ-alcohol	RT2/F81W	32%, ***S***	–1.8°	$CHCl_3$	+5.7° for ***R***[6] –6.7° for ***S***[6]
1c β-alcohol	RT2/A330W	98%, ***R***	–24.1°	$CHCl_3$	+33.6° for ***S*** –32.4° for ***R***

The data analysis revealed a few key mutations. The mutation A330P improved *S*-γ-selectivity to 92% of **1a** and 74% *ee* for (*S*)-**1a** by the RT2/A330P variant (Entry 2, Table 2.2), from 39% of **1a** and 38% *ee* for (*S*)-**1a** by RT2 (Entry 12). The A330W mutation completely changed the

product profile as R19/A330W/S72G gave the β-alcohol **1c** as the major product (48% of **1c**) with 20% of the γ-alcohol **1a** (Entry 16), compared with RT2/A330P (2% of **1c** and 91% of **1a**; Entry 2). The residue A330 sits in the active site opposite to F87 and mutations here may have large effects on substrate binding orientation (Fig. 2.1). The mutation S72W brought the preference back towards the *S*-γ-alcohol as the R19/A330W/S72W variant gave 74% of **1a** and 64% *ee* for (*S*)-**1a** (Entry 5), even though R19/A330W/S72G favoured **1c** (Entry 16).

Table 2.2 Product distribution of selected highly active variants for **1** from the initial screening (substrate:enzyme = 2000:1).

Entry	Variants	Conv.	1a	1b	1c	1d	1e&1f	others
1	KU3/A330P	88%	**92% (75% ee, *S*)**	–	4%	–	2%	–
2	RT2/A330P	90%	91% (74% ee, *S*)	–	2%	–	6%	–
3	KU3/A330P/L437S	70%	90% (10% ee, *S*)	–	3%	–	5%	–
4	R19/F87I/L437LV	100%	85% (20% ee, *S*)	–	14%	1%	–	–
5	R19/A330W/S72W	96%	74% (64% ee, *S*)	–	11%	–	15%	–
6	KU3/A330P/A328I	54%	70% (65% ee, *S*)	1%	22%	2%	5%	–
7	RP/H171L/I263G/A330W	50%	**70% (75% ee, *R*)**	–	19%	–	11%	–
8	RP/A82M/I263A	84%	69% (70% ee, *S*)	2%	24%	1%	4%	–
9	R19/F87I	97%	67% (60% ee, *S*)	1%	10%	2%	20%	–
10	RP/H171L/I263G	61%	60% (74% ee, *R*)	–	22%	1%	9%	–
11	RT2/F81W	95%	54% (31% ee, *S*)	2%	21%	–	23%	–
12	RT2	93%	39% (38% ee, *S*)	1%	28%	–	32%	2%
13	R19	86%	32% (50% ee, *S*)	1%	26%	–	39%	2%
14	RP/H171L	89%	26% (18% ee, *S*)	–	32%	–	42%	–
15	WT	40%	22% (20% ee, *S*)	–	24%	–	54%	–
16	R19/A330W/S72G	93%	20% (60% ee, *S*)	1%	**48%**	1%	27%	3%

For the *R* enantiomer, the mutation I263G promoted the selectivity to 60% of **1a** and 74% *ee* for (*R*)-**1a** by RP/H171L/I263G (Entry 10) from 26% of **1a** and 18% *ee* for (*S*)-**1a** by RP/H171L (Entry 14). The I263G/A330W combination further improved the selectivity to 70% of **1a** and 75% *ee* for (*R*)-**1a** by the variant RP/H171L/I263G/A330W (Entry 7). Another potential *R*-γ-

promoting mutation was L437S which, upon addition to KU3/A330P, dropped the *ee* for (*S*)-**1a** from 75% (KU3/A330P) to 10% (KU3/A330P/L437S) (Entries 1 and 3).

The effective mutations were combined to seek further improvement of enantioselectivity. The variants KU3/A330P/S72W, RT2/A330P/S72W and RP/H171L/I263G/A330W/L437S were thus made and tested. KU3/A330P/S72W enhanced the *S*-γ-selectivity to 90% of **1a** and 84% *ee* for (*S*)-**1a** (TTN = 8440). While the RP/H171L/I263G/A330W/L437S variant dramatically improved *R*-γ-selectivity to 90% of **1a** and 90% ee for (*R*)-**1a**, with a TTN for (*R*)-**1a** of 10060 (Fig. 2.3). TTN is the total turnover number for the dominant enantiomer in Figure 2.3 and the remainder of this thesis.

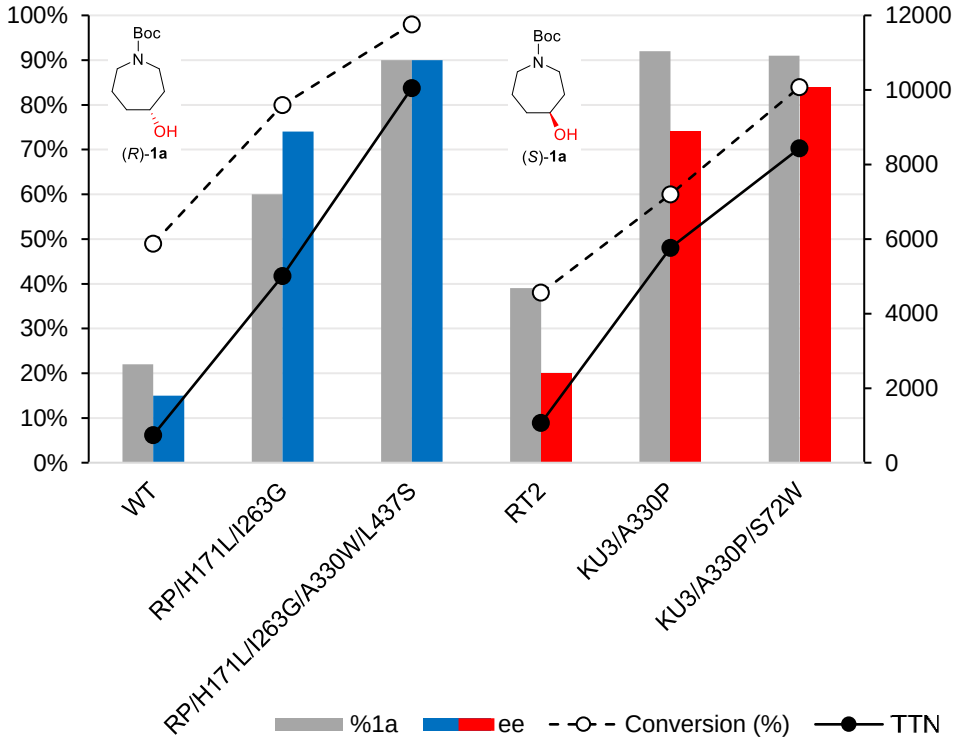


Figure 2.3 Regioselectivity, *ee*, conversion and TTN (total turnover number for the dominant enantiomer) of P450$_{BM3}$ variants for production of **1a**. Grey bars represent the regioselectivity for **1a**; Blue bars represent the *ee* for (*R*)-**1a**; Red bars represent the *ee* for (*S*)-**1a**; Black lines with black dots represent TTN for the dominant enantiomer of **1a**; Dash lines with open circles represent conversion (substrate:enzyme = 12000:1).

However, selectivity for the β-alcohol product remained low (highest 48% by R19/A330W/S72G; Entry 16, Table 2.2) and there was not enough data from the initial

screening, to guide rational design to improve the selectivity. Therefore, screening against even larger library or alternative approaches to engineer selectivity was required.

2.3 Substrate engineering for azepane

To address the issue of low β-selectivity, instead of conducting more exploratory mutagenesis or screening, the approach of substrate engineering was utilised, whereby different modifying groups or substituents for substrates are tested in order to alter the selectivity patterns without further mutagenesis. Azepane derivatives were therefore prepared with four protecting groups: Ms ($–SO_2CH_3$), Ts ($–SO_2$-*p*-tolyl), Tf ($–SO_2CF_3$) and Ips [$–SO_2CH(CH_3)_2$]. These derivatives were then screened against the same panel of 48 variants as **1** (Initial library; Table A2.1, Appendices **A2.1**) and product profiles and selectivity trends were analysed. Some further mutagenesis was also conducted based on analysis of the experimental data.

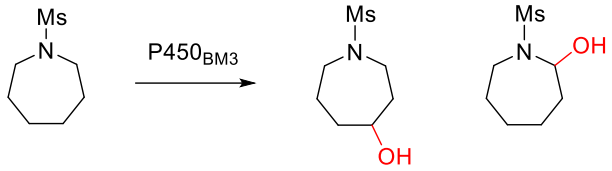
2.3.1 *N*-Ms-azepane

The relevant initial screening and product characterisation of *N*-Ms-azepane **2** was carried out by Dr. Xinxin Zhang of Jeremy Robertson's group. **2** was a good substrate, with 23 out of 48 variants showing >50% conversion. There were only two products isolated and characterised: the γ alcohol **2a** and the α alcohol **2b**. From the initial screening, the α alcohol was the predominant product: 21 out of the 23 active variants gave >70% of **2b**, with K19/F87A/A82M being the most selective at 91% (Entry 10, Table 2.3). Most of the key mutations in the initial library, *e.g.*, A184I (Entry 3), F87V (Entry 4), F81W (Entry 5), I263A (Entry 6), A330W (Entry 7), F87I (Entry 8), F87A (Entry 9), and A81M (Entry 10), favoured α oxidation of **2**.

The highest regioselectivity for the γ alcohol **2a** was observed with RP/H171L/I263G at 26% (Entry 1). The KU3/A330P variant showed similar selectivity (24%) for **2a**, and these two were

the only ones across the initial screening that gave >20% regioselectivity for γ oxidation. No β oxidation was identified. The *N*-Ms derivative was thus not pursued further.

Table 2.3 Product distribution of selected highly active variants for N-Ms-azepane **2** from the initial screening (substrate:enzyme = 2000:1).



2 **2a** **2b**

Entry	Variants	Conv.	2a	2b	others
1	RP/H171L/I263G	88%	**26%**	67%	7%
2	KU3/A330P	85%	24%	68%	8%
3	GV/A184I	90%	18%	82%	–
4	GVQ	85%	16%	82%	2%
5	RT/F81W	96%	16%	84%	–
6	RP/A82M/I263A	54%	12%	85%	3%
7	R19/A330W	70%	12%	88%	–
8	R19/F87I	84%	10%	88%	2%
9	K19/F87A	97%	8%	90%	2%
10	K19/F87A/A82M	65%	6%	**91%**	3%

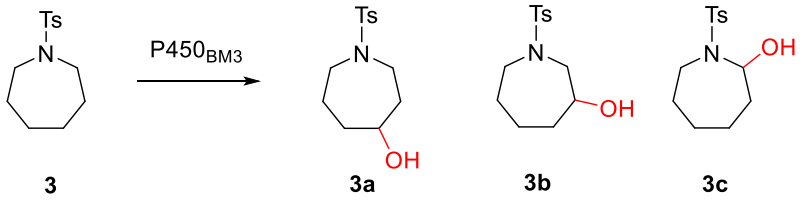
2.3.2 *N*-Ts-azepane

N-Ts-azepane (**3**) was a moderate substrate: there were 16 variants showing >50% conversion from the initial screening. These gave two remote functionalisation products, the γ alcohol **3a** and the β alcohol **3b**, together with the α alcohol **3c**.

The γ alcohol was generally less favoured compared with the *N*-Boc derivative **1**, with the highest γ-selectivity observed with the KU3/A330P variant at 56% of **3a** (Entry 1, Table 2.4), while the same variant gave 92% of the γ alcohol **1a** for **1** (Entry 1, Table 2.2). Variants RP/H171L/I263G (55% of **3a**; Entry 2, Table 2.4) and RP/A82M/I263A (51%; Entry 3, Table 2.4) also showed >50% regioselectivity for **3a**. The I263G mutation was very effective to

promote γ oxidation as demonstrated by the comparison between R19/F87A/I263G (40% of **3a**; Entry 4, Table 2.4) and R19/F87A (5% of **3a**; Entry 5). The combination KU3/A330P/I263G could lead to even higher selectivity for **3a**.

Table 2.4 Product distribution of selected highly active variants for N-Ts-azepane **3** from the initial screening (substrate:enzyme = 2000:1).



3 **3a** **3b** **3c**

Entry	Variants	Conv.	3a	3b	3c	others
1	KU3/A330P	70%	**56%**	25%	19%	–
2	RP/H171L/I263G	75%	55%	25%	20%	–
3	RP/A82M/I263A	80%	52%	27%	21%	–
4	R19/F87A/I263G	85%	40%	35%	23%	3%
5	R19/F87A	54%	5%	15%	**80%**	–
6	R19	70%	30%	40%	28%	2%
7	RT2	64%	22%	47%	26%	5%
8	K19/F87V	80%	22%	50%	24%	4%
9	R19/A330W	80%	30%	55%	10%	5%
10	R19/A330W/S72G	65%	20%	**60%**	17%	–

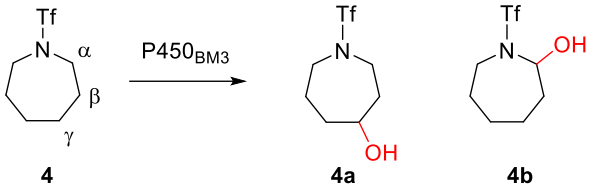
The β alcohol was most favoured by the R19/A330W/S72G variant, giving 60% of **3b** (Entry 10, Table 2.4). This regioselectivity was higher than the highest β selectivity observed with the *N*-Boc derivative **1** at 48% by the same variant (Entry 16, Table 2.2). Variants R19/A330W (Entry 9, Table 2.4) and K19/F87V (Entry 8, Table 2.4) also showed >50% regioselectivity for **3b**. The combination R19/A330W/S72G/F87V might further improve β selectivity.

2.3.3 *N*-Tf-azepane

High activity was observed across the initial screening for *N*-Tf-azepane (**4**), with 42 out of 48 variants from the library showing >50% conversion. Two major products were characterised:

the γ alcohol **4a** and the α alcohol **4b**. As with the *N*-Boc derivative, the γ oxidation products **4a** was predominant in the initial screening; 24 variants showed >50% regioselectivity for the γ alcohol, with the RP/H171L/I263G variant giving 94% of **4a** as the highest (Entry 1, Table 2.5). Variants KU3/A330P (Entry 2), R19/A330W/S72G (Entry 3), R19/F87A/I263G (Entry 4), and RP/A82M/I263A (Entry 5) also gave >80% of **4a**. On the other hand, the α alcohol was most favoured by the R19/F87A variant (80% of **4b**; Entry 10). Other highly α selective variants included GV/A184I (70% of **4b**; Entry 9) and R19/F87I (62%; Entry 8). *N*-Tf-azepane (**4**) was not further investigated since no β oxidation was observed, and high selectivity for γ alcohol has been observed with **1**.

Table 2.5 Product distribution of selected highly active variants for N-Tf-azepane **4** from the initial screening (substrate:enzyme = 2000:1).



4 **4a** **4b**

Entry	Variants	Conv.	4a	4b	others
1	RP/H171L/I263G	80%	**94%**	3%	3%
2	KU3/A330P	85%	91%	5%	4%
3	R19/A330W/S72G	60%	90%	8%	2%
4	R19/F87A/I263G	70%	85%	15%	–
5	RP/A82M/I263A	54%	80%	15%	5%
6	WT	54%	60%	35%	5%
7	R19/A330W	70%	45%	50%	5%
8	R19/F87I	84%	35%	62%	3%
9	GV/A184I	80%	20%	70%	10%
10	R19/F87A	75%	10%	**80%**	10%

2.3.4 *N*-Ips-azepane

N-Ips-azepane (**5**) was a relatively poor substrate for $P450_{BM3}$, compared with the *N*-Boc- (**1**) and *N*-Tf- (**4**) derivatives. Of the panel of 48 variants, 16 showed >50% conversion. All possible remote functionalisation products were characterised: the γ-alcohol **5a**, the γ-ketone **5b**, the β-alcohol **5c**, and the β-ketone **5d**. The α-alcohol **5e** was also formed (Fig. 2.4).

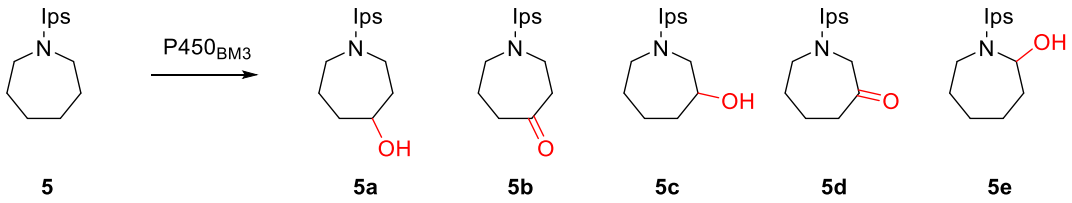


Figure 2.4 Products characterised from initial screening of N-Ips-azepane **5**.

From the initial screening, the mutation F87A favoured α-oxidation: R19/F87A gave 90% of the α-alcohol **5e** (Entry 16, Table 2.6). The mutations A330P and I263G showed great preference for γ-oxidation as the KU3/A330P gave 92% of the γ-alcohol **5a** (Entry 14) and the single mutation variant I263G afforded 97% **5a** (Entry 15). The mutation F87V promoted β-oxidation: variants GVQ (67% of the β-alcohol **5c**) and RP/F87V (64% of **5c**) both showed high β-selectivity (Entries 3 and 5). Furthermore, the variant K19/F87V/I263G favoured β-oxidation (65% of **5c**), despite the γ-oxidation preference of the mutation I263G (Entry 4).

In general, compared with the other four derivatives, **5** showed significant selectivity for β-hydroxylation, with 11 out of the 16 highly active variants giving ≥50% of the β-alcohol **5c** (Table 2.6). The highest regioselectivity for **5c** was produced by the variant GVQ/I263G at 70% (Entry 1, Table 2.6), an increase of 10% compared with the highest of 60% by R19/A330W/S72G with the *N*-Ts derivative **3** (Entry 10, Table 2.4).

Table 2.6 Product distribution of the 16 variants showing >50% conversion for **5** from initial screening (substrate:enzyme = 2000:1). ***A*** denotes other unknown products.

Entry	Variants	Conv.	5c	5d	5a	5b	5e	*A*
1	GVQ/I263G	63%	**70% (60% ee, *S*)**	–	7%	3%	6%	–
2	GVQ/I263G/A328I	58%	68%	–	22%	1%	9%	–
3	GVQ	67%	67% (75% ee, *R*)	–	10%	7%	14%	2%
4	K19/F87V/I263G	75%	65%	–	14%	8%	11%	2%
5	RP/F87V	89%	64%	2%	17%	6%	11%	1%
6	R19/A330W	71%	60% (95% ee, *R*)	2%	16%	8%	14%	2%
7	K19/F87V/P329G/A330W	97%	52%	–	18%	3%	27%	–
8	RP/H171L/I263G/F87V	100%	51%	2%	17%	20%	9%	1%
9	GVQ/A328G	58%	51% (43% ee, *S*)	–	15%	3%	27%	4%
10	GV/A184I/I263G	65%	50% (61% ee, *S*)	2%	15%	29%	2%	2%
11	GV/A184I	93%	50%	–	15%	17%	16%	2%
12	RP/S72G	85%	50% (65% ee, *R*)	–	20%	1%	29%	–
13	GAQ/I263G	97%	38%	2%	19%	1%	38%	–
14	KU3/A330P	100%	8%	–	92%	–	–	–
15	I263G	98%	3%	–	**97%**	–	–	–
16	R19/F87A	93%	3%	–	4%	–	**90%**	2%

Chiral phase GC analysis was performed for $P450_{BM3}$ oxidation products of **5** to determine the enantioselectivity of each variant. Moderate selectivity for both enantiomers were observed: high *R*-β-selectivity was produced by the variants GVQ [67% of **5c** and 75% *ee* for (*R*)-**5c**] and R19/A330W [60% of **5c** and 95% *ee* for (*R*)-**5c**] (Table 2.7). The *S*-enantiomer was most favoured by GVQ/I263G at 70% of **5c** and 60% ee for (*S*)-**5c** (Table 2.7). For the *N*-Boc derivative **1**, the highest β-selectivity was displayed by the variant R19/A330W at 48% of **1c** and 92% ee for (*R*)-**5c**. No *S*-β-favouring variant was found for **1**, as chiral phase GC analysis was not performed due to low regioselectivity. Variants GVQ and GVQ/I263G both showed little β-oxidation of **1**(Table 2.7).

Table 2.7 Selectivity for β-oxidation of some variants for N-Boc- and N-Ips-azepane

Variants	***N*-Boc-azepane, 1**		***N*-Ips-azepane, 5**	
	%1c	***ee***	**%5c**	***ee***
GVQ	4%	–	67%	75%, *R*
R19/A330W	48%	92%, *R*	60%	95%, *R*
GVQ/I263G	6%	–	70%	60%, *S*

2.3.4 Further mutagenesis for selective β-oxidation of *N*-Ips-azepane

Among the five derivatives tested, *N*-Ips-azepane (**5**) was selected for further investigation to achieve high regio- and enantioselectivity for β-functionalisation of azepane.

The selectivity data from initial screening revealed some key mutations, *e.g.*, F87V, S72G and A330W (Entries 3, 12 and 6, Table 2.6), that improved selectivity for the β-alcohol **5c** and showed preference for the *R*-isomer: the variant GVQ gave 67% of **5c** at 75% *ee* for (*R*)-**5c** (Entry 3), compared with the precursor variant GQ which showed no conversion at all. The same trend was observed from the comparison between the variants RP/S72G [50% of **5c** and 65% *ee* for (*R*)-**5c**; Entry 12] and RP (no conversion) and between RT2/A330W [60% of **5c** and 95% *ee* for (*R*)-**5c**; Entry 6] and RT2 (no conversion). Combination of these effective mutations were then tested (Fig. 2.5). Among these, the variant GVQ/A330W increased selectivity to 82% of **5c** and 90% *ee* for (*R*)-**5c** (TTN = 1250), and VQ/S72G/A330W showed the highest *R*-β-selectivity at 91% of **5c** and 96% *ee* for (*R*)-**5c** (ON = 1320) (Fig. 2.5).

High selectivity for the *S* enantiomer was observed to be promoted by the mutations I263G and A328G. The I263G mutation led to a striking inversion of enantioselectivity, with GVQ/I263G giving 60% *ee* for (*S*)-**5c** (Entry 1, Table 2.6), compared to 75% *ee* for (*R*)-**5c** by GVQ (Entry 3). The mutation A328G showed similar effect: the variant GVQ/A328G affording 51% of **5c** with 43% ee for *S* (Entry 9). The GVQ/I263G/A328G variant was attempted to be made to

combine these effective mutations. However, this variant failed to express, instead, the GV/A184I/I263G/A328G variant was generated and found to give higher *S*-β-selectivity at 76% **5c** and 79% *ee* for (*S*)-**5c** (TTN = 430). Since the variant GV/A184I/I263G gave 65% **5c** at 61% *ee* for (*S*)-**5c**, the mutation A328G was also important for *S*-β-selectivity in GV/A184I/I263G/A328G (Fig. 2.5).

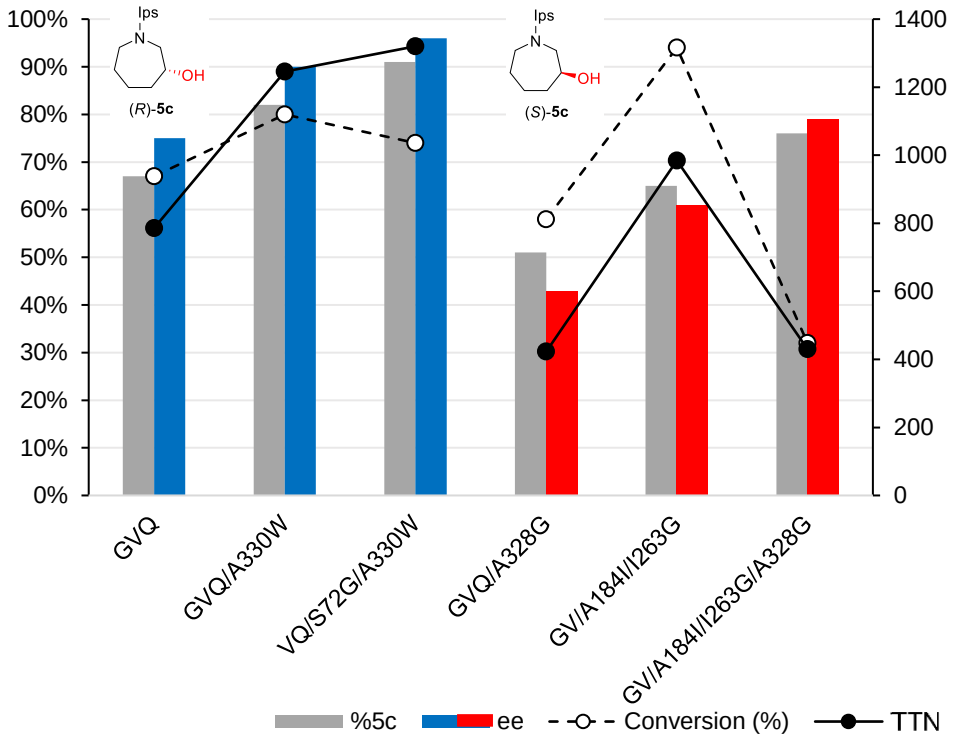


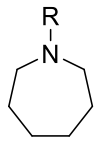
Figure 2.5 Regioselectivity, *ee*, conversion and TTN (total turnover number for the dominant enantiomer) of P450$_{BM3}$ variants for production of **5c**. Grey bars represent the regioselectivity for **5c**; Blue bars represent the *ee* for (*R*)-**5c**; Red bars represent the *ee* for (*S*)-**5c**; Black lines with black dots represent TTN for the dominant enantiomer of **5c**; Dash lines with open circles represent conversion (substrate:enzyme = 2000:1).

2.3.5 Summary of substrate engineering for azepane oxidation

The substrate engineering approach led to significant improvements in activity and selectivity by the same panel of 48 variants. Among the five derivatives tested, *N*-Boc- and *N*-Tf-azepane were good substrates for P450$_{BM3}$, with 40 and 42 variants out of 48, respectively, showing >50% conversion. The other three derivatives showed lower conversion generally, with 23 variants

for *N*-Ms-azepane, 16 for *N*-Ts-azepane, and 16 for *N*-Ips-azepane showing >50% conversion (Table 2.8).

Table 2.8 Number of variants showing oxidation at each position for each protected derivative of azepane upon the initial screening. Each derivative was tested using the same 48 variants in the initial library. Pro-α denotes variant showing >50% selectivity for the α alcohol product, etc.

	Derivative	No. Pro-α	No. Pro-β	No. Pro-γ	others	No. active variants
	*N*-Boc	8	2	**24**	6	40
	N-Ms	**21**	0	0	2	23
	N-Ts	5	5	4	2	16
	N-Tf	12	0	**24**	6	42
	N-Ips	2	**11**	3	0	16

The *N*-Boc- and *N*-Tf- derivatives showed mostly γ-hydroxylation, the *N*-Ms derivative showed predominantly α-hydroxylation, whereas the *N*-Ts derivative showed no preference for oxidation at any position. The *N*-Ips derivative showed significant selectivity for β-hydroxylation, with 11 variants of the enzyme panel giving ≥50% of the β-alcohol **5c** (Table 2.8). This provided alternative solution to tackle the issue of low selectivity for β-hydroxylation of azepane.

The data analysis revealed some key mutations (Fig. 2.6). The mutation F87A gave dominantly α-oxidation for both *N*-Boc-azepane **1** and *N*-Ips-azepane **5**; F87V promoted β-oxidation and favoured the (*R*)-enantiomer for *N*-Ips-azepane but gave mainly α-oxidation for the *N*-Boc-derivative. The F87V/I263G combination switched the preference towards (*S*)-β-oxidation for *N*-Ips-azepane. The A330W mutation favoured (*R*)-β-oxidation for both derivatives (Fig. 2.6).

For the γ-oxidation products, the mutation I263G showed high selectivity for the (*R*)-γ-alcohols for both *N*-Boc- and *N*-Ips-azepane. The combinations I263G/A330W and

I263G/A330W/L437S led to improvements on the (*R*)-γ-selectivity. The mutations A330P and S72W promoted the (*S*)-enantiomer for *N*-Boc-azepane (Fig. 2.6).

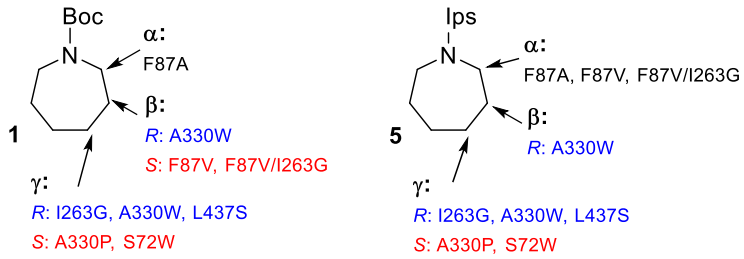


Figure 2.6 Key mutations that favoured products for *N*-Boc-azepane **1** and *N*-Ips-azepane **5**.

2.4 Azocane oxidation

2.4.1 *N*-Boc-azocane

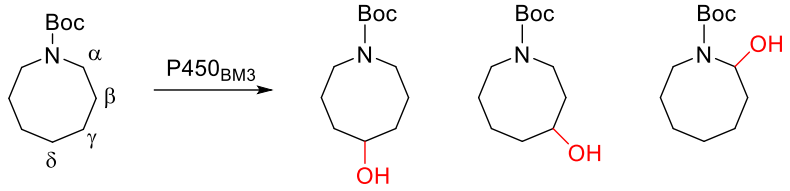
The effective variants generated in the azepane work were added to the initial library and some previous ones removed, leading to a refined library (Table A2.2, Appendices **A2.1**). *N*-Boc-azocane **6** was synthesised and tested against this refined library of 48 variants. High enzymatic activity was observed across the screening, with 30 out of the 48 variants showing ≥50% conversion at a 2000:1 substrate-to-enzyme ratio. Two remote functionalisation products, the δ alcohol **6a** and the γ alcohol **6b**, together with the α alcohol **6c**, were characterised (Table 2.9). Chiral GC analysis revealed high regio- and enantio-selectivity for the γ oxidation product **6b**. Since specific rotation data for **6b** was not available, the stereochemistry was not determined. However, tentative assignments were made by comparing the specific rotation of products for *N*-Boc-azepane **1** and *N*-Boc-azocane **6**. Product **6b** generated by the KU3/A330P/S72W variant ($[\alpha]_D = -9.2$) was tentatively assigned as the (*S*)-γ alcohol, as the γ alcohol **1a** produced

by the same variant (84% ee for *S*) showed a $[\alpha]_D$ of –5.0 (Table A3.1, Appendix **A3.1**). The γ alcohol **6b** from RLYF/I263G/A330P ($[\alpha]_D$ = +11.9) was assigned as (*R*)-γ in comparison with **1a** from RLYF/H171L/I263G/A330W (90% ee for *R*, $[\alpha]_D$ = +5.2; Table A3.1, Appendix **A3.1**). Stereochemistry and enantioselectivity of products from other variants were determined by comparative chiral GC analysis using these two as reference.

Similar as the observation with **1**, γ oxidation dominated the initial screening of **6**, with 25 out of the 30 active variants showing >50% regioselectivity for **6b**. The *S*-γ alcohol was most favoured by the KU3/A330P/S72W variant, giving 74% of **6b** and 77% ee (Entry 10, Table 2.9). Mutations F87I (75% ee; Entry 8) and A330W (67% ee; Entry 7) also showed *S*-γ-promoting effect, although both reduced the regioselectivity, to 45% and 36%, respectively. The addition of F87I to the KU3/A330P/S72W variant could lead to further improved ee for *S*, likely with the trade-off of decreased regioselectivity for **6b**.

The highest *R*-selectivity was observed with the RLYF/I263G/A330P, at 90% of **6b** and 98% ee (Entry 1, Table 2.9). The I263G mutation was highly *R*-γ-promoting, with RP/H171L/I263G giving 67% of **6b** and 95% ee. It was worth noting that the addition of I263G to the *S*-selective variant KU3/A330P (72% of **6b** and 70% ee for *S*; Entry 9) resulted in inversion of enantioselectivity to 96% ee for *R* by the KU3/A330P/I263G variant (Entry 3), while keeping the regioselectivity at 70% of **6b**. The S72G mutation also increased *R*-γ-selectivity as its addition to the RP/H171L/I263G variant (67% of **6b** and 95% ee for *R*; Entry 4) improved both the regio- and enantioselectivity to 75% of **6b** and 98% ee for *R* by RP/H171L/I263G/S72G (Entry 2). The RLYF/I263G/A330P/S72G variant might show further increased selectivity for *R*-**6b**.

Table 2.9 Product distribution of selected highly active variants for N-Ts-azepane **3** from the initial screening (substrate:enzyme = 2000:1).



6 **6a** **6b** **6c**

Entry	Variants	Conv.	6a	6b	6c	others
1	RLYF/I263G/A330P	90%	10%	**90% (98% ee, *R*)**	–	–
2	RP/H171L/I263G/S72G	75%	18%	75% (98% ee, *R*)	7%	–
3	KU3/A330P/I263G	60%	28%	70% (96% ee, *R*)	5%	–
4	RP/H171L/I263G	80%	28%	67% (95% ee, *R*)	5%	–
5	WT	54%	**35%**	26% (N.D.)	32%	7%
6	R19/F87A	90%	22%	30% (N.D.)	**45%**	2%
7	R19/A330W	65%	34%	36% (67% ee, *S*)	28%	2%
8	R19/F87I	70%	32%	45% (75% ee, *S*)	16%	6%
9	KU3/A330P	80%	20%	72% (70% ee, *S*)	8%	–
10	KU3/A330P/S72W	65%	17%	**74% (77% ee, *S*)**	17%	–

The δ alcohol was most favoured by WT $P450_{BM3}$, giving 35% of **6a** (Entry 5, Table 2.9). This could potentially be related to the fact that WT catalyses the terminal oxidation of fatty acids. The highest regioselectivity for the α alcohol was displayed by the R19/F87A variant at 45% of **6c** (Entry 6). Since no β alcohol was characterised throughout the screening with the refined library, substrate engineering was deployed to alter the selectivity patterns.

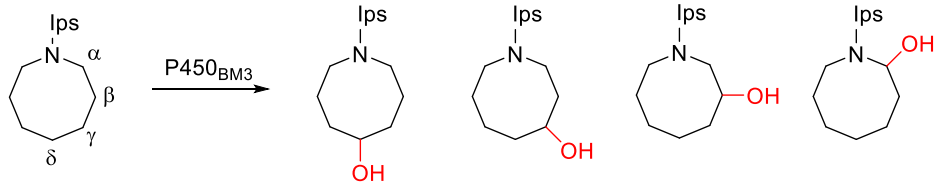
2.4.2 *N*-Ips-azocane

To address the issue with β-selectivity, *N*-Ips-azocane **7** was prepared and studied, based on the results from the azepane work that the *N*-Ips derivative revealed high β-preference. Similar to *N*-Ips-azepane **5**, **7** was a poor substrate for P450 oxidation, with only 10 variants out of the 48 from the refined library showing >50% conversion. All three remote hydroxylation products,

the δ alcohol **7a**, the γ alcohol **7b**, and the β alcohol **7c**, together with the α alcohol **7d**, were characterised (Table 2.10).

Compared with the *N*-Boc analogue **6**, significant preference for the β oxidation product was observed with the *N*-Ips derivative **7**; six variants, R19/A330W/S72A (85%), R19/A330W/S72G (84%), GV/A184I/I263G/A328G (74%), GVQ/I263A/A328G (67%), GVQ/I263G (66%), and GV/A184I/I263G (60%), showed >50% regioselectivity for **7c** (Entries 1–6, Table 2.10). The γ alcohol **7b** was most favoured by the KU3/A330P/S72W variant at 60% selectivity (Entry 10). The highest selectivity for the symmetric, δ oxidation product was observed with the R19/A328L/I263G variant at 95% of **7a** (Entry 11). The α alcohol was most favoured by the R19/F87A (75% of **7d**; Entry 8), as found with azepane derivatives and *N*-Boc-azocane,

Table 2.10 Product distribution of selected variants for N-*Ips*-azo*c*ane ***7*** from the initial screening (substrate:enzyme = 2000:1).



7 **7a** **7b** **7c** **7d**

Entry	Variants	Conv.	7a	7b	7c	7d	others
1	R19/A330W/S72A	60%	10%	5%	**85%**	–	–
2	R19/A330W/S72G	55%	9%	7%	84%	–	–
3	GV/A184I/I263G/A328G	25%	11%	8%	74%	5%	–
4	GVQ/I263A/A328G	80%	18%	5%	67%	10%	–
5	GVQ/I263G	54%	17%	8%	66%	9%	–
6	GV/A184I/I263G	60%	18%	10%	60%	12%	–
7	GVQ/A330W	65%	20%	12%	35%	31%	2%
8	R19/F87A	80%	14%	4%	2%	**75%**	5%
9	RP/H171L/I263G	70%	25%	45%	15%	9%	6%
10	KU3/A330P/S72W	65%	28%	**60%**	10%	2%	–
11	R19/A328L/I263G	60%	**95%**	5%	–	–	–

Stereochemistry and enantioselectivity analysis were then carried out for the β alcohol **7c**. No literature specific rotation data could be found for **7c**. However, single crystals of **7c** were successfully obtained using the vapour diffusion method: purified **7c** (10 mg) was dissolved in ethyl acetate (~500 μL) in a small vial, which was placed into a larger container filled with hexane. The larger container was sealed and left still for several days to afford colourless crystals. This enabled X-ray crystallographic study of **7c** to determine the absolute configuration. X-ray diffraction data was collected and analysed by Dr. Kirsten Christensen of the Department of Chemistry, University of Oxford. The major enantiomer of alcohol **7c** produced by the R19/A330W/S72A variant was determined to have an (*R*)-configuration (Figure 2.7 and Table A3.2, Appendices **A3.2**). The chiral GC analysis for this sample was used as reference to assign the stereochemistry and enantioselectivity for **7c** produced by other variants.

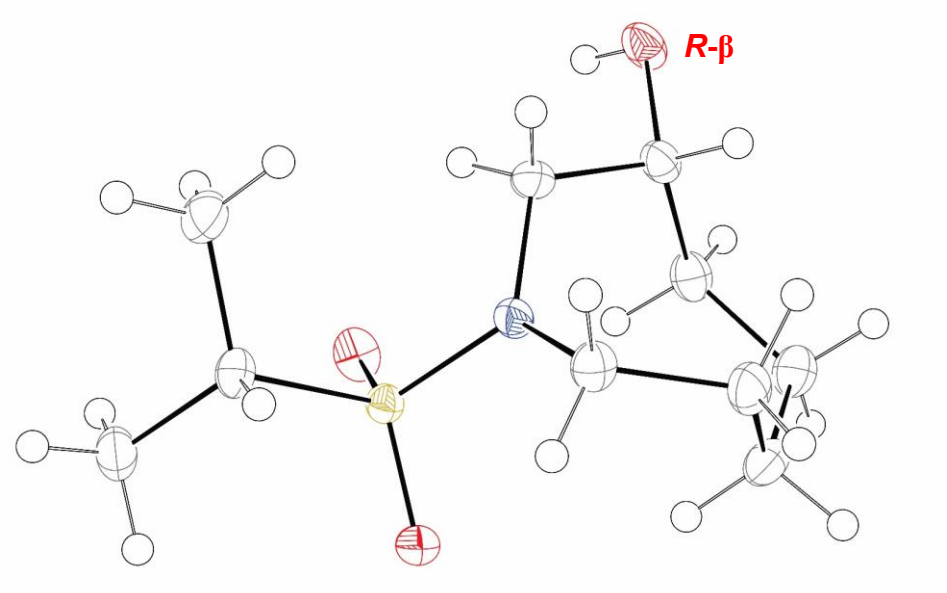


Figure 2.7 The ellipsoid plot for the crystal structure of **6c** produced by R19/A330W/S72A, as analysed from the X-ray diffraction data of the single crystal. This alcohol product is determined as the (*R*)-enantiomer.

Chiral GC analysis revealed high regio- and enantioselectivity for **7c**. The *R*-β alcohol was most favoured by the R19/A330W/S72A and R19/A330W/S72G variants, each giving ~85% of **7c**

at 99% ee (Entries 1 and 2, Figure 2.8). The F87V mutation decreased selectivity for β alcohol but maintained the ee for *R*, with the GVQ/A330W variant giving 35% of **7c** and 90% ee (Entry 3). By contrast, the I263G mutation improved regioselectivity but decreased ee for *R*, with the GVQ/I263G variant showing 66% of **7c** and 20% ee (Entry 4). The A184I mutation further decreased ee for *R* while maintaining the regioselectivity: the GV/A184I/I263G variant gave 60% of **7c** and 12% ee (Entry 5). Finally, addition of an A328G mutation to GV/A184I/I263G completely switched the enantioselectivity to *S* while increasing regioselectivity for β, with the GV/A184I/I263G/A328G variant affording 74% of **7c** and 77% ee for *S*–the highest *S*-β selectivity in the initial screening (Entry 8).

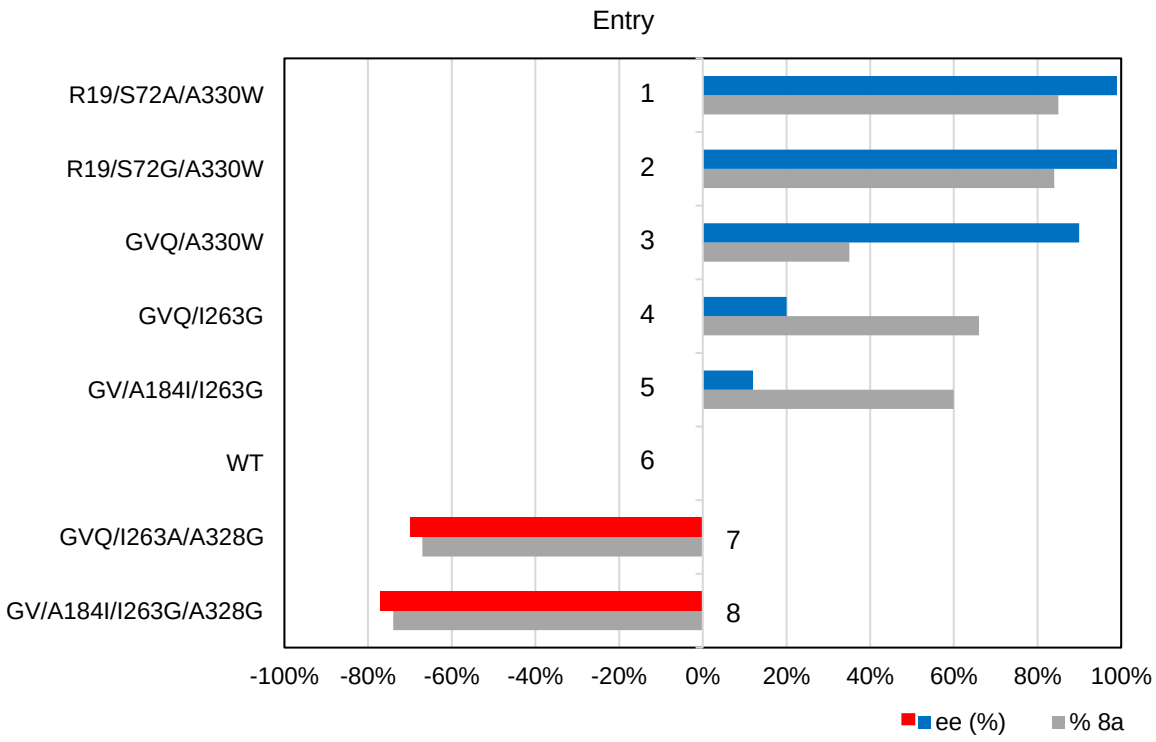


Figure 2.8 Regio- and enantio-selectivity for (R)- and (S)-**6c** of $P450_{BM3}$ variants from the initial screen. Positive values represent selectivity for (R)-**8b**. Negative values represent selectivity for (S)-**8b**.

2.5 Summary for azepane and azocane

The combined strategy of library screening and substrate engineering led to medium to high regio- and enantio-selectivity for all remote hydroxylation products of azepane and azocane. Similar selectivity trends were observed with both substrates. The *N*-Boc derivatives of both azepane (**1**) and azocane (**6**) revealed preferences for γ-oxidation upon screening with the initial library. The highest *S*-γ-selectivity was observed with KU3/A330P/S72W at 90% of **1a** and 84% ee for **1**, and with the same variant at 74% of **6b** and 77% ee for **6**. The *R*-γ alcohol was most favoured by the RLYF/H171L/I263G/A330W/L437S variant at 90% of **1a** and 90% ee for **1**, and by RLYF/I263G/A330P at 90% of **6b** and 98% ee for **6**. However, low β-selectivity was observed with *N*-Boc-azepane (highest at 48% regioselectivity by the R19/A330W/S72G variant), and no β oxidation was observed for *N*-Boc-azocane.

Substrate engineering was then deployed with the objective to alter the selectivity patterns of the 48 enzymes in the variant library. To this end, the *N*-Ms, *N*-Ts, *N*-Tf, and *N*-isopropylsulfonyl (*N*-Ips) derivatives of azepane were tested against the initial library. Different selectivity trends were observed for these derivatives against the same library of variants: the *N*-Ms derivative showed predominantly α oxidation, the *N*-Ts derivative showed no preference for any position and low selectivity, the *N*-Tf derivative favoured the γ alcohol, and the *N*-Ips derivative demonstrated significant preference for β oxidation. *N*-Ips-azocane was then prepared and tested, revealing similar β-preference. The highest *R*-β-selectivity was afforded by the VQ/S72G/A330W variant at 91% of **5b** and 96% ee for *N*-Ips-azepane **5**, and by R19/A330W/S72A at 85% of **7c** and 99% ee for *N*-Ips-azocane **7**. The *S*-enantiomer was most favoured by the GV/A184I/I263G/A328G variant at 76% of **5c** and 79% ee for the azepane *N*-Ips-derivative **5**, and the same variant gave 74% of **7c** and 77% ee for *N*-Ips-azocane **7** (Fig. 2.9).

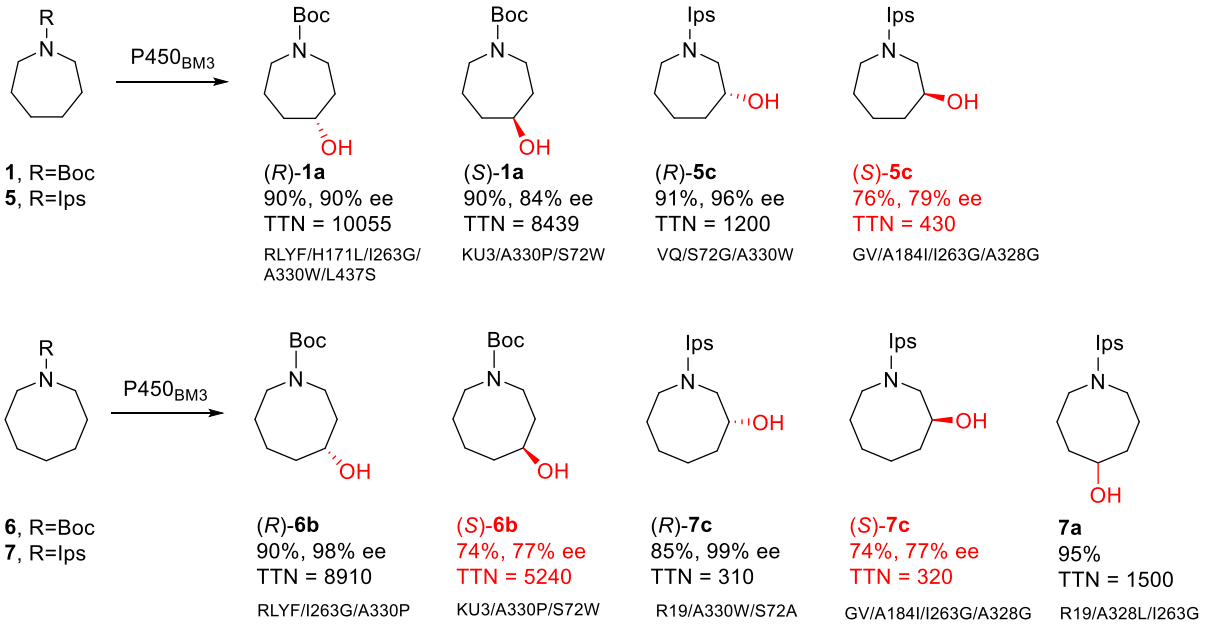


Figure 2.9 Highest selectivity and TTN achieved for all remote oxidation products of azepane and azocane and relevant variants by library screening and substrate engineering. Marked in red are the product for which further improvement in enantioselectivity is required.

High regio- and enantioselectivity were achieved for most of the remote oxidation products, except for the *S*-β alcohol for azepane and the *S*-γ and *S*-β alcohols for azocane (Fig. 2.9). To address these, a docking-guided mutagenesis approach was explored to gain insight into substate-enzyme binding interactions which might be important for mutation design and to reduce the screening effort often required for random mutagenesis.

Chapter 3

3. Docking-guided mutagenesis for azepane and azocane oxidation

3.1 Introduction

To date, in enzyme engineering, MD simulations and molecular docking have been mainly used as a tool to investigate the origin of selectivity changes caused by effective mutations. A few reports have explored the application of these to guide mutation design, mostly focusing on identifying key residues or conformational changes upon substrate binding.[172,176,235-237] There have been no report in the use of docking to guide $P450_{BM3}$ engineering to improve enantioselectivity for hydroxylation of cyclic amines.

This chapter explores the docking analysis for azepane and azocane with relevant selective variants and the potential application in mutation design. The MD simulations of variant structures were performed in GROMACS 2018.6[241] and docking calculations were conducted using Autodock Vina[220] or Autodock Flexible Receptors (ADFR),[221] as described in **5.5.1** and **5.5.2**. The simulation structures were firstly checked for stability throughout the simulation before docking calculations. All binding poses were analysed to gain insight of the binding preferences predicted by the docking calculations. The productive binding poses, which indicated the oxidation product as experimentally observed, were then overlaid with binding poses that indicated other products or were non-productive as the substrate was bound too far away from the ferryl oxygen. Mutations were then proposed such that they could interfere with these other binding poses without affecting the productive poses. The proposed variants were then tested experimentally, and the most effective ones used as template for further rounds of docking analysis and mutation design. The main objectives were to improve regio- and enantio-selectivity as well as enzymatic activity for the oxidation of azepane and azocane.

3.2 Initial docking study for azepane

3.2.1 Docking analysis with the VQ/S72G/A330W variant

3.2.1.1 MD simulations for the VQ/S72G/A330W variant

The VQ/S72G/A330W variant, which gave 91% of the β alcohol **5c** and 96% *ee* for (*R*)-**5c**, was selected as the first target to perform docking studies. This variant was a good candidate for a benchmarking study because it showed the highest enantioselectivity among the best variants for all four enantiomers of the remote hydroxylation products of azepane (Table 3.1).

Table 3.1 Best variants and selectivity for all *four* enantiomers of the remote hydroxylation products of azepane.

Products	Variant	Regioselectivity	*ee*	TTN
(*R*)-γ-ol, (*R*)-**1a**	RP/H171L/I263G/A330W/L437S	90%	90%, *R*	10060
(*S*)-γ-ol, (*S*)-**1a**	KU3/A330P/S72W	90%	84%, *S*	8440
(*R*)-β-ol, (*R*)-**5c**	VQ/S72G/A330W	91%	96%, *R*	1070
(*S*)-β-ol, (*S*)-**5c**	GV/A184I/I263G/A328G	76%	79%, *S*	290

Molecular dynamics (MD) simulation structures of the VQ/S72G/A330W variant were generated in three replicas following the protocol described in **5.5.1**. The RMSD of Cα atoms of all the recorded structures throughout the 100-ns simulation runs for all three replicas were 1–2 Å (Fig. 3.1) which showed that, during each replica simulation, most of the record conformations only had small deviations from one another. This indicated that, although they might not necessarily closely resemble the actual structure in solution, the structures generated from all three simulations were stable. These structures were then clustered following the procedure described in **5.5.1**, leading to seven clusters which constituted >90% of the total population for each of the three replica simulations (Table 3.2).

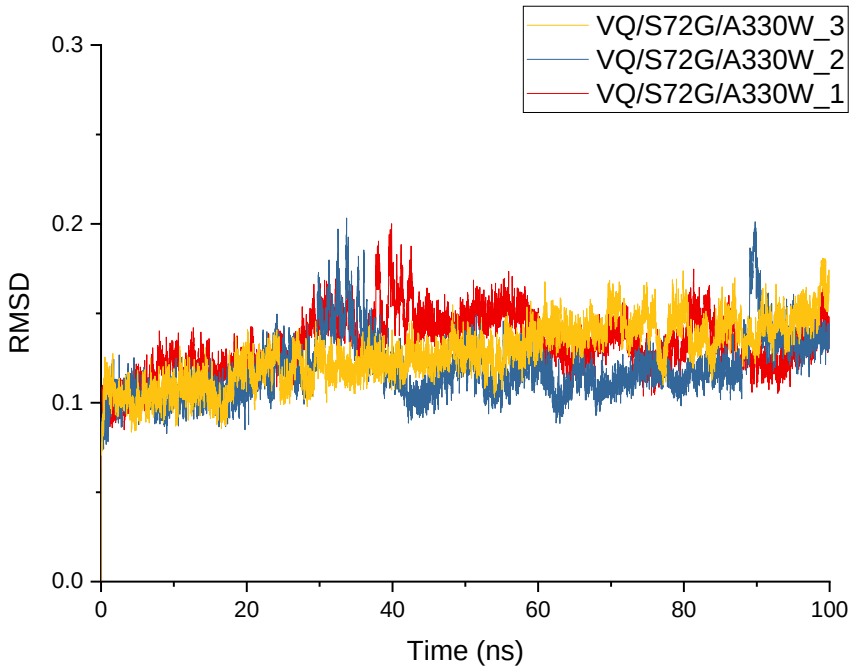


Figure 3.1. RMSD of Cα atoms of all recorded structures for each replica of the VQ/S72G/A330W variant during the three replica 100-ns MD simulations.

Table 3.2 Cluster population of the three replica MD simulation structures of VQ/S72G/A330W.

Replica	Cluster
Replica 1	C1 (8592)
	C2 (853)
Replica 2	C1 (8339)
	C2 (965)
Replica 3	C1 (6042)
	C2 (1559)
	C3 (1413)

The structures of the central member of each of these seven clusters were overlaid and compared for structural variations. The central members of the two clusters from replica 1 (R1C1 and R1C2) aligned well, with small deviations in the loop regions between helices, *e.g.*, the D/E, E/F, F/G and K′/L loops (Fig. 3.2). Similarly, the two clusters from replica 2 and the three clusters from replica 3 all aligned well with their counterparts from the same replica. The central members of the most populated cluster from each replica (R1C1, R2C1 and R3C1) were then overlaid, revealing more deviations in the loop regions, especially in the D/E and K′/L loops. Notably, however, the major helices were well-aligned (Fig. 3.3).

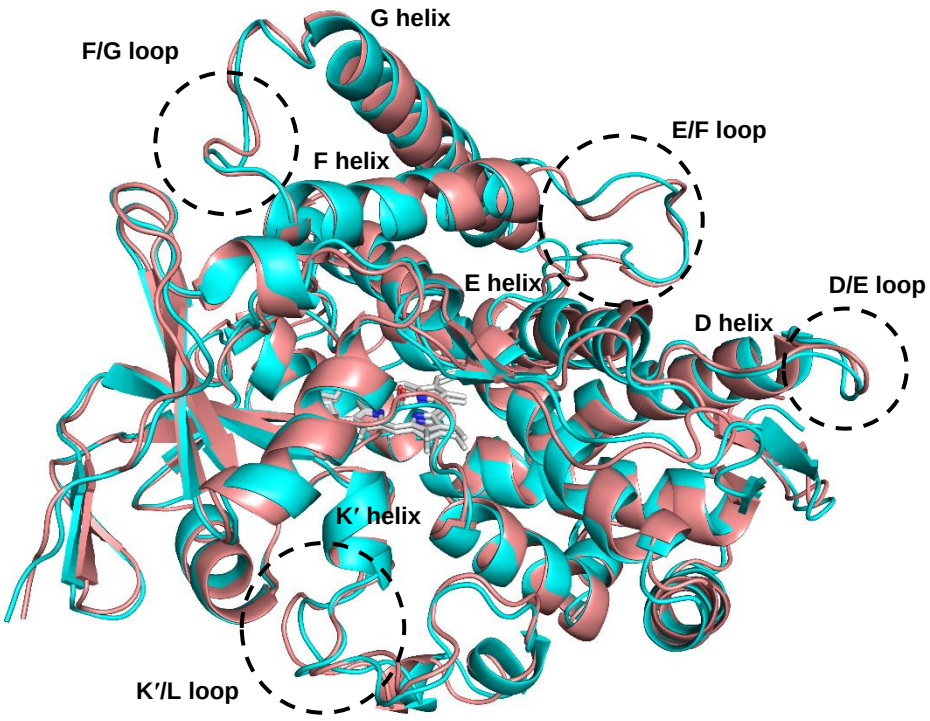


Figure 3.2. Overlay of central member of R1C1 (cyan) and R1C2 (salmon) of MD-simulated structures of VQ/S72G/A330W. Circles highlight loop regions showing the largest deviations between cluster structures (central member).

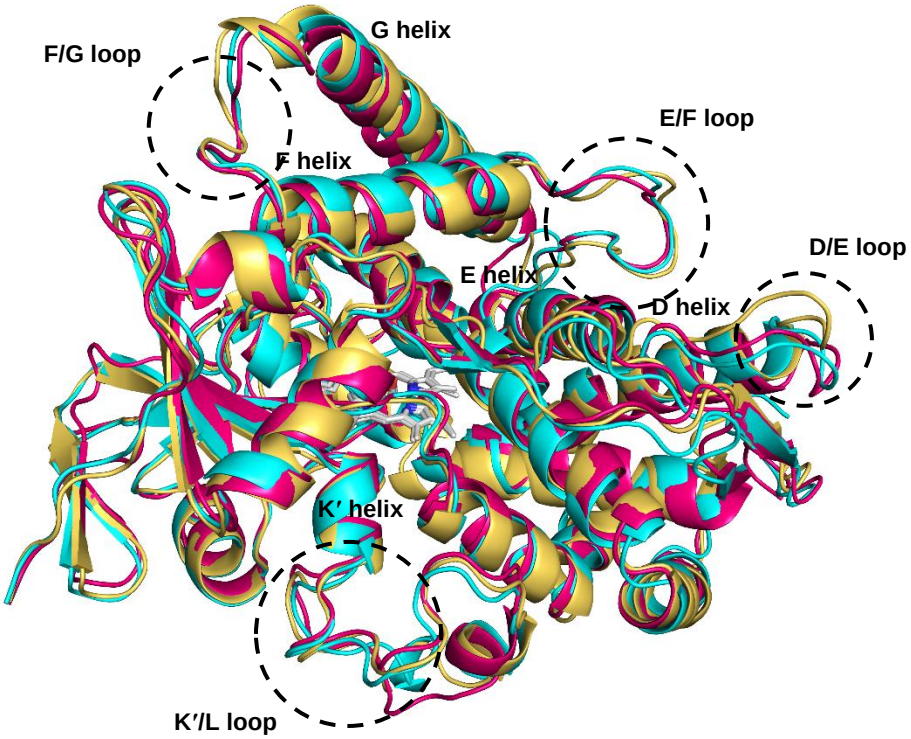


Figure 3.3. Overlay of the structures of central member of the most populated clusters from 3 replica simulations of the structure of the VQ/S72G/A330W variant. Circles highlight loop regions showing the largest deviations between replicas. R1C1 (cyan), R2C1 (red) and R3C1 (orange).

Changes in the local environment surrounding the mutated residues were then analysed between central members of R1C1, R2C1 and R3C1; structures of RxCy (cluster y of replica x) refer to structures of central member of RxCy for throughout the thesis. R1C1 and R2C1 showed closer structural alignment, with most of the residues aligning well (Fig. 3.5). For residues such as L75, V87, Q188 and I263, the main chains from R1C1 and R2C1 lined up but the side chains showed different rotamer conformations. For the W330 residue, there was a shift of the entire residue between R1C1 and R2C1, albeit mostly by less than 1 Å (Fig. 3.5). R3C1 showed more deviations from the other two counterparts, though most of the residues still aligned well with R1C1 and R2C1. Again, the W330 residue showed the most significant deviation. In R3C1, its indole side chain pointed towards the I helix, whereas in R1C1 and R2C1 the side chain faced the F helix (Fig. 3.5). Since the main chains of W330 across all clusters aligned well, these deviations likely reflected the flexibility of the W330 side chain. The diversity in W330 side-chain orientation might help expand the scope of the Autodock Vina analysis as the rigid body docking routine in Vina did not allow any flexibility of residue side chains.

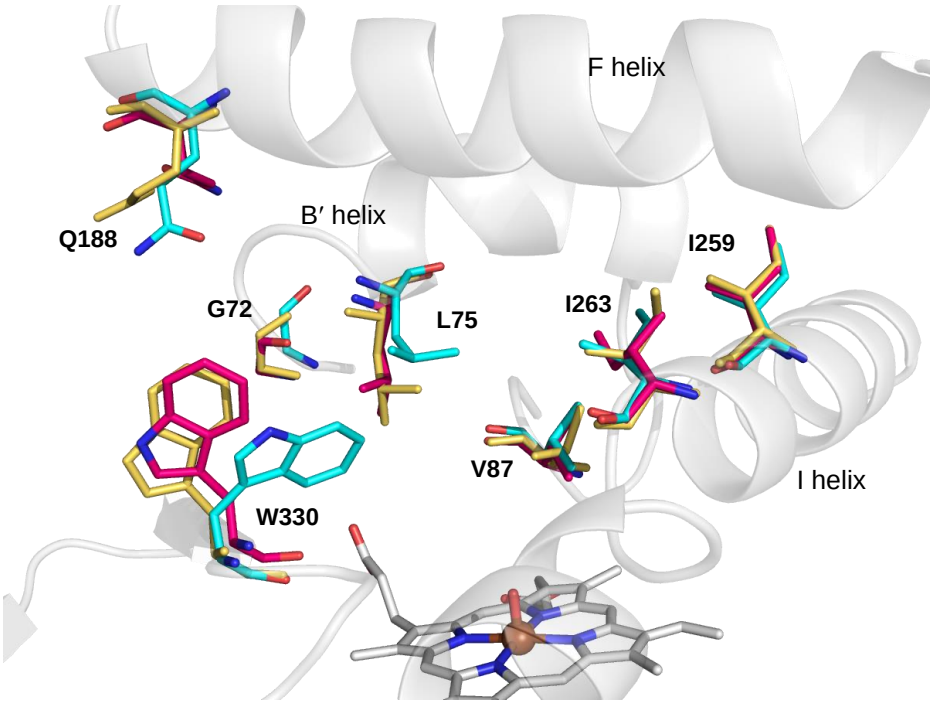


Figure 3.5. Overlay of selected residues and helices from central member of R1C1 (cyan), R2C1 (red) and R3C1 (orange) of MD-simulated structures of VQ/S72G/A330W.

The WT structure was then overlaid with these clustered structures for comparison. The main differences were around the mutated residues. The Q188 residue in the simulated structures was located closer to the haem group than the L188 in WT, rendering the binding pocket more constrained in the variant. The main chain of A330 in WT aligned well with W330 in the simulated structures but the bulky indole side chain of W330 significantly restricted the binding pocket. On the other hand, the absence of the phenyl group in V87 in the mutant created space for substrate binding. The changes at residues 87 and 330 would have significant effects on substrate binding and therefore product selectivity as they reshaped the active site pocket (Fig. 3.6).

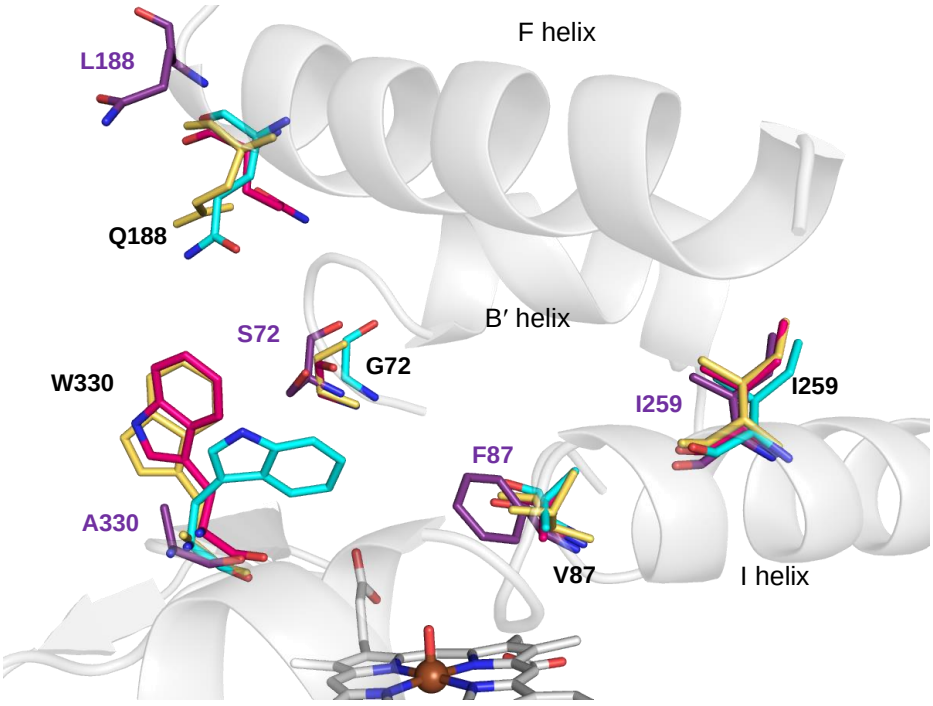


Figure 3.6. Overlay of selected residues and helices from central member of R1C1 (cyan), R2C1 (red) and R3C1 (orange) of MD-simulated structures of VQ/S72G/A330W, together with the WT structure (purple).

3.2.1.2 Docking poses of *N*-Ips-azepane (5) with the VQ/S72G/A330W variant

Substrate **5** was docked with the seven clusters by Autodock Vina which generated an output of the nine lowest affinity energy poses for each docking calculation. Among the 63 poses

generated from the docking runs, there were 18 productive poses, nine indicating (*R*)-β oxidation of the substrate, three for (*S*)-β oxidation, four for α oxidation and two for γ oxidation. Among the 45 non-productive (NP) poses, 30 indicated oxidation of the isopropyl modifying group and 15 were bound too far away for oxidation, *i.e.*, all carbon centres were more than 4.2 Å from the ferryl oxygen (Table 3.3). Surprisingly, the NP poses were dominant across all output poses (45 out of 63). It was hypothesised that this might be correlated to the lower turnover number for the oxidation of **5** by this variant. To explore this further, *N*-Boc-azepane (**1**) was docked with this variant and the output compared with that for *N*-Ips-azepane (**5**). These substrates are similar in structure but the turnover number of 1200 for the oxidation of **5** was significantly lower than that (2000) for **1**. Docking of *N*-Boc-azepane showed fewer NP poses, at 31 out of 63, compared with 45 out of 63 by for the *N*-Ips derivative **5** (Table 3.3). This supported the hypothesis that the number of NP poses observed from docking runs could reflect enzymatic activity.

Table 3.3 Pose distribution of the docking of substrates **1** and **5** with the MD-simulated structure of VQ/S72G/A330W.

***N*-Ips-azepane, 5 (TTN = 1200)**	**Pro-R-β**	**Pro-S-β**	**Pro-α**	**Pro-γ**	**Non-productive**	
					Ips-	Other
	9	3	4	2	30	15
Sub-total	Productive: 18				Non-productive: 45	
***N*-Boc-azepane, 1 (TTN = 2000)**	**Pro-R-β**	**Pro-S-β**	**Pro-α**	**Pro-γ**	Boc-	Other
	3	5	14	10	19	12
Sub-total	Productive: 32				Non-productive: 31	

Of the productive poses for *N*-Ips-azepane, the (*R*)-β poses were the most populated (9 out of 18), which was in line with the experimental observation that the (*R*)-β alcohol **5c** was the dominant product. However, the predicted selectivity, calculated based on the proportion of productive poses, did not agree with the experimental data. The docking predicted much lower enantioselectivity for (*R*)-β oxidation, with a (*R*)-β:(*S*)-β ratio of 3, compared with 45 observed

experimentally. The α- and γ- oxidation products were predicted to constitute 22% and 11% respectively, but only 5% and 4% were observed (Table 3.4).

The cluster populations of the binding poses were then taken into consideration by weighting each pose with the population of its cluster to calculate a population-weighted selectivity. There was a slight increase in the predicted (*R*)-β oxidation selectivity; the (*R*)-β:(*S*)-β ratio was increased to 4, compared with 3 without weighting. However, the selectivity for α- and γ- oxidation remained over-estimated, at 20% and 11% respectively (Table 3.4).

Table 3.4 Predicted selectivity, population-weighted and population-and-energy-weighted selectivity for oxidation products of the substrate **5** calculated from docking of **5** with MD-simulated structures of VQ/S72G/A330W and the experimental data.

	Pro-*R*-β	**Pro-*S*-β**	***R*-β/*S*-β**	**Pro-α**	**Pro-γ**
Number of poses	9	3	3	4	2
Predicted selectivity based on the number of poses	50%	17%	3	22%	11%
Population-weighted selectivity	56%	13%	4	20%	10%
Population- and energy-weighted selectivity	74%	12%	6	11%	3%
Experimental data	89%	2%	45	5%	4%

A key factor in determining product selectivity is the affinity energy of binding poses. As shown in a scatter plot of the affinity energy of all output poses (Fig. 3.7), the (*R*)-β poses had lower affinity energies, with seven out of nine poses showing affinity energies between –6.5 and –6.2 kcal/mol, while all the other productive poses were at –6.2 kcal/mol or higher. The lowest energy pose among the productive poses was a (*R*)-β pose. All productive poses were then weighted by their affinity energies. Based on $\Delta G^{o} = -RT\ln K$, if there is 0.5 kcal/mol difference in the affinity energies of two poses A and B, the population of pose A is 2.3 times that of B. If the weight of the lowest energy pose (–6.5 kcal/mol) is set as 1 (Table 3.5), then a pose with an affinity energy of –6.0 kcal/mol has a weight of 0.43, *etc*. By including the effect of the affinity energy, the population-and-energy-weighted selectivity could be calculated (Table 3.4). The

predicted (*R*)-β oxidation selectivity increased to 73% and the (*R*)-β:(*S*)-β ratio was 6, compared to 3 without any weighting. The predicted γ oxidation selectivity (3%) was practically the same as experimentally observed (4%). The α oxidation percentage remained over-estimated at 11% but was much closer to the 5% from experiments.

Table 3.5 Affinity energies of poses and their corresponding weights for the calculation of population-and-energy-weighted selectivity. A pose with an affinity energy of –6.5 kcal/mol was weighted as 1.

Affinity energy (kcal/mol)	–6.5	–6.4	–6.3	–6.2	–6.1	–6.0	–5.8	–5.5	–5.4	–5.2
Weight assigned by affinity energy	1	0.84	0.71	0.60	0.51	0.43	0.31	0.18	0.16	0.13

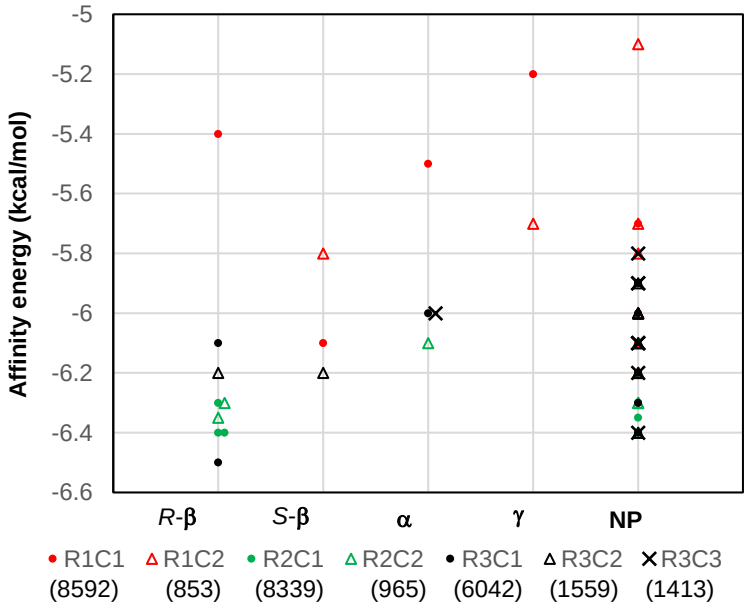


Figure 3.7 The affinity energy of output poses from docking of N-Ips-azepane, **5**, with MD-simulation structures of the variant VQ/S72G/A330W. RxCy denotes cluster y of replica x. The numbers in brackets indicate the population of each cluster.

Overall, Autodock Vina docking of **5** with the VQ/S72G/A330W variant predicted (*R*)-β oxidation as the major pathway. The lowest energy pose was that for (*R*)-β oxidation and (*R*)-β poses generally showed lower affinity energies than other productive poses. A more quantitative analysis was attempted by taking into account the cluster population and the affinity energy of the poses. It was noted, however, that Autodock Vina uses a rigid body

docking routine in vacuum, with the protein structure being fixed and no side chain movement. The affinity energy is calculated in vacuum, without structural relaxation in water, that is, enzyme–substrate interactions such as water-mediated hydrogen bonding is not possible. Since it was impractical to carry out MD simulations in water for all productive docking poses, the above analysis remains qualitative. Finally, the rate of hydrogen atom abstraction from a C–H bond by the ferryl oxygen depends on the distance between the ferryl oxygen and the carbon centre as well as the C–H···O and H···O=Fe angles. DFT calculations could estimate the relative rates of hydrogen abstraction for a range of distances and angles. However, given the un-optimised structures and binding interactions of the raw docking output, the distances and angles are unlikely to reflect catalytically relevant values. Therefore, further quantitative analysis of the docking output to predict the product profile accurately is not justified. Nevertheless, the docking results predicted qualitatively the selectivity bias, *i.e.*, mainly (*R*)-β oxidation, observed experimentally for the reaction. Therefore, the residue–substrate contacts for the different poses could provide insight into the role of mutations and potentially clues to designing mutations to enhance turnover activity and product selectivity.

3.2.1.3 The (*R*)-β poses

All 63 output poses were clustered in the pocket formed by residues G72, V87, I263, W330, L437 and T438 (Fig. 3.8). Hydrogen bonds were observed between the isopropylsulfonyl groups of some of the poses and residues W330 or T438. In general, the poses from Replica 1 were closer to the I helix than those from Replicas 2 and 3, which was likely a result of the conformational deviation of the W330 side chain in Replica 1 (Fig. 3.8). As discussed above, the side chain of W330 in Replica 1 was rotated by ~90° and pointed towards the I helix, which might push the poses closer to the I helix.

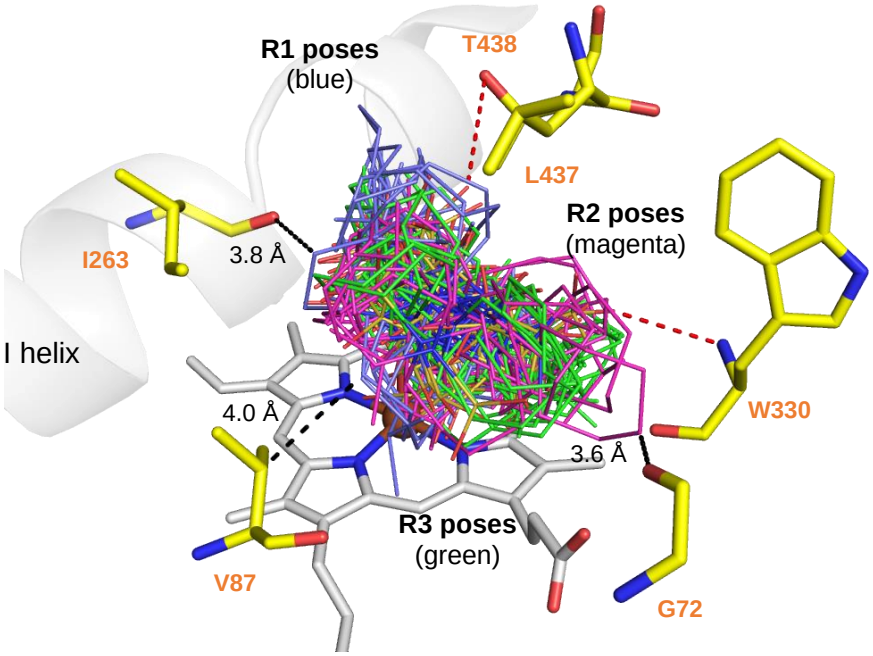


Figure 3.8 Overlay of all poses from docking of **5** with replica 1 (blue), replica 2 (magenta) and replica 3 (green) from the MD-simulation structures of the VQ/S72G/A330W variant, with key active site residues (yellow) of the R3C1 structure which gave the lowest energy (*R*)-β pose. Red dashes denote hydrogen bonds. Black dashes with labels denote distances between residues and poses.

Poses indicating the same product were clustered and analysed (Fig. 3.9-A). Most of the (*R*)-β poses adapted similar binding orientations. The sulfonyl oxygens of the isopropyl groups were in van der Waals contact with residues P329, W330, L437 and T438, and hydrogen-bonded with P329 or W330. The isopropyl group resided in the pocket formed by residues G72, L75, W330 and S332. The azepane ring pointed towards the I helix and was in van der Waals contact with residues V87 and L437 (Fig. 3.9-A). In the lowest energy (*R*)-β pose (from R3C1), the isopropyl group was in the space created by the S72G mutation but not in direct contact with the G72 residue. Cβ of the azepane ring was 3.3 Å from the ferryl oxygen while Cγ contacted V87 (Fig. 3.9-B). This pose was also highly weighted in cluster population and thus considered as the most representative binding model for (*R*)-β oxidation by VQ/S72G/A330W.

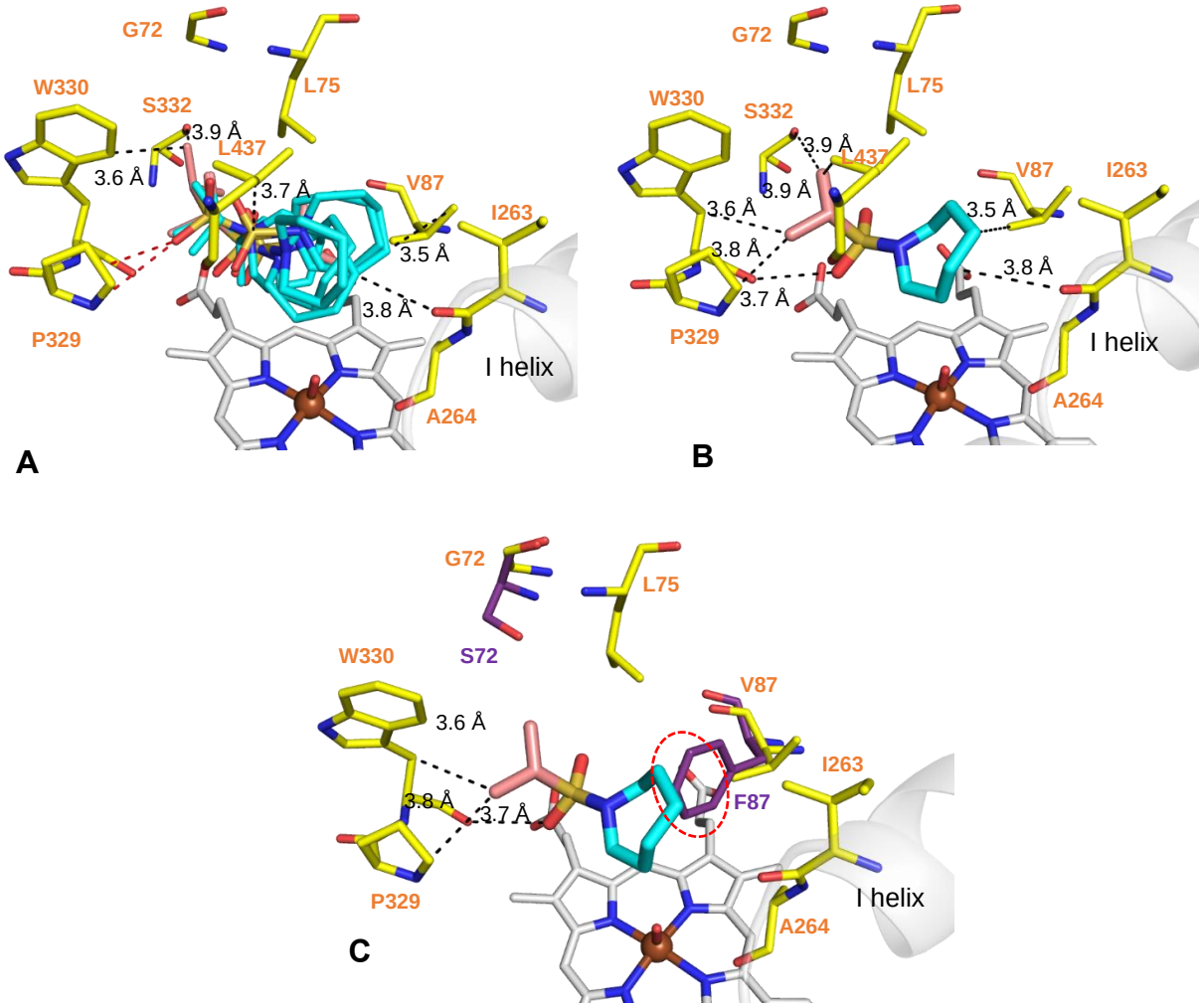


Figure 3.9 Overlay of simulated structure of VQ/S72G/A330W (yellow) with **(A)** all (*R*)-β poses (cyan), **(B)** the lowest energy (*R*)-β pose (cyan) and **(C)** the lowest energy (*R*)-β poses together with the S72 and F87 residues from the WT structure (purple). Black dashes with black labels denote distances between I263 and NP. Cyan dashes with cyan labels denote distances between I263 and the lowest energy (*R*)-β pose. Red dashed circles denote steric clashes between poses and residues.

From the overlay with the WT structure, it was clear that the (*R*)-β pose would not be stable without the S72G and F87V mutations due to the steric clashes with F87 in the WT structure (Fig. 3.9-C). Experimentally, the WT gave much lower selectivity for the (*R*)-β alcohol, at 8% of **5c** (*ee* not determined) compared with 91% of **5c** and 96% *ee* for (*R*)-**5c** by the VQ/S72G/A330W variant. The overlay and comparison suggested that the lowest energy (*R*)-β pose could serve as an indication of (*R*)-β-selectivity in docking analysis: when that pose was destabilised in a variant structure, the variant was likely to give less (*R*)-β alcohol. To explore

this further, the lowest energy (*R*)-β pose was overlaid with simulation structures of a few other variants showing varied (*R*)-β selectivity. Both RP/H171L/I263G and KU3/A330P/S72W structures showed clashes with the lowest energy (*R*)-β pose at residues F87 (both structures), A264 (RP/H171L/I263G) and T438 (KU3/A330P/S72W). Experimentally, both variants gave little β alcohol, at 3% and 4% respectively (Table 3.6). On the other hand, no clashes were observed between the lowest energy (*R*)-β pose and the GVQ/A330W structure, which was (*R*)-β selective, experimentally giving 83% of **5c** and 90% *ee* for (*R*)-**5c**. These observations supported that the (*R*)-β poses from docking with the simulated VQ/S72G/A330W replicas could serve as indicators for (*R*)-β-selectivity.

Table 3.6 Steric clashes observed from the overlay of the lowest energy (*R*)-β pose from the VQ/S72G/A330W docking and selected variants with their experimental selectivity for oxidation products of substrate **5**.

Variants	**Clashes with the (*R*)-β pose from docking with VQ/S72G/A330W**	**Experimental selectivity**		
		β-ol, 5c	**α-ol, 5e**	**γ-ol, 5a**
WT	Strong clashes with F87	8%		
RP/H171L/I263G	Strong clashes with F87 and A264	3%	1%	96%
KU3/A330P/S72W	Strong clashes with F87 and T438	4%	–	96%
GVQ/A330W	No clashes	83% of **5c**, 90% ee for *R*	9%	8%
VQ/S72G/A330W	No clashes	91% of **5c**, 96% ee for *R*	4%	5%

To summarise, the (*R*)-β poses generated from the VQ/S72G/A330W docking runs aligned well with each other. Overlay of the lowest energy (*R*)-β pose with the simulation structures of variants that gave little (*R*)-β alcohol indicated clashes between the (*R*)-β pose and the corresponding variant structure. This could potentially guide further mutagenesis to improve selectivity for (*R*)-β alcohol – mutations that do not clash with the (*R*)-β pose from docking with precursor variants could be introduced to improve enzymatic activity or product selectivity without the risk of losing (*R*)-β-selectivity.

3.2.1.4 The non-productive poses

There were two clusters of non-productive (NP) poses. For the major cluster (>90% of all non-productive poses), the azepane ring was close to the I helix and in van der Waals contact with I263 and L437, while the Ips protecting group pointed towards the porphyrin ring. The other non-productive poses had their azepane ring residing in the pocket formed by G72, W330 and S332 (Fig. 3.10-A).

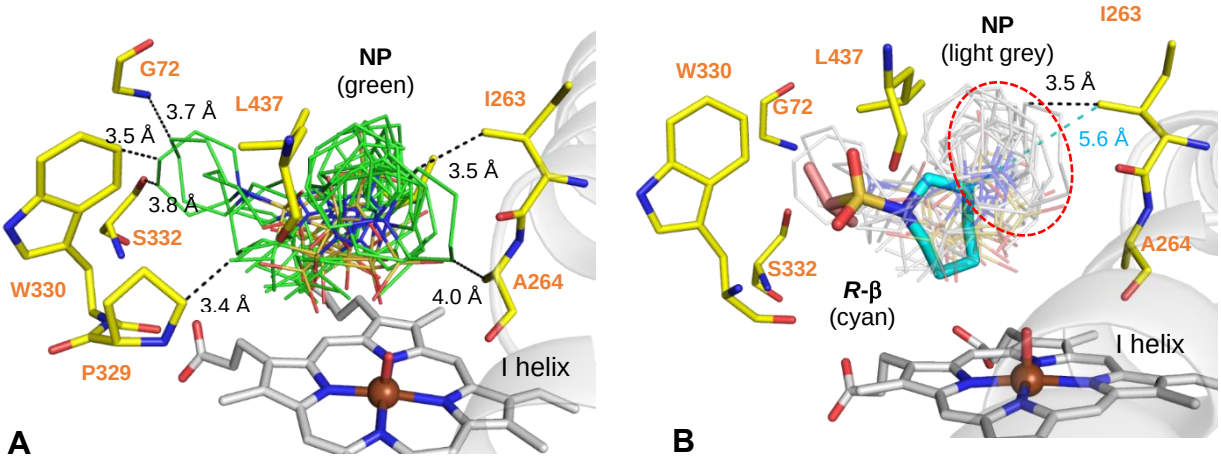


Figure 3.10 Overlay of selected residues of the simulation structure of VQ/S72G/A330W (yellow) with **(A)** all NP poses (green) and **(B)** the lowest energy (*R*)-β pose (cyan) with NP poses (light grey). Black dashes with black labels denote distances between I263 and NP. Cyan dashes with cyan labels denote distances between I263 and the lowest energy (*R*)-β pose. Red dashed circles denote the binding regions for NP poses.

From the overlay of the lowest energy (*R*)-β pose with the NP poses, the azepane rings of the major cluster of NP poses were located in a region where the (*R*)-β pose was not bound (Fig 3.10-B). It was hypothesised that if, through mutagenesis, this region of the substrate pocket for NP poses was restricted, then it would lead to destabilisation of most of these poses. The (*R*)-β pose could remain unaffected since it was not in that region of the substrate pocket. Such mutagenesis could improve the enzymatic activity since, the number of NP poses formed by Autodock Vina approach appeared to be correlated to the enzymatic activity. The physical restriction of the region for NP poses could be achieved, theoretically, by mutating residue I263 to a bulkier residue; the azepane rings of the main cluster of NP poses were typically ~3.5 Å

from the I263 side chain while the (*R*)-β pose was, at the closest point, 5.6 Å from I263 (Fig. 3.10-B).

3.2.1.5 Binding poses for other products

The (*S*)-β poses aligned well and showed significantly different orientations to the (*R*)-β poses. The Ips protecting group of the (*S*)-β poses was in van der Waals contact with the L437 residue and the I helix residues I263 and A264, while the azepane ring pointed towards the β1 strand, contacting P329 and A330: this is the opposite orientation of the (*R*)-β poses (Fig. 3.11-A). From the overlay of the lowest energy (*R*)-β pose and the (*S*)-β poses, the Ips protecting group of the pro-(*S*) poses occupied space where no part of the (*R*) pose resided (Fig. 3.11-B). The protecting group of the (*S*) poses were ~3.4 Å from the I263 side chain, while the (*R*) pose was 5.6 Å from I263 (Fig. 3.11). It could be hypothesised that, mutating I263 to bulkier groups could destabilise the (*S*) poses without affecting the (*R*) pose, potentially improving the enantioselectivity for the (*R*)-β alcohol.

In summary, the poses indicating the same product align well and comparison between different poses provide insight into the binding patterns and could guide mutagenesis designs. Overlay of the lowest energy (*R*)-β pose with the NP poses indicated there were regions of the substrate pocket where the NP poses were located but the (*R*)-β poses were not bound. This observation inspired designs to restrict the substrate binding pocket through mutagenesis to destabilise the NP poses to improve enzymatic activity of the variants. Analysis of (*R*)-β poses against other productive poses revealed similar patterns–there were significant differences between them to design mutations that would destabilise other poses without affecting the (*R*)-β pose. However, such mutagenesis was not applied on the VQ/S72G/A330W variant, as it gave very high selectivity already, at 91% of **5c** and 99% *ee* for (*S*)-**5c**. Instead, docking analysis for substrate **5** with the GV/A184I/I263G/A328G variant, which gave the other enantiomer, (*S*)-**5b**, as the

major product, was performed as a direct comparison with the VQ/S72G/A330W variant. The GV/A184I/I263G/A328G variant gave 76% of **5c** and 79% *ee* for (*S*) and could be a starting point to test docking-guided mutation design to improve enzymatic activity and product selectivity.

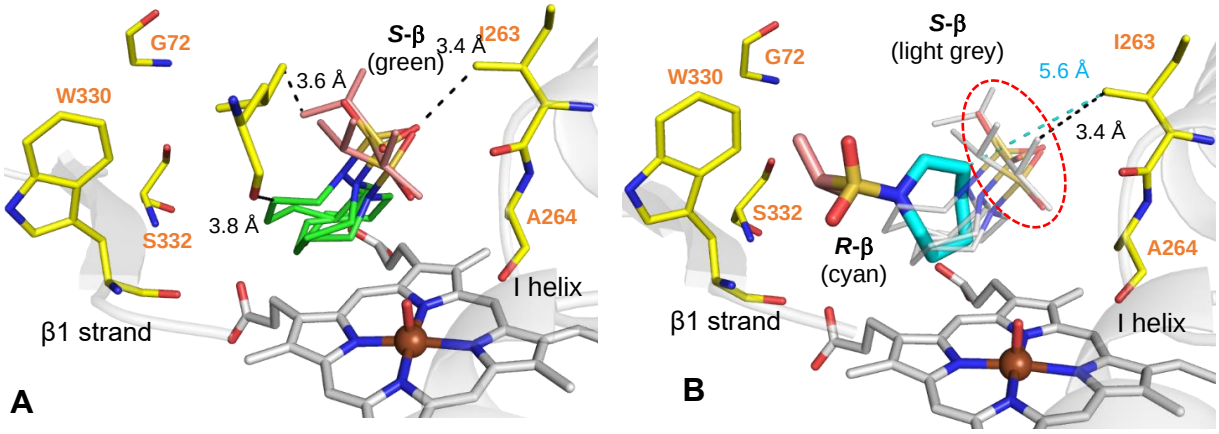


Figure 3.11 Overlay of selected residues of the simulation structure of VQ/S72G/A330W with **(A)** all (*S*)-β poses (green) and **(B)** the lowest energy (*R*)-β pose (cyan) with (*S*)-β poses (light grey). Black dashes with black labels denote distances between I263 and (*S*)-β poses. Cyan dashes with cyan labels denote distances between I263 and the lowest energy (*R*)-β pose. Red dashed circles denote the binding region for (*S*)-β poses.

3.2.2 Docking analysis for the GV/A184I/I263G/A328G variant

3.2.2.1 MD simulations of the GV/A184I/I263G/A328G variant

As with the VQ/S72G/A330W variant, three replica MD simulations were carried out for GV/A184I/I263G/A328G. The RMSD values (1–1.5 Å) indicated stable structures for all three 100-ns simulations; the Replica 1 simulation showed slightly higher RMSD towards the end of the simulation (85–100 ns), at 1.5–2 Å, which was still a rather small deviation (Fig. 3.12). Seven clusters, constituting 90% of the total population for each replica, were generated (Table 3.7).

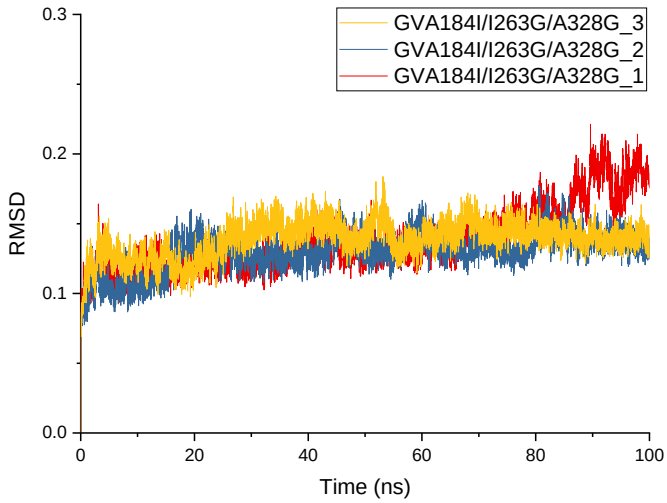


Figure 3.12 RMSD of Cα atoms of all recorded structures for each replica of the GV/A184I/I263G/A328G variant during the three replica 100-ns MD simulations.

Table 3.7 Cluster population of the three MD simulated replicas of GV/A184I/I263G/A328G.

Replica	**Cluster**
Replica 1	C1 (4323)
	C2 (3190)
	C3 (1263)
Replica 2	C1 (9043)
	C2 (606)
Replica 3	C1 (7751)
	C2 (2023)

These seven clusters of simulated structures showed better alignment with each other than those from VQ/S72G/A330W. The most populated clusters from each replica, *i.e.*, R1C1, R2C1 and R3C1, only showed small deviations in the loop regions (Fig. 3.13).

Local deviations of residues were then analysed for R1C1, R2C1 and R3C1 structures. Most of the residues showed good alignment. Rotamer conformations were observed for the side chains of V87, I184 and L188. Only a few residues showed backbone shifts and these deviations were all less than 1.5 Å (Fig. 3.14-A). Overlay of these structures with the WT revealed more deviations, especially around the mutated residues. The side chains of both F87 and I263 in the WT structure constrained the active site pocket further, compared with V87 and G263 in

GV/A184I/I263G/A328G (Fig. 3.14-B). The other mutated residues G74 and I184, as well as L188, also showed slight side-chain or backbone deviations.

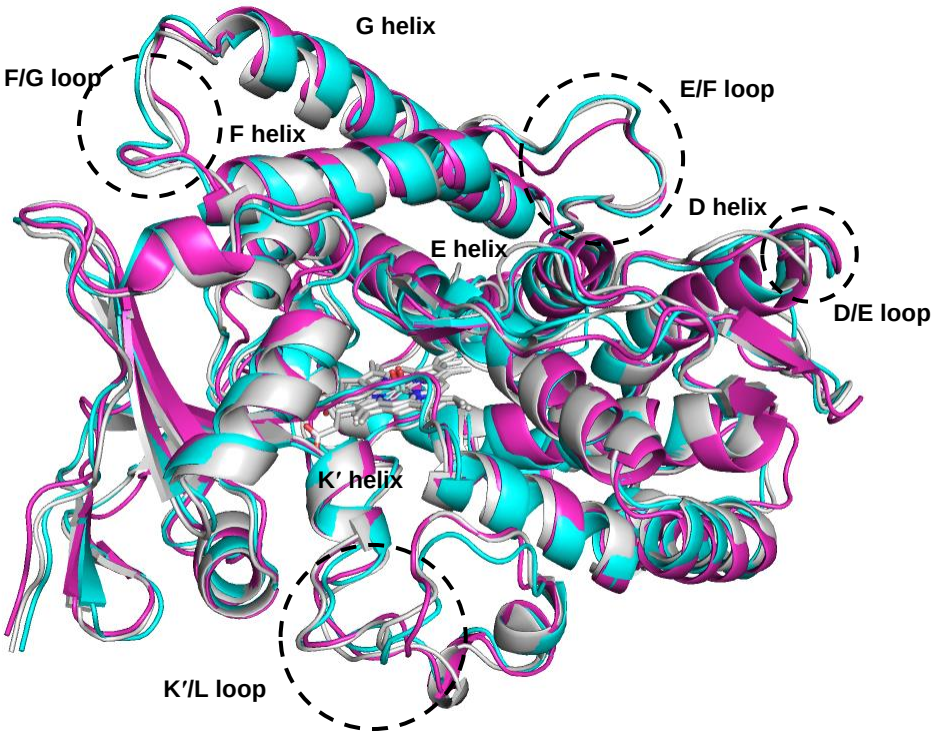


Figure 3.13. Overlay of the structures of the most populated clusters from 3 replica simulations of the structure of the GV/A184I/I263G/A328G variant. Circles highlight loop regions showing the largest deviations between replicas. R1C1 (cyan), R2C1 (grey) and R3C1 (magenta).

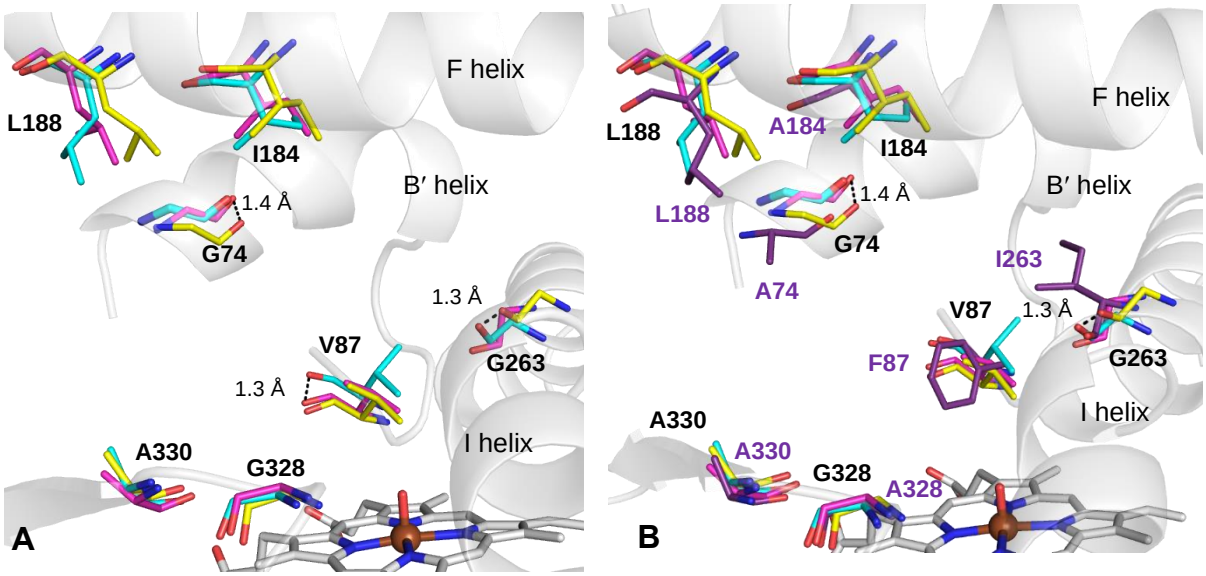


Figure 3.14. Overlay of selected residues and helices from **(A)** R1C1 (cyan), R2C1 (yellow) and R3C1 (magenta); and **(B)** R1C1 (cyan), R2C1 (yellow), R3C1 (magenta) and WT (purple) of MD-simulated structures of GV/A184I/I263G/A328G.

3.2.2.2 Docking poses of *N*-Ips-azepane (5) with the GV/A184I/I263G/A328G variant

Docking of substrate **5** was then performed with the seven clusters. Of the 63 poses generated, there were 10 poses indicating (*S*)-β oxidation of **5**, four for (*R*)-β oxidation, eight for α oxidation, six for γ oxidation, together with 35 NP poses. The lowest energy pose was for (*S*)-β oxidation (Fig. 3.15). Replica 1 dominated the lower energy (from –6 to –5.7 kcal/mol) productive poses. The lowest energy productive poses given by other replicas were two γ poses at –5.7 kcal/mol by R3C1 and one α pose at –5.7 kcal/mol by R2C1 (Fig. 3.15). To explore this, the R1C1 and R3C1 structures were overlaid, together with the lowest energy (*S*)-β pose. From the comparison, most of the residues showed small deviations; rotamer conformation was observed for side chains of L75, T268 and L437, and there were slight backbone shifts for S72, L181, G263 and A330 (Fig. 3.16). Compared with R1C1, the side chain of T268 in R3C1 was rotated by ~90°, leading to steric clashes (<1.9 Å) with the lowest energy (*S*)-β pose. This showed the importance of running replica simulations. If only one simulation was performed instead of three, we could have ended up with R2- or R3-like structures which would predict predominant α (R2-like) or γ oxidation (R3-like). Knapp *et al*. demonstrated that 100 replica MD simulations of a 827 residue T-cell receptor/peptide/MHC complex would give a much closer resemblance of the crystal structure than if two were performed.[242] However, such a large number of simulations is not practical for a routine analysis. The docking study on *N*-Ips-azepane with MD-simulated structures of VQ/S72G/A330W (Section **3.2.1.2**) demonstrated that three replicas were sufficient for qualitative agreement with experimental data.

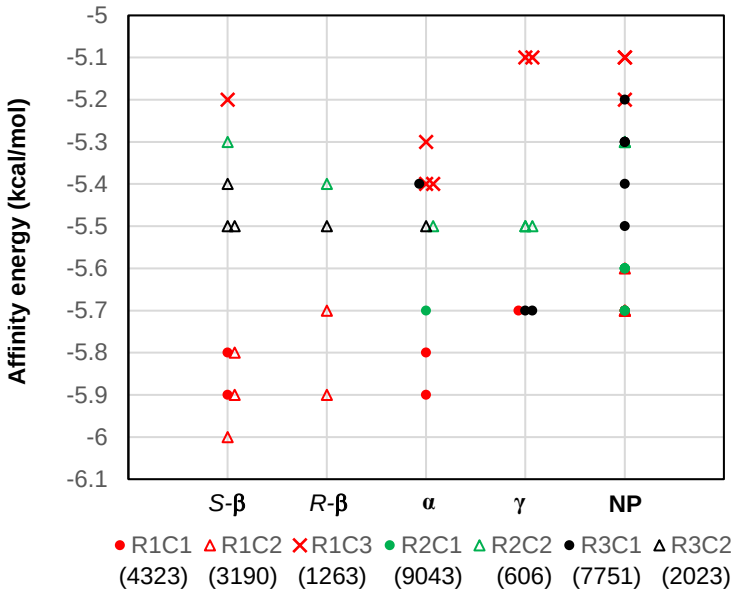


Figure 3.15. The affinity energy of output poses from docking of *N*-Ips-azepane, **5**, with MD-simulation structures of the variant GV/A184I/I263G/A328G. RxCy denotes cluster y of replica x. The numbers in brackets indicate the population of each cluster.

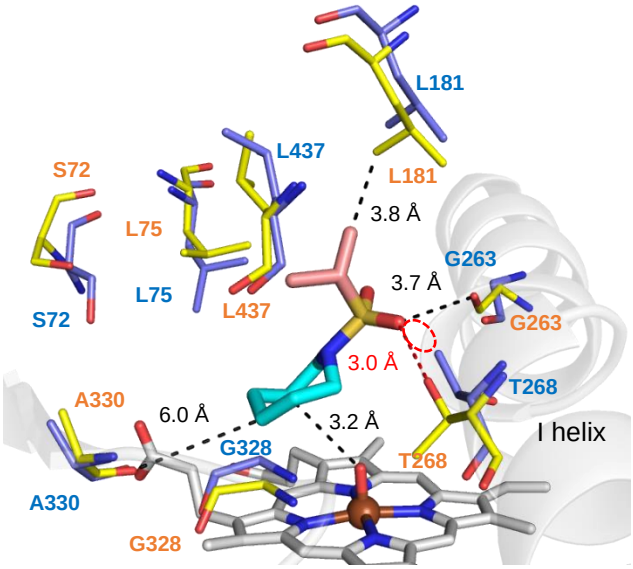


Figure 3.16. Overlay of selected residues and helices from R1C1 (yellow) and R3C1 (blue); together with the lowest energy (*S*)-β-pose (cyan). Red dashes denote hydrogen bonds. Red dashed circles denote steric clashes.

Of the productive poses, the (*S*)-β poses were the most populated, in agreement with the experimental data [76% of β alcohol **5c** and 79% *ee* for (*S*)-**5c**], although the simple pose count predicted a (*S*)-β:(*R*)-β ratio of 2.5, compared with 8.5 for 79% ee from experiment (Table 3.8). Both α and γ selectivity were over-estimated based only on the number of poses, at 29% and

21% respectively, compared with 16% and 8% experimentally (Table 3.8). The cluster population and affinity energy weighted *S*:*R* ratio of 4.0 was slightly higher. On the other hand, the predicted β oxidation selectivity of 64% was close to the 76% from experiment. The weighted α and γ selectivity were both lowered, at 23% and 13%, and closer to the experimental observations (Table 3.8).

Table 3.8 Predicted selectivity for oxidation products of the substrate **5** calculated from docking with MD-simulated structures of GV/A184I/I263G/A328G *vs*. VQ/S72G/A330W and their corresponding experimental data.

Variants		Pro-*S*-β	Pro-*R*-β	*S*-β/*R*-β	Pro-α	Pro-γ
GV/A184I/I263G/A328G (*S*-selective)	**Number of poses**	10	4	2.5	8	6
	Predicted selectivity based on number of poses	36%	14%	2.5	29%	21%
	Population/energy-weighted selectivity	51%	13%	4.0	23%	13%
	Experimental data	68%	8%	8.5	16%	8%
				***R*-β/*S*-β**		
VQ/S72G/A330W (*R*-selective)	**Number of poses**	3	9	3	4	2
	Predicted selectivity based on number of poses	17%	51%	3	22%	11%
	Population/energy-weighted selectivity	12%	74%	6	11%	3%
	Experimental data	2%	89%	45	5%	4%

Comparison of this data with that from docking of **5** with the (*R*)-selective VQ/S72G/A330W variant showed that substrate docking indicated qualitatively the correct major regio-isomer and with the correct stereoselectivity, especially with the population- and energy-weighted selectivities (Table 3.8). With this established, analysis of the binding poses and especially comparisons between poses for different products and poses from docking with (*S*)-selective *vs*. (*R*)-selective variants could provide insight into the patterns of binding interactions.

3.2.2.3 The (*S*)-β poses

The (*S*)-β poses were in similar orientations (Fig. 3.17-A). The sulfonyl oxygens of the Ips protecting group were close to the I helix and in van der Waals contact with G263, A264 and

T268, and with H-bonding to T268 in some poses. The isopropyl group was sequestered in a pocket formed by L75, L181 and L437, while the azepane ring pointed towards the β1 strand.

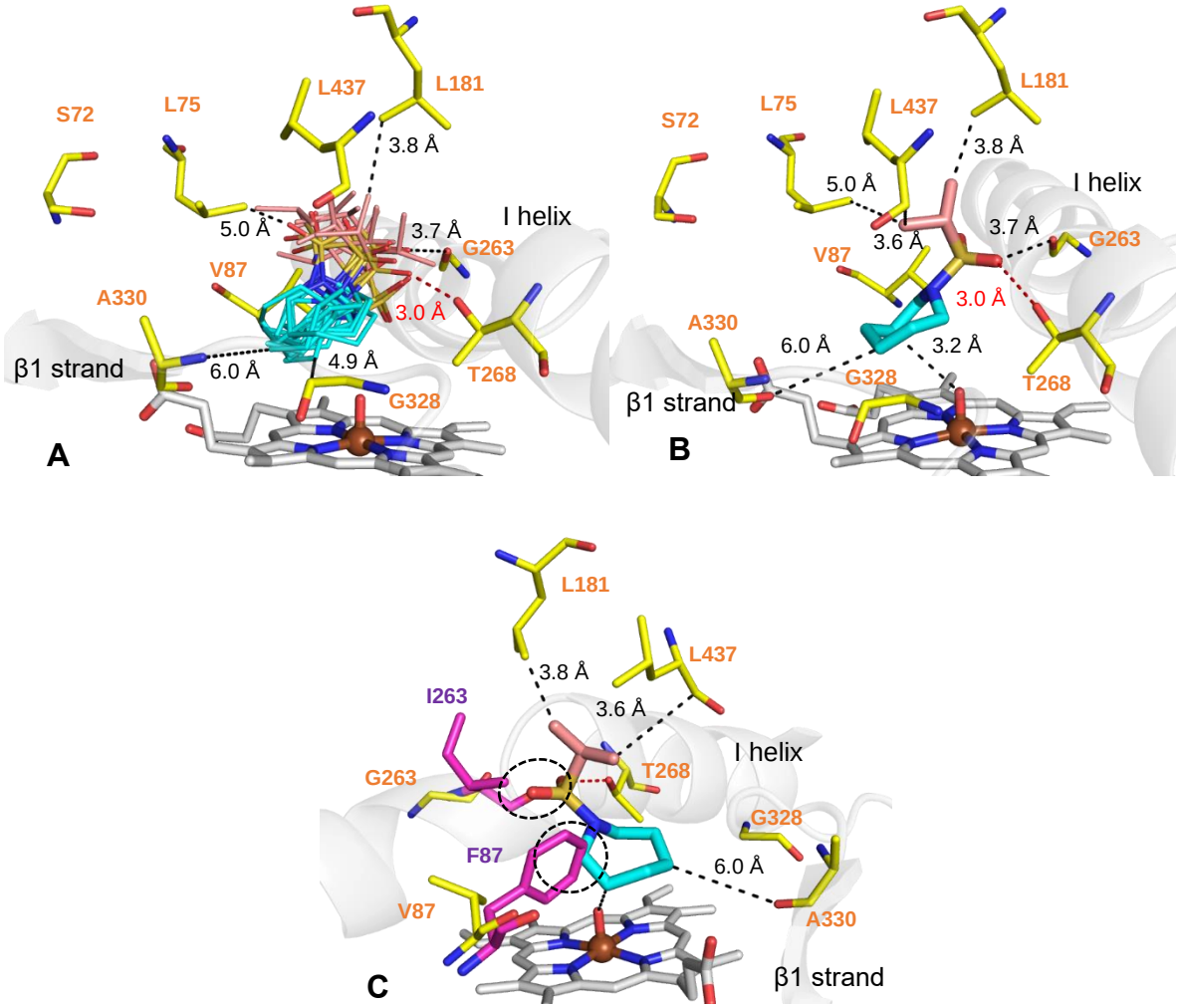


Figure 3.17 Overlay of MD-simulation structures of GV/A184I/I263G/A328G (yellow) with **(A)** all (*S*)-β poses (cyan), **(B)** the lowest energy (*S*)-β pose (cyan), and **(C)** (*S*)-β poses (cyan) with the F87 and I263 residues from the WT structure (magenta). Black dashes with black labels denote distances between residues and poses. Red dashes denote H-bonds. Black dashed circles denote steric clashes between poses and residues.

In the lowest energy (*S*)-β pose, Cβ was in the space generated by the F87V mutation and 3.2 Å from the ferryl oxygen but there was no direct contact with the V87 side chain. Other azepane ring carbons contacted G328, P329 and the porphyrin ring (Fig. 3.17-B). The lowest energy (*S*)-β pose would be destabilised without the F87V and I263G mutations because of steric clashes with the F87 and I263 side chains in the WT structure (Fig. 3.17-C).

The lowest energy (*S*)-β pose from GV/A184I/I263G/A328G was compared with the lowest energy (*R*)-β pose from VQ/S72G/A330W. The overlay showed that these two poses were in "flip-over orientation" relative to each other (Fig 3.18). It was striking that the (*S*)-β pose in the (*S*)-selective variant would not be stable in the (*R*)-selective variant, due to strong steric clashes with the I263 residue in the latter. Similarly, clashes with the L75 side chain in the GV/A184I/I263G/A328G variant meant that the (*R*)-β pose from VQ/S72G/A330W could be destabilised in this (*S*)-selective variant (Fig. 3.18). The (*R*)-β pose was also stabilised by the strong van der Waals interactions with W330 side chain of the (*R*)-selective variant, which was not found in the (*S*)-selective variant which did not have the A330W mutation.

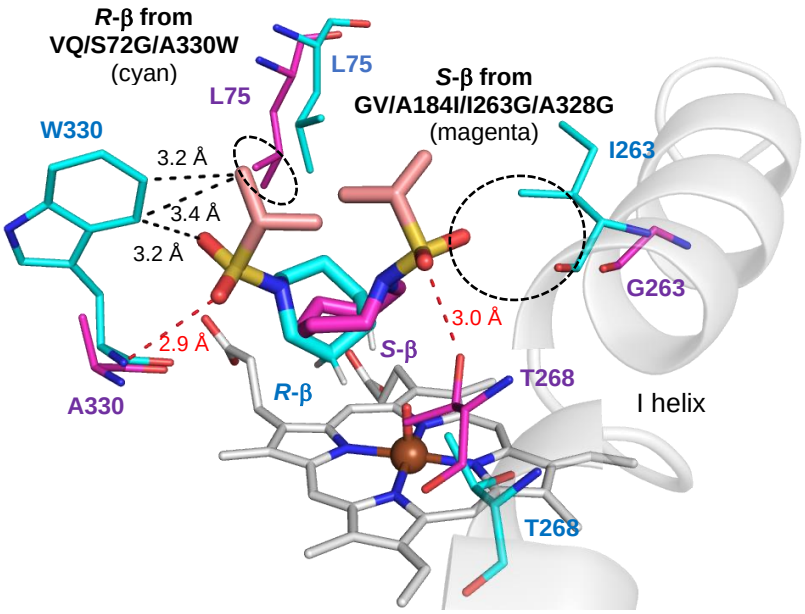


Figure 3.18 Overlay of the lowest energy (*S*)-β-pose (magenta) from docking with GV/A184I/I263G/A328G *vs*. the lowest energy (*R*)-β-pose (cyan) from docking with VQ/S72G/A330W, with the simulation structures of GV/A184I/I263G/A328G (magenta) and VQ/S72G/A330W (cyan). Red dashes denote H-bonds. Black dashed circles denote steric clashes between poses and residues.

Overall, the docking of substrate **5** with the GV/A184I/I263G/A328G variant and the comparison with the VQ/S72G/A330W docking showed that substrate docking and especially the weighted selectivity were in qualitative agreement with the major oxidation product with correct absolute stereochemistry. The overlay of the lowest energy (*S*)-β pose with other

structures supported the previous hypothesis that (*S*)-β selectivity could be maintained if the (*S*)-β pose remained unaffected by extra mutations. Therefore, the NP poses and poses for other products were examined to provide insight into the binding interactions and potentially guide mutation design.

3.2.2.4 The NP poses

Two clusters of NP poses were identified. The major cluster resided in the pocket formed by L75, G328, A330 and L437, while the other NP poses were close to the I helix and contacted V87, I263 and T268 (Fig 3.19-A).

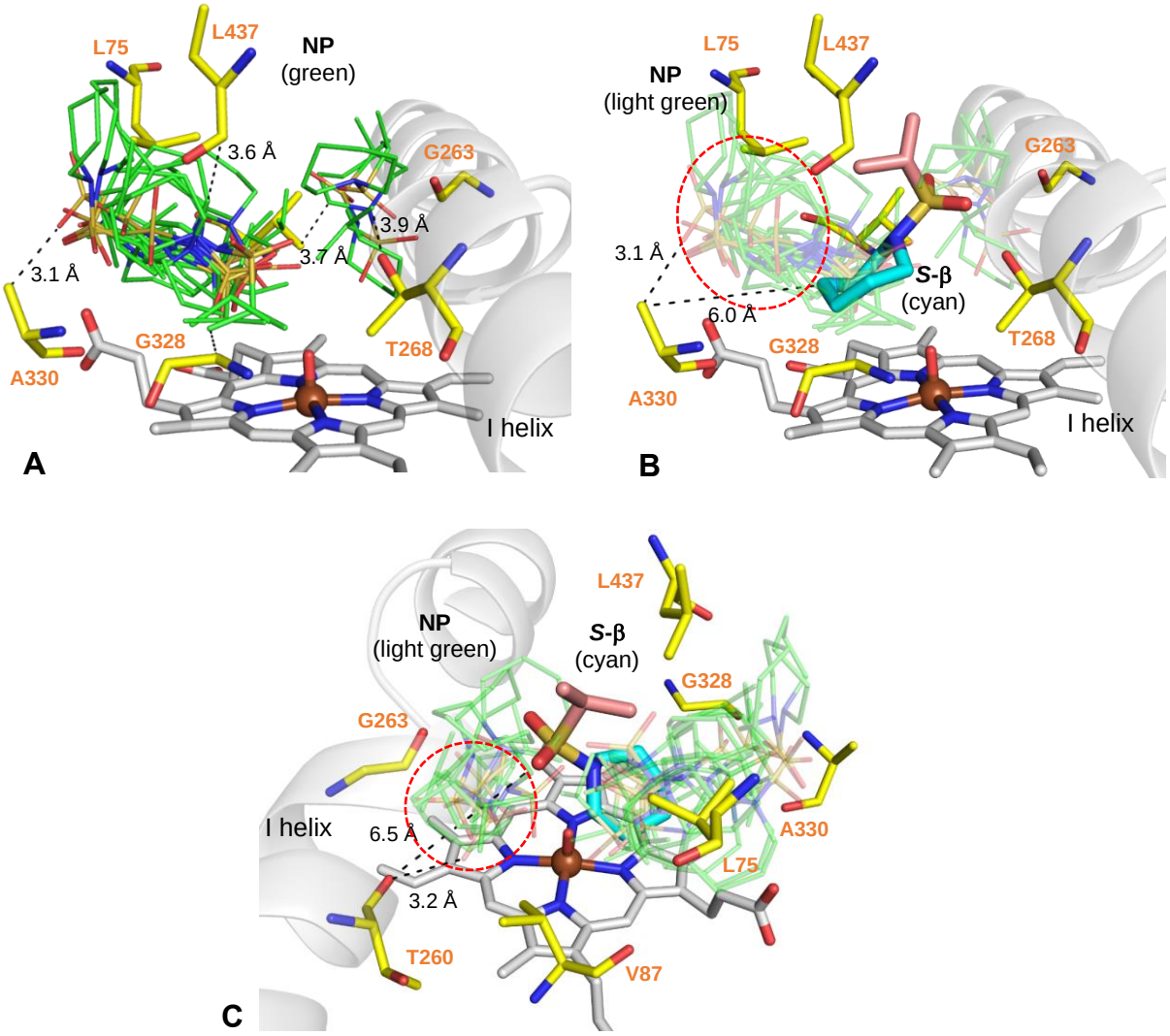


Figure 3.19 Overlay of simulation structure of GV/A184I/I263G/A328G (yellow) with **(A)** all NP poses (green) and **(B)** the lowest energy (*S*)-β pose (cyan) with NP poses (light green). **(C)** shows a different angle of **(B)**. Black dashes with black labels denote distances between residues and poses. Red dashed circles highlight the region where NP poses are bound.

The lowest energy (*S*)-β pose was then overlaid with these NP pose clusters. Similar to the observations in the VQ/S72G/A330W docking analysis, the major cluster of NP poses occupied a region of pocket where the (*S*)-β pose was not located (Fig 3.19-B). If mutations could be designed to disfavour the NP poses without interfering with the (*S*)-β poses, then it could lead to improvement of enzymatic activity without affecting the selectivity for (*S*)-β alcohol. For instance, A330 was close to the major cluster of NP poses (3.1 Å) but far away from the (*S*)-β pose (6.0 Å) (Fig. 3.19-B). Mutations A330I/L/F/M could enhance the turnover numbers without losing the (*S*)-β selectivity. Similarly, T260 was close to a cluster of NP poses (3.2 Å) but far away from the (*S*)-β pose (6.5 Å) (Fig. 3.19-C). Mutations T260F/I/L were thus proposed to promote enzymatic activity. GV/A184I/I263G/A328G would be a good template to test these ideas as it showed low activity for substrate **5** (TTN = 300).

3.2.2.5 Binding poses for other products

The (*R*)-β poses showed poorer alignment. The azepane ring mainly contacted residues S72, L75 and V87. In some poses, the Ips protecting group was close to the I helix and contacted I263, but this protecting group pointed towards the β1 strand in other poses (Fig. 3.20-A).

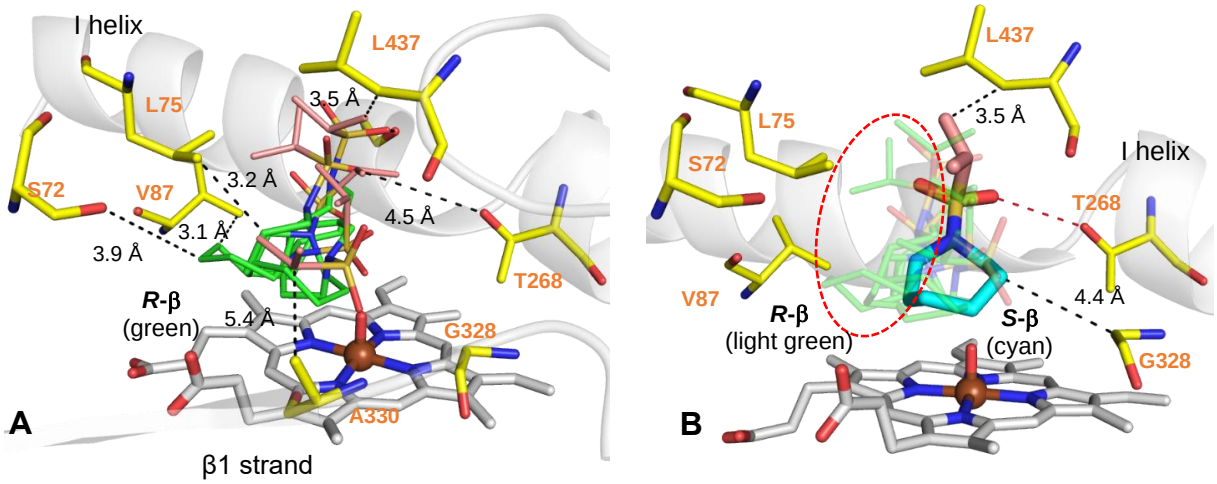


Figure 3.20 Overlay of simulation structure of GV/A184I/I263G/A328G with **(A)** all (*R*)-β poses (green) and **(B)** the lowest energy (*S*)-β pose (cyan) with (*R*)-β poses (light green). Black dashes with black labels denote distances between residues and poses. Red dashed circles highlight the region where (*R*)-β poses are bound.

From the overlay with the lowest energy (*S*)-β pose, a binding pocket for (*R*)-β poses was identified in the vicinity of S72, L75 and V87 (Fig. 3.20-B). These residues were further away from the lowest energy (*S*)-β poses (>4.5 Å). Based on the hypothesis of "destabilising unwanted poses without interfering with the productive pose to improve product selectivity", the bulky substitutions S72F/M, L75F/M and V87L/I could be proposed.

3.2.2.6 Summary of docking analysis of *N*-Ips-azepane

The GV/A814I/I263G/A328G docking predicted the (*S*)-β alcohol as the major product and the lowest energy pose was for (*S*)-β oxidation, in line with experimental data. The (*S*)-β poses aligned well. Comparison with the docking results of the (*R*)-selective VQ/S72G/A330W variant suggested that the lowest energy (*S*)-β pose generated by the (*S*)-selective variant would be destabilised in the (*R*)-selective variant, and *vice versa*. These observations suggested that the docking approach may provide a binding model for the major oxidation product with the correct bias of both regio- and stereoselectivity. Analysis of (*S*)-β poses *vs*. NP poses and poses for other products revealed significant differences in their binding interactions which led to the design of mutations to destabilise unwanted poses to improve enzymatic activity and product selectivity. Since the GV/A184I/I263G/A328G variant only showed moderate selectivity for the (*S*)-β alcohol [76% of **5c** and 79% *ee* for (*S*)-**5c**] and low activity (TTN = 300), the designed mutations of S72, L75, V87 and A330 were introduced to this variant to test their effects.

3.3 Docking guided mutagenesis study for azepane

3.3.1 Mutagenesis based on docking analysis of GV/A184I/I263G/A328G

3.3.1.1 Mutations designed for enzymatic activity

Mutations of T260 and A330 designed to disfavour NP poses from docking analysis were tested for their effects on the GV/A184I/I263G/A328G variant. The T260F, T260I and T260L variants showed slightly improved TTN, at 450, 410 and 470, respectively, compared with 300 for the template. However, they decreased the selectivity significantly to ~60% of **5c** and ~55% *ee* for (*S*)-**5c** (Fig. 3.21). In the overlay of the lowest energy (*S*)-β pose *vs*. NP poses, T260 was close to the minor cluster of NP poses and not far away from the (*S*)-β pose. The T260 mutations might have affected both the NP and (*S*)-β poses, leading to slightly improved turnover numbers but lowered (*S*)-β selectivity.

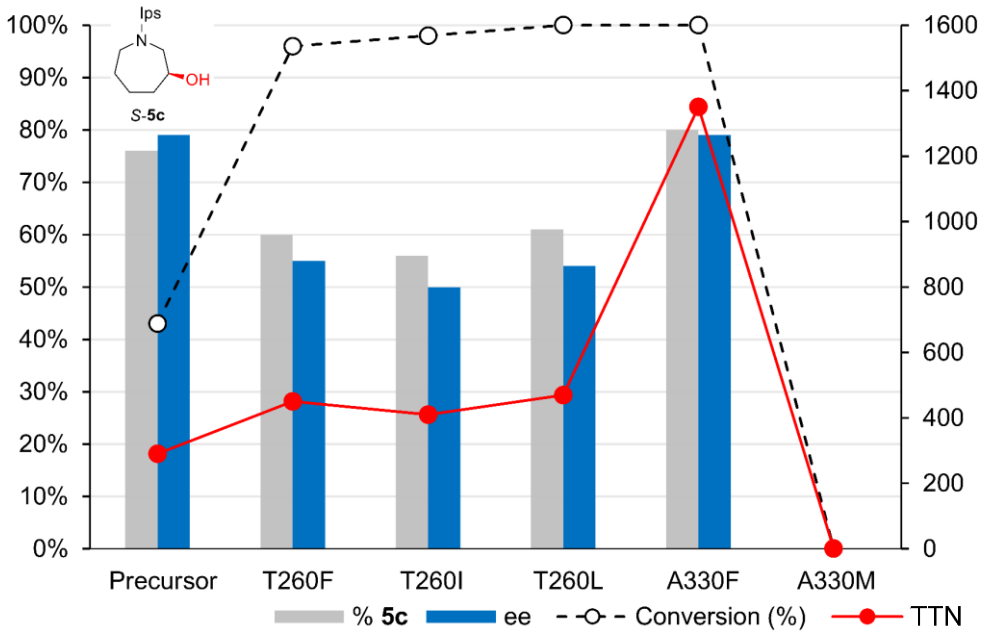


Figure 3.21 Effect of the T260F/I/L and A330F/M mutations. Precursor: GV/A184I/I263G/A328G.

It was striking that the A330F variant showed greatly enhanced activity, increasing the TTN to 1350, more than four-fold of the template, while maintaining the regio- and enantioselectivity

(80% of **5c** with 79% *ee*). This supported the proposal that bulky substitutions of A330 could destabilise NP poses without affecting the (*S*)-β poses. On the other hand, the A330M variant showed no activity at all (Fig. 3.21); the side chain of the Met residue could have forced the substrate to bind too far away from the haem.

3.3.1.2 Mutations designed for product selectivity

Mutations F87I/L, S72F/M and L75F/M were tested as they could disfavour the (*R*)-β poses and therefore improve enantio- and regioselectivity. The F87I variant showed enhanced activity and regioselectivity for **5c** and a slightly reduced ee (89% of **5c** with 74% *ee*, TTN = 510, Fig. 3.22). The F87L mutation, on the other hand, significantly improved both product selectivity and enzymatic activity, affording 94% of **5c** with 88% ee (TTN = 720).

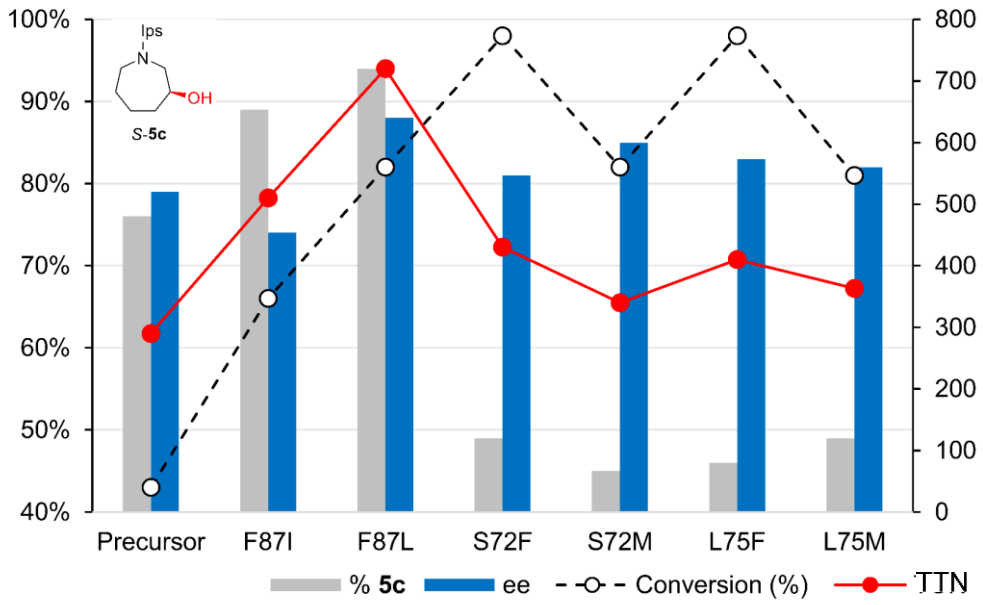


Figure 3.22 Effect of the F87I/L, S72F/M and L75F/M mutations. Precursor: GV/A184I/I263G/A328G.

The S72F, S72M, L75F and L75M mutations had similar effects, decreasing the regioselectivity to ~50% of **5c** but improving the enantioselectivity (81%–85% *ee*) and activity (TTN = 330–430) (Fig. 3.22). These results suggested that the (*R*)-β poses were destabilised by the S72 and L75 mutations with the (*S*)-β poses less affected. However, these "interrupted" (*R*)-β poses

might have adopted orientations that lead to α and γ oxidation, which contributed to the decreased regioselectivity.

Overall, the F87, S72 and L75 mutations tested all improved *ee* for (*S*)-β oxidation as predicted, although some of them decreased the regioselectivity for the β-alcohol **5c**. The F87L variant showed the most significant improvement. MD simulations of three replicas and docking with substrate **5** was performed with the F87L variant to check if the improvement in product selectivity could be reflected in the docking outputs.

Eight clusters, constituting 90% of the total population for each replica, were generated. Of the 72 poses produced, there were 12 poses indicating (*S*)-β oxidation of **5**, two for (*R*)-β oxidation, eight for α oxidation, ten for γ oxidation, together with 40 NP poses. The lowest energy pose was for (*S*)-β oxidation (Fig. 3.23). An (*S*)-β:(*R*)-β ratio of 6 was predicted from the pose count, compared to 15 from experiment. The predicted α (25%) and γ (31%) selectivities were much higher than the experimental data (both 3%) (Table 3.9). The cluster population and affinity energy weighted *S*:*R* ratio of 12 was closer to the value of 15 experimentally. The weighted β oxidation selectivity (64%) was still low compared with 94% from experiment, due to over-estimated α (17%) and γ oxidation (19%) (Table 3.9). Comparison of this data with that from the GV/A184I/I263G/A328G precursor suggested that substrate docking analysis could enable improvements in enantioselectivity of products with certain mutations introduced, but less so for the regioselectivity.

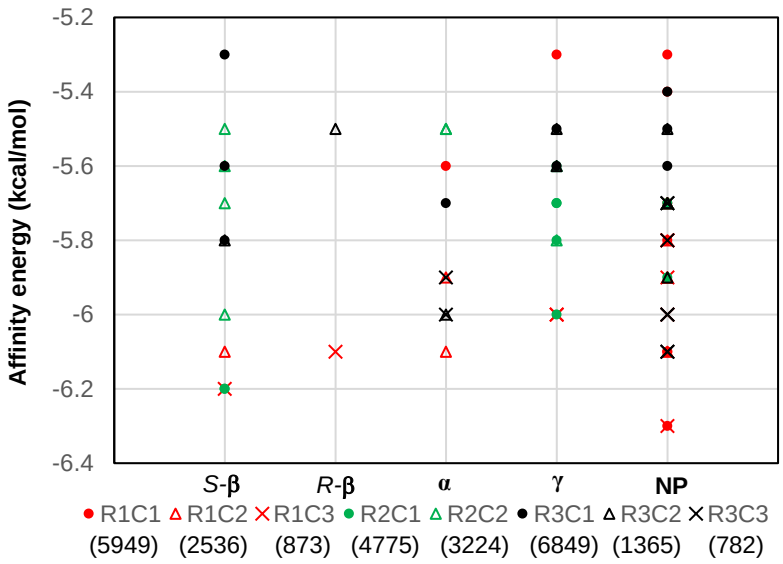


Figure 3.23. The affinity energy of output poses from Vina docking of N-Ips-azepane, **5**, with MD-simulated structures of the variant GL/A184I/I263G/A328G. RxCy denotes cluster y of replica x. The numbers in brackets indicate the population of each cluster.

Table 3.9 Predicted selectivity for oxidation products of the substrate **5** calculated from docking with MD-simulated structures of the F87L variant *vs.* GV/A184I/I263G/A328G and their corresponding experimental data.

Variants		Pro-*S*-β	Pro-*R*-β	*S*-β/*R*-β	Pro-α	Pro-γ
F87L variant	Number of poses	12	2	6	8	10
	Predicted selectivity based on number of poses	38%	6%	6	25%	31%
	Population/energy-weighted selectivity	59%	5%	12	17%	19%
	Experimental data	88%	6%	15	3%	3%
GV/A184I/I263G/A328G (*S*-selective, F87V variant)	Number of poses	10	4	2.5	8	6
	Predicted selectivity based on number of poses	36%	14%	2.5	29%	21%
	Population/energy-weighted selectivity	51%	13%	4.0	23%	13%
	Experimental data	68%	8%	8.5	16%	8%

The combinations of F87I/L mutations with S72F/M and L75F/M could be tested to explore further improvement in selectivity and especially enantioselectivity for (*S*)-**5c**. These variants could also be further combined with the A330F mutation for enzymatic activity enhancement.

3.3.1.3 Combining effective mutations

Upon addition to the F87I variant, the S72F, S72M and L75M mutations improved the *ee* but decreased the regioselectivity while maintaining the activity. The F87I/L75M, F87I/S72F and F87I/S72M variants showed very similar results. The S72F version marginally out-performed the others, affording 77% of **5c** with 86% *ee* (TTN = 500); there was a "trade-off" between regioselectivity and *ee*. In summary, the S72 and L75 mutations maintained their enantioselectivity-enhancing effect on addition to the F87I variant. The loss in regioselectivity was in an acceptable range (Fig. 3.24).

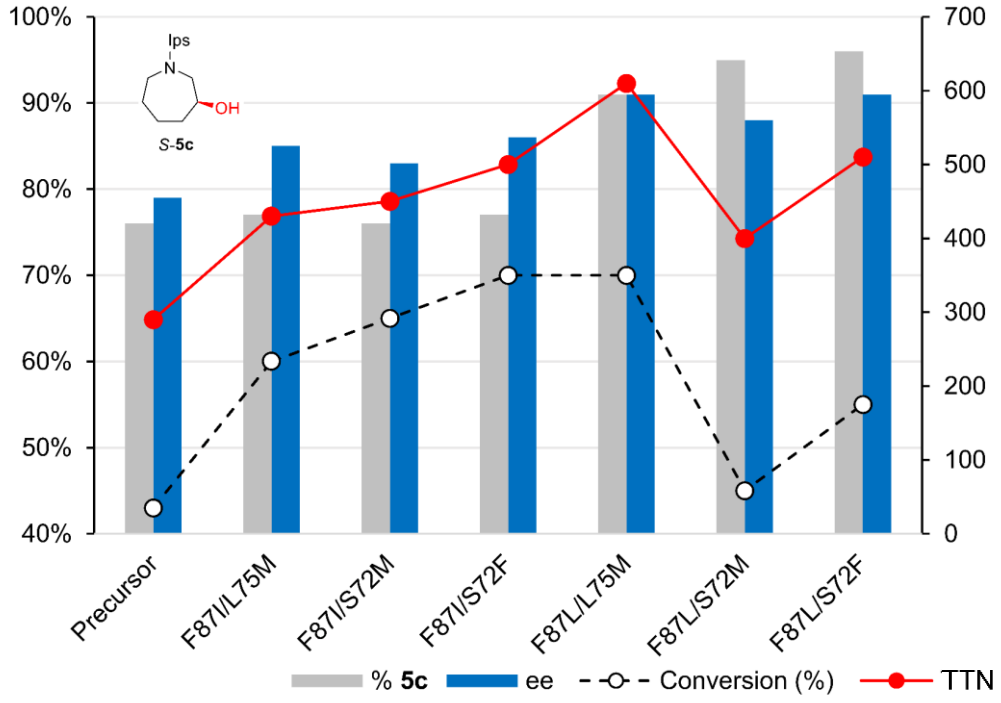


Figure 3.24 Effect of the F87I/L75, F87I/S72, F87L/L75 and F87L/S72 combined mutations. Precursor: GV/A184I/I263G/A328G.

The F87L series showed further improvement in product selectivity and especially enantioselectivity. The F87L/L75M variant gave 91% of **5c** with 91% *ee* and a TTN of 600 (Fig. 3.24), compared with 94% of **5c** with 88% *ee* (TTN = 720) by the F87L variant (Fig. 3.22). In contrast with the F87L/L75M variant, the F87L/S72F showed higher regioselectivity and the same *ee*, but lowered activity (96% of **5c**, 91% *ee*, TTN = 510). The F87L-containing variants

showed higher regio- and enantioselectivity for (*S*)-**5c**, with the F87L/S72F variant being the most selective.

The A330F mutation was then introduced to the F87L series to seek improvements of enzymatic activity. The F87L/A330F variant showed significantly improved activity and maintained the selectivity, giving 96% of **5c** with 87% ee (TTN = 1450), compared with 94% of **5c** with 88% ee (TTN = 720) by the F87L variant. However, the further combined variants F87L/L75M/A330F and F87L/S72F/A330F (95%/94% of **5c** with 84%/86% ee, TTN = 1320/1340) showed no improvements over the F87L/A330F variant (Fig. 3.25).

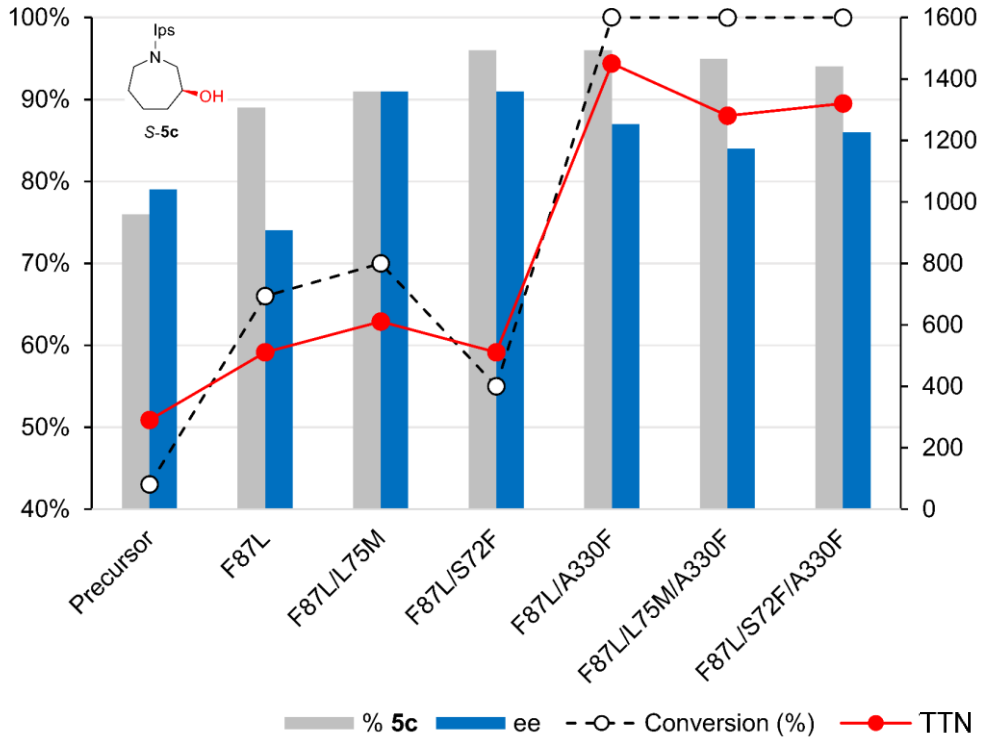


Figure 3.25 Effect of the F87L/A330F, F87L/L75M/A330F and F87L/S72F/A330F combined mutations. Precursor: GV/A184I/I263G/A328G.

Overall, single-residue mutagenesis based on docking analysis of the GV/A184/I263G/A328G variant with substrate **5** led to significant improvements in product selectivity and enzymatic activity. Combination of effective mutations achieved further improvement, with the F87L/A330F pair being the most active and with high selectivity (96% of **5c** with 87% *ee*, TTN

= 1450), a significant increase over the precursor (76% of **5c** with 79% *ee*, TTN = 300) (Fig. 3.25). However, the fact that the F87L/S72F/A330F and F87L/L75M/A330F failed to achieve further improvements suggested that combinations of individually effective mutations might not give the anticipated enhancements. This could arise from conflict between the effect of mutations. One mutation may have altered the binding orientation of the substrate, such that additional mutations designed on the basis of the precursor mutant would not show the expected effect.

3.3.2 Docking studies on the GL/A184I/I263G/A328G/S72F variant

3.3.2.1 General outputs from docking calculations

Instead of testing more combinations of effective mutations, further docking analysis was performed. The F87L/S72F variant, which was one of the most selective but showed a lower TTN, was chosen for docking study to design mutations to increase the enzymatic activity.

Substrate **5** was docked with eight clusters from three replica simulations of the F87L/S72F variant. Among the 72 poses generated, there were 11 poses indicating (*S*)-β oxidation of substrate **5**, two for *R*-β oxidation, seven for α oxidation and ten for γ oxidation, together with 42 NP poses (Fig. 3.26). (*S*)-β poses dominated the lower energy productive poses (from –6.2 to –5.8 kcal/mol) and the lowest energy productive pose was for (*S*)-β oxidation. The predicted (*S*)-β:(*R*)-β ratio based only on number of poses (5.5) was much lower than 23 from experimental data. The cluster population and affinity energy weighted the *S*:*R* ratio of 16, was closer to experimental observation. The weighted α (12%) and γ oxidation (14%) selectivity were still higher than experiment (1% and 3% respectively) (Table 3.10).

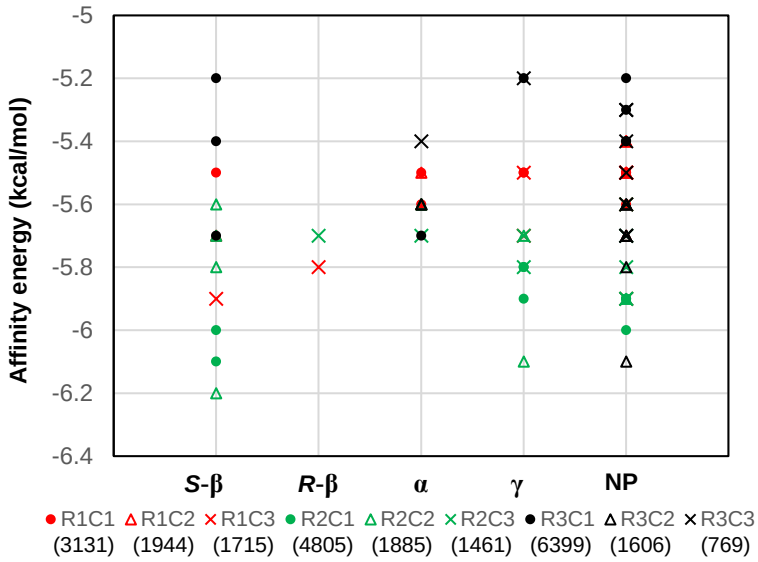


Figure 3.26 The affinity energy of output poses from Vina docking of *N*-Ips-azepane, **5**, with MD-simulation structures of the variant GL/A184I/I263G/A328G/S72F. The numbers in brackets indicate the population of each cluster.

Table 3.10 Predicted selectivity for oxidation products of the substrate **5** calculated from docking with MD-simulation structures of the F87L/S72F variant *vs.* GV/A184I/I263G/A328G and their corresponding experimental data.

Variants		Pro-*S*-β	Pro-*R*-β	*S*-β/*R*-β	Pro-α	Pro-γ
F87L/S72F variant	Number of poses	11	2	5.5	7	10
	Predicted selectivity based on number of poses	38%	6%	6	25%	31%
	Population/energy-weighted selectivity	69%	5%	14	12%	14%
	Experimental data	92%	4%	23	1%	3%
GV/A184I/I263G/A328G (*S*-selective, F87V variant)	Number of poses	10	4	2.5	8	6
	Predicted selectivity based on number of poses	36%	14%	2.5	29%	21%
	Population/energy-weighted selectivity	51%	13%	4.0	23%	13%
	Experimental data	68%	8%	8.5	16%	8%

Despite the less well-predicted regioselectivity, the enantioselectivity-enhancing effect of the F87L/S72F was successfully demonstrated by substrate docking. Analysis of the output poses and binding interactions would provide insight and guide mutation design. Since the F87L/S72F variant was already highly selective, the comparison of (*S*)-β poses *vs.* NP poses would be especially meaningful.

3.3.2.2 Analysis of poses and mutagenesis

The (*S*)-β poses adopted similar binding orientation as those from docking with the precursor, GV/A184I/I263G/A328G: The sulfonyl oxygens of the *N*-Ips protecting group were in van der Waals contact with the I helix residues G263, A264 and T268, and the azepane ring pointed towards the β1 strand (Fig. 3.27-A). In the lowest energy (*S*)-β pose, the Cβ was 3.2 Å from the ferryl oxygen. Hydrogen bonding between sulfonyl oxygen and T268 was observed (Fig. 3.27-B).

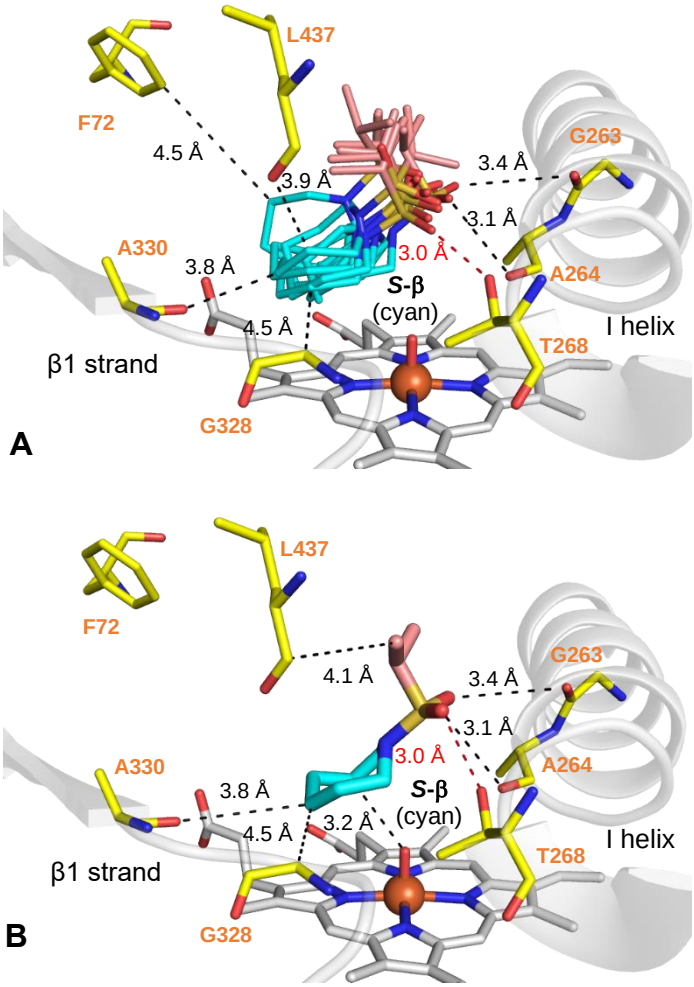


Figure 3.27 Overlay of MD-simulation structures of GL/A184I/I263G/A328G/S72F (yellow) with **(A)** all (*S*)-β poses (cyan), **(B)** the lowest energy (*S*)-β pose (cyan). Black dashes with black labels denote distances between residues and poses. Red dashes denote H-bonds. Black dashed circles denote steric clashes between poses and residues.

Two adjacent clusters of NP poses were identified: the "upper" cluster was in the pocket formed by F72, L181 and L188, while the other cluster was closer to the haem and contacted L87, A264 and T268 (Fig. 3.28-A). From the pose overlay, there was a significant region of pocket for NP poses where the (*S*)-β pose was not found. From the comparison, it was striking how bulky substitutions at L188 could destabilise the NP poses without affecting the (*S*)-β pose, as L188 was close to the main cluster NP poses (3.1 Å) but much farther away from the (*S*)-β pose (9.1 Å) (Fig. 3.28-B).

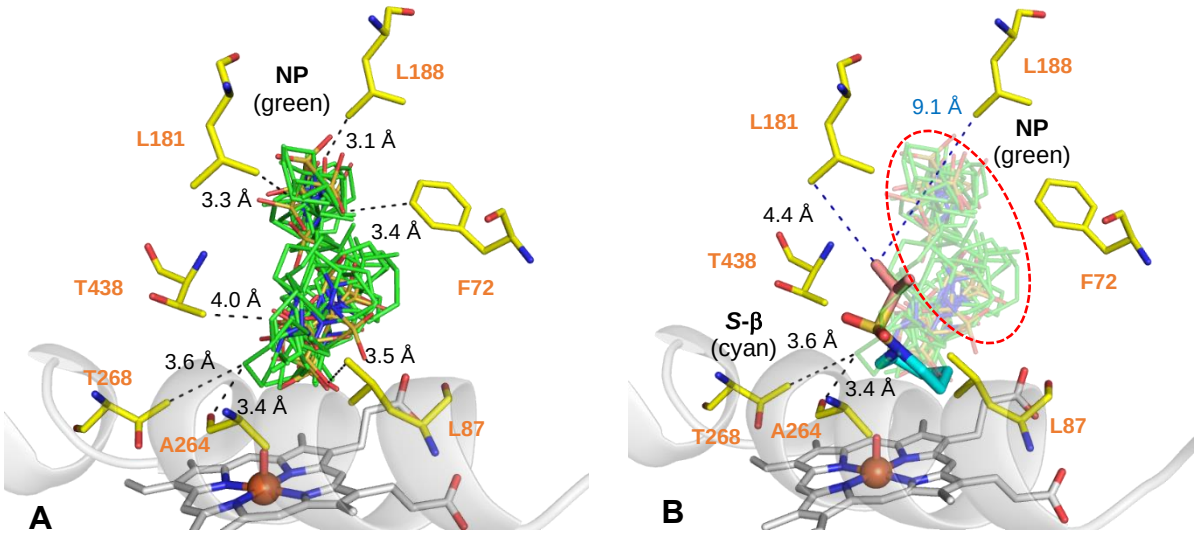


Figure 3.28 Overlay of simulation structure of GL/A184I/I263G/A328G/S72F (yellow) with **(A)** all NP poses (green) and **(B)** the lowest energy (*S*)-β pose (cyan) with NP poses (light green). Black dashes with black labels denote distances between residues and poses. Red dashed circles highlight the region where NP poses are bound.

The L188F/M/W mutations were introduced to the F87L/S72F variant. Experimentally, all three variants maintained the selectivity and improved TTN for (*S*)-**5c**. The F87L/S72F/L188F and F87L/S72F/L188W variants gave, respectively, 95% and 94% of **5c** with 91% and 92% *ee*, TTN = 900 and 990, but the GL/A184I/I263G/A328G/S72F/L188M variant showed the best selectivity and activity, giving 97% of **5c** with 92% ee, and TTN of 1490 (Fig. 3.29).

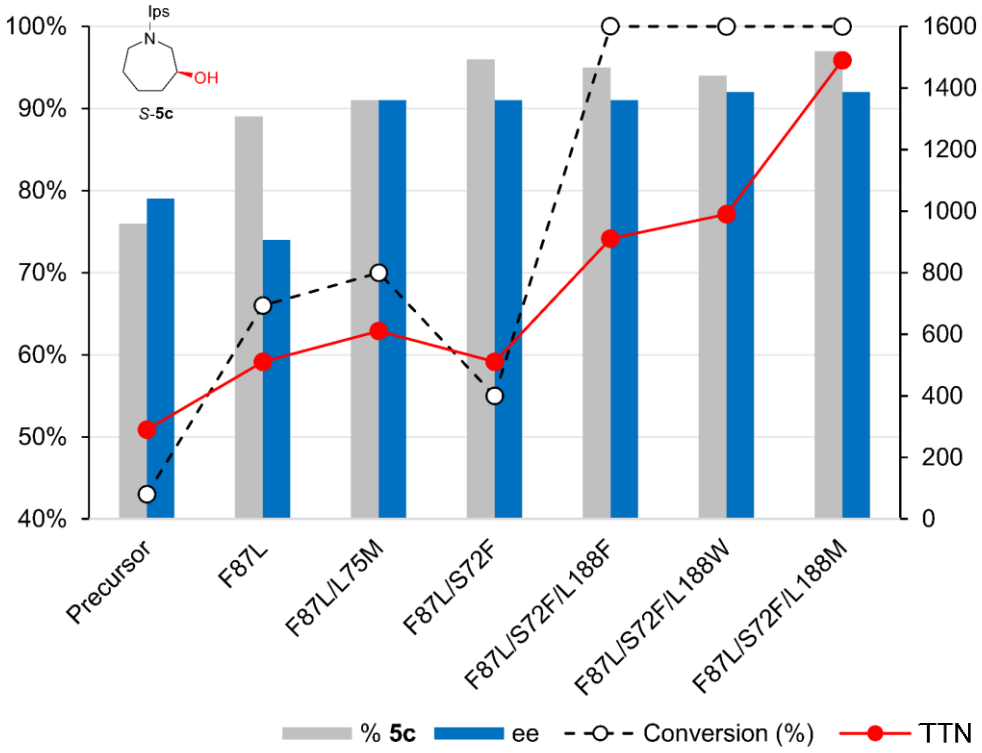


Figure 3.29 The effect of the L188 mutations. Precursor: GV/A184I/I263G/A328G.

To summarise, docking analysis of **5** with GV/A184I/I263G/A328G predicted the "correct" oxidation product and (*S*)-β stereochemistry found by experiment. These "correct" poses aligned well and comparison with NP poses and poses for other products led to mutation designs that improved product selectivity and enzymatic activity. Combination of the effective mutations showed further improvements and the F87L/S72F variant was identified as a highly selective but slightly less active variant. Docking analysis on this variant reflected the enantioselectivity-enhancing effects observed from experiments. Comparison of the (*S*)-β poses with NP poses led to the discovery of the L188M mutation which greatly enhanced the TTN. Overall, the three rounds of docking guided mutagenesis required 30 variants generated and tested; the product selectivity was improved to 97% of **5c** with 92% *ee* by the F87L/S72F/L188M variant, compared with 76% of **5c** and 79% *ee* by the template, and the TTN improved to 1490, nearly five times that of the template (300).

3.4 Docking analysis for the γ-selective variant

3.4.1 Vina docking analysis for the KU3/A330P/S72W variant

3.4.1.1 MD simulations and docking of the KU3/A330P/S72W variant

Having achieved high selectivity and activity for the (*S*)-enantiomer of the β-alcohol by docking-guided mutagenesis, docking analysis of the γ-selective variants was carried out. High selectivity was observed for the (*R*)-γ alcohol product of *N*-Boc-azepane **1** at 90% of **1a** and 90% *ee* for (*R*)-**1a** (TTN = 10060) with the RLYF/H171L/I263G/A330W/L437S variant. The (*S*)-selective variant KU3/A330P/S72W showed a high regioselectivity of 90% of **1a** but a slightly lower enantioselectivity at 84% *ee* for (*S*) and lower TTN (8440), and was chosen to perform initial docking analysis (Table 3.11).

Table 3.11 Best variants and selectivity for two enantiomers of the γ-alcohol product **1a** of azepane.

Products	Variant	Regioselectivity	*ee*	TTN
(*R*)-γ-ol, (*R*)-**1a**	RP/H171L/I263G/A330W/L437S	90%	90%, *R*	10060
(*S*)-γ-ol, (*S*)-**1a**	KU3/A330P/S72W	90%	84%, *S*	8440

Three replica 100-ns MD simulations were carried out for KU3/A330P/S72W. The low RMSD (1–2 Å) values indicated stable structures for all three simulations (Figure A5.11-A, Appendix **A5.4.2**). Eight clusters consisting of 90% of the total population for each replica were generated (Figure A5.11-B, Appendix **A5.4.2**). Substrate **1** was docked into these eight clusters. Surprisingly, among the 72 poses generated, very few productive poses were found: there were two poses for (*S*)-γ oxidation, one for (*R*)-γ, one for α, and one for β, together with 67 NP poses (Fig. 3.30). Moreover, all productive poses were from replica 3. To explore this, the structures of each cluster were analysed. As with previous observation, all clusters within the same replica showed little deviation. The most populated clusters from each replica, R1C1, R2C1 and R3C1, were overlaid and further examined.

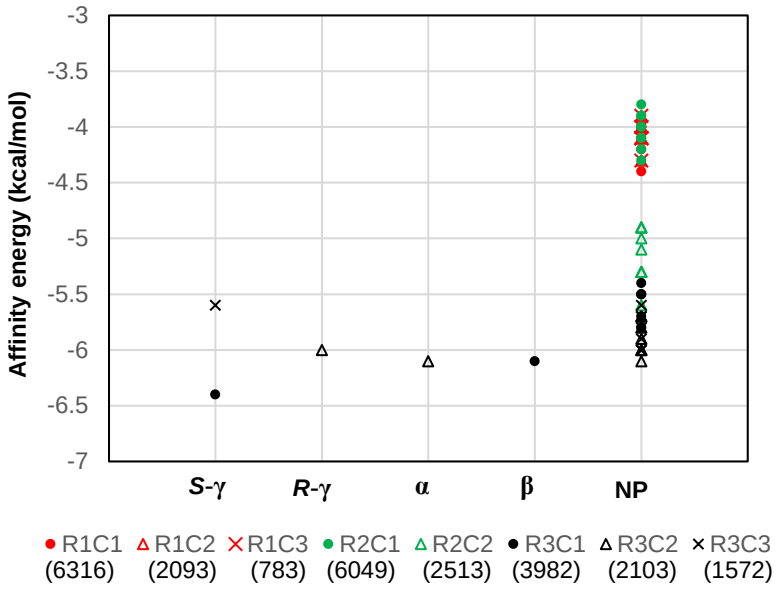


Figure 3.30 The affinity energy of output poses from Vina docking of *N*-Boc-azepane, **1**, with MD-simulated structures of the variant KU3/A330P/S72W. RxCy denotes cluster y of replica x. The numbers in brackets indicate the population of each cluster.

3.4.1.2 Structural analysis for the simulated clusters

From the overlay, the F87 side chain of R1C1 and R2C1 were directly above the haem and close to the ferryl oxygen (4.2 Å). This was the likely reason for the absence of productive poses, as the substrate would be "pushed away" by the benzene ring of F87 (Fig. 3.31). On the other hand, the F87 side chain in R3C1 was pulled back and further from ferryl oxygen (5.5 Å), which would leave more space for the substrate to approach haem. The close approach of A328 and F87 (~6.6 Å) in R1 and R2 also meant that productive poses could be destabilised by steric clashes, while in R3 these residues were farther apart (9.8 Å). Furthermore, the W72 side chain in R1C1 and R2C1 pointed away from the I helix, making a larger substrate binding pocket above the haem which might allow more NP poses. The R3C1 W72 was rotated by ~60° towards the I helix and restricted the pocket, which would force the substrate to bind closer to heme and more likely to generate productive poses (Fig. 3.31).

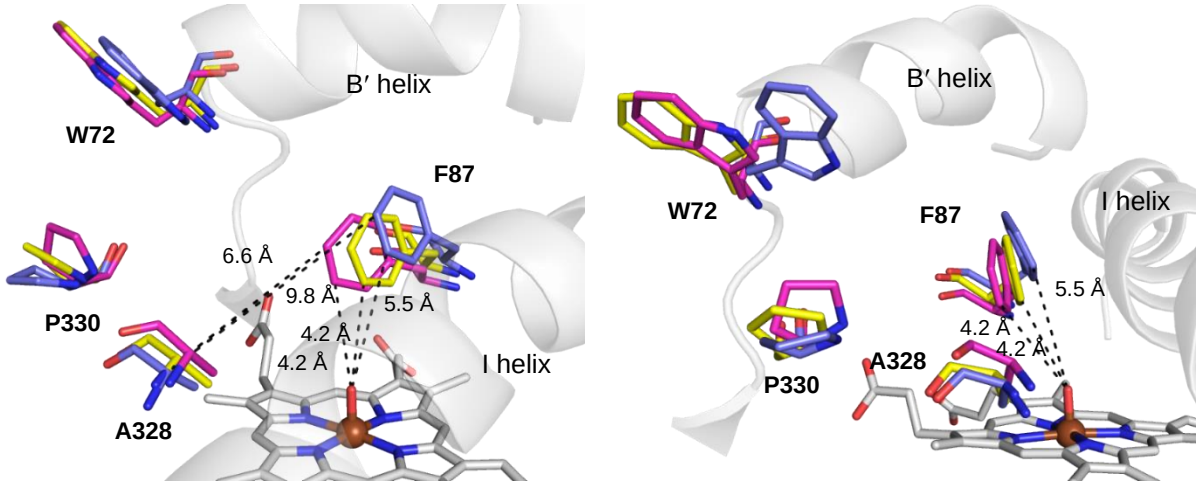


Figure 3.31 Overlay of selected residues from R1C1 (magenta), R2C1 (yellow) and R3C1 (blue) structures of KU3/A330P/S72W, with **(A)** and **(B)** showing different views. Black dashes with labels denote distances.

These observations suggested that, with Vina rigid-body docking, bulky residues around the active site, especially F87, could significantly affect the output and, in some cases, eliminate productive poses altogether. To address this, one obvious strategy was to increase the number of replicas when performing MD-simulations. However, this could be time-consuming and impractical because, as many as 100 replicas might be needed to give a close resemblance of the conformational space for the variant structure. Thus, flexible-receptor docking, where side chains of specified residues were allowed to undergo conformational changes, was explored.

To this end, the alternative docking approach of Autodock for Flexible Receptor (ADFR) was adopted, instead of the flexible-receptor docking function of Autodock Vina. Sanner *et al*. demonstrated that, compared with Vina, the ADFR approach showed greatly improved success rates when cross-docking ligands into *apo* receptors with the side-chains made flexible.[221] These cross-docking experiments were performed with datasets SEQ17, a receptor diversity set containing 17 pairs of *apo-holo* structures, and CDK2, a ligand set with one CDK2 *apo* structure and 52 known bound inhibitors. The ADFR method also down-weighted the contribution of the receptor internal energy in the scoring function, which resulted in increased ranking for the

correctly docked pose in the cross-docking experiments. ADFR docking runs were also less time-consuming: the run time increased linearly as the number of flexible side chains increased, while in Vina the run time scaled exponentially. Thus, the ADFR approach was explored for flexible-receptor docking analysis of substrate **1** with KU3/A330P/S72W.

3.4.2 The ADFR approach

In the ADFR approach, the docking site was defined by selecting residues, instead of a 30 × 30 × 30 Å box as in the Vina approach. With $P450_{BM3}$, residues 26, 72, 75, 78, 82, 87, 181, 260, 263, 328, 329, 330, 332, 354, 437 and 438 were selected to define the docking site, as these residues shaped the active site (Fig. 3.32).

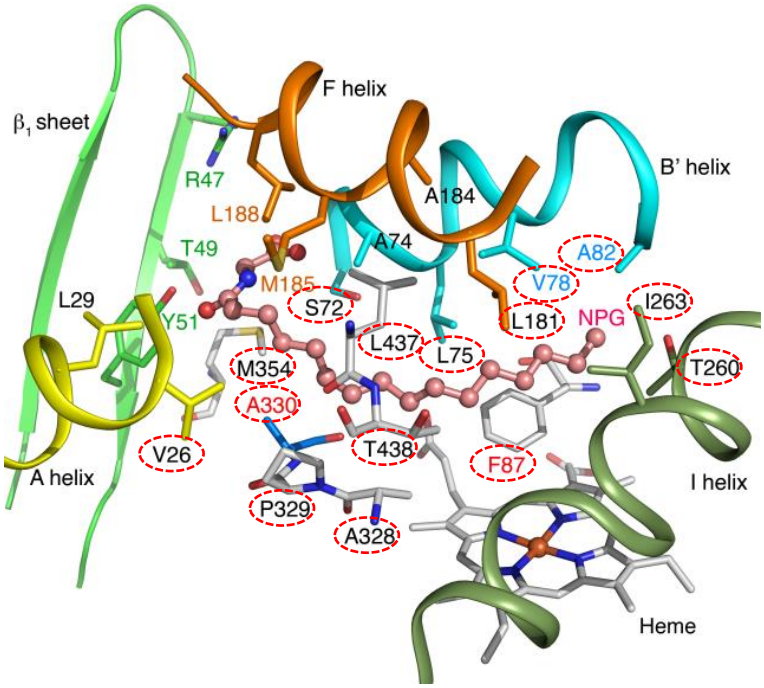


Figure 3.32 The active site structure of *N*-palmitoylglycine (NPG)-bound $P450_{BM3}$ and key amino acid residues. Red dashed circles denote the residues used to define the ADFR docking site.

Selection of flexible residues was then explored. In theory, the more flexible residues the more accurate the prediction would be. However, as the number of flexible residues increased, the docking run time also increased. Thus, a balance had to be struck between meaningful output

and run time. To explore this, substrate **1** was docked with the eight clusters using the ADFR approach with a different number of flexible residues (0, 4, 8, 12 and 16). These residues were selected from V26, W72, L75, V78, A82, F87, L181, T260, I263, A328, P329, P330, S332, M354, L437 and T438, as the side chains were in the active site and usually had significant effects on substrate binding. The outputs were examined and the number of productive poses and the "correct" pose, *i.e.* (*S*)-γ pose, were compared (Fig. 3.33). With no flexible side chains, the docking run with eight clusters took 15 mins, and few productive poses were generated: one pose for (*S*)-γ oxidation, one for (*R*)-γ, one for α, and one for β, which was similar to the Vina output. When the number of flexible residues was set as four (W72, F87, I263 and T438), the docking run time increased to 150 min. There were 28 productive poses generated, with nine (*S*)-γ and four (*R*)-γ poses. When eight residues were set to be flexible (W72, L75, F87, T260, I263 S332, L438 and T438), the number of productive poses was slightly higher, at 30, with 10 (*S*)-γ poses and four (*R*)-γ poses.

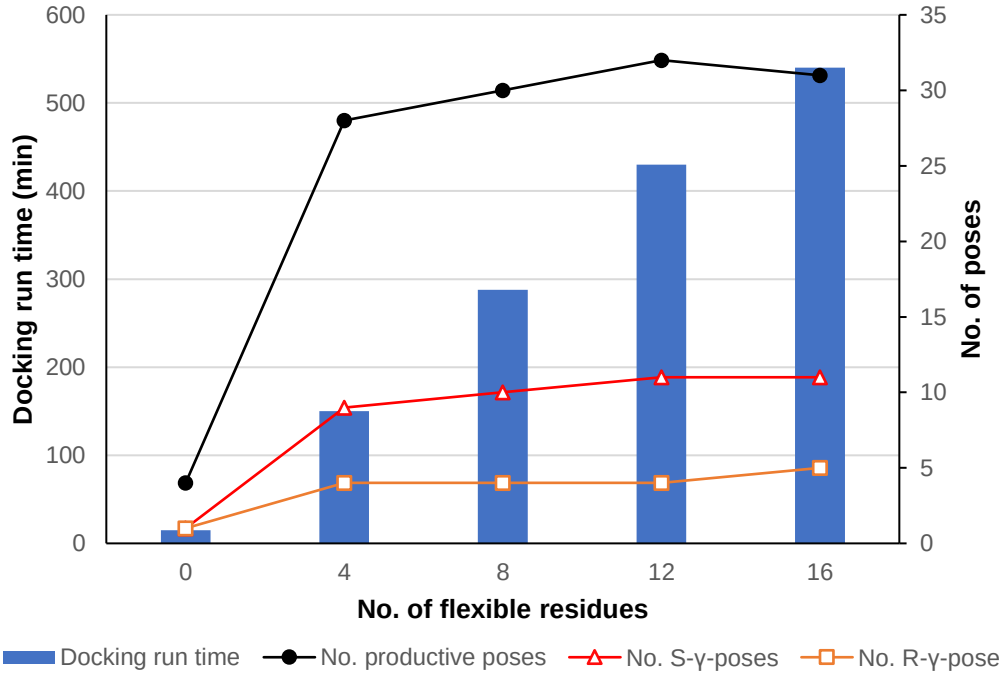


Figure 3.33 Docking run time, number of productive poses and "correct" poses generated by docking of substrate **1** with eight clusters of KU3/A330P/S72W using ADFR with 0, 4, 8,12 and 16 flexible residues.

As the number of flexible residues increased further to 12 and 16, there was no significant increase in the number of productive poses (32 and 31, respectively) and (*S*)-γ poses (both 11), but the total run time increased linearly (430 and 540 min). It was therefore concluded that having four to eight residues set as flexible was a good compromise between run time and meaningful outputs, for the docking of **1** with KU3/A330P/S72W.

3.4.3 Analysis of ADFR docking results

3.4.3.1 General outputs

The docking output for four flexible residues was analysed. In ADFR, the docking run with each cluster generated 50 binding conformations by default, which were then grouped based on affinity energy and binding orientation to give the final output poses. Docking with each cluster typically generated 5 to 20 grouped poses depending on the number of similar poses, instead of a fixed number of poses (nine for each cluster) in the Vina approach.

From the docking with eight flexible residues, there were 57 poses generated altogether. Of these, there were 10 poses for (*S*)-γ oxidation, four for (*R*)-γ, nine for α, and seven for β, together with 27 NP poses. The predicted (*S*)-γ to (*R*)-γ ratio based only on the number of poses was 2.5, much lower than 12 observed from experiment. Weighting by affinity energy and cluster population, as well as the group size of poses, increased the *S*:*R* ratio to 10, much closer to the experimental value of 12. However, the weighted α (27%) and β selectivity (18%) were both much higher than the experimental data (Fig. 3.34).

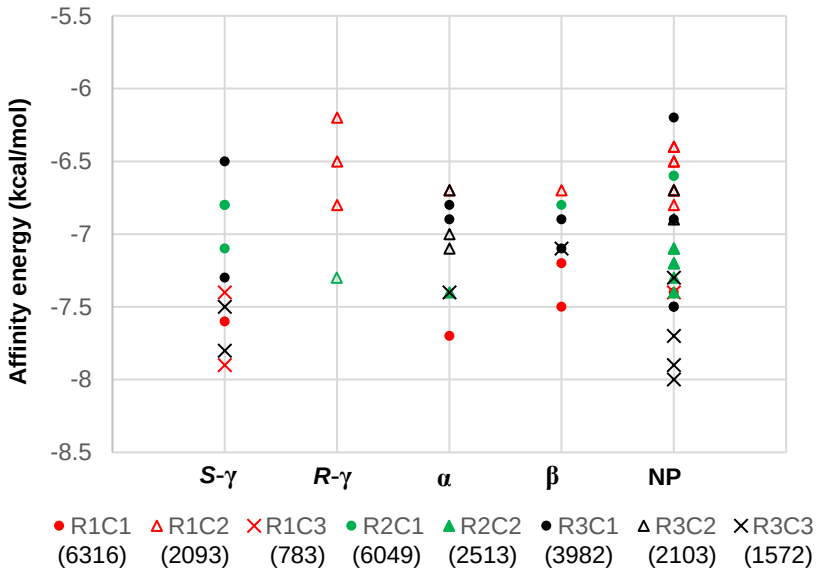


Figure 3.34. The affinity energy of output poses from docking of *N*-Ips-azepane, **5**, with MD-simulated structures of the variant KU3/A330P/S72W. RxCy denotes cluster y of replica x. The numbers in brackets indicate the population of each cluster.

Table 3.12 Predicted selectivity for oxidation products of the substrate **5** calculated from docking with MD-simulated structure of KU3/A330P/S72W and the corresponding experimental data.

	Pro-*S*-γ	**Pro-*R*- γ**	***S*- γ/*R*- γ**	**Pro-α**	**Pro-β**
Number of poses	10	4	2.5	9	7
Predicted selectivity based on number of poses	34%	13%	2.5	30%	23%
Population/energy/cluster size-weighted selectivity	50%	5%	10	27%	18%
Experimental data	83%	7%	12	4%	6%

To summarise, the ADFR docking of **1** with KU3/A330P/S72W showed many more productive poses than Vina docking with the same simulation structure. (*S*)-γ poses were observed across all replicas in ADFR docking, whereas with Vina only R3-structures showed two poses for (*S*)-γ oxidation. The R1 and R2 structures after ADFR docking were overlaid with the original structures to examine the movement of residues, especially F87.

From the overlaid R1C1 structures with the lowest energy (*S*)-γ pose, it was clear that the (*S*)-γ-pose would be destabilised by the original R1C1 F87 side chain, which was rotated by ~60° towards the B′ helix and moved out of the way after the ADFR run (Fig. 3.35-A). Furthermore,

the post-ADFR W72 side chain moved towards the (*S*)-γ pose and formed a hydrogen bond to the Boc group to stabilise the pose. The other two flexible residues L437 and I263 showed conformational changes but had no effect on the (*S*)-γ pose due to the long distances (6.8 Å and 7.1 Å).

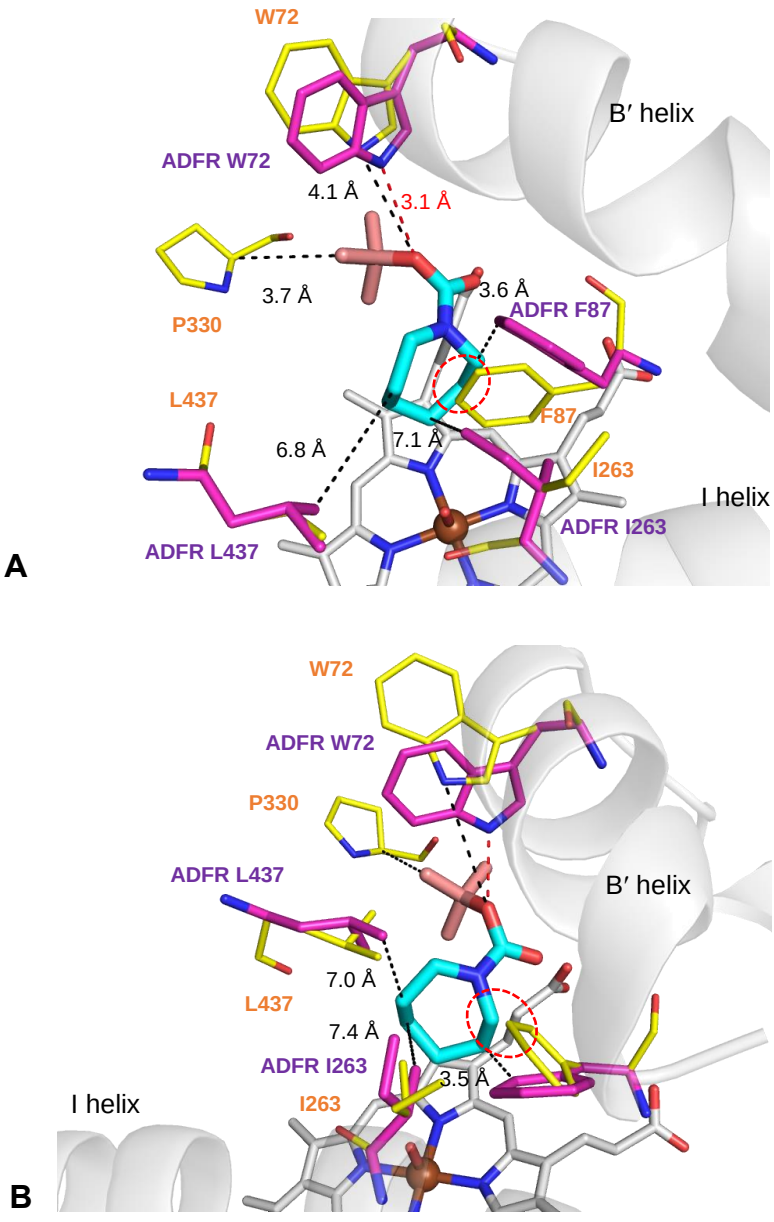


Figure 3.35 Overlay of selected residues (yellow) and their post-ADFR conformations (magenta) of the simulation **(A)** R1C1 and **(B)** R2C1 structures of KU3/A330P/S72W, together with the lowest energy (*S*)-γ-pose (cyan). Black dashes with black labels denote distances between residues and poses. Red dashes denote H-bond. Red dashed circles denote steric clashes.

Comparison of the original *vs*. post-ADFR R2C1 structures showed similar trends: the F87 side chain clashed with the (*S*)-γ pose in the simulation structure, but after ADFR run the side chain rotated away (by ~45° towards the I helix) and created space for the pose (Fig. 3.35-B). As with R1C1, the R2C1 W72 side chain was closer to the (*S*)-γ pose after ADFR run and hydrogen bonding was observed, while in the original structure W72 was farther from the pose. Both L437 and I263 were far from the (*S*)-γ pose even after ADFR (7.0 Å and 7.4 Å). These observations demonstrated that F87 was the key factor affecting the production of meaningful outputs: in R1 and R2 structures the binding regions for (*S*)-γ pose was hindered by the F87 side chain, but once conformational changes of F87 were allowed in ADFR, these regions were freed up, more productive poses were obtained.

However, due to the different parameters and algorithms, *etc*., direct comparison between the output distribution and predicted selectivity from Vina and ADFR approaches might not be meaningful. Instead, the binding poses for different products would provide insight in enzyme-substrate binding interactions and guide mutagenesis.

3.4.3.2 Analysis of binding poses

The (*S*)-γ poses showed good alignment; the Boc group was sequestered in the pocket formed by W72, P330 and S332, while the azepane ring pointed towards the I helix and contacted T260, I263 and T438. In the lowest energy pose, Cγ of the azepane ring was 3.2 Å from the ferryl oxygen (Fig. 3.36).

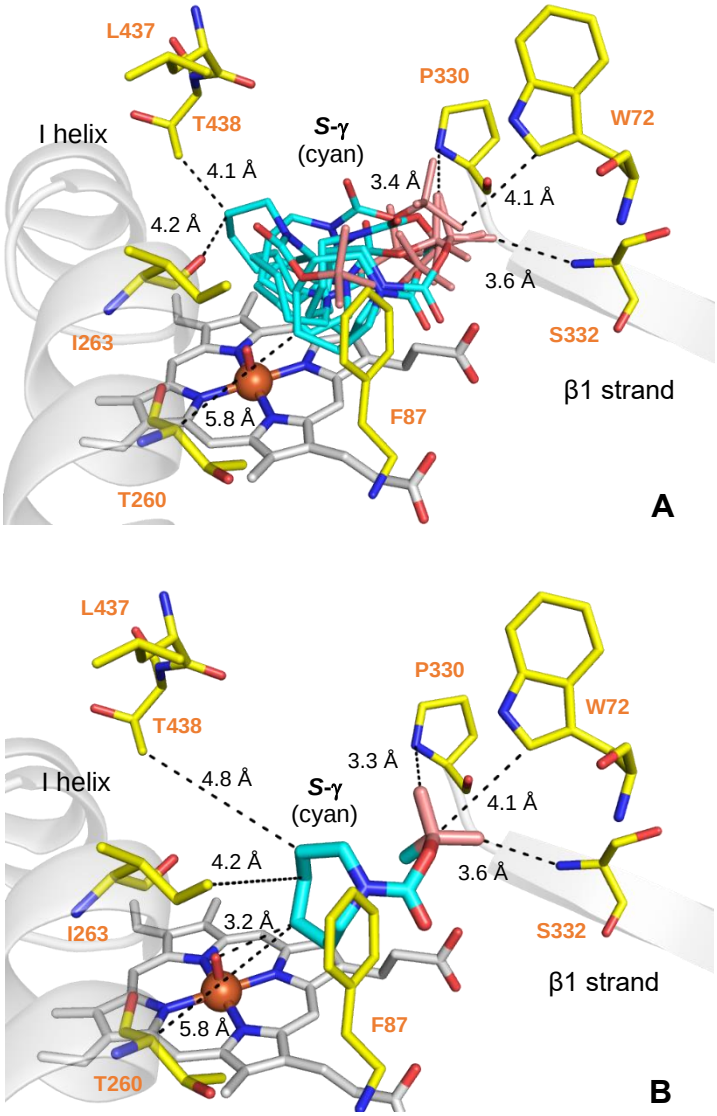


Figure 3.36 Overlay of MD-simulation structures of KU3/A330P/S72W (yellow) with **(A)** all (*S*)-γ-poses (cyan), **(B)** the lowest energy (*S*)-γ-pose (cyan). Black dashes with black labels denote distances between residues and poses.

The two (*S*)-γ poses from the Vina rigid-body docking run with the R3 structures were overlaid with the ADFR poses as comparison. From the overlay, the (*S*)-γ poses from these two approaches aligned reasonably well, with the Vina poses a little farther away from the I helix (Fig. 3.37). This indicated that given reasonable side chain conformations, as in R3 structures, Vina could give similar meaningful binding poses as ADFR, but flexible side chain docking was essential for variants with bulky residues such as F87 above the haem.

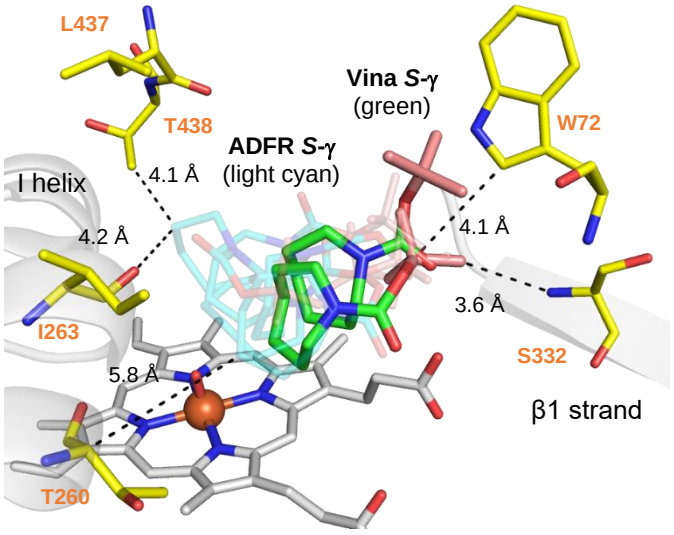


Figure 3.37 Overlay of all (*S*)-γ-poses (light cyan) generated from the ADFR approach and the (*S*)-γ-poses (green) from the Vina approach. Black dashes with black labels denote distances between residues and poses.

The NP poses were clustered above the haem, from around the I helix to the pocket formed by W72, P330 and S332 (Fig. 3.38-A). From the overlaid structure, two binding regions were identified for NP poses where the (*S*)-γ pose was not located; the major cluster was close to the I helix and contacted T438, while the other cluster was close to S332 and W72 (Fig. 3.38-B). Mutations T438L/F/M and S332F/M were thus designed to disfavour the NP poses and improve the enzymatic activity. The S332 mutations might not work as anticipated, as S332 was quite close to the (*S*)-γ pose (4.6 Å) and bulky substitutions here might thus affect the (*S*)-γ selectivity. On the other hand, the T438 mutations could be more successful as the T438 was close to the NP poses (3.3 Å) but much farther away from the (*S*)-γ pose (9.6 Å).

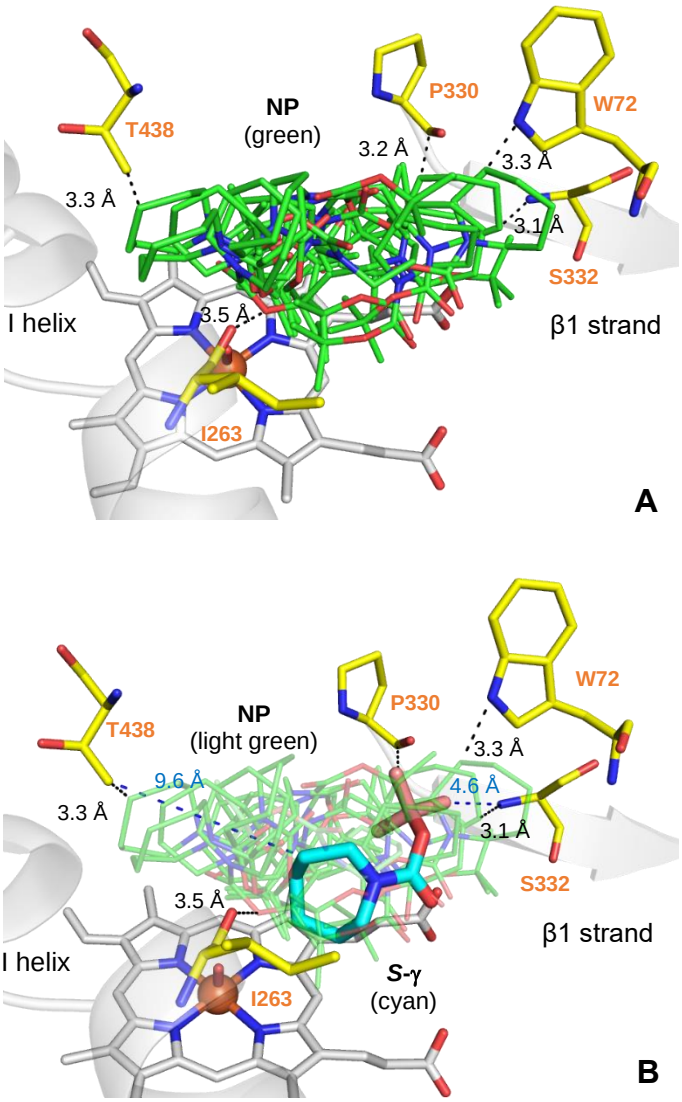


Figure 3.38 Overlay of MD-simulation structures of KU3/A330P/S72W (yellow) with **(A)** all NP poses (green), **(B)** all NP poses (light green) and the lowest energy (*S*)-γ-pose (cyan). Black dashes with black labels denote distances between residues and the (*S*)-γ-pose. Blue dashes with blue labels denote distances to the NP poses.

The (*R*)-γ poses adopted similar binding orientation as the (*S*)-γ poses. The protecting group contacted W72 and S332 while the azepane ring pointed towards the I helix. From the overlay, all (*R*)-γ poses were within the vicinity of the region of pocket formed by the (*S*)-γ poses and no specific binding region was found for (*R*)-γ poses (Fig. 3.39). Therefore, it was not straightforward to design mutations to seek improvement to enantioselectivity for (*S*)-γ alcohol. This revealed one of the limitations of the docking-guided mutagenesis approach because it relied on the presence of specific binding pockets for the "unwanted" poses. Such pockets were

easier to be identified when the variant to be studied was not already highly selective or active. Nonetheless, tests of the mutations proposed based on the analysis of NP poses *vs*. (*S*)-γ poses would be meaningful.

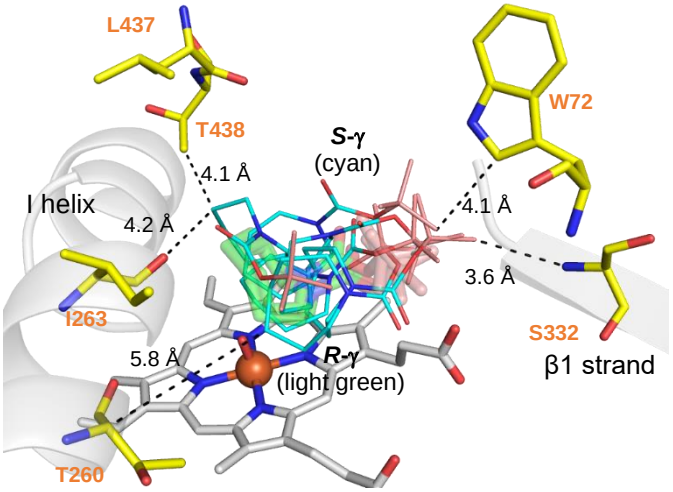


Figure 3.39 Overlay of MD-simulated structures of KU3/A330P/S72W (yellow) with all (*R*)-γ-pose (green) and (*S*)-γ-pose (light cyan). Black dashes with black labels denote distances between residues and poses.

3.4.2.3 Mutagenesis tests

The S332F/M and T438L/F/M mutations were tested for their activity-enhancing effects (Fig. 3.40). The S332 mutations slightly improved TTN to 9950 and 9100, respectively by the S332F and S332M variants, from 8440 by the precursor, while significantly reducing both regio- and enantio-selectivity to 75–78% of **1a** with 70–73% ee for (*S*)-**1a**, from 90% of **1a** with 84% *ee* by the precursor. On the other hand, the T438 mutations maintained the product selectivity and increased TTN to ~12000, with T438F being the most active (TTN = 12400). These observations were in agreement with the hypothesis that the bulky mutations at S332 might destabilise both NP and (*S*)-γ poses and that T438 mutations could disfavour the NP poses without affecting (*S*)-γ poses.

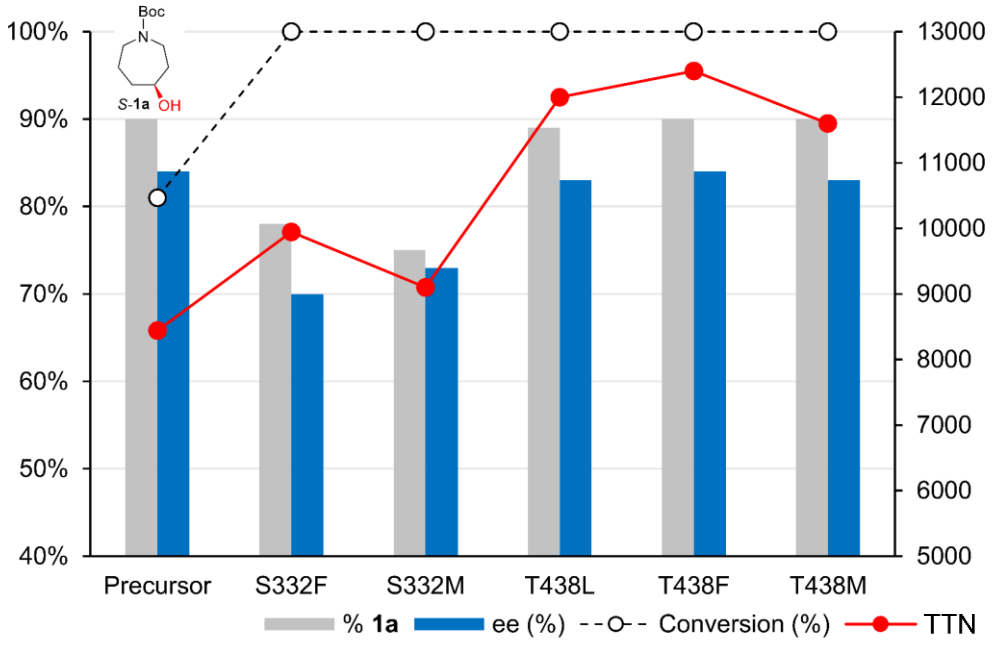


Figure 3.40 The effect of the S332F/M and T438L/F/M mutations. Precursor: KU3/A330P/S72W.

3.4.3 Summary for the docking analysis of KU3/A330P/S72W

Vina rigid-body docking of **1** with KU3/A330P/S72W showed few productive poses. The ADFR approach was adopted to perform flexible-residue docking, as it out-performed the Vina flexible-residue method in cross-docking tests.[221] Control docking experiments showed that four to eight flexible residues would be enough to generate meaningful output without long running time (< 300 min). Moreover, with no flexible residues (*i.e.*, rigid body), ADFR docking gave similar output as the Vina approach. Analysis of the post-ADFR conformations of flexible residues suggested that the positioning of F87 was key to the adoption of productive poses and having F87 as a flexible residue could increase the likelihood of "correct" poses. It was therefore concluded that for three- or four-replica docking analysis of $P450_{BM3}$ variants with F87 or bulky mutations at F87 present, the ADFR approach must be applied. For variants with small substitutions at F87, the rigid-body Vina approach could be utilised. ADFR docking with four flexible residues indicated (*S*)-γ alcohol as the major product, in agreement with the

experimental data. The analysis of (*S*)-γ poses *vs*. NP poses led to designs of S332 and T438 mutations which were found to enhance enzymatic activity; the KU3/A330P/S72W/T438F variant was the most active (TTN = 12400), increasing the TTN by ~50% from 8440 while maintaining the product selectivity (90% of **1a** with 84% *ee*). On the other hand, it was difficult to design enantioselectivity-enhancing mutations as the (*R*)-γ and (*S*)-γ poses substantially overlapped and no specific binding regions for (*R*)-γ poses was identified. This indicated the docking-guided mutagenesis approach relied on significant differences between the correct poses and poses for other products, which was difficult to identify if the variant was already highly active (>85% regioselectivity and *ee*). One approach would be to deploy site-saturation mutagenesis at S332 and T438 and present the substrate with varied steric demand from the enzyme and alter the (*S*)- and (*R*)-poses to achieve increased enantioselectivity. However, expanding the scope of the docking analysis studies by employing both Vina and ADFR approaches on other substrates could be meaningful.

3.5 Docking guided mutagenesis for azocane

3.5.1 Docking analysis for the GV/A184I/I263G/A328G variant

3.5.1.1 General docking outputs

Docking analysis was performed on the azocane substrate to expand the scope of research and seek improvements in product selectivity and enzymatic activity for azocane oxidation. From the screening and initial engineering, the highest selectivity for the two enantiomers of the β alcohol **7c** were given by similar or same variants as for azepane. The (*S*)-β alcohol was most favoured by the GV/A184I/I263G/A328G variant, giving 76% of **7c**, 77% *ee* for *S* (TTN = 180). The same variant was most (*S*)-β selective for *N*-Ips-azepane **5** oxidation (Table 3.13). The highest selectivity for the (*R*)-β alcohol was given by the R19/S72A/A330W variant (85% of **7c**, 99% *ee* for *R*, TTN = 400), which had similar mutations as the most (*R*)-β selective variant for azepane, VQ/S72G/A330W.

Table 3.13 Best variants and selectivity for two enantiomers of the β-alcohols **5c** and **7c** from initial screen.

Products	Variant	Regioselectivity	*ee*	TTN
(*S*)-β-ol, (*S*)-**7c**	GV/A184I/I263G/A328G	76%	77%, *S*	180
(*R*)-β-ol, (*R*)-**7c**	R19/S72A/A330W	85%	99%, *R*	400
(*S*)-β-ol, (*S*)-**5c**	GV/A184I/I263G/A328G	76%	79%, *S*	290
(*R*)-β-ol, (*R*)-**5c**	VQ/S72G/A330W	91%	96%, *R*	1070

The GV/A184I/I263G/A328G variant was chosen for docking analysis for *N*-Ips-azocane **7**, as it was the same variant that was docked with *N*-Ips-azepane **2** and a direct comparison could be drawn between the two. Vina rigid-body docking was performed on substrate **7** with the seven clustered structures of GV/A184I/I263G/A328G. The (*S*)-β poses were most populated among the productive poses. There were eight poses indicating (*S*)-β oxidation, three for (*R*)-β, three for α, six for γ and six for δ, together with 37 NP poses (Fig. 3.41).

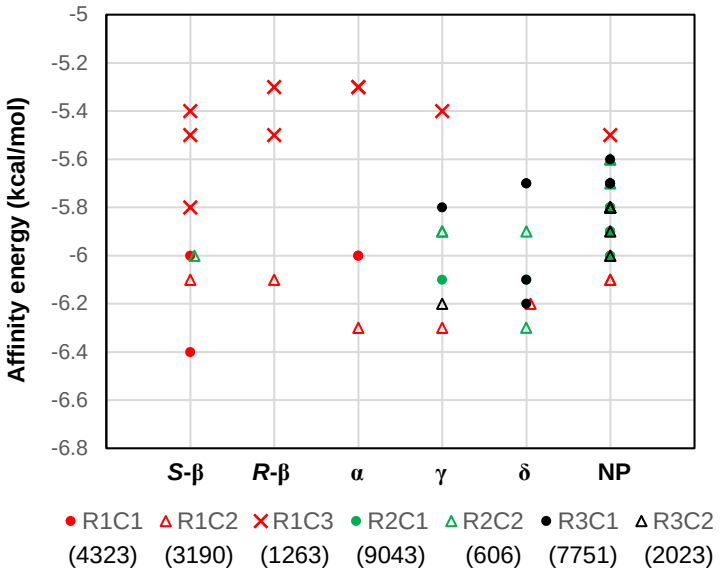


Figure 3.41 The affinity energy of output poses from Vina docking of *N*-Ips-azocane, **7**, with MD-simulated structures of the variant GV/A184I/I263G/A328G. RxCy denotes cluster y of replica x. The numbers in brackets indicate the population of each cluster.

Similar to the observation in docking with *N*-Ips-azepane, poses from replica 1 dominated the productive and especially (*S*)-β poses. Overlay of the R1C1, R2C1 and R3C1 structures with the lowest energy (*S*)-β pose indicated that the T268 side chain of R1C1 formed a hydrogen bond with the sulfonyl oxygen while the T268 side chains in the other two structures would clash with this (*S*)-β pose (Fig. 3.42). In R2C1 the T268 side chain formed a hydrogen bond with H266. Other residues, such as S72, L75, I263, G328, A330 and L437, showed rotamer conformations or backbone shifts but none of these would affect the (*S*)-β pose. These comparisons again demonstrated the importance of using multiple replicas for MD simulations and substrate docking.

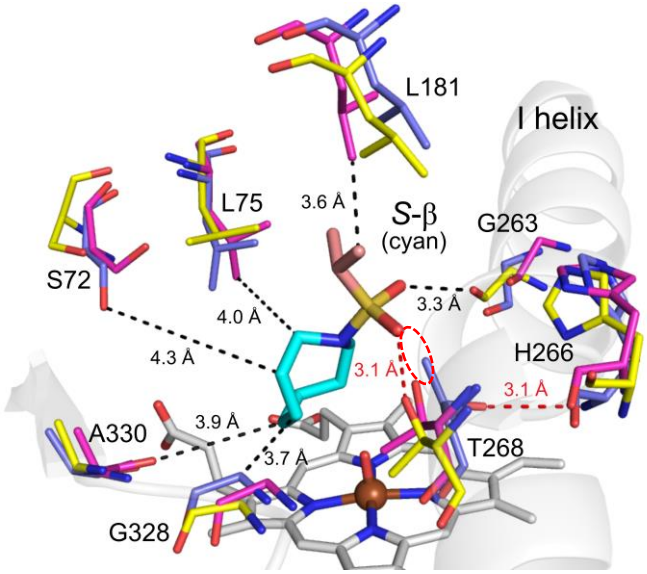


Figure 3.42 Overlay of selected residues from R1C1 (yellow), R2C1 (magenta) and R3C1 (blue) structures of GV/A184I/I263G/A328G, together with the lowest energy (S)-β pose. Black dashes with labels denote distances. Red dashes denote H-bond. Red dashed circles denote steric clashes.

3.5.1.2 Binding pose analysis and mutagenesis

The (*S*)-β poses aligned well. The Ips group was close to the I helix and contacted V87, G263, T268, while the azocane ring pointed towards the β1 strand and was in van der Waals contact with L75, V87, T268, G328 and A330. Hydrogen bonding between T268 and the sulfonyl oxygen was observed (Fig. 3.43-A). These poses were then overlaid with the lowest energy (*S*)-β-pose from docking of *N*-Ips-azepane **2** with the same simulated clusters of GV/A184I/I263G/A328G (Fig. 3.43-B). The docked azocanes mostly aligned well with the azepane pose. However, some of the Ips-azocane (*S*)-β poses were quite close to L75 and V87, compared with their azepane counterparts, which could affect mutation design, as bulky mutations at these two residues could potentially affect the ee for (*S*)-**7c** as well.

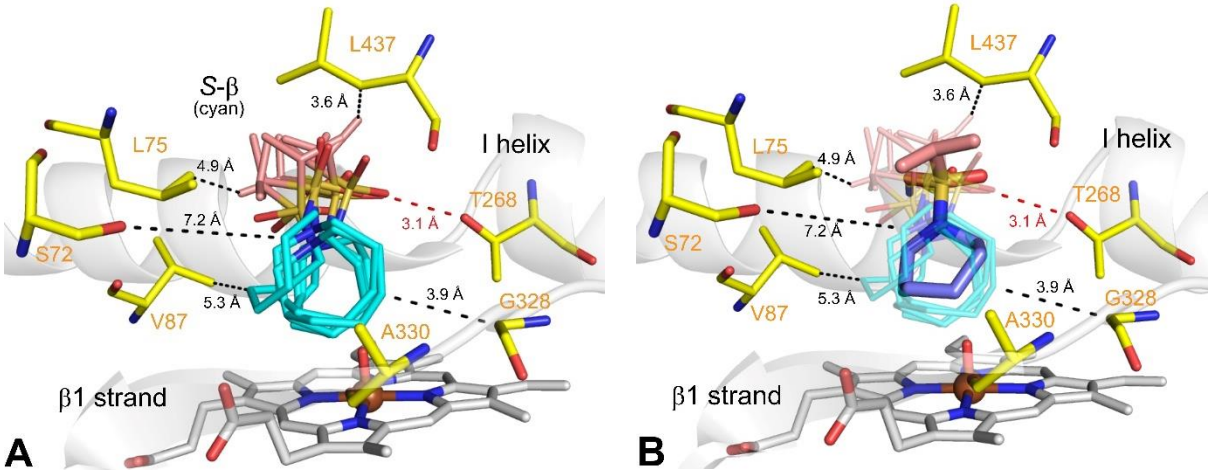


Figure 3.43 Overlay of **(A)** all (*S*)-β-poses from docking with *N*-Ips-azocane **7** (cyan) and **(B)** the azocane (S)-β poses (light cyan) with the lowest energy azepane (*S*)-β pose (blue), together with selected residues of the simulated GV/A184I/I263G/A328G structure.

Binding poses for different products were analysed to design mutations. Overlaying of the lowest energy (*S*)-β pose with (*R*)-β poses revealed specific binding pockets for the (*R*)-configured poses where the (*S*) poses were not located (Fig. 3.44). Mutations S72F/M, L75F/M and V87L/I could disfavour (*R*) poses and improve the ee for the (*S*)-β alcohol (*S*)-**7c**.

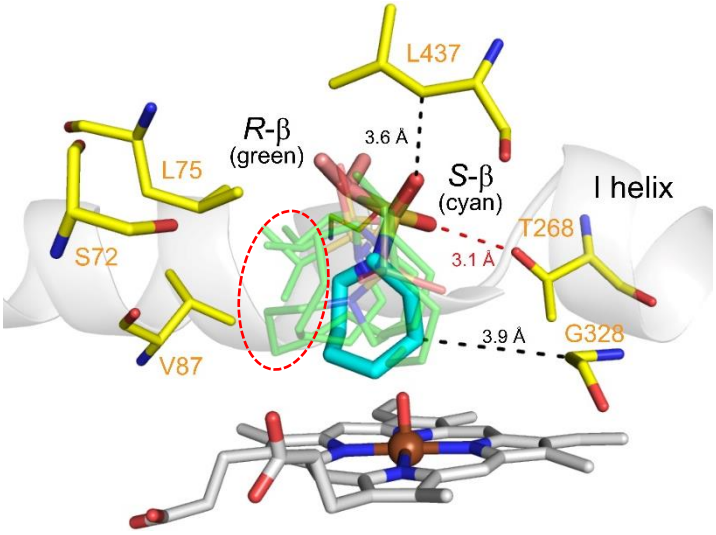


Figure 3.44 Overlay of lowest energy (*S*)-β pose (cyan) with the (*R*)-β poses (light green), together with selected residues of the MD-simulation structure of GV/A184I/I263G/A328G (yellow).

The L75F/M variants maintained the regioselectivity for **7c** (74% and 75%), but significantly decreased the ee (50% and 61%) (Fig. 3.45). Both variants also showed almost 2-fold improvements of the TTN, to 330 and 320. The F87I/L variants showed improved regioselectivity (87% and 90%) and TTN (420 and 440), but again lowered ee to ~55%. The S72F/M variants enhanced the ee to 88% and 90% while maintaining the regioselectivity (72% and 74%), with improved TTN at 410 and 420. The observations that L75 and F87 mutations decreased ee supported the hypothesis that they could interfere with the (*S*)-β poses. The S72F variant was selected as the template to conduct further mutagenesis to seek improvement in product selectivity and TTN.

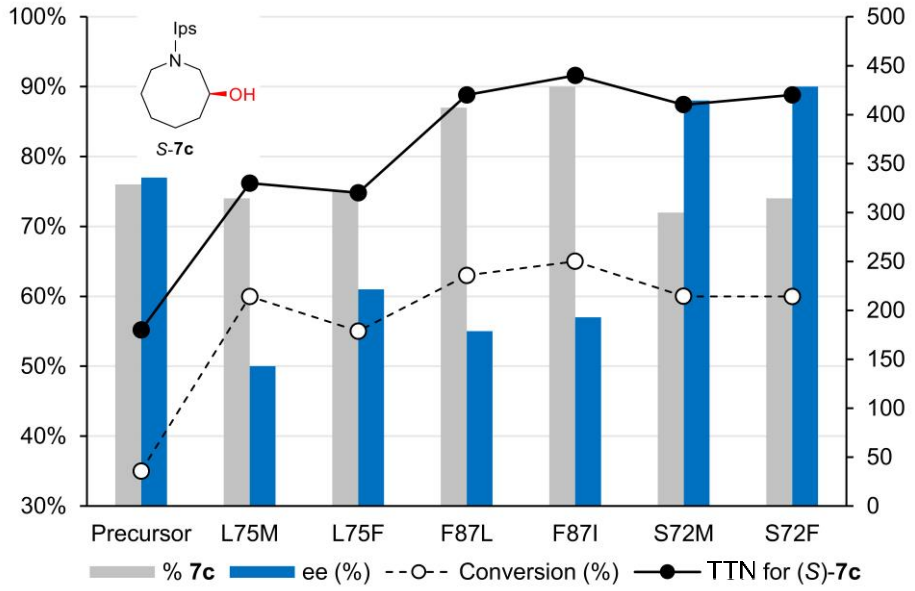


Figure 3.45 The effect of the L75M/F, F87L/I and S72M/F mutations. Precursor: GV/A184I/I263G/A328G.

3.5.1.3 Docking analysis of the GV/A184I/I263G/A328G/S72F variant

Three replica MD simulations were performed for the GV/A184I/I263G/A328G/S72F variant and substrate docking was conducted for **7** with the simulated structures. The output indicated the (*S*)-β alcohol as the major product and the (*S*)-β poses aligned well. From the overlay of the lowest energy (*S*)-β pose and NP poses, specific binding regions for NP poses were identified

where the (*S*)-β ones were not located. This "NP region" was close to residues F72, L188 and S332 (Fig. 3.46). Mutations L188F/M and S332F/M were thus designed as they could disfavour the NP poses without affecting the (*S*)-β poses.

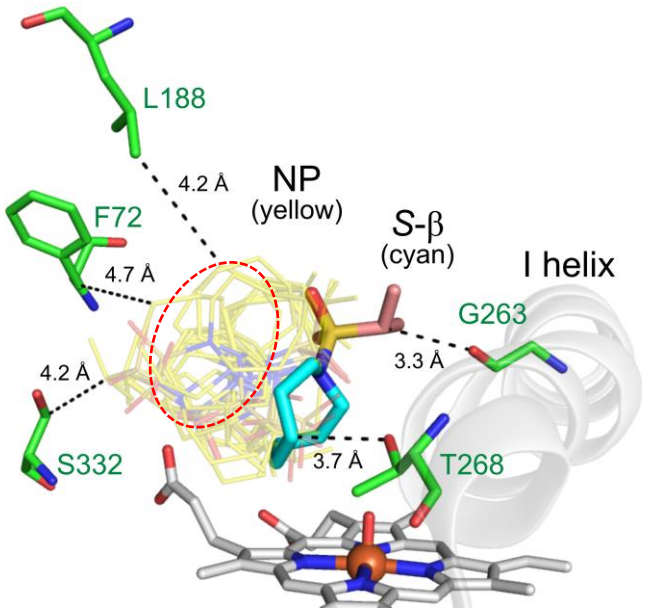


Figure 3.46 Overlay of lowest energy (*S*)-β pose (cyan) with the NP poses (light yellow), together with selected residues of the simulated GV/A184I/I263G/A328G/S72F structure (green). Red dashed circle denotes the binding pocket for NP poses.

Experimentally, the S332F/M variants showed significantly improved TTN of 900 and 930, while maintaining the regioselectivity (71%) and slightly increasing ee (93% and 95%) (Fig. 3.47). The L188F/M variants improved both the TTN (960 and 1000) and regioselectivity (81% and 80%), while maintaining the ee (90% and 91%). Among these, the GV/A184I/I263G/A328G/S72F/L188F variant showed most significant improvements compared with the precursor; the TTN was increased to 1000, five-fold of 180 by the precursor, the ee was enhanced from 77% to over 90%, and the regioselectivity was slightly higher (80%) than the precursor (76%).

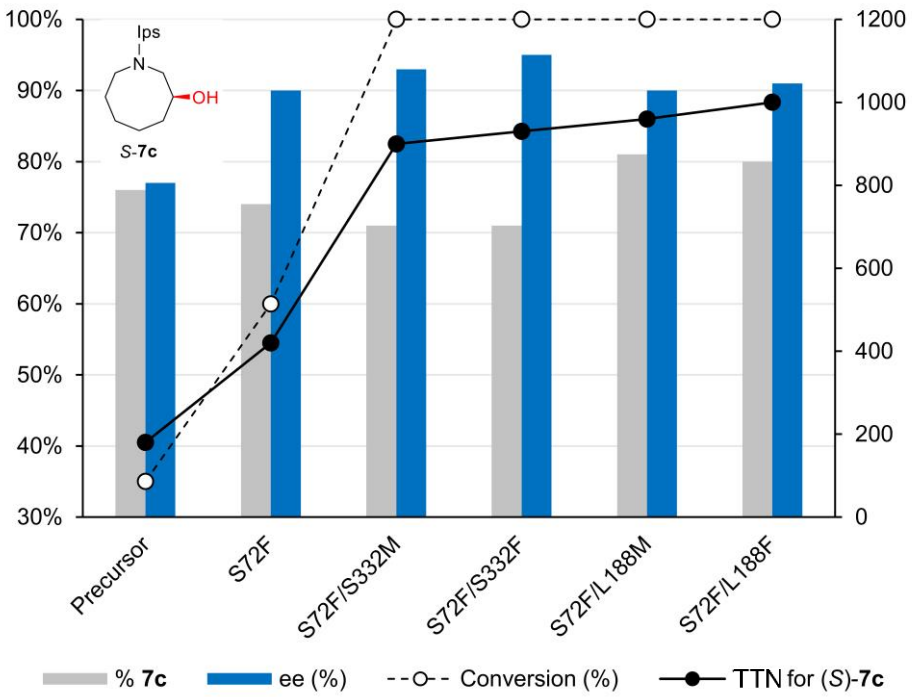


Figure 3.47 The effect of the S332F/M and L188F/M mutations. Precursor: GV/A184I/I263G/A328G.

To summarise, the docking of *N*-Ips-azocane **7** with GV/A184I/I263G/A328G suggested the (*S*)-β alcohol as the major product, in agreement with experimental data. The overlaid (*S*)-β poses showed close resemblance with the (*S*)-β poses from docking of *N*-Ips-azepane **2** with this variant. The overlay of (*S*)-β poses with poses for other products helped the design of new variants that showed improved product selectivity and enzymatic activity. Docking analysis with the best variant from the first round of mutagenesis, GV/A184I/I263G/A328G/S72F, led to the GV/A184I/I263G/A328G/S72F/L188F variant with improved TTN (1000 *vs*. 180), ee (91% *vs*. 77%), and regioselectivity (80% *vs*. 76%), compared to the GV/A184I/I263G/A328G identified from the original screening.

3.5.2 Docking analysis for the R19/S72A/A330W variant

Three replica simulations of the R19/S72A/A330W variant generated 9 clustered structures (Figure A5.7, Appendix **A5.2.2**). Due to the retention of F87 in this variant, the ADFR approach was deployed with four flexible residues (F87, I263, W330 and L437). The results indicated the (*R*)-β alcohol as the major product, in agreement with the experimental data, though the predicted (*R*)/(*S*) ratio of 11 was much lower than 199 experimentally (Figure A5.8, Appendix **A5.2.2**). The (*R*)-β poses aligned reasonably well, with the azepane ring pointed towards the I helix and the Ips group contacted A72 and W330 in most cases (Fig. 3.48). There was one (*R*)-β pose adapting the opposite binding direction, but it was within the binding region of other (*R*)-β poses.

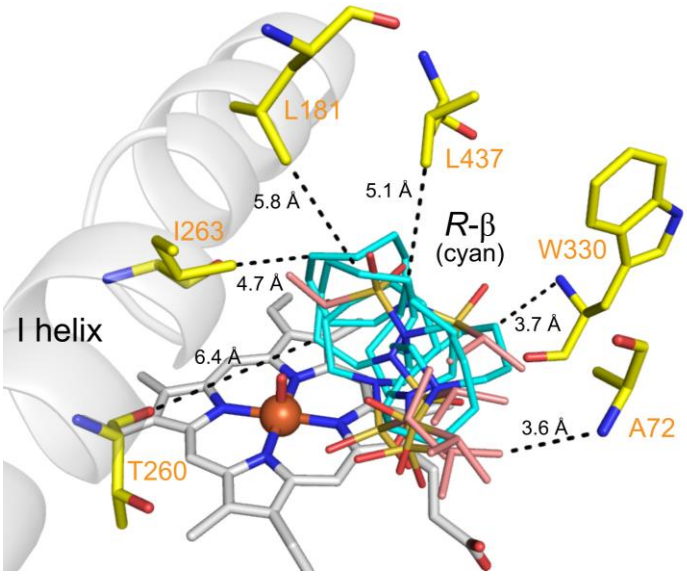


Figure 3.48 Overlay of all (*i*)-β poses with selected residues of the simulated R19/S72A/A330W structure (yellow). Black dashes with labels denote distances between residues and poses.

Since this variant showed high regioselectivity (85%) and striking enantioselectivity (99% ee) but a low TTN (340), mutations to improve enzymatic activity were of interest. The overlay of NP poses with the lowest energy (*R*)-β pose revealed a binding region for NP poses close to L181, I263 and L437 (Fig. 3.49). The mutations L181F/M and L437F/M were introduced.

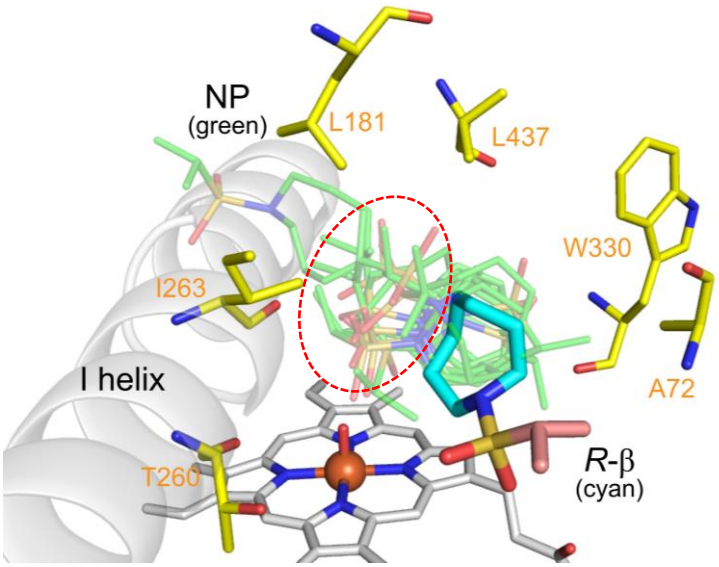


Figure 3.49 Overlay of the lowest energy (*R*)-β pose (cyan) with NP poses (green), together with selected residues of the simulated R19/S72A/A330W structure (yellow). Red dashed circle denotes the binding pocket for NP poses. Mutations L181F/M and L437F/M could disfavour the NP poses without affecting the (*R*)-β poses and thus improve TTN.

Experimentally, all the L181 and L437 mutations tested maintained the product selectivity (~85% of **7c**, 99% ee). The L437F/M variants improved TTN to >1000 while the L181F/M showed >2000 TTN. The R19/S72A/A330W/L181F variant was the most active, with a TTN of 2700, an eight-fold improvement over the precursor (Fig. 3.50).

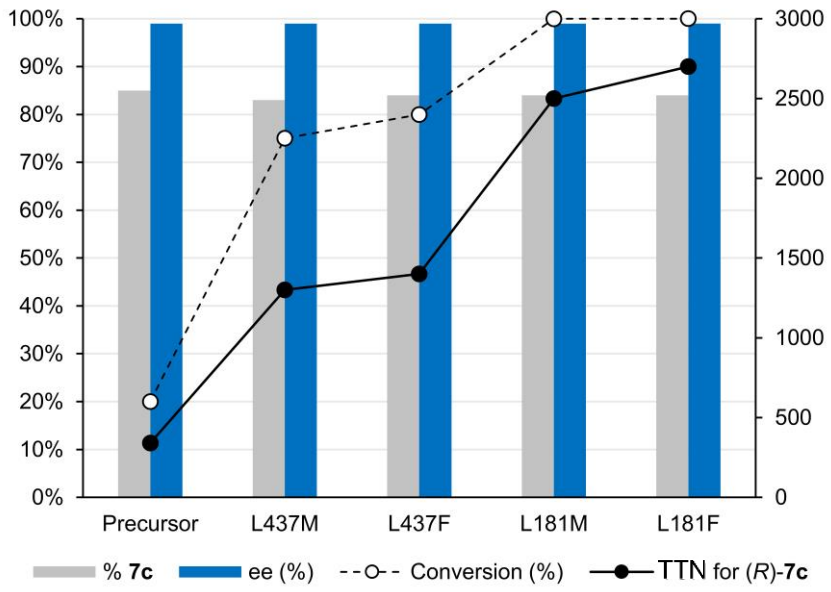


Figure 3.50 The effect of the L437F/M and L181F/M mutations. Precursor: R19/S72A/A330W.

3.6 Summary of docking-guided mutagenesis for azepane and azocane

MD simulations and substrate docking have provided insight into substrate–enzyme interactions and led to the design of new variants with significantly improved product selectivity and enzymatic activity. Two docking approaches, Autodock Vina and ADFR, were explored and compared for their performance with the cyclic amine substrates. It was concluded that, for variants with bulky residues in the $P450_{BM3}$ active site, e.g., F87 and L328, the ADFR approach was required, but otherwise, the Vina rigid-body docking could be utilised.

Although the absolute affinity energy of docking poses generated in vacuum are unlikely to be reliable, the predicted selectivity, especially when weights were assigned for each pose using cluster population and affinity energy as specified in **3.2.1.2**, was usually in agreement with the major product observed experimentally. Analysis of substrate binding interactions and overlay of poses for different products aided the design of effective mutations. Combinations of these mutations, together with iterative docking analysis on the best variants for each generation, resulted in significantly improved selectivity and TTN.

For the (*S*)-β alcohol of azepane, three rounds of guided mutagenesis improved (*S*)-β selectivity and TTN, from 76% of **5c** and 79% ee (TTN = 290) by the GV/A184I/I263G/A328G variant, to 97% of **5c** and 92% ee (TTN = 1490) by GL/A184I/I263G/A328G/S72F/L188M (Fig. 3.51). Starting from the same precursor, and by two rounds of mutagenesis, selectivity and TTN for the azocane (*S*)-β alcohol (*S*)-**7c** was improved to 80% of **7c** and 91% ee (TTN = 1000) by the GV/A184I/I263G/A328G/S72F/L188F variant, from 76% of **7c** and 77% ee (TTN = 180).

This approach was especially efficient in improving enzymatic activities. For the (*R*)-γ alcohol (*R*)-**1a** of azepane, guided mutagenesis led to the RLYF/H171L/I263G/A330W/L437S/V78F variant (94% of **1a**, 93% ee, TTN = 14500) in just one round, while the template,

RLYF/H171L/I263G/A330W/L437S, showed 90% of **1a** with 90% ee (TTN = 10300). For the other enantiomer, (*S*)-**1a**, the KU3/A330P/S72W/T438F variant (90% of **1a**, 84% ee, TTN = 12630) was also discovered by one round of guided mutagenesis, starting from the template KU3/A330P/S72W (90% of **1a**, 84% ee, TTN = 8440).

On the limitation aspects, the docking-guided mutagenesis approach relied heavily on binding orientations of poses being sufficiently different. In cases where there was too much overlap between the target productive pose(s) and others, little could be done to design mutations to improve selectivity or TTN. To tackle this, further efforts are needed to improve the basic MD simulation and substrate docking methodology; alternatively, site-saturation mutagenesis at residues that contact the cyclic amine rings can be conducted. Nevertheless, the application of docking-guided mutagenesis on more complex cyclic amine systems, such as bicyclic amines, would be of great interest.

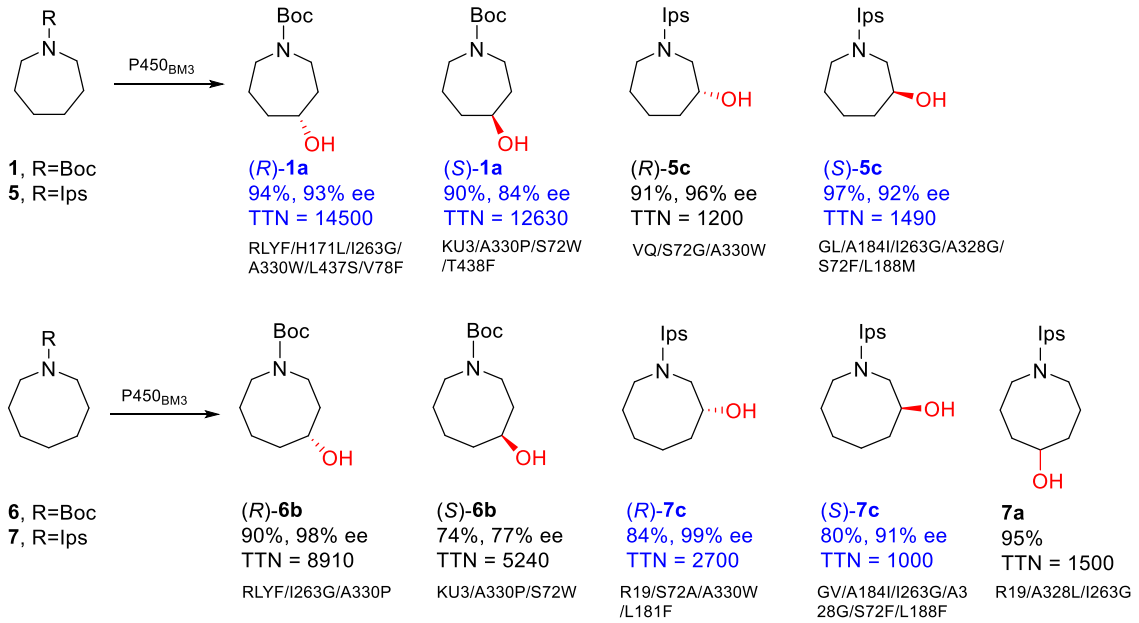


Figure 3.51 Highest selectivity and TTN achieved for all remote products of azepane and azocane and relevant variants by library screening and guided mutagenesis. Marked in blue are the ones that have been improved through docking guided mutagenesis.

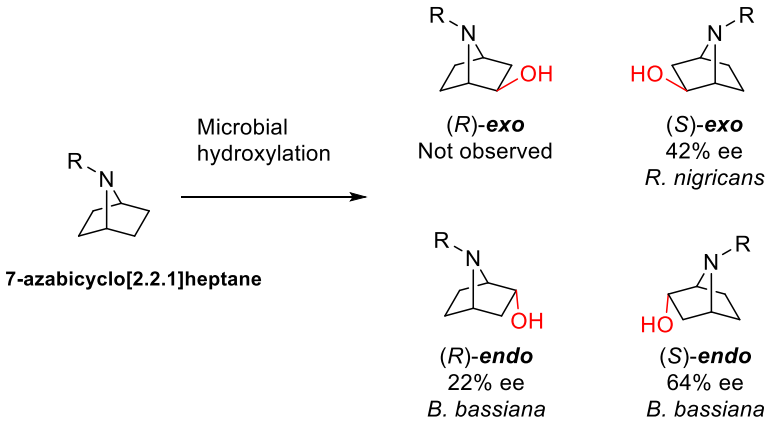
Chapter 4

4. Bicyclic and spirocyclic amines

4.1 Introduction

Recently, there has been growing interest in introducing more 3-dimensionality and saturation into building blocks in pharmaceutical and agrochemical industries.[243,244] This chapter explores the selective functionalisation of two cyclic amines with more 3D structures–the bicyclic amine 7-azabicyclo[2.2.1]heptane and the spirocyclic amine 8-azaspiro[4.5]decane. The combined strategies of library screening, substrate engineering and docking-guided mutagenesis were deployed for high regio- and enantioselectivity for the remote oxidation products.

Biocatalytic hydroxylation of 7-azabicyclo[2.2.1]heptane has been reported by Hemenway,[245] Olivo,[246] and Marshall[247] using *Beauveria bassiana* and *Rhizopus nigricans*. However, low to moderate enantioselectivities were observed; the highest enantioselectivity for the (*S*)-*endo* alcohol (64% ee) was produced by *B. bassiana*, and the same microorganism also showed highest ee for the (*R*)-*endo* product at 22%. For the *exo* alcohol product, highest ee for the (*S*)-enantiomer was 42% found with *R. nigricans*, while no (*R*)-*exo* alcohol was observed (Scheme 4.1).



Scheme 4.1 Highest enantioselectivity observed for the hydroxylation products from the work by Hemenway,[245] Olivo,[246] and Marshall.[247]

On the other hand, the direct functionalisation of 8-azaspiro[4.5]decane has not been reported. Previous work within the Wong and Robertson groups have investigated the $P450_{BM3}$ oxidation of *N*-Boc-8-azaspiro[4.5]decane, which is summarised in Section **4.4.1**. High regioselectivity was achieved for remote hydroxylation products, but the stereochemistry remained undetermined. This chapter also explores advanced Mosher ester analysis and X-ray crystallographic study to assign the absolute configuration of these products.

4.2 Initial screen and analysis of the bicyclic compound

4.2.1 Initial screen for *N*-Boc-7-aza-bicyclo[2.2.1]-heptane

N-Boc-7-aza-bicyclo[2.2.1]-heptane, **8**, was synthesised and screened against the 48 variants in the refined library (Table A2.2, Appendix **A2.1**). Of these, 29 variants showed >50% conversion. The three products were purified and characterised as the *exo*-β alcohol **8a**, the *endo*-β alcohol **8b**, and the β ketone **8c** (Fig. 4.1). Their NMR spectra were in agreement with previously reported data[248] and information can be found in Appendices **A6** and **A7**.

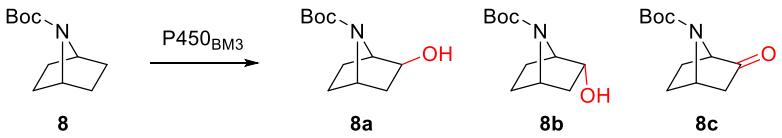


Figure 4.1 Products characterised from the initial screening of *N*-Boc-7-aza-bicyclo[2.2.1]-heptane, **8**.

The absolute configuration for enantiomers was determined by comparison with reported specific rotation data (Table 4.1).[248] The *exo*-alcohol **8a** isolated from a reaction with the R19/F87I variant was determined as the (*R*)-enantiomer, with 99% ee ($[\alpha]_D$ = +20.8), while the R19/F87A/F81W variant afforded (*S*)-**8a**, also with 99% ee ($[\alpha]_D$ = –19.9). For the *endo* alcohol **8b**, the R19/S72G/A330W variant gave the (*R*)-enantiomer with 80% ee ($[\alpha]_D$ = –4.1). The (*S*)-

configuration of product **8b** from GQ/I263G/A328L was assigned by comparison of its $[\alpha]_D$ (+2.3) with the reported data ($[\alpha]_D = -3.8$) for the (*R*)-enantiomer. The stereochemistry and ee of alcohol products given by other variants were determined by chiral-phase GC analysis and comparison with data for these four compounds.

Table 4.1 $[\alpha]_D$ of alcohol products of substrate **8** and the literature data.

Products	Variants	*ee*	$[\alpha]_D^{25}$	Solvent	Literature data
8a *exo*-β-alcohol	R19/F87I	99%, ***R***	+20.8	CH_2Cl_2	+21.1 for ***R*** (+)[248] –21.6 for ***S*** (–)[248]
8a *exo*-β-alcohol	R19/F87A/F81W	99%, ***S***	–19.9	CH_2Cl_2	+21.1 for ***R*** (+)[248] –21.6 for ***S*** (–)[248]
8b *endo*-β-alcohol	R19/S72G/A330W	80%, ***R***	–4.1	$CHCl_3$	–3.8 for ***R*** (–)[249]
8b *endo*-β-alcohol	GQ/I263G/A328L	44%, ***S***	+2.3	$CHCl_3$	–3.8 for ***R*** (–)[249]

4.2.2 Selectivity trends for *exo*-β alcohol 8a

From the initial screening, the *exo*-alcohol **8a** was the dominant product, with 23 variants showing >50% selectivity for **8a**, while six variants gave the *endo*-alcohol **8b** as the major product. The (*R*)-enantiomer was the predominant for *exo*-selective variants (>50% **8a**), with 16 out of these 23 showing >80% ee for (*R*)-**8a**, while three gave >80% ee for the (*S*)-isomer. The highest (*R*)-selectivity was observed with the R19/F87I variant, with 99% of **8a** and 99% ee (TTN = 1970), while the (*S*)-enantiomer was most favoured by the GLQ/I263G/A328G variant, at 99% of **8a** and 86% ee (TTN = 1880) (Entries 1 & 25, Figure 4.2).

Analysis of the selectivity data revealed key mutations for both enantiomers. For the (*R*)-enantiomer, the F87I mutation was sufficient to give striking regio- (99%) and enantio- (99%) selectivity (Entries 1 & 2, Fig. 4.2). The F87V mutation also promoted ee (up to 97%) but decreased regioselectivity (~70%; GVQ & K19/F87V, Entries 9 & 10). The F87V/A184I combination increased regioselectivity (82%) and maintained ee (97%; GV/A184I, Entry 8).

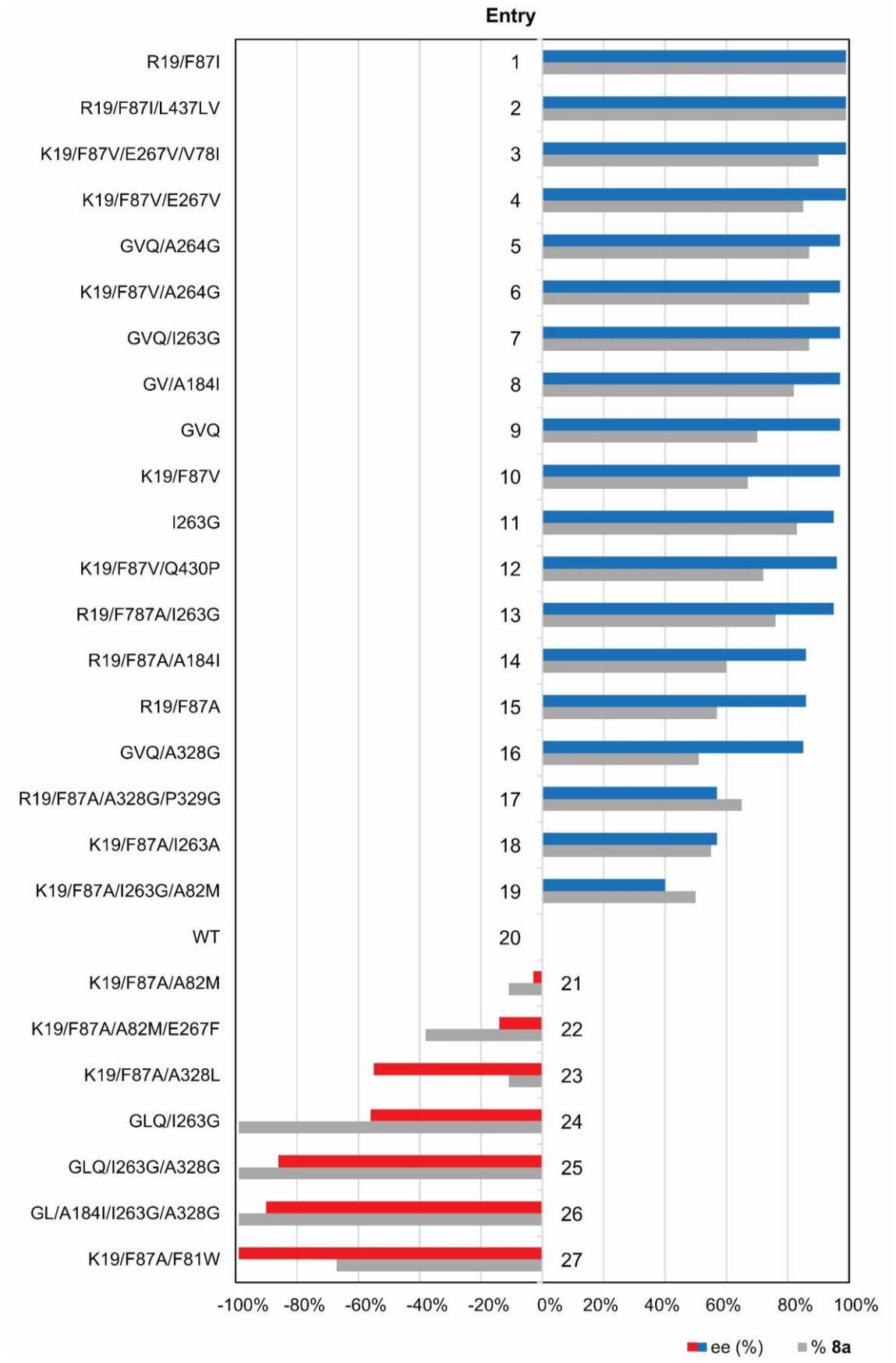


Figure 4.2 Regio- and enantio-selectivity for (*R*)- and (*S*)-**8a** of $P450_{BM3}$ variants from the initial screen. Positive values represent selectivity for (*R*)-**8a**. Negative values represent selectivity for (*S*)-**8a**.

The I263G mutation was also highly (*R*)-promoting, as the I263G variant on its own gave 83% of **8a** and 95% ee (Entry 11), and addition of I263G to the GVQ variant improved its selectivity to 87% of **8a** and 97% ee (Entry 7 & 9). Other mutations, such as E267V (Entry 4), V78I (Entry 3), A264G (Entries 5, 6, 9 & 10), and Q403P (Entries 10 & 12) also favoured the (*R*)-enantiomer.

The F87A mutation lowered the (*R*)-selectivity, with R19/F87A giving 57% of **8a** and 86% ee (Entry 15, Fig. 4.2). The selectivity could be rescued by I263G, as R19/F87A/I263G gave 76% of **8a** and 95% ee (Entry 13), but not I263A, with K19/F87A/I263A giving 55% of **8a** and 57% ee (Entry 18). The A82M mutation significantly decreased (*R*)-selectivity, *e.g.*, addition of this mutation to the R19/F87A/I263G variant lowered its selectivity to 50% of **8a** and 40% ee (Entries 13 & 19).

Strikingly, substitution of F87V with F87L reversed the enantioselectivity, as the GLQ/I263G variant gave 99% of **8a** and 56% ee for (*S*)-**8a**, while its GVQ/I263G counterpart afforded 87% of **8a** and 97% ee for *R* (Entries 7 & 24, Fig. 4.2). The addition of A328G to GLQ/I263G maintained the high regioselectivity and further increased (*S*)-selectivity, to 86% ee by GLQ/I263G/A328G (Entry 25). Considering the importance of mutations F87L, I263G and A328G to (*S*)-*exo*-selectivity, a few variants generated in the docking-guided mutagenesis for azepane were tested against **8** to investigate the effect; these are GL/A814I/I263G/A328G and a few variants derived from it with additional mutations at S72, L75 and A330. Among these, the GL/A184I/I263G/A328G variant was the most effective and further improved (*S*)-*exo*-selectivity, giving 99% of **8a** and 90% ee (Entry 26). The F81W mutation promoted (*S*)-selectivity but lowered regioselectivity, with the K19/F87A/F81W variant giving 67% of **8a** and 99% ee (Entry 27). The addition of F81W mutation to the GL/A184I/I263G/A328G variant could lead to further improved product selectivity for (*S*)-**8a**.

4.2.3 Selectivity trends for *endo* alcohol 8b

For the *endo* alcohol **8b**, the (*R*)-enantiomer was dominant: six out of the seven highly *endo*-selective variants gave high selectivity for (*R*)-**8b**, while one favoured the (*S*)-isomer. Mutations A330P/W and S72G/A were important for (*R*)-selectivity. The KU3/A330P/S72W variant was the most selective at 93% of **8b** and 96% ee for (*R*) (Entry 1, Fig. 4.3). The A330W mutation showed high regioselectivity but lower ee (95% of **8b** and 68% ee; Entry 5, Fig. 4.3). Addition of S72A/G mutations to the R19/A330W variant maintained the regioselectivity (95%) and improved ee to ~80% (Entries 3 & 4, Fig. 4.3).

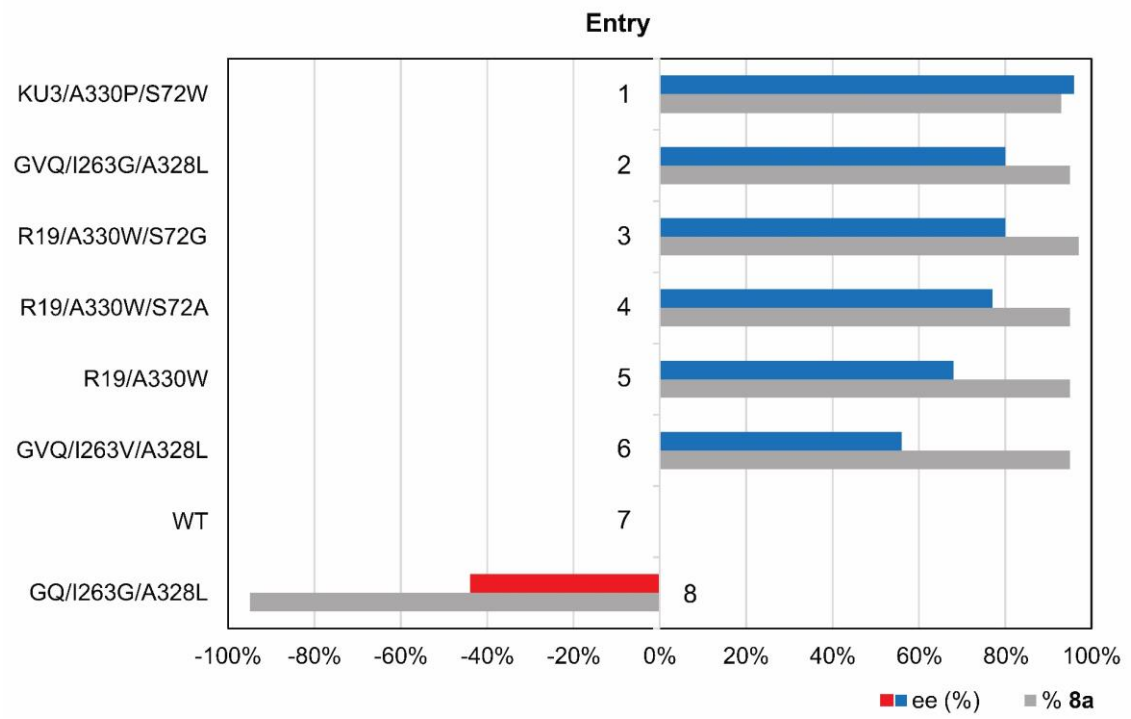


Figure 4.3 Regio- and enantio-selectivity for (*R*)- and (*S*)-**8b** of $P450_{BM3}$ variants from the initial screen. Positive values represent selectivity for (*R*)-**8b**. Negative values represent selectivity for (*S*)-**8b**.

The A328L mutation also favoured (*R*)-**8b**, as addition of this mutation to the (*R*)-*exo*-selective variant GVQ/I263G (87% of **8a** and 97% ee; Entry 7, Fig. 4.2) switched the diastereoselectivity to *endo*- with high ee (GVQ/I263G/A328L; 95% of **8b** and 80% ee; Entry 2, Fig. 4.3). Substitution of I263G with I263V maintained the regioselectivity but decreased the ee (95% of

8b and 56% ee; GVQ/I263G/A328L, Entry 2, Fig. 4.3). The reversion of the F87V mutation in the GVQ/I263G/A328L variant also inverted the enantioselectivity while keeping the regioselectivity and resulted in the only one variant with significant (*S*)-*endo*-selectivity from the initial screen, GQ/I263G/A328L (95% of **8b** and 44% ee for *S*; Entry 8, Fig. 4.3).

4.2.4 Summary of the initial screen for substrate 8

Overall, the panel of 48 variants showed good enzymatic activity and high product selectivity for substrate **8**. The (*R*)-*exo* alcohol (*R*)-**8a** was the dominant product with the highest selectivity at 99% of **8a** and 99% ee. High regio- and enantio- selectivity for the (*S*)-*exo* (99% of **8a** and 90% ee) and (*R*)-*endo* (93% of **8b** and 96% ee) products were also observed. Only the variant GQ/I263G/A328L from the initial screening favoured the (*S*)-*endo* alcohol, but with a low ee (95% of **8b** and 44% ee). Docking-guided mutagenesis was carried out with this variant to improve the (*S*)-*endo*-selectivity.

4.3 Docking-guided mutagenesis for the (*S*)-endo alcohol

4.3.1 Docking analysis with the GQ/I263G/A328L variant

Three replica MD simulations were performed for the GQ/I263G/A328L variant and nine clustered structures, constituting 90% of the total population for each replica, were generated. Docking of substrate **8** with these nine clusters revealed an (*S*)-*endo* preference over the (*R*)-*endo* poses, with a weighted (*S*):(*R*) ratio of 2.0, close to 2.5 from experimental data (Fig. A5.14, Appendix **A5.4**). However, the predicted *exo*-selectivity at 52% was much higher than experimental data at 6%. This suggested substrate docking could not be applied to predict the product regioselectivity of oxidation products for the bicyclic substrates, while the enantio-preference could be reflected correctly.

From the overlay of binding poses, there were two groups of (*S*)-*endo* poses and they were close to each other. The bicyclic ring of both groups pointed towards but did not contact the I helix (Fig. 4.4). For one of these two groups, the Boc groups were close to the β1 strand, while in the other two (*S*)-*endo* poses the protecting group was rotated by ~15° and closer to L75 and S72.

Specific binding regions for the (*R*)-*endo* poses were identified which were close to residues A330 and S332, although there was overlap between the (*R*)-*endo* and (*S*)-*endo* poses (Fig. 4.5). The mutations A330F/M and S332F/M were proposed since, although some (*S*)-*endo* poses might be sacrificed, these mutations could disfavour most of the (*R*)-*endo* poses without affecting the lowest energy (*S*)-*endo* pose, and could improve the ee for *S*.

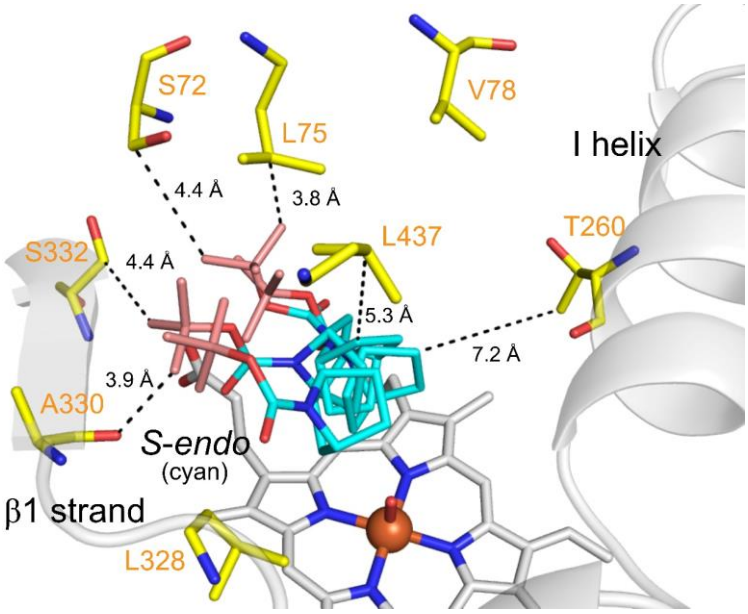


Figure 4.4 Overlay of MD-simulation structures of GQ/I263G/A328L (yellow) with all (*S*)-*endo* poses (cyan). Black dashes with black labels denote distances between residues and poses.

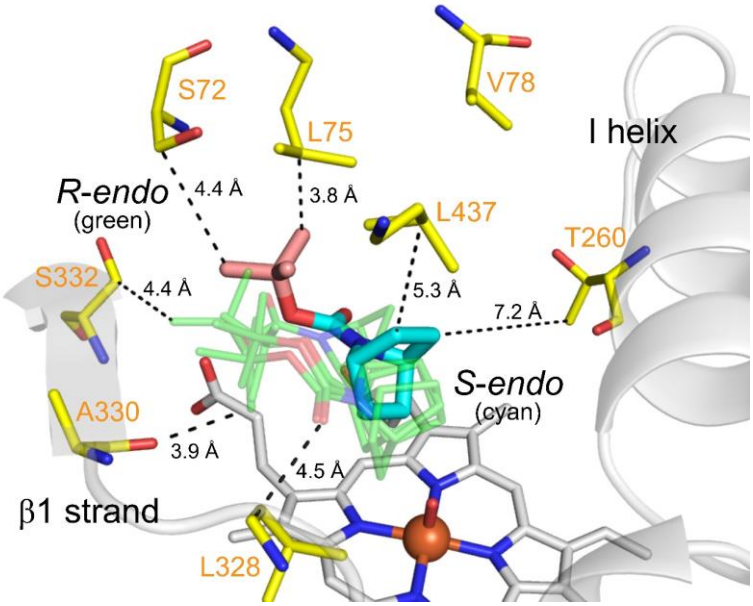


Figure 4.5 Overlay of the lowest energy (*S*)-*endo* pose (cyan) with the (*R*)-*endo* poses (green), together with the simulation structure of GQ/I263G/A328L (yellow). Black dashes with black labels denote distances between residues and poses.

Overlay of the lowest energy (*S*)-*endo* pose with the NP poses revealed significant differences in the binding orientations (Fig. 4.6). The NP poses were sequestered in the pocket formed by residues L75, V78, A82 and T260, while the (*S*)-*endo* pose was not located in this region. Bulky mutations at residues V78, A82 and T260 might improve enzymatic activity.

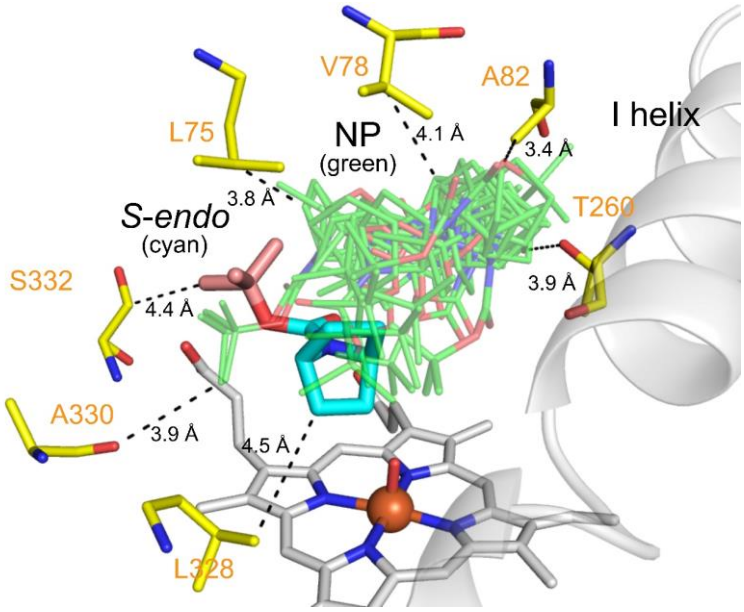


Figure 4.6 Overlay of the lowest energy (*S*)-*endo* pose (cyan) with the NP poses (green), together with the simulated structure of GQ/I263G/A328L (yellow). Black dashes with black labels denote distances between residues and poses.

4.3.2 Mutagenesis tests for the GQ/I263G/A328L docking

On the enantioselectivity aspect, the S332F/M mutations both maintained the high regioselectivity (~97% of **8b**) and slightly improved ee (~52%) and TTN (~360) (Fig. 4.7). The A330L/F variants out-performed their S332 counterparts: they showed higher ee (~63%) and TTN (~440) and maintained the regioselectivity (~95% of **8b**). The A330M variant failed to express after three trials.

The L75F/M variants increased ee (~51%) and TTN (370) slightly but reduced the regioselectivity (~86%) (Fig. 4.8). The V78 and A82 mutations showed similar effects, with the A82M variant gave the highest selectivity (95% of **8b**, 50% ee) and TTN (460) among these. The A82F variant could not be expressed. Surprisingly, the T260 variants greatly improved ee while maintaining regioselectivity and increasing TTN, with the GQ/I263G/A328L/T260L variant being the most effective, giving 76% ee, almost doubled the 44% ee by the precursor GQ/I263G/A328L, with a regioselectivity of 95% and a TTN of 480.

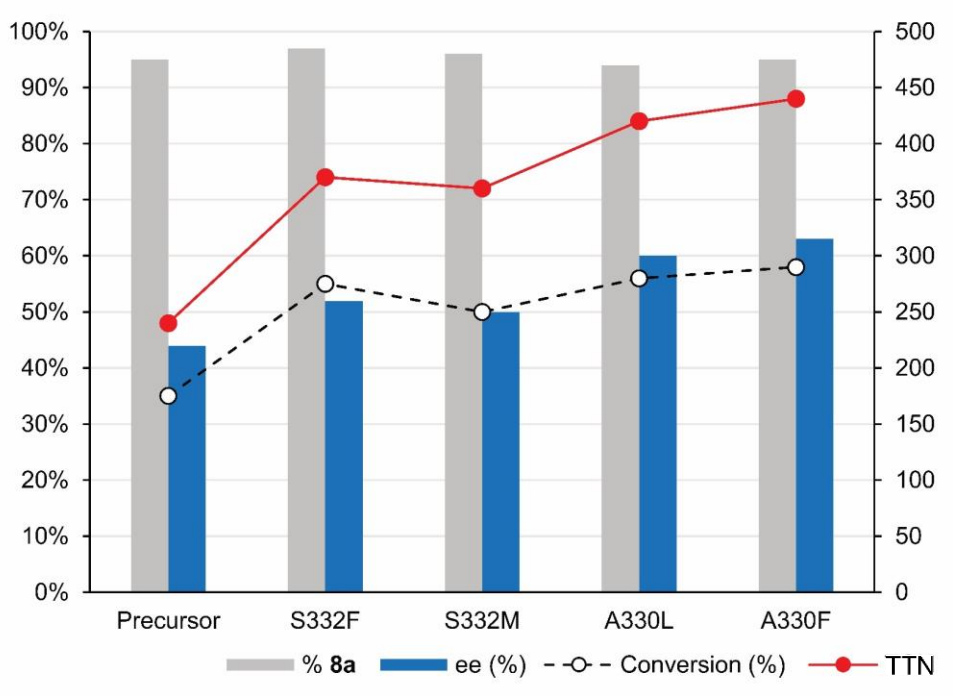


Figure 4.7 The effect of the S332F/M and A330L/F mutations. Precursor: GQ/I263G/A328L.

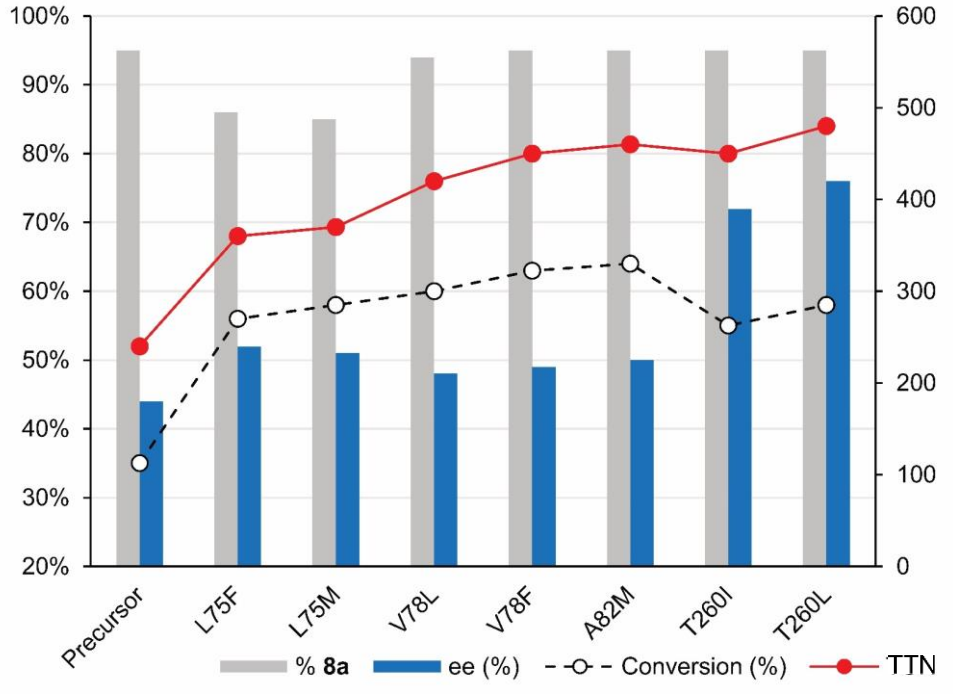


Figure 4.8 The effect of the L75F/M, V78L/F, A82M and T260I/L mutations. Precursor: GQ/I263G/A328L.

The most effective variant GQ/I263G/A328L/T260L was selected as the template for further docking analysis and mutation design.

4.3.3 Further mutagenesis on the GQ/I263G/A328L/T260L variant

Docking analysis with MD-simulation structure of GQ/I263G/A328L/T260L indicated a higher predicted (*S*):(*R*) ratio at 4.5, compared with 2.0 by docking with GQ/I263G/A328L, although there were still many more *exo* poses (42%) generated than experimentally observed (5%). The analysis of binding poses suggested that the effective mutations S332F/M and A330F (for enantioselectivity) and V78L/F, and A82M (for TTN) in the first round of mutagenesis could further improve product selectivity and enzymatic activity when introduced to the GQ/I263G/A328L/T260L variant (Figures A5.15 and A5.16, Appendix **A5.4**).

Experimentally, the S332F/M mutations further improved ee to ~86% while maintaining regioselectivity (~96%) and TTN (~540) (Fig. 4.9). The A330F mutation out-performed its S332 analogues with the GQ/I263G/A328L/T260L/A330F variant giving 98% of **8b** with 89% ee and a TTN of 920, the highest selectivity and enzymatic activity observed for *S*-**8b**.

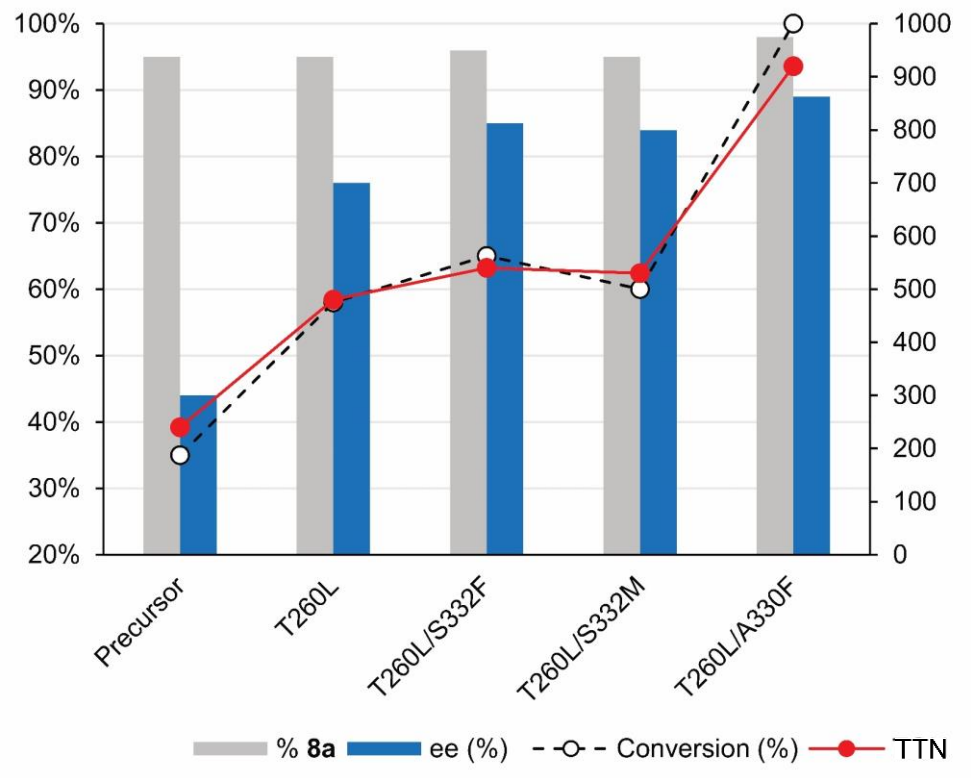


Figure 4.9 The effect of the S332F/M and A330F mutations. Precursor: GQ/I263G/A328L.

For the activity-enhancing designs, the T260L/V87L and T260L/V87F variants showed an improved TTN of ~950, twice the activity of the T260L variant, while maintaining both regioselectivity (~95%) and ee (~78%) (Fig. 4.10). The T260/V78 combinations gave much higher TTN than single mutation at both sites. The T260L/A82M variant showed similar effects as its V78L/F counterparts with even higher TTN (1100).

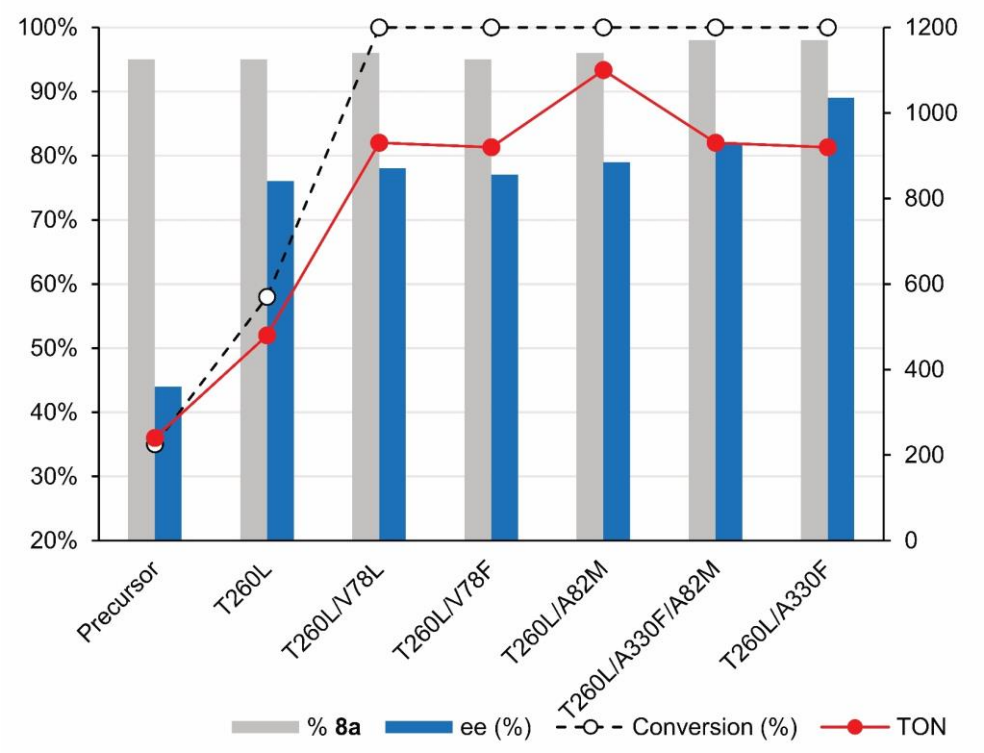


Figure 4.10 The effect of the V78L/F, A82M mutations. Precursor: GQ/I263G/A328L.

The T260L/A330F/A82M variant was then generated the combine these effective mutations. However, this variant maintained the TTN (950) and regioselectivity (96%) but decreased the ee to 82% from 89% by the T260L/A330F variant. Site-saturation mutagenesis at A82 as well as further docking analysis with the GQ/I263G/T260L/A330F variant will be of interest.

4.3.4 Summary of docking-guided mutagenesis for (*S*)-endo alcohol

Docking analysis of the bicyclic compound **8** with simulated structures of GQ/I263G/A328L indicated an *S*:*R* ratio (2.0) close to the experimental data (2.5). However, the weighted *exo*-

selectivity (52%) was over-estimated, indicating further refinement of the fundamental methodologies of MD simulations and substrate docking would be needed to give closer resemblance of experimental product selectivity of the bicyclic amines. Nevertheless, the analysis of binding poses led to mutations that improved selectivity and TTN for (*S*)-**8b**. Among these, the GQ/I263G/A328L/T260L varaint showed surprising effects and greatly improved the ee to 76% from 44% by the template variant GQ/I263G/A328L, while slightly enhancing TTN (480) and maintaining regioselectivity (95%). Docking analysis with this variant led to the GQ/I263G/A328L/T260L/A330F variant which showed the highest (*S*)-*endo*-selectivity at 98% of **8b** with 89% ee (enantiomeric ratio, er = 17, almost seven-fold of the GQ/I263G/A328L template) and a TTN of 920.

Overall, high regio- and enantio-selectivity for all four enantiomers of the remote oxidation product of substrate **8**, the (*R*)-*exo* (99% of **8a** with 99% ee), (*S*)-*exo* (99% of **8a** with 90% ee), (*R*)-*endo* (92% of **8b** with 96% ee) and (*S*)-*endo* (98% of **8b** with 89% ee) alochols, were achieved (Fig. 4.11).

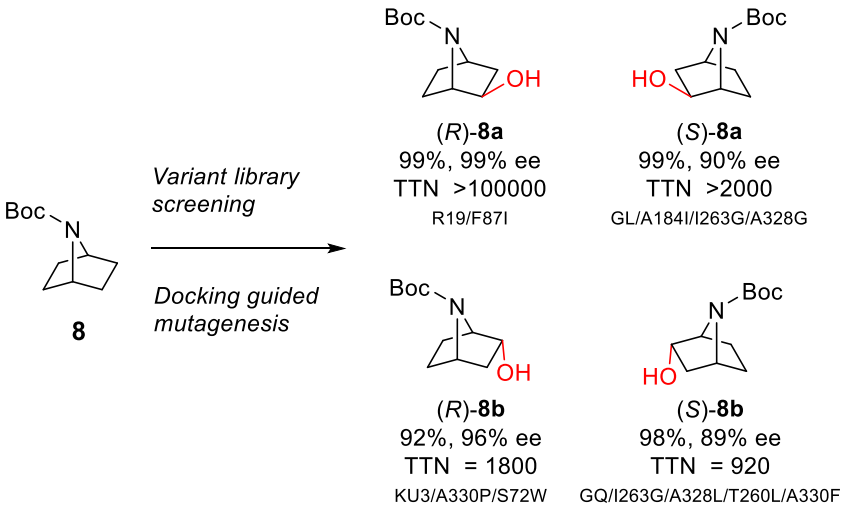


Figure 4.11 The highest selectivity achieved for four enantiomers of the remote oxidation products of substrate **8**.

4.4 The spirocyclic substrates

4.4.1 Initial screen for *N*-Boc-8-azaspiro[4.5]decane

Previously, the oxidation of *N*-Boc-8-azaspiro[4.5]decane **9** by $P450_{BM3}$ variants was studied within the Wong and Robertson groups. Initial screening of substrate **9** against 48 variants of the initial library (Table A2.1, Appendix **A2.1**) revealed high enzymatic activity. There were 34 showing >50% conversion at an enzyme:substrate ratio of 2000:1. Four remote functionalisation products were characterised: the β′ alcohol **9a**, the β′ ketone **9b**, the α′ alcohol **9c** and the α′ ketone **9d** (Fig. 4.12). There was also one unknown minor product (<10% across the screening).

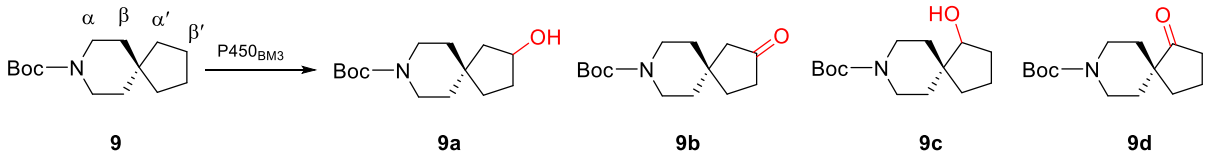


Figure 4.12 Products characterised from the initial screening of *N*-Boc-8-azaspiro[4.5]decane **9**.

High regioselectivity for alcohols **9a** and **9c** were observed. The β′ alcohol **9a** was most favoured by the R19/F87I variant, with 99% regioselectivity at 98% conversion (Entry 1, Table 4.13). Mutations F81W (Entry 2), I263G (Entry 3 & 4) and A328I (Entry 2 & 5) also promoted β′-oxidation. On the other hand, the variant K19/F87A/A82M (Entry 11) showed the highest α′-selectivity, at 99% with 99% conversion. The α′ alcohol **9c** was also favoured by mutations F87A (Entry 8), F87V (Entry 10), and I263A (Entry 9).

The absolute configuration and ee data of products **9a** and **9c** was not determined in previous work. There were no published data for comparison to determine the stereochemistry; chiral phase GC analysis did not give resolution of the different enantiomers.

Table 4.13 Regioselectivity and conversion of selected variants for hydroxylation products of **9** (substrate : enzyme = 2000:1).

Entry	Variants	Conv.	9a	9b	9c	9d	others
1	R19/F87I	98%	**99%**	–	1%	–	–
2	RT2/F81W/A328I	50%	96%	2%	2%	–	–
3	RP/H171L/I263G	97%	70%	–	30%	–	–
4	GVQ/I263G	99%	69%	–	29%	2%	–
5	K19/F87A/I263A/A328I	96%	57%	–	43%	–	–
6	RP	90%	45%	–	55%	–	–
7	RP/F87V	98%	20%	–	70%	–	10%
8	R19/F87A	94%	16%	–	82%	–	2%
9	K19/F87A/I263A	93%	2%	–	97%	1%	–
10	GVQ	15%	2%	1%	97%	–	–
11	K19/F87A/A82M	99%	1%	–	**99%**	–	–

The stereochemical outcomes were studied by advanced Mosher ester analysis and X-ray crystallography in this project. The broad signals in the ^{1}H NMR spectra of **9a** and **9c** due to the rotamer conformation of the Boc group hampered the analysis of the Mosher esters. The absence of a heavy atom in the Boc derivative would also make determination of the absolute configuration by X-ray crystallography analysis less conclusive. As discovered in the substrate engineering work for azepane and azocane, the derivatives with sulfonyl protecting groups gave sharp peaks in the ^{1}H NMR spectra. The presence of sulfonyl atom would also benefit crystallographic analysis. Thus, *N*-Ms-8-azaspiro[4.5]decane (**10**) and *N*-Ips-8-azaspiro[4.5]decane (**11**) were prepared and screened against the 48 variants in the refined library (Table A2.2, Appendices **A2.1**).

4.4.2 Initial screen for *N*-Ms-8-azaspiro[4.5]decane

The Ms-derivative **10** was a good substrate for P450$_{BM3}$ oxidation – 36 out the 48 variants in the refined library showed >50% conversion at a 2000:1 substrate-to-enzyme ratio. Two remote

hydroxylation products, the α′ alcohol **10a** and the β′ alcohol **10b** were characterised (Fig. 4.13). The α alcohol **10c** was also observed.

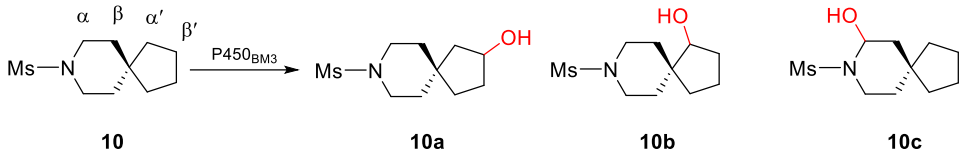


Figure 4.13 Products characterised from the initial screening of *N*-Ms-8-azaspiro[4.5]decane **10**.

The highest selectivity for the β′ alcohol was found with the R19/F87A/A328F variant, which gave 92% of **10a** but with a low conversion of 35% (Entry 1, Table. 4.14). The A328L (Entry 2) and A328I (Entry 4) variants also showed high regioselectivity (~90%) for **10a** but both with low conversion (≤50%). The I263G mutation promoted β′-selectivity as well as enzymatic activity, with the RP/H171L/I263G giving 90% of **10a** with 86% conversion. Addition of I263G to an α′-selective variant, GVQ (95% of **10b**; Entry 10), strikingly switched the regioselectivity, with the GVQ/I263G variant showing 85% of **10a** with 90% conversion (Entry 5).

Table 4.14 Regioselectivity and conversion of selected $P450_{BM3}$ variants for the hydroxylation products of **10** (substrate:enzyme = 2000:1).

Entry	Variants	Conv.	10a	10b	10c
1	R19/F87A/A328F	35%	**92%**	6%	2%
2	R19/F87A/A328L	24%	92%	5%	3%
3	RP/H171L/I263G	86%	90%	10%	–
4	K19/F87V/A328I	50%	90%	10%	–
5	GVQ/I263G	90%	85%	13%	2%
6	GVQ/I263G/V78L	60%	68%	25%	7%
7	KU3/A330P	30%	20%	10%	**70%**
8	VQ/S72G/A330W	85%	15%	85%	–
9	RP/V78I/E267V	80%	8%	92%	–
10	GVQ	85%	5%	95%	–
11	R19/F87A/F81W	65%	4%	96%	–
12	R19/F87A/A330W	73%	4%	96%	–
13	R19/F87A	99%	4%	**96%**	–

The α′ alcohol **10b** was most favoured by the R19/F87A variant, with 96% regioselectivity and 99% conversion. The F81W and A330W mutations maintained the regioselectivity (96%) but decreased conversion to 55% and 73%, respectively (Entries 11 and 12, Table 4.14). The mutation F87V (Entry 10) and the combination V78I/E267V (Entry 9) also promoted α′-selectivity.

The α alcohol **10c**, which was of less interest, was most favoured by the KU3/A330P variant, giving 70% regioselectivity (Entry 7), indicating that the A330P mutation should be avoided in the mutagenesis work to improve selectivity for **10a** and **10b**.

4.4.3 Initial screen for *N*-Ips-8-azaspiro[4.5]decane

The *N*-Ips-derivative **11** was a rather poor substrate for $P450_{BM3}$ oxidation: 10 out of the 48 variants in the refined library showed >20% conversion (substrate:enzyme = 2000:1). There were four products across the initial screening and two of these were characterised as the α′ alcohol **11a** and the β′ alcohol **11b**. The other two were minor products (<5%). High regioselectivity for **11a** and **11b** was found and the effect of mutations was similar to the trends observed with the *N*-Ms-derivative **10**.

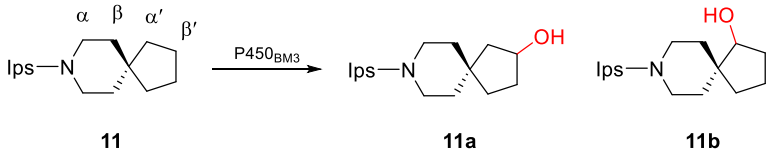
The α′ alcohol **11b** was most favoured by the R19/F87A variant, with 93% regioselectivity and 43% conversion (Entry 11, Table 4.15). Mutations F87V also promoted α′-selectivity, with the GVQ and K19/F87V variants giving ~90% **11b** and ~40% conversion (Entries 8 and 9). Addition of the E267V mutation to the K19/F87V variant slightly increased the regioselectivity to 92% but lowered the conversion to 28% (Entry 10). Mutations F81W (Entry 7) and A330W (Entry 6) also favoured **11b**.

The highest β′-selectivity was found with the RP/H171L/I263G variant, with 86% of **11a** and 21% conversion. Addition of the I263G mutation to the α′-selective GVQ variant (Entry 8)

switched the regioselectivity to β′ with the GVQ/I263G variant giving 80% of **11a** and 32% conversion (Entry 4). The A328L and A328F mutations also promoted β′-selectivity: the R19/F87A/A328L and R19/F87A/A328F variants gave ~85% of **11a** and ~24% conversion (Entries 2 and 3), while the R19/F87A precursor was highly α′-selective (Entry 11). Addition of the A328L/F mutations to the RP/H171L/I263G variant might promote the β′-selectivity further.

Interestingly, the α-selective variant for the *N*-Ms derivative **10**, KU3/A330P, showed little conversion for **11**, and the α alcohol product of **11** was not found across the screening against the same 48 variants used for **10**. These suggested the *N*-Ips derivative might suppress α-oxidation.

Table 4.15 Regioselectivity and conversion of selected $P450_{BM3}$ variants for the hydroxylation products of **11** (substrate:enzyme = 2000:1).



11 **11a** **11b**

Entry	Variants	Conversion	11a	11b	Others
1	RP/H171L/I263G	21%	**86%**	10%	4%
2	R19/F87A/A328L	24%	85%	8%	7%
3	R19/F87A/A328F	26%	83%	9%	8%
4	GVQ/I263G	32%	80%	13%	7%
5	KU3/A330P	<5%	–	–	–
6	RT2/A330W	40%	21%	77%	2%
7	K19/F87A/F81W	35%	14%	83%	3%
8	GVQ	38%	5%	90%	5%
9	K19/F87V	40%	4%	90%	6%
10	K19/F87V/E267V	28%	5%	92%	3%
11	R19/F87A	43%	6%	**93%**	1%

To summarise, the *N*-Ips derivative **11** was a much poorer substrate compared with its *N*-Ms counterpart **10**. Similar selectivity trends were observed for both substrates, but the highest selectivity for α′- (93%) and β′-(86%) alcohols found with **11** were both lower than those (96% and 92%, respectively) observed for **10**. Therefore, *N*-Ms-8-azaspiro[4.5]decane **10** was selected to conduct further stereochemistry determination studies. Mosher ester analysis was carried out for the two remote functionalisation products **10a** and **10b**.

4.4.4 Mosher ester analysis for stereochemistry and ee determination

Mosher ester analysis is an NMR-based method for deducing the absolute configuration of a secondary alcohol/amine centre.[250] Different arrays of chemical shifts in ^{1}H NMR spectra are displayed by the protons in diastereomeric (*R*-/*S*-) α-methoxy-α-trifluoromethylphenylacetic acid (MTPA) esters (as derived from conjugation of the carbinol of amino groups with MTPA), due to the shielding effect imposed by the phenyl group to protons residing above or below the plane of the aryl ring. This results in different $\Delta\delta^{RS}$ values (i.e., $\delta^R - \delta^S$, the difference in chemical shifts of a proton in the *R*- *vs*. *S*-MTPA esters) in the two configurations of the alcohol group. Comparative analysis of the experimental $\Delta\delta^{RS}$ values against the predicted values based on both (*R*)- and (*S*)-alcohols then gives a deduction of the configuration of the alcohol being examined. Furthermore, Mosher ester analysis could also be used to determine the enantiomeric ratio of the two enantiomers in an alcohol sample from the relative intensities of analogous resonances in each of the MTPA esters.[250]

The β′ alcohol **10a** afforded by the RP/H171L/I263G variant was firstly analysed. The purified product **10a** was reacted with two enantiomers of the Mosher's acid following the protocol of Hoye *et al*.[250] to give the (*R*)- and (*S*)-MTPA esters (Fig. 4.14). Assuming an (*S*)-configuration of the alcohol group, *i.e*., (2*S*)-**10a**, the two protons at C1 would experience more shielding

from the phenyl group in the (*R*)-MTPA ester than in the (*S*)-MTPA ester (Fig. 4.14). Thus, a negative $\Delta\delta^{RS}$ was predicted for the C1 protons (Table 4.16). The predicted $\Delta\delta^{RS}$ for protons at C2, C3, C4, C6 and C10 are listed in Table 4.16.

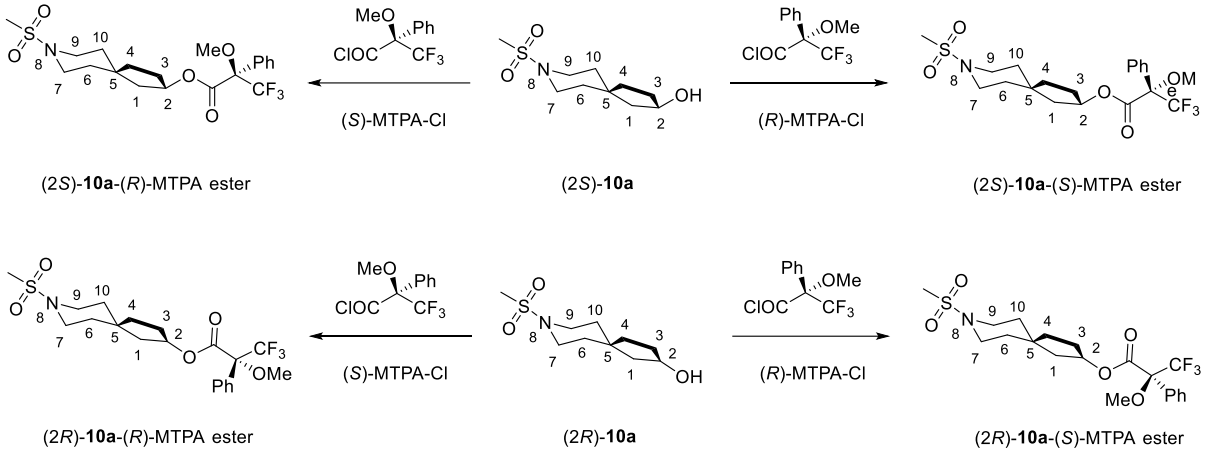


Figure 4.14 Scheme for the synthesis of the (*R*)- and (*S*)-MTPA esters from (2*R*)- and (2*S*)-**10a**.

Table 4.16 Predicted qualitive $\Delta\delta^{RS}$ values of protons at C1, C2, C3, C4 ,C6 and C10 based on (2*R*) and (2*S*) conformations of **10a**.

Assumed configuration	Predicted $\Delta\delta^{RS}$					
	C1	**C2**	**C3**	**C4**	**C6**	**C10**
(2*R*)-10a	> 0	< 0	< 0	–	> 0	< 0
(2*S*)-10a	< 0	> 0	> 0	> 0	< 0	–

The ^{1}H NMR spectra of the (*R*)- and (*S*)-MTPA esters were obtained and analysed. The chemical shift of the C2 proton was 5.44 ppm in the (*R*)-ester spectrum, while the resonance was at 5.42 ppm for the (*S*)-ester, resulting in a $\Delta\delta^{RS}$ value of +0.02. This experimental data is consistent with the predicted shift (> 0) for a (2*S*)-configuration (Fig. 4.15). Similar analysis of $\Delta\delta^{RS}$ of protons at C1, C3, C4, C6 and C10 mostly supported (2*S*)-configuration of the alcohol **10a** (Table 4.17). The major enantiomer in the product **10a** afforded by the RP/H171L/I263G was thus tentatively assigned an (*S*)-configuration.

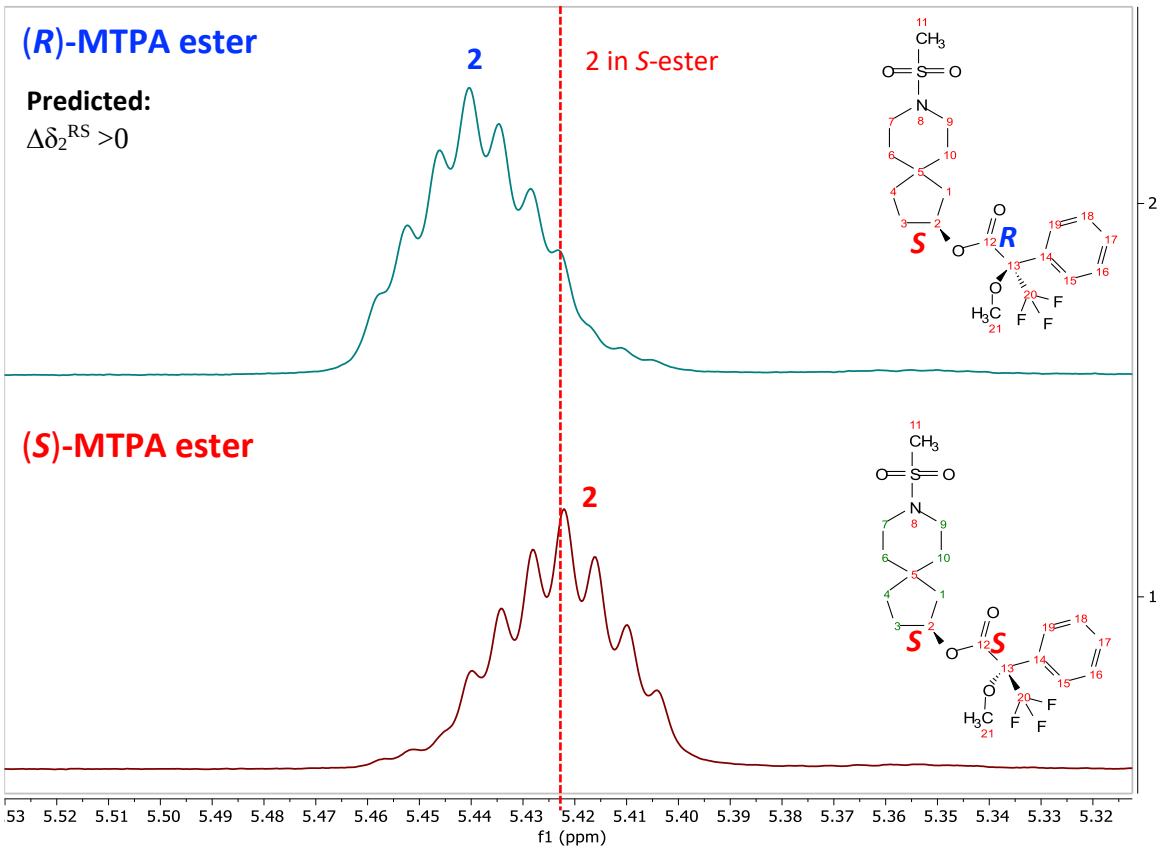


Figure 4.15. Comparative analysis of chemical shifts of proton 2 in ^{1}H NMR spectra of (*R*)- and (*S*)-MTPA esters.

Table 4.17 Comparative analysis of the $\Delta\delta^{RS}$ values of protons against the predicted values based on (2R) and (2S) configuration of **10a** formed by the RP/H171L/I263G variant.

Proton	δ in (***R***)**-ester** (ppm)	δ in (***S***)**-ester** (ppm)	$\Delta\delta^{RS}$ (ppm)	Predicted $\Delta\delta^{RS}$	
				2(*R*)-**10a**	2(*S*)-**10a**
1'	1.65	1.71	–0.06	>0	<0
1"	1.82	1.85	–0.03	>0	<0
2	5.44	5.42	+0.02	<0	>0
3'	1.90	1.84	+0.06	<0	>0
3"	2.06	2.05	+0.01	<0	–
4'	1.56	1.54	+0.02	–	>0
4"	1.61	1.54	+0.07	–	>0
6	1.55	1.56	–0.01	>0	<0
10	1.55	1.55	0	<0	–

It is worth noting, though, that this analysis is based on the assumption that the ester adopts the *S-trans* arrangement about the O–CO bond, and both the methine proton of the alcohol and the

the trifluoromethyl (CF_3) substituent of the MTPA moiety are *syn*-coplanar with the carbonyl group. However, in the spirocyclic system this assumption is not realistic, since the MTPA moiety is bonded to the cyclopentane ring of **10a** and not a chair structure as the usual Mosher's model describes.[250] Thus, the assignment of absolute stereochemistry of product **10a** remained tentative and X-ray crystallography study was required to confirm the configuration.

Nevertheless, the Mosher ester analysis remained a powerful tool to determine the ee of alcohol samples. The diastereomeric resonances of the methine group in the MTPA moiety were used to conduct the calculation, as they were relatively isolated and therefore had no overlapping issues. From integration of these resonances in both (*R*)- and (*S*)-MTPA esters, the ee of alcohol **10a** produced by the RP/H171L/I263G variant was determined to be ~78% (Fig. 4.16).

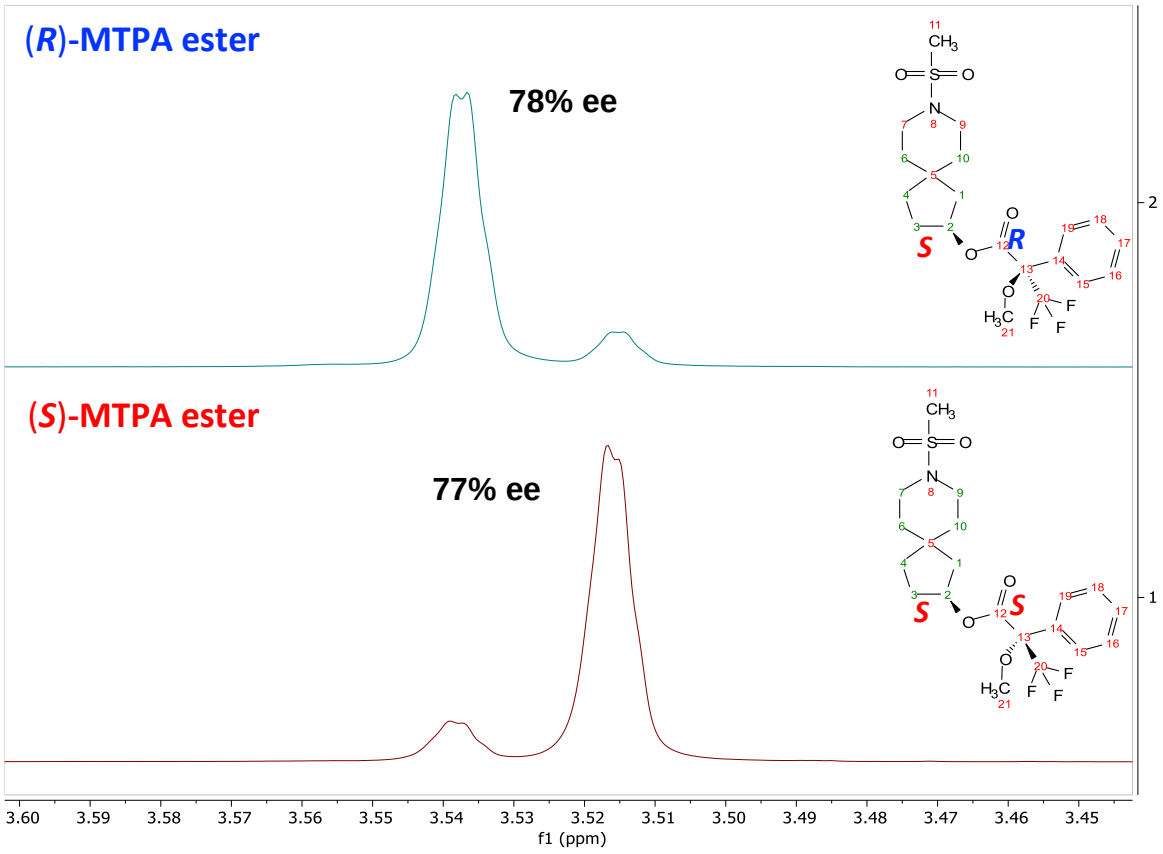


Figure 4.16. The ee calculation based on resonances of the MTPA methine groups of diastereomers .

4.4.5 X-ray crystallography study

In order to confirm the absolute stereochemistry of the alcohol **10a**, X-ray crystallography study was conducted. Single crystals of **10a** were obtained by the vapour diffusion method: purified **10a** (11 mg) was dissolved in ethyl acetate (~500 μL) in a small vial, which was placed into a larger container filled with hexane. The larger container was sealed and left still for several days to afford light yellow crystals. X-ray diffraction data was collected and analysed by Dr. Kirsten Christensen of the Department of Chemistry, University of Oxford. The major enantiomer of alcohol **10a** produced by the RP/H171L/I263G variant was determined to have an (*S*)-configuration (Fig. 4.17). This assignment supported the conclusion of the Mosher ester analysis.

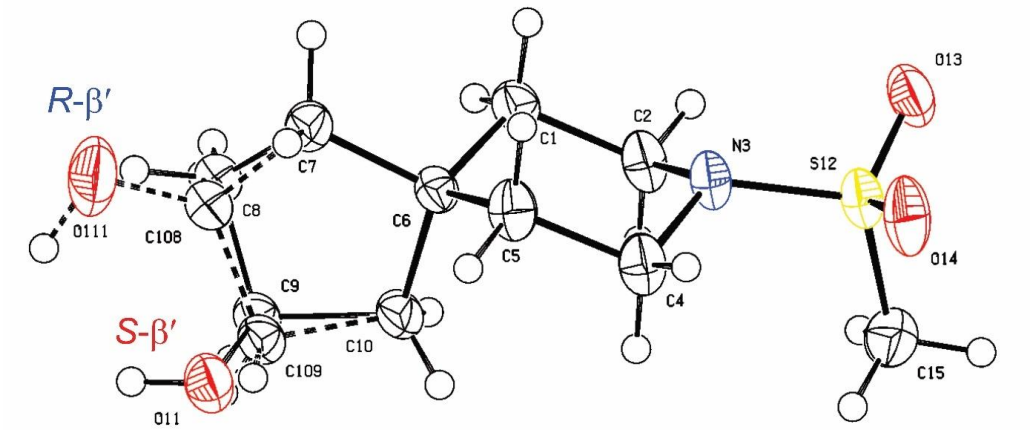


Figure 4.17. The ellipsoid plot for the crystal structure of **10a** as analysed from the X-ray diffraction data of the single crystal. Black lines represent structure of the major enantiomer (*S*-β′ configurated); Black dashed lines represent the minor enantiomer (*R*-β′ configurated).

4.4.6 Docking analysis with the RP/H171L/I263G variant

Docking analysis was conducted to understand the binding interactions and guide mutagenesis. Substrate **10** was docked with nine clusters of MD-simulation structures of RP/H171L/I263G (Fig. 5.17, Appendices **A5.5**) using the ADFR approach with residues L75, F87, P329, L437 and T438 set as flexible. Among the 95 poses generated, there were eight poses for *S*-β′

oxidation, seven for *R*-β′, ten for α′ and one for β, together with 69 non-productive poses. The *S*/*R* raito based on both pose counts (2.3) and weighted selectivity (4.8) were lower than 8.0 from experimental data (Table 4.18). The spirocyclic system has much greater complexity than the 7-azabicyclo[2.2.1]heptane and azepane substrates, which is likely the cause of the lower (*S*)/(*R*) raito and proportion of β′-oxidation in the weighted selectivity. On the other hand, there were few β-poses (3%) and none for α-oxidation, which agreed with the experimental trends, with no α- and β- alcohols detected. This suggested that docking was able to predict the correct binding direction, i.e., with the cyclopentane ring facing the ferryl oxygen, although the regio- and enantio-selectivity within the ring appeared to be more difficult to model.

Table 4.18 Predicted selectivity for oxidation products of the substrate **10** calculated from docking with MD-simulated structures of RP/H171L/I263G and their corresponding experimental data.

Variants		Pro-*S*-β′	Pro-*R*-β′	*S*-β/*R*-β	Pro-α′	Pro-α	Pro-β
RP/H171L/I263G (*S*-selective)	Number of poses	9	4	1.1	10	0	1
	Population/energy-weighted selectivity	48%	10%	4.8	34%	–	3%
	Experimental data	80%	10%	8.0	10%	–	–

The binding poses were then analysed. All *S*-β′ poses aligned surprisingly well, considering the flexibility of the substrate (Fig. 4.19-A). In these poses, the Ms groups were in the pocket formed by residues S72, A330 and L437, and the cyclopentane ring pointed towards the I helix and contacted F87 and L437. There was significant distance between the G263 residue and these poses, therefore the I263G mutation might have transmitted effects on other residues which in turn shaped the active site for *S*-β′ oxidation of **10** to be dominant. In the lowest energy *S*-β′ pose, the Cβ′ atom was 3.3 Å from the ferryl oxygen (Fig. 4.19-B).

The *R*-β′ aligned quite well and there was significant overlap between these and the lowest energy *S*-β′ pose (Fig. 4.20). As a result, it was difficult to design mutations that might interrupt *R*-β′ poses without affecting *S*-β′. Site-saturation mutagenesis at S72, A330 and L437 might bring in altered binding orientation and improved enantioselectivity. However, it remained very

time-consuming to analyse the ee of new variants since chiral-phase GC could not be used for this substrate and these analyses must be done by Mosher ester derivatisation of purified alcohol products afforded by each variant. Therefore, these mutations were not tested, but could be followed up in the future when a more convenient enantioselectivity screening analysis method is established for the spirocyclic amines.

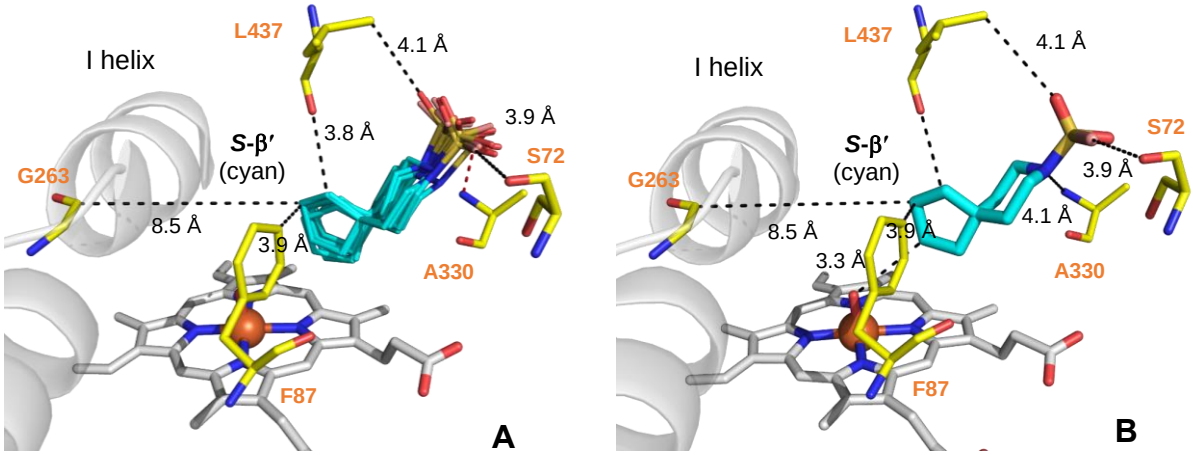


Figure 4.19 Overlay of MD-simulat*ion* structures of RP/H171L/I263G (yellow) with **(A)** all (S)-β′ poses (cyan); and **(B)** the lowest energy (S)-β′ pose (cyan). Black dashes with black labels denote distances between residues and poses. Red dashes denote H-bonds.

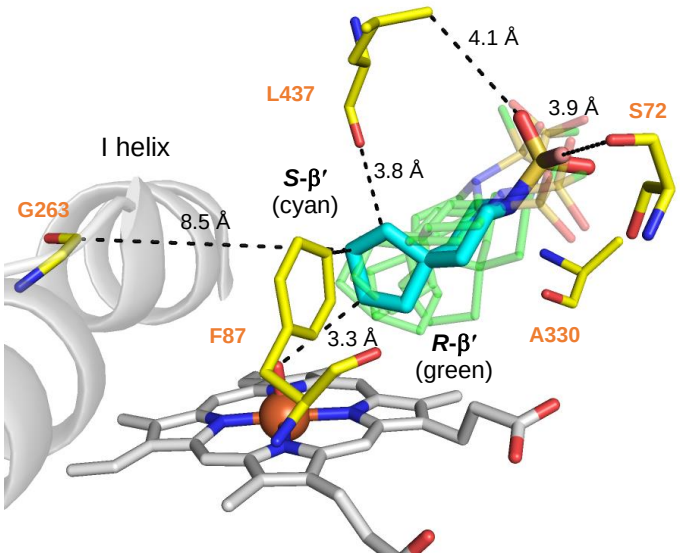


Figure 4.20 Overlay of lowest energy (S)-β′ pose (cyan) and (R)-β′ pose (green), together with the MD-simulat*ion* structures of RP/H171L/I263G (yellow). Black dashes with black labels denote distances between residues and poses.

On the other hand, the overlay of NP poses with the lowest energy *S*-β′ pose revealed significant differences–a "NP specific" region was identified where the lowest energy *S*-β′ pose was not located (Fig. 4.21). This region was close to residues M185 and L188. Mutations M185F/Y/W and L188M/F/Y/W were proposed as they could disfavour some NP poses while keep the *S*-β′ pose. These were again not tested experimentally due to the lack of an efficient screening method to analyse the enantioselectivity of new variants.

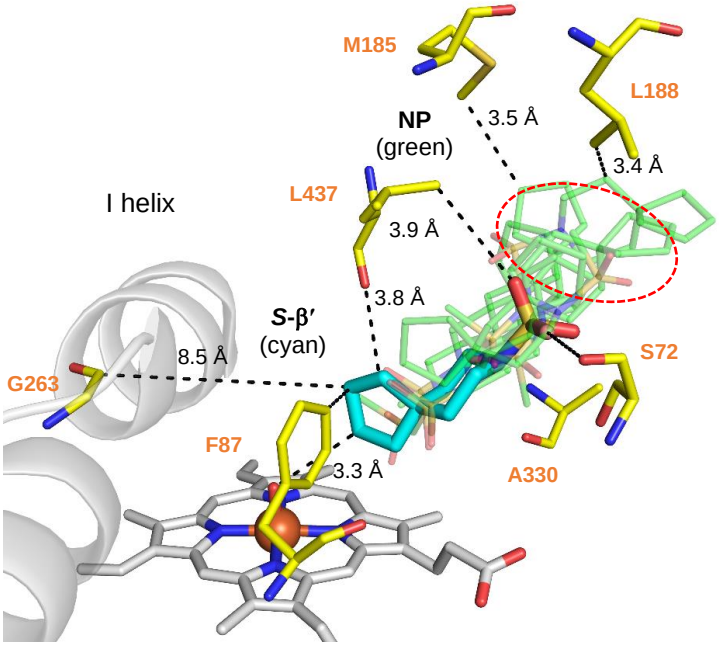


Figure 4.21 Overlay of lowest energy (S)-β′ pose (cyan) and NP pose (green), together with the MD-simulated structures of RP/H171L/I263G (yellow). Black dashes with black labels denote distances between residues and poses. Red dashed circles highlight the region where NP poses are bound but the (S)-β′ pose are not.

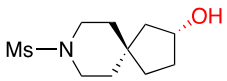
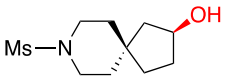
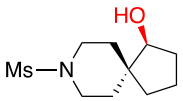
4.4.7 Mosher ester analysis for other variants

Mosher ester derivatisation was performed for a few other β′-selective variants and the stereochemistry and ee of **10a** from these variants were determined by comparison of their ^{1}H NMR spectra with those from the RP/H171L/I263G variant. Among these, the K19/F87V/A328I variant favoured the (*R*)-enantiomer, giving 90% of **10a** with 88% ee (Entry 1, Table 4.19), while its A328F analogue (Entry 2) showed higher regioselectivity (96%) but

much lower ee (60%). Interestingly, substitution of the F87V with F87A effectively switched the enantioselectivity to (*S*) with 70% ee by the R19/F87A/A328L variant (Entry 3).

The absolute configuration and ee of product **10b** were then studied. Mosher ester analysis of **10b** produced by the VQ/S72G/A330W variant suggested it had an (*S*)-configuration with 95% ee (Table A3.5, Appendix **A3.3.1**). Attempts to obtain single crystals of **10b** for X-ray diffraction study were unsuccessful. Thus, the assignment remained tentative. A screening Mosher ester analysis with a few **10b**-selective variants was conducted to determine the stereochemistry and ee. All the tested variants (Entries 6-10, Table 4.19) showed high *S*-α′-selectivity, giving >92% of **10b**, with ~95% ee for (*S*). However, no preference for (*R*)-**10b** was observed.

Table 4.19 Regioselectivity and ee for products **10a** and **10b** by selected variants from initial screen..

Product	Entry	Variants	Conversion	Regioselectivity	*ee*
 (*R*)-10**a**	1	K19/F87V/A328I	50%	90%	88%
	2	R19/F87V/A328F	50%	96%	60%
 (*S*)-10**a**	3	R19/F87A/A328L	40%	92%	70%
	4	RP/H171L/I263G	85%	90%	77%
 (*S*)-10**b**	5	VQ/S72G/A330W	100%	84%	95%
	6	R19/F87A/A82M/A330W	73%	97%	96%
	7	R19/F87A/F81W	55%	95%	94%
	8	R19/F87A	100%	96%	95%
	9	GVQ	85%	95%	95%
	10	RP/V78I/E267V	100%	92%	96%

4.4 Summary for the bicyclic and spirocyclic substrates

The combined strategies of initial screening, substrate engineering and docking-guided mutagenesis were deployed to the bicyclic amine 7-aza-bicyclo[2.2.1]-heptane and the

spirocyclic amine 8-azaspiro[4.5]decane to expand the substrate scope. From the initial and some further screening, high regio- and enantio-selectivity for three enantiomers of remote oxidation products of *N*-Boc-7-aza-bicyclo[2.2.1]-heptane **8**, the (*R*)-*exo* (99% of **8a** with 99% ee by R19/F87I), (*S*)-*exo* (99% of **8a** with 90% ee by GL/A184I/I263G/A328G), and (*R*)-*endo* (92% of **8b** with 96% ee by KU3/A330P/S72W) alcohols were observed (Fig. 4.22). While only low enantioselectivity for the (*S*)-*endo* alcohol was identified at 95% of **8b** with 44% ee by GQ/I263G/A328L, which was the only variant from the initial screen that favoured the (*S*)-*endo* product. With two rounds of docking-guided mutagenesis, the GQ/I263G/T260L/A330F variant was generated, which improved (*S*)-*endo*-selectivity to 98% of **8b** with 89% ee, doubled the 44% ee by GQ/I263G/A328L.

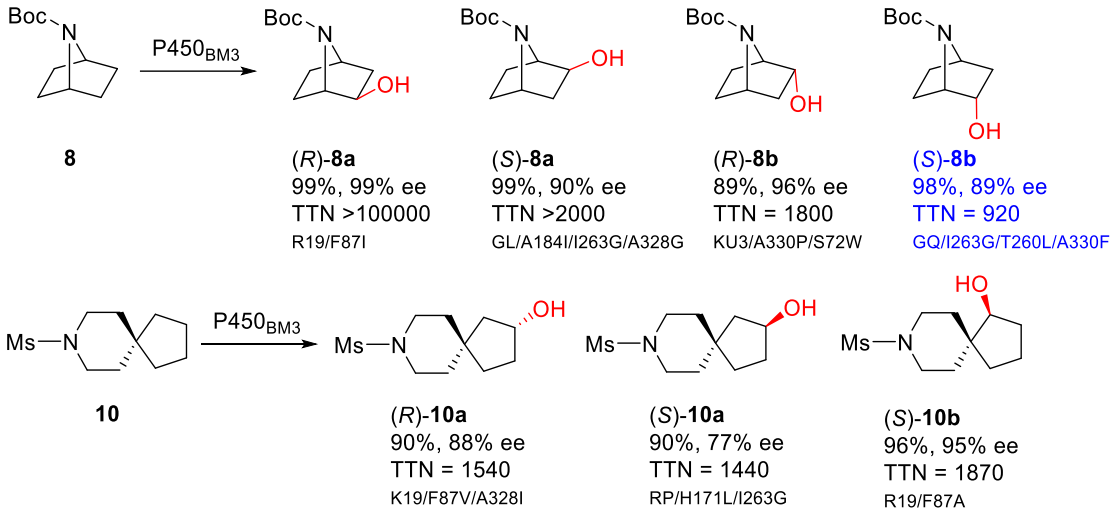


Figure 4.22 Highest selectivity and TTN achieved for all remote products of the bicyclic and spirocyclic substrates **8** and **10**. Marked in blue are products for which the selectivity and TTN have been improved through docking-guided mutagenesis.

For the spirocyclic amine, the assignment of the absolute stereochemistry of oxidation products was hardened by the fact that there was no literature data to compare with. Mosher ester analysis was explored but the relevant assignments were tentative. Instead, it was used to determine the ee of products from each variant. Single crystals were obtained for the β′ alcohol **10a** produced

by the RP/H171L/I263G variant. X-ray crystallography study conducted by Dr Kirsten Christensen confirmed the this as the (*S*)-enantiomer. Mosher ester analyses were carried out for products from selected variants to determine the enantioselectivity. High selectivity for the (*R*)-β′ (90% of **10a** and 88% ee by K19/F87V/A328I), (*S*)-β′ (90% of **10a** and 77% ee by RP/H171L/I263G), and (*S*)-α′ (96% of **10b** and 95% ee by R19/F87A) alcohols were observed (Fig. 4.22). No variant showed (*R*)-α′ preference across the screening. Docking analysis of substrate **10** with the RP/H171L/I263G variant revealed key residues and some mutations that could improve enzymatic activity and product selectivity were proposed. Future work to test these variants will be of interest.

Chapter 5

5. Experimental details

5.1 General materials, reagents, and equipment

Bacterial culture media components, kanamycin and IPTG were from Melford Laboratories, UK. Hen egg white lysozyme was from Sigma-Aldrich. $NADP^+$ was from Prozomix, UK, and glucose dehydrogenase (GDH) was supplied by Codexis, California. Oligonucleotides for site directed mutagenesis were supplied by Eurofins Genetic Service, UK. Water (ρ ~14 MΩ/cm) used for bacterial growth and buffers was from a Milipore ELIX-5 system. Buffers for molecular biology experiments were prepared using ultrapure water (ρ > 18 MΩ/cm) from a Millipore Milli-Q system. Media were autoclaved at 121 °C for 30 min. Components sensitive to high temperature were first sterilised via a 0.22 μm syringe filter prior to being added to the autoclaved media (cooled to below 40 °C). Media and buffers used in bacterial culture or enzymatic reactions were prepared as listed in Table 5.1.

Table 5.1 Components and preparation of media and buffers.

Media/buffers	Components
LB	10 g tryptone, 5 g yeast extract, 10 g NaCl, then add water up to 1L.
LB_{agar}	10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g bacto-agar, then add water up to 1L
$FB_{glycerol}$	7.4 g K_2HPO_4, 1.8 g NaH_2PO_4, 1.3 g $(NH_4)_2SO_4$, 1.0 g Na_2SO_4, 0.6 g Na_3citrate·$2H_2O$, 0.3 g NH_4Cl, 2.5 mL glycerol, then add water up to 0.5 L (after autoclaving and cooled down to 40 °C, add 1.5 mL **inoculation solution**).
Trace elements	20.1 g Na_2EDTA, 0.7 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 $ZnSO_4·7H_2O$, then add water up to 1L.
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).
Inoculation solution	15 mg kanamycin, 250 mg $MgSO_4·7H_2O$, 50 mg thiamine·HCl, add trace elements soluion to 1.5 mL.
Induction solution	380 mg IPTG, 672 mg δ-ALA·HCl, 112 mg $FeSO_4·7H_2O$, 600 mg kanamycin, then add water up to 40 ml.
200 mM KPi	2.4 g KH_2PO_4, 31.7 g K_2HPO_4, then add water up to 1L.

The amine substrates were supplied by Fluorochem, ChemCruz, Merck and Alfa Aesar, UK. Chemical reagents and solvents (HPLC grade) were from Merck, ThermoFisher Scientific and Acros, UK. Amine oxidation products were purified by flash column chromatography using Geduran Silica 60, 40–63 μm (ThermoFisher Scientific, UK), with petroleum ether (b.p. 30–40 °C)/ethyl acetate as the eluting solvent. Analytical thin-layer chromatography (TLC) was performed using petroleum ether (b.p. 30–40 °C)/ethyl acetate mixtures and bands were visualised and analysed with phosphomolybdic acid staining reagents.

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 using 1 cm pathlength quartz cuvettes. NMR spectra (^{1}H, ^{13}C, COSY, HSQC, HMBC, NOESY) 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 acquired on a Bruker microTOF spectrometer. Gas chromatographic (GC) analyses were carried out with a ThermoFisher Scientific Trace 1300 instrument. Substrate conversion and product percentages were determined from integrated peak areas. Optical rotations were measured on a UniPol 2000 digital polarimeter (Schmidt & Haensch, Germany). Measurements were taken at 25 °C using a wavelength of 589.44 nm. The pathlength of the cell was 1 dm. $[\alpha]_D^{25}$ was calculated by the formula $[\alpha]_D = \alpha/(l \times c)$, where α is the average reading taken from the polarimeter, l is the pathlength in decimetre, and c is the concentration of sample in g/100 mL.

X-ray crystallographic study was carried out by Dr Kirsten Christensen of the Department of Chemistry, University of Oxford. Low temperature single crystal X-ray diffraction data were collected using a (Rigaku) Oxford Diffraction SuperNova A diffractometer. Data were reduced using CrysAlisPro, solved using SuperFlip[251] and refined using CRYSTALS.[252,253] The Flack *x*

parameter was refined in each case.[254,255] Bayesian analysis of the Bijvoet pairs was also carried out using all the data used in the refinement.[256] This gave the Hooft *y* parameter, the P2 probability (the likelihood that the hand is correct given the crystal was enantiopure), and the P3 probability (the likelihood that the hand is correct given the crystal was enantiopure or racemic). Full crystallographic data (in CIF format) has been deposited with the Cambridge Crystallographic Data Centre (CCDC 2159261-62).

5.2 Mutagenesis and DNA preparation

5.2.1 PCR reactions

Primers for PCR reactions used in this project are listed in Table A2.4, Appendix **A2.1**. 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× Q5 polymerase reaction buffer, 1−2 μL parent DNA, 2 μL DMSO, 1.5 μL forward primer, 1.5 μL reverse primer and 1 μL Q5 polymerase to make a total volume of 50 μL. After being gently pipette-mixed and centrifuged for 5 seconds at 4000 *g*, the microcentrifuge tube was placed in the thermocycler and the program listed in Table 5.2 was run.

Table 5.2 Thermocycle program for PCR reactions.

Stage	Cycle(s)	Temperature (°C)	Duration
Initial denaturation	N.A.	98	30 s
PCR cycles	30-35	98	10 s
		T_m–5~T_m–1	30 s
		72	3 min 40 s
Final extension	N.A.	72	2 min
Final hold	N.A.	4	∞

The T_m values of primers were obtained from SnapGene and calculated in comparison with the WT $P450_{BM3}$ sequence. Once the thermocycle program was finished, the PCR mixtures were analysed by agarose gel electrophoresis.

5.2.2 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 seconds). RedSafe™ DNA gel staining (0.5 μL of a 10000× solution) 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 added until the gel was submerged. 2 μL of the loading dye was mixed with 10 μL of the PCR mixture and then carefully pipetted into a well in the gel; 10 μL of a DNA ladder (1 kbp) solution with DNA fragments of certain length was added to an adjacent well in the gel to provide a reference. A voltage difference of 90 V was applied across the gel for 30 minutes. The gel was then irradiated by a UV transilluminator. A strong band with the same mobility as a linearised 8500 bp fragment suggested a successful PCR step. Then, 1 μL of DpnI, a restriction enzyme which digests methylated DNA instead of the new DNA produced *in vitro* by PCR, was added to digest the template DNA for 2 hours at 37 °C.

5.2.3 DNA transformation

The DpnI digested PCR mixture (10 μL) was mixed with 40 μL of *Escherichia coli* (*E. coli*) DH5α competent cells in a 1.5 mL microcentrifuge tube and incubated for 45 min on ice. The cells were heat-shocked at 42 °C for 60 seconds and then placed back on ice for 2 min. Filter-sterilised SOC media (600 μL) was added to the mixture which was then incubated for 2 h in a shaker at 200 rpm, 37 °C. The culture was then spread onto an LB_{kan} agar plate and placed in a 37 °C oven to grow overnight.

5.2.4 DNA preparation

A single colony was picked from the agar plate and inoculated into 5 mL liquid LB_{kan} in a 15-mL centrifuge tube. The tube was left in a shaker at 180 rpm, 37 °C for 16 h. The culture was centrifuged at 9000 *g* and 4 °C for 5 min. Plasmid DNA was then extracted and purified using a GeneJET plasmid miniprep kit from ThermoFisher Scientific following the protocol provided by the manufacturer. The concentration (μg/μL) of the purified DNA 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 this 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 Sanger sequencing by Source Bioscience, UK. The sequencing was done using two primers, T7F and BM3FP2 (Table 5.3), 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 **5.2.3**.

Table 5.3 The DNA sequences of the T7F and BM3FP2 primers.

Primers	Sequence
T7F	TAATACGACTCACTATAGG
BM3FP2	CTGCAGCGAGCAAATCCAGACGAC

5.3 Enzyme expression and quantification

5.3.1 Over-expression of P450 enzyme

A single colony of transformed *E. coli* BL21 (DE3) from a LB_{kan} agar plate was inoculated into 500 mL FB_{gly} media containing 50 μg/mL kanamycin in a 2-L flask and shaken at 120 rpm for ~24 h at 37 °C. A 5 mL aliquot of the culture was mixed with 5 mL sterilised glycerol in a 15-mL centrifuge tube to prepare a cell bank which was then stored as 1-mL aliquots at −80 °C. Once the culture reached OD ~0.6, the incubator temperature was lowered to 20 °C and the

culture left to shake for 1 h. Protein production was then induced by adding 40 μL of 0.4 M IPTG, and haem synthesis supplemented with 500 μL of 0.1 M δ-ALA and 500 μL of 10 mM $FeCl_2$. The culture was shaking a further 48−72 hours, at 120 rpm, 20 °C, with supplements (kanamycin, IPTG, δ-ALA, and $FeCl_2$) added every 24 h. Cells were harvested by centrifugation at 9,000 *g*, 4 °C for 10 min.

The cell pellet was re-suspended in 20 mL potassium phosphate buffer (50 mM, pH = 8.0) per 500 mL culture, transferred to a beaker on ice and the cells were lysed by sonication via 10 cycles of 15-second pulses (55% amplification) on with 30-second off. The lysate was centrifuged for 15 min at 9,000 *g*, 4 °C to remove the cell debris. The P450 enzyme was recovered from the supernatant by ammonium sulfate precipitation. The amount of ammonium sulfate required to take the solution to 40% saturation was calculated (http://www.encorbio.com/protocols/AM-SO4.html) and added to the solution, which was then stirred at 4 °C until all the solids had dissolved. The solution was centrifuged at 9,000 *g,* 4 °C for 15 min. The supernatant was retained and to this was added the amount of ammonium sulfate that take the solution from 40% to 60% saturation to precipitate the P450 enzyme. After the ammonium sulfate had all dissolved the solution was centrifuged at 9,000 *g* for 5 min at 4 °C, after which the supernatant was discarded. The retained pellet was then dissolved in 5 ml potassium phosphate buffer (200 mM, pH = 8.0). The purified enzyme was quantified as described in **5.3.2** and divided into 1-mL aliquots and stored at −20 °C.

5.3.2 P450 quantification

The P450 enzyme stock solution was defrosted on ice and diluted 20-fold by adding 50 μL to 950 μL of potassium phosphate buffer. A few crystals of sodium dithionite were added to reduce the P450 enzyme and the UV–vis spectrum was measured and used as the baseline. CO (~20 bubbles) was then gently bubbled through the solution, and a new spectrum was recorded. More

CO was bubbled through the solution and the spectrum recorded until the 450 nm peak height was maximised. The UV–vis absorbance at 450 and 490 nm, A_{450} and A_{490}, are recorded and the concentration of the protein was then calculated from the following equation:

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

Where $\varepsilon_{450-490}$ = 91,000 M^{-1} cm^{-1} is the extinction coefficient.[81]

5.4 Substrate screening and product characterisation

5.4.1 Initial screen

Amine substrates were screened with the initial library containing 48 $P450_{BM3}$ variants on a 1 mL scale. These reactions were conducted in two 24-well plates in 200 mM phosphate buffer (pH 8.0), along with 4 μM $NADP^+$, 100 mM glucose, 2 U/mL GDH, 2 mM substrate, and 1 μM enzyme. Screening plates were prepared by aliquoting 100 μL of the 10 μM enzyme solution into each well along with 900 μL of the solution listed in the Table 5.4. The plates were then put on a platform shaker and shaken at 140 rpm and ambient temperature for 16 hours. After this time, 300 μL of each reaction mixture was pipetted into a 1.5 mL microcentrifuge tube along with 300 μL of ethyl acetate. The mixtures were vortexed and then centrifuged for 2 minutes at 14,300 g to separate the phases. 150 μL of the organic layer was transferred into a 250 μL glass GC insert and GC analyses were carried out with a ThermoFisher Scientific Trace 1300 instrument equipped with a flame ionisation detector (FID) and an AI1310 autosampler using a DB-1 fused silica column (30 m × 0.25 mm i.d. × 0.25 μm film thickness, Agilent Technology, UK) or for chiral-phase analysis a Cyclosil-B fused silica column (30 m × 0.25 mm × 0.25 μm, Agilent Technology, UK) using helium as the carrier gas at a flow rate of 1.5 mL/min. For normal-phase analyses, both the injector and the FID were held at 250 °C. For

chiral-phase analyses, the injector was held at 200 °C and the FID at 250 °C. The oven settings and retention times for various products can be found in Appendi*x* **A4.1.1**.

Table 5.4 Components for screening reactions.

Substance	Stock concentration	Volume/μL	Final concentration
Glucose	1 M	100	100 mM
Substrate	200 mM	10	5 mM
GDH	2 U/μL	10	2 U/mL
$NADP^+$	1 M	10	80 μM
Enzyme	10 μM	100	2 μM
Phosphate buffer	N/A	770	200 mM

5.4.2 Preparative scale reactions and product characterisation

Preparative scale reactions (20 mL − 200 mL) were carried out for 16–24 h in 200- or 500-mL beaker with a stirring bar, with the same concentrations as in the initial screening. Progress of the reaction was monitored by GC. After the reaction had proceeded to >90% conversion (GC), the reaction mixture was extracted with ethyl acetate (3 × 50 mL). The organic extracts were combined and washed with 50 mL saturated aq $NaHCO_3$ and then 50 mL brine, dried over Na_2SO_4, filtered, and the solvent removed by rotary evaporation. The residue was dissolved in ~5 mL of EtOAc 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 be used for purification. Column chromatography purification was carried out with Geduran Silica 60, 40−63 μm. The fractions were analysed by TLC, stained with a solution of phosphomolybdic acid in ethanol, and by GC. Fractions containing the same compounds were combined, solvents removed by rotary evaporation, and the purified compound characterised by NMR and MS.

5.4.3 Preparation of Mosher ester derivatives

Based on the procedure of Hoye *et al*.,[257] The alcohol product to be derivatised (0.04 mmol.), was mixed with dry pyridine (10 μL, 0.12 mmol) in a screw-capped 10 mL glass vial charged

with a magnetic stirring bar. 1 mL of anhydrous dichloromethane was then added to the mixture. The (*R*)- or (*S*)-(+)-α-methoxy-α-(trifluoromethyl)phenylacetyl chloride (15 μL, 0.08 mmol) was then added dropwise. The vial was then capped and the reaction mixture stirred at ambient temperature for 2 h. After 2 h, once the reaction had proceeded to completion (TLC and GC). The solution was washed with 1 mL saturated aq $NaHCO_3$ and then 1 mL brine, dried over Na_2SO_4, filtered, and the solvent removed by rotary evaporation. The mixture was purified by silica gel column chromatography (30% ethyl acetate in petroleum ether) to afford the relevant Mosher esters of the alcohols.

5.5 MD simulations and molecular docking

5.5.1 MD simulations

MD simulations were performed for the haem domain of wild type $P450_{BM3}$ and variants that showed high selectivity for relevant products. All structures were relaxed via MD simulations performed in GROMACS 2018.6.[241] Mutations were added to the WT $P450_{BM3}$ (PDB code: 1jpz)[258] or the A330P variant (PDB code: 3m4v)[165] crystal structures via Pymol. The co-crystallised substrate for the WT structure, *N*-palmitoylglycine, was removed prior to adding mutations. The haem moiety was simulated as the ferryl, Compound I species, using parameters from Shahrokh, *et al*.[259] The protein was prepared using the Amber 99SB*-ILDN force field[260] with TIP3P water.[261] The protein was placed into the centre of an octahedral box, with a minimum distance of 10 nm to any box edge, and then solvated with *ca.* 17500 water molecules and the net charge neutralised by addition of corresponding Na^+ ions. Steepest descent energy minimisation was performed until the maximum force was <500 kJ mol^{-1} nm^{-2}. Periodic boundary conditions were used to model the system, with electrostatic interactions treated by the particle mesh Ewald method,[262] and bond lengths involving hydrogen atoms constrained

with the LINCS algorithm.[199] Short-range non-bonded interactions were calculated with 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.[263] This was followed by an NPT equilibration step for 1250 ps utilising the Berendsen barostat with a time constant of 1 ps.[264] All positional restraints were then removed and 100 ns of production MD was performed in triplicate or quadruplicate using the Parrinello-Rahman barostat[265] 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.

5.5.2 Molecular docking

The individual MD trajectories were clustered using the Daura algorithm[266] with a RMSD cut-off of 1.2 Å. The backbone Cα atoms of the full sequence of the haem domain of $P450_{BM3}$ were used to calculate the RMSD values for clustering. The clustering resulted in typically ~15 clusters, each containing different number of similar conformations of the enzyme. The central member of the three most populated clusters, with a population cut-off of 5% of all trajectories, were used as receptor structures for docking studies, which were performed in Autodock Vina (for rigid-body docking),[220] or ADFR (for flexible receptor docking).[221] All water molecules were removed from the structure prior to the docking calculation. Substrate structures were obtained from PubChem and energy-minimisation was performed by Chem3D. The resulting structures were processed by Autodock Tool to generate the PDBQT file for docking runs. In Autodock Vina, the docking site was defined as a 30 × 30 × 30 Å box centred on the ferryl oxygen, and the affinity energies of binding poses were calculated using the Autodock Vina scoring function. The exhaustiveness parameter was set as 20, meaning 20 independent calculations, starting from random conformations, were performed for each receptor-ligand pair, giving 20 binding poses. These poses were ranked based on their affinity energies and the nine

poses with the lowest energies were output. In ADFR, the entire active site is accessible and the side chain of residues at positions 75, 78, 87, 260, 263, 267, 328, 330, 437 and 438 were selected to have full flexibility on the basis that these were most likely to contact a bound substrate. By default, an ADFR docking experiment performed 50 independent calculations, generating 50 binding poses. These poses were then clustered with the objective to remove duplicated poses and the lowest energy pose from each cluster was reported.

The binding poses output from both Vina and ADFR were analysed using the same criteria. A binding pose was identified as "productive" if at least one of the carbons had 2.9−4.2 Å distance to the ferryl oxygen. Otherwise, the pose was classified as "non-productive". For productive poses, the closest carbon to ferryl oxygen was determined as the hydroxylation site. The C–H···O and H···O=Fe angles of the two hydrogens on the closest carbon were then examined. The hydrogen of which the C–H···O angle was closest to 150° (accepted range 130−170°) and the H···O=Fe angle was closest to 130° (accepted range 110−130°) was assigned as most likely to be abstracted.

Chapter 6

6. Conclusion and further work

The main goal of this research was to achieve selective remote functionalisation of cyclic amines by engineered variants of $P450_{BM3}$. The combined strategies of variant library screening and substrate engineering were deployed together with docking-guided mutagenesis. High product selectivity and enzymatic activity were targeted, with the aim of 90% regioselectivity, 90% enantioselectivity and 90% conversion at an enzyme:substrate ratio of ≥1000:1 for each product (90/90/90). This goal was achieved for most of the remoted oxidation products of azepane, azocane, 7-aza-bicyclo[2.2.1]-heptane and the 8-azaspiro[4.5]decane (Fig. 6.1).

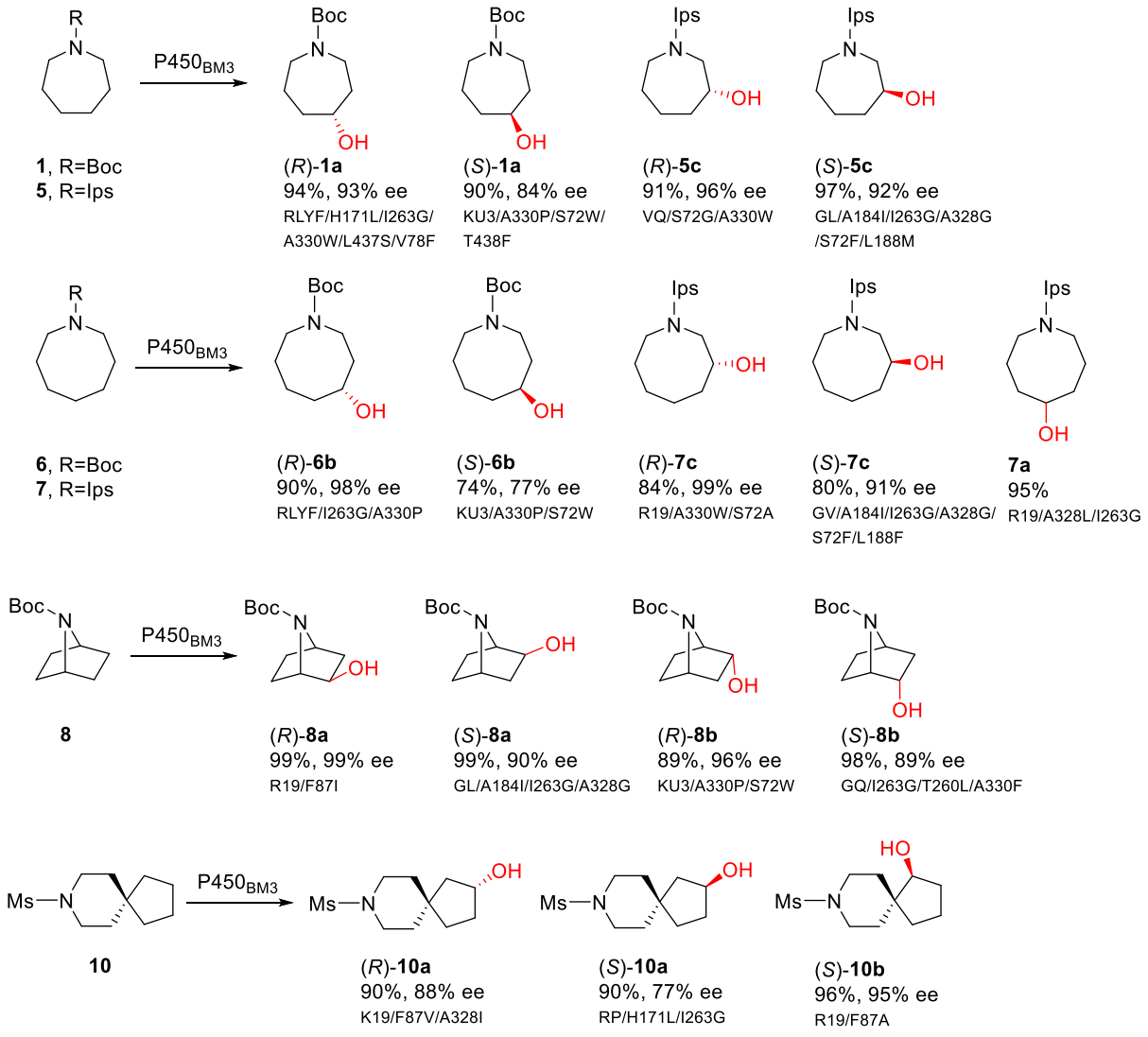


Figure 6.1. Highest selectivity for the major hydroxylation products of cyclic amines.

The initial screening and substrate engineering for azepane and azocane led to medium to high regio- and enantio-selectivity for both enantiomers of all the remote functionalisation products (up to 90% regioselectivity and 90% ee). Different selectivity patterns were observed for differently protected derivatives screened against the same panel of 48 $P450_{BM3}$ variants. Notably, the *N*-isopropylsulfonyl (*N*-Ips) derivatives (**5** and **7**) revealed high preference for β alcohols **5c** and **7c**. From screening data, some common residues/mutations were identified that promoted the same product of azepane and azocane, such as, I263G for (*R*)-γ alcohols, A330P and S72W for (*S*)-γ, A330W for (*R*)-β, and I263G/A328G for (*S*)-β. Rational design based on these selectivity trends improved selectivity for oxidation products of azepane and azocane as exemplified in **2.2** and **2.3.4**. Iterative docking-guided mutagenesis (IDGM) then led to further improvement of product selectivity and enzymatic activity and achieved 90/90/90 goal for most of the oxidation products of azepane and azocane.

The substrate scope was then expanded and the bicyclic amine 7-aza-bicyclo[2.2.1]-heptane and the spirocyclic amine 8-azaspiro[4.5]decane were studied by deploying these combined strategies. High regio- and enantio-selectivity for most of the remote oxidation products were observed after initial screening and substrate engineering. No variant showed a preference for (*R*)-α′ oxidation of *N*-Ms-8-azaspiro[4.5]decane **10**, and only low enantioselectivity was found for the (*S*)-*endo* alcohol of *N*-Boc-7-aza-bicyclo[2.2.1]-heptane **8** (95% of **8b** with 44% ee) with GQ/I263G/A328L – the only variant from the initial screen that favoured the (*S*)-*endo* product. Without any indications from the screening results of residues and mutations that could promote (*S*)-selectivity for **8b**, the value of docking-guided mutagenesis was illustrated by the generation of the GQ/I263G/T260L/A330F variant after two rounds of IDGM, forming 98% of **8b** with 89% ee.

For most of the products, screening against the 48 variants in the initial or refined library were sufficient to give high selectivity, demonstrating the diversity of mutations and substrate pocket

topologies in the variant library. Substrate engineering led to significantly altered selectivity patterns and provided alternative routes to access certain products. Rational design based on experimental selectivity trends could improve selectivity; however, this required relevant effective mutations to be present in the library in the first place. This suggested a wider library (96 or 144 variants *etc*., instead of 48), with more residues and combination of mutations included, would be beneficial. Additionally, site saturation mutagenesis at key residues could be conducted to seek further improvement of enantioselectivities.

The IDGM approach facilitated mutation design to improve enzymatic activity and product selectivity and only ~10 variants were required in each round of mutagenesis. This approach only requires the unwanted poses to be sufficiently distinct from the poses for the desired product. It is simple to apply and complements more comprehensive approaches such as random mutagenesis, CAST and ISM in engineering enzymes for application in synthesis. The simplest and most efficient application of IDGM would be to improve enzymatic activities, since this only requires the NP poses, which are typically distinct from the correct poses, to be disfavoured. There were many such successful examples in this work, the most significant one being the discovery of R19/A330W/S72A/L181F variant, which showed a TTN for product (*R*)-**7c** of 2700, eight-fold of 340 by R19/A330W/S72A. This was achieved by a single round of mutagenesis. Notably, the IDGM approach has also allowed discovery of effective mutations at residues that were previously not targeted, such as V26, L181, L188, S332, and T438, which added a dimension to the mutation design.

However, shortcomings of the IDGM approach are also recognised. When the correct poses and other poses adopt similar binding orientation, it is not possible to design mutations to alter selectivity or enzymatic activity. Another limiting aspect is that the analysis of the outputs of docking calculation is time-consuming. Automation of these processes, such as measurement of distances and angles, and sorting binding poses into corresponding products, would speed

up the docking analysis and facilitate more efficient mutation design and enzyme evolution. Moreover, it is worth noting that the IDGM approach will not always work as expected and initial hits are typically required. Thus, a stress test will be of interest. For instance, docking analysis could be performed for spirocyclic amine **10** with an (*S*)-α′-selective variant and mutations that disfavoured (*S*)-α′ poses could be tested to seek any (*R*)-α′-preference, which was not observed experimentally. Finally, for molecules that are hydrophobic, symmetrical and lacking substituents, molecular docking is likely to give many poses of similar binding orientations and affinity energies but indicating more products, making data analysis and mutation design more difficult. To this end, improvement in the methodologies of docking calculations will be needed.

Future work to deploy the IDGM approach to other substrates, such as terpenes, steroids, lactams, and more complex amines, will be of interest. Further developments of this approach would benefit tailored application for different substrates. A more *in silico* version of IDGM could be explored. For this, analysis of binding poses and identification of key residues could be processed computationally. Mutations would be added to the template variant, MD-simulation performed, and docking with the simulation variant structure conducted to predict the outcome. Effective variants could be used to perform further rounds of *in silico* evolution via docking analysis and mutagenesis. The variant, evolved *in silico*, with a few mutations added, could be tested experimentally. To establish this, automation of docking analysis and improvement in docking calculation methodologies are highly desirable.

References

1. Chen, W.; Paul, A.; Abboud, K. A., *et al.*, Rapid functionalization of multiple C-H bonds in unprotected alicyclic amines. *Nat Chem* **2020,** *12* (6), 545-550.

2. Trindade, A. F.; Faulkner, E. L.; Leach, A. G., *et al.*, Fragment-oriented synthesis: beta-elaboration of cyclic amine fragments using enecarbamates as platform intermediates. *Chem Commun (Camb)* **2020,** *56* (62), 8802-8805.

3. Chen, W.; Ma, L.; Paul, A., *et al.*, Direct alpha-C-H bond functionalization of unprotected cyclic amines. *Nat Chem* **2018,** *10* (2), 165-169.

4. Topczewski, J. J.; Cabrera, P. J.; Saper, N. I., *et al.*, Palladium-catalysed transannular C-H functionalization of alicyclic amines. *Nature* **2016,** *531* (7593), 220-224.

5. Kang, Y.; Richers, M. T.; Sawicki, C. H., *et al.*, C-H functionalization of cyclic amines: redox-annulations with alpha,beta-unsaturated carbonyl compounds. *Chem Commun (Camb)* **2015,** *51* (53), 10648-51.

6. Li, C.; Liu, Y.; Pei, X.-Q., *et al.*, Stereo-complementary bioreduction of saturated N -heterocyclic ketones. *Process Biochem* **2017,** *56*, 90-97.

7. Vitaku, E.; Smith, D. T.; Njardarson, J. T., Analysis of the structural diversity, substitution patterns, and frequency of nitrogen heterocycles among U.S. FDA approved pharmaceuticals. *J Med Chem* **2014,** *57* (24), 10257-74.

8. Mitchell, E. A.; Peschiulli, A.; Lefevre, N., *et al.*, Direct alpha-functionalization of saturated cyclic amines. *Chem Eur J* **2012,** *18* (33), 10092-142.

9. Campos, K. R., Direct sp3 C-H bond activation adjacent to nitrogen in heterocycles. *Chem Soc Rev* **2007,** *36* (7), 1069-84.

10. Hameed, A.; Javed, S.; Noreen, R., *et al.*, Facile and Green Synthesis of Saturated Cyclic Amines. *Molecules* **2017,** *22* (10).

11. Lee, S.; Sperry, J., Isolation and biological activity of azocine and azocane alkaloids. *Biorg Med Chem* **2022,** *54*.

12. Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H., *et al.*, Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **2019,** *24* (23).

13. Price, A. J.; Howard, S.; Cons, B. D., Fragment-based drug discovery and its application to challenging drug targets. *Essays Biochem* **2017,** *61* (5), 475-484.

14. Murray, C. W.; Rees, D. C., The rise of fragment-based drug discovery. *Nat Chem* **2009,** *1* (3), 187-92.

15. St Denis, J. D.; Hall, R. J.; Murray, C. W., *et al.*, Fragment-based drug discovery: opportunities for organic synthesis. *RSC Med Chem* **2020,** *12* (3), 321-329.

16. Grainger, R.; Heightman, T. D.; Ley, S. V., *et al.*, Enabling synthesis in fragment-based drug discovery by reactivity mapping: photoredox-mediated cross-dehydrogenative heteroarylation of cyclic amines. *Chem Sci* **2019,** *10* (8), 2264-2271.

17. de Esch, I. J. P.; Erlanson, D. A.; Jahnke, W., *et al.*, Fragment-to-Lead Medicinal Chemistry Publications in 2020. *J Med Chem* **2022,** *65* (1), 84-99.

18. Jahnke, W.; Erlanson, D. A.; de Esch, I. J. P., *et al.*, Fragment-to-Lead Medicinal Chemistry Publications in 2019. *J Med Chem* **2020,** *63* (24), 15494-15507.

19. Schmidt, M. F.; Rademann, J., Dynamic template-assisted strategies in fragment-based drug discovery. *Trends Biotechnol* **2009,** *27* (9), 512-21.

20. Gradler, U.; Schwarz, D.; Blaesse, M., *et al.*, Discovery of novel Cyclophilin D inhibitors starting from three dimensional fragments with millimolar potencies. *Bioorg Med Chem Lett* **2019,** *29* (23), 126717.

21. Hubbard, R. E., Fragment approaches in structure-based drug discovery. *J Synchrotron Radiat* **2008,** *15* (Pt 3), 227-30.

22. Atobe, M.; Serizawa, T.; Yamakawa, N., *et al.*, Discovery of 4,6- and 5,7-Disubstituted Isoquinoline Derivatives as a Novel Class of Protein Kinase C zeta Inhibitors with Fragment-Merging Strategy. *J Med Chem* **2020,** *63* (13), 7143-7162.

23. Dutta, S.; Li, B.; Rickertsen, D. R. L., *et al.*, C-H Bond Functionalization of Amines: A Graphical Overview of Diverse Methods. *SynOpen* **2021,** *5* (3), 173-228.

24. Vo, C. V.; Bode, J. W., Synthesis of saturated N-heterocycles. *J Org Chem* **2014,** *79* (7), 2809-15.

25. Kumar, G.; Pradhan, S.; Chatterjee, I., N-Centered Radical Directed Remote C-H Bond Functionalization via Hydrogen Atom Transfer. *Chem Asian J* **2020,** *15* (6), 651-672.

26. Beak, P.; Lee, W. K., α-Lithioamine synthetic equivalents from dipole-stabilised carbanions: the t-Boc group as an activator for α′-lithiation of carbamates. *Tetrahedron Lett* **1989,** *30* (10), 1197-1200.

27. Beak, P.; Basu, A.; Gallagher, D. J., *et al.*, Regioselective, diastereoselective, and enantioselective lithiation-substitution sequences: reaction pathways and synthetic applications. *Acc Chem Res* **1996,** *29* (11), 552-560.

28. O'Brien, P., Basic instinct: design, synthesis and evaluation of (+)-sparteine surrogates for asymmetric synthesis. *Chem Commun (Camb)* **2008,** (6), 655-67.

29. Snieckus, V.; Cuevas, J. C.; Sloan, C. P., *et al.*, Intramolecular α-amidoyl to aryl 1,5-hydrogen atom transfer reaction. Heteroannulation and α-nitrogen functionalisation by radical translocation. *J Am Chem Soc* **1990,** *112* (2), 896-898.

30. Kamon, T.; Irifune, Y.; Tanaka, T., *et al.*, Total synthesis of (±)-kainic acid: a photochemical C–H carbamoylation approach. **2011,** *13* (10), 2674-2677.

31. Marinkovic, S.; Hoffmann, N., Efficient radical addition of tertiary amines to electron-deficient alkenes using semiconductors as photochemical sensitisers. *Chem Commun (Camb)* **2001,** (17), 1576-7.

32. Norinder, J.; Matsumoto, A.; Yoshikai, N., *et al.*, Iron-catalysed direct arylation through directed C–H bond activation. *J Am Chem Soc* **2008,** *130* (18), 5858-5859.

33. McNally, A.; Prier, C. K.; MacMillan, D. W., Discovery of an alpha-amino C-H arylation reaction using the strategy of accelerated serendipity. *Science* **2011,** *334* (6059), 1114-7.

34. Chu, L.; Qing, F. L., Benzoyl peroxide (BPO)-promoted oxidative trifluoromethylation of tertiary amines with trimethyl(trifluoromethyl)silane. *Chem Commun (Camb)* **2010,** *46* (34), 6285-7.

35. Zhang, C.; De, C. K.; Mal, R., *et al.*, α-Amination of nitrogen heterocycles: ring-fused aminals. *J Am Chem Soc* **2008,** *130* (2), 416-417.

36. Maruyama, T.; Mizuno, Y.; Shimizu, I., *et al.*, Reactions of a N-Acyliminium Ion Pool with Benzylsilanes. Implication of a Radical/Cation/Radical Cation Chain Mechanism Involving Oxidative C−Si Bond Cleavage. *J Am Chem Soc* **2007,** *129* (7), 1902-1903.

37. Condie, A. G.; González-Gómez, J. C.; Stephenson, C. R. J., Visible-Light Photoredox Catalysis: Aza-Henry Reactions via C−H Functionalization. *J Am Chem Soc* **2010,** *132* (5), 1464-1465.

38. Sun, P.; Sun, C.; Weinreb, S. M., Stereoselective Total Syntheses of the Racemic Form and the Natural Enantiomer of the Marine Alkaloid Lepadiformine via a Novel N-Acyliminium Ion/Allylsilane Spirocyclization Strategy. *J Org Chem* **2002,** *67* (12), 4337-4345.

39. Ghobrial, M.; Harhammer, K.; Mihovilovic, M. D., *et al.*, Facile, solvent and ligand free iron catalyzed direct functionalization of N-protected tetrahydroisoquinolines and isochroman. *Chem Commun (Camb)* **2010,** *46* (46), 8836-8.

40. Murahashi, S.-I.; Komiya, N.; Terai, H., *et al.*, Aerobic Ruthenium-Catalyzed Oxidative Cyanation of Tertiary Amines with Sodium Cyanide. *J Am Chem Soc* **2003,** *125* (50), 15312-15313.

41. Payne, P. R.; Garcia, P.; Eisenberger, P., *et al.*, Tantalum Catalyzed Hydroaminoalkylation for the Synthesis of α- and β-Substituted N-Heterocycles. *Org Lett* **2013,** *15* (9), 2182-2185.

42. DeBoef, B.; Pastine, S. J.; Sames, D., Cross-Coupling of sp3 C−H Bonds and Alkenes: Catalytic Cyclization of Alkene−Amide Substrates. *J Am Chem Soc* **2004,** *126* (21), 6556-6557.

43. Chatani, N.; Asaumi, T.; Yorimitsu, S., *et al.*, Ru3(CO)12-Catalyzed Coupling Reaction of sp3 C−H Bonds Adjacent to a Nitrogen Atom in Alkylamines with Alkenes. *J Am Chem Soc* **2001,** *123* (44), 10935-10941.

44. Davies Huw, M. L.; Jin, Q., Catalytic asymmetric reactions for organic synthesis: The combined C—H activation/Cope rearrangement. *Proc Natl Acad Sci USA* **2004,** *101* (15), 5472-5475.

45. Klapars, A.; Campos, K. R.; Waldman, J. H., *et al.*, Enantioselective Pd-Catalyzed α-Arylation of N-Boc-Pyrrolidine: The Key to an Efficient and Practical Synthesis of a Glucokinase Activator. *J Org Chem* **2008,** *73* (13), 4986-4993.

46. Campos, K. R.; Klapars, A.; Waldman, J. H., *et al.*, Enantioselective, Palladium-Catalyzed α-Arylation of N-Boc-pyrrolidine. *J Am Chem Soc* **2006,** *128* (11), 3538-3539.

47. Sulikowski, G. A.; Lee, S., Kinetic isotope effects in the intramolecular insertion of a metal carbene into a C□H bond adjacent to nitrogen. *Tetrahedron Lett* **1999,** *40* (46), 8035-8038.

48. Millet, A.; Larini, P.; Clot, E., *et al.*, Ligand-controlled β-selective C(sp3)–H arylation of N-Boc-piperidines. *Chem Sci* **2013,** *4* (5).

49. Millet, A.; Dailler, D.; Larini, P., *et al.*, Ligand-controlled alpha- and beta-arylation of acyclic N-Boc amines. *Angew Chem Int Ed* **2014,** *53* (10), 2678-82.

50. Zhang, J.; Park, S.; Chang, S., Catalytic Access to Bridged Sila- N-heterocycles from Piperidines via Cascade sp(3) and sp(2) C-Si Bond Formation. *J Am Chem Soc* **2018,** *140* (41), 13209-13213.

51. Oeschger, R.; Su, B.; Yu, I., *et al.*, Diverse functionalization of strong alkyl C–H bonds by undirected borylation. *Science* **2020,** *368* (6492), 736-741.

52. Griffiths, R. J.; Kong, W. C.; Richards, S. A., *et al.*, Oxidative beta-C-H sulfonylation of cyclic amines. *Chem Sci* **2018,** *9* (8), 2295-2300.

53. Yi, D.; Bayer, T.; Badenhorst, C. P. S., *et al.*, Recent trends in biocatalysis. *Chem Soc Rev* **2021,** *50* (14), 8003-8049.

54. Wu, S.; Snajdrova, R.; Moore, J. C., *et al.*, Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew Chem Int Ed* **2021,** *60* (1), 88-119.

55. Wang, Y.; Xue, P.; Cao, M., *et al.*, Directed Evolution: Methodologies and Applications. *Chem Rev* **2021,** *121* (20), 12384-12444.

56. Sheldon, R. A.; Brady, D.; Bode, M. L., The Hitchhiker's guide to biocatalysis: recent advances in the use of enzymes in organic synthesis. *Chem Sci* **2020,** *11* (10), 2587-2605.

57. Sharma, A.; Gupta, G.; Ahmad, T., *et al.*, Enzyme Engineering: Current Trends and Future Perspectives. *Food Rev Int* **2019,** *37* (2), 121-154.

58. Bornscheuer, U. T., The fourth wave of biocatalysis is approaching. *Philos Trans Royal Soc A* **2018,** *376* (2110).

59. Behrendorff, J. B.; Huang, W.; Gillam, E. M., Directed evolution of cytochrome P450 enzymes for biocatalysis: exploiting the catalytic versatility of enzymes with relaxed substrate specificity. *Biochem J* **2015,** *467* (1), 1-15.

60. Hammer, S. C.; Knight, A. M.; Arnold, F. H., Design and evolution of enzymes for non-natural chemistry. *Curr Opin Green Sustain Chem* **2017,** *7*, 23-30.

61. Strohmeier, G. A.; Pichler, H.; May, O., *et al.*, Application of designed enzymes in organic synthesis. *Chem Rev* **2011,** *111* (7), 4141-64.

62. Li, Z.; Jiang, Y.; Guengerich, F. P., *et al.*, Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. *J Biol Chem* **2020,** *295* (3), 833-849.

63. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J., *et al.*, Engineering the third wave of biocatalysis. *Nature* **2012,** *485* (7397), 185-94.

64. Rosenthaler, L. In *Durch Enzyme bewirkte asymmetrische Synthesen*.

65. Estell, D. A.; Graycar, T. P.; Wells, J. A., Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J Biol Chem* **1985,** *260* (11), 6518-6521.

66. Stemmer, W. P. C., Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **1994,** *370* (6488), 389-391.

67. Arnold, F. H.; Volkov, A. A., Directed evolution of biocatalysts. *Curr Opin Chem Biol* **1999,** *3* (1), 54-59.

68. Dong, J.; Fernandez-Fueyo, E.; Hollmann, F., *et al.*, Biocatalytic Oxidation Reactions: A Chemist's Perspective. *Angew Chem Int Ed* **2018,** *57* (30), 9238-9261.

69. Bhosale, S. H.; Rao, M. B.; Deshpande, V. V., Molecular and industrial aspects of glucose isomerase. *Microbiol Rev* **1996,** *60* (2), 280-300.

70. Sambyal, K.; Singh, R. V., Exploitation of E. coli for the production of penicillin G amidase: a tool for the synthesis of semisynthetic beta-lactam antibiotics. *J Genet Eng Biotechnol* **2021,** *19* (1), 156.

71. Ro, D. K.; Paradise, E. M.; Ouellet, M., *et al.*, Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006,** *440* (7086), 940-3.

72. Roessner, C. A.; Santander, P. J.; Scott, A. I., Multiple biosynthetic pathways for vitamin B12: variations on a central theme. *Vitam Horm* **2001,** *61*, 267-97.

73. Khan, A. G.; Eswaran, S. V., Woodward's synthesis of vitamin B12. *Resonance* **2003,** *8* (6), 8-16.

74. Whitehouse, C. J.; Bell, S. G.; Wong, L. L., P450(BM3) (CYP102A1): connecting the dots. *Chem Soc Rev* **2012,** *41* (3), 1218-60.

75. Whitehouse, C. J.; Bell, S. G.; Tufton, H. G., *et al.*, Evolved CYP102A1 (P450BM3) variants oxidise a range of non-natural substrates and offer new selectivity options. *Chem Commun (Camb)* **2008,** (8), 966-8.

76. Fasan, R., Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catal* **2012,** *2* (4), 647-666.

77. Ilia G. Denisov; Thomas M. Makris; Stephen G. Sligar, *et al.*, Structure and chemistry of cytochrome P450. *Chem Rev* **2005,** *105*, 2253-2277.

78. Nelson, D. R., Cytochrome P450 diversity in the tree of life. *Biochim Biophys Acta Proteins Proteom* **2018,** *1866* (1), 141-154.

79. Klingenberg, M., Pigments of rat liver microsomes. *Arch Biochem Biophys* **1958,** *75* (2), 376-386.

80. Omura, T.; Sato, R., The Carbon Monoxide-binding Pigment of Liver Microsomes: I. EVIDENCE FOR ITS HEMOPROTEIN NATURE. *J Biol Chem* **1964,** *239* (7), 2370-2378.

81. Omura, T.; Sato, R., The Carbon Monoxide-binding Pigment of Liver Microsomes. *J Biol Chem* **1964,** *239* (7), 2379-2385.

82. Tian, X.; Liu, W. Q.; Xu, H., *et al.*, Cell-free expression of NO synthase and P450 enzyme for the biosynthesis of an unnatural amino acid L-4-nitrotryptophan. *Synth Syst Biotechnol* **2022,** *7* (2), 775-783.

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

84. Henrot, M.; Richter, M. E.; Maddaluno, J., *et al.*, Convergent asymmetric synthesis of (+)-aureothin employing an oxygenase-mediated resolution step. *Angew Chem Int Ed* **2012,** *51* (38), 9587-91.

85. Kim, T. W.; Hwang, J. Y.; Kim, Y. S., *et al.*, Arabidopsis CYP85A2, a cytochrome P450, mediates the Baeyer-Villiger oxidation of castasterone to brassinolide in brassinosteroid biosynthesis. *Plant Cell* **2005,** *17* (8), 2397-412.

86. Werck-Reichhart, D.; Feyereisen, R., Cytochromes P450: a success story. *Genome Biol* **2000,** *1* (6), REVIEWS3003-REVIEWS3003.

87. Hannemann, F.; Bichet, A.; Ewen, K. M., *et al.*, Cytochrome P450 systems--biological variations of electron transport chains. *Biochim Biophys Acta* **2007,** *1770* (3), 330-44.

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

89. 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), 11934-11944.

90. 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.

91. Child, S. A.; Reddish, M. J.; Glass, S. M., *et al.*, Functional interactions of adrenodoxin with several human mitochondrial cytochrome P450 enzymes. *Arch Biochem Biophys* **2020,** *694*, 108596.

92. Hanukoglu, I., Electron Transfer Proteins of Cytochrome P450 Systems. In *Advances in Molecular and Cell Biology*, Bittar, E. E., Ed. JAI Press Inc.: Stamford, CT, 1996; Vol. 14, pp 29-56.

93. Smith, G. C.; Tew, D. G.; Wolf, C. R., Dissection of NADPH-cytochrome P450 oxidoreductase into distinct functional domains. *Proc Natl Acad Sci USA* **1994,** *91* (18), 8710-8714.

94. Wang, M.; Roberts, D. L.; Paschke, R., *et al.*, Three-dimensional structure of NADPH-cytochrome P450 reductase: Prototype for FMN- and FAD-containing enzymes. *Proc Natl Acad Sci USA* **1997,** *94* (16), 8411-8416.

95. Hasler, J. A.; Estabrook, R.; Murray, M., *et al.*, Human cytochromes P450. *Mol Aspects Med* **1999,** *20* (1), 1-137.

96. Hansen, C. C.; Nelson, D. R.; Moller, B. L., *et al.*, Plant cytochrome P450 plasticity and evolution. *Mol Plant* **2021,** *14* (8), 1244-1265.

97. Pandian, B. A.; Sathishraj, R.; Djanaguiraman, M., *et al.*, Role of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants* **2020,** *9* (5).

98. Hawkes, D. B.; Adams, G. W.; Burlingame, A. L., *et al.*, Cytochrome P450cin (CYP176A), Isolation, Expression, and Characterization*. *J Biol Chem* **2002,** *277*, 27725 - 27732.

99. 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.

100. McLean, M. A.; Maves, S. A.; Weiss, K. E., *et al.*, Characterization of a Cytochrome P450 from the Acidothermophilic ArchaeaSulfolobus solfataricus. *Biochem Biophys Res Commun* **1998,** *252* (1), 166-172.

101. Koo, L. S.; Tschirret-Guth, R. A.; Straub, W. E., *et al.*, The Active Site of the Thermophilic CYP119 from Sulfolobus solfataricus *. *J Biol Chem* **2000,** *275* (19), 14112-14123.

102. Jackson, C. J.; Lamb, D. C.; Marczylo, T. H., *et al.*, A Novel Sterol 14α-Demethylase/Ferredoxin Fusion Protein (MCCYP51FX) from Methylococcus capsulatusRepresents a New Class of the Cytochrome P450 Superfamily*. *J Biol Chem* **2002,** *277* (49), 46959-46965.

103. Seth-Smith, H. M. B.; Edwards, J.; Rosser, S. J., *et al.*, The Explosive-Degrading Cytochrome P450 System Is Highly Conserved among Strains of Rhodococcus spp. *Appl Environ Microbiol* **2008,** *74* (14), 4550-4552.

104. Hunter, D. J.; Roberts, G. A.; Ost, T. W., *et al.*, Analysis of the domain properties of the novel cytochrome P450 RhF. *FEBS Lett* **2005,** *579* (10), 2215-20.

105. Nodate, M.; Kubota, M.; Misawa, N., Functional expression system for cytochrome P450 genes using the reductase domain of self-sufficient P450RhF from Rhodococcus sp. NCIMB 9784. *Appl Microbiol Biotechnol* **2006,** *71* (4), 455-62.

106. Narhi, L. O.; Fulco, A. J., Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in Bacillus megaterium. *J Biol Chem* **1987,** *262* (14), 6683-6690.

107. Ost, T. W. B.; Clark, J.; Mowat, C. G., *et al.*, Oxygen Activation and Electron Transfer in Flavocytochrome P450 BM3. *J Am Chem Soc* **2003,** *125* (49), 15010-15020.

108. Takaya, N.; Suzuki, S.; Kuwazaki, S., *et al.*, Cytochrome P450nor, a Novel Class of Mitochondrial Cytochrome P450 Involved in Nitrate Respiration in the Fungus Fusarium oxysporum. *Arch Biochem Biophys* **1999,** *372* (2), 340-346.

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

110. 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* (1), 306-17.

111. Itoh, A.; Howe, G. A., Molecular Cloning of a Divinyl Ether Synthase: IDENTIFICATION AS A CYP74 CYTOCHROME P-450*. *J Biol Chem* **2001,** *276* (5), 3620-3627.

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

113. Miura, Y.; Fulco, A. J., (Omega -2) hydroxylation of fatty acids by a soluble system from bacillus megaterium. *J Biol Chem* **1974,** *249* (6), 1880-8.

114. Fulco, A. J., P450BM-3 and other Inducible Bacterial P450 Cytochromes: Biochemistry and Regulation. *Annu Rev Pharmacol Toxicol* **1991,** *31* (1), 177-203.

115. Omura, K.; Aiba, Y.; Onoda, H., *et al.*, Reconstitution of full-length P450BM3 with an artificial metal complex by utilising the transpeptidase Sortase A. *Chem Commun (Camb)* **2018,** *54* (57), 7892-7895.

116. Kitazume, T.; Haines, D. C.; Estabrook, R. W., *et al.*, Obligatory Intermolecular Electron-Transfer from FAD to FMN in Dimeric P450BM-3. *Biochemistry* **2007,** *46* (42), 11892-11901.

117. Zhang, H.; Yokom, A. L.; Cheng, S., *et al.*, 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.

118. Neeli, R.; Girvan, H. M.; Lawrence, A., *et al.*, The dimeric form of flavocytochrome P450 BM3 is catalytically functional as a fatty acid hydroxylase. *FEBS Lett* **2005,** *579* (25), 5582-8.

119. Girvan, H. M.; Dunford, A. J.; Neeli, R., *et al.*, Flavocytochrome P450 BM3 mutant W1046A is a NADH-dependent fatty acid hydroxylase: implications for the mechanism of electron transfer in the P450 BM3 dimer. *Arch Biochem Biophys* **2011,** *507* (1), 75-85.

120. Boddupalli, S. S.; Estabrook, R. W.; Peterson, J. A., Fatty acid monooxygenation by cytochrome P-450BM-3. *J Biol Chem* **1990,** *265* (8), 4233-4239.

121. Capdevila, J. H.; Wei, S.; Helvig, C., *et al.*, The highly stereoselective oxidation of polyunsaturated fatty acids by cytochrome P450BM-3. *J Biol Chem* **1996,** *271* (37), 22663-71.

122. Noble, M. A.; Miles, C. S.; Chapman, S. K., *et al.*, Roles of key active-site residues in flavocytochrome P450 BM3. *Biochem J* **1999,** *339 (Pt 2)* (Pt 2), 371-9.

123. Gotoh, O., Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **1992,** *267* (1), 83-90.

124. Sevrioukova, I. F.; Li, H.; Zhang, H., *et al.*, Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci USA* **1999,** *96* (5), 1863-8.

125. Williams, P. A.; Cosme, J.; Vinkovic, D. M., *et al.*, Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004,** *305* (5684), 683-6.

126. Butler, C. F.; Peet, C.; Mason, A. E., *et al.*, Key mutations alter the cytochrome P450 BM3 conformational landscape and remove inherent substrate bias. *J Biol Chem* **2013,** *288* (35), 25387-25399.

127. Cowart, L. A.; Falck, J. R.; Capdevila, J. H., Structural Determinants of Active Site Binding Affinity and Metabolism by Cytochrome P450 BM-3. *Arch Biochem Biophys* **2001,** *387* (1), 117-124.

128. Carmichael, A. B.; Wong, L. L., Protein engineering of Bacillus megaterium CYP102. The oxidation of polycyclic aromatic hydrocarbons. *Eur J Biochem* **2001,** *268* (10), 3117-25.

129. Rea, V.; Dragovic, S.; Boerma, J. S., *et al.*, 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.

130. Vottero, E.; Rea, V.; Lastdrager, J., *et al.*, Role of residue 87 in substrate selectivity and regioselectivity of drug-metabolizing cytochrome P450 CYP102A1 M11. *J Biol Inorg Chem* **2011,** *16* (6), 899-912.

131. Graham-Lorence, S.; Truan, G.; Peterson, J. A., *et al.*, 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.

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

133. Huang, W. C.; Westlake, A. C.; Marechal, J. D., *et al.*, Filling a hole in cytochrome P450 BM3 improves substrate binding and catalytic efficiency. *J Mol Biol* **2007,** *373* (3), 633-51.

134. Truan, G.; Peterson, J. A., Thr268 in Substrate Binding and Catalysis in P450BM-3. *Arch Biochem Biophys* **1998,** *349* (1), 53-64.

135. Yeom, H.; Sligar, S. G.; Li, H., *et al.*, The role of Thr268 in oxygen activation of cytochrome P450BM-3. *Biochemistry* **1995,** *34* (45), 14733-14740.

136. Yeom, H.; Sligar, S. G., Oxygen Activation by Cytochrome P450BM-3: Effects of Mutating an Active Site Acidic Residue. *Arch Biochem Biophys* **1997,** *337* (2), 209-216.

137. Ravichandran, K. G.; Boddupalli, S. S.; Hasermann, C. A., *et al.*, Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science* **1993,** *261* (5122), 731-6.

138. Falck, J. R.; Reddy, Y. K.; Haines, D. C., *et al.*, Practical, enantiospecific syntheses of 14,15-EET and leukotoxin B (vernolic acid). *Tetrahedron Lett* **2001,** *42* (25), 4131-4133.

139. Celik, A.; Sperandio, D.; Speight, R. E., *et al.*, Enantioselective epoxidation of linolenic acid catalysed by cytochrome P450(BM3) from Bacillus megaterium. *Org Biomol Chem* **2005,** *3* (15), 2688-90.

140. Alworth, W. L.; Xia, Q. W.; Liu, H. M., Organochlorine substrates and inhibitors of P450BM-3. *FASEB J* **1997,** *11*, A804-A804.

141. Coon, M.; McGinnity, D.; Vaz, A., *et al.*, Novel substrates for mechanistic studies with cytochrome P450 BM-3. *FASEB J* **1997,** *11*, 221.

142. Urlacher, V. B.; Makhsumkhanov, A.; Schmid, R. D., Biotransformation of beta-ionone by engineered cytochrome P450 BM-3. *Appl Microbiol Biotechnol* **2006,** *70* (1), 53-9.

143. Watanabe, Y.; Laschat, S.; Budde, M., *et al.*, Oxidation of acyclic monoterpenes by P450 BM-3 monooxygenase: influence of the substrate E/Z-isomerism on enzyme chemo- and regioselectivity. *Tetrahedron* **2007,** *63* (38), 9413-9422.

144. Kubo, T.; Peters, M. W.; Meinhold, P., *et al.*, Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450 BM-3. *Chem Eur J* **2006,** *12* (4), 1216-20.

145. Maves, S. A.; Yeom, H.; McLean, M. A., *et al.*, Decreased substrate affinity upon alteration of the substrate-docking region in cytochrome P450(BM-3). *FEBS Lett* **1997,** *414* (2), 213-8.

146. Gillam, E. M. J.; Aguinaldo, A. M. A.; Notley, L. M., *et al.*, Formation of Indigo by Recombinant Mammalian Cytochrome P450. *Biochem Biophys Res Commun* **1999,** *265* (2), 469-472.

147. Li, Q.-S.; Schwaneberg, U.; Fischer, P., *et al.*, Directed Evolution of the Fatty-Acid Hydroxylase P450 BM-3 into an Indole-Hydroxylating Catalyst. *Chem Eur J* **2000,** *6* (9), 1531-1536.

148. Li, Q. S.; Ogawa, J.; Schmid, R. D., *et al.*, Engineering cytochrome P450 BM-3 for oxidation of polycyclic aromatic hydrocarbons. *Appl Environ Microbiol* **2001,** *67* (12), 5735-9.

149. Sulistyaningdyah, W. T.; Ogawa, J.; Li, Q.-S., *et al.*, Metabolism of polychlorinated dibenzo-p-dioxins by cytochrome P450 BM-3 and its mutant. *Biotechnol Lett* **2004,** *26*, 1857-1860.

150. Schulze, H.; Schmid, R. D.; Bachmann, T. T., Activation of Phosphorothionate Pesticides Based on a Cytochrome P450 BM-3 (CYP102 A1) Mutant for Expanded Neurotoxin Detection in Food Using Acetylcholinesterase Biosensors. *Anal Chem* **2004,** *76* (6), 1720-1725.

151. Alejo-Gonzalez, K.; Hanson-Viana, E.; Vazquez-Duhalt, R., Enzymatic detoxification of organophosphorus pesticides and related toxicants. *J Pestic Sci* **2018,** *43* (1), 1-9.

152. Giver, L.; Gershenson, A.; Freskgard, P. O., *et al.*, Directed evolution of a thermostable esterase. *Proc Natl Acad Sci USA* **1998,** *95* (22), 12809-13.

153. 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.

154. Farinas, Edgardo T.; Schwaneberg, U.; Glieder, A., *et al.*, Directed Evolution of a Cytochrome P450 Monooxygenase for Alkane Oxidation. *Adv Synth Catal* **2001,** *343* (6-7), 601-606.

155. Meinhold, P.; Peters, M. W.; Chen, M. M., *et al.*, Direct conversion of ethane to ethanol by engineered cytochrome P450 BM3. *ChemBioChem* **2005,** *6* (10), 1765-8.

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

157. Fasan, R.; Meharenna, Y. T.; Snow, C. D., *et al.*, Evolutionary history of a specialized p450 propane monooxygenase. *J Mol Biol* **2008,** *383* (5), 1069-80.

158. Landwehr, M.; Hochrein, L.; Otey, C. R., *et al.*, Enantioselective alpha-hydroxylation of 2-arylacetic acid derivatives and buspirone catalyzed by engineered cytochrome P450 BM-3. *J Am Chem Soc* **2006,** *128* (18), 6058-9.

159. Sawayama, A. M.; Chen, M. M.; Kulanthaivel, P., *et al.*, A panel of cytochrome P450 BM3 variants to produce drug metabolites and diversify lead compounds. *Chem Eur J* **2009,** *15* (43), 11723-9.

160. Coelho, P. S.; Wang, Z. J.; Ener, M. E., *et al.*, A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat Chem Biol* **2013,** *9* (8), 485-7.

161. Brandenberg, O. F.; Chen, K.; Arnold, F. H., Directed Evolution of a Cytochrome P450 Carbene Transferase for Selective Functionalization of Cyclic Compounds. *J Am Chem Soc* **2019,** *141* (22), 8989-8995.

162. Ren, X.; Yorke, J. A.; Taylor, E., *et al.*, Drug Oxidation by Cytochrome P450BM3 : Metabolite Synthesis and Discovering New P450 Reaction Types. *Chem Eur J* **2015,** *21* (42), 15039-47.

163. 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.

164. Whitehouse, C. J.; Bell, S. G.; Yang, W., *et al.*, A highly active single-mutation variant of P450BM3 (CYP102A1). *ChemBioChem* **2009,** *10* (10), 1654-6.

165. Whitehouse, C. J.; Yang, W.; Yorke, J. A., *et al.*, Structural basis for the properties of two single-site proline mutants of CYP102A1 (P450BM3). *ChemBioChem* **2010,** *11* (18), 2549-56.

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

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

168. Kille, S.; Zilly, F. E.; Acevedo, J. P., *et al.*, Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat Chem* **2011,** *3* (9), 738-43.

169. Hecht, M.; Bromberg, Y.; Rost, B., News from the protein mutability landscape. *J Mol Biol* **2013,** *425* (21), 3937-48.

170. van der Meer, J. Y.; Biewenga, L.; Poelarends, G. J., The Generation and Exploitation of Protein Mutability Landscapes for Enzyme Engineering. *ChemBioChem* **2016,** *17* (19), 1792-1799.

171. Acevedo-Rocha, C. G.; Gamble, C. G.; Lonsdale, R., *et al.*, P450-Catalyzed Regio- and Diastereoselective Steroid Hydroxylation: Efficient Directed Evolution Enabled by Mutability Landscaping. *ACS Catal* **2018,** *8* (4), 3395-3410.

172. Li, A.; Acevedo-Rocha, C. G.; D'Amore, L., *et al.*, Regio- and Stereoselective Steroid Hydroxylation at C7 by Cytochrome P450 Monooxygenase Mutants. *Angew Chem Int Ed* **2020,** *59* (30), 12499-12505.

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

174. Whitehouse, C. J. C.; Yang, W.; Yorke, J. A., *et al.*, Structure, electronic properties and catalytic behaviour of an activity-enhancing CYP102A1 (P450BM3) variant. *Dalton Trans* **2011,** *40* (40), 10383-10396.

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

176. Chen, W.; Fisher, M. J.; Leung, A., *et al.*, Oxidative Diversification of Steroids by Nature-Inspired Scanning Glycine Mutagenesis of P450BM3 (CYP102A1). *ACS Catal* **2020,** *10* (15), 8334-8343.

177. Zhang, K.; Shafer, B. M.; Demars, M. D., 2nd, *et al.*, 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.

178. 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.

179. Braunegg, G.; de Raadt, A.; Feichtenhofer, S., *et al.*, The Concept of Docking/Protecting Groups in Biohydroxylation. *Angew Chem Int Ed* **1999,** *38* (18), 2763-2766.

180. Munzer, D. F.; Meinhold, P.; Peters, M. W., *et al.*, Stereoselective hydroxylation of an achiral cyclopentanecarboxylic acid derivative using engineered P450s BM-3. *Chem Commun (Camb)* **2005,** (20), 2597-9.

181. Vickers, C.; Backfisch, G.; Oellien, F., *et al.*, Enzymatic Late-Stage Oxidation of Lead Compounds with Solubilizing Biomimetic Docking/Protecting groups. *Chem Eur J* **2018,** *24* (68), 17936-17947.

182. Le-Huu, P.; Rekow, D.; Kruger, C., *et al.*, Chemoenzymatic Route to Oxyfunctionalized Cembranoids Facilitated by Substrate and Protein Engineering. *Chem Eur J* **2018,** *24* (46), 12010-12021.

183. Karplus, M., Molecular Dynamics Simulations of Biomolecules. *Acc Chem Res* **2002,** *35* (6), 321-323.

184. Hollingsworth, S. A.; Dror, R. O., Molecular Dynamics Simulation for All. *Neuron* **2018,** *99* (6), 1129-1143.

185. Alder, B. J.; Wainwright, T. E., Phase Transition for a Hard Sphere System. *J Chem Phys* **1957,** *27* (5), 1208-1209.

186. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977,** *267* (5612), 585-590.

187. Hospital, A.; Goni, J. R.; Orozco, M., *et al.*, Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem* **2015,** *8*, 37-47.

188. Lazaridis, T.; Karplus, M., Effective energy function for proteins in solution. *Proteins* **1999,** *35* (2), 133-152.

189. Kawata, M.; Nagashima, U., Particle mesh Ewald method for three-dimensional systems with two-dimensional periodicity. *Chem Phys Lett* **2001,** *340* (1), 165-172.

190. Orozco, M.; Luque, F. J., Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems. *Chem Rev* **2000,** *100* (11), 4187-4226.

191. Roux, B. t.; Simonson, T., Implicit solvent models. *Biophys Chem* **1999,** *78* (1), 1-20.

192. Bashford, D.; Case, D. A., Generalized born models of macromolecular solvation effects. *Annu Rev Phys Chem* **2000,** *51*, 129-52.

193. Hermans, J.; Berendsen, H. J. C.; Van Gunsteren, W. F., *et al.*, A consistent empirical potential for water–protein interactions. *Biopolymers* **1984,** *23* (8), 1513-1518.

194. Ott, K.-H.; Meyer, B., Parametrization of GROMOS force field for oligosaccharides and assessment of efficiency of molecular dynamics simulations. *J Comput Chem* **1996,** *17* (8), 1068-1084.

195. MacKerell, A. D.; Bashford, D.; Bellott, M., *et al.*, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **1998,** *102* (18), 3586-3616.

196. Cornell, W. D.; Cieplak, P.; Bayly, C. I., *et al.*, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **1995,** *117* (19), 5179-5197.

197. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J., *et al.*, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J Phys Chem B* **2001,** *105* (28), 6474-6487.

198. Moffett, A. S.; Shukla, D., Using molecular simulation to explore the nanoscale dynamics of the plant kinome. *Biochem J* **2018,** *475* (5), 905-921.

199. Hess, B., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **2008,** *4*, 435-447.

200. Phillips, J. C.; Braun, R.; Wang, W., *et al.*, Scalable molecular dynamics with NAMD. *J Comput Chem* **2005,** *26* (16), 1781-802.

201. Case, D.; Cerutti, D. S.; Cheatham, T., *et al.*, *Amber 2017, University of California, San Francisco*. 2017.

202. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr., *et al.*, CHARMM: the biomolecular simulation program. *J Comput Chem* **2009,** *30* (10), 1545-614.

203. Eastman, P.; Swails, J.; Chodera, J. D., *et al.*, OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **2017,** *13* (7), e1005659.

204. McConkey, B.; Sobolev, V.; Edelman, M., The performance of current methods in ligand-protein docking. *Curr Sci* **2002,** *83*.

205. Meng, X.-Y.; Zhang, H.-X.; Mezei, M., *et al.*, Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des* **2011,** *7* (2), 146-157.

206. Norel, R.; Fischer, D.; Wolfson, H. J., *et al.*, Molecular surface recognition by a computer vision-based technique. *Protein Eng* **1994,** *7* (1), 39-46.

207. Fischer, D.; Norel, R.; Wolfson, H., *et al.*, Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition. *Proteins* **1993,** *16* (3), 278-92.

208. DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S., *et al.*, Docking flexible ligands to macromolecular receptors by molecular shape. *J Med Chem* **1986,** *29* (11), 2149-2153.

209. Leach, A. R.; Kuntz, I. D., Conformational analysis of flexible ligands in macromolecular receptor sites. *J Comput Chem* **1992,** *13* (6), 730-748.

210. Eisen, M. B.; Wiley, D. C.; Karplus, M., *et al.*, HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins* **1994,** *19* (3), 199-221.

211. Böhm, H.-J., LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* **1992,** *6* (6), 593-606.

212. Böhm, H. J., The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J Comput Aided Mol Des* **1992,** *6* (1), 61-78.

213. Hart, T. N.; Read, R. J., A multiple-start Monte Carlo docking method. *Proteins* **1992,** *13* (3), 206-22.

214. Goodsell, D. S.; Lauble, H.; Stout, C. D., *et al.*, Automated docking in crystallography: Analysis of the substrates of aconitase. *Proteins* **1993,** *17* (1), 1-10.

215. Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S., Flexible ligand docking using a genetic algorithm. *J Comput Aided Mol Des* **1995,** *9* (2), 113-30.

216. Morris, G. M.; Goodsell, D. S.; Halliday, R. S., *et al.*, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* **1998,** *19* (14), 1639-1662.

217. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J., *et al.*, Improved protein–ligand docking using GOLD. *Proteins* **2003,** *52* (4), 609-623.

218. Clark, K. P.; Ajay, Flexible ligand docking without parameter adjustment across four ligand–receptor complexes. *J Comput Chem* **1995,** *16* (10), 1210-1226.

219. Taylor, J. S.; Burnett, R. M., DARWIN: A program for docking flexible molecules. *Proteins* **2000,** *41* (2), 173-191.

220. 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.

221. Ravindranath, P. A.; Forli, S.; Goodsell, D. S., *et al.*, AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLoS Comput Biol* **2015,** *11* (12), e1004586.

222. Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O., Ligand Binding Affinities from MD Simulations. *Acc Chem Res* **2002,** *35* (6), 358-365.

223. Carlson, H. A.; Jorgensen, W. L., An Extended Linear Response Method for Determining Free Energies of Hydration. *J Phys Chem* **1995,** *99* (26), 10667-10673.

224. Jain, A. N., Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J Comput Aided Mol Des* **1996,** *10* (5), 427-440.

225. Head, R. D.; Smythe, M. L.; Oprea, T. I., *et al.*, VALIDATE: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands. *J Am Chem Soc* **1996,** *118* (16), 3959-3969.

226. Muegge, I.; Martin, Y. C., A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J Med Chem* **1999,** *42* (5), 791-804.

227. Ishchenko, A. V.; Shakhnovich, E. I., SMall Molecule Growth 2001 (SMoG2001): An Improved Knowledge-Based Scoring Function for Protein−Ligand Interactions. *J Med Chem* **2002,** *45* (13), 2770-2780.

228. Feher, M.; Deretey, E.; Roy, S., BHB: A Simple Knowledge-Based Scoring Function to Improve the Efficiency of Database Screening. *J Chem Inf Comput Sci* **2003,** *43* (4), 1316-1327.

229. Verkhivker, G.; Appelt, K.; Freer, S. T., *et al.*, Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng* **1995,** *8* (7), 677-91.

230. Ewing, T. J.; Makino, S.; Skillman, A. G., *et al.*, DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **2001,** *15* (5), 411-28.

231. Miller, M. D.; Kearsley, S. K.; Underwood, D. J., *et al.*, FLOG: A system to select@ quasi-flexibl… ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des* **1994,** *8*, 153-174.

232. Gabb, H. A.; Jackson, R. M.; Sternberg, M. J., Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* **1997,** *272* (1), 106-120.

233. Paulsen, M. D.; Ornstein, R. L., Dramatic differences in the motions of the mouth of open and closed cytochrome P450BM-3 by molecular dynamics simulations. *Proteins* **1995,** *21* (3), 237-243.

234. Pleiss, J., Systematic Analysis of Large Enzyme Families: Identification of Specificity- and Selectivity-Determining Hotspots. *ChemCatChem* **2014,** *6* (4), 944-950.

235. Dodani, S. C.; Kiss, G.; Cahn, J. K., *et al.*, Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. *Nat Chem* **2016,** *8* (5), 419-25.

236. Dodani, S. C.; Cahn, J. K.; Heinisch, T., *et al.*, Structural, functional, and spectroscopic characterization of the substrate scope of the novel nitrating cytochrome P450 TxtE. *ChemBioChem* **2014,** *15* (15), 2259-67.

237. Ebert, M. C. C. J. C.; Guzman Espinola, J.; Lamoureux, G., *et al.*, Substrate-Specific Screening for Mutational Hotspots Using Biased Molecular Dynamics Simulations. *ACS Catal* **2017,** *7* (10), 6786-6797.

238. Peng, Y.; Gao, C.; Zhang, Z., *et al.*, A Chemoenzymatic Strategy for the Synthesis of Steroid Drugs Enabled by P450 Monooxygenase-Mediated Steroidal Core Modification. *ACS Catal* **2022,** *12* (5), 2907-2914.

239. Taylor, R. D.; MacCoss, M.; Lawson, A. D., Rings in drugs. *J Med Chem* **2014,** *57* (14), 5845-59.

240. Wijdeven, M. A.; Willemsen, J.; Rutjes, F. P. J. T., The 3-Hydroxypiperidine Skeleton: Key Element in Natural Product Synthesis. *Eur J Org Chem* **2010,** *2010* (15), 2831-2844.

241. Abraham, M. J.; Murtola, T.; Schulz, R., *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

242. Knapp, B.; Ospina, L.; Deane, C. M., Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J Chem Theory Comput* **2018,** *14* (12), 6127-6138.

243. Saruengkhanphasit, R.; Collier, D.; Coldham, I., Synthesis of Spirocyclic Amines by Using Dipolar Cycloadditions of Nitrones. *J Org Chem* **2017,** *82* (12), 6489-6496.

244. Lovering, F.; Bikker, J.; Humblet, C., Escape from flatland: increasing saturation as an approach to improving clinical success. *J Med Chem* **2009,** *52* (21), 6752-6.

245. Olivo, H. F.; Hemenway, M. S., Total Synthesis of (±)-Epibatidine Using a Biocatalytic Approach. *J Org Chem* **1999,** *64* (24), 8968-8969.

246. Hemenway, M. S.; Olivo, H. F., Syntheses of New Phosphorus-Containing Azabicycloalkanes and Their Microbial Hydroxylation Using Beauveria bassiana. *J Org Chem* **1999,** *64* (17), 6312-6318.

247. Davis, C. R.; Johnson, R. A.; Cialdella, J. I., *et al.*, Microbiological Oxygenation of Bridgehead Azabicycloalkanes. *J Org Chem* **1997,** *62* (7), 2244-2251.

248. Stephen R. Fletcher; Raymond Baker; Mark S. Chambers, *et al.*, total synthesis and determination of the absolute configuration of epibatidine. *J Org Chem* **1994,** *59*, 1771-1778.

249. Albertini, E.; Barco, A.; Benetti, S., *et al.*, Enantioselective approach to 7-azabicyclo[2.2.1]heptane ring systems using D-(−)-quinic acid as the chiral educt: Application to the formal synthesis of (+)-epibatidine. *Tetrahedron Lett* **1997,** *38* (4), 681-684.

250. Hoye, T. R.; Jeffrey, C. S.; Shao, F., Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat Protoc* **2007,** *2* (10), 2451-8.

251. Palatinus, L.; Chapuis, G., SUPERFLIP - a computer program for the solution of crystal structures by charge flipping in arbitrary dimensions. *J Appl Crystallogr* **2007,** *40*, 786-790.

252. Betteridge, P. W.; Carruthers, J. R.; Cooper, R. I., *et al.*, CRYSTALS version 12: software for guided crystal structure analysis. *J Appl Crystallogr* **2003,** *36*, 1487-1487.

253. Cooper, R. I.; Thompson, A. L.; Watkin, D. J., CRYSTALS enhancements: dealing with hydrogen atoms in refinement. *J Appl Crystallogr* **2010,** *43*, 1100-1107.

254. Flack, H. D., On Enantiomorph-Polarity Estimation. *Acta Crystallogr A* **1983,** *39* (Nov), 876-881.

255. Flack, H. D.; Bernardinelli, G., Reporting and evaluating absolute-structure and absolute-configuration determinations. *J Appl Crystallogr* **2000,** *33*, 1143-1148.

256. Hooft, R. W. W.; Straver, L. H.; Spek, A. L., Determination of absolute structure using Bayesian statistics on Bijvoet differences. *J Appl Crystallogr* **2008,** *41*, 96-103.

257. Hoye, T. R.; Jeffrey, C. S.; Shao, F., Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat Protoc* **2007,** *2* (10), 2451-2458.

258. Haines, D. C.; Tomchick, D. R.; Machius, M., *et al.*, Pivotal Role of Water in the Mechanism of P450BM3. *Biochemistry* **2001,** *40* (40), 13456-13465.

259. Shahrokh, K.; Orendt, A.; Yost, G. S., *et al.*, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J Comput Chem* **2012,** *33* (2), 119-33.

260. Lindorff-Larsen, K.; Piana, S.; Palmo, K., *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78* (8), 1950-8.

261. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D., *et al.*, Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **1983,** *79* (2), 926-935.

262. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald—An N-Log(N) method for Ewal sums in large systems. *J Chem Phys* **1993,** *98* (98), 10089-10092.

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

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

265. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **1981,** *52* (12), 7182-7190.

266. Daura, X.; Gademann, K.; Jaun, B., *et al.*, Peptide Folding When Simulation Meets Experiment. *Angew Chem Int Ed* **1999,** *38*, 236-240.