From drug design to drug metabolism and biocatalysis: exploring the realm of proteins for pharmaceutical applications

Paolo Schiavini

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Chemistry,
McGill University
Montréal, Québec, Canada
November 2016

Abstract

Being implicated in all aspects of life, proteins are at the center of pharmaceutical development. Not only do they represent the most common pharmacological target, but they also take part in several processes that regulate drug absorption, distribution, metabolism and secretion. In addition to that, other fields of science and technology, e.g. material chemistry or catalysis, are seeing an exciting growth in the application of proteins.

Despite the tremendous progress of the last decades, many efforts are still being directed towards a better understanding of the way these macromolecules work at the atomic level.

The relationship between structure and function of proteins remains an important subject of investigations from research groups all around the world. Advances in the methodologies for the determination of the three dimensional structure of proteins have contributed enormously to the field, but they still cannot provide scientists with all the desired answers. In fact, proteins are dynamic entities whose properties change depending on several variables including temperature, pH, presence of ions, type of solvent and interaction with other molecules. As a result, much emphasis has been placed on the development of computational tools to simulate protein systems and predict their behaviour based on a set of given conditions. In the end, full understanding of a biochemical process is often achievable only through a combination of approaches at the interface of chemistry, physics, biochemistry and computer science.

This doctorate thesis describes a series of studies where diverse strategies, from the use of small molecule as probes to investigate conformational dynamics, to computer-assisted protein engineering, were used to investigate different protein systems. Organic synthesis, an array of analytical procedures, computational tools and site-directed mutagenesis include some of the many techniques that were combined to tackle projects relevant to the pharmaceutical field.

In chapter 1, the topic of protein science is introduced. Protein structure, dynamics and protein engineering are discussed. Two types of protein families, kainate receptor (KAR) and cytochrome P450 (P450s), are presented.

In chapter 2, a study on kainate receptor GluK2, a glutamate-type receptor involved in neurodegenerative diseases, is described. Analogues of L-glutamate were designed and used as molecular probes to investigate the conformational flexibility of the receptor. The findings

shed light on the importance of a tyrosine residue, in the receptor binding site, in inducing receptor inhibition.

In chapter 3, the focus is shifted to cytochromes P450s and their potential for biocatalytic application. In particular, P450 3A4, a human enzyme responsible for metabolizing > 50 % of current drugs, was chosen as a starting point for the development of a versatile biocatalyst for hydroxylation reactions. Computer-aided protein engineering was used in the attempt to develop a chemo-enzymatic synthesis for the preparation of latanaprost, a prostaglandin analogue widely used for the treatment of glaucoma.

In chapter 4, a semi-rational enzyme engineering approach was used to investigate the effect of "active site crowding" of P450 3A4 on its regio- and stereoselectivity, with the final goal of developing a biocatalyst for the preparation of useful hydroxylated products. General guidelines for the remodelling of the enzyme active site were established, and a biochemical assay for the screening of several mutants was developed. The strategy was applied for the total synthesis of (*R*)-lisofylline, a compound under investigation for its antinflammatory properties, and the functionalization of the steroid sex hormone progesterone.

In chapter 5, a complex case of P450-mediated metabolism is presented. A series of prolyl oligopeptidase (POP) inhibitors were found to undergo a series of unexpected chemical transformation following enzymatic sulphur oxidation.
Through a combination of LC–MS analysis, synthetic work, deuterium exchange studies, and computational predictions, a mixture of interconverting stereoisomers was characterized.

Résumé

Les protéines, impliquées dans tous les aspects de la vie, sont au centre du développement pharmaceutique. Elles ne sont pas seulement la principale cible pharmacologique, mais elles participent aussi à plusieurs processus qui régulent l'absorption, la distribution, le métabolisme and la sécrétion des médicaments. De plus, d'autres domaines de la science et des technologies, pour exemple la chimie des matériaux et la catalyse, voient maintenant une croissance remarquable facilitée en partie par l'utilisation des protéines.

Malgré le progrès des dernières décennies, chimistes at biochimistes aspirent toujours à mieux comprendre le fonctionnement de ces macromolécules au niveau moléculaire. La relation entre la structure and la fonction des protéines reste un sujet important des investigations de plusieurs groupes de recherche dans le monde. Les avancées dans la méthodologie pour la détermination de la structure tridimensionnelle des protéines ont beaucoup contribué au domaine, mais nombreuses questions restent sans réponse. En fait, les protéines sont des entités dynamiques avec des propriétés qui changent à l'effet de plusieurs variables comme la température, le pH, la présence d'ions, le type de solvant et l'interaction avec d'autre molécules. Par conséquent, beaucoup d'efforts sont actuellement dédiés au développement d'outils de calcul pour simuler et prévoir *in silico* le comportement des protéines. En effet, la compréhension complète d'un processus biochimique est possible seulement à travers une combinaison d'approches à l'interface de la chimie, la physique, la biochimie et les sciences informatiques.

Cette thèse de doctorat décrit une série d'études de différentes protéines. Pour ce faire, une variété de stratégies comprenant l'utilisation de petites molécules comme sondes pour enquêter la dynamique conformationnelle, ou l'ingénierie des protéines aidée par l'informatique, ont été utilisées. La synthèse organique, un assortiment de procédures analytiques, la chimie informatique et la mutagénèse contrôlée sont quelques unes des techniques qui ont été combinées pour faire face à des projets pertinents dans le domaine pharmaceutique.

Dans le chapitre 1, le sujet de la science des protéines est introduit. La structure, la dynamique et l'ingénierie des protéines sont discutés. Deux familles de protéines, le récepteurs kaïnate (KAR) et les cytochromes P450 (P450s) sont présentés.

Dans le chapitre 2, GluK2, un récepteur kaïnate impliqué dans les maladies neurodégénératives, est étudié. Des analogues de L-glutamate ont été conçus et utilisés comme sondes moléculaires pour examiner la flexibilité conformationnelle du récepteur. Les

découvertes ont révélé l'importance d'une tyrosine dans le site actif pour l'inhibition du récepteur.

Dans le chapitre 3, l'attention est tournée vers les cytochromes P450s et leur potentiel dans des applications biocatalytiques. En particulier, P450 3A4, une enzyme humaine responsable du métabolisme de plus de 50% des médicaments sur le marché, a été choisi comme point de départ pour le développement d'un biocatalyseur de réactions d'hydroxylation. L'ingénierie de protéines a été utilisé pour développer une synthèse chimio-enzymatique du latanoprost, un analogue des prostaglandines communément utilisé pour le traitement du glaucome.

Dans le chapitre 4, une stratégie semi-rationnelle pour l'ingénierie des protéines a été utilisée pour examiner l'effet de « l'encombrement du site actif » de P450 3A4 sur sa régio- et stéreosélectivité, afin de développer un biocatalyseur pour la production d'intéressants composés hydroxylés.
Des lignes directrices générales ont été établies pour le criblage de mutants. La stratégie a été appliquée à la synthèse de (*R*)-lisofylline, un composé étudié pour ses propriétés antiinflammatoires, et pour la fonctionnalisation de la progestérone, une hormone sexuelle stéroïdienne.

Dans le chapitre 5, une étude du métabolisme d'inhibiteurs de la prolyl oligopeptidase (POP) est décrite. Les investigations ont révélé une chaine de transformations chimiques inattendues à la suite de l'oxydation enzymatique d'un atome de soufre catalysé par les P450s humaines. En combinant des analyses de chromatographie couplée à la spectrométrie de masse, de la synthèse organique, des études d'échange hydrogène/deutérium et des modélisations moléculaires, un mélange de stéréoisomères en équilibre a été caractérisé.

Acknowledgements

I would like to thank my PhD supervisors, Karine Auclair and Nicolas Moitessier, for their incredible support and guidance throughout these past five years.

A special thanks to all current and previous members of the Moitessier and Auclair group for teaching me all sorts of chemistry tricks, for the useful discussions, for sharing with me the joy and sorrow of this challenging yet rewarding life experience.

I dedicate this personal achievement to my family whose love and support go beyond any person could ask for. Thanks to my parents, who care only for my happiness and never clipped my wings, and to my brothers, as together we are one.

TABLE OF CONTENTS

Abstract iii
Résumé iv
Acknowledgements vi
Table of Contents vii
List of Figures xi
List of Tables xv
Abbreviations xvi

CHAPTER 1 - PROTEIN STRUCTURE, DYNAMICS AND TWO RELEVANT PROTEIN FAMILIES 1

1.1 THE ROLE OF PROTEINS IN HUMAN SOCIETY 1
1.2 A LOOK INTO THE STRUCTURE OF PROTEINS 2
1.3 PROTEIN-DRUG INTERACTIONS 2
1.4 PROTEINS IN DRUG METABOLISM 5
1.5 PROTEINS IN BIOCATALYSIS 8
1.5.1 Applications of biocatalysis in the pharmaceutical industry 9
1.6 PROTEIN ENGINEERING 11
1.6.1 Directed evolution 12
1.6.2 Rational design 13
1.7 KAINATE RECEPTORS, A SUBTYPE OF GLUTAMATE-GATED CHANNELS 14
1.7.1 Conformational flexibility of KARs 16
1.8 CYTOCHROME P450 ENZYMES 19
1.8.1 Introduction to cytochrome P450 enzymes 19
1.8.2 P450 reaction mechanism 20
1.8.3 Industrial applications of P450s 22
1.8.4 Engineering of bacterial P450s 23
1.8.5 Substrate promiscuity of human P450s 24
1.9 P450 3A4 25
1.9.1 Mutagenesis of P450 3A4 28
1.10 MOLECULAR FORECASTER 29
1.11 OBJECTIVES OF THIS THESIS 30

CHAPTER 2 - USE OF SMALL MOLECULES TO PROBE THE CONFORMATIONAL FLEXIBILITY OF KAINATE RECEPTOR GLUK2 31

2.1 PREFACE 31
2.2 INTRODUCTION 32
2.3 RATIONAL DESIGN OF MOLECULAR PROBES 34
2.4 SYNTHESIS OF COMPOUNDS **2.4**-**2.8** 35
2.5 BIOLOGICAL STUDIES 36
2.6 DISCUSSION 39
2.7 PROPOSED FUTURE WORK 40
2.8 CONCLUSIONS 41

CHAPTER 3 - TOWARDS THE RATIONAL ENGINEERING OF P450 3A4 FOR THE CHEMO-ENZYMATIC SYNTHESIS OF LATANOPROST 42

3.1 PREFACE 42
3.2 AN INTERESTING SYNTHETIC TARGET: LATANOPROST 43
3.3 CURRENT SYNTHESES OF LATANOPROST 43
3.4 HYPOTHESES ON THE USE OF P450S FOR THE HYDROXYLATION OF A LATANOPROST PRECURSOR 45
3.5 SYNTHESIS OF COMPOUNDS **3.19** 47
3.6 TRANSFORMATION OF **3.19** BY WILD-TYPE P450 3A4 48
3.7 SELECTION OF P450 3A4 MUTATIONS 51
3.8 BIOCHEMICAL ASSAY 53
3.9 DISCUSSION 54
3.10 CONCLUSIONS 55

CHAPTER 4 - ACTIVE SITE CROWDING OF P450 3A4 AS A STRATEGY TO ALTER ITS SELECTIVITY 56

4.1 PREFACE 56
4.2 INTRODUCTION 57
4.3 DEVELOPMENT OF GUIDELINES FOR ACTIVE SITE CROWDING 57
4.4 OXIDATION OF A PRECURSOR OF (*R*)-LISOFYLLINE 56
4.5 OXIDATION OF COMPOUND **4.7** 65
4.6 OXIDATION OF PROGESTERONE 68

4.7 CONCLUSIONS 69

CHAPTER 5 - P450-MEDIATED METABOLIC INSTABILITY OF INHIBITORS OF PROLYL OLIGOPEPTIDASE 70

5.1 PREFACE 70
5.2 METABOLISM STUDIES OF TWO INHIBITORS OF POP 71
5.3 INVESTIGATING SULFUR OXIDATION 72
5.4 STRUCTURE DETERMINATION OF **5.3**-**5.9** 73
5.5 LC-MS/MS FRAGMENTATION STUDIES 74
5.6 EQUILIBRATION STUDIES 76
5.7 TIME-BASED METABOLISM STUDIES 79
5.8 A CLOSER LOOK AT THE SYNTHETIC MIXTURES 80
5.9 DEUTERIUM STUDIES 81
5.10 SITE OF METABOLISM PREDICTION AND RESOLUTION OF THE PROBLEM 83
5.11 GROUND STATE ENERGIES 84
5.12 CONCLUSIONS 85
5.13 ADDITIONAL CONTRIBUTIONS TO THE STUDY OF THE METABOLISM OF POP INHIBITORS..... 86

CHAPTER 6 - CONTRIBUTIONS AND FUTURE DIRECTIONS 87

6.1 STUDY OF THE CONFORMATIONAL FLEXIBILITY OF KARS 87
6.2 RATIONAL ENGINEERING OF P450S FOR BIOCATALYTIC APPLICATIONS 88
6.3 METABOLISM OF POP INHIBITORS AND P450-CATALYZED SULFOXIDATION REACTIONS.. 90
6.4 PUBLICATIONS 91

CHAPTER 7 - EXPERIMENTAL METHODS 92

7.1 INSTRUMENTS 92
7.2 CHEMICALS 92
7.3 SYNTHETIC PROCEDURES 92
7.4 BIOLOGICAL STUDIES 102
7.4.1 Site-directed mutagenesis and DNA primers 102

7.4.2 Expression and purification of P450 enzymes *104*
7.4.3 P450 enzyme assay using purified enzyme *104*
7.4.4 P450 enzyme assay with cell lysate *105*
7.4.5 Metabolism studies with human liver microsomes (HLM) *105*
7.4.6 Testosterone assay for HLM activity *105*
7.4.7 Measurement of intrinsic clearance Cl_{int} *106*
7.4.8 Initial velocities of P450-catalyzed reactions *106*
*7.4.9 Equilibration studies for **5.3**-**5.9*** *106*
*7.4.10 Deuterium exchange studies for **5.8** and **5.9*** *107*
7.5 LC-MS METHODS 108
*7.5.1 Oxidation of **3.19**, latanoprost precursor* *108*
*7.5.2 Oxidation of **4.1**, deoxy-precursor of (R)-lisofylline* *108*
7.5.3 Oxidation of testosterone for HLM activity assay *108*
7.5.4 Oxidation of progesterone *109*
*7.5.5 Metabolism of **5.1** and **5.2*** *109*
*7.5.6 Deuterium exchange studies for **5.8** and **5.9*** *109*
*7.5.7 Metabolism of **5.11*** *110*
7.5.8 Protein LC-MS *110*
7.6 COMPUTUTATIONAL STUDIES 110
7.6.1 Flexible docking for KAR *110*
7.6.2 Prediction of transition states for mutant P450s *111*

REFERENCES **112**

APPENDIX **123**

LIST OF FIGURES

Figure 1.1 Overview of current and predicted enzyme market 2

Figure 1.2 Representation of the hierarchical structure of proteins 3

Figure 1.3 Random high-throughput screening vs rational drug design 4

Figure 1.4 Examples of drug metabolism 6

Figure 1.5 Predicted sites of metabolism of diazepam 7

Figure 1.6 12 principles of green chemistry 8

Figure 1.7 Synthesis of (*R*)-epichlorohydrine 9

Figure 1.8 Synthesis of sitagliptin 9

Figure 1.9 Synthesis of montelukast 10

Figure 1.10 Directed evolution cycle 12

Figure 1.11 Example of how rational design can be used for protein engineering 14

Figure 1.12 Glutamate receptor subunit structure 15

Figure 1.13 Accepted model for the structural basis of KAR activation 17

Figure 1.14 Flexible docking 18

Figure 1.15 Examples of human P450-catalyzed reactions 19

Figure 1.16 Schematic organization of mammalian P450 systems 20

Figure 1.17 Catalytic cycle of cytochrome P450 21

Figure 1.18 Example of P450 applications in industrial processes 23

Figure 1.19 Relative importance of P450s in drug metabolism 24

Figure 1.20 Structural similarities among P450s from different domains of life 25

Figure 1.21 P450 3A4 structural features 26

Figure 1.22 Chemical diversity of P450 3A4 substrates 27

Figure 1.23 Active site of P450 3A4 (PDB 3NXU) 28

Figure 2.1 Different conformational states of GluK2 ABD 33

Figure 2.2 List of compounds designed to study the conformational flexibility of KARs. 34

Figure 2.3 (3*S*)-β-substituted-glutamates scaffold 35

Figure 2.4 Synthesis of **2.4**-**2.8** 35

Figure 2.5 Proposed reaction of cyclization for β-substituted glutamic acids 36

Figure 2.6 Schematic representation of the electrophysological assay used to study the antagonist behaviour of compounds **2.4**-**2.8** 37

Figure 2.7 Electrophysiological assays for the study of the agonist/antagonist behaviour of **2.4**-**2.8** towards GluK2 36

Figure 2.8 Docking of previously reported antagonists of GluK1 36

Figure 3.1 Structure of latanoprost (**3.1**) 43

Figure 3.2 Synthesis of latanoprost by Prevost *et al.*[2] 44

Figure 3.3 Structures of possible precursors of latanoprost in the proposed P450-based chemo-enzymatic synthesis 45

Figure 3.4 Proposed chemo-enzymatic synthesis of latanoprost 46

Figure 3.5 Synthesis of compound **3.19** (precursor to enzymatic oxidation) 47

Figure 3.6 Oxidation of **3.19** by P450 3A4. 48

Figure 3.7 Analysis of the products resulting from the transformation of **3.19** by P450 3A4. 49

Figure 3.8 Unlikely transformation of **3.19** by P450 3A4 wild-type 50

Figure 3.9 Suggested process explaining the observed m/z = 287 associated with the HPLC peak for compound **3.27** 50

Figure 3.10 Modelling carried out to guide the mutagenesis of P450 3A4 to favor formation of **3.24** over **3.27** 52

Figure 3.11 Schematic representation of the biochemical assay developed for the expression and testing of P450 3A4 mutants 53

Figure 4.1 Schematic representation of "active site crowding" 58

Figure 4.2 Proposed chemo-enzymatic synthesis of (*R*)-lisofylline 59

Figure 4.3 Comparison between regio- and steroselectivity of wild-type and A370I. 64

Figure 4.4 Structure of **4.7** 65

Figure 4.5 Computer modelling of increasing P450 3A4 active site crowding 66

Figure 4.6 Selectivity change, as a result of active site crowding, in the oxidation of **4.7** by P450 wild-type and mutants I369F/L482F, I301F/I369F/L482F, and L211F/I301F/I369F/G481W/L482F 67

Figure 4.7 Oxidation of progesterone (**4.10**) by P450 3A4 wild-type and mutants I369F/L482F, I301F/I369F/L482F, and L211F/I301F/I369F/G481W/L482F 68

Figure 5.1 Metabolism studies of **5.1** and **5.2**. 71

Figure 5.2 Synthesis of **5.3**-**5.9** by chemical oxidation of **5.1** and **5.2** with ammonium persulfate and comparison with HLM microsomal oxidation. 72

Figure 5.3 Schematic representation of the strategy used to assign structures to **5.3**-**5.9** 73

Figure 5.4 Representative fragmentation pattern shared by compounds **5.3**–**5.9** 74

Figure 5.5 Proposed structures and fragmentation mechanisms for the main fragments observed during LC-MS/MS analysis of **5.3**-**5.9** 75

Figure 5.6 Equilibration of each of compounds **5.3**-**5.9** separately when incubated in PKi at 37˚C over time. 76

Figure 5.7 Proposed mechanisms of epimerization for **5.3**-**5.9** 77

Figure 5.8 Equilibration studies for **5.1** and **5.2** 78

Figure 5.9 Time-based metabolism studies of **5.1** and **5.2** 79

Figure 5.10 Equilibration profiles of **5.9** and **5.8** when isolated and incubated in PKi for 1h 80

Figure 5.11 Hypotheses on deuterium exchange studies 81

Figure 5.12 H/D exchange studies 82

Figure 5.13 IMPACTS metabolism predictions for **5.1** and **5.2**. 83

Figure 5.14 Calculated energies for **5.3**-**5.10** (kcal) 84

Figure 5.15 Structure and potency vs POP of **5.11** 86

Figure 5.16 LC-MS analysis of metabolism of **5.11** 86

Figure 6.1 Suggested optimization strategy to increase crossing of the blood brain barrier ... 88

Figure 6.2 Suggested strategies to further improve active site crowding of P450 3A4 89

Figure 6.3 Examples of two commonly used drugs containing a sulfoxide moiety 90

LIST OF TABLES

Table 1-1 Relevant examples of applications of biocatalysis in the synthesis of pharmaceuticals 10

Table 1-2 Applications of P450s to bioconversion processes 22

Table 2-1 Diastereomeric ratios and reaction yields for the synthesis of **2.4**-**2.8** 36

Table 3-1 Performance of P450 3A4 mutants in the oxidation of compound **3.19** 54

Table 4-1 P450 3A4 mutant activity and selectivity towards hydroxylation of **4.1** 60

Table 4-2 Synthesis of (*R*)-lisofylline and regioselectivity of 13 selected P450 mutants 63

Table 4-3 Activity of A370I vs wild-type towards hydroxylation of **4.1** 65

Abbreviations

ABD = Agonist Binding Domain

Ala = Alanine

Arg = Arginine

Asn = Asparagine

Asp = Aspartic acid

Cbz = Carboxybenzyl

Cys = Cysteine

C-H = Carbon-Hydrogen

C-O = Carbon-Oxygen

CHP = Cumene Hydroperoxide

CNS = Central Nervous System

CPR = Cytochrome P450 Reductase

CYP = Cytochrome P450

DCM = Dichloromethane

DFT = Density functional theory

DMSO = Dimethyl Sulfoxide

DNA = Deoxyribonucleic Acid

DPP = DiPeptidyl Peptidase

d-PKI = Deuterated Phosphate Buffer

E_a = Energy of activation

er = Enantiomeric Ratio

ESI = Electrospray Ionization

FF = Force field

Gly = Glycine

Gln = Glutamine

Glu = Glutamic acid

GluK2 = Glutamate Receptor Kainate 2

His = Histidine

HLM = Human Liver Microsomes

HPLC = High Pressure Liquid Chromatography

HPLC-MS = High Pressure Liquid Chromatography – Mass Spectrometry

HRMS = High-Resolution Mass Spectrometry

iGluR = Ionotropic Glutamate Receptor

Ile = Isoleucine

KAR = Kainate Receptor

LC = Liquid Chromatography

LC-MS = Liquid Chromatography-Mass Spectrometry

LC-UV-MS = Liquid Chromatography- Ultraviolet-Mass Spectrometry

Lys = Lysine

MD = Molecular dynamics

Met = Methionine

MM = Molecular Mechanics

MS = Mass Spectrometry

NAD(P)H = Nicotinamide Adenine Dinucleotide Phosphate

NMR = Nuclear Magnetic Resonance

PCR = Polymerase Chain Reaction

PDB = Protein Data bBank

Phe = Phenylalanine

POP = Prolyl Olipopeptidase

Pro = Proline

PKi = Phosphate Buffer

QM = Quantum Mechanics

Rf = Retention Factor

RT = Room Temperature

Ser = Serine

SoM = Site of Metabolism

Tb = Theobromine

TFA = trifluoroacetic acid

THF = Tetrahydrofuran

Thr = Threonine

TLC = Thin Layer Chromatography

Trp = Tryptophan

TS = Transition State

Tyr = Tyrosine

UV = Ultraviolet

Val = Valine

WT = Wild-type

CHAPTER 1 - PROTEIN STRUCTURE, DYNAMICS AND TWO RELEVANT PROTEIN FAMILIES

1.1 The role of proteins in human society

Proteins are macromolecules consisting of one or more long chains of amino acid residues linked in head-to-tail fashion via amide bonds. The word protein (from *Proteus*, Greek god) was introduced ~ 180 years ago by Johannes Mulder who recognized the existence of 'animal substances' made of a large numbers of atoms. Many landmarks followed, including Sumner's crystallization of the first protein (urease) in 1926 and pioneering work in X-ray crystallography in the 1960s[3]. The last 50 years have seen a tremendous progress in the field of protein science, thanks to technological advances in the physical sciences as as well as computer modelling[4]. Nowadays, proteins are involved in many sectors of human society. More then 90 % of all drugs currently on the market target proteins[5], with respective sales of tens of billion of dollars. Huge investments by companies and governments are in place for the discovery of new chemicals that are able to affect protein function.

The last two decades have also seen a remarkable growth in the use of protein therapeutics. In 2015, 13 of the 41 FDA-approved drugs were protein biologics[6], and at present more then 170 proteins have been approved for the treatment of many diseases including cancer, diabetes, anemia and viral infections[7].

Outside the pharmaceutical field, proteins are employed in the food industry, paper making, detergents, biofuel production, degradation of waste, textile, leather and cosmetic industry[8]. Enzymes, a particular class of proteins with the ability of catalysing chemical reactions, are the biggest players in the protein market, with sales projected to double by 2021 (Figure 1.1). A rapidly expanding sector at the interface of the chemical and pharmaceutical industry is the use of biocatalysis for the manufacturing of drugs and other fine chemicals[9].

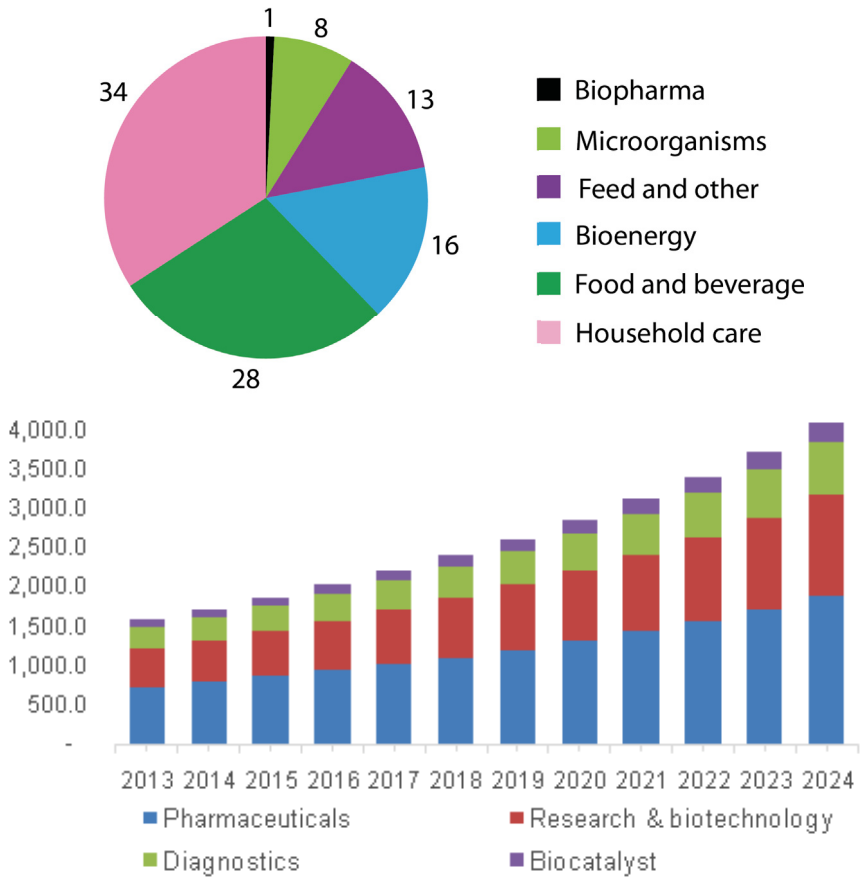


Figure 1.1 Overview of current and predicted enzyme market. Top: Current use of enzymes in the different industrial sectors (*Source: BCC Res Jan 2011: Enzymes in industrial applications: global markets).* Bottom: Predictions of enzyme sales, in North America, in different specialty sectors, 2013 – 2024 (USD Million)

1.2 A look into the structure of proteins

Thousands of different types of proteins, from a multitude of organisms, have already been identified, isolated and characterized[10]. The human genome alone contains about 20000 protein-coding genes, 95 % of which show alternative splicing, meaning that the same gene can encode for different proteins[11]. So far, 18,000 different protein products have been detected in at least one human tissue[12].

After decades of research, what remains fascinating is the uniqueness of each protein as a direct consequence of its three dimensional structure. The immense variety of 3D structures can be attributed to the chemical diversity of the twenty natural amino acids building blocks and the possible number of sequences that they can be arranged in. The structure of proteins is organized into hierarchical levels denoted as primary, secondary, tertiary, and quaternary structure (Figure 1.2). The primary structure of a protein is defined by its amino acid sequence, the secondary structure describes local conformations of the polyamide chain, the

tertiary structure represents the three-dimensional shape of a protein and the quaternary structure describes the non-covalent interaction among several polypeptide strands into oligomeric assemblies consisting of more than one peptide molecule.

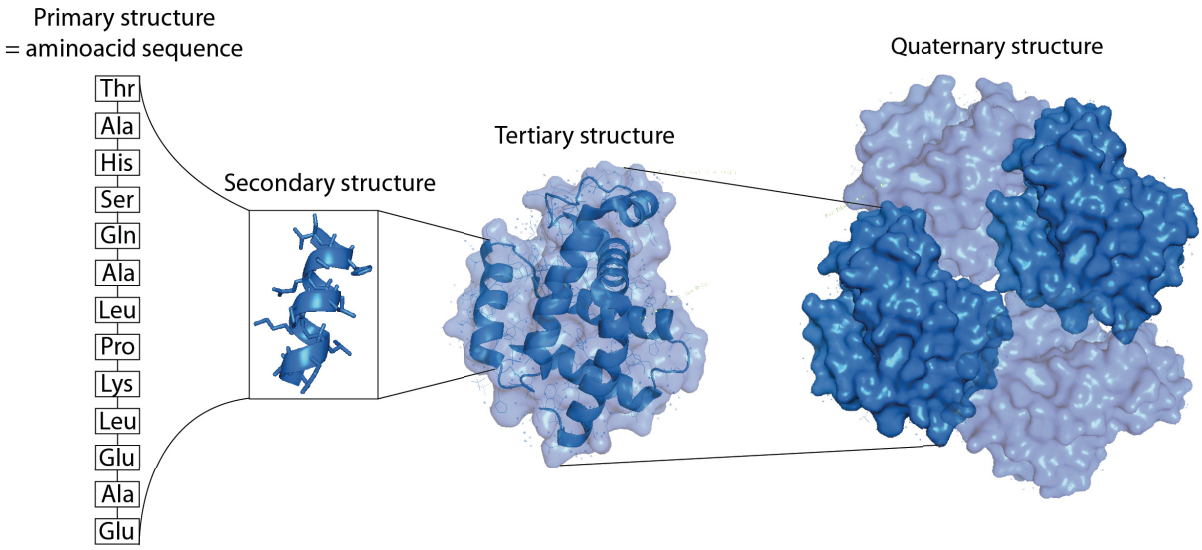


Figure 1.2 Representation of the hierarchical structure of proteins exemplified by hemoglobin

Knowledge of the three dimensional structure of a protein is important in understanding its function. X-ray crystallography, NMR and, more recently, high resolution EM (electron microscopy) have emerged as remarkable tools for structural elucidation, providing scientists with precious insights into the molecular underpinnings of the functioning of many proteins. Many of the studies presented in this thesis could not have been possible without previously determined X-ray crystal structures, which provided the basis for several experimental hypotheses. However, it is important to remember that under experimental conditions for crystallographic studies, proteins are taken out of their native environment and they may exhibit very different features. Temperature, solvent, ionic strength, presence of metal ions, interaction with other biomolecules, aggregation into a crystal lattice, can all play a role in determining the folding of a polypeptide chain and cannot be ignored. Although one can attempt to draw parallels between *in vivo* and *in vitro* studies, it is always important to be aware of the limitations imposed by the experimental techniques for structural elucidation.

1.3 Protein-drug interactions

Proteins are involved in virtually all biological processes and many human diseases are associated with protein malfunctioning resulting in either the reduction or the enhancement of

certain biochemical pathways. As a consequence, much of the drug discovery process relies on our ability to find chemical entities that are able to alter protein function.

It is estimated that an average of 3-6 years are needed for the identification of a lead compound fit for clinical studies[13]. Several strategies have been attempted and are currently used to reduce the time necessary for early stage drug discovery. The struggle to discover drugs has sparked, in the 1990s, the rise of high-throughput screening (HTS). HTS involves a brute force approach where tens of thousands of compounds (compound libraries) are tested against a particular target using a quantitative bioassay via the use of automation, miniaturized assays, microfluidic chips, subnanolitre dispensing and large-scale data analysis[14]. The purpose of HTS is to identify molecules active on the target (most often a protein) and that can then be further converted by chemical optimization to a genuine lead compound (with appropriate potency and selectivity) as candidate for clinical development. Despite the initial excitement, this strategy has not delivered the anticipated results. Experts have come to acknowledge that HTS-supplied hits are often difficult to transform into drugs because of effects such as excessive functionalization or presence of functional groups with known liabilities in terms of ADMETox (adsorption, disposition, metabolism, excretion, toxicity)[15]. In addition, HTS provides no information about the way molecules interact with the target receptor, thereby precluding efficient optimization.

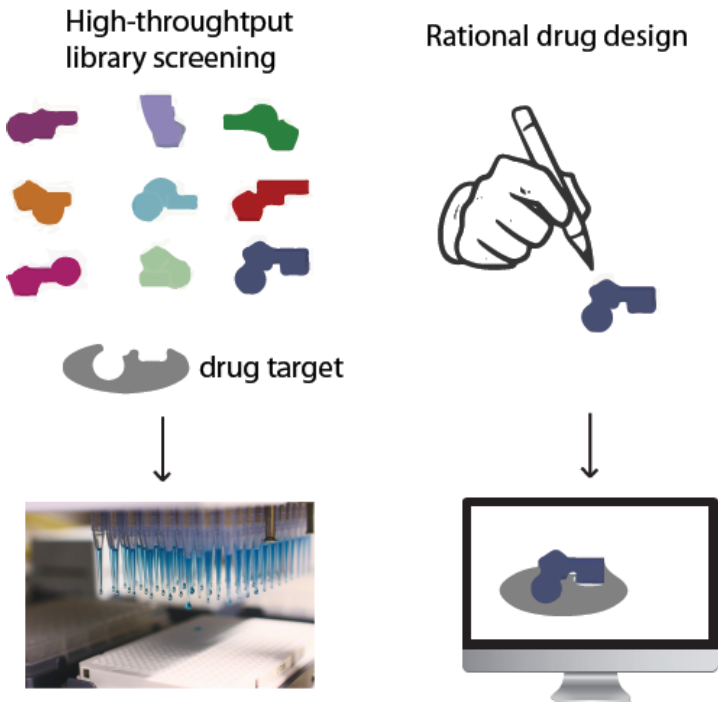


Figure 1.3 Random high-throughput screening vs rational drug design

In response, many research groups throughout the world have dedicated efforts toward other strategies such as structure-based drug design (SBDD), an approach which relies on the knowledge of the three-dimensional structure of a protein and on the use of computational tools to aid the process of drug design in a rational fashion (Figure 1.3). The wide spreading use of molecular modelling is partly the result of the development of user-friendly software

that has provided non-specialists with tools to design or rationalize their own experiments. Despite the ever-increasing power of computers, one of the biggest challenges that remains in the field of computational drug design is the ability to provide accurate predictions in a limited amount of time while taking into account protein dynamics and flexibility.

Proteins are complex molecular systems composed of hundreds (or thousands) of atoms that constantly move and rearrange themselves based on many factors including temperature, solvent properties and interaction with other molecules. At thermodynamic equilibrium, proteins exist in an ensemble of conformational states, the relative abundance of which can, for example, shift upon binding of a ligand. Being able to predict these changes, based on the ligand structural features (shape, electronic properties, etc.) has now become an important subject of research in rational drug design. Part of the research efforts described in this thesis focus on this very topic and will be presented in section 1.7 and in Chapter 2.

1.4 Proteins in drug metabolism

Proteins play an important role in drug development not only because they represent the major pharmacological target, but also due to their involvement in drug metabolism.

Drug metabolism is the process whereby a chemical entity administered as therapeutic is transformed by specialized enzymes in the body in order to facilitate excretion. The products of these enzymatic transformations are called metabolites.

Metabolites can have biological properties that differ from the patent drug and can linger in the body long enough to induce pharmacological effects. In some cases, they exhibit properties similar to the original drug in terms of potency against the desired target, and can therefore contribute in a positive way to the treatment, like in the case of the anti-migraine drug zolmitriptan[16] (Figure 1.4A). In other cases, they are biologically inactive and don't result in any significant effect, as for the metabolism of nifedipine, a blood pressure lowering drug[17] (Figure 1.4B). In a third scenario, metabolites can interfere with other biochemical pathways and cause toxicity. Metabolism-induced toxicity is one of the most common types of adverse side effects of drugs, and is one of the greatest concerns in the process of drug discovery and development[18]. It has been estimated that over 25% of drug candidates fail clinical trials due to unexpected adverse drug properties[19]. Furthermore, there are examples of new drugs entering the market that are withdrawn due to severe adverse effects that are not detected in preclinical or clinical trials[20]. All these scenarios have a dramatic impact on both the patients' health and on delays in the development of new medicines.

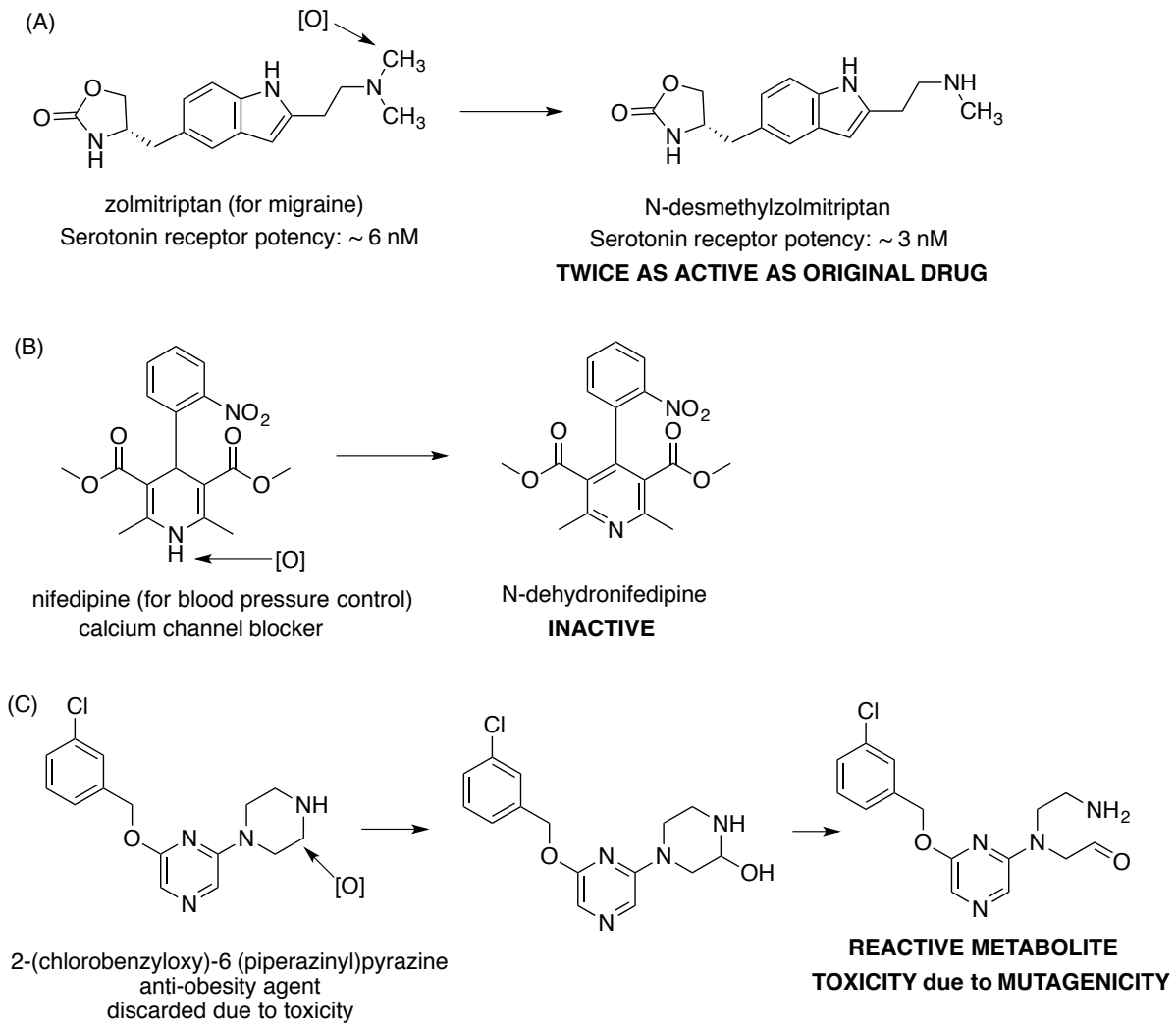


Figure 1.4 Examples of drug metabolism. The metabolites can exhibit potencies similar to the original drug (A), no potency (B) or cause toxicity. [O] = enzymatic oxidation

The major pathway for drug metabolism involves the cytochrome P450 monooxygenases (P450s or CYPs), a family of enzymes able to catalyse a variety of oxidation reactions such as C-H hydroxylation, heteroatom oxidation, double bond epoxidation, among others. The high oxidation potential of P450s can cause formation of so-called "reactive" metabolites, species that are able to covalently bind to cellular macromolecules[21]. From a toxicological assessment, this phenomenon is a major liability and usually results in the ending of a molecular entity as drug candidate, as was the case for 2-(3-chlorobenzyloxy)-6-(piperazinyl)pyrazine, a drug candidate for the treatment of obesity discontinued due to genotoxicity[22] (Figure 1.4C).

Often, reactive metabolites contain electrophilic species (e.g. aldehydes, epoxides, Michael acceptors), known to react with nucleophilic protein residues (e.g. cysteine, lysine). Reactive metabolites may also react with the P450s themselves and inhibit them. P450 inhibition can interfere with the metabolism of other drugs that are present in the body, therefore causing so-

called drug-drug interactions. For all these reasons, great importance is given to early drug metabolism studies in the process of drug discovery.

For many years, optimization of the metabolism of new chemical entities in drug discovery stage has been achieved by trial and error. In this approach, compounds are synthesized and tested for transformation by P450s. Structural modifications during the optimization process are based on empirical methods and previous experience.

More recently, the necessity for minimizing costs and speeding up the drug discovery process has pushed pharmaceutical companies to implement drug metabolism assessments at an even earlier stage, through *in silico* methods[23]. Most of these methods rely on the use of molecular docking to predict protein-ligand interactions between the major P450s and the drug candidate and from these compute the most likely site of metabolism (SoM) (Figure 1.5).

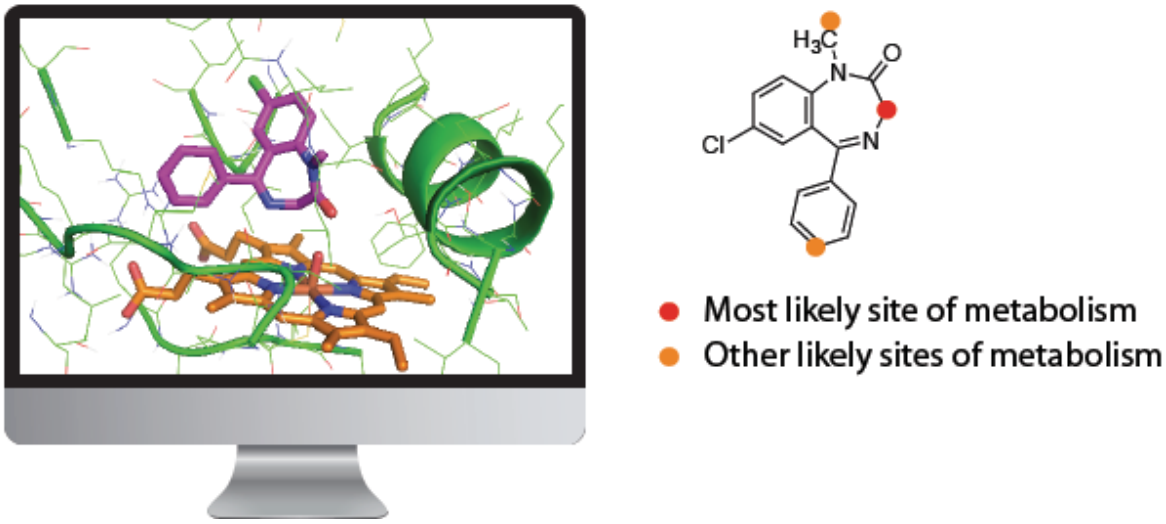


Figure 1.5 Predicted sites of metabolism of diazepam. Docking of diazepam into the active site of human P450 3A4 provides the most likely site of metabolism and other possible options

Despite the advances in computational power, none of the available software can predict accurately sites of metabolism by P450s for most known drugs. Efforts are being directed towards the development of programs that are able to produce, with little computations, reliable predictions. In order to do this, experimental data on biotransformation of new molecules is key in validating these programs.

In reality, drug metabolism is a complicated process that only begins with an enzymatic reaction. Metabolites can exhibit completely different properties from their parent drug, not just in terms of affinity towards the pharmacological target, but also in terms of chemical reactivity. In the end, a combination of computational predictions and experimental

procedures is the best approach. This is a constant theme in the research described in this thesis, and will be dealt with in a very interesting case of metabolism described in chapter 5.

1.5 Proteins in biocatalysis

Biocatalysis is defined as the process through which biological catalysts are used to speed up chemical reactions. It is a practice that humans have known for a long time, with the oldest records of brewing dating back to about 6000 years ago.

From an industrial point of view, biocatalytic processes are attractive for several reasons: 1) enzymes, the most common biocatalysts, are usually produced from renewable feedstocks and are biodegradable; 2) enzymes are usually non-toxic catalysts, unlike many of the metal and organometallic catalysts; 3) biocatalytic processes are generally operated at moderate reaction conditions (temperature, pressure and pH levels) leading to low energy consumption; 4) enzyme-catalyzed reactions are usually very selective leading to high product purity, decreased waste production, facilitating the downstream process; 5) biocatalytic processes are commonly run in aqueous reaction media, preventing large consumption of organic solvents; and 6) recombinant DNA technologies offer the possibility of improving the performance of biocatalysts. These advantages put forward a number of potential economic and environmental benefits, in line with the 12 Principles of Green Chemistry[24] (Figure 1.6).

All these reasons have pushed many fine chemical and pharmaceutical companies to focus on acquiring biocatalysis expertise.

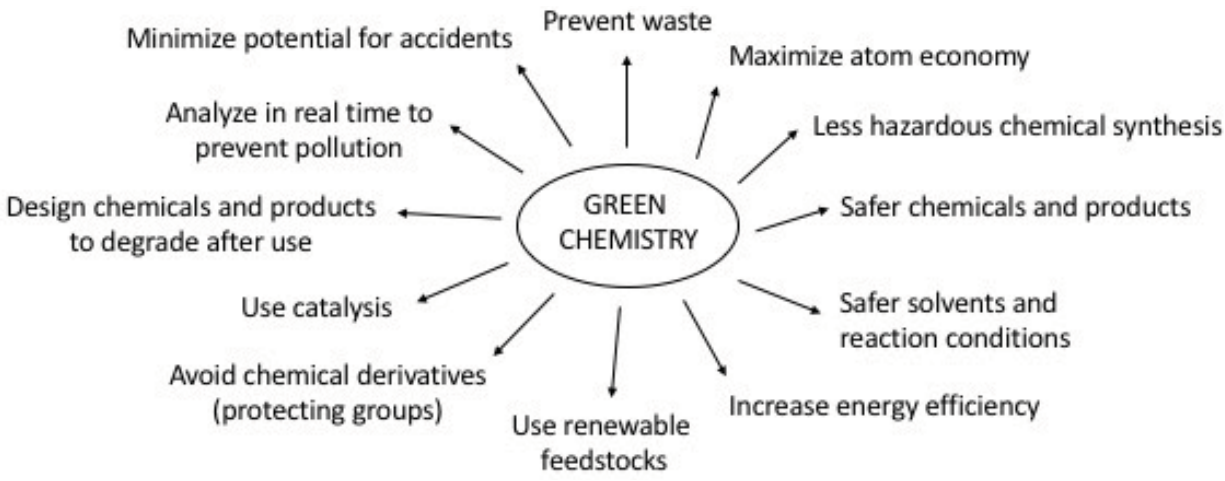


Figure 1.6 12 principles of green chemistry

1.5.1 Applications of biocatalysis in the pharmaceutical industry

Presently, the majority of enzyme-catalyzed reactions in the pharmaceutical industry make use of hydrolases (acylases, amidases, esterases, lipases, proteases). An interesting application of this class of enzymes is for the preparation of chiral compounds through resolutions, deracemizations and desymmetrizations[25]. (*R*)-epichlorohydrine, produced in more than a million ton every year, is a classic example of a molecule generated by enzymatic resolution (Figure 1.7).

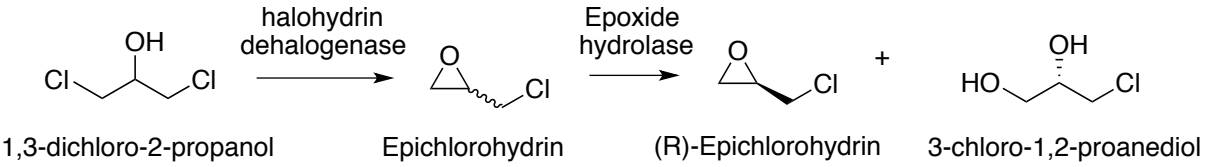


Figure 1.7 Synthesis of (*R*)-epichlorohydrine using halohydrin dehalogenase and epoxide hydrolase

Other interesting enzyme-catalyzed reactions that are increasingly used include asymmetric C–C bond formation using lyases and the synthesis of chiral amines catalyzed by transaminases[26], as in the synthesis of sitagliptin (Figure 1.8).

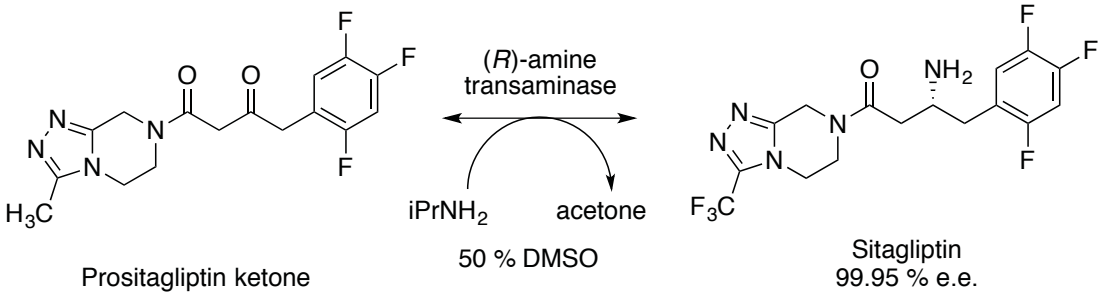


Figure 1.8 Synthesis of sitagliptin using an engineered (*R*)-selective amine transaminase

Dehydrogenases have also demonstrated great potential for the synthesis of many industrial organic products[27]. The industrial applications of oxidoreductases cover the synthesis of amino acids (natural and non-natural), chiral alcohols, amines and amides[28]. Perhaps the most established biocatalytic redox reaction is the ketone reduction to a chiral alcohol, using alcohol dehydrogenases (ADH) or ketoreductases (KRED). This process has gained a lot of interest due to the high enantioselectivity of the enzymes[29], and it has been applied, for example, to prepare one of the key intermediates in the synthesis of the drug montelukast (Figure 1.9).

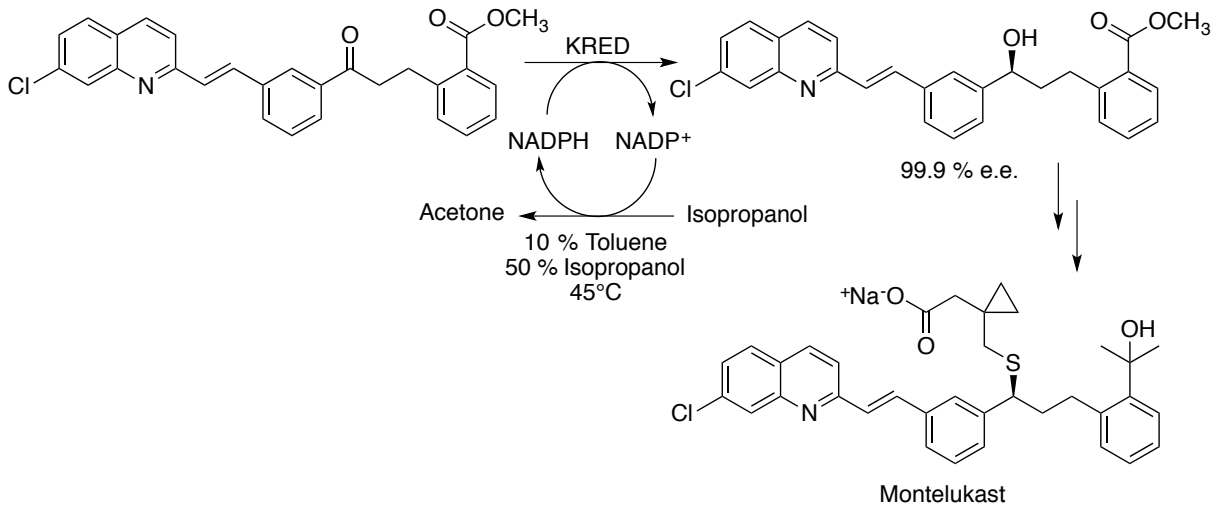


Figure 1.9 Stereoselective reduction catalysed by a ketoreductase in the synthesis of montelukast

Among other interesting enzymes catalyzing redox reactions, Baeyer-Villiger monooxygenases (BVMO) are notable for their ability to enable the selective insertion of an oxygen atom into a cyclic ketone to create a lactone[30], an important reaction for the synthesis of multi-cyclic lactones, which are common pharmaceutical intermediates. Finally, the heme-containing P450 oxygenases are generally recognized as having an enormous potential[31-32] which will be discussed in more detailed in section 1.8.

Additional examples of industrial applications of enzymes are listed below in Table 1.1.

Table 1-1 Relevant examples of applications of biocatalysis in the synthesis of pharmaceuticals

Target compound	Company	Reaction	Enzyme	Organism
SCH 56592	Schering Plough	desymmetrization	lipase	*Candida antartica*
β-Lactams	GSK[a]	desymmetrization	lipase	*Candida antartica*
Lotrafiban	GSK[a]	Hydrolysis	lipase	*Candida antartica*
Paclitaxel	BMS[b]	Hydrolysis	lipase	*Pseudomonas AK* *P. cepacia PS-30*
SCH 66336	Schering Plough	desymmetrization	lipase	*P. aeruginosa*
Xemilofiban	Monsanto	Hydrolysis	penicillin acylase	*E. coli*
Renin inhibitor	Hoffmann La Roche	Hydrolysis	subtilisin	*B. licheniformis*
Lamivudine	GSK	Hydrolysis	deaminase	*E. coli*
AG7088	Pfizer	Reduction	Lactate,formate dehydrogenases	*L. mesenteroides* *C. boidinii*
ACE inhibitor	Ciba-Geigy	Reduction	D-lactate dehydrogenase	*S. epidermis*
LY300164	Eli Lilly	Reduction	dehydrogenase	*Z. rouxii*

Omapatrilat	BMS	Reductive amination	phenylalanine dehydrogenase	*T. intermedius*
N-Butyl DNJ	Pharmacia	Oxidation	dehydrogenase	*G. oxydans*
2-Quinoxaline-carboxylic acid	Pfizer	Oxidation	monooxygenase	*A. repens*
Lobucavir prodrug	BMS	acylation	subtilisin	*B. licheniformis*

[a]GSK = Glaxo Smith Kline; [b]BMS = Bristol-Meyers Squibb

One of the major challenges to the successful development of a commercial biocatalytic process is the inherent specificity of enzymes for their natural substrate, which significantly restricts applications of one given enzyme to different biotransformations. This problem has been addressed through different approaches. Advances in genetics have significantly contributed to the discovery of new enzymes, thereby expanded the pool of biocatalysts from which synthetic processes can be built on. Besides identifying new enzymes, a common approach consists of altering the features of known enzymes using protein engineering techniques. The issue of enzyme specificity is at the centre of the field of biocatalysis and will be the subject of Chapters 3 and 4 in this thesis. An introduction to protein engineering is presented below.

1.6 Protein engineering

Protein engineering is a relatively young discipline that aims at modifying protein structure with the goal of improving existing functions or introducing new ones. It has proven to be a valuable tool to increase our basic understanding of protein structure-function relationships.

The first engineered enzymes were commercialized (subtilisins) at the beginning of the 1990s[33], which marked the beginning of a new era for the use of biological systems in chemical applications. Since then, the use of engineered macromolecules has been growing in all areas of biotechnology, from pharmaceuticals to biomaterials and biofuels.

Protein engineering can be achieved both through chemical modification and mutagenesis. In the former strategy, also referred to as bioconjugation, the properties of a protein are modified by forming a covalent link with other molecules, biological or newly designed. This approach has been used extensively, for example, in the preparation of antibody-drug conjugates[34], attachment of fluorescence probes or immobilization on solid supports[35].

On the other side, protein engineering via mutagenesis consists of altering protein function by modifying its amino acid sequence. The two main approaches to achieve it are rational design and directed evolution.

1.6.1 Directed evolution

Directed evolution is a technique that mimics the mechanisms that underlie biological evolution, i.e. the production of a large number of gene variants, followed by selection of the most suitable protein variants and amplification of their corresponding genes[36]. In contrast to *in vivo* evolution, which needs thousands or millions of years to leave a mark, directed evolution generates results in a much shorter timeframe, from days to weeks.

The major advantage of directed evolution is that it does not require any prior knowledge of the three-dimensional structures of a protein, nor of its mechanism of action. Instead, a random-based iterative strategy is used. At the beginning of the cycle, a large library of mutants is generated by introducing random DNA mutations (Figure 1.10). Then, through a high-throughput screening, the best mutants are selected and used as starting point for the next cycle. At the end of the optimization, an improved (evolved) variant of the lead enzyme is generated. The engineered product will contain a variable number of mutations, depending on the number of cycles performed as well as the methodology used to introduce the mutations.

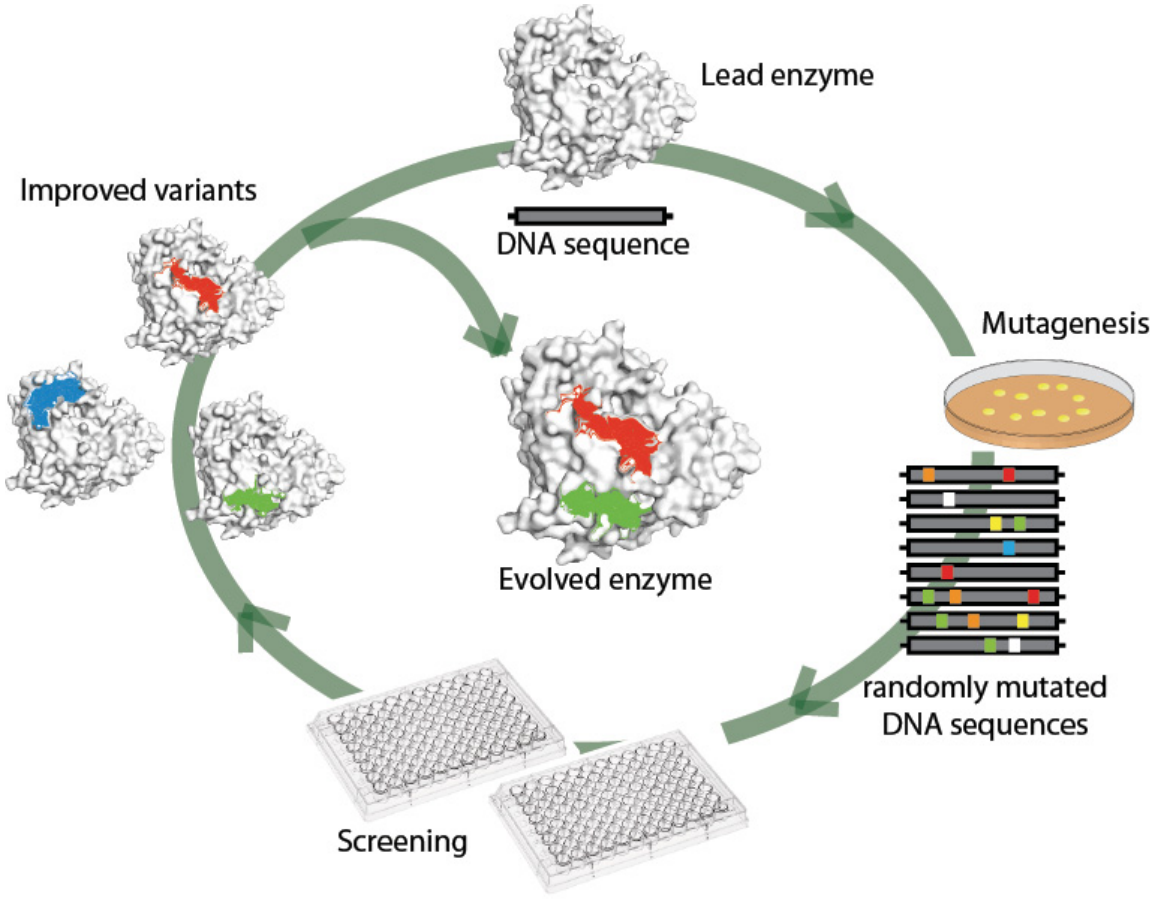


Figure 1.10 Directed evolution cycle

The two natural evolutionary processes that have been adapted for *in vitro* evolution are random mutagenesis and gene recombination. Random mutagenesis was initially achieved using chemical compounds or UV irradiation to introduce genetic mutations, while currently the most commonly used system uses PCR-based mutagenesis. Gene recombination is based on the fragmentation of parental DNA into small nucleotide sequences and the following reconstitution of the entire gene through shuffling of such nucleotide fragments. This method exploits natural variations between naturally existing genes to create new ones, and it often generates correctly folded and functional chimeric proteins.

Directed evolution has been used with great success in recent years for optimizing enzyme function. Since the late 1990's directed evolution has been used to enhance enzyme properties such as thermostability[37-39], tolerance to organic solvents[40] and pH profile[41]. Combining both rational design and directed evolution it has been possible to change stereoselectivities[42-43] and to expand enzymes' substrate specificity[44-45].

1.6.2 Rational design

The term "rational" comes in opposition to other "random" strategies, like directed evolution. Quoting Willem Stemmer, the father of DNA shuffling, "Rational design […] attempts to modify or create [protein] molecules for specific applications by predicting which amino acid sequence will produce a protein with the desired properties". This strategy has the advantage of being relatively less expensive than directed evolution, but it requires knowledge of the three-dimensional structure of the desired protein. If the protein structure has not been experimentally determined and its amino acid sequence displays a certain similarity with that of fully characterized proteins, a model of the structure of the target protein can be built by using other similar proteins as template (so called homology modelling).

Rational design is performed by taking advantage of computational tools to predict how protein function will be affected by specific mutations. Recombinant DNA technologies are used to produce the promising mutants *in vitro*. For example, PCR-based site-specific mutagenesis is an established and easy to implement technique to introduce the desired mutations (deletions, insertions, substitutions) in the gene of interest. The mutated protein can then be expressed in a heterologous host, the most common being *E. coli*, and finally analyzed for the desired feature. A protein can, for example, be engineered to increase its affinity for a specific ligand, by introducing amino acid residues that are predicted to form productive interactions (Figure 1.11).

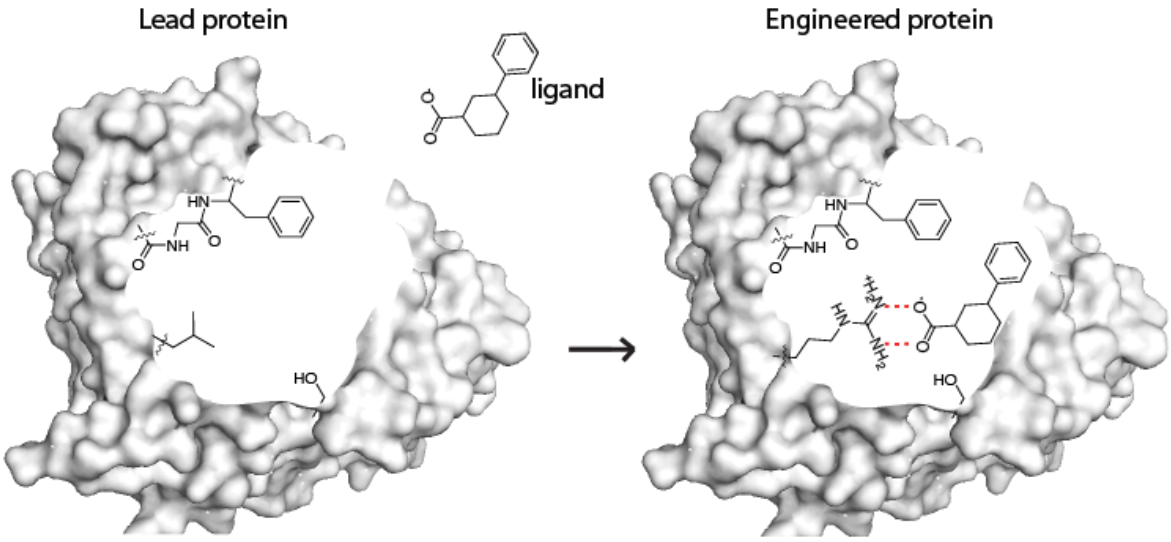


Figure 1.11 Example of how rational design can be used for protein engineering. Mutating an isoleucine residue into an arginine can orient a carboxylic acid-containing ligand

Several enzymes commonly used in industrial processes are engineered through rational design to increase thermal stability[46], to achieve reaction stereoselectivity[47] and for activity in organic solvent[48]. Protein rational design is also becoming an increasingly useful tool for optimizing protein drugs and creating novel biotherapeutics[49].

Choice of two families of proteins

Proteins can be classified according to many different parameters such as their role within an organism (structural, catabolic etc.) or locations within cellular compartments (membrane-bound, cytosolic, etc.). One way to differentiate proteins is to distinguish between catalytic (enzymes) and non-catalytic (receptors, antibodies, among others) proteins. This thesis looks at both an enzyme, P450 3A4, and a receptor, GluK2. An introduction to both is presented in the following sections.

1.7 Kainate receptors, a subtype of glutamate-gated channels

The glutamate-gated channels (iGluR) are ligand-gated tetrameric ion channels that mediate synaptic transmission in the central nervous system. They are not considered enzymes because they don't catalyze any reaction, but their ability to trigger cellular responses, following interactions with an agonist, characterizes them as receptors. Several subtypes of iGluRs can be distinguished, according to the nature of the ligand. The main subtypes are (*S*)-2-amino-3-

(3-hydroxy-5-methyl-4-isoxazolyl)propionic acid (AMPA) receptors, *N*-methyl-D-aspartate (NMDA) receptors, and kainate-receptors (KAR). Each of the four subunits constituting iGluRs share a similar structure (Figure 1.12), consisting of:

- An amino-terminal domain (ATD)
- An agonist binding domain (ABD) formed by lobe D1 and D2. The two lobes are separated by a cleft in which the ligand binds
- A transmembrane domain
- A carboxyl terminus, located inside the cells and interacting with other proteins

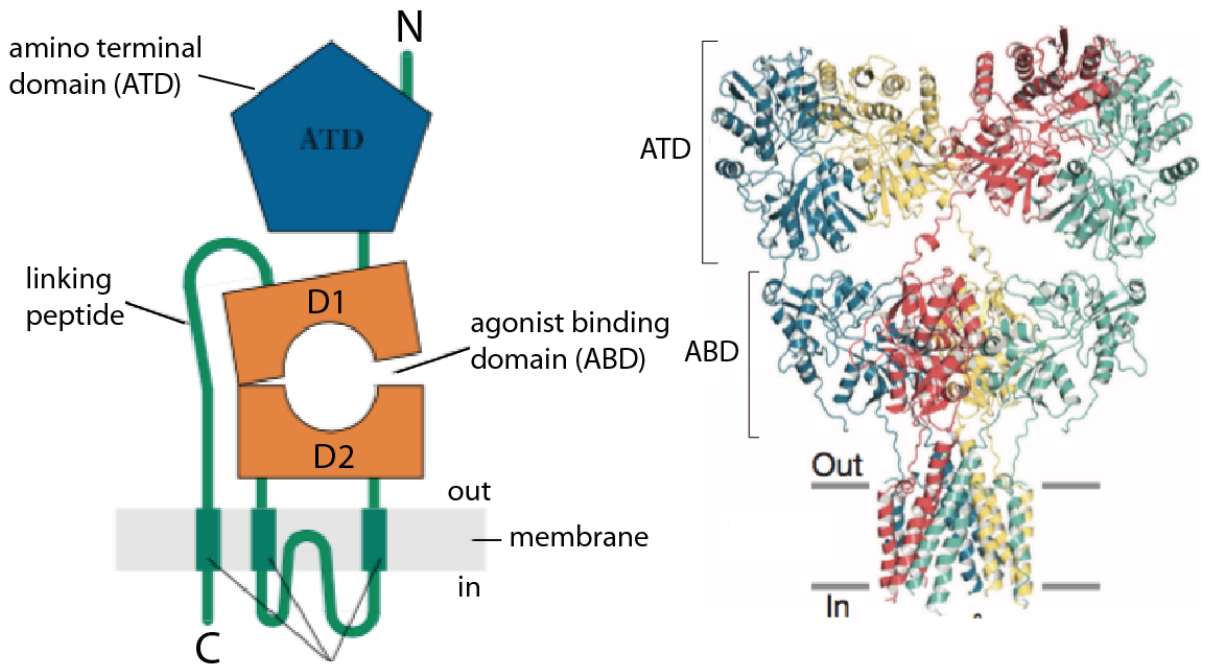


Figure 1.12 Glutamate receptor subunit structure. On the left, schematic representation of glutamate receptor subunit structure. On the right, experimental structure of KAR GluK2. ATD = amino terminal domain; ABD = agonist binding domain; D1, D2 = ABD lobes

Four subunits assemble to form the ion channel in a dimer-of-dimer model. The assembly has been proposed to be sequential: two monomers dimerize first and then two of the newly formed dimers dimerize again. For kainate receptors, specifically, 5 subtypes have been identified: GluK1, GluK2, GluK3, GluK4 and GluK5 which co-assemble in various combinations to form functional receptors. Subunits of KARs assemble to form homomere or heteromere dimers. Homomeric GluK2 will be the focus of Chapter 2.

Abnormalities in glutamatergic neurotransmission are considered to be an important factor contributing to neurodegenerative and mental disorders[50].

In particular, there is evidence strongly suggesting that KARs might be involved in

imbalances in the excitatory/inhibitory processes linked to epilepsy. In fact, the data available from post-mortem human epileptic tissue indicates an upregulation of kainate receptors in the hippocampus of pharmacoresistant TLE (temporal lobe epilepsy) patients[51]. In addition, a consortium of academic and industry groups[52] showed that antagonists of GluK1 prevent the development of epileptic activity in certain areas of hippocampal slices in an epilepsy model.

1.7.1 Conformational flexibility of KARs

Despite showing promise in animal models of these conditions, no KAR antagonist has yet been approved for therapeutic intervention in humans[53]. This failure is partially attributable to the difficult development of KAR subunit-specific antagonists[54], however, a more general obstacle is a limited knowledge concerning how iGluR structure relates to function[55]. In particular, it is difficult to predict how compounds rationally designed to target KAR will modulate channel activity.

An early structural model describing iGluR activation proposed that the degree of closure of the extracellular agonist binding cleft around agonist molecules correlated with increased agonist efficacy[56]. In this sense, the dimeric agonist binding domain (ABD) can be thought of as a clamshell, comprised of an upper lobe (D1) and lower lobe (D2) that both form the agonist cleft (Figure 1.13). When an agonist binds to the ABD, it is thought that the D2 lobes are pulled upward, closing the agonist binding cleft and generating tension on residues linking the ABD to the transmembrane pore. This tension has been shown to correlate with channel opening, whereby cations pass through the cell membrane to generate a physiological response[57]. For KAR, channel opening is very transient, occurring for just milliseconds before the agonist-bound receptor enters a desensitized state characterized by a long duration channel closures[58-59]. Alternatively, channel closure can be triggered when the agonist unbinds from the receptor, leading the protein to revert to an unliganded apo state.

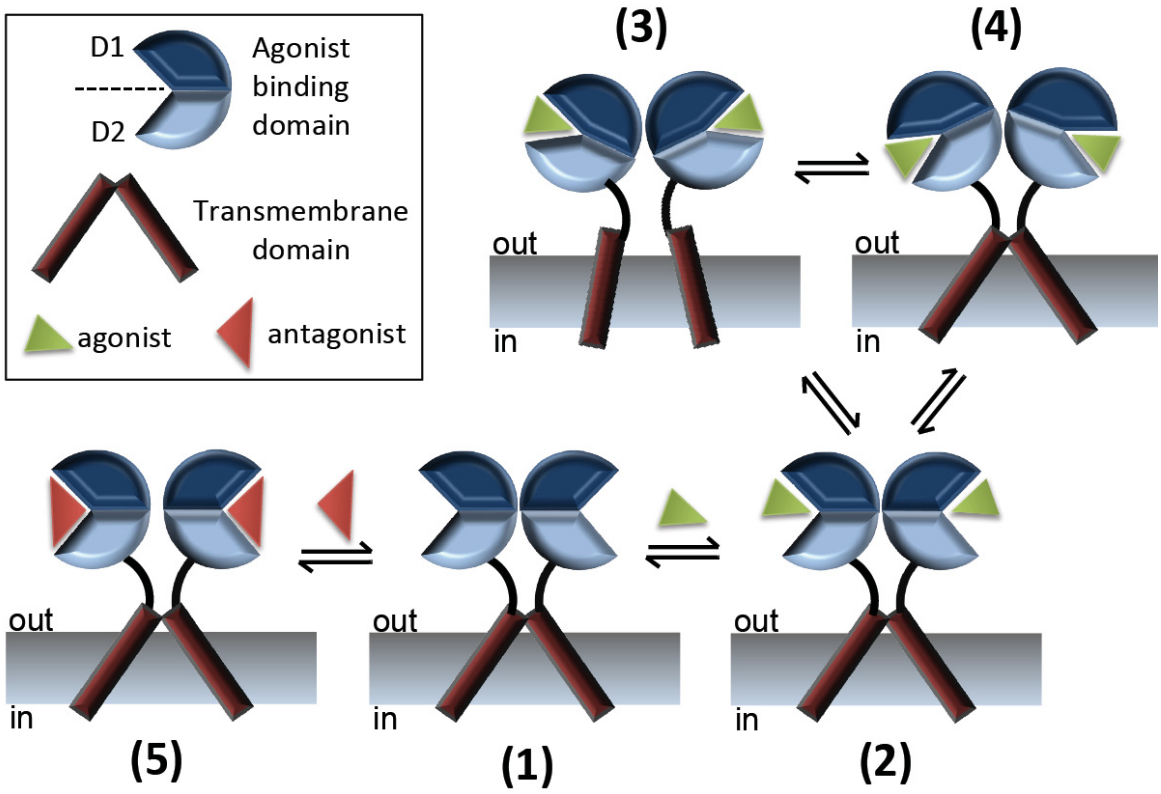


Figure 1.13 Accepted model for the structural basis of KAR activation. In its resting state (1), the receptor ion channel is closed and the agonist binding domain (ABD) is ready to accept ligands. Binding of agonist molecules (2) favors closure of the ABD, which is correlated with increased frequency of channel opening (3). Receptor activation is in equilibrium with desensitization, where the ABD adopts a different conformation that inhibits opening of the ion channel (4). When binding, an antagonist interferes with closure of the ABD and consequently prevents activation of the receptor (5).

Incorporating protein flexibility in drug design is still one of the big challenges today[60]. Many efforts are being directed towards capturing the motion of proteins experimentally. Nuclear magnetic resonance (NMR), with its ability to deliver timescales between conformational states, has emerged as a powerful tool to study protein dynamics, but it still suffers from high experimental costs and intensive labor.[61] In addition, NMR methods are limited by the possibilities of studying protein with molecular masses < 70 KDa. Some of these inconveniences can be circumvented by the use of *in silico* techniques. The most accurate simulations of protein motion rely solely on the laws of quantum mechanics (QM), in so-called *ab initio* methods. Unfortunately, this approach can be computationally intensive and extremely time demanding even with the most powerful supercomputers. In fact, many QM methods are limited to 200 atoms by default.

Molecular dynamics (MD) simulations seek to overcome this limitation by using simple approximations based on Newtonian physics to simulate atomic motions. Chemical bonds and

atomic angles are modeled using simple virtual springs, while the forces acting on each of the atoms are estimated from Newtonian-like equations. Collectively, these parameters are called a 'force field'. AMBER[62], CHARMM[63], and GROMOS[64] are example of commonly used force fields. Once the forces acting on each of the atoms have been calculated, the positions of these atoms are moved according to Newton's laws of motion, in accordance to molecular mechanics (MM) principles. The simulation time is then advanced, often by only 1 or 2 femtoseconds. Even by relying on MM principles, the computational demands for MD can be very high. Many biochemical processes, including receptor conformational shifts relevant to drug binding, occur on the millesecond-to-second time scales that are much longer than those amenable to simulation[65].

A less-computationally demanding solution to the problem of receptor flexibility is the use of MM-based "flexible docking" (Figure 1.14). This strategy is possible when multiple receptor conformations have been already defined. It is used to rank ligands in terms of affinity by sequentially docking them into each receptor structure. Each member of the conformational ensemble can be considered to contribute to the docking score either equally or with some weighing factor. In a different scenario, flexible docking can be used to estimate affinity of one ligand to the different conformations. This concept was used in Chapter 2 of this thesis to study the effect of structural features of glutamate analogues on the conformational flexibility of glutamate-type receptor, GluK2.

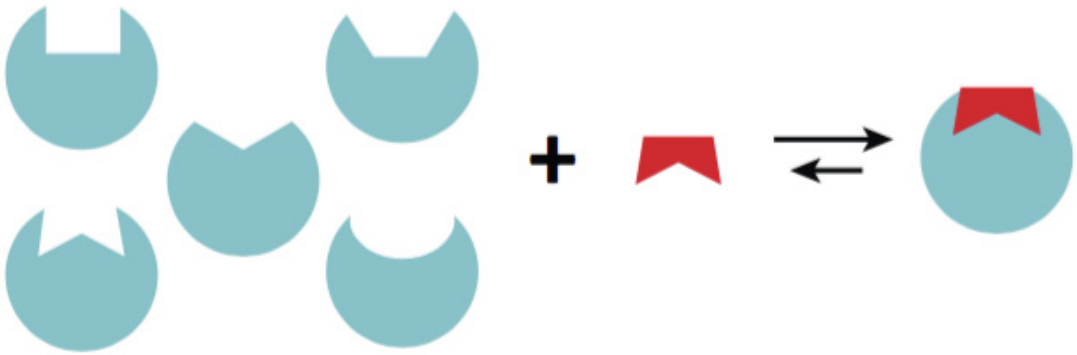


Figure 1.14 Flexible docking

1.8 Cytochrome P450 enzymes

1.8.1 Introduction to cytochrome P450 enzymes

Discovered more than 70 years ago, cytochrome P450 enzymes (CYPs or P450s) constitute a superfamily of heme-containing monooxygenases that are widespread in nature. They are found in mammalian tissue and organs as well as in plants, bacteria, yeast and insects, with more than 20,000 P450 sequences known today[66]. These enzymes have received their name based on their spectral properties exhibiting a typical absorption maximum at 450 nm for the reduced carbon monoxide-bound complex[67]. The prosthetic group present in P450s is the heme b (iron protoporphyrin IX). In mammals, P450s have two main roles[68]: the metabolism of xenobiotics (compounds exogenous to the organism), as well as the synthesis of critical signaling molecules like steroid hormones and polyunsaturated fatty acids. P450s exhibit a remarkable diversity in the type of reactions that they are able to catalyze. Analysis of the literature reveals up to twenty different reactions including hydroxylations, heteroatom oxidations, epoxidation of alkenes, isomerizations[69-71] (Figure 1.15).

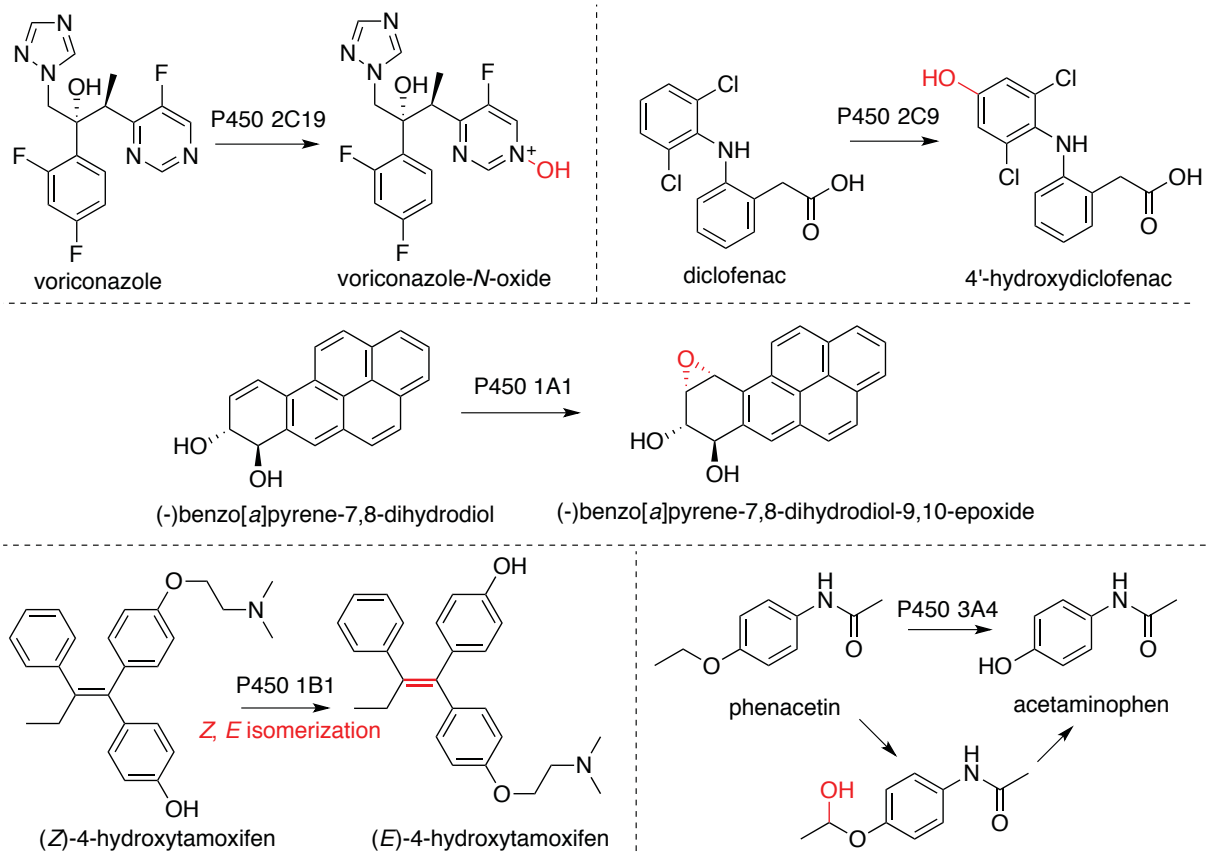


Figure 1.15 Examples of human P450-catalyzed reactions

1.8.2 P450 reaction mechanism

A very interesting reaction catalyzed by P450s is the insertion of one atom of oxygen into a non-activated C-H bonds. This reaction involves the activation of molecular oxygen by the heme-iron center. One oxygen atom is inserted into the substrate while the other is reduced to water, obeying the stoichiometry:

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

The electrons are derived from NAD(P)H (nicotinamide adenine dinucleotide (phosphate)) and are delivered to the P450 by an enzymatic redox partner. In microsomal P450s, the redox partner is called cytochrome P450 reductase (CPR)[72]. In mammalian cells, both the P450 and its redox partner are bound to a membrane (Figure 1.16)

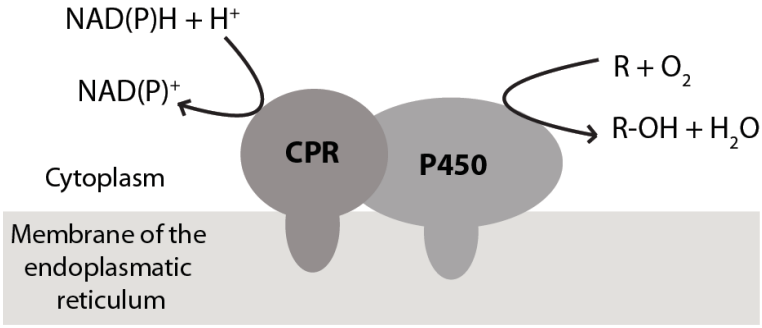


Figure 1.16 Schematic organization of mammalian P450 systems. CPR = cytochrome P450 reductase

The mechanism of P450s was initially introduced in 1968[73] but it was later extended to accommodate newly discovered reaction pathways and intermediates[68, 74-75]. The catalytic cycle can be described in eight consecutive steps (Figure 1.17). In its resting state, the heme-iron is coordinated to four nitrogens from the porphyrin ring, to a cysteinate axial ligand and to a water molecule distal ligand. Iron is in the +3 oxidation state (ferric). Binding of a ligand in the substrate pocket causes a conformational change in the heme iron binding pocket leading to a spin shift of the heme iron from a low-spin (6-coordinate) state to a high-spin (5-coordinate) state (step 1) with water leaving. Next, reduction of the iron to its ferrous state (+2) by the transfer of a first electron from NADPH via the redox partner (step 2). Subsequently the iron binds oxygen to form an oxy-ferrous complex (step 3). The second electron transfer reduces the oxy-ferrous complex and produces a negatively charged iron-peroxo species (step 4) which is then protonated at the distal oxygen atom to form an iron-hydroxyperoxo complex, also referred to as compound 0 (step 5). Further protonation at the distal oxygen facilitates O-O bond cleavage to form the iron-oxo species (ferryl with a radical

on the porphyrin) also referred to as compound I, with the formation of a molecule of water (step 6). The next step of the reaction involves oxidation of the substrate via abstraction of H• by the reactive compound I (step 7), before recombination of the radicals to form the product (step 8).

An important process that regulates product formation is electron coupling or "uncoupling", meaning that not all electrons delivered to the heme iron are contributing to product formation. There are three uncoupling pathways: the oxidase shunt (A), the peroxide shunt (B) and the autoxidation shunt (C). The oxidase shunt occurs when the iron-oxo complex does not oxidize the substrate but is reduced instead to form water. The peroxide shunt is caused by the disintegration of the iron-hydroxyperoxo complex and release of hydrogen peroxide. The autoxidation shunt takes place when superoxide is released (O_2^-) from the oxy-ferrous complex. The uncoupling reactions do not only decrease the turnover number but the uncoupling products (H_2O_2, O_2^-) can damage the enzyme[76-77].

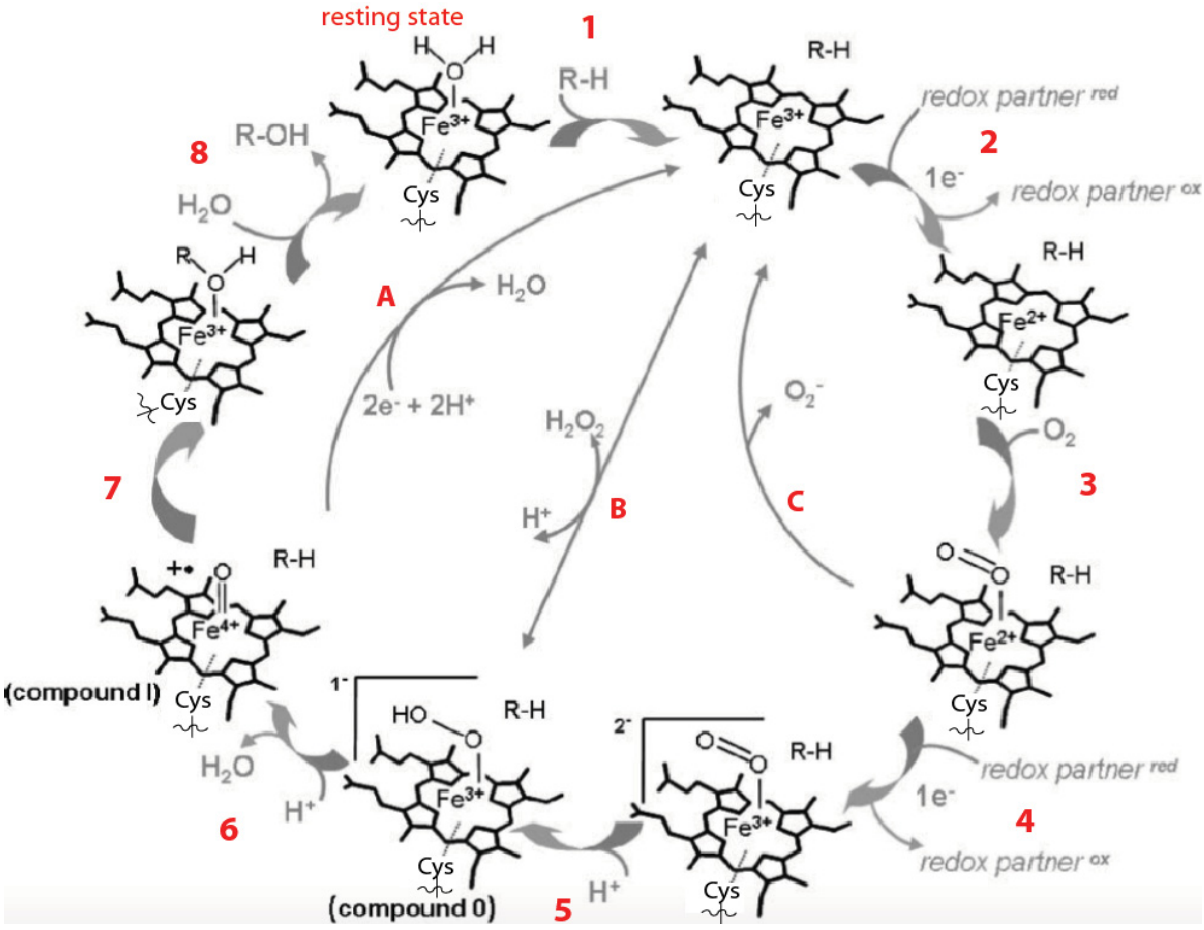


Figure 1.17 Catalytic cycle of cytochrome P450 where R-H is the substrate and R-OH is the product. Steps of the main cycle are numbered in a clockwise fashion and shunt pathways are indicated with letters. For simplification purposes, the P450 active site is depicted by a cartoon representation of heme b (a protoporphyrin IX-iron complex) ligated to the cysteine thiolate ligand

1.8.3 Industrial applications of P450s

The majority of industrial biotransformations catalyzed by P450s (Table 1-2) are performed using whole cells, meaning the substrates are incubated with metabolically active microorganisms, in fermentation plants. In contrast to purified enzymes, whole cell systems are preferable since redox cofactors, such as NADPH, may be regenerated by cellular metabolism[78]. Whole cell systems also offer the advantages of (1) providing the native environment to keep the enzymes stable and (2) avoiding costly isolation and purification of reaction biocatalysts, intermediates, and cofactors. However, research is being directed towards the development of technologies (such as immobilization on solid support) to be able to facilitate the use of purified enzymes, which can potentially lead to higher productivity owing to higher catalyst concentration and simpler product purification[79].

Regio- and stereo-selective hydroxylations are the most common type of reactions catalyzed by industrial P450s for the production of fine chemicals such as pharmaceuticals, flavors, and fragrances[80]. A well-established commercial application of P450s is the hydroxylation of steroid hormones, for example for the synthesis of hydrocortisone with *Curvularia sp*. (Figure 1.18) at a scale of ~100 tons per year[81] (Bayer Pharma). Further examples include the production of pravastatin (a potent cholesterol-lowering agent) by Bristol-Myers Squibb using *Streptomyces carbophilus*[82], the P450-catalyzed hydroxylation of alkanes for the production of dicarboxylic acids, as well as the hydroxylation of important aromatic precursors[83].

Table 1-2 Applications of P450s to bioconversion processes

Substrate	**P450**	**Reaction**	**organism**
Compactin	105A3	6β-hydroxylation	*S. carbophilus*
Vitamin D3	107	25- and 1α-hydroxylation	*P. autotrophica*
Vitamin D3	105A1	25- and 1α-hydroxylation	*S. griseolus*
Progesterone	106A2	11α-hydroxylation	*B. megaterium*
Alkanes	102A1	hydroxylation	*B.megaterium*
(+)-Valencene	109B1	C2-oxidation	*B. subtilis*
Fatty acids	152A1, B1	α- or β-hydroxylation	*B. subtilis*
Styrene	152A1	epoxidation	*B. subtilis*
11-Deoxycortisol	P450lun	11β-hydroxylation	*C. lunata*
11-Deoxycortisol	11V1	11β-hydroxylation	*bovine*
progesterone	17A1, 21B2	17α- and 21-hydroxylation	*bovine*
Endogeneous sterols	11A1, 17A1	hydroxylations to cortisol	*bovine*

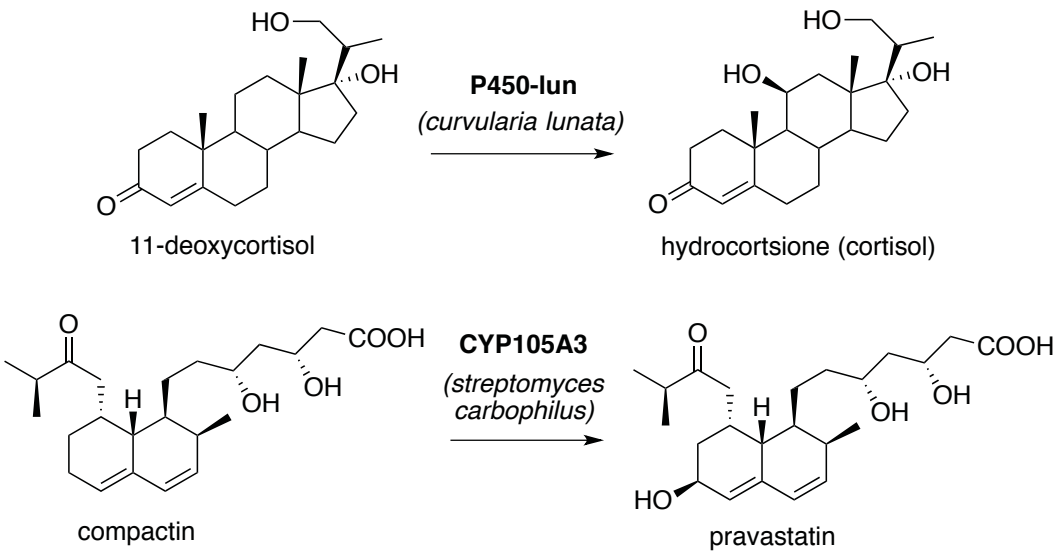


Figure 1.18 Example of P450 applications in industrial processes

1.8.4 Engineering of bacterial P450s

Most of the protein engineering efforts in the P450 area has been focused on bacterial P450cam and P450BM3[166]. In particular, with just one to three active site mutations, P450cam variants have been generated to hydroxylate (+)-R-pinene[167], a structural relative to the natural substrate (+)-camphor, a variety of aromatic compounds[168-169] and indole to form indigo[170]. Another valuable example of substrate scope expansion for P450cam was provided by Xu *et al.* with the engineering of P450cam to hydroxylate short-chain alkanes[171]. Furthermore, multiple attempts have been made to fuse the electron transfer components of P450cam to its reductase domain to increase enzyme activity[172].

P450BM3 is one of the most studied and frequently engineered P450s, as it has the fastest known catalytic rate for a P450[173]. The P450BM3 protein backbone undergoes large structural changes (> 10 Å) during catalysis and the large active site volume reflects the poor regioselectivity with which BM-3 hydroxylates its preferred C12-C20 fatty acids substrates[174]. Much like P450cam, P450BM3 has been engineered to accept a variety of substrates. By mutating the two residues (R47, Y51) located at the opening of the substrate pocket P450BM3 variants were isolated with increased activity for hydroxylation of shorter chain fatty acids[175], and epoxidation of the anti-malarial drug-precursor amorphadiene[176]. Screening with a simple NADPH consumption assay to monitor cofactor consumption in the presence of a substrate, directed evolution of P450BM3 yielded a highly promiscuous variant F87V/L188Q/A74G

with enhanced activity for a variety of substrates such as indole, alkanes, arenes, and polycyclic, aromatic hydrocarbons[177]. Similar engineering efforts generated variants with activity for β-ionone, a carotenoid intermediate[178], and valencene for (+)-nootkatone production[179].

Arnold's research group has focused on using directed evolution to force P450BM3 to accept small alkanes as substrates with the ultimate target of methane hydroxylation[180]. In addition, P450BM3 variants were found with stereoselective secondary hydroxylation of linear alkanes and stereoselective epoxidation of alkenes[181].

1.8.5 Substrate promiscuity of human P450s

P450s play a remarkable role in human drug metabolism, catalyzing the transformation of ~ 75 % of all drugs on the market[84].

The human genome contains 57 genes encoding for P450 enzymes[85]. However, most of the known metabolic reactions are performed by only eleven enzymes: P4503A4, 3A5, 2D6, 1A, 1A2, 2C19, 2C8, 2C9, 2E1, 2B6, 2A6. P450 3A4 alone is responsible for the metabolism of ~50% of all drugs[86] (Figure 1.19).

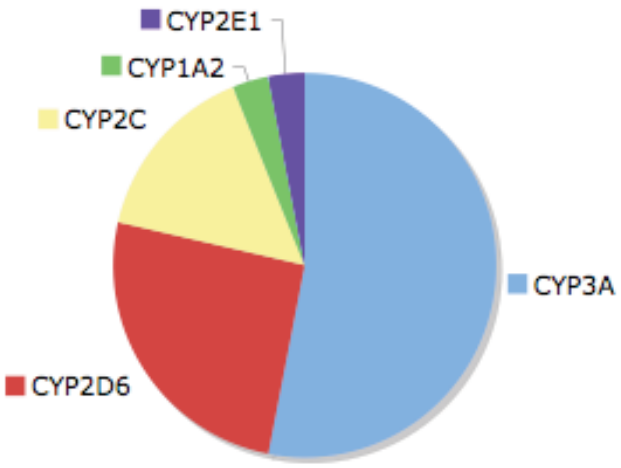


Figure 1.19 Relative importance of P450s in human drug metabolism

The remarkable ability of P450s to accept a wide range of substrates places them in a special position as potential biocatalysts for oxidation reactions. In terms of biocatalytic applications, human P450s have received much less attention than their bacterial cousins, in part due to their lower stability. Nevertheless, small-scale applications in research environments could greatly benefit from versatile oxidation biocatalysts especially if the substrate specificity could be tuned for a desired application.

1.9 P450 3A4

P450 3A4 is the most abundant P450 in most adult humans. It adopts the typical fold of the P450 superfamily with a structural core formed by a four-α-helix bundle comprised of three parallel and one antiparallel helix[68, 87] (Figure 1.20).

The F-F'-G-G'-sequence of α-helices, also found in many P450s, is located in the upper part of the active site and is assumed to be the primary interaction site of P450 3A4 with substrates (Figure 1.21). Molecular dynamics simulations have predicted this sequence to undergo large-scale movements upon interaction with the lipid bilayer, which allow for opening and closing of the substrate access channel[88-89].

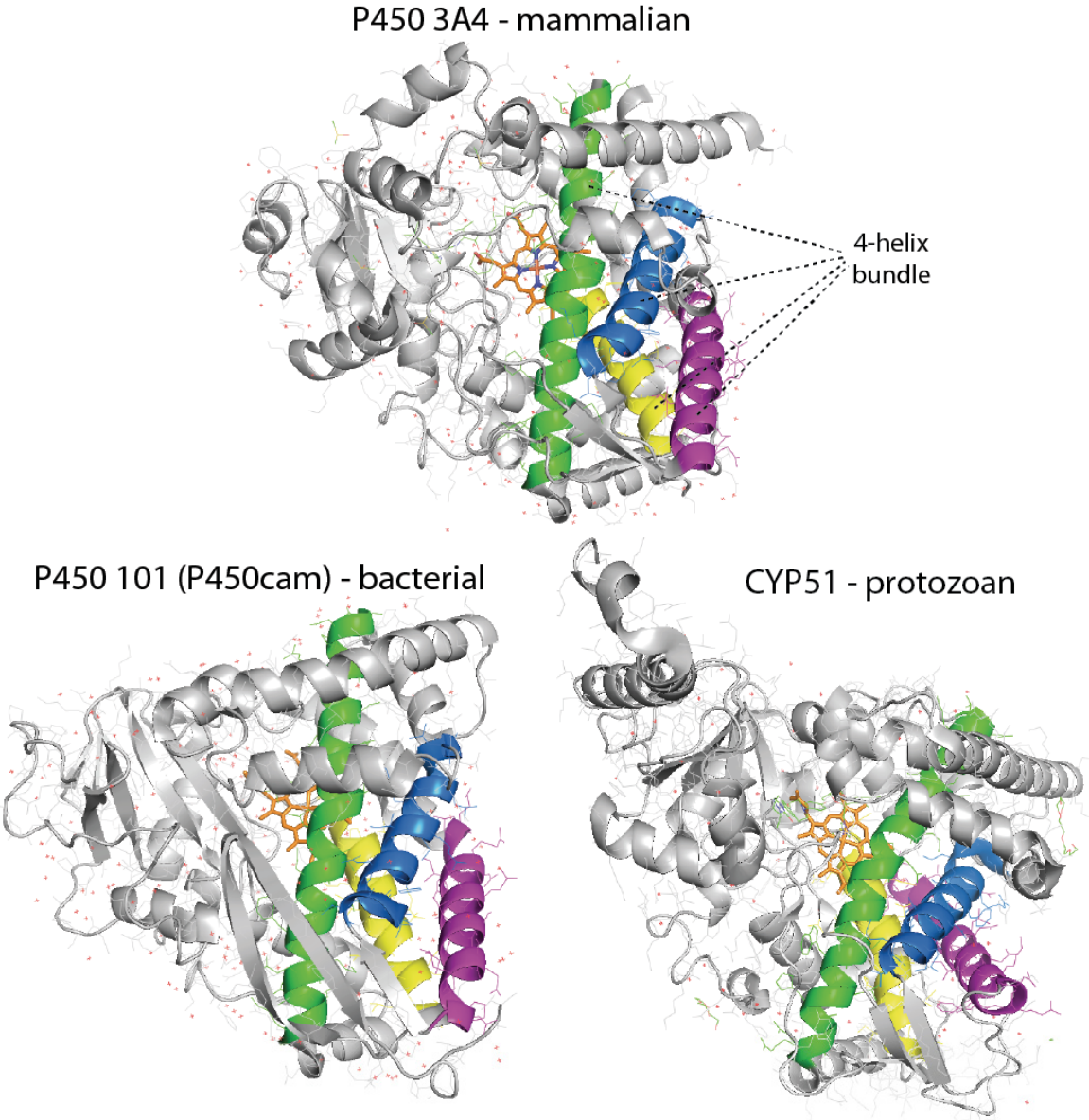


Figure 1.20 Structural similarities among P450s from different domains of life. The four-helix bundle is shown in colors.

Compared to other P450 structures, the F-helix of P450 3A4 is markedly short and it is followed by an ordered polypeptide stretch that does not conform to any secondary structure motif[90-91]. This stretch contains several amino acid residues that play a direct or indirect role in P450 3A4 function. Leu210, Leu211 and Asp214, which are implicated in the regio- and stereoselectivity of P450 3A4[92-93], are situated in this region.

Another characteristic feature of P450 3A4 is the "phenylalanine cluster", located above the active site and formed by residues Phe213, Phe215, Phe108, Phe219, Phe220, Phe241, and Phe304 (Figure 1.21). The aromatic side chains of these residues form a highly ordered hydrophobic core, which contributes to the high affinity of the enzyme for greasy molecules.

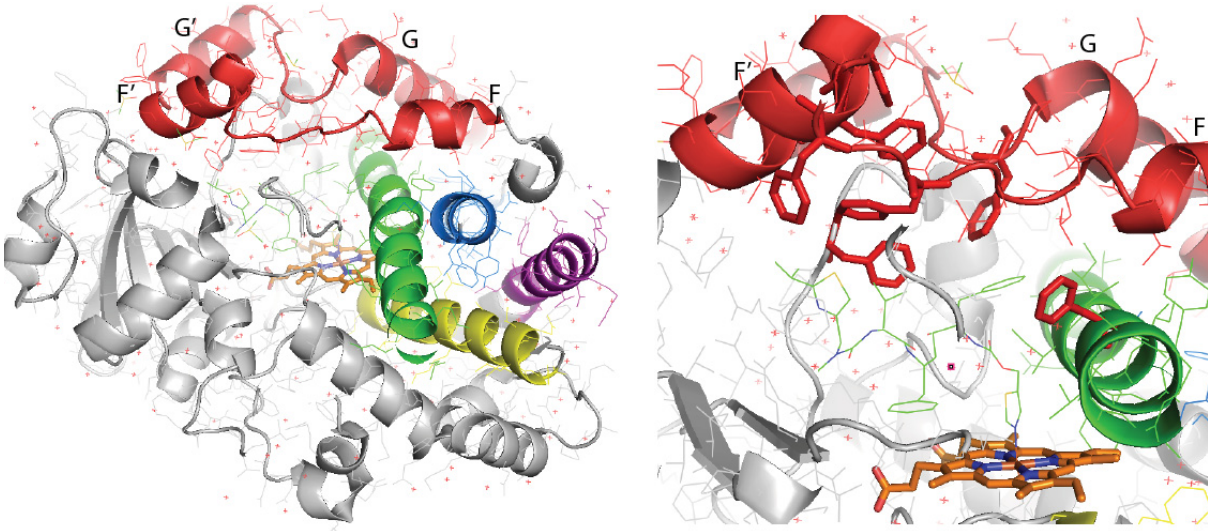


Figure 1.21 P450 3A4 structural features. In red, the F-F'-G-G'-sequence of α-helices believed to play a key role in substrate binding. On the right, the phenylalanine cluster is highlighted in sticks.

An important feature is the accessibility to the heme by substrates of P450 3A4, which is greater than in, for instance, P450 2C9[90]. This allows P450 3A4 to accommodate large ligands such as erythromycin[94] (Figure 1.22). The total solvent-accessible volume of the active site undergoes large changes depending on the presence and identity of the ligand. The volume in the ligand-free structures was estimated to amount to ≈ 950 Å^3. In P450 3A4 complexes with ligands, the volume increases to 1650 Å^3 for erythromycin (PDB code 2J0D) and to ≈ 2000 Å^3 in the presence of two ketoconazole molecules in the binding pocket (PDB code 2V0M)[94].

As a consequence of its large flexible active site, P450 3A4 shows exceptional substrate-promiscuity. This makes P450 3A4 a key player in drug metabolism and as a potential biocatalyst. A representative sample of reactions catalyzed by P450 3A4 is given below (Figure 1.22).

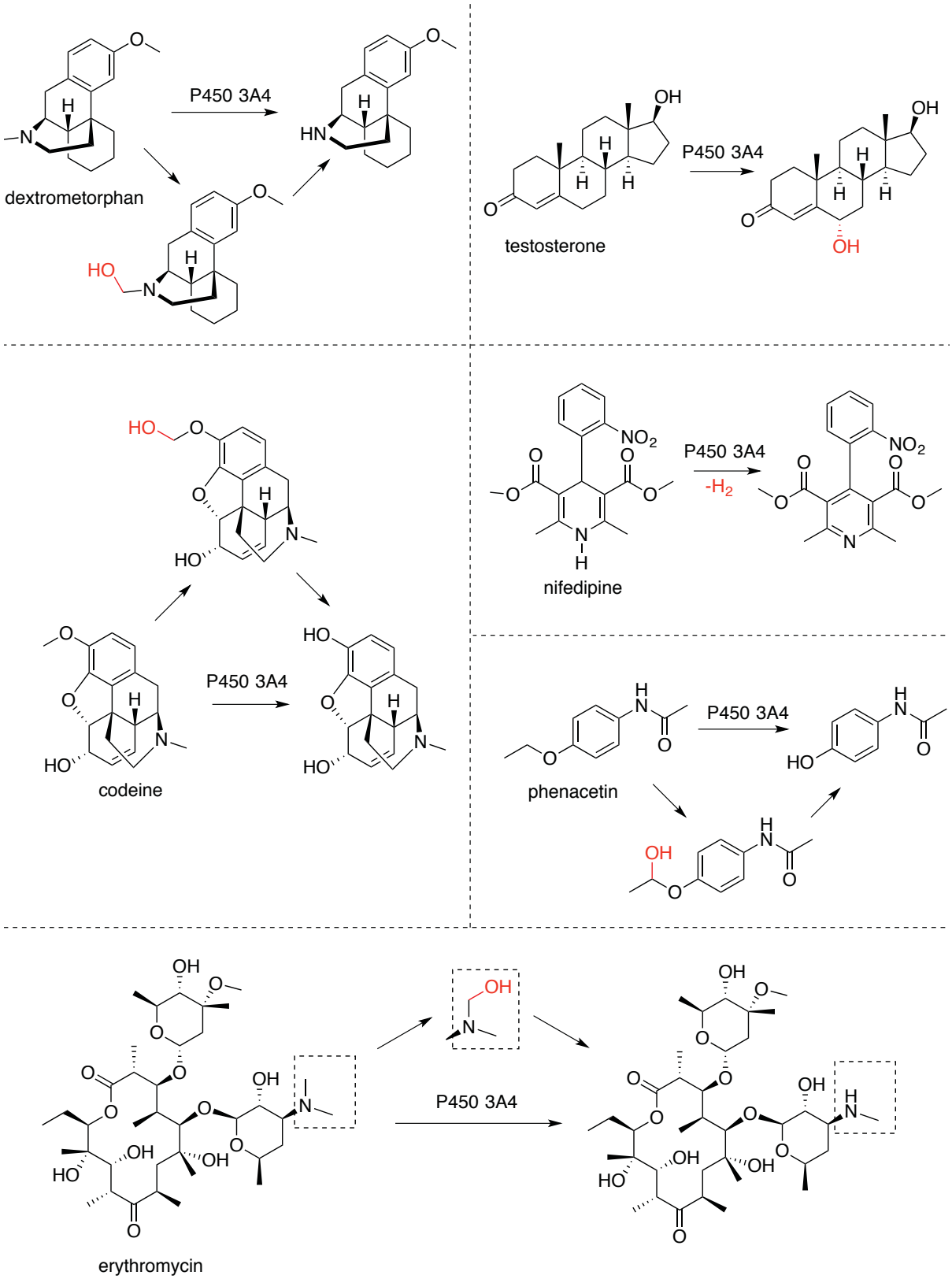


Figure 1.22 Chemical diversity of P450 3A4 substrates and reactions

1.9.1 Mutagenesis of P450 3A4

The literature contains a large amount of data on P450 3A4 engineering. With a few exceptions, such as the use of directed evolution by Kumar *et al*. to enhance peroxide-supported metabolism of several substrates[95], the majority of the mutagenesis studies have been conducted by targeting specific amino acid residues in the enzyme active site (Figure 1.23). For example, Halpert *et al*. published a series of studies that illustrate the role of active site residues in determining enzyme specificity towards different substrates. Alanine, phenylalanine and tryptophan scanning revealed the importance of residues Leu210, Leu211, Ile369, Ala370, and Leu373 in determining substrate specificity and regioselectivity of oxidation towards testosterone and progesterone[93, 96]. The same residues, and neighboring others such as Phe107, Ser119, Ile301, Phe304, Thr309 were also showed to play an important role in the regioselectivity of diazepam and midazolam oxidation[97 92]. In a different study, Leu211 was mutated into a phenylalanine or tyrosine to investigate the effect of active site volume modification on the potency of a series of inhibitors[98].

All these studies reveal how small changes in the active site of P450 3A4 can have a dramatic impact on its selectivity and confirm the potential of using P450 3A4 as a starting point for the development of biocatalysts for desired applications. The mutagenesis data reported in the literature was used to guide the remodeling of the active site of P450 3A4 discussed in Chapter 4.

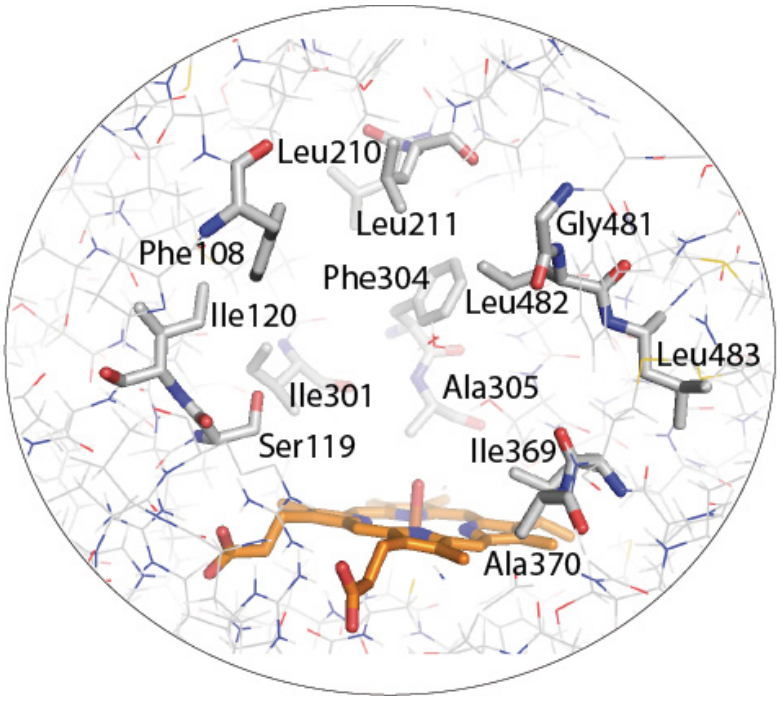


Figure 1.23 Active site of P450 3A4 (PDB 3NXU)

1.10 FORECASTER

Our research group has an established expertise in the development of software for (bio)organic chemistry[183]. Over the years, a suite of programs for drug design, predictions of drug metabolism and protein engineering has been reported and is now accessible through FORECASTER[112], an integrated, web-based platform already used by several academic groups around the world. Relying on this expertise, the use of computational tools to design or rationalize experiments is a key component of every project described in this thesis.

A feature of some of the programs within FORECASTER is the use of MM principles for the prediction of properties of chemical systems. Rather than utilizing complex equations derived from quantum physics, MM uses "force fields" for calculating the potential energy of a system. A force field can be defined as a set of parameters for different types of atoms, chemical bonds, angles and interactions (covalent, van Der Waals, ionic, etc.) that are derived empirically and used to feed equations derived from classical (Newtonian) physics.

The most commonly used of the FORECASTER suite of programs is FITTED[184], a docking program for structure-based drug design that stands out for its ability to consider protein flexibility and the presence of water molecules. In addition to having been validated for the study of kainate receptors, discussed in chapter 2, it has been applied for the identification of potential HCV polymerase inhibitors[185] and for the virtual screening of covalent POP inhibitors[183]. In an independent comparative study using six proteins, FITTED outperformed 15 other methods proving to be a valuable tool to investigate protein flexibility[186].

IMPACTS is a different program developed for the prediction of sites of metabolism by P450 enzymes[141]. It models the transition state (TS) of substrates reacting with a P450 by combining a docking component and a reactivity-based score that ranks the likelihood of different functional groups (sulphur atoms, double bond, etc.) to undergo oxidation. Through MM principles, the geometry of the TS is then evaluated. IMPACTS has shown to outperform medicinal chemistry experts by 6-7 % and an example of its predicting power will be described in Chapter 5.

One of the latest additions to FORECASTER is a tool for the prediction of side-chain conformation of proteins upon single-point mutation[113]. This program was developed by combining MM principles and statistical data from a library of thousands of PDB structures. Predictions are carried out by considering different rotamers of an amino acid residue and calculating the potential energy based on the interactions with surrounding residues. When applied to a library of mutant kinases, the program showed prediction accuracy of approximately 2 Å (RMSD metric).

1.11 Objectives of this thesis

The goal of the research presented in this thesis was to investigate the relationship between structure and function of proteins in a series of relevant case studies, using different approaches.

First, the structural flexibility of a glutamate-type receptor, GluK2, was investigated through the use of small molecules as probes. The goal was to demonstrate how the process of drug design can benefit from the understanding of the conformational dynamics of a receptor upon small molecule binding.

The focus was then shifted to enzymes, in particular P450s. Protein engineering was used to understand and modify the catalytic properties of human P450 3A4. Mutagenesis of the active site was performed to gain insight into the role of certain amino acid residues, with the final goal of developing novel biocatalysts for the synthesis of pharmaceutically relevant molecules.

In the first approach to engineer P450 3A4, a small number of mutations were rationally designed with the use of computational tools, in the attempts of creating a biocatalyst for the synthesis of the drug latanaprost. Mutagenesis was guided by looking at specific ligand-protein interactions, e.g. hydrogen bonds, phenyl-phenyl stacking. In parallel, all data acquired was shared with colleagues as a training set for the development of a new software for rational engineering of P450s. In the second approach, a semi-rational strategy was used for the remodelling of the active site of P450 3A4 in order to alter its regio- and stereoselectivity. In particular, the goal was to investigate the effect of "crowding" of the active site, i.e. reduction of its volume by deliberately mutating small residues to bulkier ones. General guidelines for this strategy were established and implemented experimentally towards the synthesis of useful hydroxylated products.

Finally, the unique features of P450s were studied in the context of drug metabolism. A series of sulphur-containing inhibitors of prolyl oligopeptidase (POP) demonstrated an unexpected chemical behaviour following P450 oxidation. Analytical and computational techniques were used with the goal of developing a strategy for the structural characterization of a mixture of rapidly interconverting P450 metabolites.

CHAPTER 2 - USE OF SMALL MOLECULES TO PROBE THE CONFORMATIONAL FLEXIBILITY OF KAINATE RECEPTOR GLUK2

2.1 Preface

Understanding protein flexibility is one of the main challenges of protein science today.
Since crystal structures fail at providing a true representation of protein dynamics, many efforts have been directed towards the development of computational tools to predict protein motion based on a number of parameters, e.g. temperature, pH and interactions with other molecules. Despite the advances in computer science, simulating protein motion in real time is often extremely time consuming, even for supercomputers. Alternative techniques, such as flexible docking, can represent a good compromise in terms of predictive accuracy and computational time.
Among cases of protein flexibility that have drawn the attention of the scientific community, ionotropic glutamate receptors (iGluRs) are very interesting because of the fundamental role that these proteins play in the central nervous system (CNS). Kainate-type receptors (KARs) are the least understood subtypes of iGluRs, yet recent evidence has linked them to neurodegenerative states such as schizophrenia and epilepsy[99]. Although KARs represent a promising pharmacological target, the limited understanding of the structural basis of receptor activation has hindered the development of new therapeutics.
This study targets the relationship between the activity of newly designed ligands and the conformational state induced at the receptor level. By relying on our docking tool FITTED, rational design was used to investigate particular interactions between the ligand and key amino acid residues in the binding site of the receptor. A small library of ligands was prepared and tested against KAR subtype GluK2. I designed and synthesized all the compounds, whereas the electrophysiological recordings were carried out by Brent Dawe, as part of a collaboration with Derek Bowie's research group (McGill department of pharmacology).
The results were published in the following article: Schiavini, P.; Dawe, B.; Bowie, D.; Moitessier, N., Discovery of novel small-molecule antagonists for GluK2, *Bioorg. Med. Chem. Lett.* (**2015**), 25, 11, 2416-2420.

2.2 Introduction

Over the years, our group has reported the development of several pieces of software for computer-aided drug design. Among these programs, FITTED can dock ligands to flexible proteins and can therefore predict not only the binding mode of the ligand but also the most favored conformation of the protein. In particular, a number of known protein conformations are given as input files, and the software ranks the likelihood of those conformations induced by a given ligand.

The capabilities of FITTED are particularly relevant to the study of KARs because of the critical role that protein flexibility plays in determining the mechanism of action of these receptors, as discussed in Chapter 1. The crystallographic studies reported for KARs provide critical information for the application of FITTED to these protein systems. In particular, insights into the flexible properties of KARs have been obtained by resolution of the crystal structures[100] of KAR subunit GluK2 in complex with either L-glutamate (**2.1**, Figure 2.1), kainate (**2.2**) and domoate (**2.3**). While the amino acid L-glutamate is the natural agonist of KARs, kainate and domoate, neurotoxins found in seaweed, have shown to act as partial agonists for this type of receptors. Careful examination of the crystal structures of GluK2 bound to **2.1**, **2.2** and **2.3** has revealed different conformational states of the agonist binding domain (ABD): a 'closed' state for L-glutamate, an "intermediate" state for kainate and an "open" state for domoate (Figure 2.1).

In a previous study by our group[101], Fay *et al*. have looked closely at these three states and they were able to correlate the degree of opening of the GluK2 ABD to the translational movement of a tyrosine residue (Tyr488) in close proximity to the ligands bound (Figure 2.1). Based on this observation, it was hypothesized that the clash between the Tyr488 and a bulky group on the nearby ligand could be responsible for hindering full closure of the ABD. In the same study, the relationship between closure of the ABD and agonist efficacy in GluK2 had been investigated. In particular, a library of ligands exhibiting a wide range of agonist behaviour towards GluK2 were docked using FITTED, to predict the most favorable conformational state induced by each ligand[101]. Unexpectedly, all partial agonists were predicted to bind to the closed state of the receptor, in analogy to the natural agonist L-glutamate. From these results, the role of an "open" conformational state of the GluK2 ABD, and its relationship with the receptor activation, remained to be clarified. Thus, we hypothesized that small molecular probes could be rationally designed to induce opening of the ABD and that, through electrophysiology studies, we could correlate such conformational state to ligand efficacy.

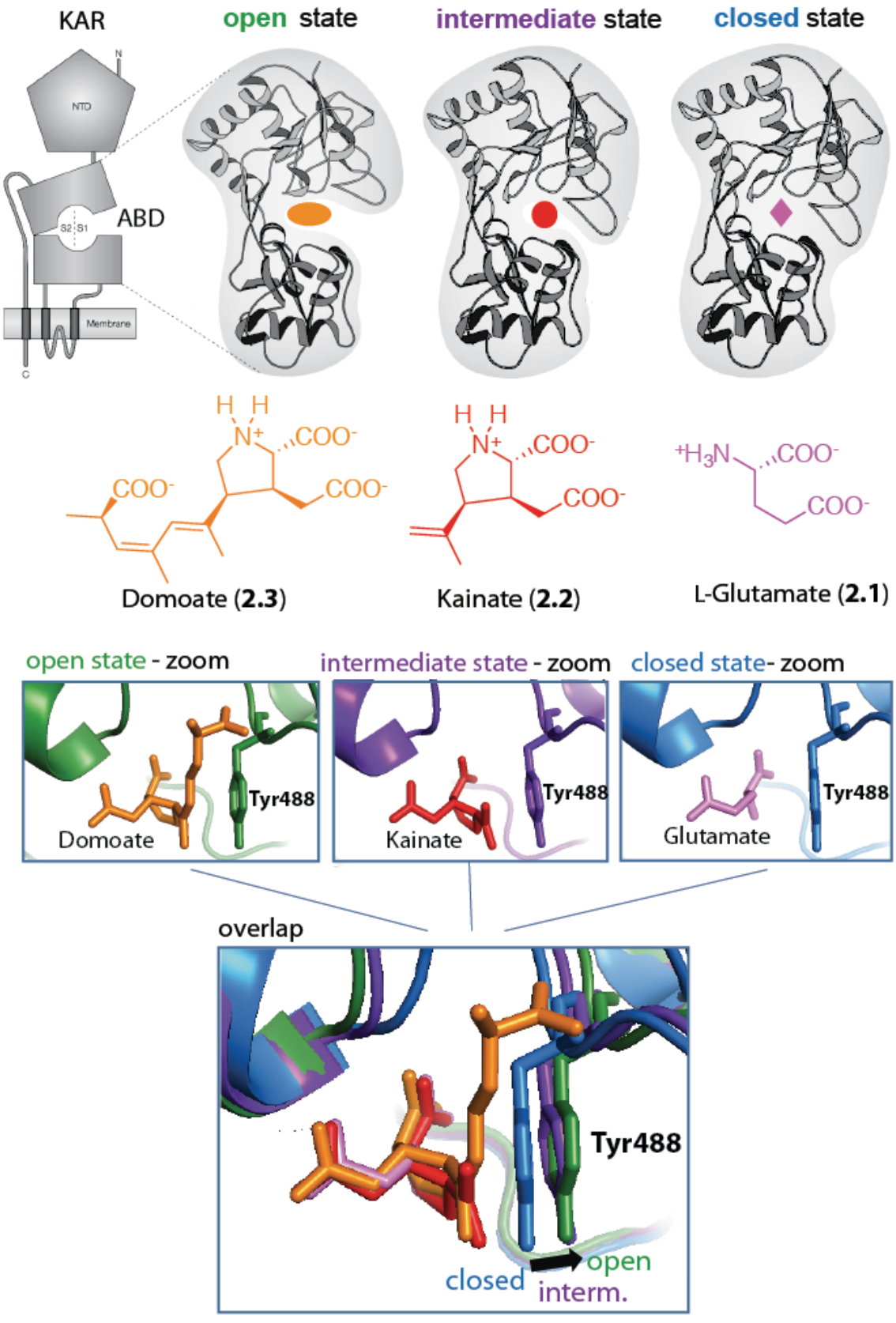
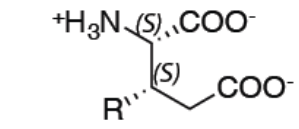


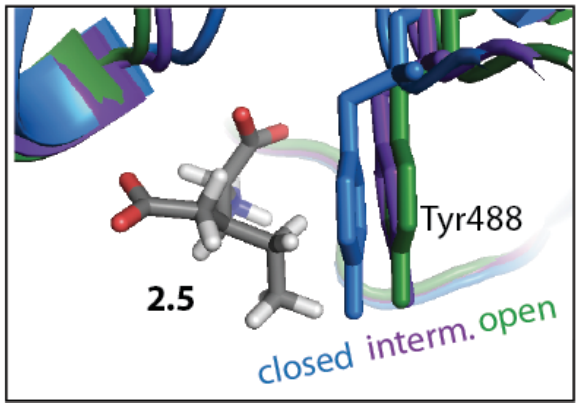
Figure 2.1 Different conformational states of GluK2 ABD that appear from the crystal structures of the complexes GluK2-glutamate (PDB 1S7Y), GluK2-kainate (PDB 1TT1) and GluK2-domoate (PDB 1YAE). At the bottom, the overlap of the three crystal structures shows the translational movement of Tyr488 in going from a closed state to intermediate and open state of GluK2 ABD.

2.3 Rational design of molecular probes

Following the previous work, we wanted to investigate further the effect of ligands that are specifically designed to constrain GluK2 ABD in an intermediate or open state (similar to each other but significantly different from the closed state) through a predicted interaction with the aforementioned Tyr488. In line with traditional structure-based drug design, we started from the visual inspection of the available crystal structures of GluK2. When GluK2 ABD is bound to L-glutamate, it appears that a substituent at the β position of the amino acid would point directly towards Tyr488 (Figure 2.2). To probe this hypothesis, we first considered the introduction of alkyl groups. We then relied on the FITTED software to predict which conformational state was favored upon binding of each new potential ligand. Interestingly, the stereochemistry at the β carbon appeared to play an important role in the outcome of the docking. It was found that all (3*S*)-β-substituted-glutamates, regardless of the size of the β-substituent, were predicted to induce an intermediate or open state. On the contrary, (3*R*)-analogues showed a less consistent pattern, with some poses being more similar to the closed state. Since our interest was to probe intermediate/open states, compounds with a (3*S*) stereochemistry were carried forward. Among them, five were synthesized (Figure 2.2).



R = Me, Et, Pr, iBu, Ph



Compound	R	Predicted conformation
L-Glutamate (**2.1**)	H	Closed
2.4	Me	Intermediate
2.5	Et	Intermediate
2.6	Pr	Open
2.7	iBu	Open
2.8	Ph	Open

Figure 2.2 List of compounds designed to study the conformational flexibility of GluK2. All compounds are predicted by FITTED to induce either an intermediate or an open conformational state. The 3D picture shows the predicted binding pose of **2.5** in proximity to the Tyr488 in the three conformational states overlapped

2.4 Synthesis of compounds 2.4-2.8

For the synthesis we relied on a procedure that was developed by Wehbe *et al.*[102]. This methodology relies on the use of a chiral auxiliary, (1*R*,2*R*,5*R*)-(+)-2-hydroxy-3-pinanone (**2.9**), for installment of the (*S*) stereochemistry at the α carbon on the glutamate scaffold (Figure 2.3). In particular, the chiral auxiliary, in its ketone form, is condensed with glycine *tert*-butyl ester (**2.10**) to give the imine intermediate **2.11** (Figure 2.4). The following step is a Michael addition of the conjugated base of **2.11** on a α,β-unsaturated ester (**2.12**-**2.16**) that carries the substituent R. This reaction is characterized by low stereoselectivity at the β carbon (d.r. ratios given in Table 2-1), however the diastereomeric products can be separated by column chromatography to give the desired intermediate (**2.17**-**2.21).** Finally, acid hydrolysis yields the final compound (**2.4**-**2.8**).

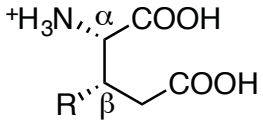


Figure 2.3 (3*S*)-β-substituted-glutamates

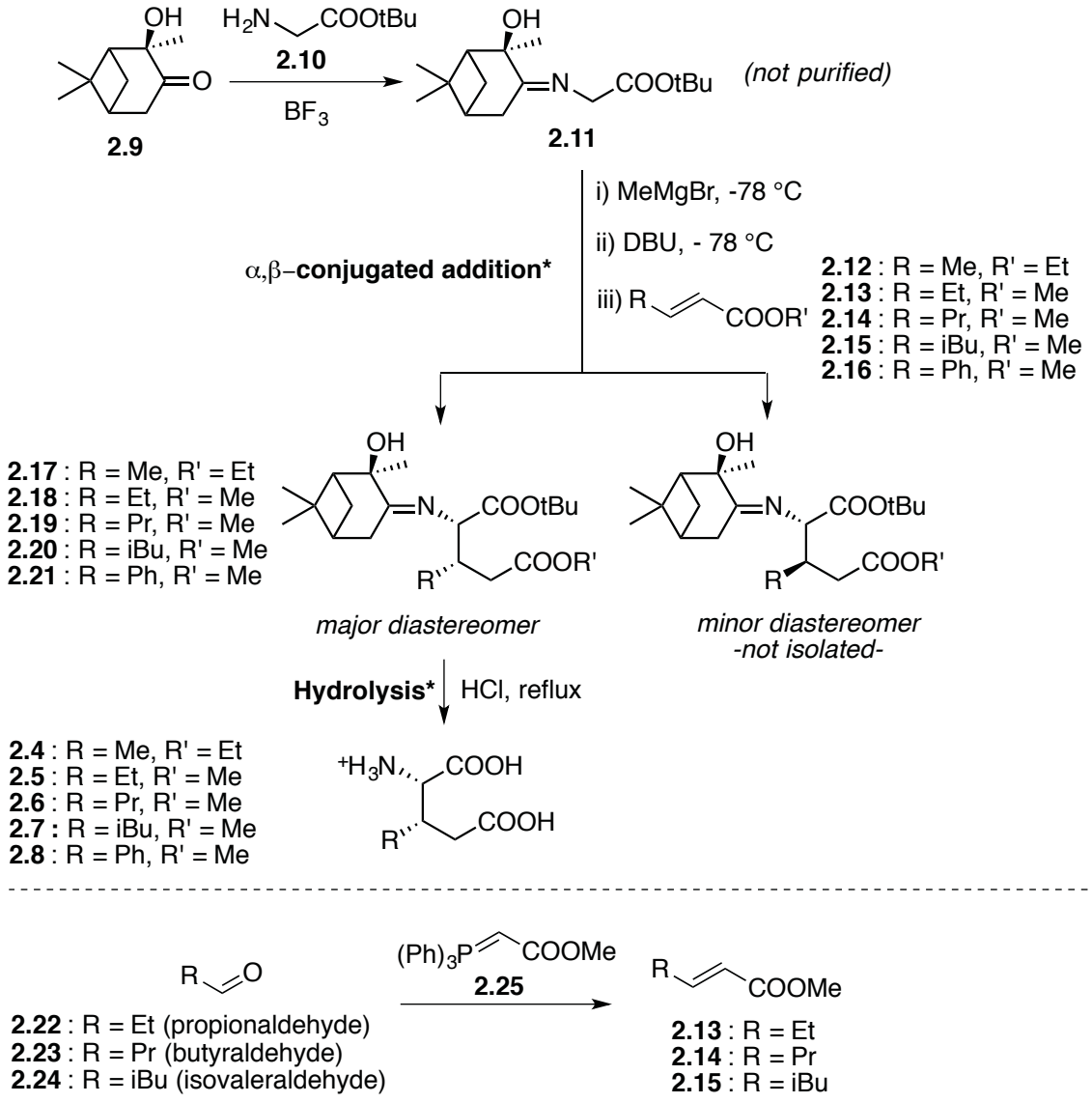


Figure 2.4 Synthesis of **2.4-2.8**

Table 2-1 Diastereomeric ratios and reaction yields for the synthesis of **2.4-2.8**

		α,β-conjugated addition*		Hydrolysis*
Final compound	**R**	**d.r (NMR)**	**Yield (major)**	**yield**
2.4	Me	6:4	34	89
2.5	Et	6:4	40	85
2.6	Pr	6:4	40	75
2.7	*i*Bu	6:4	41	87
2.8	Ph	6:4	42	96

Our attempts to introduce secondary or tertiary carbons at the β position (e.g. isopropyl, cyclopropyl, *tert*-butyl) were unproductive. Interestingly, with such bulky groups, the α,β-conjugated addition appeared to be working (data not shown), but hydrolysis of the protected intermediate did not yield the desired product. We believe this can be attributed to the high tendency of these β-glutamates to cyclize to pyroglutamatic acids in the conditions used for the experiment, a phenomenon reported by Wehbe for methyl analogues (Figure 2.5).

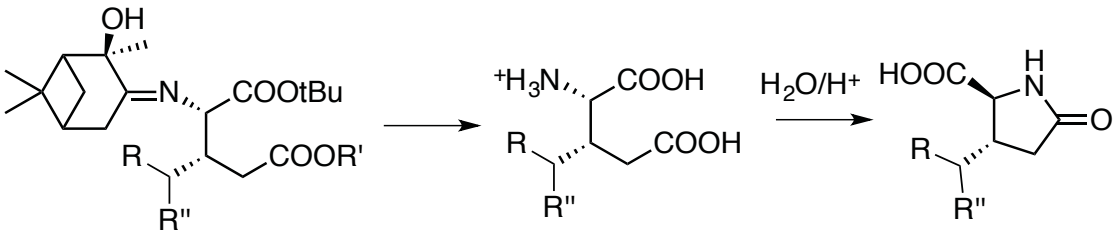


Figure 2.5 Proposed reaction of cyclization for β-substituted glutamic acids

2.5 Biological studies

The studies described in this paragraph were carried out by Brent Dawe, as part of a collaboration with Derek Bowie's research group at McGill University.

The five synthesized compounds were tested in a functional assay to determine whether they could competitively bind to GluK2, thereby interfering with L-glutamate binding and the ability of the receptor to be activated (Figure 2.6). Briefly, a solution containing GluK2-expressing cells (cell line HEK293T) was exposed to L-glutamate and the current generated by the activation of the receptor was measured and set to 100 %. The presence of a GluK2 antagonists in the system caused a reduction in the generated current, which was used to estimate the potency of the antagonist.

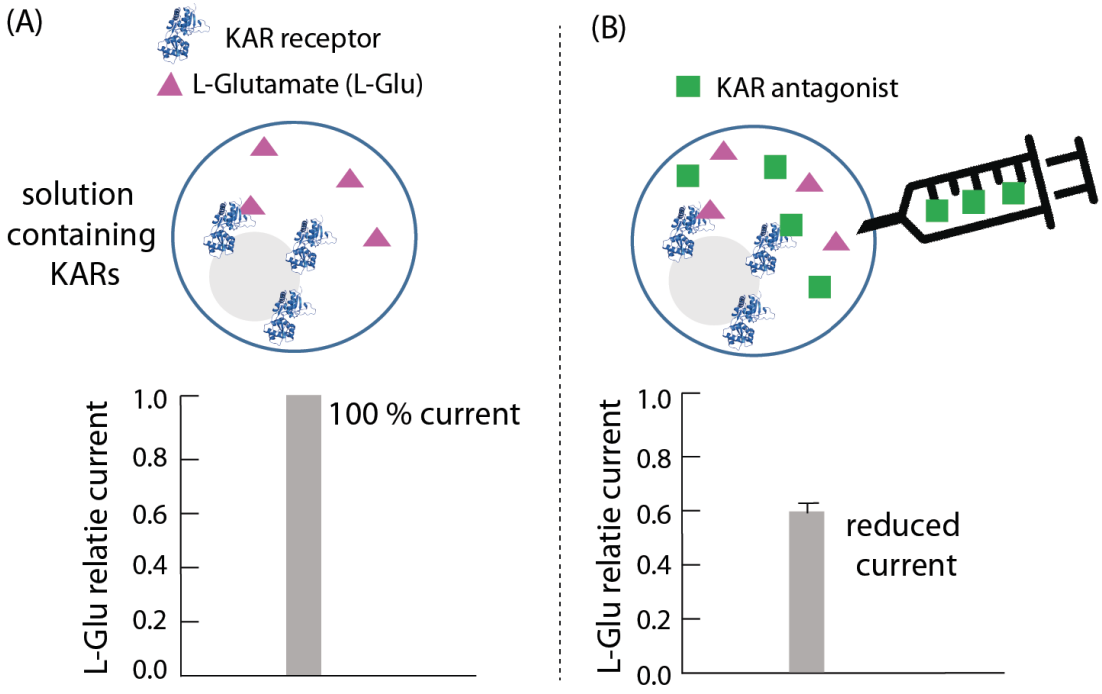


Figure 2.6 Schematic representation of the electrophysiological assay used to study the antagonist behaviour of compounds **2.4-2.8**

The antagonist behaviour of compounds **2.4-2.8** was assessed by measuring the reduction in GluK2 peak current response. It was observed that the first two members of the series, **2.4** and **2.5**, when used at 1 mM concentration, inhibited the response of GluK2 to saturating (10 mM) L-glutamate to 38 ± 3 % (n = 4) and 46 ± 6 % (n = 4) of its maximum amplitude (Figure 2.7A-B). Compound **2.6**, when used at the same concentration, slightly inhibited L-glutamate responses to 95 ± 3 % of their original amplitude (n = 4) (Figure 2.7C). The last two bulkier compounds, **2.7** and **2.8**, did not affect the current response to glutamate under the same conditions (Figure 2.7E-F), indicative of reduced (or minimal) affinity for the receptor, relative to their counterparts with smaller substituent groups.

Due to time restrictions, no experiment was performed to confirm that the molecules under study act as competitive antagonists, however their high structural similarity to the natural agonist strongly suggests this is the case. This hypothesis is further supported by the observation that the inhibitory activity decreases with a clear trend in going from **2.4**, which differs L-glutamate by only a methyl group, to **2.5** and **2.6** which are less alike analogues.

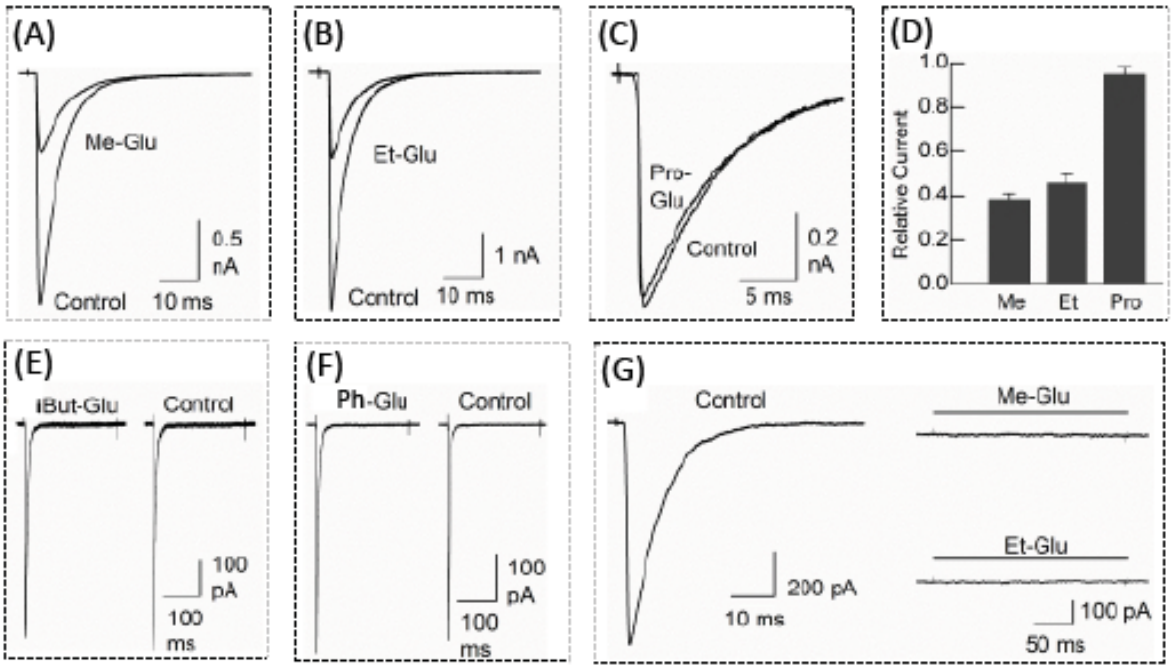


Figure 2.7 Electrophysiological assays for the study of the agonist/antagonist behaviour of **2.4**-**2.8** towards GluK2. Inhibition of GluK2 current responses to 10 mM L-glutamate measured for **2.4** (A), **2.5** (B) **2.6** (C). (D) Summary of inhibition produced by compounds **2.4-2.6**, shown relative to the L-glutamate induced current measured following a return to the control background solution. (E) In contrast, **2.7** and **2.8** did not appear to inhibit responses to 10 mM L-glutamate. (G) When GluK2-expressing patches were exposed to 1 mM **2.4** (n = 5) and **2.5** (n = 4), no current responses were detected, despite the same patches having displayed robust responses to 10 mM L-glutamate (control). Details on the described biological assays can be found in the experimental section of this thesis.

Other trials aiming at measuring agonist behaviour towards GluK2 receptors, indicated that all the compounds lacked detectable agonist activity (Figure 2.7 G).

In summary, from these results it appears that that compounds **2.4**, **2.5** and **2.6** were acting as antagonists and that these are likely to compete with the natural ligand for occupancy of the agonist binding cleft.

The lack of activity for **2.7** and **2.8** can be rationalized based on the size of the R substituent for these β-substituted-glutamates. It is likely that increasing the size of R beyond a propyl group induces significant clashes between the protein and the ligand that is no longer able to compete with the natural ligand under the conditions of the experiments.

This outcome could not have been predicted by FITTED, which at the current stage of development cannot provide an estimate of the energy of binding, let alone K_d values.

2.6 Discussion

The literature contains several reports that can be helpful to rationalize the data. Mayer and colleagues provided insight into the structural basis of KA receptor competitive antagonism by crystallizing GluK1 ABD complexes with novel antagonists. The authors showed that UBP301, UBP318 (**2.26**, **2.27**, Figure 2.8) induced a similar, generally large opening of the GluK1 agonist binding cleft in comparison to the closed conformation elicited by the full agonist glutamate.[100, 103] Although no structural work is available for GluK2 bound to the same or other antagonists, it is reasonable to assume that it behaves similarly to GluK1 from a structural point of view. The two subunits share ~80% sequence identity, with the most conserved region being around the glutamate binding site. When we look more closely at the available structures of GluK1 ABD bound to L-glutamate and UBP antagonists, and we align the structures, we notice that the tyrosine residue in the ligand proximity undergoes a significant shift (Figure 2.8). This indicates that there may be a direct relationship between antagonist behavior and an open conformational state induced in the ABD.

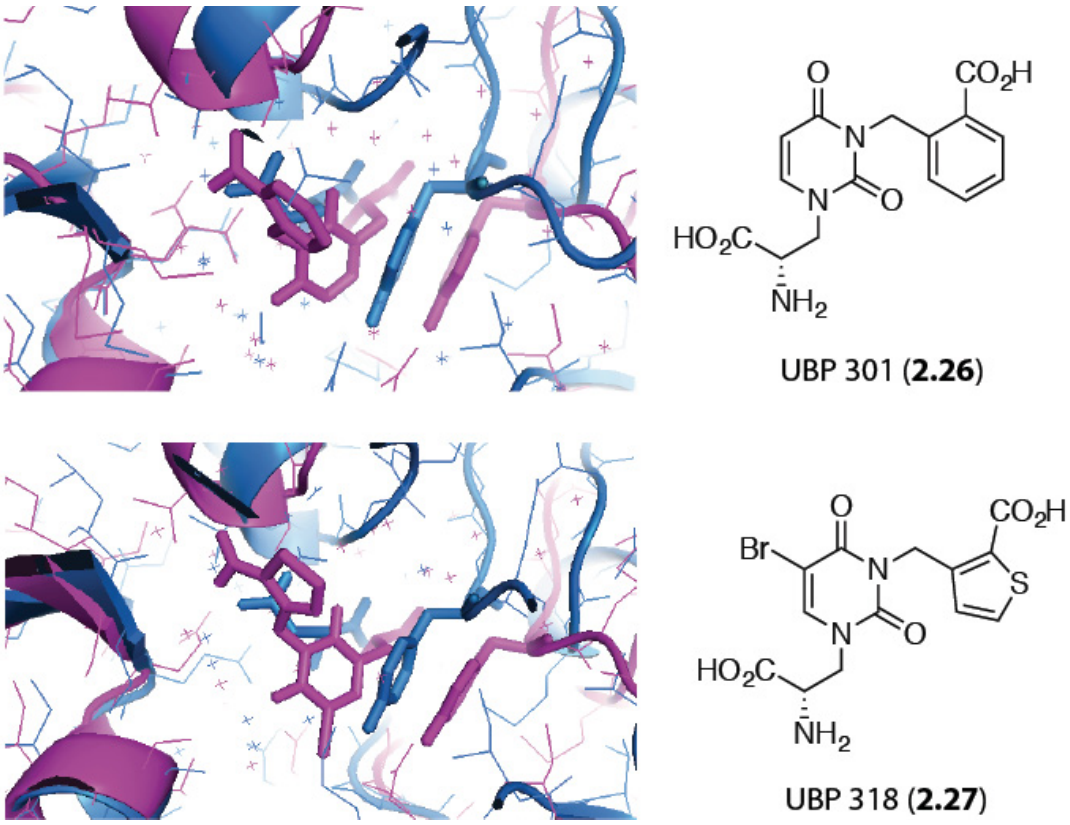


Figure 2.8 Docking of previously reported antagonists of GluK1.Superimposition of crystal structures of GluK1 ABD bound to glutamate (in blue) and to antagonist UBP318 (in magenta, PDB 2QS2); On the right, chemical structures of UBP302 and UBP318.

It must be emphasized that the extent of the movement of the aforementioned tyrosine is based on how structural alignments are performed. In our case, MATCH-UP, a piece of our computational platform FITTED, was used to superimpose the available structures of the ABD without specific structural constraints. In other words, no definitive parallels can be made between GluK1 and GluK2 in the absence of full-length crystal structures of the receptor bound to an antagonist. It is possible that in the receptor physiological state the positions of some binding site residues are more or less constrained than in the isolated ABD, therefore causing unpredicted movements.

Furthermore, based on our observations, it is perplexing that kainate (**2.2**) and domoate (**2.3**) (Figure 2.1) do not act as GluK2 antagonists, since they induce an intermediate and open conformation, respectively. Even though for most structures the trend remains 'greater closure, greater agonistic efficacy', no definitive connection can be made. Indeed, there are examples of antagonists inducing full closure in KAR structures[104], and of partial agonists inducing varying degrees of closure[105]. One must remember that crystal structures are only snapshots of the receptor at one point in time and do not display other conformations that may be entered less frequently but critical for functional behavior. KARs are thought to be very dynamic and one must take into account the weighted time the receptor spends in closed cleft versus open cleft conformations in regulating efficacy.

2.7 Proposed future work

The conflicting evidence of the structural features that determines what makes a ligand a KAR agonist or an antagonist could be addressed using chemical-physical techniques aimed at providing a more representative picture of the dynamics of KAR receptors. In this regard, one technique that is gaining momentum in drug discovery is Hydrogen/Deuterium Exchange Mass Spectrometry (HDX-MS)[187]. In HDX-MS, a protein is incubated into D_2O-reconstituted buffer for seconds to minutes, allowing exchangeable protons (such as those in the amide backbone of the protein) to exchange with deuterium. The reaction is stopped by lowering the pH, the protein is digested into individual peptides, and these are analyzed by mass spectrometry to assess the extent of exchange. By doing so, each peptide becomes a conformational probe allowing to identify which regions of the proteins are more solvent accessible or less structured (i.e. featuring a high degree of motion as a result of the absence of a strong hydrogen-bond network). By doing a comparative study between an apo-protein and the same protein in presence of a ligand, HDX-MS data can be mapped into a crystal structure to generate a heat-map of the regions more affected by ligand binding.

Since the binding of an agonist to a receptor causes in the latter conformational changes that are different from the ones induced by binding of an antagonist, these two different classes of ligands generate very distinctive HDX-MS signals. This possibility allows for studying protein dynamics and its relationship with ligand efficacy. This approach has been successfully used, for example, by Griffith *et al*. to study the molecular details of ligand activation of human β adrenergic G-protein coupled receptors (GPCRs)[188] or for conformational epitope mapping of antibody-antigen interactions[189].

In the case of KARs, HDX-MS could shed light on the importance of conformational changes that are far from the agonist binding region but still may be critical for the receptor activation/deactivation. These experiments could reveal different dynamics between different KAR subunits and provide, for example, an explanation of why the same ABD "open" state is associated with agonism in case of GluK1 and antagonism in the case of GluK2.

2.8 Conclusions

Despite the uncertainty that is always correlated with docking studies based on crystal structures, this study refines the existing knowledge about structure-activity relationships for KAR inhibition. Most of the antagonists reported so far have a molecular weight significantly higher than glutamate, which agrees with the theory that an effective antagonist must hinder full closure of the ABD acting like a jamming object. Our study reveals that smaller molecules could be as effective, if properly designed to interact with amino acid residues that play a key role in the the activation of the receptor.

We have also demonstrated how flexible docking can be used to study protein flexibility and overcome limitations that might arise with highly time-consuming protein dynamic simulations.

The limitations of relying on a limited number of crystal structures to study KAR dynamics have been addressed and suggestions on the use of complementary techniques, such as HDX-MS, have been put forward. As final remark, it is worth mentioning that as MS technologies progress, HDX-MS is moving towards the high-throughput screening of ligand-induced protein dynamics. It is hoped that this kind of HTS screenings, aimed at understanding the structural basis of ligand efficacy, rather than randomly fishing for activity, could be used in concomitance to flexible docking for the advancements of rational approaches in drug discovery and structure-based drug design.

CHAPTER 3 – TOWARDS THE RATIONAL ENGINEERING OF P450 3A4 FOR THE CHEMO-ENZYMATIC SYNTHESIS OF LATANOPROST

3.1 Preface

The rational design of new enzymes for biocatalysis is among the most exciting outcomes of the advances in protein science, however much research is still necessary for the development of accurate computational tools to predict the outcome of mutations.

The research efforts described in this chapter are part of a long-term project for the development of a new piece of software for the rational engineering of cytochrome P450 enzymes (P450s), based on a collaboration between experimental and computational chemists in the Auclair and Moitessier groups.

As discussed in the Chapter 1, P450s are a unique family of enzymes with remarkable potential in terms of biocatalytic applications. Human P450 3A4, characterized by a high degree of substrate promiscuity and structural plasticity, is a particularly promising starting point for the development of a versatile biocatalyst for hydroxylation reactions. Herein, we describe attempts to engineer P450 3A4 into a biocatalyst for the chemo-enzymatic synthesis of latanoprost, a widely used drug for the treatment of glaucoma. In particular, we envisaged to modify a recently reported synthesis of latanoprost by introducing a P450 3A4-catalyzed hydroxylation at a non-activated C-H bond. The desired intermediate to be used as the enzyme substrate was synthesized and fully characterized. By relying on a first version of the software for P450 biocatalysis, three mutants of P450 3A4 were designed and produced through recombinant technologies (expression in *E. coli*). A biochemical assay that does not require protein purification was developed and applied to the testing of each mutant as biocatalyst for the desired biotransformation. Unfortunately, none of the mutants showed the desired activity, a finding that exposed some of the limitations of the current software.

I designed and performed most of the experimental work with the help an undergraduate student, Lekha Puri, who contributed to the preparation of one of the key intermediates (compound **3.5**). The software used for the design of the mutations was previously developed by a former member of the Moitessier group (Joshua Pottel).

The results presented in this chapter are not published.

3.2 An interesting synthetic target: latanoprost

Latanoprost (**3.1**, full name: isopropyl-(*Z*)-7[(1*R*,2*R*,3*R*,5*S*)3,5-dihydroxy-2-[(3*R*)-3-hydroxy-5-phenylpentyl]cyclopentyl]-5-heptenoate) is a drug used to control the progression of glaucoma or ocular hypertension[106]. It is administered into the eye as a solution and it acts by increasing the outflow of aqueous fluid from the eyes. From a structural point of view, it is a prostaglandin analogue (Figure 3.1), more specifically an analogue of prostaglandin F2α (**3.2**). Latanoprost is an isopropyl ester prodrug, meaning that it is inactive until it is hydrolyzed by esterases in the cornea to the biologically active acid.

It is on the World Health Organization's List of Essential Medicines, with annual sales of ~ $1.6 billion. It is also known by the brand name of Xalatan manufactured by Pfizer.

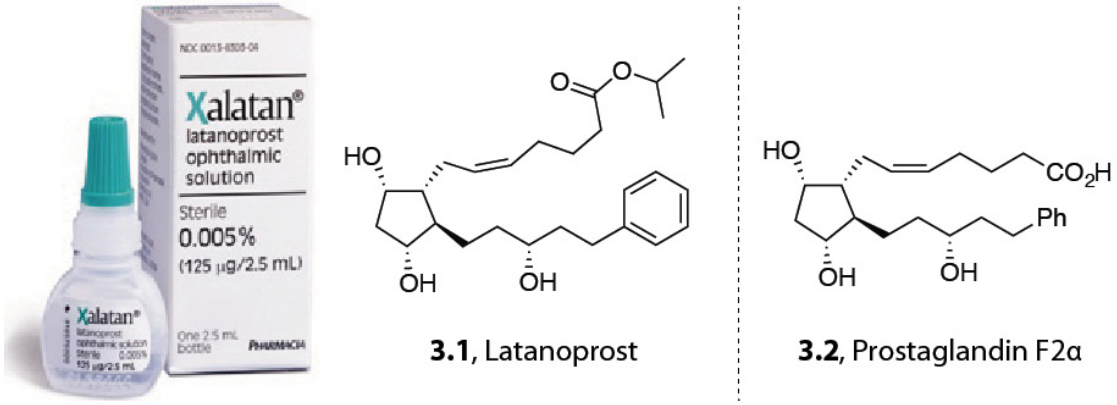


Figure 3.1 Structure of latanoprost (**3.1**), distributed as 0.005% solution by Pfizer, and prostaglandin F2α (**3.2**)

3.3 Current syntheses of latanoprost

Latanoprost is currently manufactured using a method developed by Corey in 1969[107], a process that requires 20 steps. In 2015, a shorter methodology (12 steps) was reported by Prevost *et al.*[2], following a publication describing a similar process for prostaglandin F2α[108-109]. In this new method (Figure 3.2), the first key step consists on the condensation of two molecules of succinaldehyde (**3.4**) in the presence of L-proline and dibenzylammonium trifluoroacetate. The resulting key intermediate **3.5** is a bicyclic hemiacetal containing an unsaturated aldehyde moiety and bearing two of the five final stereogenic centres. After oxidation of **3.5** to lactone **3.6**, the first of the two long side chains is introduced using mixed cuprate **3.7** as a nucleophile attacking at the beta position of the unsaturated aldehyde. The preparation of the mixed cuprate begins with the asymmetric iridium-catalyzed allylation reaction between allyl acetate **3.8** and 3-phenyl-1-propanol **3.9** (Figure 3.2, bottom), which

yields homoallylic alcohol **3.10** in 98:2 er. The alcohol is then protected as a silyl ether **3.11**, which undergoes ozonolysis to **3.12** and iodination to **3.13** before cupration to form **3.7**. This mixed cuprate is then used in a stereoselective conjugated addition resulting in the formation of an enolate which is immediately trapped as silylther **3.14**. Subsequent ozonolysis followed by addition of $NaBH_4$ yields the key lactone intermediate **3.15** with high stereoselectivity. The lactone is reduced back to hemiacetal **3.16**, which is finally subjected to a Wittig reaction with phosphonium salt **3.17**, thereby introducing the second long side chain and leading to the final compound **3.1** after esterification.

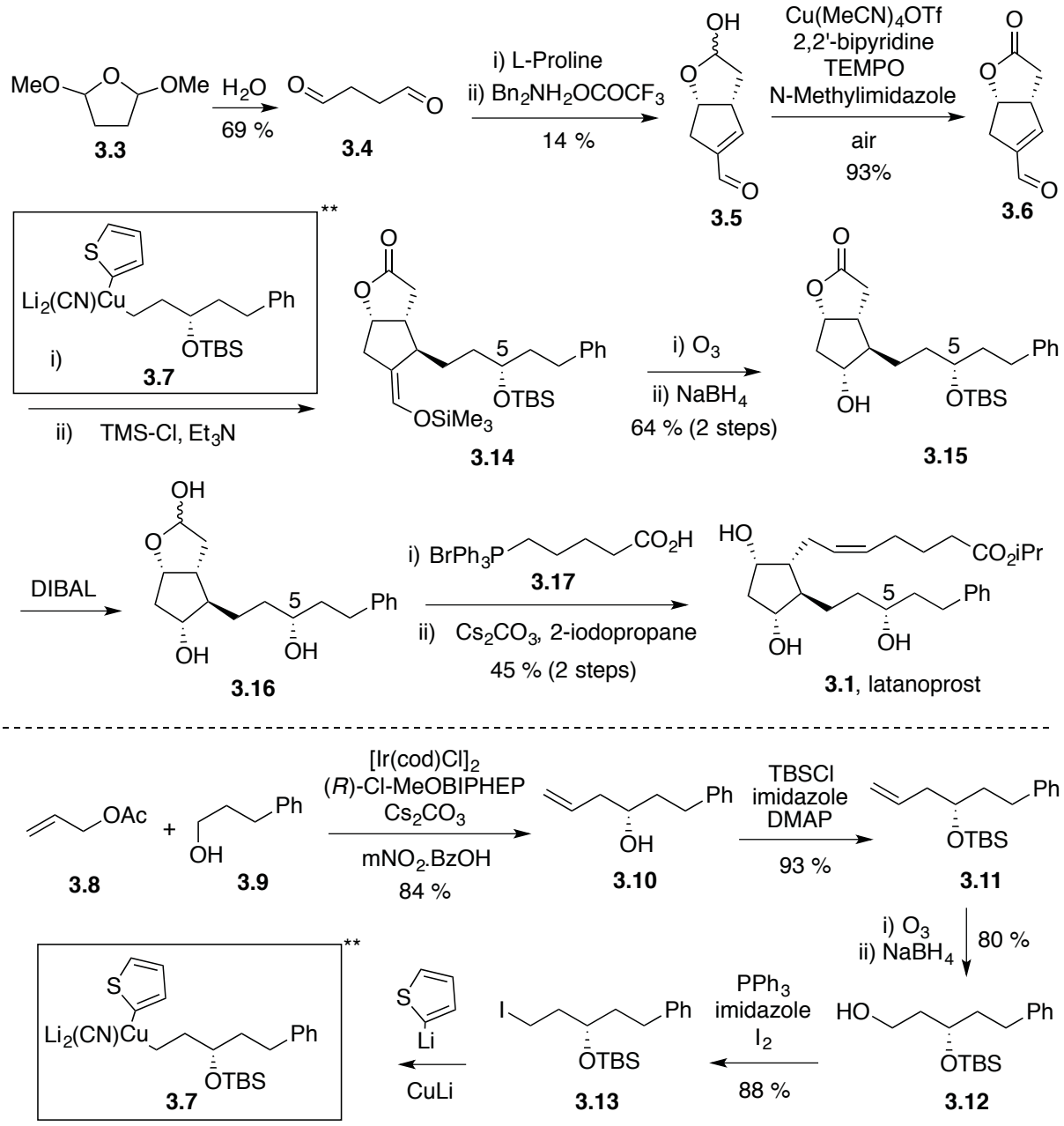


Figure 3.2 Synthesis of latanoprost by Prevost *et al.*[2]

3.4 Hypotheses on the use of P450s for the hydroxylation of a latanoprost precursor

The synthesis of latanoprost by Prevost *et al.*[2] has attracted our attention because it provides an excellent opportunity to investigate the potential of P450s to catalyse the regio- and stereoselective hydroxylation of unactivated C-H bonds towards the synthesis of a compound of pharmaceutical interest. In particular, it can be envisioned to utilize a P450 enzyme for hydroxylation at position C-5 of the bottom chain (Figure 3.3). In a first scenario, the enzymatic oxidation could be carried out as the final step in the synthesis, using compound **3.18** (Figure 3.3) as substrate for the enzyme. In a second possible route, P450 biocatalysis could be used at an earlier stage of the synthetic process, for the hydroxylation of compound **3.19**. In regard to this last option, it is interesting that the acidity of the hydroxyl group at C-5 does not prevent the following Wittig reaction in the original synthesis (Figure 3.2). In both envisioned chemo-enzymatic synthetic routes, fully illustrated in Figure 3.4, a total of two steps are saved compared with the original synthesis by Prevost *et al.*[2] Although the number of steps is not the sole parameter in evaluating the overall efficacy of a synthetic process, our study aims at demonstrating that biocatalysis can limit excessive derivatization and protection/deprotection steps, by leveraging the ability of enzymes to catalyse reactions with high regio- and stereoselectivity.

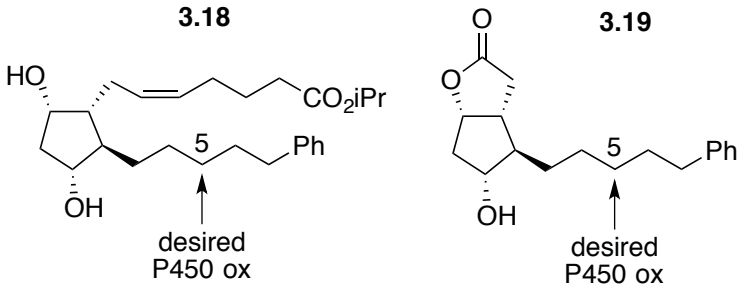


Figure 3.3 Structures of possible precursors in the proposed P450-based chemo-enzymatic synthesis of latanoprost

Between Route 1 and 2 (Figure 3.4), the former appears much easier to implement for several reasons. First, to successfully engineer an enzyme for a particular biotransformation, one must consider the structural features of the substrate. In our case, **3.19** contains several less C-H bonds than **3.18** and therefore presents less challenges in the development of a regio- and stereoselective reaction. Second, **3.18** contains both a double bond and an ester moiety, which are known to be easily targeted by P450 and therefore might compete as sites of oxidation. Finally, **3.18** is characterized by a higher conformational flexibility than **3.19**, due to the presence of two "floppy" side chains instead of only one. These chains contain many rotatable bonds that increase the number of possible binding poses inside the enzyme active site and consequently selectivity is more difficult to achieve.

For all these reasons, synthetic efforts were focused on the preparation of **3.19**.

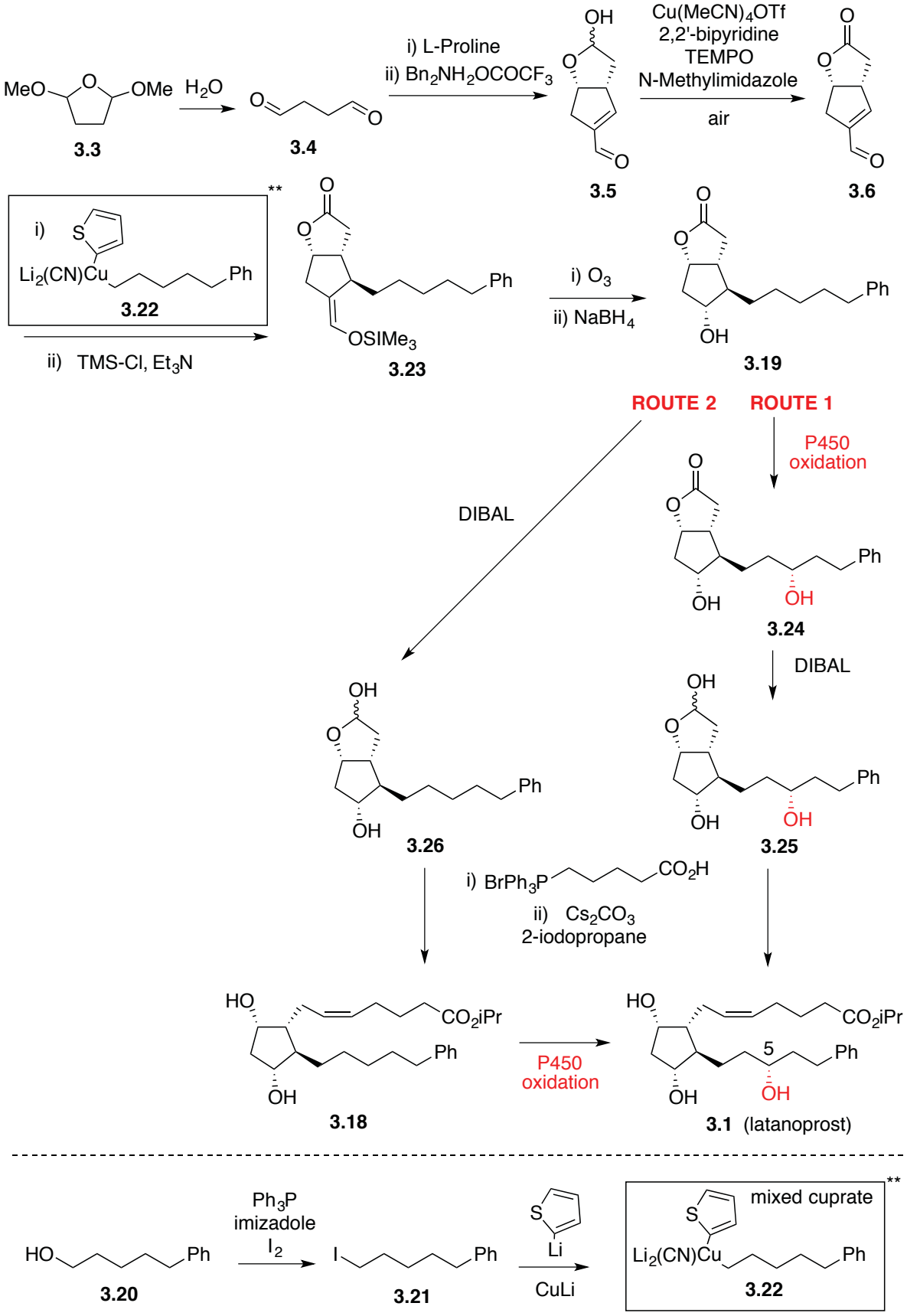


Figure 3.4 Proposed chemo-enzymatic synthesis of latanoprost based on the P450-catalyzed regio- and stereoselective hydroxylation of a non-activated C-H bond

3.5 Synthesis of compounds 3.19

Compound **3.19** was synthesized following Prevost's published procedures[2] (Figure 3.5). Although the protocols were followed accurately, some challenges were encountered. The reaction yield for the key step **3.4** → **3.5,** reported to be as low as 14 %, was found to be irreproducible. Over several attempts, we were only able to obtain yields in the range 2-4 %. During experiments designed to understand the reason of this irreproducibility, **3.5** appeared to be relatively stable, able to tolerate temperatures of ~ 60 °C, significantly higher than the ones used for the reaction work up, ~ 30°C. In addition, once purified, **3.5** was found to be stable at room temperature for several months. Instead, it is possible that the reaction **3.4** → **3.5** is more temperature-sensitive than what Prevost *et al.* have originally observed. In fact, the reaction was run at "room temperature" for 44 hours, however this is not a specific T value and it can very greatly depending on the laboratory conditions. As future directions, more controlled experiments set a particular temperature value may indicate the optimal conditions for both reaction yields and reproducibility.

In the following steps, deprotonation of thiophene, to be used as ligand for cuprate **3.22**, was

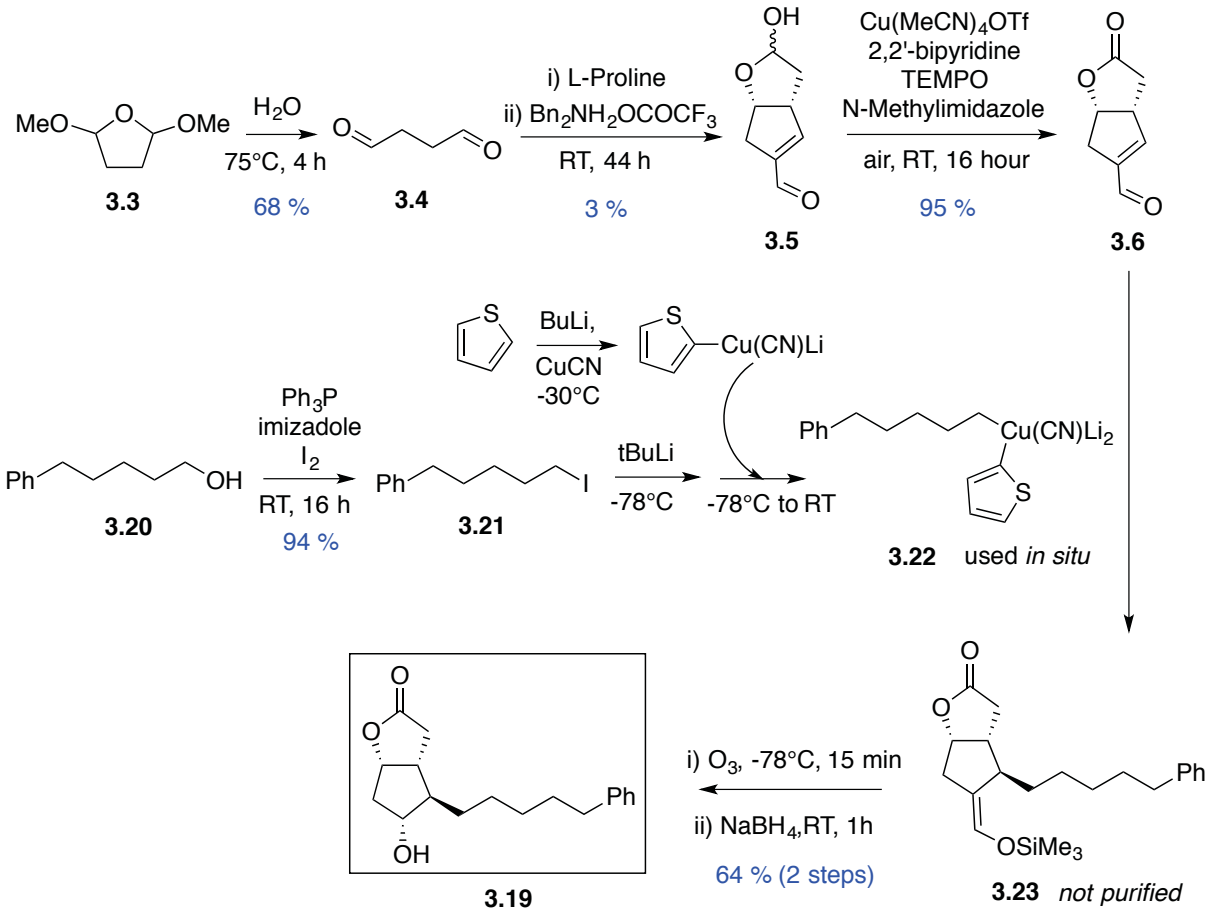


Figure 3.5 Synthesis of compound **3.19** (precursor to enzymatic oxidation)

found to be very sensitive to the quality of the solvent THF. A rapid blackening of the solution, after BuLi addition, was observed with dry THF obtained from different sources. We believe small impurities in the solvent can catalyse a rapid polimerization of thiophene, which we were able to avoid only by using > 99.9 % pure, freshly opened dry THF (from Sigma).

3.6 Transformation of 3.19 by wild-type P450 3A4

To find out if and how the substrate **3.19** is converted by the wild-type P450 3A4, a previously reported procedure for expression of P450 3A4 in *E. coli* and purification using a protein Histidine-Tag (His-Tag) and nickel chromatography was used[110]. Once characterized, the enzyme was incubated with compound **3.19** and the reaction followed by HPLC-MS.

Due to the challenges in producing large quantities of enzyme, the transformation of **3.19** by P450 3A4 was studied only at a μg-scale (more precisely, 18 μg of **3.19** per reaction). With such a small amount of material, identification of the reaction products is extremely difficult using NMR and therefore structural information was collected through HPLC-MS. The desired product of the enzymatic oxidation, compound **3.24** (Figure 3.4) is commercially available and was then used as a HPLC standard to monitor whether P450 3A4 is able to oxidize **3.19** at the desired site. Enzymatic oxidation may also occur at different positions on **3.19**, which turned out to be the case. To assign the structures of the observed enzymatic products **3.27** and **3.28**, analysis of the HPLC retention times and MS fragmentation experiments were used (Figure 3.7). Before discussing the observations that led us assign the structures to **3.27** and **3.28**, results are summarized in Figure 3.6.

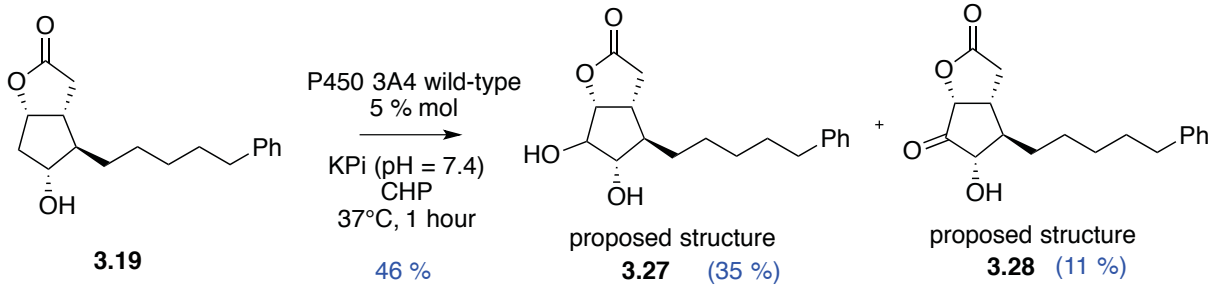


Figure 3.6 Oxidation of **3.19** by P450 3A4. KPi = phosphate buffer, CHP = cumene hydroperoxide

Wild-type P450 3A4 transforms **3.19** with good conversion yields. Regardless of the site of oxidation, this is a positive result demonstrating that **3.19** is indeed a substrate for the enzyme.

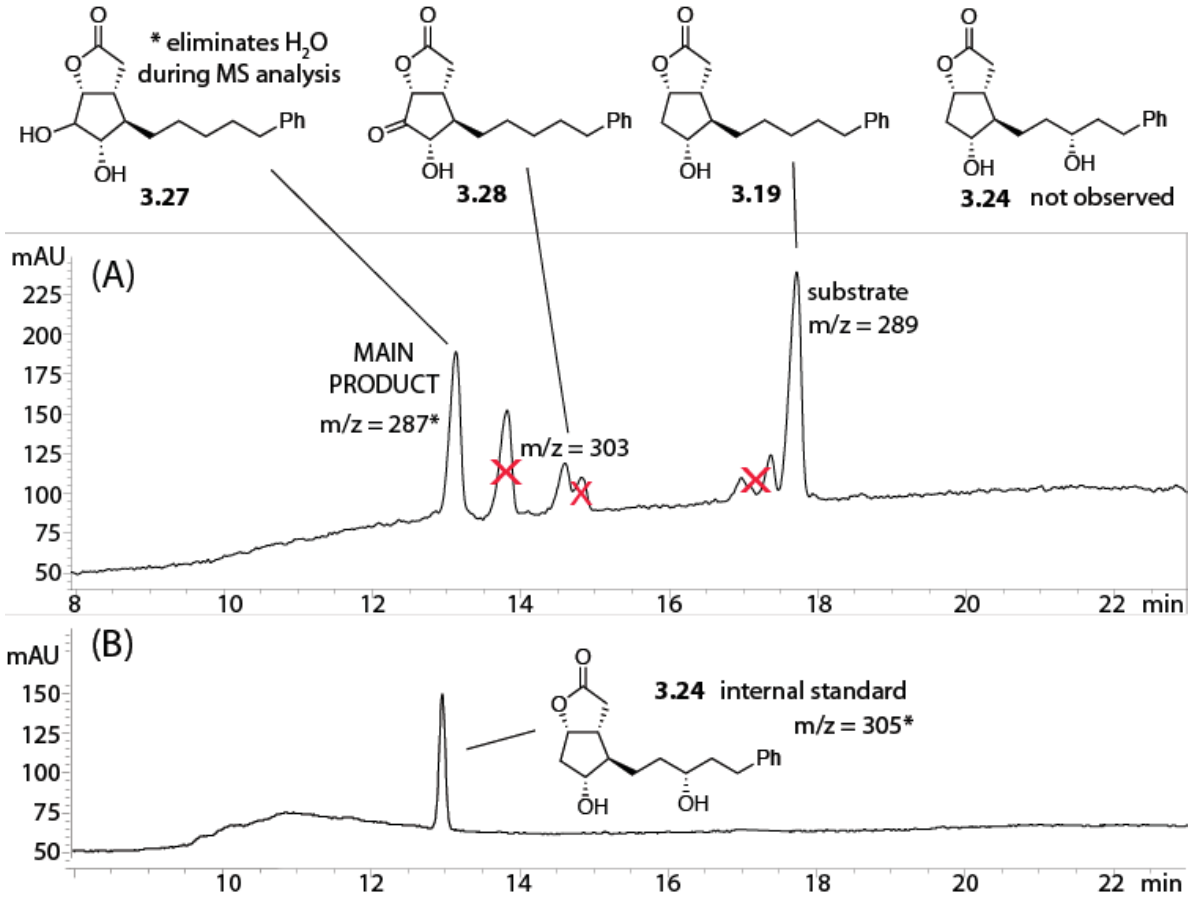


Figure 3.7 Analysis of the products resulting from the transformation of **3.19** by P450 3A4. (A) UV trace of the reaction mixture after extraction in DCM. Barred peaks (X) correspond to compounds already present in the enzyme stock in the absence of the substrate. Proposed structure of the oxidation products are indicated. (B) UV trace for **3.24**. MS spectra are reported in the experimental section.

Experimental observations:

1. Analysis of the retention time of **3.24** (desired product of oxidation, and available as standard) reveals that **3.24** is NOT produced by wild-type P450 3A4
2. The retention time of **3.27**, the major product, indicates polarity similar to that of **3.24**
3. The detected mass of **3.24** (m/z 305) equals the mass of **3.19** (m/z 289) + 16 Da
4. The detected mass of **3.27** (m/z 287) corresponds to a loss of 2 Da (- 2H) from **3.19**
5. **3.28**, the minor product, features lower polarity than **3.27** but greater than **3.19**
6. The detected mass for **3.28** (m/z 303) equals the mass of **3.19** +14 Da (+ 1O, - 2H)

Conclusions:

1. The similar polarity between internal standard **3.24** and major product observed, **3.27**, suggests the latter one is also a diol, yet an isomeric form of the first one
2. At first glance, the mass detected for **3.27** (m/z = 287) could indicate hydroxyl oxidation to ketone (Figure 3.8), yet this is in strong disagreement with polarity observations previously outlined (ketone groups are far less polar than hydroxyls)

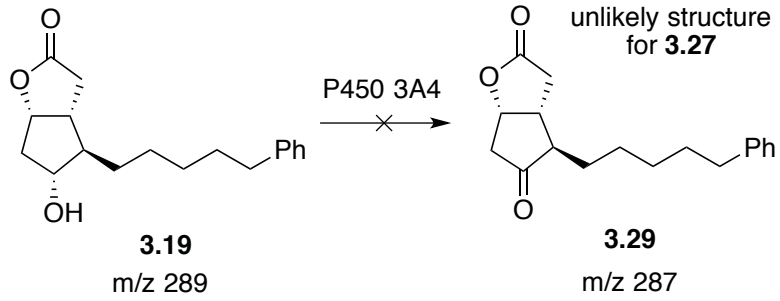


Figure 3.8 Unlikely transformation of **3.19** by P450 3A4 wild-type based on polarity observations from HPLC-MS analysis

3. A more plausible scenario explaining the m/z = 287 associated with the HPLC peak for **3.27** is that the molecular ion of **3.27** (m/z = 305) is not observed, and instead a quantitative loss of H_2O (- 18 Da) occurs spontaneously inside the mass spectrometer (Figure 3.9). According to the literature[111] vicinal diols are known to easily eliminate water when exposed to the high energy conditions inside the MS.

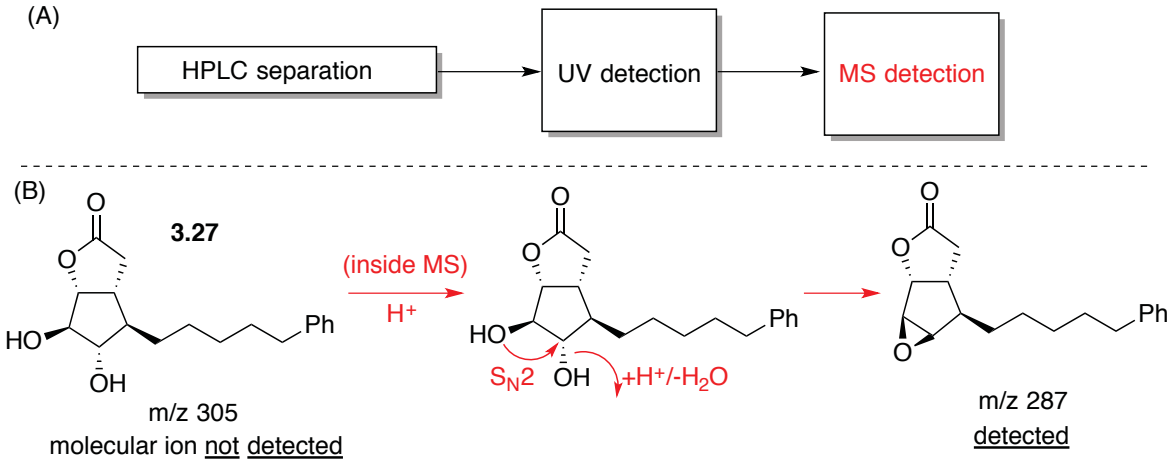


Figure 3.9 Suggested process explaining the observed m/z = 287 associated with the HPLC peak for compound **3.27.** (A) During HPLC-MS analysis, chemical transformations can happen inside the mass spectrometer after UV detection. (B) Proposed mechanism of H_2O elimination

4. Hypotheses on the stereochemistry of the proximal diol group in **3.27** can be made. If we assume that the H_2O elimination process observed during HPLC-MS analysis occurs through an S_N2 mechanism, then *anti* stereochemistry is required (Figure 3.9B).

5. Finally, the structure assigned to **3.28** (Figure 3.7) matches the observations on the polarity and the detected mass (water elimination not as likely as in diols). This product is assumed to be the result of a double-oxidation process by P450 3A4, which first oxidizes **3.19** to the alcohol **3.27**, which is further oxidized to the ketone **3.28**.

3.7 Selection of P450 3A4 mutations

Once the main product of oxidation of **3.19** by P450 3A4 has been identified, computational protocols were used to design mutations aimed at modifying the enzyme selectivity to favor hydroxylation at position C-5 (Figure 3.3).

First, we obtained predictions for the transition state (TS) leading to **3.24** and **3.27**, the desired and observed products of **3.19** respectively (Figure 3.10A-B). To achieve this, we applied constraints to force oxidation by the heme-iron at specific sites on the substrate. The binding poses of the ligand in the active site of P450 3A4 were obtained by using procedures implemented in our computational platform FORECASTER[112].

From the visual analysis of the predicted TS, we identified five amino acid residues, Ser119, Ile301, Ala360, Ile369, Leu482 surrounding the ligand that, once mutated, might help stabilizing the TS leading to the desired product **3.24**. Then, we set out to mutate *in silico* each of these residues to each of the natural amino acids that seemed tolerated in terms of size, or in other words that did not seem to block the ligand from reaching the heme-iron. To screen the mutations, we relied on a specific program for protein engineering that was recently reported by our group[113]. The program, as mentioned in section 1.10, is designed to predict the conformation of the side chain of amino acid residues upon mutation, based on MM principles to explore the potential energy of side chain rotamers. The resulting protein structure can then be used for the docking of ligands and analysis of predicted protein-ligand interactions. Refinement of this program for the rational engineering of P450s is still ongoing. In fact, one of the main goals of the study presented in this chapter was to contribute to the experimental validation of the current version of the program and provide data than can be used for its refinement. For this reasons, it was extremely interesting to test experimentally the mutants that appeared to be promising, in terms of computed energy, from *in silico* predictions.

The predicted structures of the P450 3A4 mutants were analyzed visually. Out of all the mutations, two stood out as particular promising to shift the selectivity of P450 3A4 and favor formation of **3.24**. The first one, mutation of serine 119 to phenylalanine (S119F), was predicted to induce the formation of a phenyl-phenyl stacking interaction between the newly mutated residue and the ligand (Figure 3.10C). These types of interactions can contribute considerably to the enthalpy of binding (1-3 kcal) and are often sought after in traditional drug design[114]. Another promising mutation, I369T (isoleucine to threonine), was predicted to introduce a hydrogen bond between the OH of the (new) threonine and the carbonyl moiety of the substrate. Following these predictions, single mutants S119F, I369T and double mutant S119F/I369T were chosen to be produced and tested experimentally.

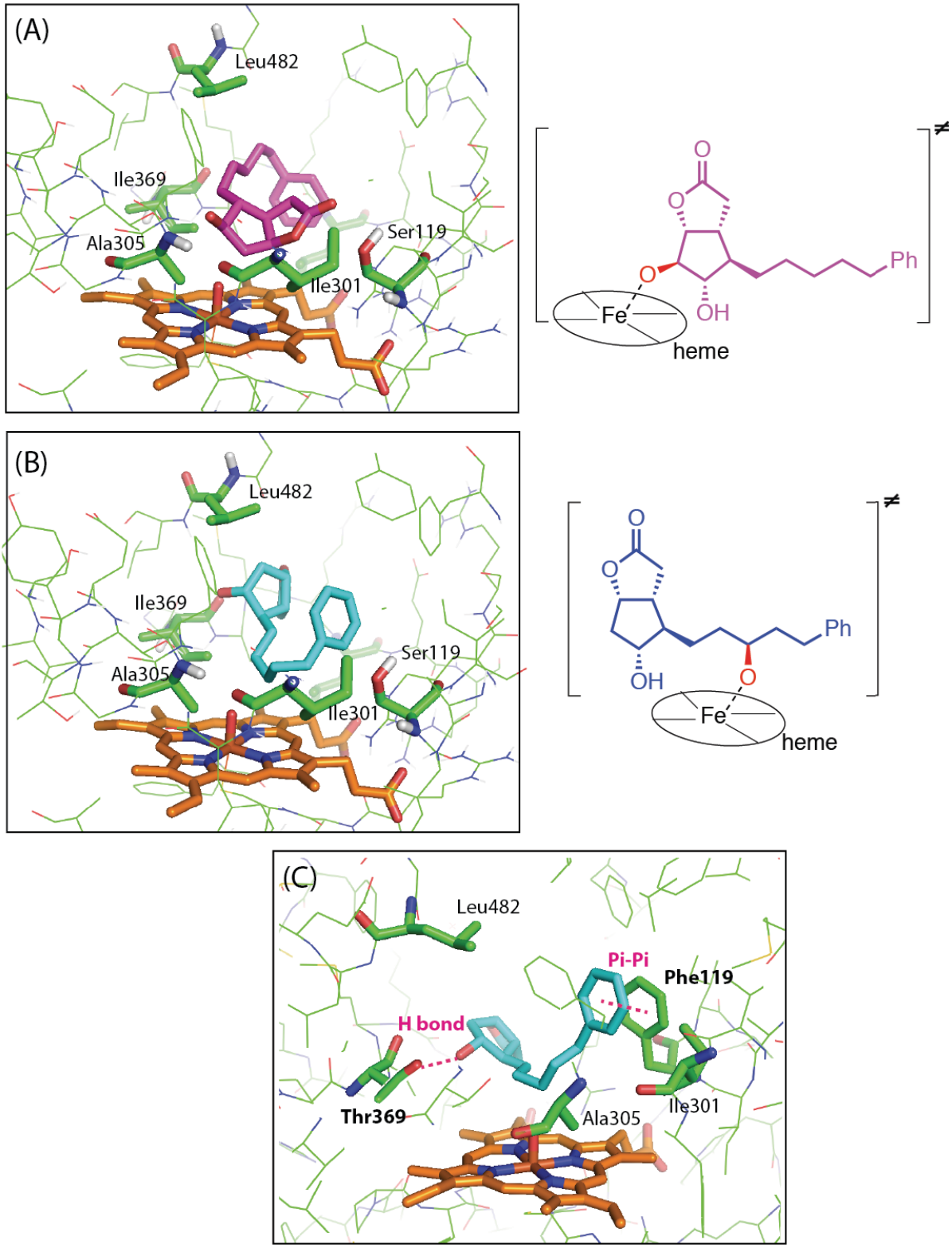


Figure 3.10 Modelling carried out to guide the mutagenesis of P450 3A4 to favor formation of **3.24** over **3.27**. (A) Predicted transition state leading to **3.27** with wild-type P450 3A4. (B) predicted transition state leading to **3.24** with wild-type P450 3A4. (C) Predicted transition state leading to **3.27** with double mutant S119F/I369T, showing favourable Pi-Pi stacking and a new hydrogen bond. In pink, compound **3.27**; in cyan, compound **3.24**; in green wireframe, protein residues investigated for the mutagenesis; in orange wireframe, the heme prosthetic group

3.8 Biochemical assay

PCR-based site-directed mutagenesis was used to introduce the mutations in the P450 3A4 wild-type gene. Following protein expression in *E. coli*, efforts were focused on the development of an assay that does not require protein purification.

HisTag affinity chromatography has been used to purify many P450s, but it can be very time consuming. Another inconvenience specific to human P450s is that the biocatalytic cycle requires the presence of the redox partner cytochrome P450 reductase (CPR) and the cofactor NADPH. The latter is relatively expensive and CPR is a protein that must be expressed and purified. To bypass this last problem, it has been demonstrated by the Auclair lab that cumene hydroperoxide (CHP) can be used as an efficient peroxide shunt to prompt P450 3A4 into its highly reactive oxidative state, therefore avoiding the need for CPR and NADPH[110]. As part of the research conducted for this study, we were pleased to find out that CHP can be used in complex matrices like the cell lysate, once the cell debris been removed by centrifugation. This finding allowed us to develop an assay for the screening of mutants without the need to purify either CPR or the mutant P450 (Figure 3.11)

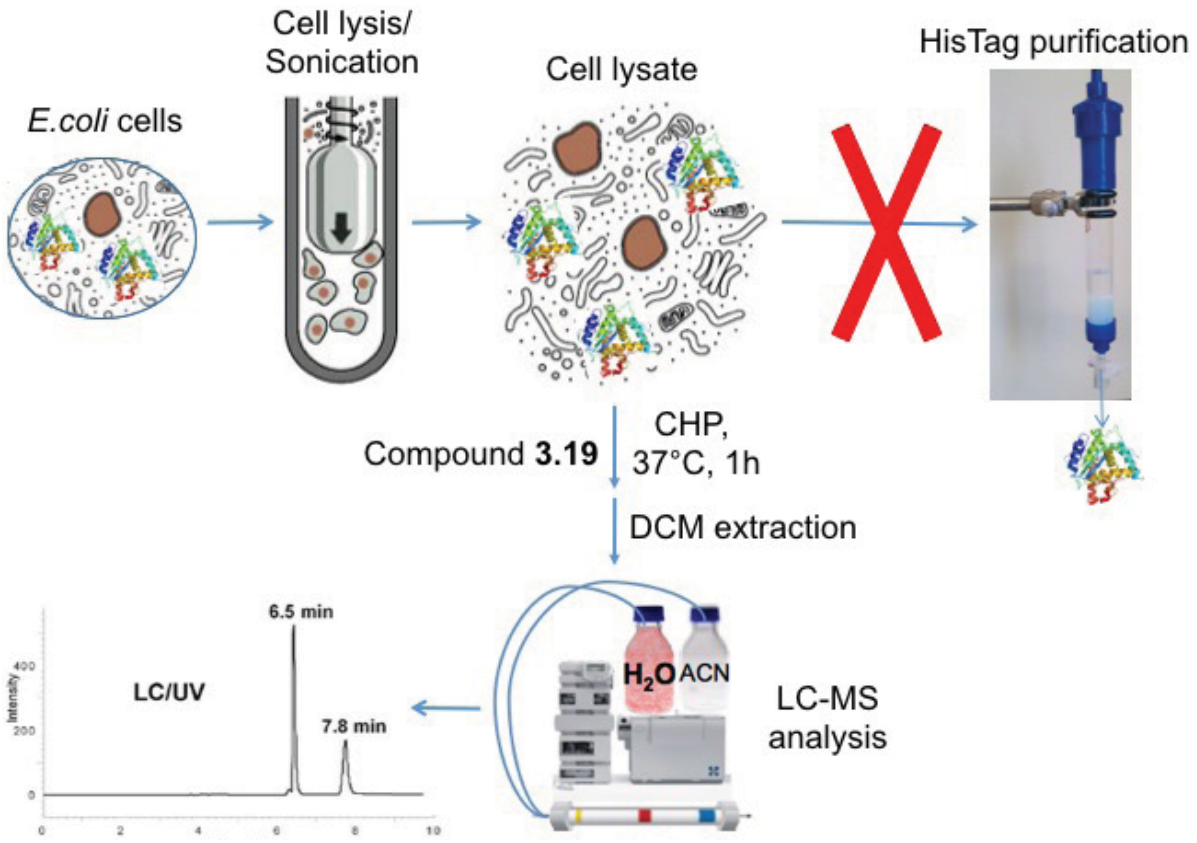


Figure 3.11 Schematic representation of the biochemical assay developed for the expression and testing of P450 3A4 mutants. The protein of interest is first expressed in *E. coli* cells, which are then lysed using sonication. The resulting cell lysate is mixed with the substrate **3.19** and CHP, and the reaction is analysed by LC-MS, after extraction of the substrate and products in DCM.

Following incubation of substrate **3.19** with P450-containing cell lysate and CHP, the products can be extracted in organic solvent and analyzed by HPLC-MS. Analysis of the UV peak areas for each product gives an estimate of the activity of each mutant P450.
The described protocol was used for the experimental testing of the P450 3A4 mutants that were designed as potential biocatalyst for the oxidation of compound **3.19**.
Unfortunately, none of the mutants was able to catalyze the desired reaction. Instead, the same product observed for wild-type was detected. Results are summarized in Table 3-1.

Table 3-1 Performance of P450 3A4 mutants in the oxidation of compound **3.19**

P450	Compound 3.24 detected	Main metabolite	Activity
Wild-type	NO	**3.27**	+++
S119F	NO	**3.27**	++
I369T	NO	**3.27**	+
S119F/I369T	NO	**3.27**	+

3.9 Discussion

First of all, it is interesting to notice that the mutation I359T appears to reduce considerably the activity of the enzyme. This could be explained by the fact the P450 3A4 is known to have a clear preference for lipophilic molecules, most likely due to its hydrophobic active site. It is possible that the introduction of polar groups (such as the threonine OH) reduces the affinity of the enzyme for **3.19**, which is characterized by a theoretical cLogP = 2. The replacement of non-polar side chains to polar ones in P450 3A4 was further investigated in Chapter 4.
In regards to our attempts to modify the site of oxidation, we can speculate on the reasons why rational engineering of P450 3A4 did not work. In general, it must be emphasized that one of the primary goals of this study was to contribute to the development of a computational tool which is still under development. Predictions of the mutant protein structures are based on a number of approximations, most of which relate to the current challenges in protein science discussed in the introduction chapter. First, protein dynamics is not taken into account when ligands are docked to either the wild-type or the mutants. As mentioned in paragraph 1.2 (Chapter 1), experimental crystal structures cannot render accurately the physical reality and docking inaccuracies might arise as a result. Efforts to include backbone motions through NMA (normal mode analysis)[115] are currently being made. Using the NMA approach, the

protein is represented as a set of beads with different masses interconnected by virtual springs, which mimics protein flexibility. Once implemented, this algorithm could provide better predictions for biocatalysis by P450 3A4.

As a second approximation, amino acid residues are considered isolated from the rest of the protein when side chain conformations are being optimized. This means that new mutations are assumed not to impact the conformations of any other residue. Although this might be the case for distant ones, neighboring residues can influence each other and overall have an effect on the binding of a ligand. In addition, in the current version of the software, residue conformations are optimized before docking of the ligand, which is assumed not to affect the structure of the protein. In reality, the process of binding is more complex and implies a series of re-adjustments of the protein until an optimal pose is reached.

Finally, it is not to be excluded the possibility of multiple substrates binding in the active site at the same time, a phenomenon that would complicate considerably any predictive attempts. P450 3A4 has shown high levels of complexity in terms of cooperativity with some experimental conditions showing up to four copies of the same substrate bound in the active site and allosteric pockets[116-118].

3.10 Conclusions

This research project was conceived as a collaboration with computational chemists. In particular, the regio- and stereoselective hydroxylation of an aliphatic chain, applied to synthesis of latanoprost, was chosen to validate a computational tool for biocatalysis under development in our laboratories. Following the synthesis of the envisioned synthetic precursor for the biocatalytic step, initial results revealed this precursor to be good substrate for P450 3A4 wild-type. After structure determination of the main product of oxidation by P450 3A4, three mutants were selected based on predicted protein-substrate interactions. A biochemical assay that does not require protein purification was developed and used for estimating the performance of the newly designed mutants as biocatalysts. Unfortunately, none of the mutations revealed successful in directing P450 oxidation at the desired position on the substrate. Despite the disappointment, this study was an important first step in evaluating the predicting capabilities of the current version of our software.

It is also worth pointing out that the acquired knowledge in the biochemistry of P450 3A4, as well as the establishment of reproducible biochemical procedures for the testing of mutants were key factors contributing to the success of a following study (Chapter 4).

CHAPTER 4 - ACTIVE SITE CROWDING OF P450 3A4 AS A STRATEGY TO ALTER ITS SELECTIVITY

4.1 Preface

The availability of tunable, selective, hydroxylation biocatalysts would significantly streamline the synthesis of complex biological molecules. However, producing such enzymes remains arduous and could greatly benefit from additional protein engineering strategies. The fundamental challenge in the development of novel biocatalytic processes is that the majority of enzymes found in nature are highly specific for one substrate. This is obviously an obstacle to the preparation of any compound that is not the result of a known biochemical pathway.
Based on this premise, substrate-promiscuous enzymes are a promising starting point for the development of versatile biocatalysts.
In this study, human cytochrome P450 3A4 (P450 3A4), known for its ability to metabolize hundreds of drugs, was engineered to alter its regio- and stereoselectivity. Rational mutagenesis was used to specifically introduce steric hindrance in the large active site of P450 3A4 and favour oxidation at a more sterically accessible position on the substrate.
65 mutants were designed to achieve "active site crowding" and, after being generated *in vitro* using standard molecular biology methods, they were tested as biocatalysts for three P450-caralyzed transformations. Hydroxylation of a synthetic precursor of (*R*)-lisofylline, a compound under investigation for its anti-inflammatory properties, was chosen as a first proof-of-principle application of our protein engineering strategy. Increasing active site crowding was then studied with respect to the selectivity of oxidation from an internal double bond to a terminal phenyl group in a derivative of theobromine. The same correlation between crowding and selectivity was found in a final case focused on the hydroxylation of the steroid sex hormone progesterone.
I designed, produced and tested all the mutants of P450 3A4. Synthetic standards for **4.4** and **4.5** were synthesized by a former member of the Auclair group (Aaron Larsen). Standards for **4.7** and **4.8** were synthesized by another member of the Auclair group (Kin Meng Cheong).
The results presented in this chapter were published in the following article: Schiavini, P.; Cheong, K. J., Moitessier, N.; Auclair, K., Active site crowding of P450 3A4 as a strategy to alter its selectivity, *ChemBioChem* (**2017**), 18(3), 248-252

4.2 Introduction

In the last 15 years, many efforts have been directed towards controlling the regio- and stereoselectivity of P450-catalyzed oxidation reactions and a number of reviews on this topic have been written[119-121]. Directed evolution has been used extensively, often in conjunction with rational approaches. The widely studied P450 BM-3[122], and other bacterial P450s have been engineered for the oxidation of terpenes[123-125], alkaloids[126], steroids[127-128] as well as non-natural products[129-130]. However, the vast majority of the efforts have been mostly directed towards increasing the substrate scope of naturally occurring selective enzymes[131-132]. For example, recently Güclu *et al*. engineered a D-fructose-6-phosphate aldolase (FSA) from *Escherichia coli* by switching two leucine residues for two alanines[133]. The expansion of the active site increased the substrate promiscuity of the enzyme which was then able to accept substrates containing up to seven atoms compared with only three atom-substrates accepted by the wild-type.

In this chapter we have investigated a significantly different approach: modifying a substrate-promiscuous enzyme with a large active site to create a smaller, more selective substrate-binding pocket. Examples of substrate promiscuous enzymes include bacterial hydrolases such as *Klebsiella pneumonia* NMD-1[134], a variety of sugar biosynthetic enzymes and glycosyltransferases, and human drug metabolizing P450 enzymes (P450s)[85, 135]. P450 3A4 alone transforms ~50% of all clinical drugs and is known to accept substrates with sizes ranging from ~150 Da (e.g. acetaminophen) to ~1200 Da (e.g. cyclosporine)[86]. Reasons for the broad substrate specificity of P450 3A4 are still unclear but may include the enzyme's large active site, structural plasticity[94, 136], and the non-polar nature of the substrate binding site.

4.3 Development of guidelines for active site crowding

To deal with the known flexibility of P450 3A4, a number of crystal structures of this enzyme in complex with ligands of different sizes (PDB: 3NXU, 1TQN, 1W0G, 2J0D) were analyzed and superimposed to evaluate the different degree of active site opening. One structure exhibiting an average state of opening (3NXU), the most common in all available structures, was selected and a number of residues lining the active site were identified. In order of proximity from the heme cofactor, these residues include: Ala305, Ala370, Ser119, Ile301, Ile369, Ile120, Leu482, Phe304, Gly481, Phe108, Leu210, and Leu211 (Figure 4.1).

The mutagenesis was guided by two basic principles. First, we hypothesized that enzyme activity was going to be best maintained by replacing non-polar side chains with non-polar analogues. Moreover, the mutations were designed to introduce larger residues. This approach was based on the premise that the introduction of bulky groups around the heme would reduce the active site volume and thus decrease the freedom of motion of the substrate, minimizing the number of possible poses and consequently enhancing both regio- and stereoselectivity of hydroxylation. In the selection of mutants, the conformational states of the newly mutated residues were optimized using a recently reported software for single-point mutations[113], after removal of the native ligand. Mutations associated with high destabilization energies, calculated with the aforementioned program, were discarded. It was envisaged that active site crowding may favor oxidation at less-hindered positions on the substrate, as if a hand could reach the bottom of a packed container with only the tip of the fingers.

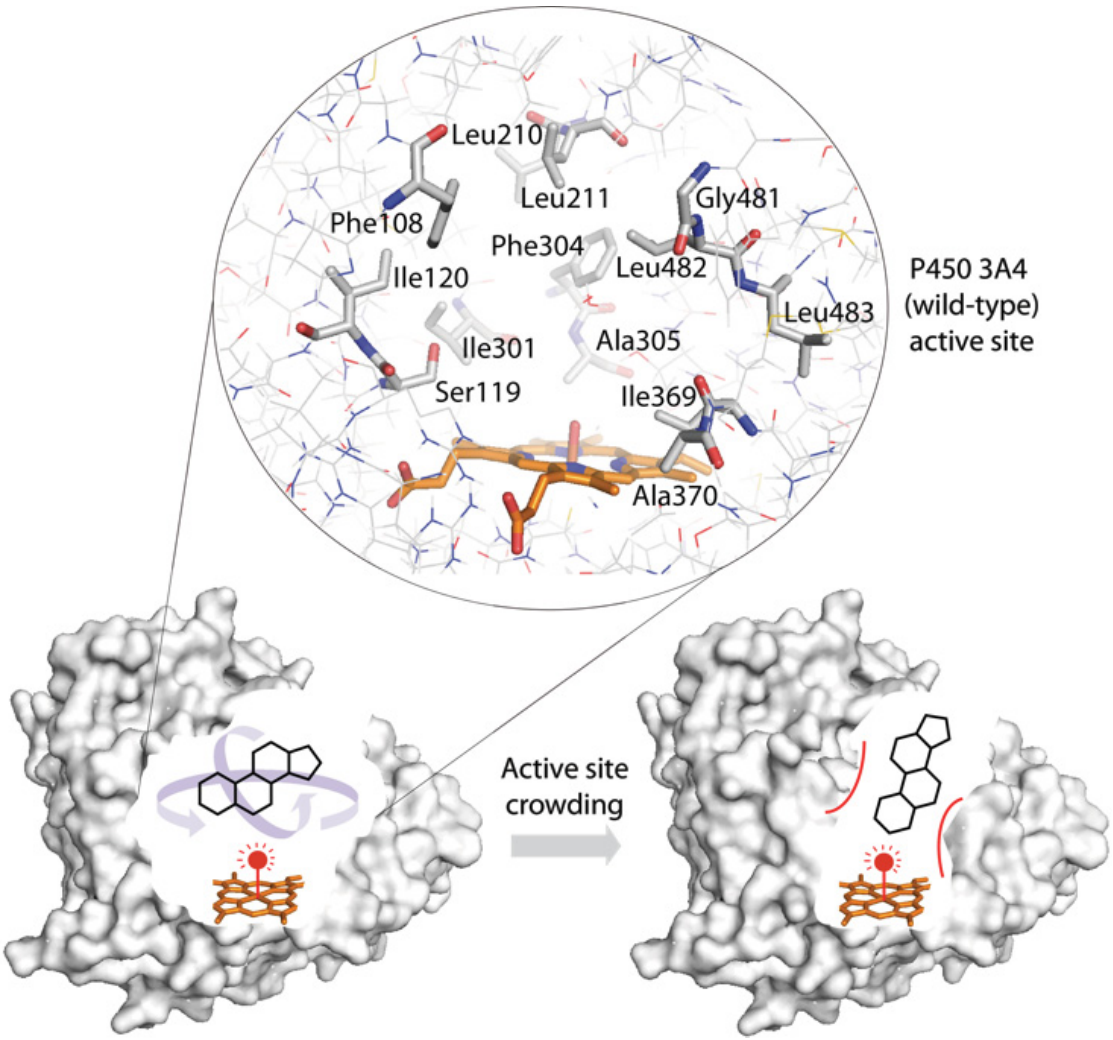


Figure 4.1 Schematic representation of "active site crowding" with the heme shown in orange wireframe and the oxygen oxidizing atom in red. In the zoomed area, residues that were selected for mutagenesis of P450 3A4 are displayed as wireframe colored by atom type

4.4 Oxidation of a precursor of (*R*)-lisofylline

To demonstrate our hypotheses, three P450-catalyzed transformations were investigated. The former builds upon a previous study reporting the use of theobromine as a chemical auxiliary to control reactions of P450 3A4[137-138]. Herein, we hypothesized that compound **4.1** (Figure 4.2), an alkyl derivative of theobromine (**4.6**), could be used as synthetic precursor of (*R*)-lisofylline (**4.2**) an antiinflammatory substance that has shown potential both in the treatment of type 1 diabetes and in the recovery of hematopoiesis following cytotoxic therapies[139-140]. This proposed two-step synthetic route of (*R*)-lisofylline relies on the development of a method for the stereoselective hydroxylation of a specific unactivated C-H bond.

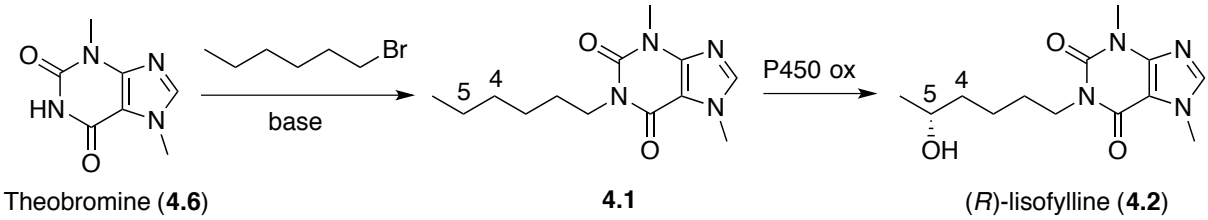


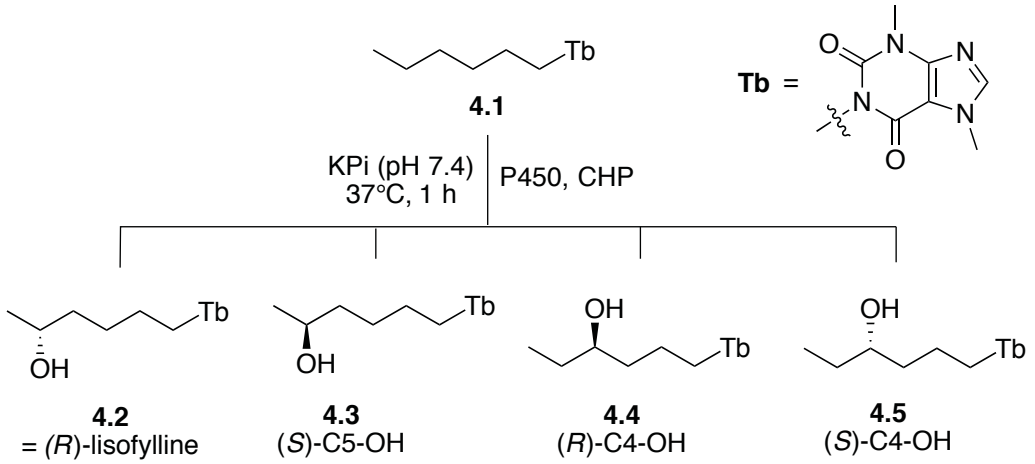
Figure 4.2 Proposed chemo-enzymatic synthesis of (*R*)-lisofylline

Compound **4.1** was previously reported to be hydroxylated mainly at C-4 by the wild type (WT) enzyme[137], however C-5 hydroxylation (with *R* stereochemistry) is required for the synthesis of (*R*)-lisofylline.

Based on the active site crowding concept, 65 mutants were designed and generated *in vitro*. This list contains 29 single-point mutants, 21 double-mutants, 3 triple-mutants, 5 tetra-mutants, 4 penta-mutants, 2 hexa-mutants and 1 hepta-mutant. The mutations were added progressively, one at a time, by discarding all the inactive enzymes and evolving all the ones displaying moderate to high activity compared to wild-type.

The mutants were generated using standard molecular biology with protein expression *in E. coli*, following a procedure described in the experimental section. The full list of mutant enzymes with relative activity, regio- and steroselectivity towards hydroxylation of **4.1** is reported in Table 4-1, whereas a reduced subset in Table 4-2 has been chosen to highlight the the most interesting mutants in terms of regioselectivity.

Table 4-1 P450 3A4 mutant activity and selectivity towards hydroxylation of **4.1**



Mut.	Mutation	**4.2**[a]	**4.3**[a]	**4.4**[a]	**4.5**[a]	Activity[b]
Wild-type	No mutation	6	13	61	20	+++
M1	F108W	8	18	55	19	+++
M2	M114F	-	-	-	-	-
M3	S119D	-	-	-	-	-
M4	S119Q	-	-	-	-	-
M5	S119T	12	26	33	29	++
M6	S119Y	34	23	20	23	++
M7	S119F	24	26	27	23	++
M8	I120F	12	23	36	29	++
M9	L210F	6	9	71	14	+++
M10	L211F	5	9	74	12	+++
M11	I301F	10	28	13	49	+++
M12	I301W	-	-	-	-	-
M13	I301K	-	-	-	-	-
M14	F304W	10	15	42	33	+++
M15	A305V	12	17	49	22	+
M16	I369T	-	-	-	-	-
M17	I369F	18	15	46	21	++

Mut.	Mutation	**4.2**[a]	**4.3**[a]	**4.4**[a]	**4.5**[a]	Activity[b]
M18	A370V	27	26	21	26	+++
M20	A370I	61	21	8	10	++
M21	A370F	-	-	-	-	-
M22	G481V	4	10	72	14	+++
M23	G481Q	3	7	78	12	+++
M24	G481W	4	12	68	16	++++
M25	G481L	3	10	73	13	+++
M26	L482F	15	14	47	24	++
M27	L482Q	14	12	58	16	++
M28	L482R	15	27	15	43	+
M29	L483F	19	15	48	18	+++
M30	S119T/I369F	24	19	27	30	+++
M31	S119T/L482F	24	23	28	25	++
M32	S119T/A370I	38	16	14	32	+++
M33	S119Y/A370V	25	8	29	38	++
M34	S119Y/I369F	34	25	24	17	+
M35	S119Y/L482F	37	22	28	13	++
M36	I120F/L482F	25	21	31	23	++
M37	I120F/G481Q	9	18	45	28	+++
M38	I120F/I301F	15	30	14	41	+++
M39	I120F/I369F	21	26	33	20	++
M40	L211F/G481W	11	5	66	18	+++
M41	I301F/L482F	16	28	15	41	++
M42	I301F/I369F	17	49	7	27	+
M43	I301F/A370I	14	23	6	57	++
M44	F304W/A370I	32	33	20	15	+

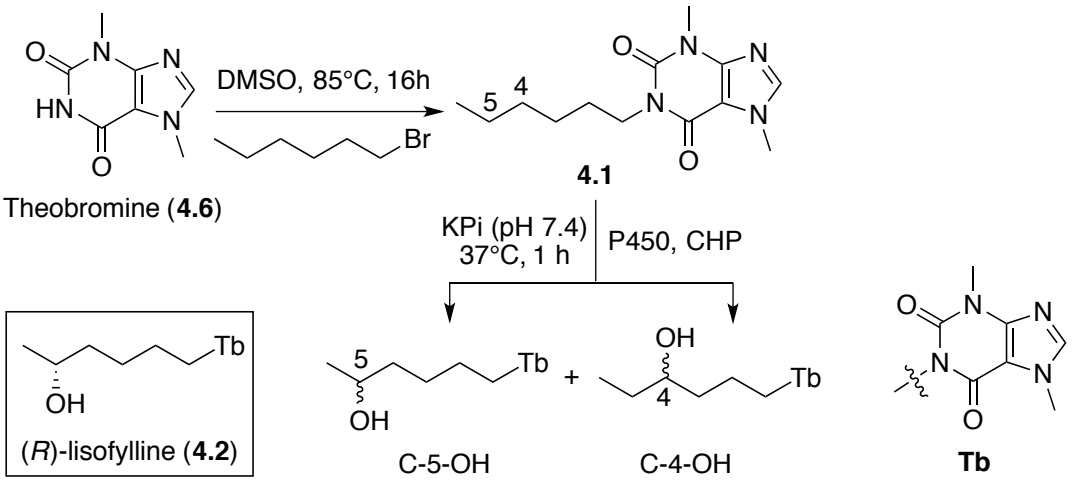
Mut.	Mutation	**4.2**[a]	**4.3**[a]	**4.4**[a]	**4.5**[a]	Activity[b]
M45	A305V/L482F	-	-	-	-	-
M46	I369F/A370I	46	19	16	19	++
M47	I369F/L482F	14	19	43	24	+++
M48	F304W/I369F	20	21	38	21	++
M49	A301F/A370V	6	19	6	69	+++
M50	G481W/L482F	17	18	45	20	++
M51	I120F/I301F/L482F	-	-	-	-	-
M52	I120F/I369F/F304W	19	23	36	22	+
M53	I301F/I369F/L482F	8	40	6	46	++++
M54	I301F/F304W/I369F/L482F	17	25	26	32	+
M55	L211F/I301F/I369F/L482F	9	41	7	43	+++
M56	F108W/I301F/I369F/L482F	8	53	8	31	+++
M57	I301F/I369F/G481W/L482F	11	47	5	37	+++
M58	I301F/I369F/A370I/L482F	-	-	-	-	-
M59	F108W/I301F/I369F/ G481W/L482F	11	48	6	35	+++
M60	S119T/I301F/I369F/ G481W/L482F	6	48	3	43	+++
M61	L210F/I301F/I369F/ G481W/L482F	12	43	6	39	+++
M62	L211F/I301F/I369F/ G481W/L482F	14	52	4	30	++++
M63	S119T/L210F/I301F/ I369F/G481W/L482F	7	43	4	46	+++
M64	S119T/L211F/I301F/ I369F/G481W/L482F	7	53	3	37	++++
M65	F108W/S119T/L210F/ I301F/I369F/G481W/L482F	7	51	3	39	+++

[a] % of each isomer. Standard deviation ± 1

[b] +, ++, +++, ++++ indicate activity estimates relative to the wild-type

Shifts in the C-5/C-4 ratio of hydroxylation were observed with several single-point mutants in which a smaller residue is mutated to a larger one such as Trp, Phe, Leu or Ile (Table 4-2). As expected, changes in the regioselectivity are less notable for mutants at residues further away from the heme group (Leu211, Gly481, Leu482).

Table 4-2 Synthesis of (*R*)-lisofylline (**4.2**) and regioselectivity of 13 selected P450



Mutation	C-5-OH[a]	C-4-OH[a]	Activity[b]
Wild-type	19	81	+++
L211F	14	86	+++
G481W	16	84	++++
F304W	25	75	++
L482F	29	71	+++
I369F	33	67	+++
I301F	38	62	+++
S119T	38	62	+++
S119Y	57	43	++
A370V	53	47	++
A370L	75	25	++
A370I	**82**	**18**	**++**
A370I/I301F	65	35	++
A370I/I369F	47	53	++

[a] % of each isomer. Standard deviation ± 1

[b] ++, +++, ++++ indicate activity estimates relative to the wild-type

Of all the explored residues, Ala370 appeared to be the hottest spot. Replacing the existing Ala with a larger Leu or Ile led to C-5/C-4 hydroxylation ratios of, respectively, 75:25 and 82:18. In terms of enantioselectivity, A370I again performed best, resulting in an enantiomeric ratio (*R*:*S*) of 75:25 versus 52:48 for A370L (Table 4.1).

At this point it became relevant to carry out some modelling to visually compare the transition state (TS) of the oxidation of **4.1** by wild-type P450 3A4 leading to the known product **4.4**, and the TS of **4.1** leading to (*R*)-lisofylline (**4.2**) with mutant A370I (Figure 4.3).

The complexity of the system under investigation is an obstacle in carrying out extensive dynamic simulation studies and a number of approximations had to be used. In practice, the desired substrate was docked by forcing oxidation at desired sites of reaction. For each reaction, multiple transition states were generated and the best ones were chosen based on calculated energies. This was achieved using IMPACTS, a program for substrate-P450 transition state modeling developed by our group and previously validated[141].

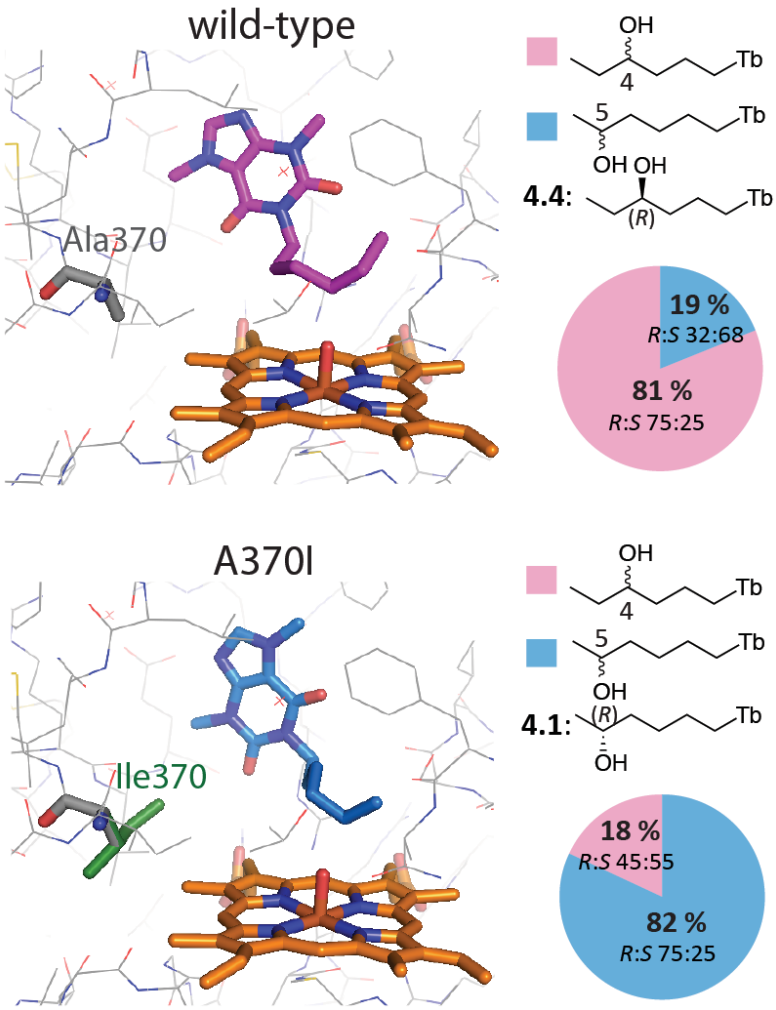


Figure 4.3 Comparison between regio- and steroselectivity of wild-type and A370I. Docking pictures illustrate the predicted transition state for the conversion of **4.1** to **4.4** by wild-type (top) and of **4.1** to (*R*)-lisofylline (**4.2**) by A370I (bottom). Selectivity values deviation standards ± 1

From the visual analysis of the predicted TS for A370I (Figure 4.3), a weak interaction between the newly introduced Ile370 and the bicyclic moiety of the ligand seems to be taking place, causing the latter to assume a more vertical position in the active site, in agreement with our initial model (Figure 4.1). In this binding pose, the alkyl side chain of the ligand appears to be more stretched out thereby exposing the -CH_2 group at position 5 for oxidation by the heme-iron.

Kinetic experiments were carried out with purified enzyme and a single substrate concentration in saturating conditions. These revealed an initial velocity (v_0) of 80 pmol/min for A370I compared with v_0 = 105 pmol/min for wild-type (Table 4-3).

Table 4-3 Activity of A370I vs wild-type towards hydroxylation of **4.1**

Enzyme	substrate	v_0 (pmol/min)
Wild-type	**4.1**	105 ± 4
A370I	**4.1**	80 ± 3

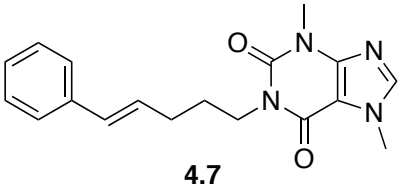
Also, these results were replicated when the natural cofactors NADPH and the redox partner CPR were used instead of CHP.

To further enhance selectivity at position C-5, several single-point mutations were combined, albeit with no improvement. Nevertheless, the almost complete switch of regioselectivity and the high level of stereoselectivity achieved with only one mutation, A370I, is remarkable considering the large number of highly similar C-H bonds.

Remarkably, the quadruple mutant F108W/I301F/I369F/L482F was found to favor the formation of the (*S*) enantiomer of lisofylline with a (*R*:*S*) ratio of ~ 1:6 (Table 4-1). Despite (*S*)-lisofylline being biologically inactive, this last finding emphasizes the versatility of P450 3A4 as a tunable biocatalyst to access different stereoisomers.

4.5 Oxidation of compound 4.7

Next we decided to explore the selectivity of our mutants with a substrate containing multiple functional groups known to be soft spots for P450 oxidation. Thus efforts were focused on the transformation of compound **4.7** (Figure 4.4), a derivative of theobromine (**4.6**) which had been previously synthesized by a co-worker in the Auclair



4.7

Figure 4.4 Structure of **4.7**

Lab (Kin Meng Cheong). From the list of P450 mutants prepared (Table 4-1), three were selected for their increasing level of active site crowding, namely I369F/L482F, I301F/I369F/L482F and L211F/I301F/I369F/G481W/L482F (Figure 4.5) which were purified and characterized by LC-MS to confirm mutagenesis together with gene sequencing.

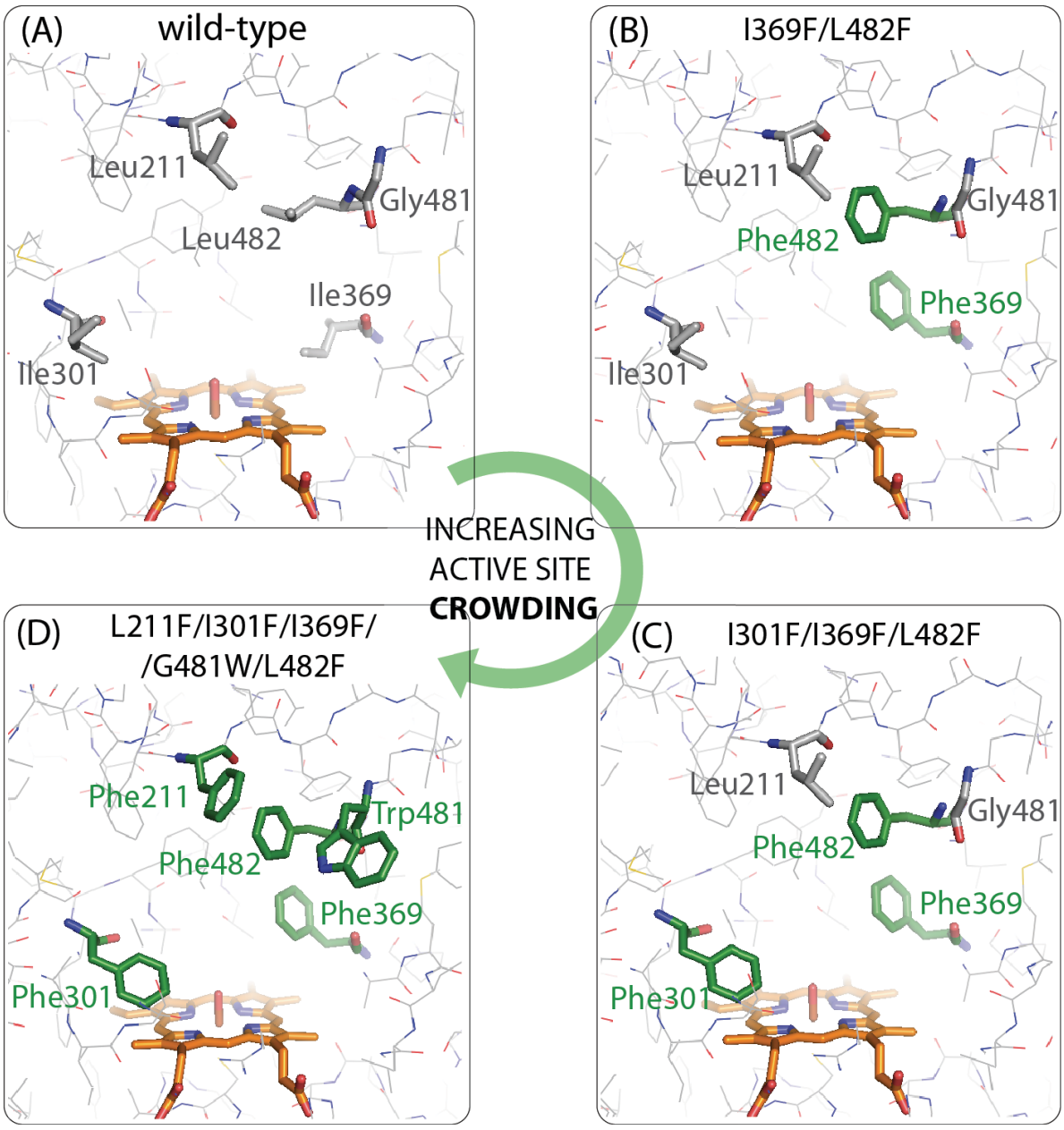


Figure 4.5 Computer modelling of increasing P450 3A4 active site crowding in the absence of any ligand. The structure of the wild-type (A) was retrieved from the protein data bank (PDB 3NXU) with the respective ligand being concealed. The structures of mutants I369F/L482F (B), I301F/I369F/L482F (C) and L211F/I301F/I369F/G481W/L482F (D) were obtained by *in silico* prediction of the conformational states of the mutated residues (in green wireframe).

These three P450 variants and the wild-type were tested for the oxidation of **4.7**. LC-MS analysis was used to identify all the products. Epoxidation of the double bond and phenyl oxidation were identified by comparison with synthetic standards and by MS fragmentation experiments respectively (supporting information). Gratifyingly, a clear correlation between selectivity and the degree of active site crowding was observed (Figure 4.6). In particular, increasing active site crowding appears to be correlated with an incremental shift in selectivity towards less-hindered positions on the substrate, from the internal double bond to the phenyl group at the extremity of the molecule.

These results are in agreement with our initial hypothesis according to which the introduction of bulky groups in the active site of the enzyme should favor oxidation at the less sterically hindered positions.

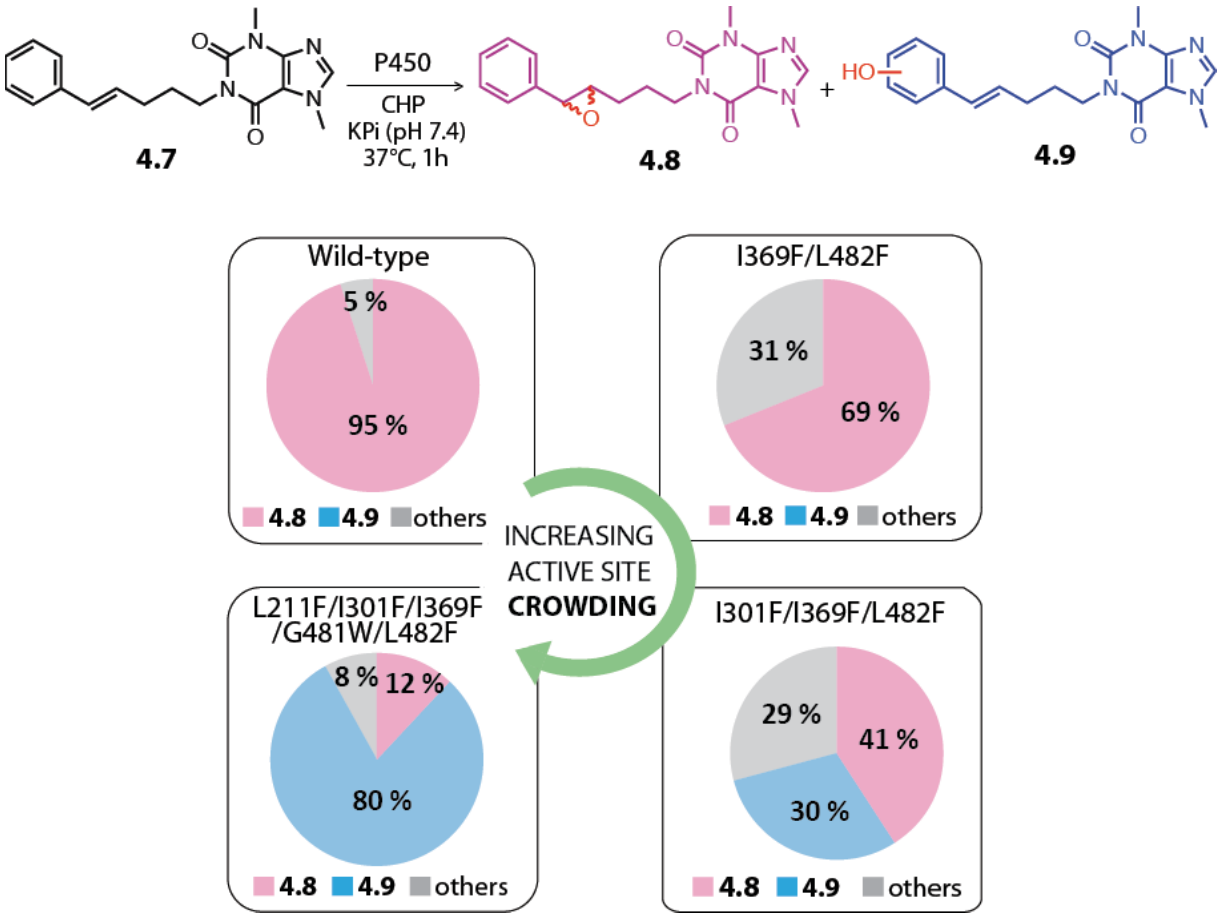


Figure 4.6 Selectivity change, as a result of active site crowding, in the oxidation of **4.7** by P450 wild-type and mutants I369F/L482F, I301F/I369F/L482F, and L211F/I301F/I369F/G481W/L482F. Increasing crowding causes a shift from epoxidation of the internal double bond to hydroxylation of the terminal phenyl group. Standard deviation for the selectivity values: ± 1

4.6 Oxidation of progesterone

As a third application of our strategy, efforts were focused on progesterone (**4.10**, Figure 4.7), a steroid hormone exhibiting high structural rigidity. Steroids have important pharmacological properties and synthetic methodologies to selectively functionalize them can be very useful[142-143]. Here again the same four P450 3A4 variants were compared: wild-type, I369F/L482F, I301F/I369F/L482F and L211F/I301F/I369F/G481W/L482F. In analogy to the transformation of compound **4.7**, a clear correlation between selectivity and the degree of active site crowding is observed for progesterone oxidation (Figure 4.7).

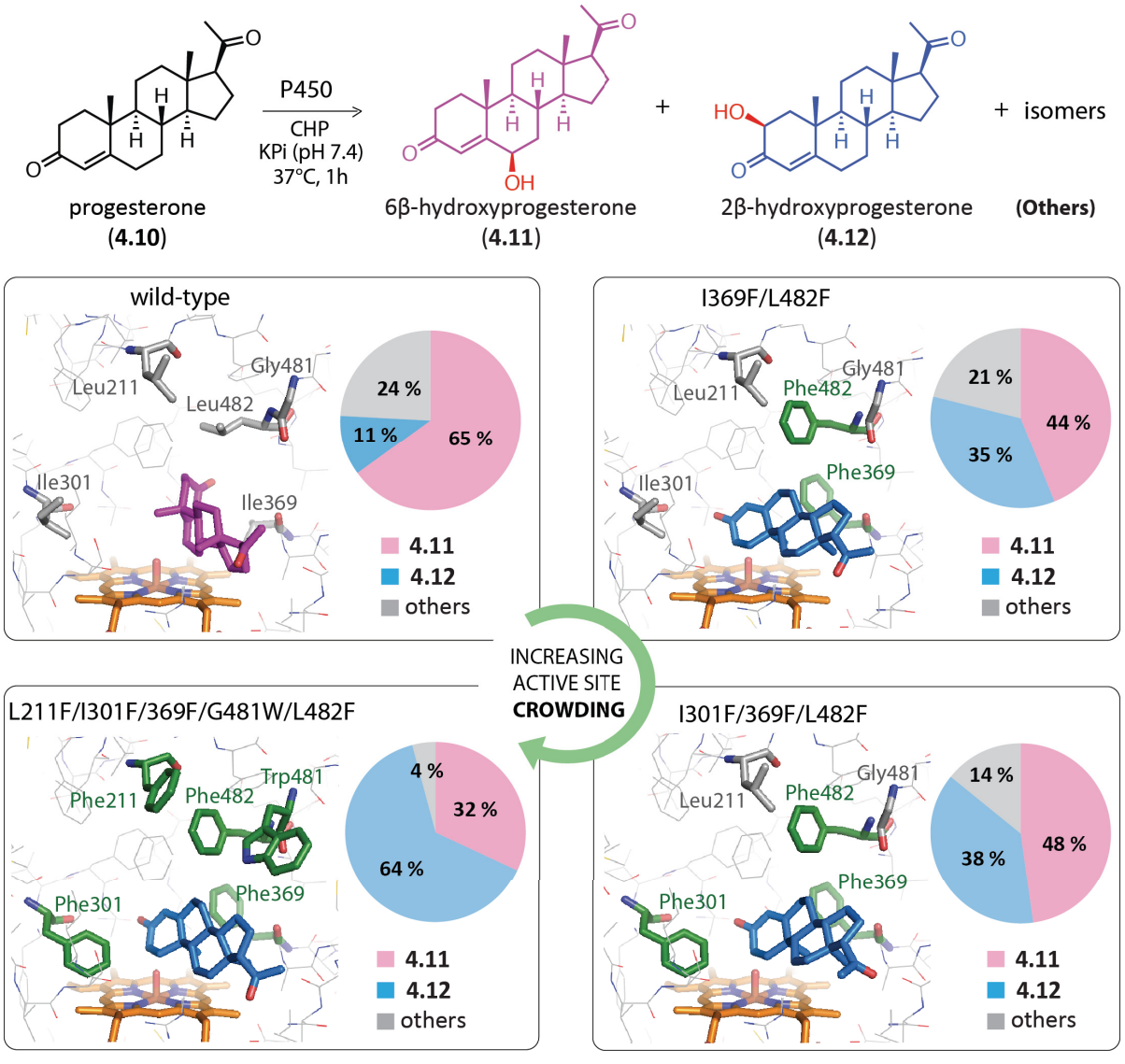


Figure 4.7 Oxidation of progesterone (**4.10**) by P450 3A4 wild-type and mutants I369F/L482F, I301F/I369F/L482F, and L211F/I301F/I369F/G481W/L482F. The docking pictures illustrate the predicted transition states for the formation and 6β-hydroxyprogesterone (**4.11**) by wild-type and 2β-hydroxyprogesterone (**4.12**) for the three mutants. Selectivity values standard deviation ± 1

In particular, the relative amount of 6β-hydroxyprogesterone (**4.11**) diminishes in favor of the 2β-hydroxyprogesterone (**4.12**) diastereoisomer as the number of mutations increases and the active site becomes smaller.

In addition, incremental crowding leads to higher homogeneity of the product, as other isomers become less and less favored. We propose that this is due to a decrease in the number of possible binding modes for the substrate.

4.7 Conclusions

In conclusion, we have presented a structure-guided approach for biocatalyst engineering which complements existing methods. While the common approach is to start from enzymes that only accept one or a few substrates and mutate them to switch or broaden their substrate selectivity, the strategy studied here takes advantage of a substrate-promiscuous enzyme with a large active site and introduces multiple mutations to shrink the substrate binding pocket. In addition, because it focuses efforts on the substrate binding pocket, this approach is particularly interesting because it requires no high throughput assay. Dozens of mutants are tested as opposed to the thousands that are usually generated with random-based protein engineering techniques such as directed evolution.

As the choice of mutations is based on general guiding principles which are not substrate-specific, "active site crowding" could be theoretically applied to other promiscuous enzymes as well as other substrates. In this respect, the role of substrate structural rigidity appears to play an important role in determining the reaction selectivity as crowding increases. In fact, when the highly flexible side of chain of **4.1** was partially rigidified through introduction of a double bond (in **4.7**), a clear trend linking selectivity and crowding was observed. This correlation appeared again in the hydroxylation of progesterone, a highly rigid molecule. A possible explanation for these results is that structural rigidification decreases the number of possible binding poses of the ligand, which is consequently more affected by the constraining of the active site. Based on our results, we foresee that higher structural rigidity of other substrates will likely correspond to higher correlation between shift in selectivity and the number of "crowding" mutations.

As a final remark, it is worth pointing out that structure-activity relationships of drug metabolizing enzymes, such as P450 3A4, can be very useful in the context of early drug development. The important of P450s in drug metabolism is at center of the next chapter of this thesis.

CHAPTER 5 - P450-MEDIATED METABOLIC INSTABILITY OF INHIBITORS OF PROLYL OLIGOPEPTIDASE

5.1 Preface

The importance of thorough drug metabolism studies is well recognized in the pharmaceutical field. In the early stages of drug development, major metabolites are tested for activity and toxicity, and the resulting findings are carefully considered before promoting a drug candidate for expensive and time-consuming clinical trials.

Liquid chromatography coupled to mass spectrometry (LC-MS) is undoubtedly the most widely used tool for metabolite identification. Despite its extensive use, LC-MS cannot always reveal the full spectrum of chemical transformations happening to the drug candidates within a biological system. In fact, the metabolites detected can be the result of a multistep metabolic path and are not necessarily the primary products of enzymatic reaction. In order to fully elucidate a metabolic process, an array of techniques, from computational techniques aimed at predicting the most likely site of metabolism (SoM) must sometimes be combined in a synergistic fashion.

The study described in this chapter exemplifies how a combination of computational and experimental techniques may be necessary for solving the most difficult cases of drug metabolism. In the process of characterizing a mixture of rapidly interconverting P450 metabolites, the unique catalytic properties of P450s were once again exposed.

I designed and performed all the experiments, except for the computational work. Creation of the MatLab protocol for implementation of the "exclusion rules" was carried out by Joshua Pottel, a former member of the Moitessier group. The ground state energies of compounds **5.3**-**5.9** were also computed by Joshua Pottel.

The results described in this chapter were published in the following article: Schiavini, P.; Pottel, J.; Moitessier, N.; Auclair, K., Metabolic Instability of Cyanothiazolidine-Based Prolyl Oligopeptidase Inhibitors: a Structural Assignment Challenge and Potential Medicinal Chemistry Implications, *ChemMedChem* (**2015**), 10, 1174-1183

5.2 Metabolism studies of two inhibitors of POP

Prolyl oligopeptidase (POP) is a post-proline serine endopeptidase that has become a promising drug target for the treatment of neurodegenerative states such as Alzheimer's disease[1, 144-146] and some psychiatric disorders (e.g., bipolar disorder)[147].

In previous work by Lawandi *et al.*[1], compounds **5.1** and **5.2** (Figure 5.1) were synthesized and tested as POP covalent inhibitors. **5.1**, in particular, showed promising activity in living cells, with IC_{50} in the high nanomolar. Prior to beginning a lead optimization study, **5.1** and **5.2** were incubated with human liver microsomes (HLM) and their metabolites examined by LC-MS. The chromatographic profiles revealed the formation of three mono-oxygenated products for **5.1** (Figure 5.1B), and four for **5.2** (Figure 5.1C). No other metabolites were detected. Analysis of the retention times and co-elution controls (Figure 5.1D) confirmed the different chemical nature of these seven compounds, which are here labelled in order of increasing elution times **5.3**, **5.4**, **5.5**, **5.6**, **5.7**, **5.8**, **5.9**.

Chemical intuition and a vast body of literature suggested that the sulfur atom of **5.1** and **5.2** would be the most likely site of oxidation by P450s[148], the main metabolic enzymes found in HLM, although this would lead to only two diastereomeric metabolites for each substrate molecule. Alternatively, aromatic oxidation of the benzyl ring could have produced three regioisomers each.

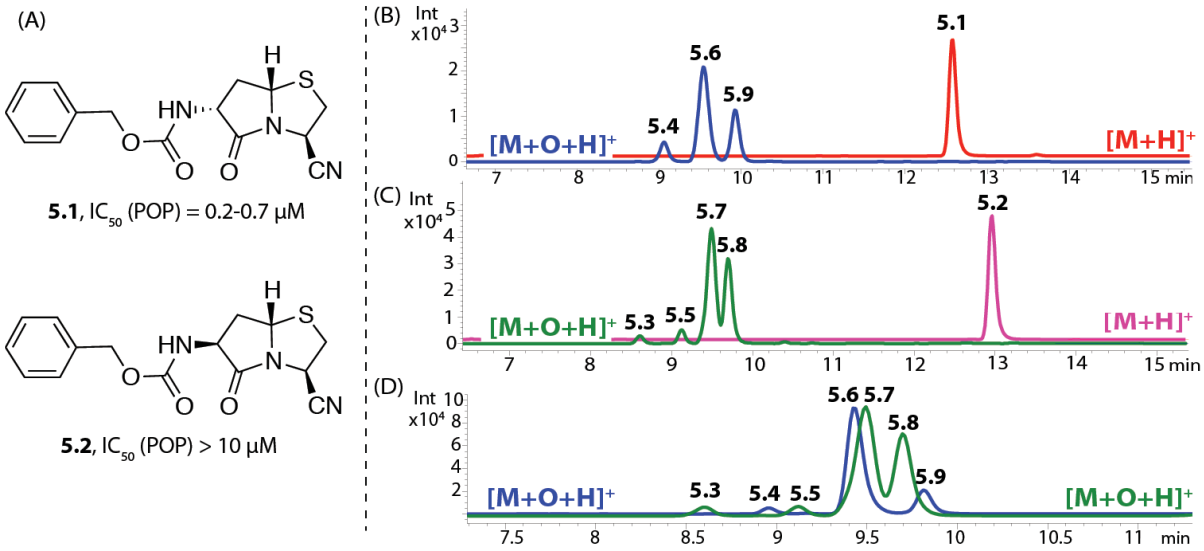


Figure 5.1 Metabolism studies of **5.1** and **5.2.** A) Structure of **5.1** and **5.2** with respective inhibitory potencies vs human POP[1]. B) Red trace: monitoring of $[M+H]^+$ ion (m/z = 318) during LC–MS analysis of **5.1**; blue trace: monitoring of $[M+O+H]^+$ ions (m/z = 334) after incubation of **5.1** with HLM (1 $mgmL^{-1}$) in potassium phosphate buffer (PKi, 0.1 mm, pH 7.4) for 1 h at 37°C. C) Pink trace: monitoring of $[M+H]^+$ ion (m/z = 318) during LC–MS analysis of **5.2**; green trace: monitoring of $[M+O+H]^+$ ions (m/z = 334) after incubation of **5.2** with HLM (same conditions used for **5.1**). D) Overlay of the LC–MS traces of products of monooxidation of **5.1** and **5.2**

5.3 Investigating sulfur oxidation

To identify the correct site of metabolism, we first focused our efforts on preparing synthetic standards for comparison of HPLC retention times with those of the observed metabolites. Ammonium persulfate was used to perform the sulfoxidation of **5.1** and **5.2** while avoiding formation of the corresponding sulfones[149]. Interestingly, oxidation of **5.1** led to the formation **5.4**, **5.6**, **5.9** (Figure 5.2A), whereas **5.2** led to the formation of a mixture of **5.3**, **5.4**, **5.5**, **5.6**, **5.7** and **5.8** (Figure 5.2B). In summary, the two separate reactions provided the same seven compounds previously observed in the metabolism of **5.1** and **5.2**. Isolation of compounds **5.3** through **5.9** for NMR characterization or X-ray crystallography turned out to be unsuccessful. In fact, whenever isolation of each of these compounds was attempted through semi-preparative HPLC, a new mixture was regenerated upon solvent removal. This observation was key in revealing the scenario of a quick equilibration among the molecules under study.

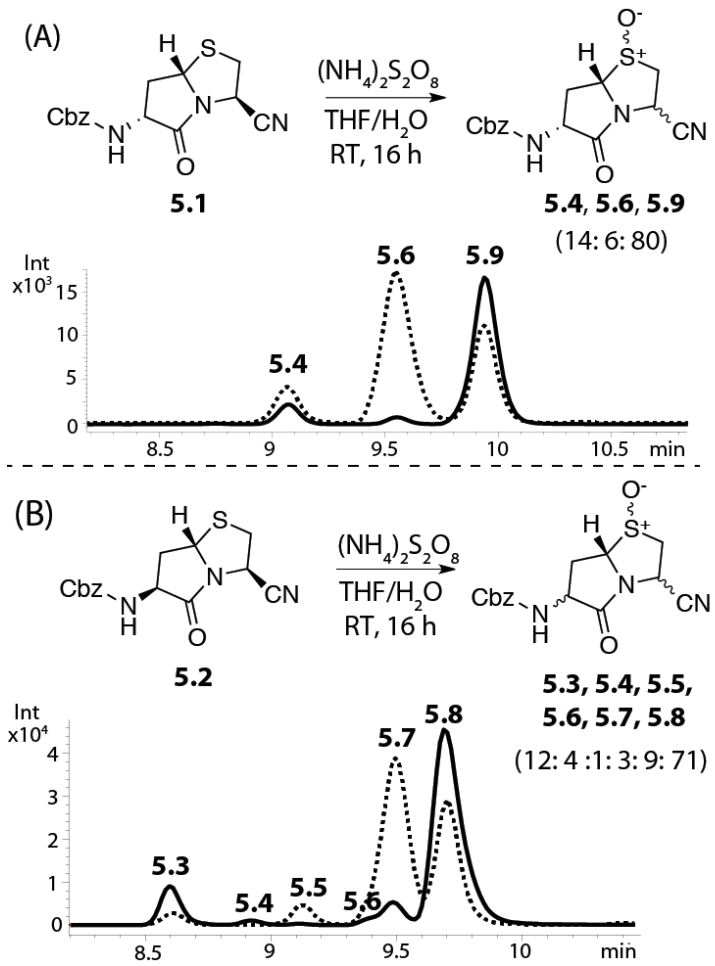


Figure 5.2 Synthesis of **5.3-5.9** by chemical oxidation of **5.1** and **5.2** with ammonium persulfate and comparison with HLM microsomal oxidation. (A) LC-MS trace overlap of the products of chemical (solid line) and HLM microsomal (dashed line) oxidation of **5.1**. (B) LC-MS trace overlap of the products of chemical (solid line) and HLM microsomal (dashed line) oxidation of **5.2**.

5.4 Structure determination of 5.3-5.9

In order to fully characterize all seven observed compounds, we envisioned a strategy where different techniques could yield complementary pieces of structural data. Figure 5.3 gives an overview of the overall path followed and the tools used to arrive at the final solution. Each individual experiment and the information it provided will be discussed in details in the following sections. In summary, we first clarified what type of isomers (regio- or stereo-) we were dealing with: LC-MS/MS fragmentation studies revealed that **5.3**-**5.9** all contained a sulfoxide moiety and that they differed only by their stereochemistry.

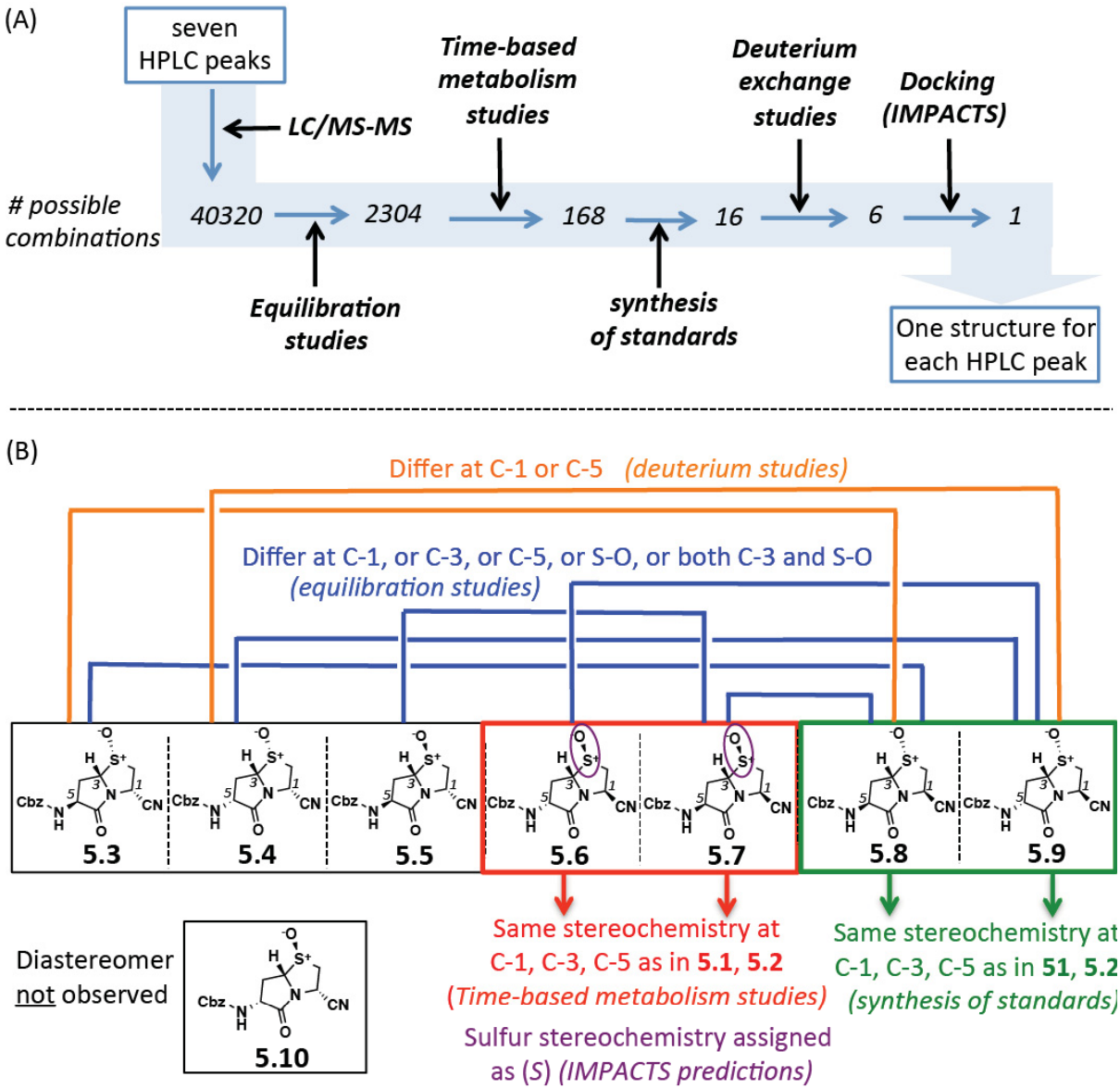


Figure 5.3 Schematic representation of the strategy used to assign structures to **5.3-5.9**. (A) Flowchart summarizing all the techniques used to to assign a structure to each of the seven HPLC peaks from metabolism analysis of **5.1** and **5.2**. B) Graphical representation of the partial structural information collected through each of the experiments and *in silico* calculations.

At this point, we assumed that no information on the configuration of any stereocenter was available, and we considered all possible sets of stereoisomers that could be assigned to the seven HPLC peaks. In the presence of four stereogenic centers, eight diastereomers are possible (although we observed only seven). From a mathematical perspective, 8 different numbers can be ordered in 40320 (=8!) unique arrangements (i.e. [1,2,3,4,5,6,7,8] or [2,1,3,4,5,6,7,8] or [1,8,2,7,3,6,4,5] etc.). In case of **5.3**-**5.10**, one arrangement could be, for example, [**5.3**, **5.4**, **5.5**, **5.6**, **5.7**, **5.8**, **5.9**, **5.10**], another one [**5.8**, **5.3**, **5.5**, **5.6**, **5.7**, **5.4**, **5.9**, **5.10**], etc., each of which we will call henceforth 'combinations'. Thus the order of compounds within each combination represents the order of elution time on the HPLC trace (Figure 5.1D). Based on this hypothesis, a series of experiments was designed to negate certain combinations and, by a process of elimination, reduce the number of possibilities to one and only one. This one solution would contain not only the structures of **5.3**-**5.9**, but also that of the unobserved eighth diastereomer **5.10** (Figure 5.3).

In order to keep track of all of the experimental observations and evaluate their consequences on the structure assignments, we relied on a logic-testing protocol written in MatLab (v. 2012b) for which experimental results were encoded as "exclusion rules".

5.5 LC–MS/MS fragmentation studies

When LC separation was coupled to ESI-MS (Electrospray Ionization Mass Spectrometry), an almost identical fragmentation pattern was observed for compounds **5.3** through **5.9** (Figure 5.4), which was taken as evidence that the seven compounds were diastereomers.

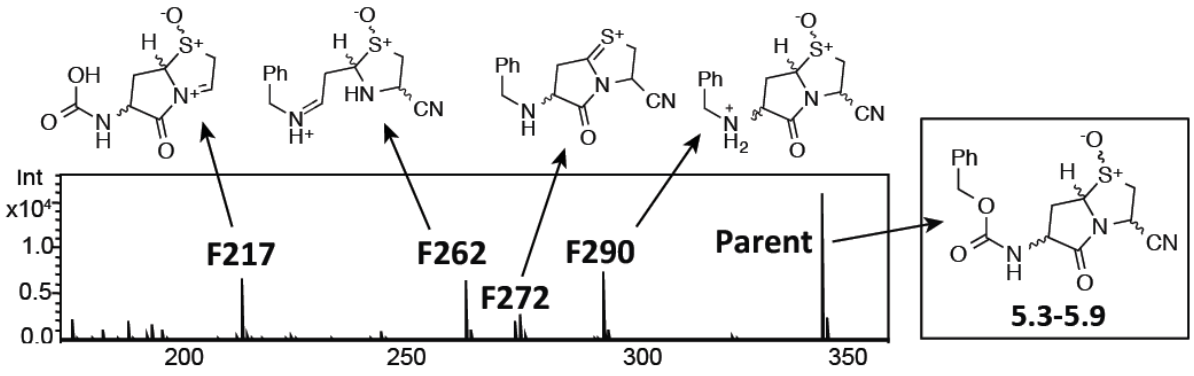


Figure 5.4 Representative fragmentation pattern shared by compounds **5.3**–**5.9**. The proposed structures for the main fragments strongly suggest that **5.3**–**5.9** are diastereomeric sulfoxides. The labelling of the fragments refers to their m/z value, for example F217 = fragment with m/z = 217.

By taking advantage of high resolution MS (time-of-flight), we assigned structures to the main fragments, which allowed a partial structural characterization (Figure 5.5). Each fragment can be rationalized by a mechanism that involves protonation followed by elimination of a very stable molecule (e.g. CO_2, CO, HCN, H_2O), a common phenomenon in MS[150]. Formation of the fragment F290 was rationalized by a 4-membered ring rearrangement leading to loss of CO_2, as previously reported for scaffolds bearing a Cbz-protected amine[151]. The presence of a sulfoxide moiety was confirmed by the formation of fragment F272 which corresponded to a loss of H_2O from fragment F290 in all compounds analyzed. Loss of the oxygen atom of the sulfoxide had been reported in the literature for similar structures analyzed with the same fragmentation technique[152].

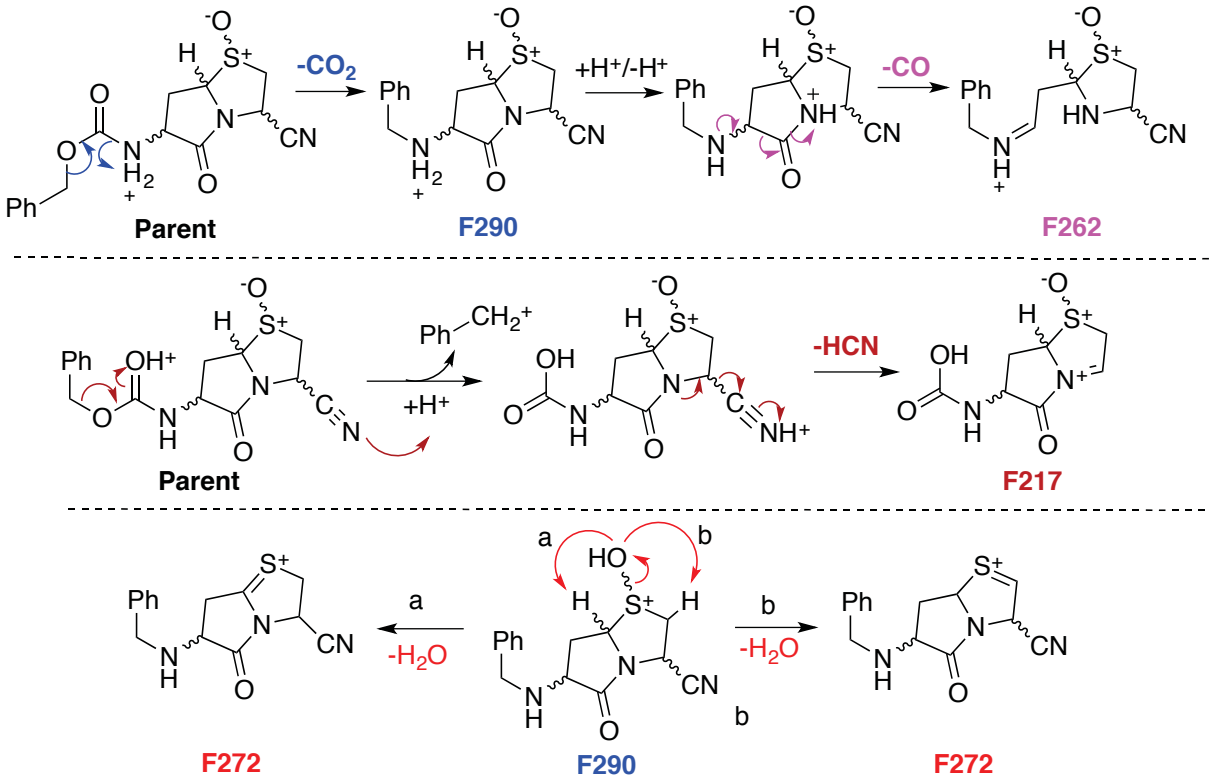


Figure 5.5 Proposed structures and fragmentation mechanisms for the main fragments observed during LC-MS/MS analysis of **5.3-5.9**

5.6 Equilibration studies

As previously mentioned, attempts to synthesize **5.3**-**5.9** revealed the existence of a rapid equilibration among these compounds. To further investigate this phenomenon and draw connections with the results obtained from P450 metabolism studies (Figure 5.1), thorough equilibration studies were carried out. In practice, each of the seven compounds **5.3-5.9** was rapidly isolated by analytical HPLC and without prior removal of the eluent it was allowed to equilibrate under the conditions used to study P450 metabolism: potassium phosphate buffer (KPi) 0.1 M, pH = 7.4 at 37°C, under gentle shaking (250 rpm). Isomerization was monitored by LC-MS after 1 min, 10 min, 1 hour, 6 hour, 24 hours and 3 weeks of incubation time, and as a result a complex pattern of interconversion phenomena was discovered (Figure 5.6).

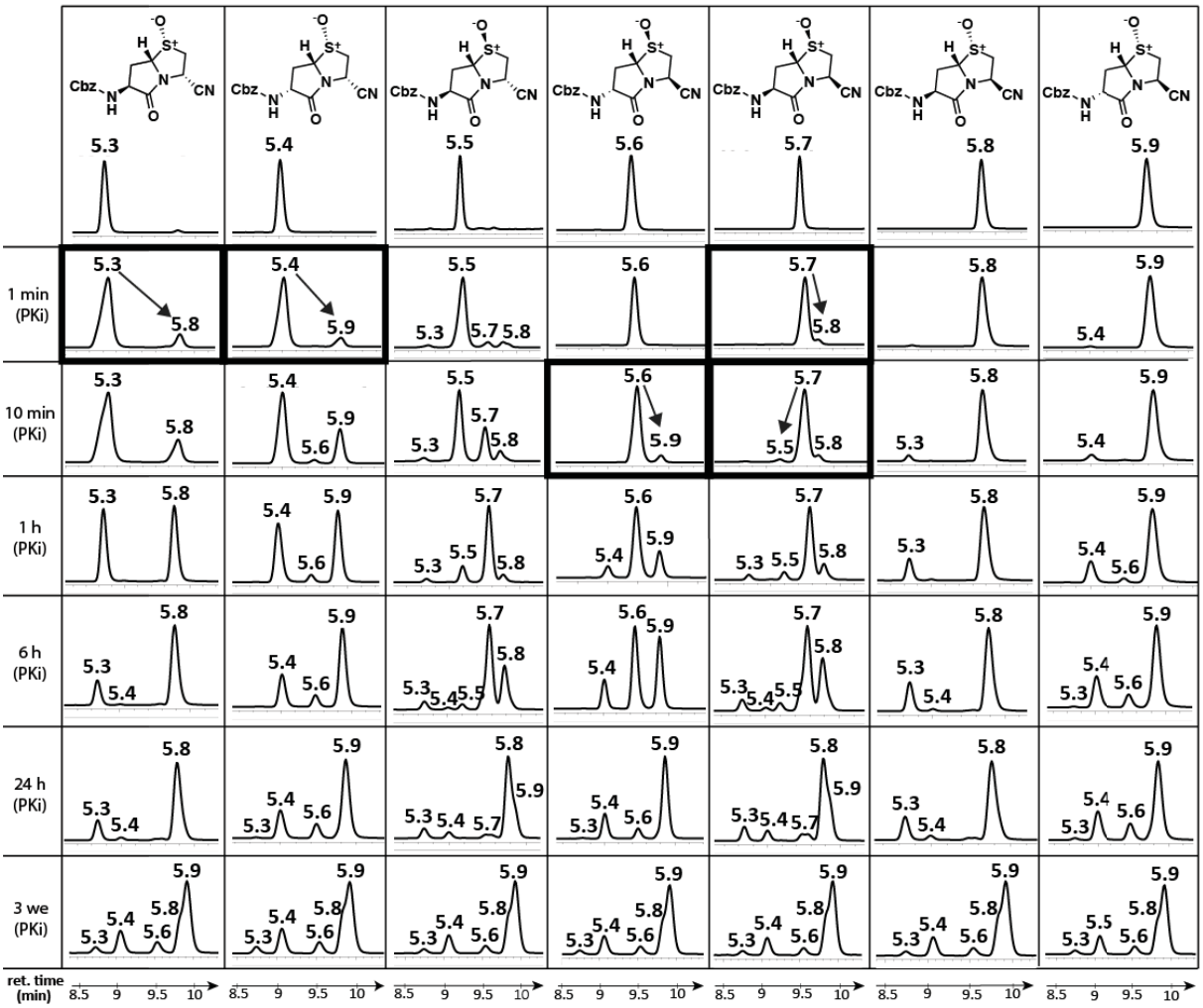


Figure 5.6 Equilibration of each of compounds **5.3**-**5.9** separately when incubated in PKi at 37°C over time. The arrows in the highlighted graphs show pairs of compounds that are assumed to differ at only one stereocenter. All the LC-MS traces are reported as SIM (Single Ion Monitoring m/z = 334).

In addition to the presence of clear kinetic trends (for example, the rate of interconversion between **5.3** and **5.8** parallels the one between **5.4** and **5.9**, in both directions), it is remarkable that each of the seven compounds generates exactly the same mixture after 3 weeks, an observation that confirmed that all seven molecules existed in equilibrium with each other under the experimental conditions.

To rationalize the interconversion pattern, we analyzed the likelihood of epimerization at each of the four stereogenic centers of **5.3**-**5.9**. The C-1 α to the CN group (Figure 5.7A) was deemed fairly prone to epimerization due to the slightly acidic character of the α-proton. Inversions of stereochemistry at the sulfoxide moiety and at position C-3 (ring junction) were also considered realistic. Indeed, previous studies on thiazolidine sulfoxides have shown that these systems can undergo ring opening, allowing the neighboring C-S bond to freely rotate and the sulfur to re-attack the C-3 on both faces of the cycle[153] (Figure 5.7B). Finally, inversion of the stereochemistry at C-5, for the same class of compounds, has been previously reported by Hanessian *et al.*[154]. A proposed mechanism of epimerization at this position is illustrated in Figure 5.7D.

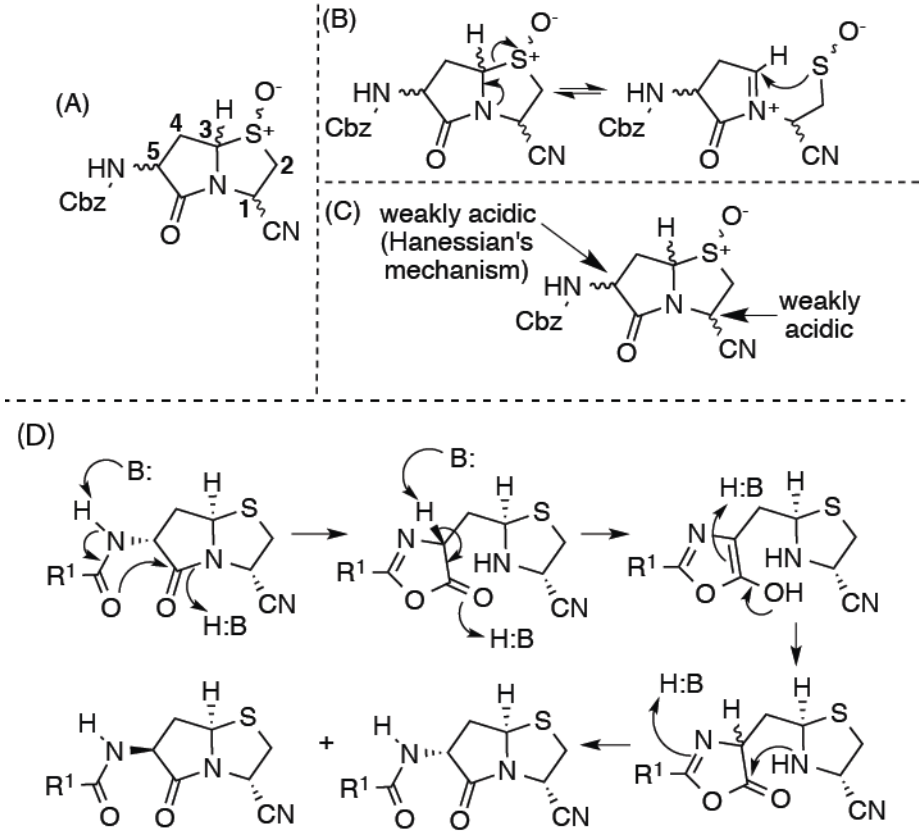


Figure 5.7 Proposed mechanisms of epimerization for **5.3-5.9**. (A) Carbon labeling for the structural scaffold of **5.3-5.9**. (B) Proposed mechanism of epimerization at S-O and C-3 for **5.3-5.9**. (C) Additional stereocenters considered prone to epimerization. (D) Proposed mechanism of epimerization at C-5 based on Hanessian's theory for related compounds

We envisaged that our quest for structural assignment could benefit from looking at the chronological order of appearance of the HPLC peaks in the equilibration studies. In fact, when decrease of a peak is concomitant with appearance of a new one, one can assume that the corresponding compounds have structures differing at C-1 only, C-5 only, C-3 only, S-O only, or both C-3 and S-O. In particular, if we compare the equilibration profiles in the timeframe 1-10 min (Figure 5.5), the following pairs of compounds are assumed to have such correlation: (**5.3** and **5.8**); (**5.4** and **5.9**); (**5.6** and **5.9**); (**5.7** and **5.8**); (**5.5** and **5.7**). This assumption was implemented into the MatLab protocol, and allowed us to decrease the number of combinations from 40320 to 2304 (Figure 5.3A).

In a control experiment, substrates **5.1** and **5.2** were subjected to the same conditions that triggered epimerization of **5.3-5.9** (PKi, 37˚C, 1 h shaking). Surprisingly, LC-MS analysis revealed significantly higher stability of **5.1** and **5.2** (Figure 5.8) compared to the respective sulfoxide derivatives.

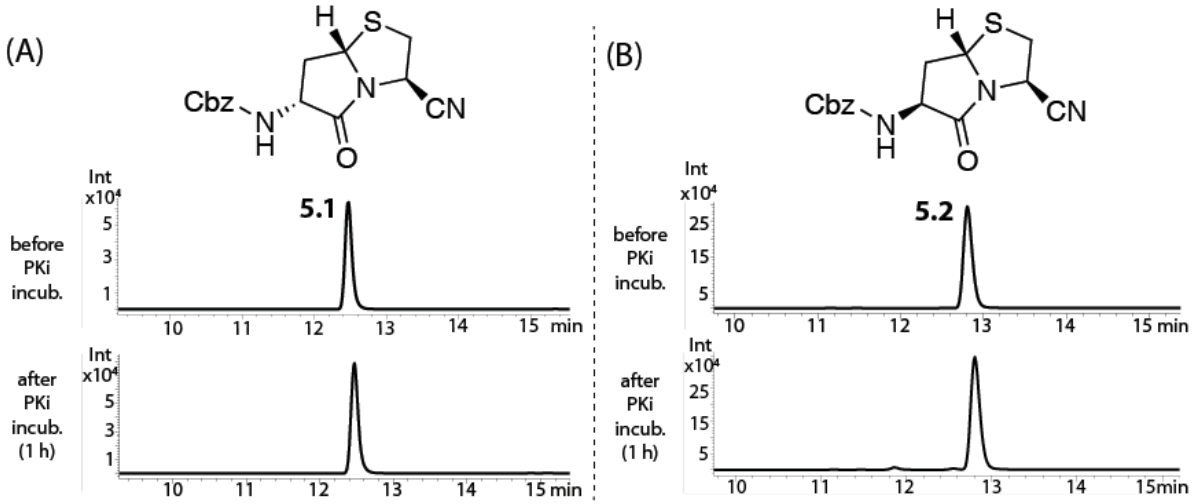


Figure 5.8 Equilibration studies for **5.1** and **5.2**. Monitoring of $[M+H]^+$ (m/z = 318) by LC-MS during analysis of the stability of compounds **5.1** (A) and **5.2** (B) towards epimerization, upon incubation in PKi at 37˚C for 1 hour.

Although no definitive conclusions can be made, one can speculate on the role of the additional oxygen in increasing the isomerization rate. First of all, the H α to the CN group is thought to acquire a stronger acidic character in the presence of the sulfoxide moitety (sulfur has a partial positive charge). Second, the hypothesized intramolecular ring opening would not be as favorable in the absence of the S-O oxygen that is able to delocalize the lone pair of the nitrogen at the ring junction. On the contrary, we believe that epimerization at C-5, further

away from the sulfur, is only slightly affected by the presence of the sulfoxide. As Figure 5.6 illustrates, < 1 % of **5.4** is generated from **5.3** after 6 h incubation in PKi.

5.7 Time-based metabolism studies

In the following experiment, the P450 metabolism analysis for **5.1** and **5.2** was repeated with the exception that the formation of the oxidation products was monitored over time (Figure 5.9). For both **5.1** and **5.2** mainly one compound each (**5.6** and **5.7**, respectively) was rapidly produced in the first 10 minutes of incubation, whereas secondary peaks appeared only at a later stage, which highly resembled what had been previously observed during equilibration studies (Figure 5.6). We thus concluded that compounds **5.6** and **5.7** were the major metabolites of **5.1** and **5.2**, respectively, and that the mixtures of products originally observed (Figure 5.1B-C) were the result of the isomerization of the primary metabolites in the medium used for the metabolism assay. These findings were translated into the following statement: **5.6** and **5.7** must have unchanged stereochemistry at position C1, C3 and C-5 with respect to corresponding P450 substrates **5.1** and **5.2**. By adding these constraints as exclusion rules to the MatLab protocol, the number of possible combinations was further decreased to 168.

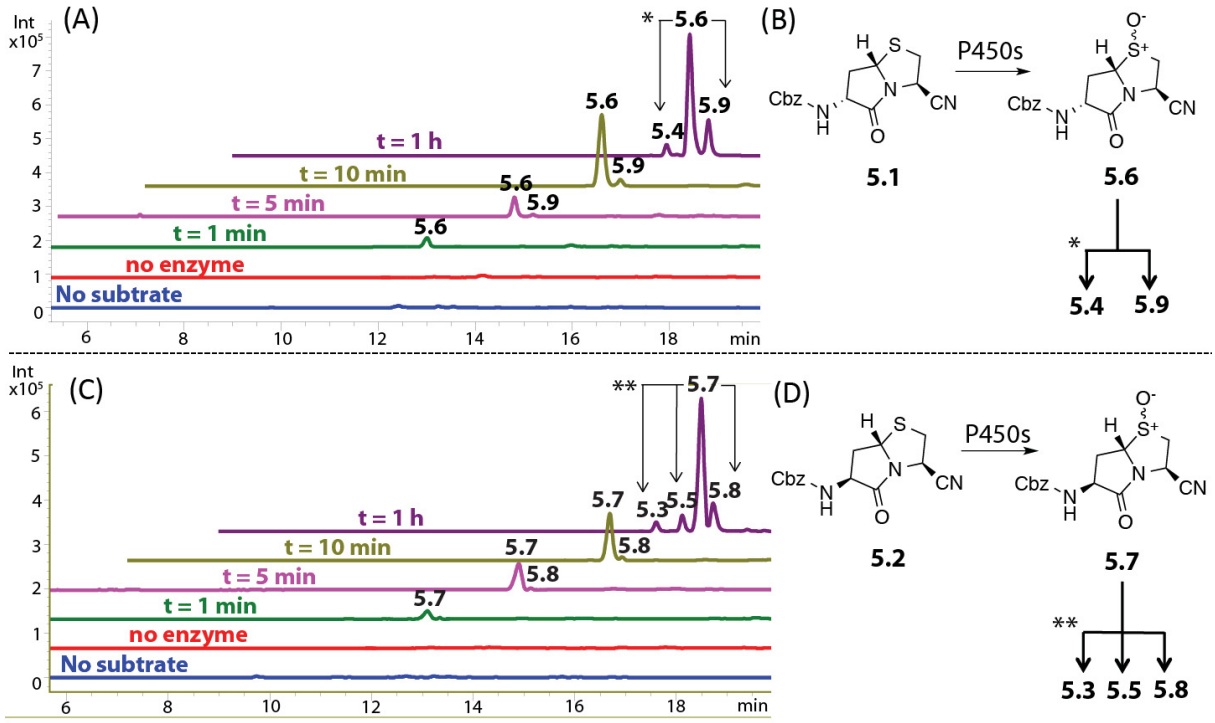


Figure 5.9 Time-based metabolism studies of **5.1** and **5.2**. Monitoring of $[M+O+H]^+$ ions (m/z = 334) after incubation of **5.1** (A) and **5.2** (C) with human liver microsomes (1 mg/ml) in PKi (0.1 mM, pH 7.4), for different periods of time, at 37°C. (B, D) Schematic representation of the sequence of chemical transformations happening during incubation of **5.1** and **5.2** with human liver microsomes.

5.8 A closer look at the synthetic mixtures

As mentioned, attempts to use NMR to characterize the synthetic standards **5.3**-**5.9** (obtained, as mixtures, Figure 5.2) were unsuccessful. In fact, the simple process of removing the solvent after semi-preparative HPLC purification was sufficient to observe ≥30% isomerization. However, valuable information was retrieved by carefully analyzing the profiles of the synthetic products mixtures and by comparing them with previously collected data. In particular, if we look at the equilibration profile of **5.9** after 1 hour of incubation in PKi (Figure 5.10A, extracted from Figure 5.6), we notice its high resemblance to the mixture profile from chemical oxidation of **5.1** (Figure 5.10B). This suggests that the reaction of **5.1** with ammonium persulfate initially produces **5.9**, which later undergoes isomerization to yield **5.4** and **5.6**. In other words, **5.9** is a sulfoxide derivative of **5.1** but with no change in the initial stereochemistry at C-1, C3 or C-5 (Figure 5.10C). This information can be easily translated in additional exclusion rules for the MatLab protocol. Similarly, comparison between the equilibration profile of **5.8** after 1 hour incubation in PKi (Figure 5.10D), and the mixture

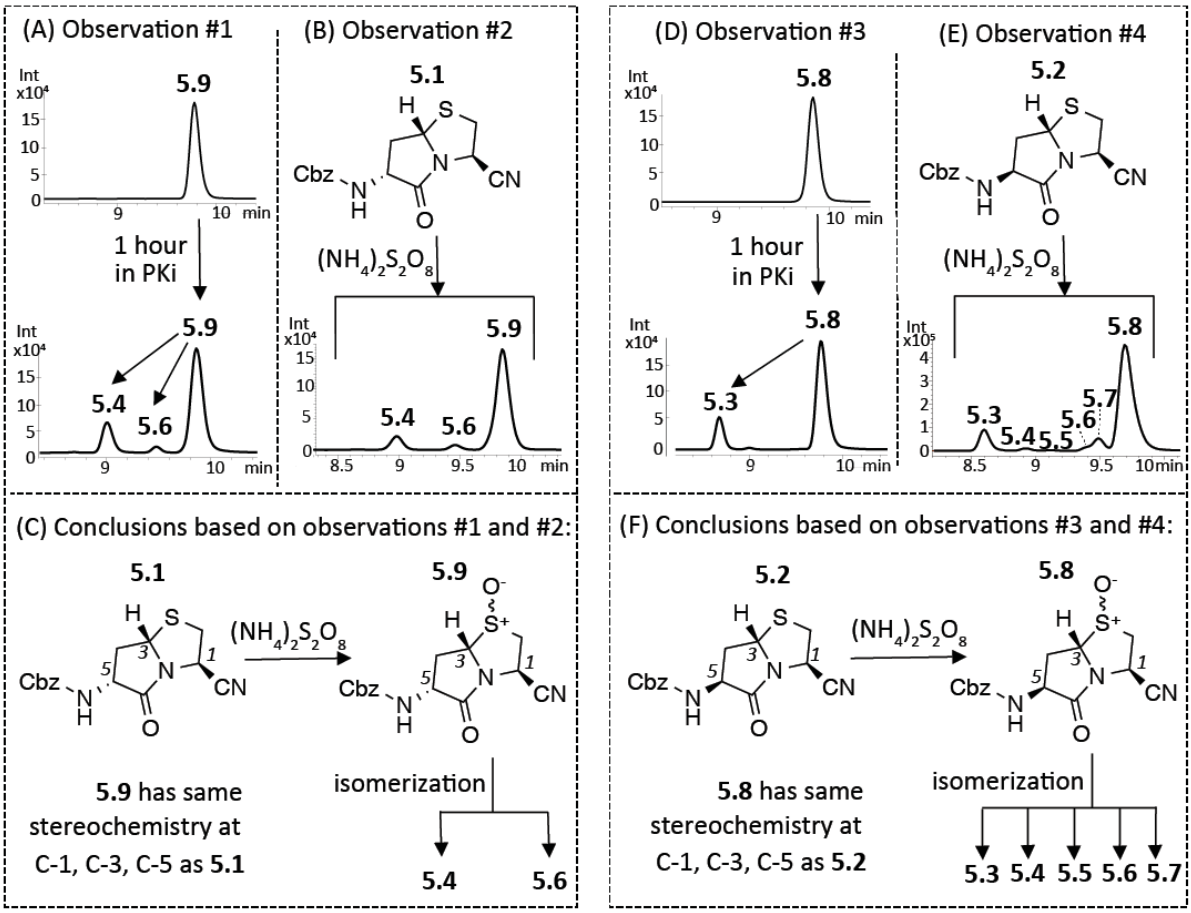


Figure 5.10 Equilibration profiles of **5.9** (A) and **5.8** (D) when isolated and incubated in PKi for 1h; mixture of products from chemical oxidation of **5.1** (B) and **5.2** (E). Partial assignment of the stereochemistry of **5.9** (C) and **5.8** (F) based on observations illustrated in panels A-D.

profile from chemical oxidation of **5.2** (Figure 5.10E), suggests that **5.8** is a sulfoxide derivative of **5.2** but with no change in the initial stereochemistry (Figure 5.10F). These last conclusions about the nature of **5.8** and **5.9** led to a decrease in the number of combinations from 168 to 16. At this stage of the study we knew that **5.6** (the product of HLM microsomal oxidation of **5.1**) and **5.9** (the product of chemical oxidation of **5.1**) differed only by the stereochemistry of the sulfur atom. Interestingly, this means that P450s are able to oxidize **5.1** with a stereoselectivity opposite to that of standard chemical methodology. The same can be concluded for the **5.2** which produces **5.7** with P450 catalysis, and **5.8** with $(NH_4)_2S_2O_8$.

5.9 Deuterium studies

When **5.3**-**5.9** are incubated separately in phospate buffer prepared with D_2O (d-PKi) instead of H_2O, replacement of hydrogen with deuterium atoms within the compounds can be monitored by mass spectrometry. Since initially the sole source of deuterium atom is the deuterated solvent, it is possible to use LC-MS to determine whether conversion of one compound into another involves intermolecular exchanges between metabolites and the solvent, or if the equilibration happens intramolecularly (Figure 5.11).

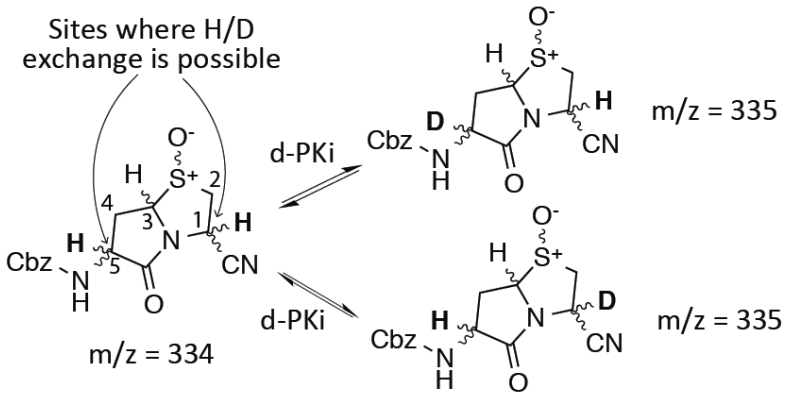


Figure 5.11 When **5.3**-**5.9** are incubated in d-PKi, H/D exchange can only involve positions C-1, C-5 and carbamate N-H

One of the hurdles that we encountered in the development of this assay was elucidating the behavior of the carbamate N-H. When compounds **5.3**-**5.9** are isolated and analyzed by LC-MS using acetonitrile and D_2O as eluents for the LC separation, an increase of two mass units (compared with the mass of the molecules analyzed using regular H_2O) is observed (Figure 5.12B). Since this difference in mass is observed without the appearance of another chromatographic peak, it was concluded that the +1 mass increase was due to the replacement of a H^+ with a D^+ in the ionization step, whereas the second +1 was the result of H/D exchange at the carbamate moiety. This is, in fact, the only reasonable site where H/D exchange does not result in production of a different isomer. By relying on the assumption that the rate of N-H/D exchange is comparable to the rate of N-D/H, N-D is expected to be almost quantitatively replaced by N-H during the LC separation when H_2O is used as one of the two eluents. It was

thus concluded that H/D exchange at the carbamate moiety does happen, but that it does not significantly affect the results of the assay when HPLC separation is performed using an acetonitrile/H_2O mixture which reverts the exchange back to hydrogen.

Thus, when the four stereogenic centers are considered, only epimerization at C-1 or C-5 is expected to proceed via incorporation of a D atom from D_2O. These are the only positions where inversion of stereochemistry involves H exchange with the solvent, whereas a lack of incorporation of deuterium is only consistent with epimerization at either C-3 or S-O.

As shown in Figure 5.11D, upon incubation in d-PKi for 1 hour, isomer **5.8** equilibrated to **5.3** with the incorporation of one D atom, suggesting that the two compounds differ only by the stereochemistry at C-1 or C-5.

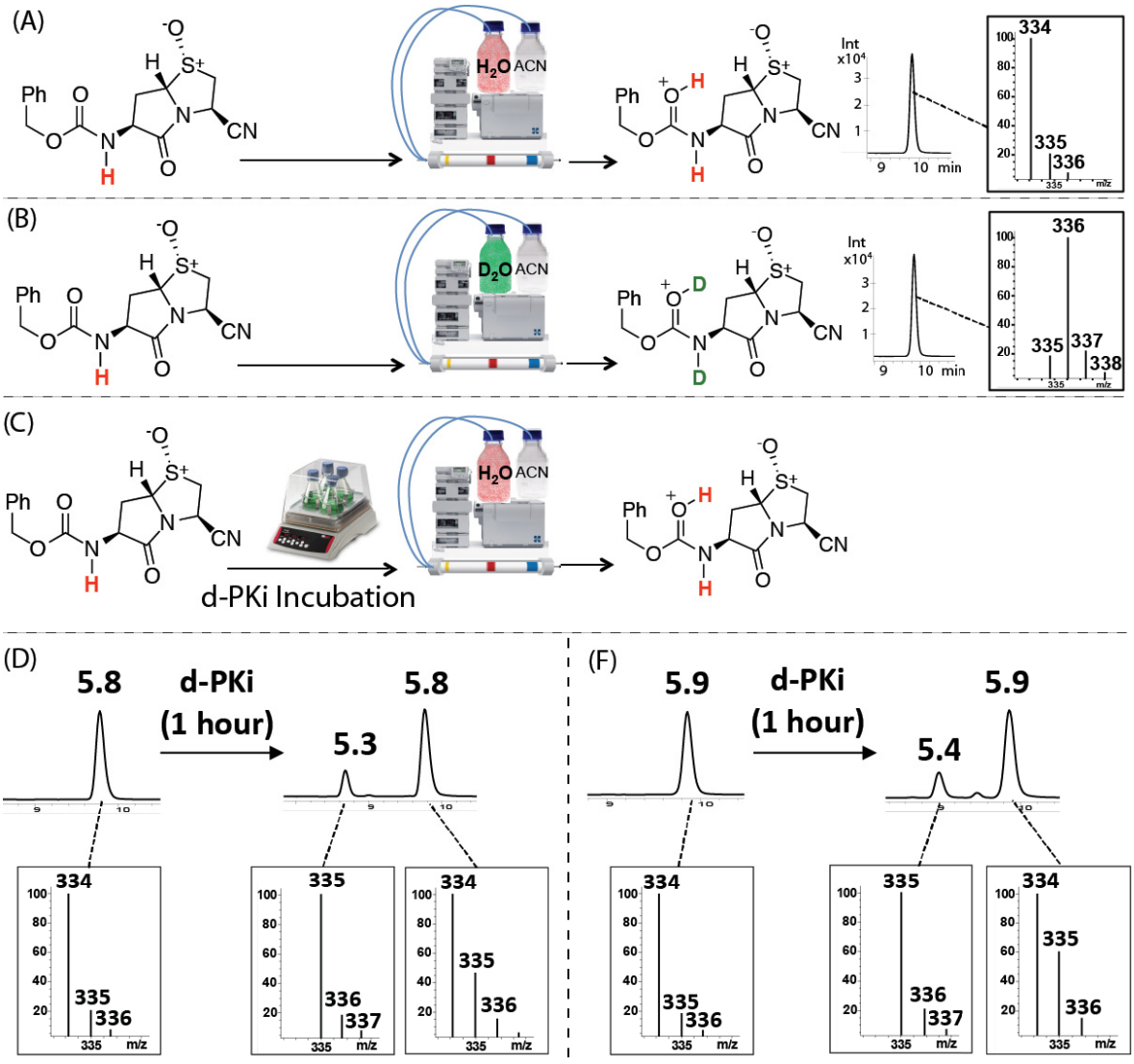


Figure 5.12 H/D exchange studies. (A, B, C) Control experiments assessing H/D exchange at the carbamate moiety exemplified for **5.8**. (A) LC-MS analysis of **5.8** using H_2O and ACN as HPLC eluents. (B) LC-MS analysis of **5.8** using D_2O and ACN as HPLC eluents. (C) Hypothesized N-H/D behaviour for **5.8**, based on experiments (A) and (B). (D, F) LC-MS traces revealing incorporation of D when **5.8** and **5.9** are incubated in d-PKi for 1 hour, at 37ºC. The HPLC separation was performed using H_2O and acetonitrile as eluents.

The interconversion between **5.4** and **5.9** paralleled that of **5.3** and **5.8**, with one D atom incorporated. The deduced structural relationship between **5.3** and **5.8**, **5.4** and **5.9**, further narrowed down the number of combinations from 16 to 6.

5.10 Site of metabolism prediction and resolution of the problem

To narrow it down to one possible combination we took advantage of our recently reported software IMPACTS (In-silico Metabolism Prediction by Activated Cytochromes and Transition States)[141]. As described in chapter 1 (section 1.10), this program is designed to predict the site of metabolism (SoM) of xenobiotics by P450s based on a combination of docking, transition state modeling and rule-based substrate reactivity prediction, and it has been validated on a wide variety of drugs and small molecules including a diverse set of functional groups. IMPACTS was applied to substrate **5.1** and **5.2** with P450s 3A4, 2C19, 2C9, 2D6 and 1A2, separately. For all five P450s, the sulfur atom was predicted to be the most likely SoM (Figure 5.13), in agreement with the experiments, while other possible chemical transformations were ranked as significantly less favorable.

In the case of prochiral centers, IMPACTS also provides the most favorable stereoisomer formed. Figure 5.13 shows the predicted metabolites of **5.1** and **5.2** (**5.6** and **5.7**, respectively). Relying on this program for the assignment of the stereochemistry at the sulfur atom in the structures of the primary metabolites allowed us to arrive at a final solution: only one diastereomeric structure for each of the seven HPLC peaks.

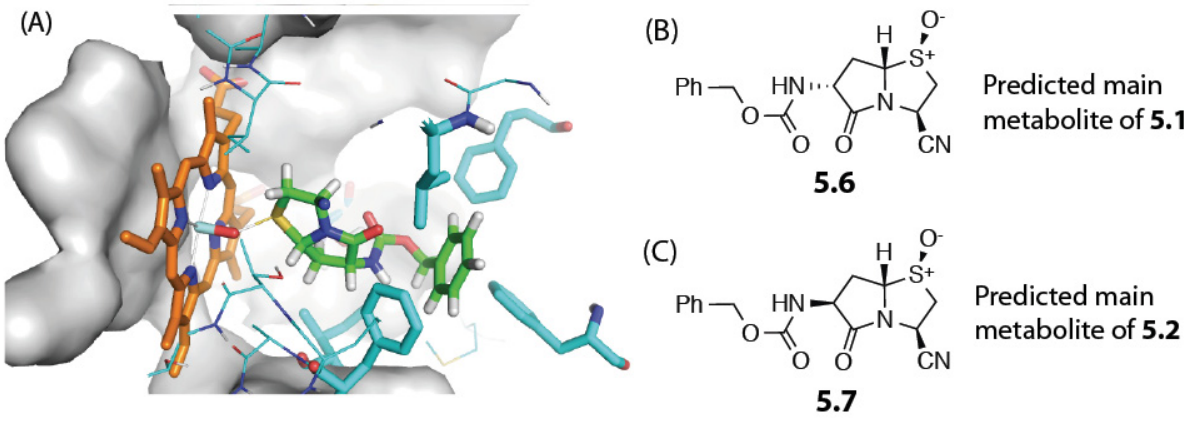


Figure 5.13 IMPACTS metabolism predictions for **5.1** and **5.2**. (A) Predicted transition state for the reaction of **5.1** with P450 3A4 (orange: enzyme heme group; green: compound **5.1**; light blue: protein residues surrounding the ligand; grey: protein binding site surface); predicted metabolites of **5.1** (B) and **5.2** (C) for the five major P450s.

5.11 Ground state energies

To further validate our conclusions based on the predictions by IMPACTS, additional computations were carried out (Figure 5.14) The equilibration studies revealed that each compound generated the same mixture after 3 weeks of incubation. We rationalized that this carried information about the concentrations at the thermodynamic equilibrium, and that the relative concentrations could be explained by ground state energy differences. A number of actions from our in-house computational platform FORECASTER[112] were then used to carry out a conformational search on the 8 possible diastereomers and their 8 enantiomers followed by energy minimization to obtain a ranking from most to least energetically favorable.

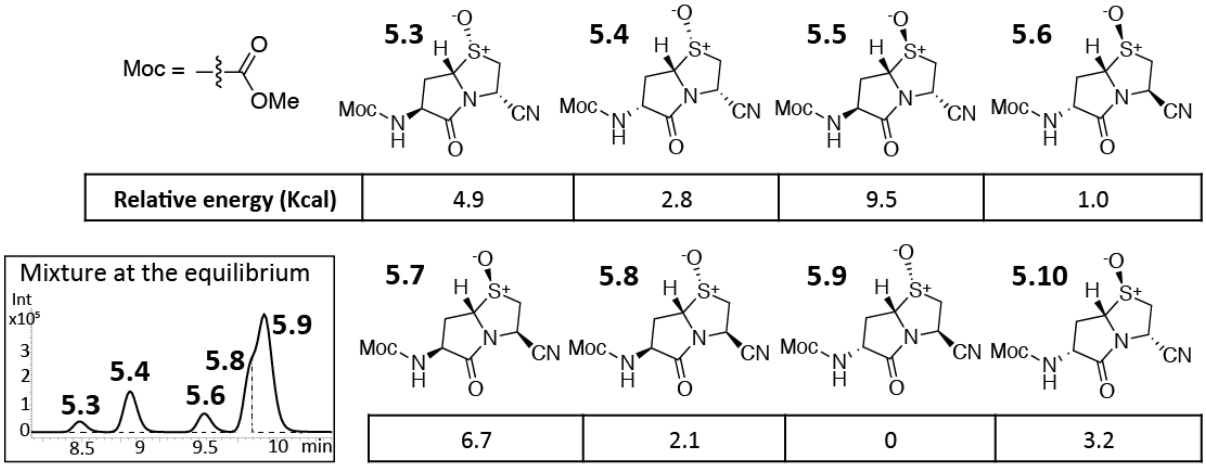


Figure 5.14 Calculated energies for **5.3**-**5.10** (kcal), reported as relative to the predicted most stable isomer, **5.9**. The experimental order of stability can be deduced from the LC-MS peak areas after 3 weeks of incubation in PKi, when thermodynamic equilibrium is reached.

In general, the predictions are in good agreement with our structural assignments. This is particularly true for the lowest-in-energy compound, **5.9**, which is experimentally the major one (> 50%) at the thermodynamic equilibrium, and for compounds **5.5** and **5.7** which have very high potential energies and are not detected experimentally when the thermodynamic equilibrium is reached. The difference between the experimental and calculated data is larger for the other compounds, probably because of the intrinsic errors associated with the computational method itself. Another explanation for the slight mismatch between experiments and computations could be related to **5.10**, the unobserved diastereomer. In fact,

it is possible that the peak for **5.10** was hiding beneath one of the other seven, under the experimental conditions. Different HPLC separation methods were attempted to investigate this hypothesis (data not shown), but never more than seven peaks were observed. Encouragingly, the best match between experimental concentrations and calculated energies overlaps with the predictions by IMPACTS. The agreement between the two very different computational methods used here (docking-based method IMPACTS and DFT) provides a high-confidence in the computational study and in the final structural assignment.

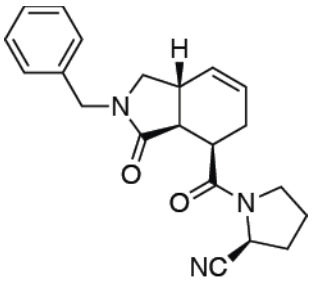
5.12 Conclusions

In the study described herein, we have shed light on a complex oxidation/epimerization process leading to several metabolites. Although this study was centered on a small number of inhibitors specifically designed to target POP, our discoveries may have wider implications by revealing some of the issues associated with the use of cyanothiazolidine moieties as drug candidates. For example, the bicyclic scaffold of **5.1** and **5.2** has been patented as part of a series of compounds[155] targeting DPP-IV, a prolyl peptidase associated with type-2 diabetes[156]. Here too, one might expect sulfur oxidation followed by a sequence of isomerizations similar to the ones that we have discovered. Such a large number of metabolites is an important liability in drug development. Not only would toxicological studies on every isomer be challenging, but the process of ring opening could also trigger subsequent harmful covalent modifications in biological systems. In fact, the hypothesized iminium ion intermediate could exist in solution long enough to react with surrounding biomolecules. It is also reasonable to think that cyanothiazolidine moieties alone, even when not contained within a bicyclic scaffold, could be prone to the same instability issues following sulfur oxidation. These moieties are found in other previously reported DDP-IV inhibitors for example[157-158].

The results presented here have provided guidance for others in designing a second generation of POP inhibitors lacking the liabilities associated with bicyclic cyanothiazolidine moieties, while maintaining high potencies. For example, a series of hexahydroisoindoles is looking very promising[159].

5.13 Additional contributions to the study of the metabolism of POP inhibitors

As part of the development of a new series of covalent inhibitors of POP, compound **5.11** (Figure 5.15) was designed, synthesized and tested vs POP by members of the Moitessier research group. Despite the high potency of **5.11** (K_i (POP) = 1 nM), one of the concerns was the numerous possible sites of metabolism.



5.11

K_i (POP) = 1 nM

Figure 5.15 Structure and potency vs POP of **5.11**

In fact, both the double bond and the *N*-benzyl moiety may lead to reactive metabolites and/or low metabolic stability. The former may be prone to epoxidation, while the latter may undergo *N*-debenzylation.

Unexpectedly, **5.11** was found to be very stable towards human liver microsomes, with a measured Cl_{int} = 4 µL/min/mg protein. Under the experimental conditions used, no *N*-debenzylated products were observed, and only trace amounts of mono-oxidation metabolites were detected (Figure 5.16)

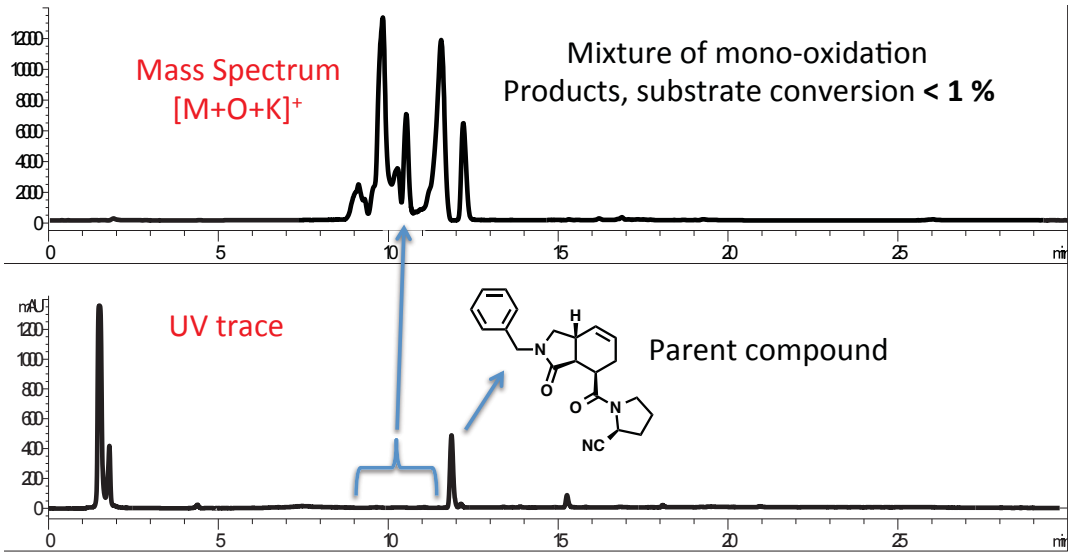


Figure 5.16 LC-MS analysis of metabolism of **5.11**. Top: monitoring of $[M+O+K]^+$ ions (m/z = 404) after incubation of **5.11** with human liver microsomes (1 $mgmL^{-1}$) in potassium phosphate buffer (PKi, 0.1 M, pH 7.4) for 1 h at 37°C. Bottom: UV trace and structure of **5.11**

The results presented in this section was published in the following article:

Mariaule, G.; De Cesco, S.; Airaghi, F.; Kurian, J.; Schiavini, P.; Rocheleau, S.; Huskić, I.; Auclair, K.; Mittermaier, A.; Moitessier, N., 3-Oxo-hexahydro-1H-isoindole-4-carboxylic Acid as a Drug Chiral Bicyclic Scaffold: Structure-Based Design and Preparation of Conformationally Constrained Covalent and Noncovalent Prolyl Oligopeptidase Inhibitors, *J. Med. Chem.* (**2015**), 59 (9), 4221–4234

CHAPTER 6 - CONTRIBUTIONS AND FUTURE DIRECTIONS

The goal of this work was to contribute to the current understanding of how protein structure relates to function and, at the same time, demonstrate that experimental and computational approaches are synergistic. The results presented here contributed to the scientific knowledge in several ways. First, we demonstrated how the conformational flexibility of KAR receptors can be investigated by combining the use of rationally designed small molecular probes and flexible docking. Next, rational mutagenesis was used to investigate the potential of P450s for biocatalytic applications and to contribute to the development of a new computational tool for rational protein engineering. By establishing a new strategy coined "active site crowding" we also demonstrated how the regio- and stereoselectivity of P450s can be altered for the preparation of useful hydroxylated products. Finally, we explored the unique properties of P450s in the context of drug metabolism. By breaking down a complex case of metabolic instability, we revealed the ability of these enzymes to catalyze highly stereoselective sulfoxidation reactions, while at same time we validated a program under refinement for the prediction of the site of metabolism by P450s.

6.1 Study of the conformational flexibility of KARs

Incorporating protein dynamics into rational drug design is still one of the major challenges in drug discovery. To account for this variable without the need of highly time-consuming computations, a promising *in silico* strategy is the use of flexible docking.

The results presented in Chapter 2 demonstrate how flexible docking can be used for the study of KAR receptors. In particular, by taking advantage of our program FITTED, we designed a small library of L-glutamate analogues to probe the conformational flexibility of GluK2 and investigate the role of a particular residue in the active site.

In particular, we showed that targeting Tyr488 in the ligand binding domain could be important in the development of effective antagonists. Ensuring that this interaction takes place could represent a new potential approach to develop molecules that are able to antagonize GluK2 and eventually be useful in the treatment of neuropathological disease states. For this purpose, the chemico-physical properties of the antagonist will have to be optimized to facilitate crossing of the blood brain barrier (BBB), one of the main obstacles in

the development of drugs targeting the central nervous system. A common way to favor this process is to increase the lipid solubility of the drug candidate. Starting from our β-substituted glutamic acids, this could be achieved, for example, by replacing the γ carboxyl group with an ester group within a five-membered ring (Figure 6.1). This rigidification could also help decreasing the entropic energy of binding.

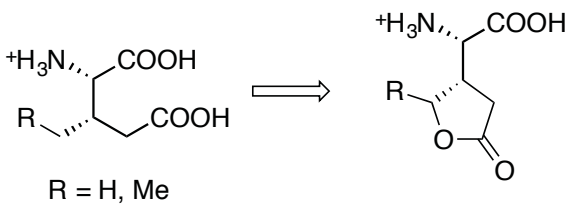


Figure 6.1 Suggested optimization strategy to increase crossing of the blood brain barrier

6.2 Rational engineering of P450s for biocatalytic applications

Biocatalysis is a powerful approach for the preparation of organic compounds under environmentally-friendly reaction conditions.

The results described in Chapter 3 contribute to the field of biocatalysis by 1) investigating the use of P450 enzymes as biocatalysts for regio- and stereoselective hydroxylation reactions; 2) contributing to the development of a new computational tool for the rational engineering of P450s. In an attempt to develop a chemo-enzymatic synthesis of latanoprost, the remarkable capabilities of human P450 3A4 were highlighted. The reaction catalyzed, a stereoselective hydroxylation of a non-activated C-H bond, in position α to a chiral alcohol group, was not the desired one for the preparation of latanoprost, but it still highlights the potential of P450s to carry out reactions that are very challenging using traditional methodologies.

Computer-aided engineering of P450 3A4 was attempted, and despite being unsuccessful, important lessons were learned. It is worth remembering that during the development of a new computational tool, initial predictions are made on a limited amount of existing information, new experimental data is used to validate and/or refine the software, which in turn can lead to better predictions.

Based on our results, it appears that further improvement is required for the software under development in our lab. This will require a large amount of experimental data from high-throughput screening of mutants of P450 3A4. This could be achieved by taking advantage of our purification-free experimental protocol described in Chapter 3. Mutant libraries could be selected for mutations in the active site only, in order to reduce the complexity of the *in silico* model. Investigations could be extended to different P450s including not only mammalian but also microbial P450s such as P450 BM-3, which is one of the most studied P450s.

The findings reported in Chapter 4 demonstrate how the field of enzyme engineering can greatly benefit from rational approaches that do not rely solely on computations, but are instead founded on a series of general principles guided by chemical intuition. "Active site crowding" was shown to be a powerful way of altering the selectivity of P450 3A4, in particular by favoring oxidation at less-hindered positions on the substrate.

As a continuation of this study, further crowding of P450 3A4 could be explored in order to shift/improve even more the selectivities. To do this, mutant L211F/I301F/I369F/G481W/L482F (M62), which has shown activity comparable to wild-type, could be used as a starting point for additional mutations.

Although the majority of the active site residues closest to the heme prosthetic group have already been investigated, mutating some residues further away could be beneficial. Among these, Val240, Ile300 andVal313 (Figure 6.2A) could, for example, be mutated to Phe.

An alternative strategy to further increase active site crowding could be focusing on the residues that have been already mutated to Phe and replacing them with Trp instead, for example in mutations such as L211W, I369W, L482W. According to preliminary computations, no particular steric clash is predicted for "super crowded" P450 3A4 mutant L211F/I301F/I369F/G481W/L482F (Figure 6.2B), yet this variant may still bind substrates of size comparable to (*R*)-lisofylline or progesterone.

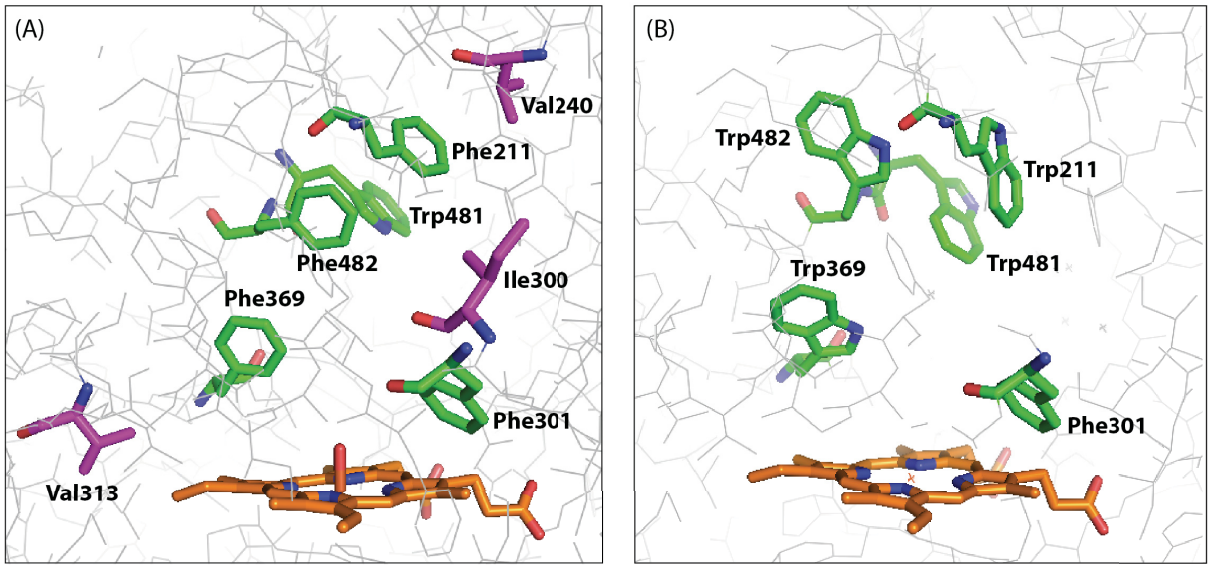


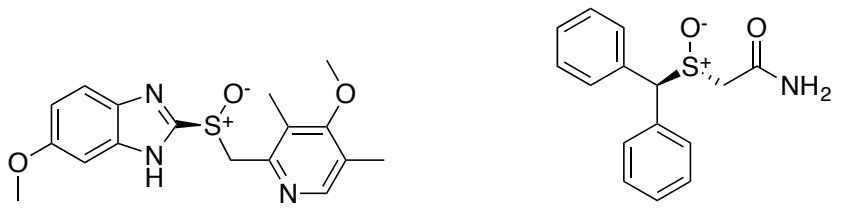
Figure 6.2 Suggested strategies to further improve active site crowding of P450 3A4. (A) Predicted conformation of mutant L211F/I301F/I369F/G481W/L482F with additional residues worth mutating highlighted in magenta. (B) Predicted conformation of mutant L211W/I301F/I369W/G481W/L482W

6.3 Metabolism of POP inhibitors and P450-catalyzed sulfoxidation reactions

In the last 10 years, the availability of crystal structures for the most important human P450s (P450 3A4, 2D6, 2C9, 2C19, 1A2, 2E1) have allowed the development of many computer programs for the prediction of the site of metabolism. However, combining accuracy and low computational time remains a challenge, and experimental validations of new programs is always necessary. The results presented in Chapter 5 have provided additional validation to our software IMPACTS, which was able to predict the correct site of metabolism for the compounds under study.

By complementing *in silico* calculations with a vast assay of experimental techniques, we were able to unveil a complex phenomenon of metabolic instability. Despite this study being focused on a small number of POP inhibitors, our findings may have wider implications in medicinal chemistry. In fact, they revealed the potential pitfalls of cyanothiazolidine-based scaffolds which are found in several drug candidates currently under investigation.

Finally, an interesting link between the observed metabolism of POP inhibitors and the potential use of human P450s as biocatalysts can be made. In fact, the excellent stereoselectivity of the P450-catalzyed reaction of sulfoxidation of POP inhibitors is remarkable. Although several methodologies for enantioselective sulfoxidation have been already reported[160], the use of enzymes to perform such reactions is an interesting alternative[161]. Our results demonstrated the potential of P450s to catalyze, with high stereoselectivity, the formation of compounds with opposite stereochemistry to the one favored through standard chemical methods. Further investigations on the use of human P450s as biocatalysts for enantioselective sulfoxidation reactions would be worth pursuing. In fact, sulfoxide moieties are found in several drugs on the market such as gastric controller (*S*)-omeprazole or wakefulness-promoting agent (*R*)-modafinil (Figure 6.3).



(*S*)-Omeprazole (Nexium)
Gastric controller

(*R*)-Modafinil
wakefulness-promoting agent

Figure 6.3 Examples of two commonly used drugs containing a sulfoxide moiety

6.4 Publications

1. Schiavini, P.; Cheong, K. J., Moitessier, N.; Auclair, K., Active site crowding of P450 3A4 as a strategy to alter its selectivity, *ChemBioChem* (**2017**), 18(3), 248-252

2. Schiavini, P.; Dawe, B.; Bowie, D.; Moitessier, N., Discovery of novel small-molecule antagonists for GluK2, *Bioorg. Med. Chem. Lett.* (**2015**), 25, 11, 2416-2420.

3. Schiavini, P.; Pottel, J.; Moitessier, N.; Auclair, K., Metabolic Instability of Cyanothiazolidine-Based Prolyl Oligopeptidase Inhibitors: a Structural Assignment Challenge and Potential Medicinal Chemistry Implications, *ChemMedChem* (**2015**), 10, 1174-1183

4. Mariaule, G.; De Cesco, S.; Airaghi, F.; Kurian, J.; Schiavini, P.; Rocheleau, S.; Huskić, I.; Auclair, K.; Mittermaier, A.; Moitessier, N., 3-Oxo-hexahydro-1H-isoindole-4-carboxylic Acid as a Drug Chiral Bicyclic Scaffold: Structure-Based Design and Preparation of Conformationally Constrained Covalent and Noncovalent Prolyl Oligopeptidase Inhibitors, *J. Med. Chem.* (**2015**), 59 (9), 4221–4234

CHAPTER 7 - Experimental methods

7.1 Instruments

HPLC–UV-MS analysis of the enzymatic reaction mixtures was performed on an Agilent 1100 modular system equipped with an autosampler, a quaternary pump system, a photodiode array UV detector, a quadrupole MS detector, and a ChemStation (for LC 3D A.09.03) data system.

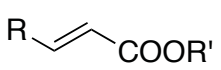
LC-MS/MS fragmentation studies and protein HPLC–MS studies were performed using a Dionex Ultimate 3000 UHPLC coupled to a Bruker Maxis Impact Quadrupole-Time of Flight mass spectrometer in positive ESI mode. Data were processed with Bruker Data Analysis software, version 4.1.

7.2 Chemicals

All the chemicals for synthetic procedures were purchased from Sigma-Aldrich unless otherwise stated. (*R*)-lisofylline, (±)-lisofylline, 2β-hydroxyprogesterone, 6β-hydroxyprogesterone were purchased from Santa Cruz Pharmaceuticals.

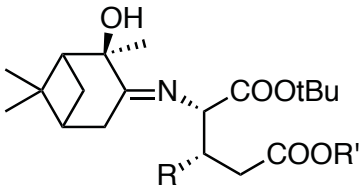
Reagents for site-directed mutagenesis were purchased from Agilent, except for DNA Primers that were purchased from Life Technologies. Human Liver Microsomes (Ultra PoolTM 150) were purchased from BD Gentest.

7.3 Synthetic procedures

General procedure for synthesis of compounds 2.13-2.15. 

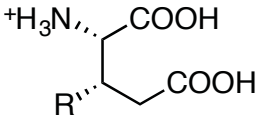
Methyl (triphenylphosphoranylidene)acetate (**2.25**, 5 g, 15 mmol) was suspended in benzene (30 ml) and the aldehyde (**2.22**-**2.24**, 3.3 ml, 44 mmol, 3 eq) was added. After refluxing for 11 h under stirring, the mixture was washed with water (30 × 2 ml) and the organic extract was concentrated in vacuo. The white solid formed upon concentration was filtered, rinsed with petroleum ether and discarded. The remaining oil was purified by flash column chromatography (hexane / ethyl acetate, gradient from 99 / 1 to 90 /10).

General procedure for synthesis of compounds 2.17-2.21

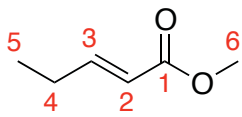


Glycine *t*-butyl ester in its free amine form (1.2 g, 8.9 mmol, 1.5 eq) was added to a solution of the chiral auxiliary (1*R*, 2*R*, 5*R*)-(+)-2-hydroxy-3-pinanone (1 g, 7.1 mmol) in benzene (30 ml). BF_3•Et_2O was added and the mixture was refluxed using a Dean-Stark apparatus for 3 h under stirring. The solvent was then evaporated under high vacuum and the crude reaction crude used within 2 h for the next synthetic step. The crude was dissolved in anhydrous THF and CH_3MgBr (3 M solution in Et_2O, 2.4 ml, 1.2 eq) was added at -30°C. After 3 min DBU (1 ml, 1.1 eq) was added and the reaction stirred for 20 min before the unsaturated ester (**2.12**-**2.16**, methyl cinnamate or ethyl *trans*-but-2-enoate, 1.2 eq) was added at -30°C. After stirring for 30 min at room temperature a NH_4Cl saturated solution (30 ml) was added and the aqueous phase extracted with EtOAc. The combined extracts were dried over Na_2SO_4, evaporated under reduced pressure and the two diastereomers purified by flash chromatography (ether / petroleum ether, gradient from 7 / 93 to 70 /30).

General procedure for synthesis of compounds 2.4-2.8

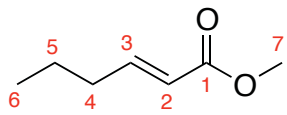


The desired Michael adduct **2.17-2.21** (300 g) was suspended in a 3 M HCl aqueous solution (20 ml). The mixture was refluxed for 3 h, then washed with DCM (3 × 40 ml) and finally concentrated under vacuum. Crude compounds were purified by preparative HPLC. The column used was a Luna 5µ CN 100A, packed 250 × 21.20 mm purchased from Phenomenex. Separation was achieved using mobile phases A (Milli-Q water + 0.05% TFA) and B (acetonitrile + 0.05% TFA) at a flow rate of 3 mL/min with the UV detector set to monitor at 220 nm. Upon equilibration of the column with 99% phase A, elution consisted of gradient step from 99% phase A to 99% phase B over 40 min.

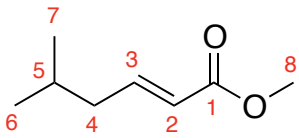


Methyl *trans*-pent-2-enoate (2.13) White oil. R_f 0.4 (2% ethyl acetate in petroleum ether); IR: υ max 2969, 1723, 1661, 1436, 1269, 1176 cm^{-1}; ^1H NMR (300 MHz, $CDCl_3$) δ 7.03 (dt, 1H, J = 15, 7 Hz, H-3), 5.81 (dt, 1H, J = 15, 2 Hz, H-2), 3.73 (s, 3H, H-6), 2.23 (m, 2H, H-4), 1.07 (t, 3H, J = 6 Hz, H-5) ppm; ^{13}C NMR (125 MHz, $CDCl_3$) δ 167.4, 151.2, 120.1, 51.6, 25.5, 12.3; HRMS (ESI):

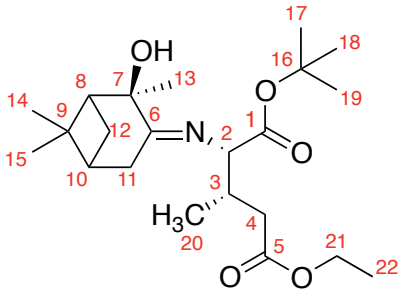
Calculated for $C_6H_{10}O_2Na$ $[M + Na]^+$: 137.0578, found 137.0570. Spectroscopic properties for this known compound match literature values[162].



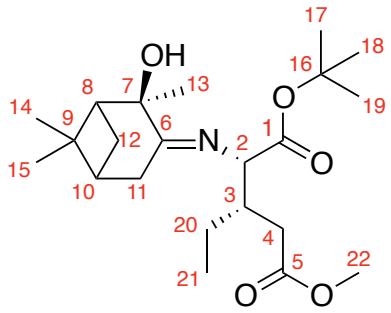
Methyl *trans*-hex-2-enoate (2.14) White oil. R_f 0.43 (2% ethyl acetate in petroleum ether); IR: υ max 2960, 1723, 1658, 1436, 1270, 1173, 979 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 6.95 (dt, 1H, J = 15, 8 Hz, H-3), 5.80 (dd, 1H, J = 15, 2 Hz, H-2), 3.71 (s, 3H, H-7), 2.16 (ddt, 2H, J = 7, 7, 1.5 Hz, H-4), 1.47 (tq, 2H, J = 7, 7 Hz, H-5), 0.93 (t, J = 7 Hz, 3H, H-6) ppm; ^{13}C NMR (125 MHz, $CDCl_3$) δ 167.3, 149.7, 121.1, 51.5, 34.3, 21.4, 13.5; HRMS (ESI): Calculated for $C_7H_{12}O_2Na$ $[M + H]^+$: 151.0735, found 151.0725.



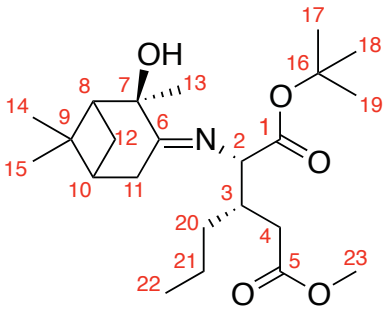
Methyl *trans*-5-methylhex-2-enoate (2.15) White oil. R_f 0.4 (2% ethyl acetate in petroleum ether); IR: υ max 2956, 1723, 1655, 1436, 1269, 1174, 979 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$); δ 6.95 (dt, 1H, J = 15, 7 Hz, H-3), 5.81 (dt, 1H, J = 15, 1.5 Hz, H-2), 3.73 (s, 3H, H-8), 2.09 (ddt, 2H, J = 7, 7, 1.5 Hz, H-4), 1.76 (tq, 2H, J = 7, 7 Hz, H-5), 0.92 (d, J = 7 Hz, 6H, H-6, H-7) ppm; ^{13}C NMR (125 MHz, $CDCl_3$) δ 167.3, 148.8, 122.0, 51.6, 41.7, 27.9, 22.5; HRMS (ESI): Calculated for $C_7H_{12}O_2Na$ $[M + H]^+$: 165.08915, found 165.0886.



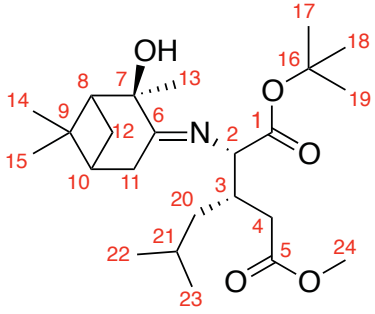
Compound 2.17 White oil.; R_f 0.28 (diethyl ether / petroleum ether: 30: 70); ^1H NMR (400 MHz, C_6D_6) δ 4.08 (d, 1H, J = 6.5 Hz, H-2), 3.98 (m, 2H, H-21), 3.01 (m, 1H, H-3), 2.74 (dd, 1H, J = 15, 4 Hz, H-4), 2.59 (s, 1H, OH), 2.55-2.45 (m, 3H, H-4', H-11, H-12), 2.2 (m, 1H, H-11'), 2.11 (t, 1H, H-8), 1.82 (d, 1H, H-12'), 1.73 (m, 1H, H-10), 1.54 (s, 3H, H-13), 1.33 (s, 9H, H-17,18,19), 1.15 (d, 1H, J = 6.5 Hz, H-20), 1.11 (s, 3H, H-14), 0.96 (t, 3H, J = 6.5 Hz, H-22), 0.68 (s, 3H, H-15) ppm; Spectroscopic properties for this known compound match literature values.[163]



Compound 2.18 White oil.; R_f 0.29 (diethyl ether / petroleum ether: 30: 70); IR: υ max 2920, 1732, 1368, 1151 cm^{-1}; ^1H NMR (500 MHz, C_6D_6) δ 4.32 (d, 1H, J = 4.0 Hz, H-2), 3.36 (s, 3H, H-26), 2.86-2.75 (m, 2H, H-3, H-4), 2.60 (dd, 1H, OH), 2.52 (dd, 1H, J = 14, 7 Hz, H-11), 2.42 (dt, 1H, J = 18, 3 Hz, H-11'), 2.23 (m, 1H, H-12), 2.12 (t, 1H, J = 6 Hz, H-8), 1.84 (d, 1H, H-12'), 1.74 (m, 1H, H-10), 1.56 (s, 3H, H-13), 1.58-1.47 (m, 2H, H-20,20'), 1.35 (s, 9H, H-17,18,19), 1.11 (s, 3H, H-14), 0.89 (t, J = 7 Hz, 3H, H-22), 0.68 (s, 3H, H-15) ppm; ^{13}C NMR (125 MHz, C_6D_6) δ 179.9, 173.2, 170.0, 81.1, 76.6, 64.9, 51.0, 50.6, 40.7, 38.7, 38.3, 34.9, 33.5, 28.7, 28.6, 27.0, 27.4, 24.7, 22.7, 11.9 ppm; HRMS (+ESI): Calculated for $C_{22}H_{37}NO_5Na$ $[M + Na]^+$: 418.2569, found 418.2569.

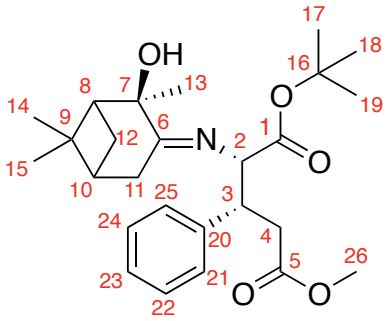


Compound 2.19 White oil.; R_f 0.23 (diethyl ether / petroleum ether: 30: 70); IR: υ max 2922, 1733, 1368, 1151 cm^{-1}; ^1H NMR (500 MHz, C_6D_6) δ 4.32 (d, 1H, J = 4.0 Hz, H-2), 3.36 (s, 3H, H-23), 2.92 (m, 1H, H-3), 2.80 (dd, 1H, J = 14.0, 4.5 Hz, H-4), 2.70 (broad s, 1H, OH), 2.61 (dd, 1H, J = 14.0, 7.5 Hz, H-4'), 2.52 (dd, 1H, J = 18, 3.0 Hz, H-11), 2.43 (dd, 1H, J = 18.0, 3.0 Hz, H-11'), 2.23 (m, 1H, H-12), 2.13 (t, 1H, J = 6 Hz, H-8), 1.85 (d, 1H, H-12'), 1.74 (m, 1H, H-10), 1.56 (s, 3H, H-13), 1.54-1.41 (m, 2H, H-20,20'), 1.39-1.23 (m, 2H, H-21,21'), 1.35 (s, 9H, H-17,18,19), 1.11 (s, 3H, H-14), 0.88 (t, J = 7.5 Hz, 3H, H-22), 0.69 (s, 3H, H-15) ppm; ^{13}C NMR (125 MHz, C_6D_6) δ 179.7, 173.2, 170.0, 81.0, 76.6, 65.1, 51.0, 50.6, 38.8, 38.7, 38.3, 35.2, 34.1, 33.5, 28.7, 28.6, 28.0, 27.4, 22.7, 20.7, 14.4 ppm; HRMS (+ESI): Calculated for $C_{23}H_{39}NO_5Na$ $[M + Na]^+$: 432.2720, found 432.2720.

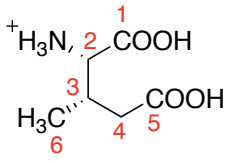


Compound 2.20 White oil.; R_f 0.3 (diethyl ether / petroleum ether: 30: 70); IR: υ max 2954, 1733, 1368, 1151 cm^{-1}; ^1H NMR (500 MHz, C_6D_6) δ 4.34 (d, 1H, J = 5.0 Hz, H-2), 3.36 (s, 3H, H-24), 2.99 (m, 1H, H-3), 2.79 (dd, 1H, J = 14.0, 4.5 Hz, H-4), 2.70 (broad s, 1H, OH), 2.63 (dd, 1H, J = 14.0, 7.5 Hz, H-4'), 2.53 (dd, 1H, J = 18.0, 3.0 Hz, H-11),

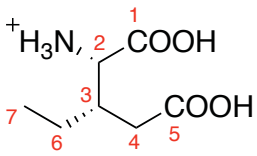
2.46 (dt, 1H, J = 18.0, 3.0 Hz, H-11'), 2.24 (m, 1H, H-12), 2.13 (t, 1H, J = 6 Hz, H-8), 1.85 (d, 1H, H-12'), 1.74 (m, 1H, H-10), 1.61 (m, 1H, H-21), 1.56 (s, 3H, H-13), 1.38 (m, 2H, H-20,20'), 1.35 (s, 9H, H-17,18,19), 1.11 (s, 3H, H-14), 0.97 (d, J = 6.5 Hz, 3H, H-22), 0.91 (d, J = 6.5 Hz, 3H, H-23), (s, 3H, H-15) ppm; ^{13}C NMR (125 MHz, C_6D_6) δ 179.8, 173.2, 170.0, 81.0, 76.6, 64.9, 51.0, 50.5, 41.0, 38.7, 38.3, 36.8, 35.4, 33.5, 28.7, 28.6, 27.4, 25.7, 23.0, 22.9, 22.7 ppm; HRMS (+ESI): Calculated for $C_{24}H_{41}NO_5Na$ $[M + Na]^+$: 446.2877, found 446.2877.



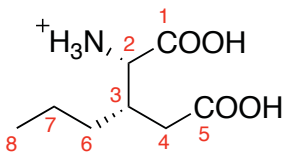
Compound 2.21 White oil.; R_f 0.20 (diethyl ether / petroleum ether: 35: 65); IR: υ max 2919, 1731, 1368, 1148, 699 cm^{-1} ^{1}H NMR (400 MHz, C_6D_6) δ 7.3-7.0 (m, 5H, H-21,22,23,24,25) 4.44 (d, 1H, J = 6.0 Hz, H-2), 4.35 (m, 1H, H-3), 3.20 (s, 3H, H-26), 3.15 (s, 1H, H-4), 3.13 (d, J = 5.5 Hz, 1H, H-4'), 2.64 (broad s, 1H, OH), 2.43 (dd, J = 18.0, 2.8 Hz, 1H, H-11), 2.17 (m, 1H, H-12), 2.09 (t, J = 6 Hz, 1H, H-8), 2.02 (dt, J = 18.0, 2.8 Hz, 1H, H-11'), 1.77 (d, 1H, J = 10.4 Hz, H-12'), 1.62 (m, 1H, H-10), 1.56 (s, 3H, H-13), 1.25 (s, 9H, H-17,18,19), 1.06 (s, 3H, H-14), 0.46 (s, 3H, H-15) ppm; ^{13}C NMR (125 MHz, C_6D_6) δ 180.3, 172.3, 169.1, 141.8, 128.8, 128.6, 127.1, 81.2, 76.7, 67.9, 51.1, 50.5, 45.4, 38.6, 38.1, 36.0, 33.5, 28.7, 28.6, 27.9, 27.3, 22.6 ppm; HRMS (+ESI): Calculated for $C_{26}H_{37}NO_5Na$ $[M + Na]^+$: 466.2564, found 466.2564.



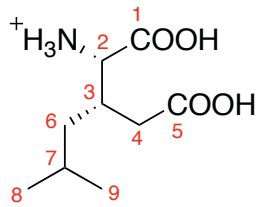
(2*S*,3*S*)-3-Methylglutamic acid (2.4) White solid. m. p. 146 – 149 °C; R_f 0.76 (^{n}butanol / acetic acid / water: 5:3:2); IR: υ max 1698, 1608, 1397, 1194, 1135 cm^{-1}; ^{1}H NMR (400 MHz, D_2O) δ 4.09 (d, 1H, J = 3 Hz, H-2), 2.65 (dd, 1H, J = 15, 6 Hz, H-4a), 2.56 (m, 1H, H-3), 2.48 (dd, 1H, J = 15, 8 Hz, H-4b), 1.02 (d, 3H, J = 6.8 Hz, H-6) ppm; ^{13}C NMR (125 MHz, D_2O) δ 175.5, 170.5, 56.0, 36.8, 30.3, 13.6 ppm -trifluoroacetate peaks at 162.6 (d), 115.9 (q) ppm -acetone reference peaks at 215.9, 30.9 ppm; HRMS (ESI): Calculated for $C_6H_{11}NO_4Na$ $[M + Na]^+$: 184.0586, found 184.0580.



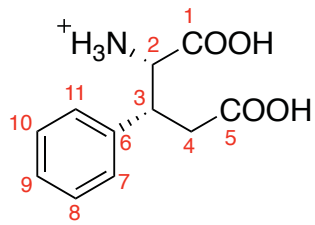
(2*S*,3*S*)-3-Ethylglutamic acid (2.5) White solid; m. p. 131 – 133 °C R_f 0.70 (^{n}butanol / acetic acid / water: 5:3:2); IR: υ max 3159, 1704, 1615, 1198, 1123 cm^{-1}; ^{1}H NMR (500 MHz, D_2O) δ 4.16 (d, J = 3 Hz, 1H, H-2), 2.57 (m, 2H, H-4), 2.41 (m, 1H, H-3), 1.53 (m, 1H, H-6a), 1.37 (m, 1H, H-6b), 0.95 (t, 3H, J = 7.5 Hz, H-7) ppm; ^{13}C NMR (125 MHz, D_2O) δ 175.5, 171.1, 54.8, 36.9, 33.9, 21.8, 10.3 ppm; HRMS (ESI): Calculated for $C_7H_{13}NO_4Na$ $[M + Na]^+$: 198.0742, found 198.0737



(2*S*,3*S*)-3-Propylglutamic acid (2.6) White solid; m. p. 122 – 123 °C R_f 0.74 (^{n}butanol / acetic acid / water: 5:3:2); IR: υ max 3111, 1711, 1618, 1198, 1176 cm^{-1}; ^{1}H NMR (500 MHz, D_2O) δ 4.12 (d, J = 3 Hz, 1H, H-2), 2.62-2.47 (m, 3H, H-3, H-4), 1.53 (m, 1H, H-6), 1.47-1.27 (m, 4H, H-6, H-7), 0.89 (t, 3H, J = 7 Hz, H-3) ppm; ^{13}C NMR (125 MHz, D_2O) δ 176.8, 172.2, 50.1, 35.7, 35.1, 31.2, 19.9, 13.4 ppm; HRMS (ESI): Calculated for $C_8H_{15}NNaO_4$ $[M + Na]^+$: 212.0899, found 212.0893:

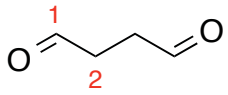


(2*S*,3*S*)-3-Isobutylglutamic acid (2.7) White solid, very hygroscopic; m. p. 132 – 133 °C; R_f 0.70 (^{n}butanol / acetic acid / water: 5:3:2); IR υ max 3134, 1704, 1618, 1177 cm^{-1}; ^{1}H NMR (500 MHz, D_2O) δ 4.16 (d, J = 3 Hz, 1H, H-2), 2.61-2.50 (m, 3H, H-3, H-4), 1.68-1.57 (m, 1H, H-7), 1.27 (m, 2H, H-6), 0.88 (dd, 6H, J = 15, 7 Hz, H-8, H-9) ppm; ^{13}C NMR (125 MHz, D_2O) δ 176.7, 172.0, 56.0, 38.5, 35.1, 33.8, 24.8, 22.4, 21.4 ppm; HRMS (ESI): Calculated for $C_9H_{17}NNaO_4$ $[M + Na]^+$: 226.1055, found 226.1050



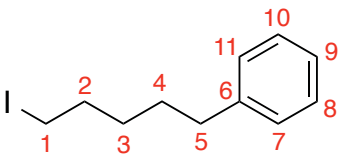
(2*S*,3*R*)-3-phenylglutamic acid (2.8) White solid; m. p. 140 – 145 °C; R_f 0.5 (^{n}butanol / acetic acid / water: 5:3:2); IR: υ max 1712, 1614, 1413, 1200, 1125 cm^{-1}; ^{1}H NMR (500 MHz, D_2O) δ 7.41-7.30 (m, 5H, H7-H11), 4.29 (d, 1H, J = 5 Hz, H-2), 3.74 (m, 1H, H-3), 3.10 (m, 2H, H-4) ppm; ^{13}C NMR (125 MHz, D_2O)

δ 174.9, 170.4, 135.4, 128.9, 128.2, 127.9, 56.7, 41.5, 35.4 ppm; HRMS (+ESI): Calculated for $C_{11}H_{13}NO_4Na$ $[M + Na]^+$: 246.0742, found 246.0737



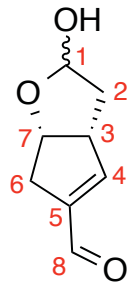
Succinaldehyde (3.4). A solution of 2,5-dimethoxytetrahydrofuran (90 ml, 697 mmol) in water (180 ml) was stirred at 75 °C (oil bath temperature) for 4 h. The temperature was then increased to 120 °C (oil bath temperature) and 150 ml of distillate collected (over a period of 4h). The residue was allowed to cool before being extracted in DCM (20 × 10 ml). The combined organic phases were dried ($MgSO_4$), filtered, and concentrated to give succinaldehyde as a yellow liquid in 68 % crude yield. The product was used immediately in the following step. Spectroscopic properties for this known compound match literature values[108].

Yellow oil. R_f 0.36 (Petroleum ether/EtOAc 1:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.82 (s, 1H, H-1), 2.81 (s, 2H, H-2)



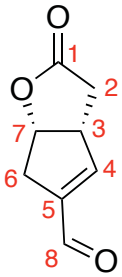
5-Phenylpentyl iodide (3.21). To a solution of 5-phenyl-1-pentanol (3.0 g, 18.3 mmol) in DCM (70 mL) was added triphenylphosphine (9.5 g, 36.6 mmol) and imidazole (3.7 g, 55 mmol), followed by iodine (8.5 g, 36.6 mmol) in three portions over 30 min. The mixture was stirred at r.t for 3 h (until completion), filtered through Celite, then washed twice with 10% sodium thiosulfate (2 × 20 mL). The organic layer was further washed with deionized water (20 mL). The organic layer was dried over anhydrous sodium sulfate and evaporated to near dryness. The crude product was purified by silica gel chromatography using petroleum ether (100 %) to yield the pure product as a clear oil. Spectroscopic properties for this compound match literature values[164].

Colorless oil. R_f = 0.5 ($CHCl_3$); ^{1}***H NMR*** (400 MHz, $CDCl_3$) δ 7.30-7.16 (m, 5H, Ph), 3.19 (t, 2H, J = 7 Hz, H-1), 2.63 (t, 2H, J = 7 Hz, H-5), 1.87 (m, 3H, H-2), 1.65 (m, 2H, H-4), 1.45 (m, 2H, H-3).



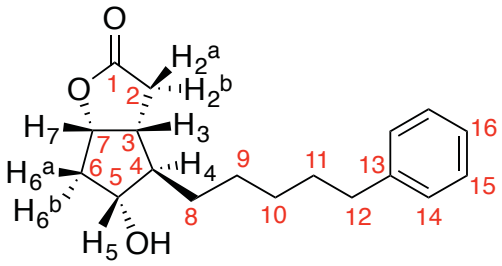
(3a*R*,6a*S*)-2-Hydroxy-3,3a,6,6a-tetrahydro-2H-cyclopenta[b]furan-5-carbaldehyde (3.5). To a solution of freshly prepared succinaldehyde (20 g, 28 mmol) in THF (140 ml) (*S*)-Proline (540 mg, 0.47 mmol) was added as a solid and the reaction stirred at r.t. for 20 h. Additional THF (140 ml) was added, followed by $[Bn_2NH_2][OCOCF_3]$ (2.2 g, 6.6 mmol). The reaction was stirred for a further 14 h. After addition of charcoal (50 g) the volume of the reaction mixture was reduced by half by evaporation under reduced pressure. *tert*-Butyl methyl ether (TBME) (140 ml) was added and the mixture was stirred for 20 min before filtration of the resulting solids (through a sinter funnel). The solids were washed with TBME (3 × 50 ml) and the filtrate was concentrated under reduced pressure. The material was purified by column chromatography (eluting with hexanes/EtOAc, gradient 90:10 to 10:90), then further washed with heptanol at 0°C to provide the pure product (2:1 mixture of diastereoisomers) as a white solid in 11 % yield. Spectroscopic properties for this known compound match literature values.[108]

White solid. R_f 0.16 (Petroleum ether/EtOAc 1:1); ^{1}H NMR (400 MHz, $CDCl_3$) δ (mixture of 2 diastereoisomers, signals of minor indicated by *) 9.80* (s, 1H, H-8), 9.78 (s, 1H, H-8), 6.80* (m, 1H, H-4), 6.66 (m, 1H, H-4), 5.58 (m, 1H, H-1), 5.54* (m, 1H, H-1), 4.97 (m, 1H, H-7), 4.91* (m, 1H, H-7), 3.67 (m, 1H), 3.58* (m, 1H), 2.83-2.68 (m, 3H), 2.83-2.68* (m, 3H), 2.28-2.12 (m, 2H), 2.28-2.12* (m, 2H), 1.97-1.92 (m, 1H), 2.28-2.12* (m, 1H).



(3a*R*,6a*S*)-2-Oxo-3,3a,6,6a-tetrahydro-2H-cyclopenta[b]furan-5-carbaldehyde (3.6). To a solution of lactol **3.5** (2.0 g, 13.0 mmol) in acetonitrile (65 mL) were added tetrakisacetonitrile copper triflate (246 mg, 650 µmol, 20 mol %), 2,2'-bipyridine (102 mg, 650 µmol, 5 mol %), TEMPO (102 mg, 650 mmol, 5 mol %)) and *N*-methylimidazole (104 µL, 1.30 mmol, 10 mol %). The solution was stirred at r.t. for 16 h and then filtered through a silica plug, washed with EtOAc (10 ml × 3). The filtrate was concentrated and the crude material was purified by column chromatography (SiO_2, hexanes/EtOAc (1:1→3:7) to give the lactone as a white solid in 95 % yield. Spectroscopic properties for this known compound match literature values.[108]

White solid. R_f 0.44 (EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.83 (s, 1H, H-8), 6.62 (m, 1H, H-4), 5.23 (t, 1H, J = 6 Hz, H-7), 3.79 (m, 1H, H-3), 3.01-2.84 (m, 3H), 2.58 (dd, 1H).



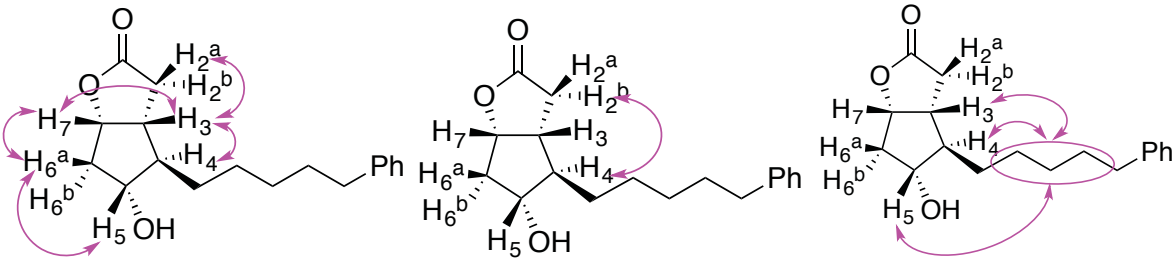
(3a*R*,4*R*,5*R*,6aS)-4-(5-Phenylpentyl)-5-hydroxyhexahydro-2H-cyclopenta[b]furan-2-one (3.19). 5-iodo-1-phenylpentane **3.21** (1 mL, 1.240 mg, 4.5 mmol) was added via syringe to a flame dried 100 ml-Schlenk flask, and further dried under vacuum for 30 min. Dry diethyl ether (10 ml) was added and the solution cooled to –78 °C. A solution of t-BuLi (1.7 M in pentane, 4.50 ml) was added dropwise (the formation of a white precipitate, LiI, indicates lithiation). The reaction mixture was stirred at –78 °C for 1.5 h, then at –40 °C for 1.5 h before being cooled back to –78 °C. To a separate 50-ml flask, CuCN was added as a solid (340 mg, 3.9 mmol) and dried under vacuum for 3 h, then suspended in anhydrous THF (10 ml) and cooled to –78 °C. Meanwhile, a solution of freshly distilled thiophene (300 ml, 315 mg, 3.7 mmol) in anhydrous THF (10.0 mL) was added via syringe to a flame dried Schlenk flask and cooled to –30 °C. To this, *n*-BuLi (2.5 M in hexane, 1.45 ml) was added dropwise and the solution was stirred at –30 °C for 30 min, then cooled to –78 °C and added via cannula to the CuCN suspension. The cooling bath was removed and the suspension was allowed to warm to r.t. The resulting clear, brown solution of the cuprate was added dropwise via cannula to the Schlenk flask containing the alkyl lithium kept at –78 °C. The mixture was stirred at –20 °C for 1 h to allow formation of the mixed cuprate. The reaction mixture was cooled to –78 °C and a solution of enal **3.6** (500 mg, 3.3 mmol) in anhydrous THF (10.0 mL) was added dropwise. The mixture was stirred at – 78 °C for 1 h and then allowed to warm slowly to –20 °C. TMSCl (2 mL, 16.4 mmol) was added via syringe followed by NEt_3 (3 mL, 19.7 mmol) and the reaction was stirred for 30 min at this temperature. The reaction was quenched by the addition of saturated aqueous NH_4Cl solution (80mL) and extracted with Et_2O (3 × 60 mL). The combined organic phases were washed with saturated NH_4Cl solution (40 mL) before being dried over Na_2SO_4, filtered, and concentrated to give the crude material as a blue-green oil, which was used directly in the subsequent transformation within 24 hours.

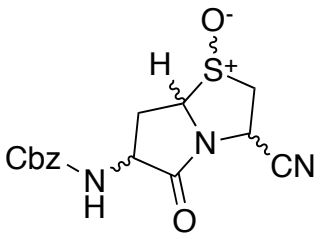
The crude material from the conjugate addition/trapping experiment, was dissolved in DCM/MeOH (3:1, 40 ml total) and cooled to –78 °C. A stream of ozone was passed through the stirred solution. After reaction completion (~15 min, judged by TLC) the flask was purged with a stream of oxygen for 10 min/ $NaBH_4$ (250 mg, 6.6 mmol) was next added slowly as a solid. The reaction mixture was stirred at –78 °C for 1.5 h, then allowed to warm to r.t. and stirred at this temperature for 1 h. Additional $NaBH_4$ (50 mg) was added and the reaction was stirred for another 40 min. The solvent was removed under vaccum and the residue was

redissolved in EtOAc (50 ml). The solution was poured into a saturated aqueous NaCl solution (30 mL) and extracted in EtOAc (3 × 50 mL). The combined organic phases were dried over Na_2SO_4 and concentrated in vacuum. The compound was purified by column chromatography ($CHCl_3$/ethyl acetate: 100:0→70:30 over 25 column volumes) to yield the final alcohol as a colourless oil in 62 % yield.

Colorless oil. R_f 0.2 (chloroform/EtOAc 7:3); IR: υ max 3541, 2928, 2856, 1765, 1240, 1176, 1040, 750 cm^{-1}; ^1H NMR (500 MHz, $CDCl_3$) δ 7.29-7.16 (m, 5H, phenyl), 4.95 (dt, 1H, J = 7 Hz, 2 Hz, H-7), 4.01 (aq, 1H, J = 5 Hz, H-5), 2.80 (dd, 1H, J = 11 Hz, 7 Hz, H-2a), 2.60 (t, 2H, J = 8 Hz, H-12), 2.52 (m, 2H, H-3), 2.50 (m, 2H, H-2b), 2.26 (ddd, 1H, J = 15 Hz, 7 Hz, 5 Hz, H-6a), 2.04 (m, 1H, H-6b), 1.82 (m, 1H, H-4), 1.71 (bs, 1H, OH), 1.62 (m, 2H, H-8), 1.42-1.15 (m, 6H, H-9,10,11) ppm; ^{13}C NMR (125 MHz, $CDCl_3$) δ 178.0, 142.6, 128.5, 128.4, 128.3, 125.7, 84.4, 77.7, 54.4, 43.1, 40.6, 36.3, 36.0, 31.3, 29.4, 27.7 ; HRMS (ESI): Calculated for $C_{18}H_{24}O_3Na$ $[M + Na]^+$: 311.1623, found 311.1614

<u>NOESY CORRELATIONS</u> (no correlation other than these ones is observed):





Synthesis of mixtures of 5.3-5.9. The starting material, **5.1** or **5.2** (25 mg, 0.079 mol), was dissolved in a mixture of 70:30 THF: water (2 ml) before addition of ammonium persulfate (54 mg, 0.24 mol). The reaction mixture was stirred at room temperature for 16 hours before evaporation of the solvent. Water (5 ml) and DCM (5 ml) were added to the residue with mixing. The organic layer was collected, dried over sodium sulfate before evaporation of the solvent under vacuum. The products were purified as a mixture through semi-preparative HPLC on a 25 cm × 9.4 mm ZORBAX Rx-C18 column (Agilent). Elution consisted of a gradient step from 95% mobile phase A (water) in 5% mobile phase B (acetonitrile), to 95% phase B over 45 min, at a flow rate of 2 ml/min. The absorption was recorded at 220 nm.

7.4 Biological studies

7.4.1 Site-directed mutagenesis and DNA primers

All P450 3A4 mutants were generated using the QuikChange XL Site-Directed Mutagenesis Kit (Agilent) and pSE3A4His encoding *N*-terminally truncated and tetrahistidine-tagged CYP3A4[165], referred here as the wild-type. Sequences for all DNA primers used are listed in the Tables 7.1

Table 7-1

Mut.	Mutation	Plasmid*	Primers 5'→3'
Single mutation (vs WT)			
M1	F108W	WT	cttcacaaaccggaggccttggggtccagtgggatttatg
M2	M114F	WT	ccttttggtccagtgggatttttcaaaagtgccatctctatagc
M3	S119D	WT	gggatttatgaaaagtgccatcgatatagctgaggatgaagaatgg
M4	S119Q	WT	gtgggatttatgaaaagtgccatccagatagctgaggatgaagaatggaag
M5	S119T	WT	ggatttatgaaaagtgccatcactatagctgaggatgaagaat
M6	S119Y	WT	ggatttatgaaaagtgccatctatatagctgaggatgaagaatgg
M7	S119F	WT	ggatttatgaaaagtgccatctttatagctgaggatgaagaatgg
M8	I120F	WT	gtgggatttatgaaaagtgccatctctttcgctgaggatgaaga
M9	L210F	WT	ccctttgtggaaaacaccaagaagtttttaagatttgatttttggatc
M10	L211F	WT	ttgtggaaaacaccaagaagcttttcagatttgatttttggatcc
M11	I301F	WT	ctggagctcgtggcccaatcaattttctttatttttgct
M12	I301W	WT	tggagctcgtggcccaatcaatttggtttatttttgctggctatgaaac
M13	I301K	WT	gagctcgtggcccaatcaattaagtttatttttgctggctatgaaa
M14	F304W	WT	gctcgtggcccaatcaattatctttatttgggctggctatgaaacc
M15	A305V	WT	gcccaatcaattatctttatttttgttggctatgaaaccacgag
M16	I369T	WT	gaaacgctcagattattcccaactgctatgagacttgaga
M17	I369F	WT	aacgctcagattattcccatttgctatgagacttgagag
M18	A370V	WT	ctcagattattcccaattgttatgagacttgagagggtc
M19	A370L	WT	gaatgaaacgctcagattattcccaattctaatgagacttgagagggtctg
M20	A370I	WT	aacgctcagattattcccaattattatgagacttgagagggtctg

M21	A370F	WT	aacgctcagattattcccaatttttatgagacttgagagggtctg
M22	G481V	WT	ccctgaaattaagcttaggagtacttcttcaaccagaaaaacc
M23	G481Q	WT	ccccctgaaattaagcttaggacagcttcttcaaccagaaaaacccg
M24	G481W	WT	cccctgaaattaagcttaggatggcttcttcaaccagaaaaaccc
M25	G481L	WT	atccccctgaaattaagcttaggattacttcttcaaccagaaaaacc
M26	L482F	WT	ctgaaattaagcttaggaggatttcttcaaccagaaaaacccg
M27	L482Q	WT	ctgaaattaagcttaggaggacagcttcaaccagaaaaacccgttg
M28	L482R	WT	gaaattaagcttaggaggacgtcttcaaccagaaaaacccg
M29	L483F	WT	gaaattaagcttaggaggacttttttcaaccagaaaaacccgt

Multiple mutations

Mut.	**Mutations**	**Plasmid***	**Primers 5'→3'**
M30	S119T/I369F	M18	same as M5
M31	S119T/L482F	M27	same as M5
M32	S119T/A370I	M21	same as M5
M33	S119Y/A370V	M19	same as M6
M34	S119Y/I369F	M18	same as M6
M35	S119Y/L482F	M27	same as M6
M36	I120F/L482F	M27	same as M8
M37	I120F/G481Q	M24	same as M8
M38	I120F/I301F	M11	same as M8
M39	I120F/I369F	M18	same as M8
M40	L211F/G481W	M25	same as M10
M41	I301F/L482F	M27	same as M11
M42	I301F/I369F	M18	same as M11
M43	I301F/A370I	M21	same as M11
M44	F304W/A370I	M21	same as M14
M45	A305V/L482F	M27	same as M15
M46	I369F/A370I	M18	

forward primer for M46 only **5'→3'**: aacgctcagattattcccatttattatgagacttgagagggtctg

M47	I369F/L482F	M27	same as M18
M48	F304W/I369F	M14	same as M18
M49	A301F/A370V	M11	same as M19
M50	G481W/L482F	M27	same as M25
M51	I120F/I301F/L482F	M44	same as M8
M52	I120F/I369F/F304W	M41	same as M14
M53	I301F/I369F/L482F	M44	same as M18
M54	I301F/F304W/I369F/L482F	M59	
	forward primer for M54 only **5'→3'**: gctcgtggcccaatcaattttctttatttgggctggctatgaaacc		
M55	L211F/I301F/I369F/L482F	M59	same as M10
M56	F108W/I301F/I369F/L482F	M59	same as M1
M57	I301F/I369F/G481W/L482F	M59	
	forward primer for M57 only **5'→3'**: cccctgaaattaagcttaggatggtttcttcaaccagaaaaaccc		
M58	I301F/I369F/A370I/L482F	M59	same as M50
M59	F108W/I301F/I369F/G481W/L482F	M63	same as M1
M60	S119T/I301F/I369F/G481W/L482F	M63	same as M5
M61	L210F/I301F/I369F/G481W/L482F	M63	same as M9
M62	L211F/I301F/I369F/G481W/L482F	M63	same as M10
M63	S119T/L210F/I301F/I369F/G481W/L482F	M67	same as M5
M64	S119T/L211F/I301F/I369F/G481W/L482F	M68	same as M5
M65	F108W/S119T/L210F/ I301F/I369F/G481W/L482F	M69	same as M1

7.4.2 Expression and purification of P450 enzymes

A previously reported protocol was used for the expression and purification of P450 3A4 wild-type and selected mutant P450s[110].

7.4.3 P450 enzyme assay using purified enzyme

The desired substrate (600 μM, 60 nmol) and P450 (10 μM, 3 nmol) were mixed in potassium phosphate buffer (0.1 M, pH = 7.4, 10 % glycerol, total volume 100 μL) and pre-incubated at

37°C for 5 min. Reactions were initiated with the addition of CHP (100 μM, 40 pmol) and allowed to proceed at 37°C with orbital shaking (250 RPM). After 60 min, the reaction was terminated with the addition of DCM (0.3 mL × 3) and extraction. After DCM evaporation, the residue was re-dissolved in CH_3CN (150 μL) for HPLC-UV-MS analysis.

7.4.4 P450 enzyme assay with cell lysate

The pellet obtained from a 750 mL cell culture (~10 g of cells) was resuspended in phosphate buffer (10 ml, 0.1 M potassium phosphate, pH 7.4, 10% glycerol). The following protease inhibitors were added: phenylmethylsulfonyl fluoride (PMSF, 2 mM), leupeptin (1.6 μg/mL), aprotinin (1 μg/mL), bestatin (0.8 μg/mL), and pepstatin A (0.7 μg/ mL), along with lysozyme (10 mg per g of cells). The resulting mixture was stirred slowly for 30 min and then sonicated at 60% cycle duty and power 8 for 4 cycles of 30 s each. The lysate was centrifuged for 1 h at 54 000 × g and the pellet was discarded. The desired substrate (300 μM, 300 nmol) was added to freshly prepared cell lysate and the mixture pre-incubated at 37°C for 5 min. The reaction was initiated by the addition of CHP (5 mM, 5 μmol) and allowed to proceed at 37°C with orbital shaking (250 RPM). After 15 min, additional CHP was added (5 μmol, final total concentration 10 mM) and the reaction was terminated with the addition of dichloromethane (DCM, 0.7 mL × 3). The combined organic extracts were evaporated, and the residue was re-dissolved in CH_3CN (150 μL) for HPLC-UV-MS analysis.

7.4.5 Metabolism studies with human liver microsomes (HLM)

The desired substrate (100 μM) and HLM (1 mg/ml) was mixed in potassium phosphate buffer (0.1 M, pH = 7.4, total volume 100 μL) and pre-incubated at 37°C for 5 minutes. The reaction was initiated by addition of NADPH (500 μM) and the mixture was incubated at 37ºC for 1 hour. At the end of the reaction, substrates and metabolites were extracted in DCM (0.3 ml × 3). The combined organic extracts were evaporated and the residue was redissolved in acetonitrile (100 μL) before LC-MS analysis.

7.4.6 Testosterone assay for HLM activity

Prior to any metabolism assay, the activity of HLM was measured from the conversion of testosterone to 6β-hydroxytestosterone. Each enzymatic reaction contained HLM (1 mg/ml),

testosterone (100 μM) and NADPH (500 μM) in KPi buffer (0.1 M at pH 7.4) in a final volume of 100 μL. The reactions were initiated by the addition of NADPH, and then allowed to proceed for 1 h at 37ºC with shaking at 250 RPM. At the end of the reaction, the tubes were spiked with internal standard cortexolone (5 μL from a 200 μM stock solution in MeOH) and extracted in EtOAc (0.3 mL x 3). The extracts were evaporated and redissolved in acetonitrile (75 μL for the resin) before HPLC analysis.

7.4.7 Measurement of intrinsic clearance Cl_{int}

The test substance (10 μM) was mixed with 0.1 M phosphate buffer at pH 7.4 and HLM (0.5 mg/mL) and incubated at 37°C. NADPH at 2 mM concentration was added to start the reaction. After 0, 3, 7, 12, 20 and 30 min the reaction was stopped by addition of ice-cold acetonitrile, and immediately vortexed for 30 seconds. The resulting solution was store at 4°C for 1 hour to facilitate protein precipitation and later centrifuged for 30 min at 2000 g. The supernatants were then analyzed by HPLC/MS.
For the measurement of Cl_{int}, the UV peak for **5.11**, at the different time points, was integrated and the area values used to calculate % remaining compound. Each reaction was performed in triplicate and the corresponding standard deviation used to calculate experimental errors.

7.4.8 Initial velocities of P450-catalyzed reactions

The desired substrate (500 μM) and purified P450 3A4 variant (5 μM) were mixed in potassium phosphate buffer (0.1 M, pH = 7.4, 10 % glycerol, 2 ml total volume) and pre-incubated at 37°C for 5 minutes. CHP (final concentration 100 μM) was added to initiate the reaction and the same amount was re-added every 15 minutes. At the chosen time-points (1, 2, 4, 6, 8, 10, 15, 20, 25, 30, 40, 50, 60, 70 min) 50 μL aliquots of reaction mixture were collected and quenched by addition of acetonitrile (50 μL). After vortexing for 30 sec, each sample was centrifuged at 14,000 RPM for 30 min, then filtered and analysed by LC-MS. UV peak areas were used to monitor the progression of the reaction.

7.4.9 Equilibration studies for 5.3-5.9

Compounds **5.3**-**5.9** were isolated by collecting the respective peaks separated by HPLC using the method described in section 7.5.5. Upon solvent evaporation, compounds **5.3**-**5.9**,

separately, were incubated in 0.1 M potassium phosphate buffer at pH 7.4 and 37ºC for 1 min, 10 min, 1 hour, 24 hours, or 3 weeks. At the end of the incubation period, compounds **5.3-5.9** were extracted in DCM (0.3 ml × 3). The combined organic extracts were evaporated and the residue was redissolved in acetonitrile (30 μl) before HPLC-MS analysis.

7.4.10 Deuterium exchange studies for 5.8 and 5.9

Compounds **5.8** and **5.9** were isolated by collecting the respective peaks separated by HPLC using the method described in section 7.5.5. Once isolated, the compounds were incubated in 0.1 M phosphate buffer, which was prepared with deuterium oxide and adjusted to pH 7.4. After 1 hour incubation, the compounds were isolated and analyzed by HPLC-MS.

7.4.11 Electrophysiology assays for GluK2 activation

The studies described in this paragraph were carried out by Brent Dawe, as part of a collaboration with Derek Bowie's research group at McGill University.

HEK293T cells were co-transfected with cDNA encoding GluK2 (Q/R unedited) KAR subunits and enhanced green fluorescent protein ($eGFP_{S65T}$) for 4 - 8 hrs using the calcium phosphate precipitation method. This model for KAR studies has been previously reported and validated[190]. Agonist solutions were prepared by dissolving the agonist in external solution and correcting the pH to 7.4. All experiments were performed on excised membrane patches in the outside-out configuration. Thin-walled borosilicate glass pipettes (3-5 MΩ, King Precision Glass, Inc.) were coated with dental wax to serve as recording pipettes. Agonist-containing external solutions were applied to outside-out patches for 250 ms at -60 mV (unless otherwise stated) using a piezo-stack driven perfusion system. Sufficient time between applications of L-glutamate was allowed for complete recovery from macroscopic desensitization. Solution exchange time was determined routinely after experiments by measuring the liquid junction current (10-90 % rise-time = 100-400 µs). Series resistances (3-15 MΩ) were routinely compensated by 95%. Current records were filtered at 10 kHz and digitized at 25-100 kHz. The reference electrode was connected to the bath via an agar bridge of 3M KCl.

7.5 LC-MS methods

7.5.1 Oxidation of 3.19, latanoprost precursor

The column used for was an analytical 4.6 × 250 mm, 4 μm SYNERGI 4μ Hydro-RP 80 A (Phenomenex, Torrance, CA). Elution consisted of a gradient step from 99% mobile phase A (H_2O + 0.05 % formic acid) and 1% mobile phase B (CH_3CN + 0.05 % formic acid), to 99 % phase B over 20 min, at a flow rate of 1 ml/min. The absorption was recorded at 220 nm. Compounds **3.24**, **3.26**, **3.27**, **3.19** eluted with retention times of 12.9 min, 13.1 min, 14.6 min, 17.7 min, respectively. MS analysis included total ion current (TIC) and single-ion-monitoring (SIM) at m/z 287, 289, 303, 305.

7.5.2 Oxidation of 4.1, deoxy-precursor of (*R*)-lisofylline

For non-chiral separation, the column used was an analytical 4.6 × 250 mm, 4 μm SYNERGI 4μ Hydro-RP 80 A (Phenomenex, Torrance, CA). Elution consisted of a gradient step from 50% mobile phase A (H_2O) and 50% mobile phase B (CH_3CN), to 95 % phase B over 20 min, at a flow rate of 0.5 ml/min. The absorption was recorded at 273 nm. The racemic mixture **4.5**+**4.6** and the mixture **4.2**+**4.3** eluted with retention times of 14.6 and 14.9 min, respectively. MS analysis included total ion current (TIC) and single-ion-monitoring (SIM) for the hydroxylation products (m/z 281).

For chiral separation after cell-lysate oxidation assay, samples were pre-purified by analytical HPLC using the same conditions as for non-chiral analysis. Chiral analysis was achieved with an analytical Lux® 100 × 4.6 mm, 5 μm Amylose-1A (Phenomenex, Torrance, CA). Elution consisted of a gradient step from 80% mobile phase A (hexanes) and 20% mobile phase B (isopropanol), to 75 % phase B over 25 min, at a flow rate of 1 ml/min. The absorption was recorded at 273 nm. Compounds **4.5**, **4.2**, **4.3** and **4.4** eluted at retention times of, respectively, 16.4 min, 17.5 min, 19 min, 20.3 min.

7.5.3 Oxidation of testosterone for HLM activity assay

The column used was an analytical 4.6 × 250 mm, 4 μm SYNERGI 4μ Hydro-RP 80 A (Phenomenex, Torrance, CA). Separation and quantification of the product (relative to the internal standard cortexolone) was achieved using mobile phases A (Milli-Q water) and B

(acetonitrile) at a flow rate of 0.5 mL/min with the UV detector set to monitor at 244 nm. Elution consisted of an initial isocratic step at 50% phase B for 5 min, followed by an increase to 70% phase B over the next 7 min and finally to 95% phase B over the final 8 min. The retention times were 7.1 min for the product 6β-hydroxytestosterone, 10.1 min for the internal standard cortexolone and 15.8 min for the substrate testosterone.

7.5.4 Oxidation of progesterone

The column used was an analytical 4.6 × 250 mm, 4 μm SYNERGI 4μ Hydro-RP 80 A (Phenomenex, Torrance, CA). Elution consisted of a gradient step from 99% mobile phase A (H_2O) and 1% mobile phase B (MeOH), to 99 % phase B over 50 min, at a flow rate of 0.5 ml/min. The absorption was recorded at 256 nm. 6β-hydroxyprogesterone and 2β-hydroxyprogesterone eluted with retention times of, respectively, 47.5 min and 48.2 min. MS analysis included total ion current (TIC) and single-ion-monitoring (SIM) for the hydroxylation products (m/z 331).

7.5.5 Metabolism of 5.1 and 5.2

The column used was a 150 × 4.6 mm, 5 μM, Eclipse XDB-C8 (Agilent). Elution of metabolites **5.3-5.9** consisted of a gradient step from 95% mobile phase A (water + 0.05% formic acid) and 5 % mobile phase B (acetonitrile + 0.05% formic acid), to 95% phase B over 20 min, at a flow rate of 1 ml/min. The absorption was recorded at 220 nm. MS analysis included total ion current (TIC) and single-ion-monitoring (SIM) for m/z 334.

7.5.6 Deuterium exchange studies for 5.8 and 5.9

HPLC separation of **5.3**-**5.9** was achieved using the same method described in section 7.5.5. MS analysis included total ion current (TIC) and single-ion-monitoring (SIM) for m/z 334 (no deuterium), m/z 335 (+ 1 D), m/z 336 (+ 2 D).

7.5.7 Metabolism of 5.11

Separation was achieved using a Zorbax Eclipse XDB-C18 150 mm x 4.6 mm, 5 μm. Elution consisted of a gradient step from 99% mobile phase A (H_2O) and 1% mobile phase B (CH_3CN), to 99 % phase B over 20 min, at a flow rate of 1 mL/min. The absorption was recorded at 220 nm. Compound **5.11** eluted at 11.9 min and a mixture of unidentified monooxidation products eluted between 9 min and 12.5 min.

For LC-MS metabolite identification, the following ions were monitored: m/z 366 for $[M+O+H]^+$ and m/z 404 $[M+O+K]^+$ (mono-oxodation), m/z = 382 for $[M+2O+H]^+$ and m/z = 420 $[M+2O+K]^+$ (double-oxidation), m/z = 235 for $[M-Bz+H]^+$ and m/z =273 $[M-Bz+K]^+$ (debenzylation).

7.5.8 Protein LC-MS

After removal of phosphate salts through dialysis in water (membrane cut-off 30000 Da), P450 samples were run through a Poros R2/10 column from Applied Biosystems (10 μm, 2000 Å, 2.1 × 100 mm) using a gradient of 90% mobile phase A (0.1% formic acid in water) and 10 % mobile phase B (0.1% formic acid in acetonitrile) to 0 % mobile phase A and 100% mobile phase B in 15 minutes.

7.6 Compututational studies

7.6.1 Flexible docking for KAR

The X-ray structures of GluK1 and GluK2 complexes were retrieved from the Protein Data Bank (codes: 1S7Y for the GluK2-glutamic acid complex; 1TT1 for the GluK2-kainic acid complex; 1YAE for the GluK2-domoic acid complex; 1TXF for the GluK1-glutamic acid complex; 2F35 for the GluK1-UBP302 complex; 2QS2 for the GluK1-UBP318 complex). Hydrogen atoms were added with their position optimized through energy minimization. Protein superimposition was achieved using MATCH-UP, a module of FITTED that aligns the α-carbons of the residues found with at least one atom within 10 Å from the ligand. All protein structures for the docking were prepared using PROCESS (a module of FITTED), and the ligands were fully ionized and prepared with SMART (a module of FITTED). The β-substituted glutamic acid derivatives (Me-Glu, Et-Glu, Pr-Glu, iBu-Glu, Ph-Glu) were docked using three protein structures as input files (1S7Y, 1TT1, 1YAE). As mentioned in the main manuscript,

comparison of the crystal structures reveals three distinct protein conformations that we will refer to as closed, intermediate, and open, based on the degree of the cleft closure formed by GluK2 ABD half domains. Additional details and validation of FITTED on GluK2 KARs have been previously reported[101].

7.6.2 Prediction of transition states for mutant P450s

A combination of the IMPACTS software for prediction of metabolism and a second software for single-point mutations was used. Briefly, the protein structure (PDB 3NXU) was prepared by removing the native ligand, followed by optimization of the conformational states of the mutated residues. The desired substrate was docked forcing oxidation at desired sites of reaction. The regio and stereochemistry of the product was taken into account by discriminating between the different C-H bonds. For each reaction, multiple transition states were generated and the best ones were chosen based on calculated energies. The fully-automated protocol is accessible through our online platform FORECASTER, under the workflow 'biocatalysis with P450s'.

REFERENCES

1. Lawandi, J.; Toumieux, S.; Seyer, V.; Campbell, P.; Thielges, S.; Juillerat-Jeanneret, L.; Moitessier, N., Constrained peptidomimetics reveal detailed geometric requirements of covalent prolyl oligopeptidase inhibitors. *J Med Chem* **2009,** *52* (21), 6672-84.
2. Prevost, S.; Thai, K.; Schutzenmeister, N.; Coulthard, G.; Erb, W.; Aggarwal, V. K., Synthesis of prostaglandin analogues, latanoprost and bimatoprost, using organocatalysis via a key bicyclic enal intermediate. *Org Lett* **2015,** *17* (3), 504-7.
3. Whitford, D., *Proteins: structure and function*. John Wiley & Sons: 2005.
4. Mittenthal, G. C.-A. s. M. W. D. C.-A. s. J. E., The origin, evolution and structure of the protein world. *Biochem J* **2009,** (417), 621-637.
5. Rask-Andersen, M.; Almen, M. S.; Schioth, H. B., Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov* **2011,** *10* (8), 579-90.
6. Mullard, A., 2015 FDA drug approvals. *Nat Rev Drug Discov* **2016,** *15* (2), 73-6.
7. Kinch, M. S., An overview of FDA-approved biologics medicines. *Drug discovery today* **2015,** *20* (4), 393-8.
8. Li, S.; Yang, X.; Yang, S.; Zhu, M.; Wang, X., Technology prospecting on enzymes: application, marketing and engineering. *Comput Struct Biotechnol J* **2012,** *2*, e201209017.
9. Hermann, B. G.; Patel, M., Today's and tomorrow's bio-based bulk chemicals from white biotechnology: a techno-economic analysis. *Appl Biochem Biotechnol* **2007,** *136* (3), 361-88.
10. Burley, C. Z. S. D. D. S. G. M. V. S. K., RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules. *J. Chem Educ* **2016,** (93), 569-575.
11. Blencowe, Q. P. O. S. L. J. L. B. J. F. B. J., Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* **2008,** *40*, 1413-1415.
12. Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabuddhe, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D.; Patil, A. H.; Nanjappa, V.; Radhakrishnan, A.; Prasad, S.; Subbannayya, T.; Raju, R.; Kumar, M.; Sreenivasamurthy, S. K.; Marimuthu, A.; Sathe, G. J.; Chavan, S.; Datta, K. K.; Subbannayya, Y.; Sahu, A.; Yelamanchi, S. D.; Jayaram, S.; Rajagopalan, P.; Sharma, J.; Murthy, K. R.; Syed, N.; Goel, R.; Khan, A. A.; Ahmad, S.; Dey, G.; Mudgal, K.; Chatterjee, A.; Huang, T. C.; Zhong, J.; Wu, X.; Shaw, P. G.; Freed, D.; Zahari, M. S.; Mukherjee, K. K.; Shankar, S.; Mahadevan, A.; Lam, H.; Mitchell, C. J.; Shankar, S. K.; Satishchandra, P.; Schroeder, J. T.; Sirdeshmukh, R.; Maitra, A.; Leach, S. D.; Drake, C. G.; Halushka, M. K.; Prasad, T. S.; Hruban, R. H.; Kerr, C. L.; Bader, G. D.; Iacobuzio-Donahue, C. A.; Gowda, H.; Pandey, A., A draft map of the human proteome. *Nature* **2014,** *509* (7502), 575-81.
13. Fishman, M. C.; Porter, J. A., Pharmaceuticals: a new grammar for drug discovery. *Nature* **2005,** *437* (7058), 491-3.
14. Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S., Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* **2011,** *10* (3), 188-95.

15. Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N., Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput Struct Biotechnol J* **2013,** *5*, e201302011.
16. Obach, R. S., Pharmacologically active drug metabolites: impact on drug discovery and pharmacotherapy. *Pharmacol Rev* **2013,** *65* (2), 578-640.
17. Raemsch, K. D.; Sommer, J., Pharmacokinetics and metabolism of nifedipine. *Hypertension* **1983,** *5* (4 Pt 2), II18-24.
18. Kirchmair, J.; Goller, A. H.; Lang, D.; Kunze, J.; Testa, B.; Wilson, I. D.; Glen, R. C.; Schneider, G., Predicting drug metabolism: experiment and/or computation? *Nat Rev Drug Discov* **2015,** *14* (6), 387-404.
19. Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A., An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov* **2015,** *14* (7), 475-86.
20. Onakpoya, I. J.; Heneghan, C. J.; Aronson, J. K., Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Med* **2016,** *14*, 10.
21. Attia, S. M., Deleterious effects of reactive metabolites. *Oxid Med Cell Longev* **2010,** *3* (4), 238-53.
22. Kalgutkar, A. S.; Dalvie, D. K.; Aubrecht, J.; Smith, E. B.; Coffing, S. L.; Cheung, J. R.; Vage, C.; Lame, M. E.; Chiang, P.; McClure, K. F.; Maurer, T. S.; Coelho, R. V., Jr.; Soliman, V. F.; Schildknegt, K., Genotoxicity of 2-(3-chlorobenzyloxy)-6-(piperazinyl)pyrazine, a novel 5-hydroxytryptamine2c receptor agonist for the treatment of obesity: role of metabolic activation. *Drug Metab Dispos* **2007,** *35* (6), 848-58.
23. Andrade, C. H.; Silva, D. C.; Braga, R. C., In silico prediction of drug metabolism by P450. *Curr Drug Metab* **2014,** *15* (5), 514-25.
24. Anastas, P.; Eghbali, N., Green chemistry: principles and practice. *Chem Soc Rev* **2010,** *39* (1), 301-12.
25. Wohlgemuth, R., Biocatalysis--key to sustainable industrial chemistry. *Curr Opin Biotechnol* **2010,** *21* (6), 713-24.
26. Meyer, H.-P.; Eichhorn, E.; Hanlon, S.; Lütz, S.; Schürmann, M.; Wohlgemuth, R.; Coppolecchia, R., The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC). *Catal Sci Technol* **2013,** *3* (1), 29-40.
27. Woodley, J. M., New opportunities for biocatalysis: making pharmaceutical processes greener. *Trends Biotechnol* **2008,** *26* (6), 321-7.
28. Straathof, A. J.; Panke, S.; Schmid, A., The production of fine chemicals by biotransformations. *Curr Opin Biotechnol* **2002,** *13* (6), 548-56.
29. Hollmann, F.; Arends, I. W. C. E.; Holtmann, D., Enzymatic reductions for the chemist. *Green Chem* **2011,** *13* (9), 2285.
30. Mihovilovic, M. D., Enzyme Mediated Baeyer-Villiger Oxidations. *Curr Org Che* **2006,** *10*, 1265-1287.
31. Julsing, M. K.; Cornelissen, S.; Buhler, B.; Schmid, A., Heme-iron oxygenases: powerful industrial biocatalysts? *Curr Opin Chem Biol* **2008,** *12* (2), 177-86.
32. O'Reilly, E.; Kohler, V.; Flitsch, S. L.; Turner, N. J., Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chem Commun (Camb)* **2011,** *47* (9), 2490-501.
33. Maurer, K. H., Detergent proteases. *Curr Opin Biotechnol* **2004,** *15* (4), 330-4.
34. Sochaj, A. M.; Swiderska, K. W.; Otlewski, J., Current methods for the synthesis of homogeneous antibody-drug conjugates. *Biotechnol Adv* **2015,** *33* (6 Pt 1), 775-84.
35. Kalia, J.; Raines, R. T., Advances in Bioconjugation. *Curr Org Chem* **2010,** *14* (2), 138-147.

36. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat Rev Genet* **2015,** *16* (7), 379-94.
37. Moore, J. C.; Arnold, F. H., Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nat Biotechnol* **1996,** *14* (4), 458-67.
38. Siddiqui, K. S.; Cavicchioli, R., Improved thermal stability and activity in the cold-adapted lipase B from Candida antarctica following chemical modification with oxidized polysaccharides. *Extremophiles* **2005,** *9* (6), 471-6.
39. Diaz, J. E.; Lin, C. S.; Kunishiro, K.; Feld, B. K.; Avrantinis, S. K.; Bronson, J.; Greaves, J.; Saven, J. G.; Weiss, G. A., Computational design and selections for an engineered, thermostable terpene synthase. *Protein Sci* **2011,** *20* (9), 1597-606.
40. Song, J. K.; Rhee, J. S., Enhancement of stability and activity of phospholipase A(1) in organic solvents by directed evolution. *Biochim Biophys Acta* **2001,** *1547* (2), 370-8.
41. Bessler, C.; Schmitt, J.; Maurer, K. H.; Schmid, R. D., Directed evolution of a bacterial alpha-amylase: toward enhanced pH-performance and higher specific activity. *Protein Sci* **2003,** *12* (10), 2141-9.
42. Zhang, N.; Stewart, B. G.; Moore, J. C.; Greasham, R. L.; Robinson, D. K.; Buckland, B. C.; Lee, C., Directed evolution of toluene dioxygenase from Pseudomonas putida for improved selectivity toward cis-indandiol during indene bioconversion. *Metab Eng* **2000,** *2* (4), 339-48.
43. Sacchi, S.; Rosini, E.; Molla, G.; Pilone, M. S.; Pollegioni, L., Modulating D-amino acid oxidase substrate specificity: production of an enzyme for analytical determination of all D-amino acids by directed evolution. *Protein Eng Des Sel* **2004,** *17* (6), 517-25.
44. Bolt, A.; Berry, A.; Nelson, A., Directed evolution of aldolases for exploitation in synthetic organic chemistry. *Arch Biochem Biophys* **2008,** *474* (2), 318-30.
45. Cheriyan, M.; Walters, M. J.; Kang, B. D.; Anzaldi, L. L.; Toone, E. J.; Fierke, C. A., Directed evolution of a pyruvate aldolase to recognize a long chain acyl substrate. *Bioorg Med Chem* **2011,** *19* (21), 6447-53.
46. Lin, Y. S., Using a strategy based on the concept of convergent evolution to identify residue substitutions responsible for thermal adaptation. *Proteins* **2008,** *73* (1), 53-62.
47. Vallin, M.; Syren, P. O.; Hult, K., Mutant lipase-catalyzed kinetic resolution of bulky phenyl alkyl sec-alcohols: a thermodynamic analysis of enantioselectivity. *Chembiochem* **2010,** *11* (3), 411-6.
48. Castro, G. R.; Knubovets, T., Homogeneous biocatalysis in organic solvents and water-organic mixtures. *Crit Rev Biotechnol* **2003,** *23* (3), 195-231.
49. Lazar, G. A.; Marshall, S. A.; Plecs, J. J.; Mayo, S. L.; Desjarlais, J. R., Designing proteins for therapeutic applications. *Curr Opin Struct Biol* **2003,** *13* (4), 513-8.
50. Frankle, W. G.; Lerma, J.; Laruelle, M., The synaptic hypothesis of schizophrenia. *Neuron* **2003,** *39* (2), 205-16.
51. Li, J. M.; Zeng, Y. J.; Peng, F.; Li, L.; Yang, T. H.; Hong, Z.; Lei, D.; Chen, Z.; Zhou, D., Aberrant glutamate receptor 5 expression in temporal lobe epilepsy lesions. *Brain Res* **2010,** *1311*, 166-74.
52. Smolders, I.; Bortolotto, Z. A.; Clarke, V. R.; Warre, R.; Khan, G. M.; O'Neill, M. J.; Ornstein, P. L.; Bleakman, D.; Ogden, A.; Weiss, B.; Stables, J. P.; Ho, K. H.; Ebinger, G.; Collingridge, G. L.; Lodge, D.; Michotte, Y., Antagonists of GLU(K5)-containing kainate receptors prevent pilocarpine-induced limbic seizures. *Nat neurosci* **2002,** *5* (8), 796-804.
53. Lerma, J.; Marques, J. M., Kainate receptors in health and disease. *Neuron* **2013,** *80* (2), 292-311.
54. Larsen, A. M.; Bunch, L., Medicinal chemistry of competitive kainate receptor antagonists. *ACS chemical neuroscience* **2011,** *2* (2), 60-74.
55. Dawe, G. B.; Aurousseau, M. R.; Daniels, B. A.; Bowie, D., Retour aux sources: defining the structural basis of glutamate receptor activation. *J physiol* **2014,** *593* (1), 97-110.

56. Armstrong. N.; Gouaux, E., Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor crystal structures of the Gl. *Neuron* **2000,** *28*, 165-181.
57. Kazi, R.; Dai, J.; Sweeney, C.; Zhou, H. X.; Wollmuth, L. P., Mechanical coupling maintains the fidelity of NMDA receptor-mediated currents. *Nat neurosci* **2014,** *17* (7), 914-22.
58. Zhang, W.; St-Gelais, F.; Grabner, C. P.; Trinidad, J. C.; Sumioka, A.; Morimoto-Tomita, M.; Kim, K. S.; Straub, C.; Burlingame, A. L.; Howe, J. R.; Tomita, S., A transmembrane accessory subunit that modulates kainate-type glutamate receptors. *Neuron* **2009,** *61* (3), 385-96.
59. Dawe, G. B.; Musgaard, M.; Andrews, E. D.; Daniels, B. A.; Aurousseau, M. R.; Biggin, P. C.; Bowie, D., Defining the structural relationship between kainate-receptor deactivation and desensitization. *Nat Struct Mol Biol* **2013,** *20* (9), 1054-61.
60. Teague, S. J., Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* **2003,** *2* (7), 527-41.
61. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007,** *450* (7172), 964-72.
62. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* (9), 1157-74.
63. Brooks, B. R. B., R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M., CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* **1983,** *4*, 187-217.
64. Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F., The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* **2005,** *26* (16), 1719-51.
65. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biol* **2011,** *9*, 71.
66. Nelson, D. R.; Zeldin, D. C.; Hoffman, S. M.; Maltais, L. J.; Wain, H. M.; Nebert, D. W., Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics* **2004,** *14* (1), 1-18.
67. Omura, T.; Sato, R., The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence for Its Hemoprotein Nature. *J Biol Chem* **1964,** *239*, 2370-8.
68. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I., Structure and chemistry of cytochrome P450. *Chem Rev* **2005,** *105* (6), 2253-77.
69. Guengerich, F. P., Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* **2001,** *14* (6), 611-50.
70. Isin, E. M.; Guengerich, F. P., Complex reactions catalyzed by cytochrome P450 enzymes. *Biochim Biophys Acta* **2007,** *1770* (3), 314-29.
71. Gillam, E. M., Exploring the potential of xenobiotic-metabolising enzymes as biocatalysts: evolving designer catalysts from polyfunctional cytochrome P450 enzymes. *Clin Exp Pharmacol Physiol* **2005,** *32* (3), 147-52.
72. Lewis, D. F.; Hlavica, P., Interactions between redox partners in various cytochrome P450 systems: functional and structural aspects. *Biochim Biophys Acta* **2000,** *1460* (2-3), 353-74.
73. Katagiri, M.; Ganguli, B. N.; Gunsalus, I. C., A soluble cytochrome P-450 functional in methylene hydroxylation. *J Biol Chem* **1968,** *243* (12), 3543-6.
74. Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G., The catalytic pathway of cytochrome p450cam at atomic resolution. *Science* **2000,** *287* (5458), 1615-22.

75. Ortiz de Montellano, P. R.; De Voss, J. J., Oxidizing species in the mechanism of cytochrome P450. *Nat Prod Rep* **2002,** *19* (4), 477-93.
76. Kadkhodayan, S.; Coulter, E. D.; Maryniak, D. M.; Bryson, T. A.; Dawson, J. H., Uncoupling oxygen transfer and electron transfer in the oxygenation of camphor analogues by cytochrome P450-CAM. Direct observation of an intermolecular isotope effect for substrate C-H activation. *J Biol Chem* **1995,** *270* (47), 28042-8.
77. Vidal-Limon, A.; Aguila, S.; Ayala, M.; Batista, C. V.; Vazquez-Duhalt, R., Peroxidase activity stabilization of cytochrome P450(BM3) by rational analysis of intramolecular electron transfer. *J Inorg Biochem* **2013,** *122*, 18-26.
78. Zhao, H.; van der Donk, W. A., Regeneration of cofactors for use in biocatalysis. *Curr Opin Biotechnol* **2003,** *14* (6), 583-9.
79. Barbosa, O.; Ortiz, C.; Berenguer-Murcia, A.; Torres, R.; Rodrigues, R. C.; Fernandez-Lafuente, R., Strategies for the one-step immobilization-purification of enzymes as industrial biocatalysts. *Biotechnol Adv* **2015,** *33* (5), 435-56.
80. Yoshida, Y., The 50th anniversary and new horizons of cytochrome P450 research: expanding knowledge on the multiplicity and versatility of P450 and its industrial applications. Foreword. *Biol Pharm Bull* **2012,** *35* (6), 811.
81. Sonomoto, K.; Hoq, M. M.; Tanaka, A.; Fukui, S., 11beta-Hydroxylation of Cortexolone (Reichstein Compound S) to Hydrocortisone by Curvularia lunata Entrapped in Photo-Cross-Linked Resin Gels. *Appl Environ Microbiol* **1983,** *45* (2), 436-43.
82. Watanabe, I.; Nara, F.; Serizawa, N., Cloning, characterization and expression of the gene encoding cytochrome P-450sca-2 from Streptomyces carbophilus involved in production of pravastatin, a specific HMG-CoA reductase inhibitor. *Gene* **1995,** *163* (1), 81-5.
83. Urlacher, V. B.; Lutz-Wahl, S.; Schmid, R. D., Microbial P450 enzymes in biotechnology. *Appl Microbiol Biotechnol* **2004,** *64* (3), 317-25.
84. Rittle, J.; Green, M. T., Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science* **2010,** *330* (6006), 933-7.
85. Zanger, U. M.; Schwab, M., Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* **2013,** *138* (1), 103-41.
86. Guengerich, F. P., Cytochrome P-450 3A4: regulation and role in drug metabolism. *Annu Rev Pharmacol Toxicol* **1999,** *39*, 1-17.
87. Peterson, J. A.; Graham, S. E., A close family resemblance: the importance of structure in understanding cytochromes P450. *Structure* **1998,** *6* (9), 1079-85.
88. Denisov, I. G.; Shih, A. Y.; Sligar, S. G., Structural differences between soluble and membrane bound cytochrome P450s. *J Inorg Biochem* **2012,** *108*, 150-8.
89. Zhao, Y.; White, M. A.; Muralidhara, B. K.; Sun, L.; Halpert, J. R.; Stout, C. D., Structure of microsomal cytochrome P450 2B4 complexed with the antifungal drug bifonazole: insight into P450 conformational plasticity and membrane interaction. *J Biol Chem* **2006,** *281* (9), 5973-81.
90. Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H., Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004,** *305* (5684), 683-6.
91. Park, H.; Lee, S.; Suh, J., Structural and dynamical basis of broad substrate specificity, catalytic mechanism, and inhibition of cytochrome P450 3A4. *J Am Chem Soc* **2005,** *127* (39), 13634-42.
92. Khan, K. K.; He, Y. Q.; Domanski, T. L.; Halpert, J. R., Midazolam oxidation by cytochrome P450 3A4 and active-site mutants: an evaluation of multiple binding sites and of the metabolic pathway that leads to enzyme inactivation. *Mol Pharmacol* **2002,** *61* (3), 495-506.

93. Harlow, G. R.; Halpert, J. R., Alanine-scanning mutagenesis of a putative substrate recognition site in human cytochrome P450 3A4. Role of residues 210 and 211 in flavonoid activation and substrate specificity. *J Biol Chem* **1997,** *272* (9), 5396-402.
94. Ekroos, M.; Sjogren, T., Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc Natl Acad Sci U S A* **2006,** *103* (37), 13682-7.
95. Kumar, S.; Liu, H.; Halpert, J. R., Engineering of cytochrome P450 3A4 for enhanced peroxide-mediated substrate oxidation using directed evolution and site-directed mutagenesis. *Drug Metab Dispos* **2006,** *34* (12), 1958-65.
96. He, Y. A.; He, Y. Q.; Szklarz, G. D.; Halpert, J. R., Identification of three key residues in substrate recognition site 5 of human cytochrome P450 3A4 by cassette and site-directed mutagenesis. *Biochemistry* **1997,** *36* (29), 8831-9.
97. He, Y. A.; Roussel, F.; Halpert, J. R., Analysis of homotropic and heterotropic cooperativity of diazepam oxidation by CYP3A4 using site-directed mutagenesis and kinetic modeling. *Arch Biochem Biophys* **2003,** *409* (1), 92-101.
98. Fowler, S. M.; Taylor, J. M.; Friedberg, T.; Wolf, C. R.; Riley, R. J., CYP3A4 active site volume modification by mutagenesis of leucine 211. *Drug Metab Dispos* **2002,** *30* (4), 452-6.
99. Bowie, D., Ionotropic Glutamate Receptors & CNS Disorders. *CNS Neurol Disord Drug Targets* **2008,** *7(2)*, 129-143.
100. Mayer, M. L.; Ghosal, A.; Dolman, N. P.; Jane, D. E., Crystal structures of the kainate receptor GluR5 ligand binding core dimer with novel GluR5-selective antagonists. *J Neurosci* **2006,** *26* (11), 2852-61.
101. Fay, A. M.; Corbeil, C. R.; Brown, P.; Moitessier, N.; Bowie, D., Functional characterization and in silico docking of full and partial GluK2 kainate receptor agonists. *Mol Pharmacol* **2009,** *75* (5), 1096-107.
102. Wehbe, J.; Rolland, V.; Roumestant, M. L.; Martinez, J., Glutamate transporter blockers: enantiomerically pure (2S,3S)- and (2S,3R)-3-methyl glutamic acids. *Tetrahedron: Asymmetry* **2003,** *14* (9), 1123-1126.
103. Alushin, G. M.; Jane, D.; Mayer, M. L., Binding site and ligand flexibility revealed by high resolution crystal structures of GluK1 competitive antagonists. *Neuropharmacol* **2011,** *60* (1), 126-34.
104. Frydenvang, K.; Lash, L. L.; Naur, P.; Postila, P. A.; Pickering, D. S.; Smith, C. M.; Gajhede, M.; Sasaki, M.; Sakai, R.; Pentikainen, O. T.; Swanson, G. T.; Kastrup, J. S., Full domain closure of the ligand-binding core of the ionotropic glutamate receptor iGluR5 induced by the high affinity agonist dysiherbaine and the functional antagonist 8,9-dideoxyneodysiherbaine. *J Biol Chem* **2009,** *284* (21), 14219-29.
105. Venskutonyte, R.; Frydenvang, K.; Hald, H.; Rabassa, A. C.; Gajhede, M.; Ahring, P. K.; Kastrup, J. S., Kainate induces various domain closures in AMPA and kainate receptors. *Neurochem Int* **2012,** *61* (4), 536-45.
106. Patel, S. S.; Spencer, C. M., Latanoprost. A review of its pharmacological properties, clinical efficacy and tolerability in the management of primary open-angle glaucoma and ocular hypertension. *Drugs Aging* **1996,** *9* (5), 363-78.
107. Corey, E. J. W., N. M.; Schaaf, T. K.; Huber, W., Stereo-controlled synthesis of prostaglandins F-2a and E-2 (dl). *J Am Chem Soc* **1969,** *91*, 5675.
108. Coulthard, G.; Erb, W.; Aggarwal, V. K., Stereocontrolled organocatalytic synthesis of prostaglandin PGF2alpha in seven steps. *Nature* **2012,** *489* (7415), 278-81.
109. Aggarwal, V. K. C., M. G., Erb, W. 2013.
110. Chefson, A.; Zhao, J.; Auclair, K., Replacement of natural cofactors by selected hydrogen peroxide donors or organic peroxides results in improved activity for CYP3A4 and CYP2D6. *Chembiochem* **2006,** *7* (6), 916-9.

111. Horn, D. H. S., *The ecdysones. In "Naturally occurring insecticides"*. Dekker: **1971**; 333-459.
112. Therrien, E.; Englebienne, P.; Arrowsmith, A. G.; Mendoza-Sanchez, R.; Corbeil, C. R.; Weill, N.; Campagna-Slater, V.; Moitessier, N., Integrating medicinal chemistry, organic/combinatorial chemistry, and computational chemistry for the discovery of selective estrogen receptor modulators with Forecaster, a novel platform for drug discovery. *J Chem Inf Model* **2012,** *52* (1), 210-24.
113. Pottel, J.; Moitessier, N., Single-Point Mutation with a Rotamer Library Toolkit: Toward Protein Engineering. *J Chem Inf Model* **2015,** *55* (12), 2657-71.
114. Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D., Estimates of the ab initio limit for pi-pi interactions: the benzene dimer. *J Am Chem Soc* **2002,** *124* (36), 10887-93.
115. Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H., Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem Rev* **2010,** *110* (3), 1463-97.
116. Denisov, I. G.; Baas, B. J.; Grinkova, Y. V.; Sligar, S. G., Cooperativity in cytochrome P450 3A4: linkages in substrate binding, spin state, uncoupling, and product formation. *J Biol Chem* **2007,** *282* (10), 7066-76.
117. Frank, D. J.; Denisov, I. G.; Sligar, S. G., Analysis of heterotropic cooperativity in cytochrome P450 3A4 using alpha-naphthoflavone and testosterone. *J Biol Chem* **2011,** *286* (7), 5540-5.
118. Muller, C. S.; Knehans, T.; Davydov, D. R.; Bounds, P. L.; von Mandach, U.; Halpert, J. R.; Caflisch, A.; Koppenol, W. H., Concurrent cooperativity and substrate inhibition in the epoxidation of carbamazepine by cytochrome P450 3A4 active site mutants inspired by molecular dynamics simulations. *Biochemistry* **2015,** *54* (3), 711-21.
119. Fasan, R., Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catalysis* **2012,** *2*, 647–666.
120. Roiban, G. D.; Reetz, M. T., Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem Commun (Camb)* **2015,** *51* (12), 2208-24.
121. Urlacher, V. B.; Girhard, M., Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol* **2012,** *30* (1), 26-36.
122. Whitehouse, C. J.; Bell, S. G.; Wong, L. L., P450(BM3) (CYP102A1): connecting the dots. *Chem Soc Rev* **2012,** *41* (3), 1218-60.
123. Khatri, Y.; Girhard, M.; Romankiewicz, A.; Ringle, M.; Hannemann, F.; Urlacher, V. B.; Hutter, M. C.; Bernhardt, R., Regioselective hydroxylation of norisoprenoids by CYP109D1 from Sorangium cellulosum So ce56. *Appl Microbiol Biotechnol* **2010,** *88* (2), 485-95.
124. 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.
125. Bogazkaya, A. M.; von Buhler, C. J.; Kriening, S.; Busch, A.; Seifert, A.; Pleiss, J.; Laschat, S.; Urlacher, V. B., Selective allylic hydroxylation of acyclic terpenoids by CYP154E1 from Thermobifida fusca YX. *Beilstein J Org Chem* **2014,** *10*, 1347-53.
126. Zhang, K.; Shafer, B. M.; Demars, M. D., 2nd; Stern, H. A.; Fasan, R., Controlled oxidation of remote sp3 C-H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity. *J Am Chem Soc* **2012,** *134* (45), 18695-704.
127. Kille, S.; Zilly, F. E.; Acevedo, J. P.; Reetz, M. T., Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat Chem* **2011,** *3* (9), 738-43.

128. Jozwik, I. K.; Kiss, F. M.; Gricman, L.; Abdulmughni, A.; Brill, E.; Zapp, J.; Pleiss, J.; Bernhardt, R.; Thunnissen, A. W., Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 from Bacillus megaterium. *FEBS J* **2016**.
129. Lewis, J. C. A. F. H., Catalysts on Demand: Selective Oxidations by Laboratory-Evolved Cytochrome P450 BM3. *Biotransformations* **2009**.
130. Weber, E.; Seifert, A.; Antonovici, M.; Geinitz, C.; Pleiss, J.; Urlacher, V. B., Screening of a minimal enriched P450 BM3 mutant library for hydroxylation of cyclic and acyclic alkanes. *Chem Commun (Camb)* **2011,** *47* (3), 944-6.
131. Ghislieri, D.; Green, A. P.; Pontini, M.; Willies, S. C.; Rowles, I.; Frank, A.; Grogan, G.; Turner, N. J., Engineering an enantioselective amine oxidase for the synthesis of pharmaceutical building blocks and alkaloid natural products. *J Am Chem Soc* **2013,** *135* (29), 10863-9.
132. Shepherd, S. A.; Karthikeyan, C.; Latham, J.; Struck, A.-W.; Thompson, M. L.; Menon, B. R. K.; Styles, M. Q.; Levy, C.; Leys, D.; Micklefield, J., Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem. Sci.* **2015,** *6* (6), 3454-3460.
133. Güclü, D.; Szekrenyi, A.; Garrabou, X.; Kickstein, M.; Junker, S.; Clapés, P.; Fessner, W.-D., Minimalist Protein Engineering of an Aldolase Provokes Unprecedented Substrate Promiscuity. *ACS Catalysis* **2016,** *6* (3), 1848-1852.
134. Kim, Y.; Cunningham, M. A.; Mire, J.; Tesar, C.; Sacchettini, J.; Joachimiak, A., NDM-1, the ultimate promiscuous enzyme: substrate recognition and catalytic mechanism. *FASEB J* **2013,** *27* (5), 1917-27.
135. Nath, A.; Atkins, W. M., A quantitative index of substrate promiscuity. *Biochem* **2008,** *47* (1), 157-66.
136. Skopalik, J.; Anzenbacher, P.; Otyepka, M., Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. *J Phys Chem B* **2008,** *112* (27), 8165-73.
137. Larsen, A. T.; May, E. M.; Auclair, K., Predictable stereoselective and chemoselective hydroxylations and epoxidations with P450 3A4. *J Am Chem Soc* **2011,** *133* (20), 7853-8.
138. Larsen, A. T.; Lai, T.; Polic, V.; Auclair, K., Dual use of a chemical auxiliary: molecularly imprinted polymers for the selective recovery of products from biocatalytic reaction mixtures. *Green Chem* **2012,** *14* (8), 2206.
139. Cui, P.; Macdonald, T. L.; Chen, M.; Nadler, J. L., Synthesis and biological evaluation of lisofylline (LSF) analogs as a potential treatment for Type 1 diabetes. *Bioorg Med Chem Lett* **2006,** *16* (13), 3401-5.
140. de Vries, P.; Singer, J. W., Lisofylline suppresses ex vivo release by murine spleen cells of hematopoietic inhibitors induced by cancer chemotherapeutic agents. *Exp Hematol* **2000,** *28* (8), 916-23.
141. Campagna-Slater, V.; Pottel, J.; Therrien, E.; Cantin, L. D.; Moitessier, N., Development of a computational tool to rival experts in the prediction of sites of metabolism of xenobiotics by p450s. *J Chem Inf Model* **2012,** *52* (9), 2471-83.
142. Le Bideau, F.; Dagorne, S., Synthesis of transition-metal steroid derivatives. *Chem Rev* **2013,** *113* (10), 7793-850.
143. Biellmann, J. F., Enantiomeric steroids: synthesis, physical, and biological properties. *Chem Rev* **2003,** *103* (5), 2019-33.
144. Szeltner, Z.; Polgar, L., Structure, function and biological relevance of prolyl oligopeptidase. *Curr Protein Pept Sci* **2008,** *9* (1), 96-107.
145. Mannisto, P. T.; Venalainen, J.; Jalkanen, A.; Garcia-Horsman, J. A., Prolyl oligopeptidase: a potential target for the treatment of cognitive disorders. *Drug News Perspect* **2007,** *20* (5), 293-305.

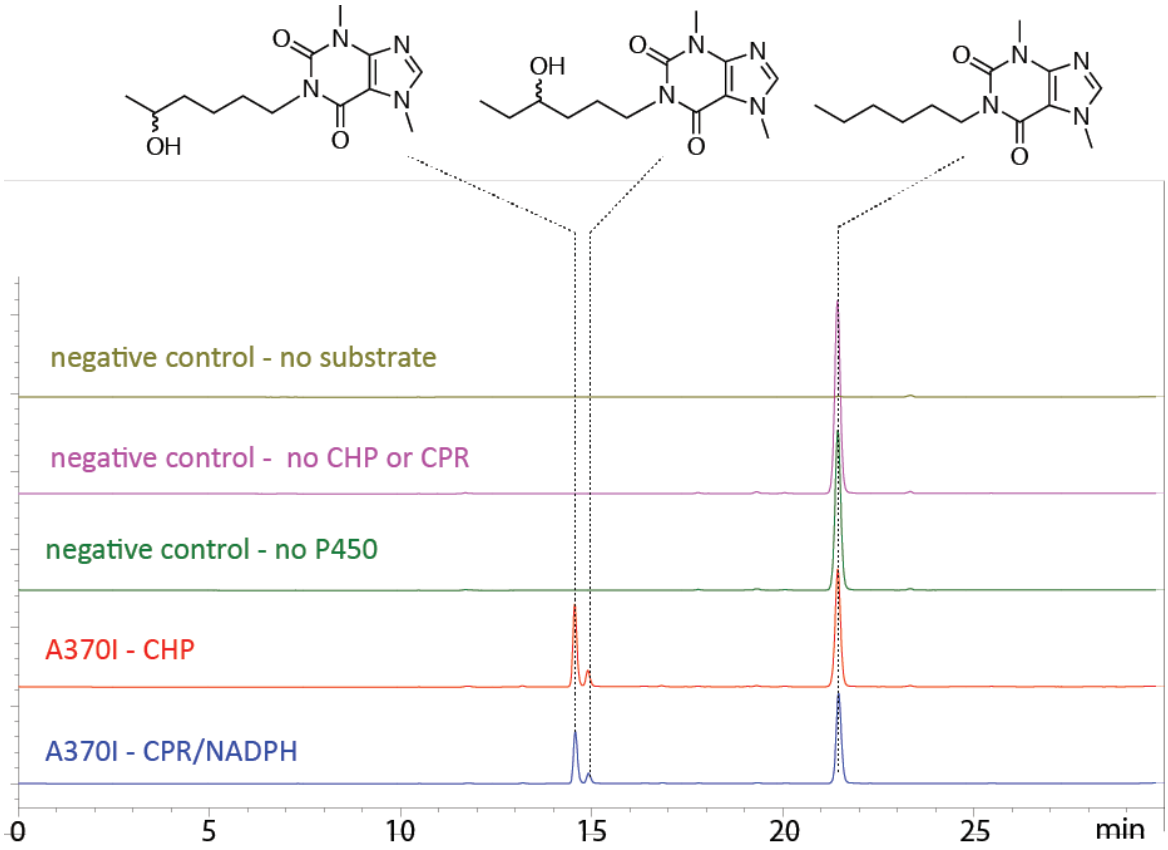
146. Lawandi, J.; Gerber-Lemaire, S.; Juillerat-Jeanneret, L.; Moitessier, N., Inhibitors of prolyl oligopeptidases for the therapy of human diseases: defining diseases and inhibitors. *J Med Chem* **2010,** *53* (9), 3423-38.
147. Williams, D. H. O. B., D. P.; Sandercock, A. M.; Stephens, E., Order Changes within Receptor Systems upon Ligand Binding: Receptor Tightening/Oligomerisation and the Interpretation of Binding Parameters. *J Mol Biol* **2004,** *340*, 373–383.
148. Sheridan, R. P.; Korzekwa, K. R.; Torres, R. A.; Walker, M. J., Empirical regioselectivity models for human cytochromes P450 3A4, 2D6, and 2C9. *J Med Chem* **2007,** *50* (14), 3173-84.
149. F. Gao, X. Y., O. Zahr, A. Larsen, K. Vong, K. Auclair, Synthesis and use of sulfonamide-, sulfoxide-, or sulfone-containing aminoglycoside–CoA bisubstrates as mechanistic probes for aminoglycoside N-6′-acetyltransferase. *Bioorg Med Chem Lett* **2008,** *50*, 3173–3184.
150. Levsen, K.; Schiebel, H. M.; Behnke, B.; Dotzer, R.; Dreher, W.; Elend, M.; Thiele, H., Structure elucidation of phase II metabolites by tandem mass spectrometry: an overview. *J Chromatogr A* **2005,** *1067* (1-2), 55-72.
151. Zhao, Y. M. W. L. W. Y., Intramolecular four-membered ring rearrangements of 1-(N-benzyloxycarbonylamino)arylmethylphosphonate monoesters under electrospray ionization conditions. *Int J Mass* **2006,** *247* (1-2), 17.
152. Y.-H. Lim, D.-H. P., Y.-Y. Youn, K.-H. Kim, H.-S. Cho, Identification of Ceftiofur Oxidation Products by High-Performance Liquid Chromatography/Electrospray Ionization/Tandem Mass Spectrometry. *Mass Spectrom Lett* **2011,** *2*, 16-19.
153. Herchen, S. R., Syntheses and thermal transformations of N-alkyl thiazolidine sulfoxides. *Tetrahedron Lett* **1989,** *30*, 425-428.
154. S. Hanessian, G. M.-S., H.-G. Lombart, W. D. Lubell, Design and synthesis of conformationally constrained amino acids as versatile scaffolds and peptide mimetics. *Tetrahedron* **1997,** *53*, 12789-12854.
155. H. Wagner, K. S., G. Jaehne, H. Gaul, C. Buning, G. Tschank, U. Werner 2005.
156. T. Vilsbøll, F. K. K., DPP IV inhibitors - current evidence and future directions. *Br J Diabetes Vasc Dis* **2007,** *7*, 69-74.
157. A. Thomas, B. G., P. R. Lingam, D. M. Shah 2006.
158. Gupta, R. C.; Chhipa, L.; Mandhare, A. B.; Zambad, S. P.; Chauthaiwale, V.; Nadkarni, S. S.; Dutt, C., Novel N-substituted 4-hydrazino piperidine derivative as a dipeptidyl peptidase IV inhibitor. *Bioorg Med Chem Lett* **2009,** *19* (17), 5021-5.
159. Mariaule, G.; De Cesco, S.; Airaghi, F.; Kurian, J.; Schiavini, P.; Rocheleau, S.; Huskic, I.; Auclair, K.; Mittermaier, A.; Moitessier, N., 3-Oxo-hexahydro-1H-isoindole-4-carboxylic Acid as a Drug Chiral Bicyclic Scaffold: Structure-Based Design and Preparation of Conformationally Constrained Covalent and Noncovalent Prolyl Oligopeptidase Inhibitors. *J Med Chem* **2016,** *59* (9), 4221-34.
160. Kagan, H., *Asymmetric synthesis of chiral sulfoxides*. 2008.
161. Gao, P. L., A.; H. H. Lee; Wang, D. I. C.; Li, Z., Enhancing Enantioselectivity and Productivity of P450-Catalyzed Asymmetric Sulfoxidation with an Aqueous/Ionic Liquid Biphasic System. *ACS Catalysis* **2014,** *4*, 3763-3771.
162. Subba Rao, V. K. K., P., *Tetrahedron: Asymmetry* **2005,** *16*, 3268–3274.
163. Wehbe, J. R. V. R., M. L.; Martinez, J., Glutamate transporter blockers: enantiomerically pure (2S,3S)- and (2S,3R)-3-methyl glutamic acids. *Tetrahedron: Asymmetry* **2003,** (9), 1123-1126.
164. Hahn, R. C., Tompkins, J., One-pot, one- and multi-carbon homologation of alkyl halides; reaction of grignard reagents with chloroiodomethane. *Tetrahedron Lett* **1990,** *31* (7), 937–940.

165. Domanski, T. L.; Liu, J.; Harlow, G. R.; Halpert, J. R., Analysis of four residues within substrate recognition site 4 of human cytochrome P450 3A4: role in steroid hydroxylase activity and alpha-naphthoflavone stimulation. *Arch Biochem Biophys* **1998,** *350* (2), 223-32.
166. Rabe, K. S., Gandubert, V. J., Spengler, M., Erkelenz, M., and Niemeyer, C. M. Engineering and assaying of cytochrome P450 biocatalysts, *Anal Bioanal Chem* **2008**, 392, 1059-1073.
167. Bell, S. G., Chen, X. H., Sowden, R. J., Xu, F., Williams, J. N., Wong, L. L., and Rao, Z. H. Molecular recognition in (+)-alpha-pinene oxidation by cytochrome P450(cam), *J Am Chem Soc* **2003,** 125, 705-714.
168. Harford-Cross, C. F., Carmichael, A. B., Allan, F. K., England, P. A., Rouch, D. A., and Wong, L. L. Protein engineering of cytochrome P450(cam) (CYP101) for the oxidation of polycyclic aromatic hydrocarbons, *Protein Eng*. **2000**, 13, 121-128.
169. Jones, J. P., O'Hare, E. J., and Wong, L. L. The oxidation of polychlorinated benzenes by genetically engineered cytochrome P450(cam): potential applications in bioremediation, *Chem Commun*, **2000,** 247-248.
170. Manna, S. K., and Mazumdar, S. Tuning the substrate specificity by engineering the active site of cytochrome P450cam: A rational approach, *Dalton Trans*. **2010**, 39, 3115- 3123.
171. Xu, F., Bell, S. G., Lednik, J., Insley, A., Rao, Z. H., and Wong, L. L. The heme monooxygenase cytochrome P450(cam) can be engineered to oxidize ethane to ethanol, *Angew Chem Int Ed* **2005**, 44, 4029-4032.
172. Hirakawa, H., and Nagamune, T. Molecular assembly of P450 with ferredoxin and ferredoxin reductase by fusion to PCNA, *Chembiochem* **2010**, 11, 1517-1520.
173. Munro, A. W., Girvan, H. M., and McLean, K. J. Variations on a (t)heme - novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily, *Natural Product Reports* **2007**, 24, 585-609
174. Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A., and Deisenhofer, J. Crystal-structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's, *Science* **1993**, 261, 731-736
175. Ost, T. W. B., Miles, C. S., Murdoch, J., Cheung, Y. F., Reid, G. A., Chapman, S. K., and Munro, A. W. Rational re-design of the substrate binding site of flavocytochrome P450BM3, *FEBS Lett*. **2000**, 486, 173-177.
176. Dietrich, J. A., Yoshikuni, Y., Fisher, K. J., Woolard, F. X., Ockey, D., McPhee, D. J., Renninger, N. S., Chang, M. C. Y., Baker, D., and Keasling, J. D. A Novel Semi- biosynthetic Route for Artemisinin Production Using Engineered Substrate-Promiscuous P450(BM3), *ACS Chem Biol* **2009**, 4, 261-267.
177. Appel, D., Lutz-Wahl, S., Fischer, P., Schwaneberg, U., and Schmid, R. D. A P450BM-3 mutant hydroxylates alkanes, cycloalkanes, arenes and heteroarenes, *J Biotechnol* **2001**, 88, 167-171.
178. Urlacher, V. B., Makhsumkhanov, A., and Schmid, R. D. Biotransformation of beta-ionone by engineered cytochrome P450BM-3, *Appl Microbiol Biotechnol* **2006**, 70, 53-59.
179. Sowden, R. J., Yasmin, S., Rees, N. H., Bell, S. G., and Wong, L. L. Biotransformation of the sesquiterpene (+)-valencene by cytochrome P450(cam) and P450(BM-3), *Org Biomol Chem* **2005**, 3, 57-64.
180. Meinhold, P., Peters, M. W., Hartwick, A., Hernandez, A. R., and Arnold, F. H. Engineering cytochrome P450BM3 for terminal alkane hydroxylation, *Adv Synth Catal,* **2006**, 348, 763-772.
181. Kubo, T., Peters, M. W., Meinhold, P., and Arnold, F. H. Enantioselective epoxidation of terminal alkenes to (R)- and (S)-epoxides by engineered cytochromes P450BM-3, *Chem Eur J* **2006**, 12, 1216-1220.
182. Prier C. K., Hyster, T. K., Farwell, C. C. Huang, A., Arnold, F. H. Asymmetric Enzymatic Synthesis of Allylic Amines: A Sigmatropic Rearrangement Strategy" *Angew*

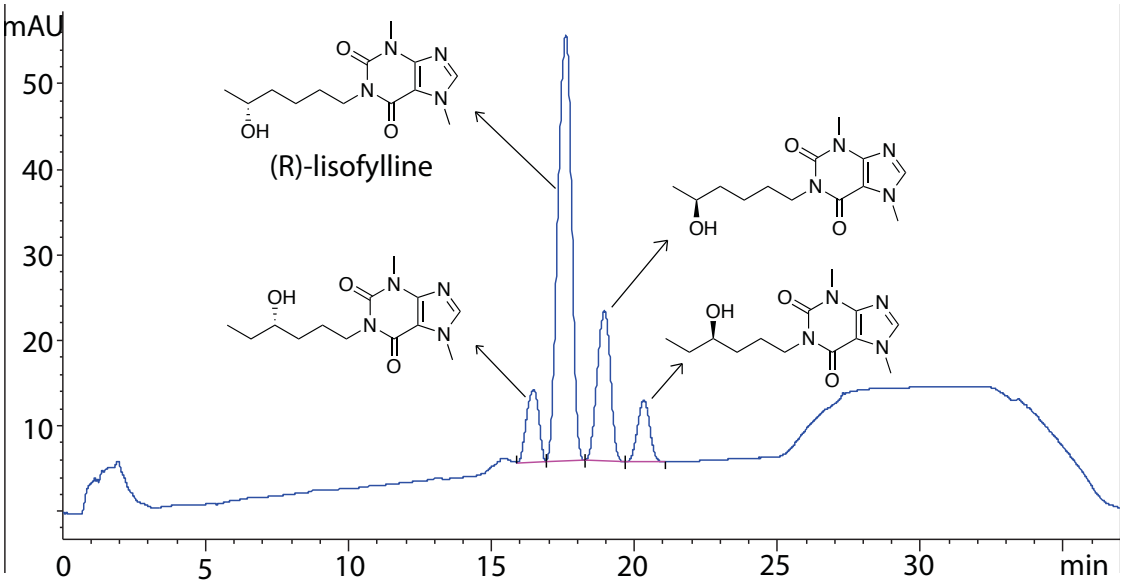
Chem **2016**, 55, 4711-4715
183. Moitessier N., Pottel J., Therrien E., Englebienne P., Liu Z., Tomberg A., Corbeil C.R. Medicinal Chemistry Projects Requiring Imaginative Structure-Based Drug Design Methods. *Acc Chem Res* **2016**, 49 (9), 1646–1657
184. Corbeil C.R., Englebienne P., Moitessier N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0, *J Chem Inf Mod* **2007**, 47, 2, 435-449
185. Corbeil, C. R.; Englebienne, P.; Yannopoulos, C. G.; Chan, L.; Das, S. K.; Bilimoria, D.; L'Heureux, L.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 2. Development and Application of FITTED 1.5 to the Virtual Screening of Potential HCV Polymerase Inhibitors. *J. Chem. Inf. Model.* **2008**, 48, 902–909.
186. Xu, W.; Lucke, A. J.; Fairlie, D. P. Comparing sixteen scoring functions for predicting biological activities of ligands for protein targets. *J Mol Graphics Modell.***2015**, 57, 76–88
187. Marciano D. P., Dharmarajan V., Griffin P. R., HDX-MS guided drug discovery: small molecules and biopharmaceuticals, *Curr Opin Struct Biol* **2014**, 28, 105-111
188. Zhang, X., Chien, E. Y. T., Chalmers, M. J., Pascal, B. D., Gatchalian, J., Stevens R. C., Griffin, P. R., Dynamics of the β2-Adrenergic G-Protein Coupled Receptor Revealed by Hydrogen–Deuterium Exchange, *Anal Chem,* **2010**, 82 (3), 1100–1108
189. Zhang H, Cui W, Gross ML: Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Lett* **2014**, 588, 308-317
190. Bowie, D. External anions and cations distinguish between AMPA and kainate receptor gating mechanisms. *J Physiol* **2002**, 539, 725-733

Appendix – HPLC and LC-MS chromatograms

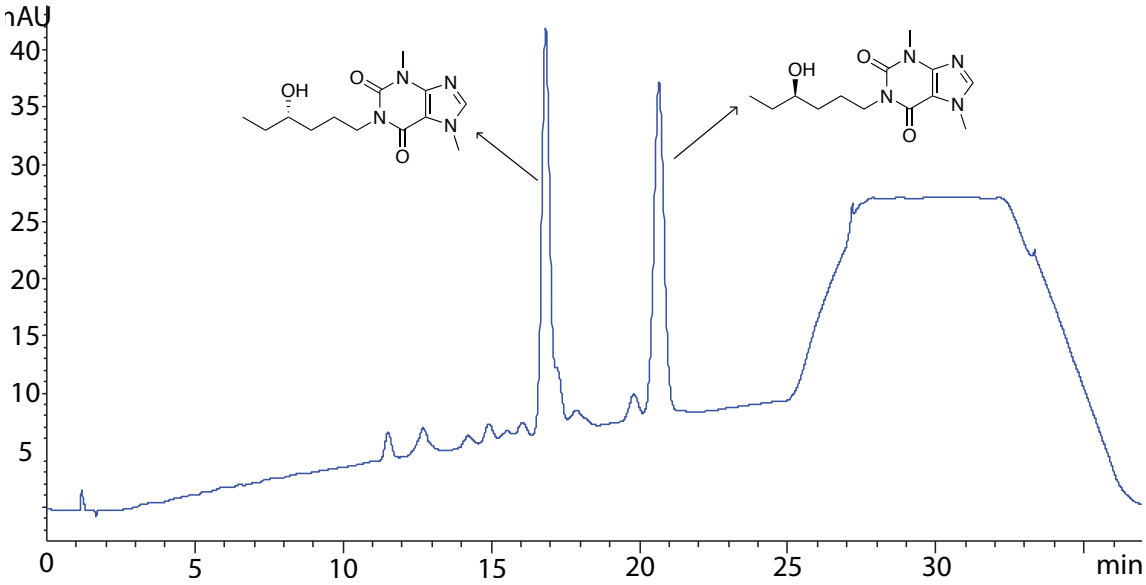
Oxidation of (*R*)-lisofylline precursor **4.1** with purified enzyme, non-chiral separation

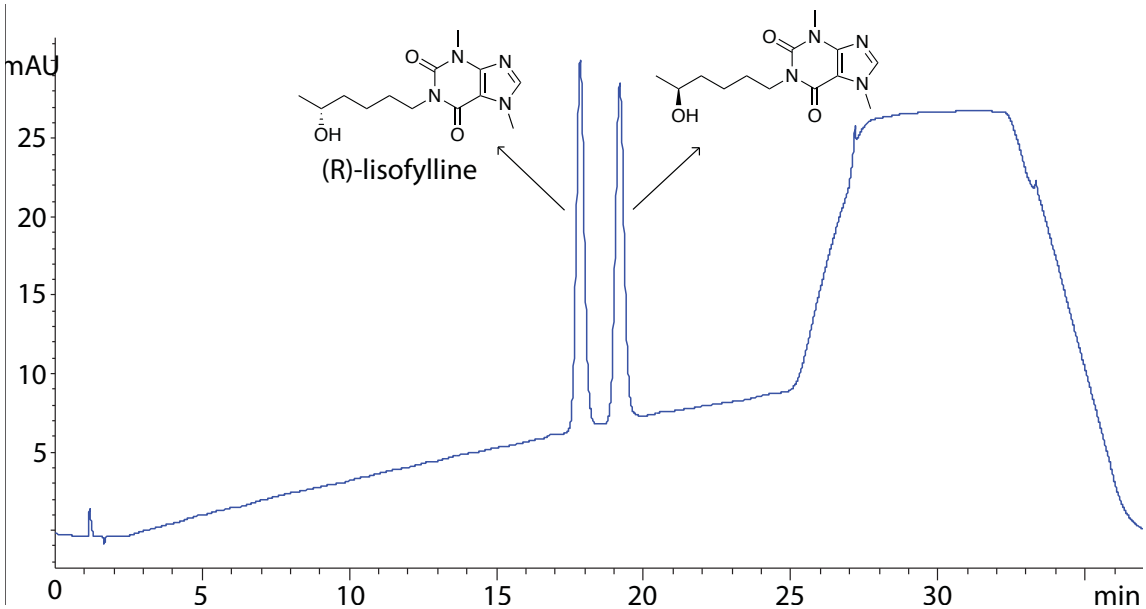


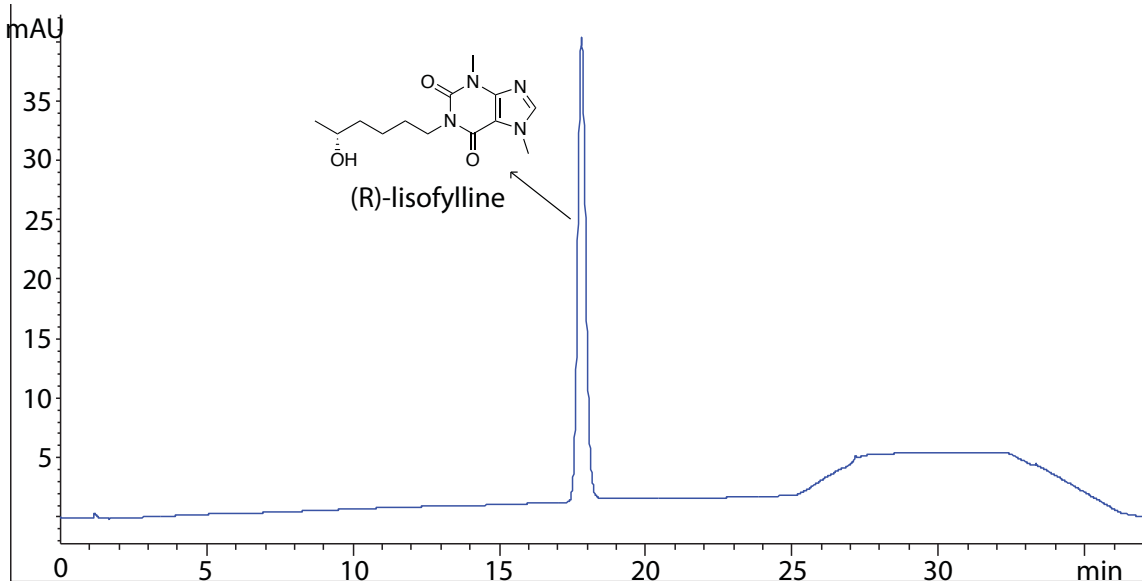
Oxidation of (*R*)-lisofylline precursor **4.1** with purified enzyme, chiral separation



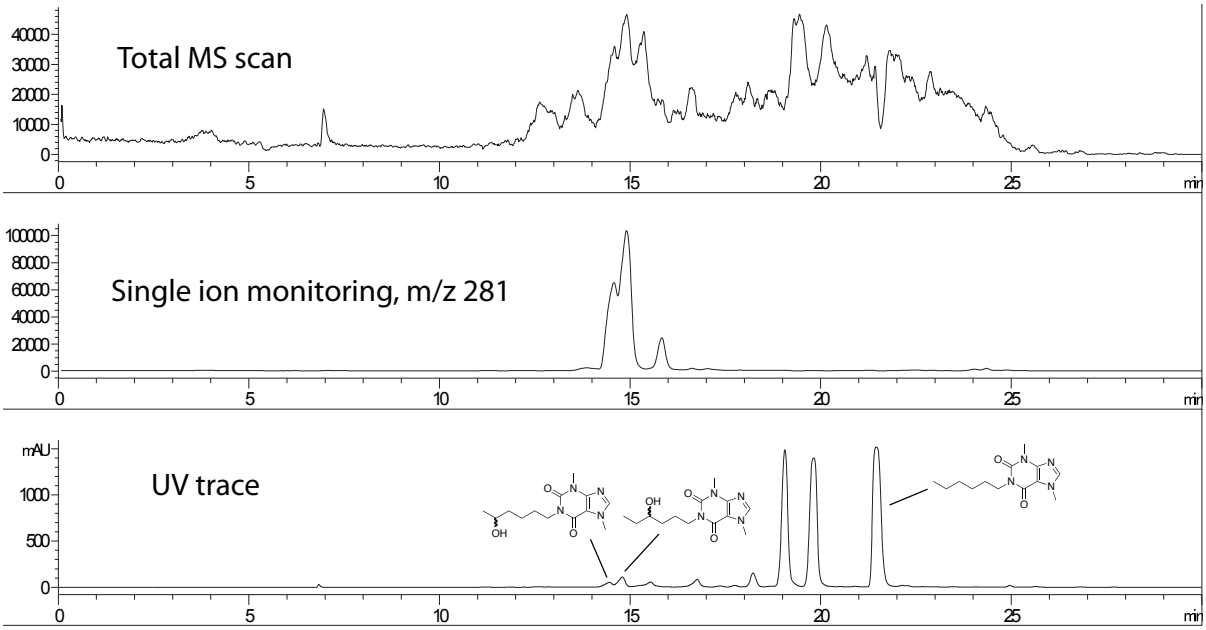
Chemical standards for the oxidation products of (*R*)-lisofylline precursor **4.1**



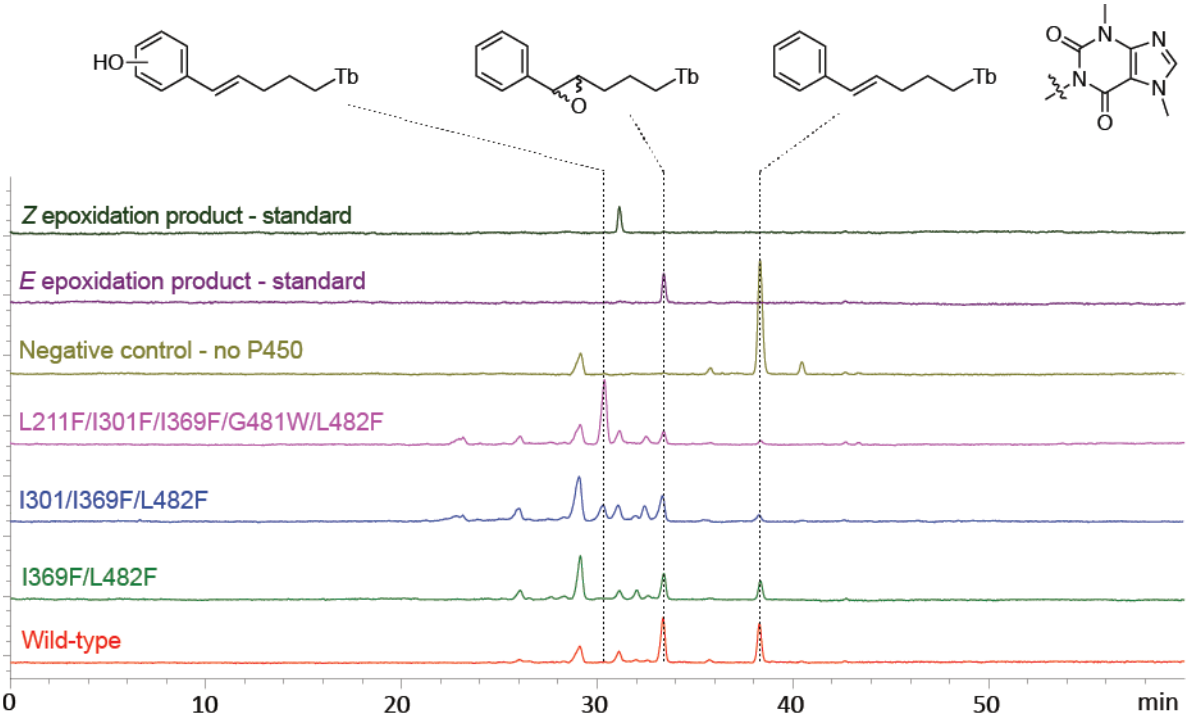




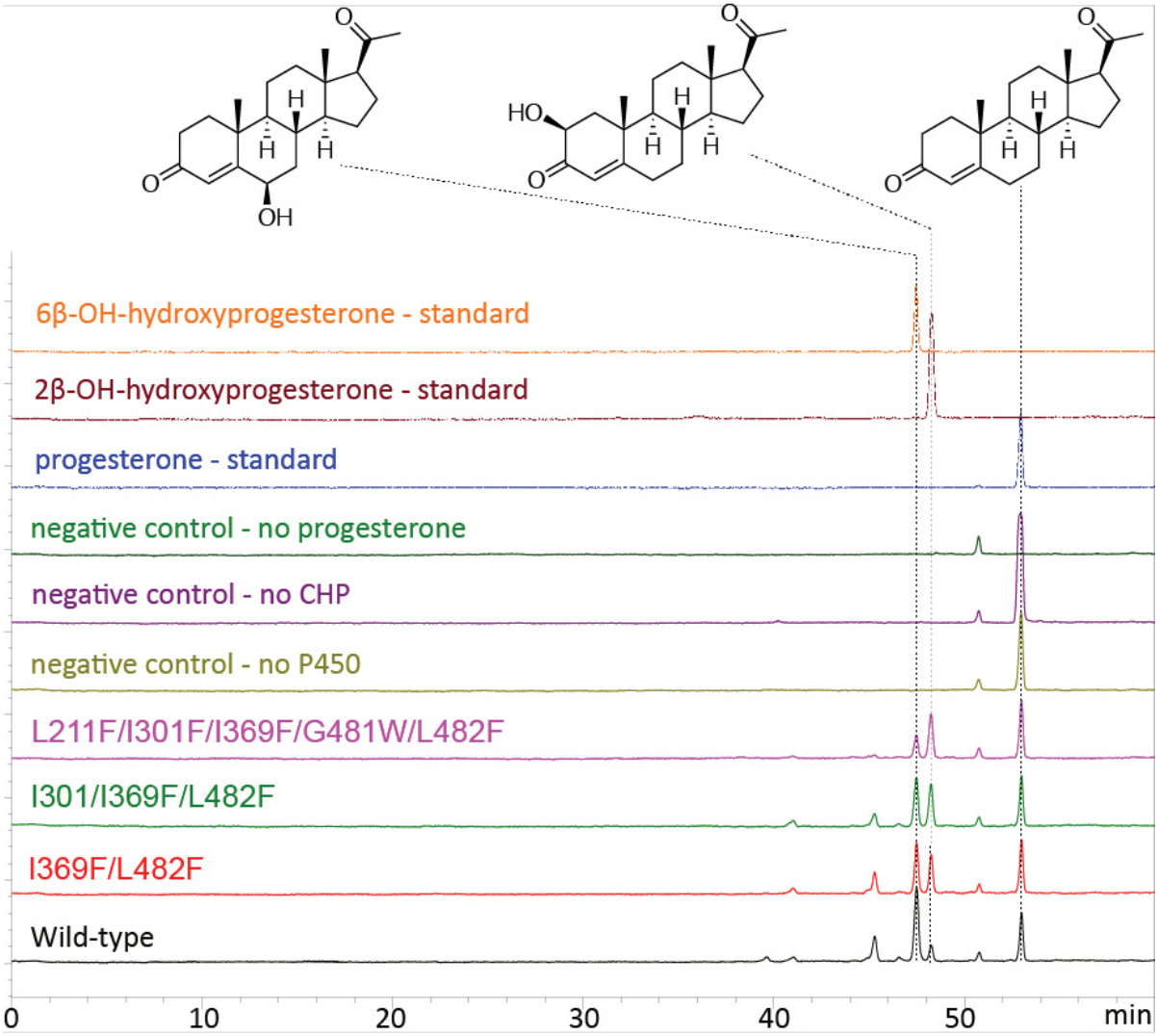
Oxidation of (*R*)-lisofylline precursor **4.1** with P450-containing cell lysate



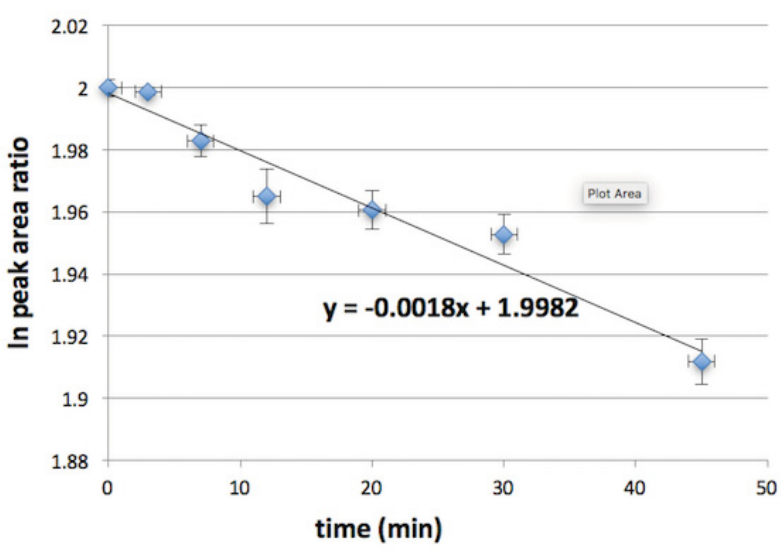
Oxidation of **4.7** with purified enzyme

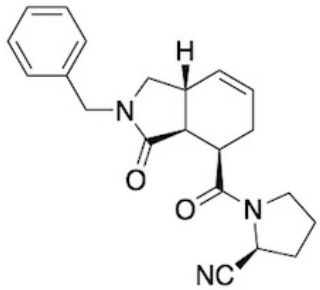


Oxidation of progesterone with purified enzyme



Disappearing of **5.11** over time during incubation with human liver microsomes, monitored as ln ratio (peak area at t = 0)/(peak area at t = X)



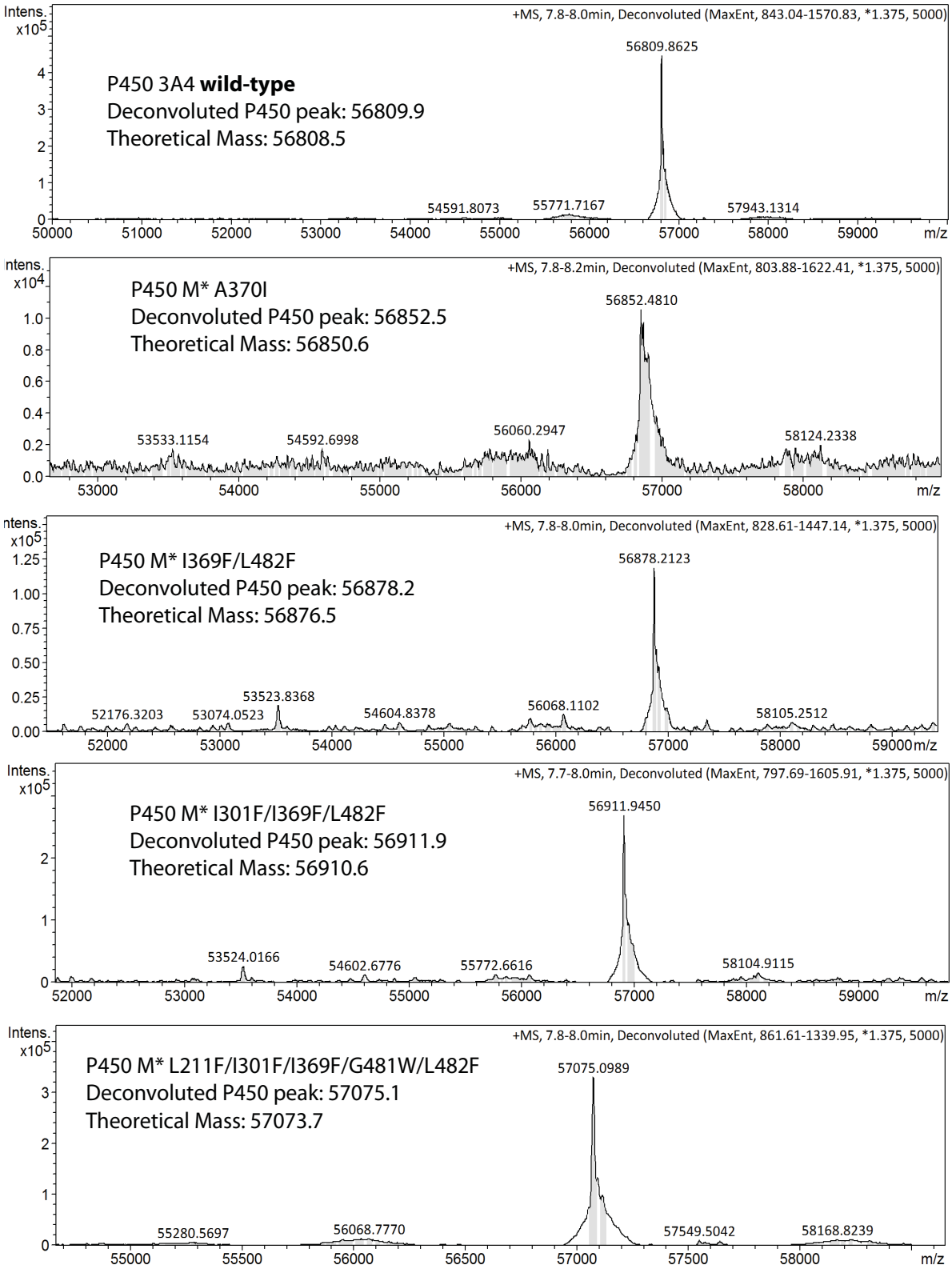


k = + 0.0018

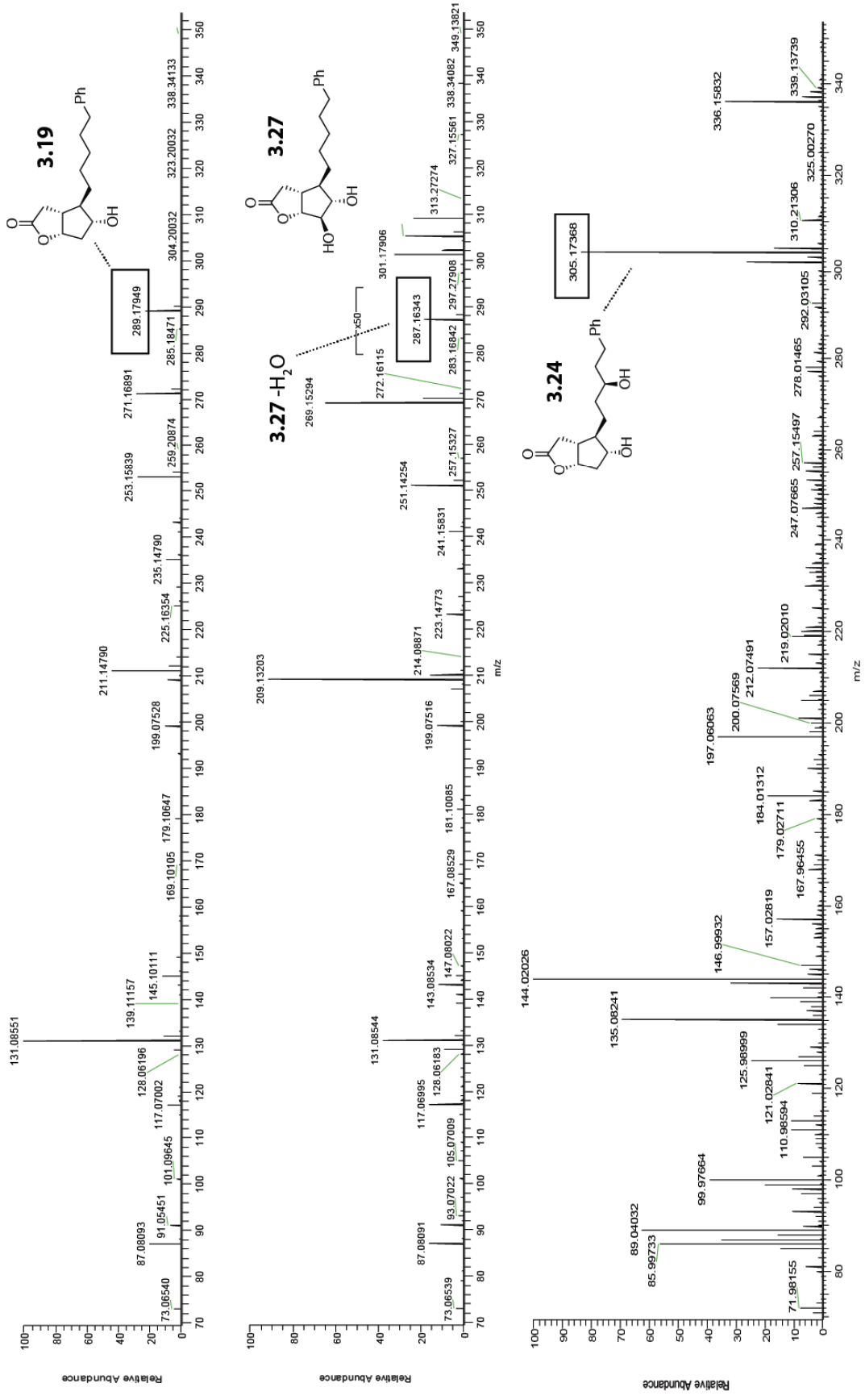
$t_{1/2}$ = 385 min^{-1}

Cl_{int} = 4 ul/min/mg protein

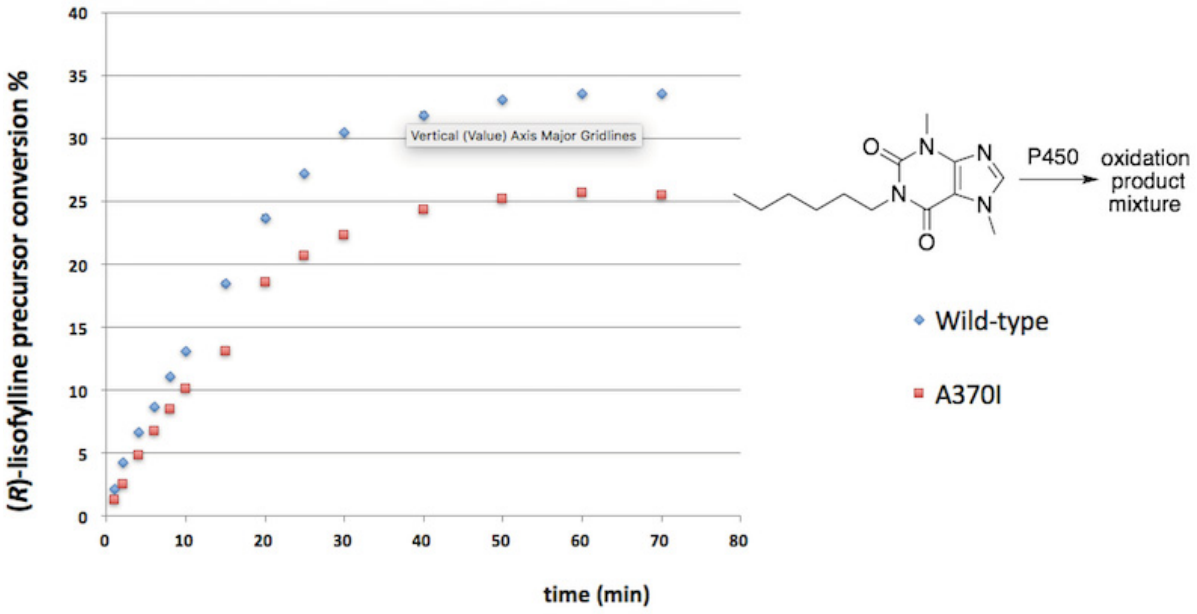
Protein HPLC-MS analysis of P450 3A4 wild-type and mutants A370I, I369F/L482F, I301F/I369F/L482F, L211F/I301F/I369F/G481W/L482F

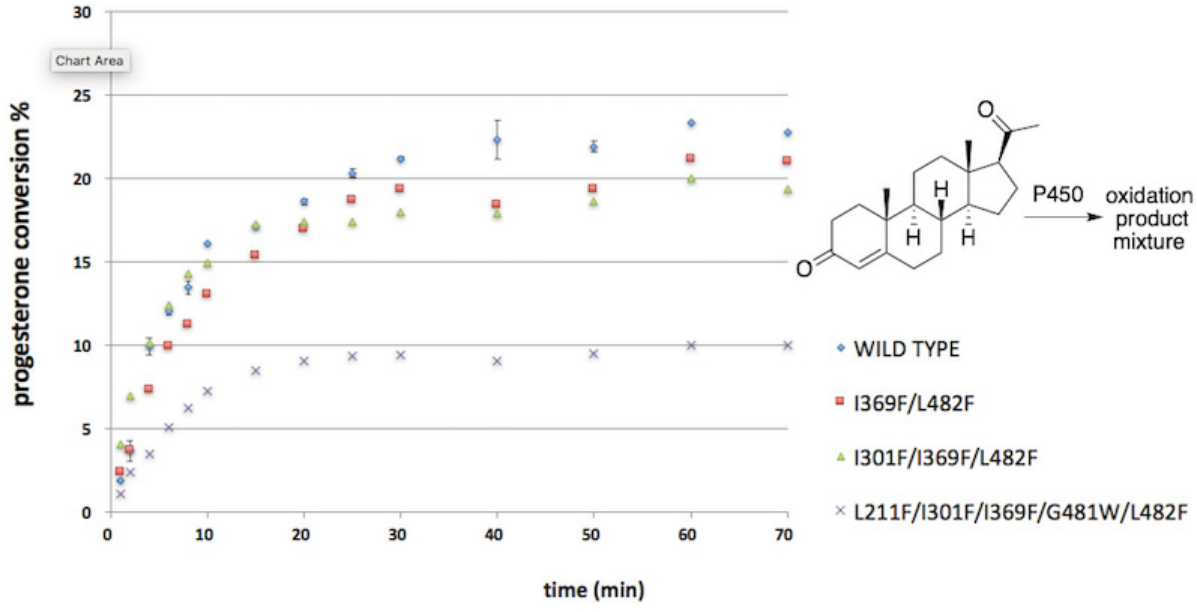


Mass Spectrometry fragmentation data for **3.19**, **3.24** and **3.27**



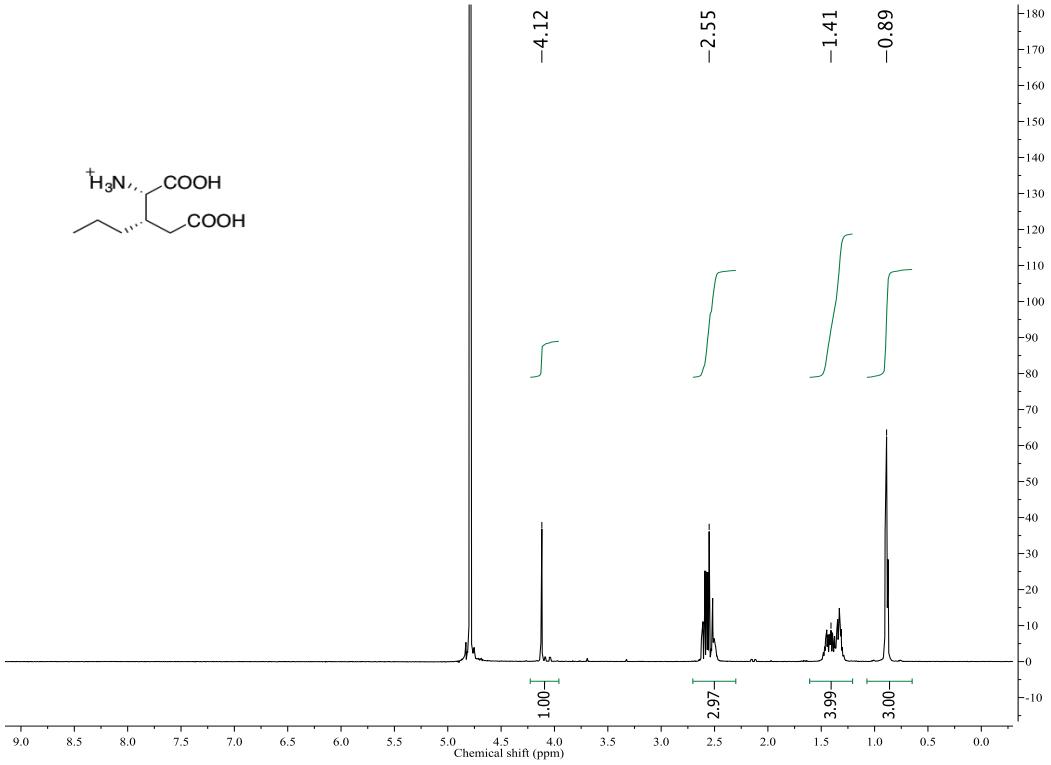
Enzymes activity curves



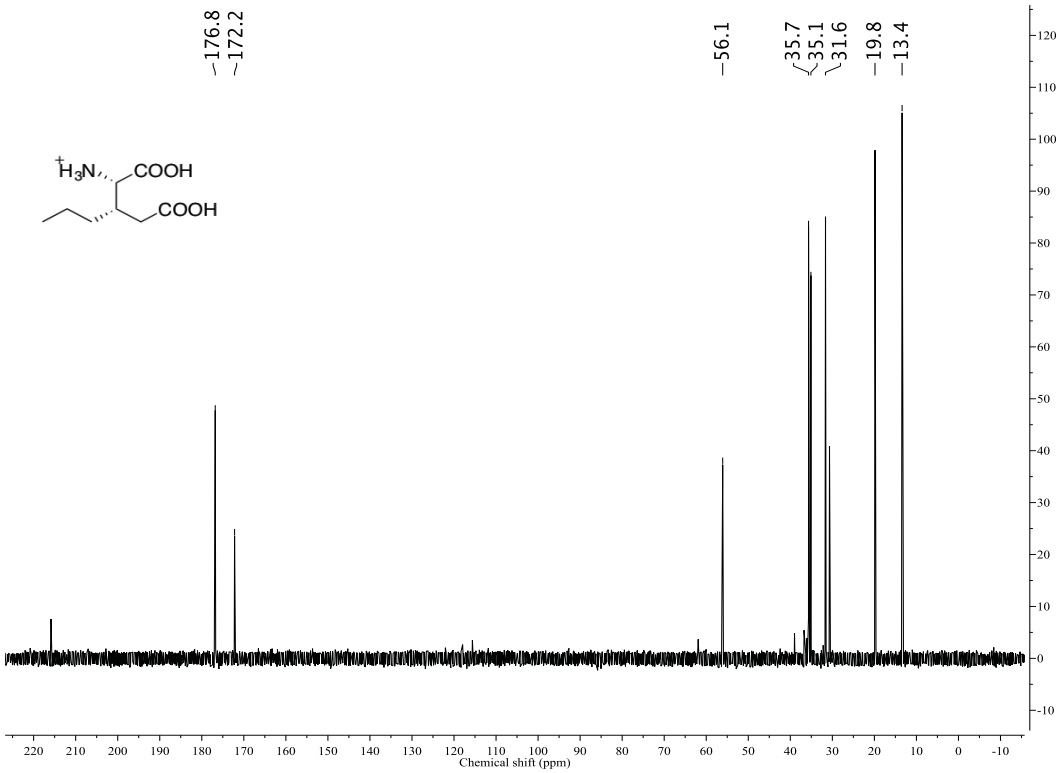


Appendix – selected NMR spectra

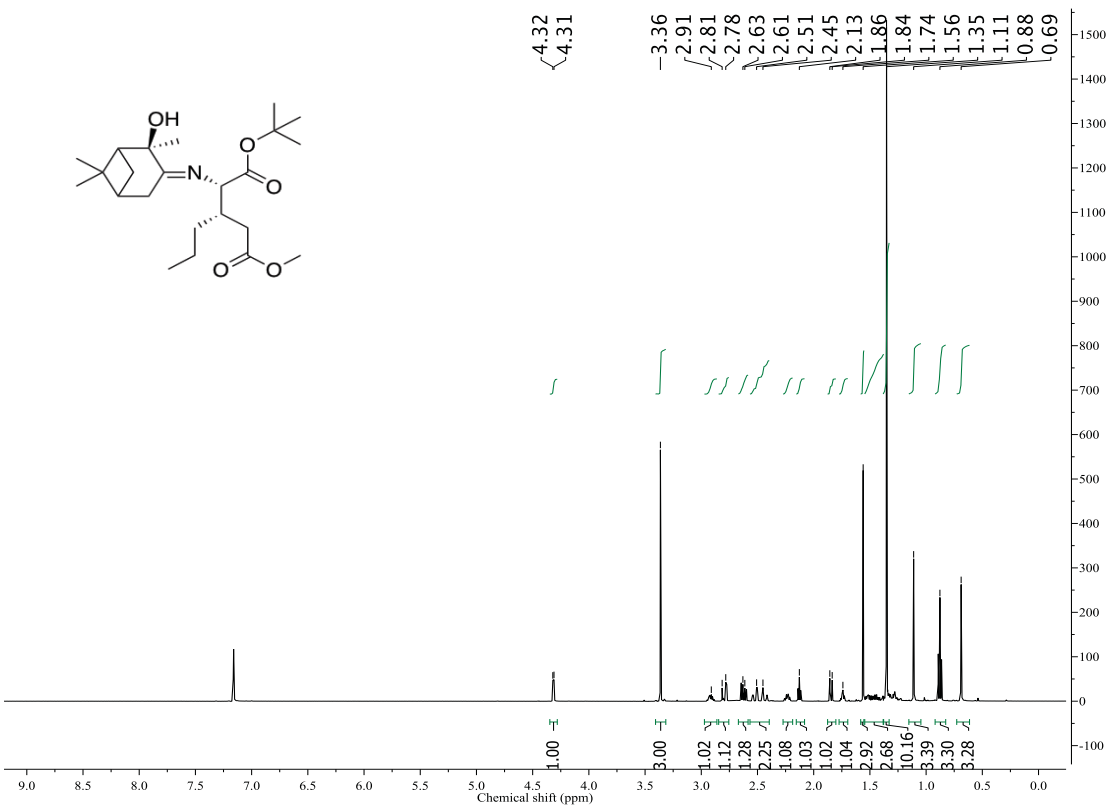
^{1}H NMR of compound **2.6**



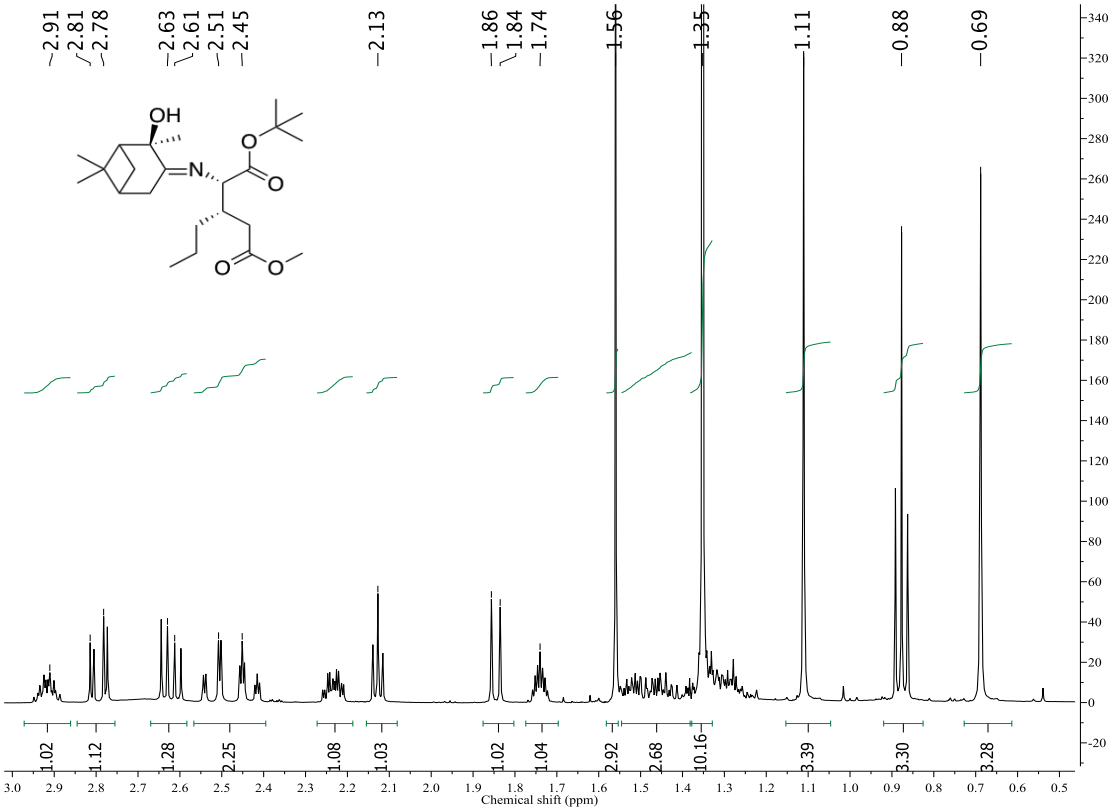
^{13}C NMR of compound **2.6**



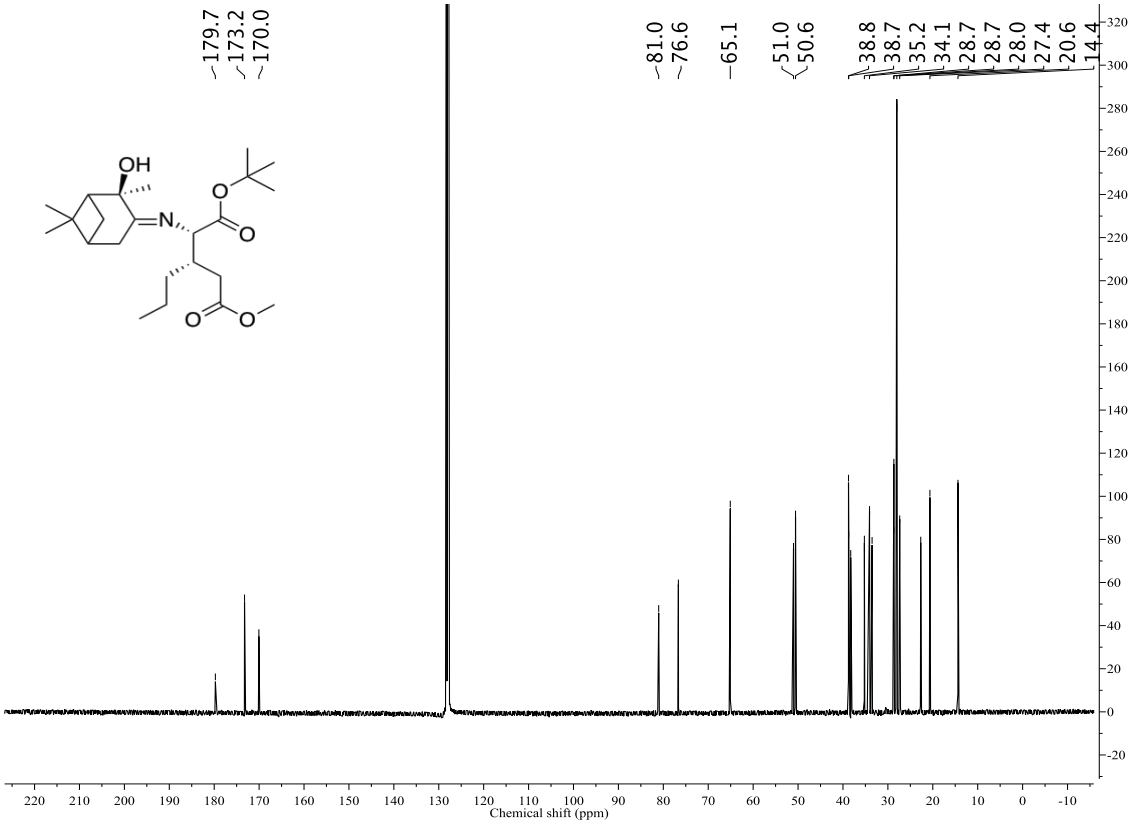
^{1}H NMR of compound **2.19**



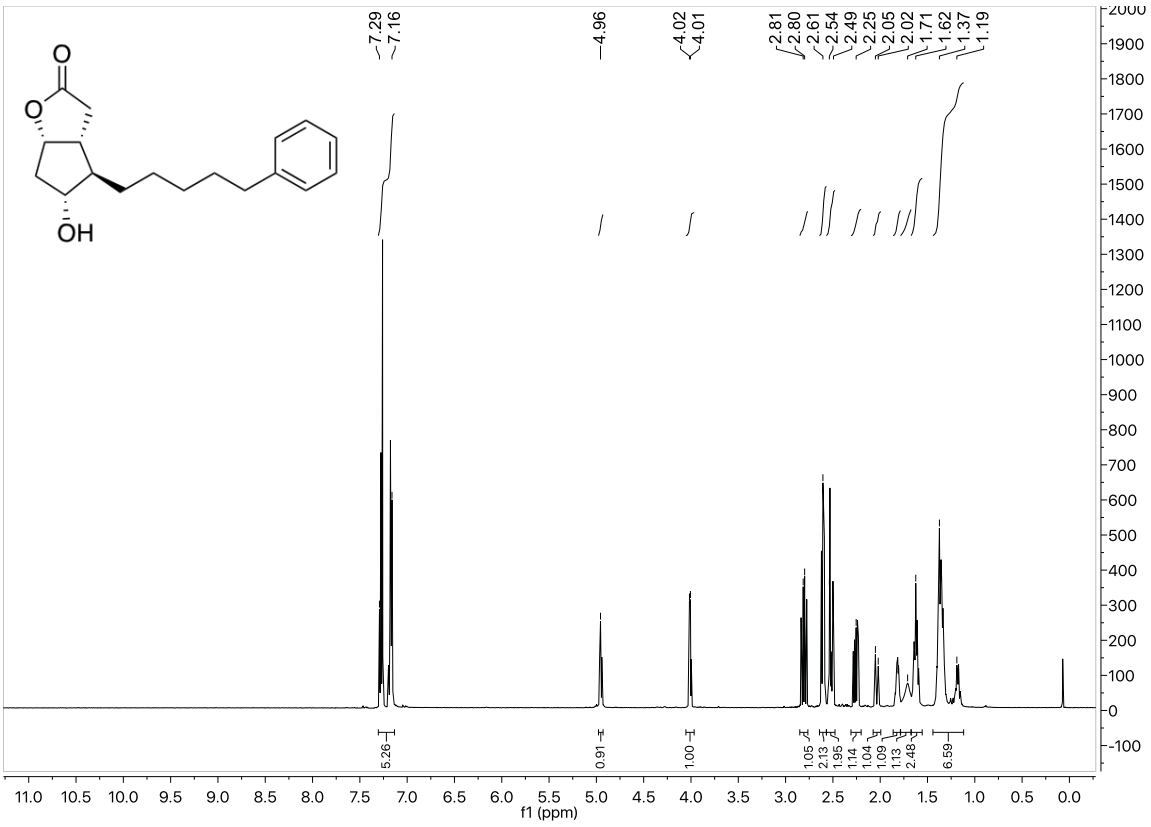
^{1}H NMR of compound **2.19**, zoom 0.5-3 ppm



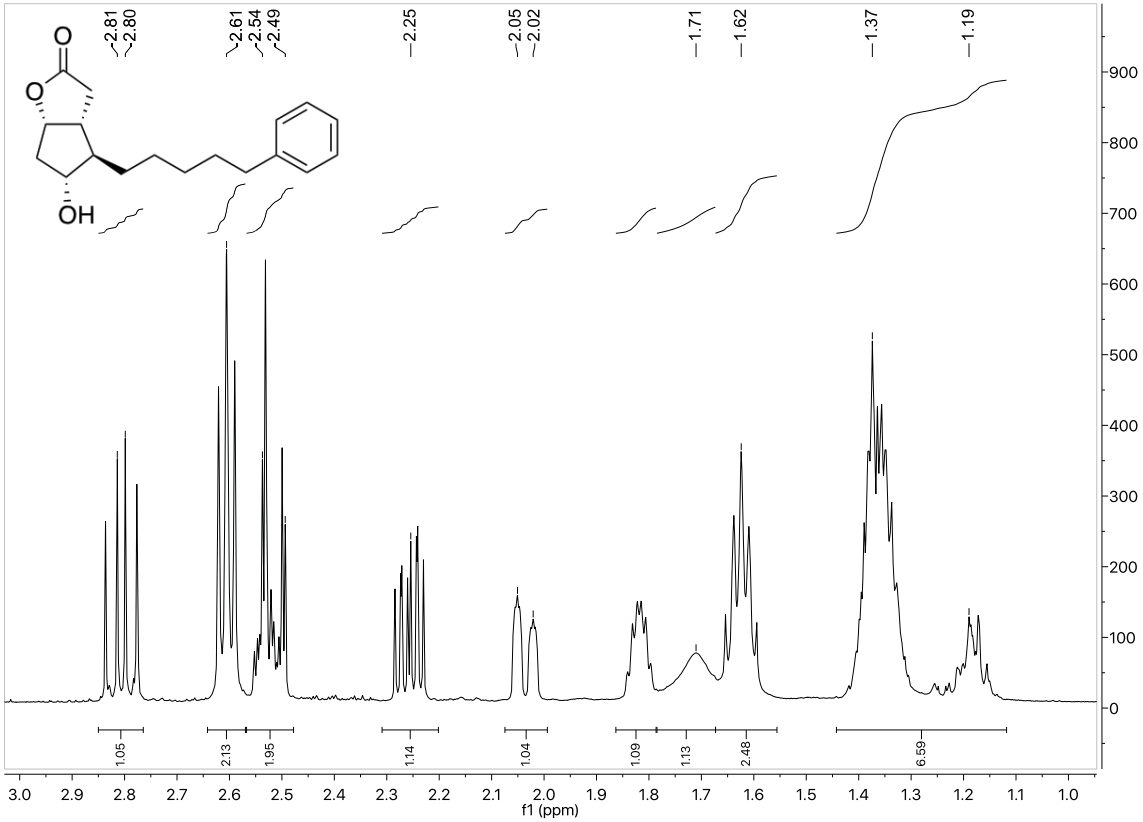
^{13}C NMR of compound **2.19**



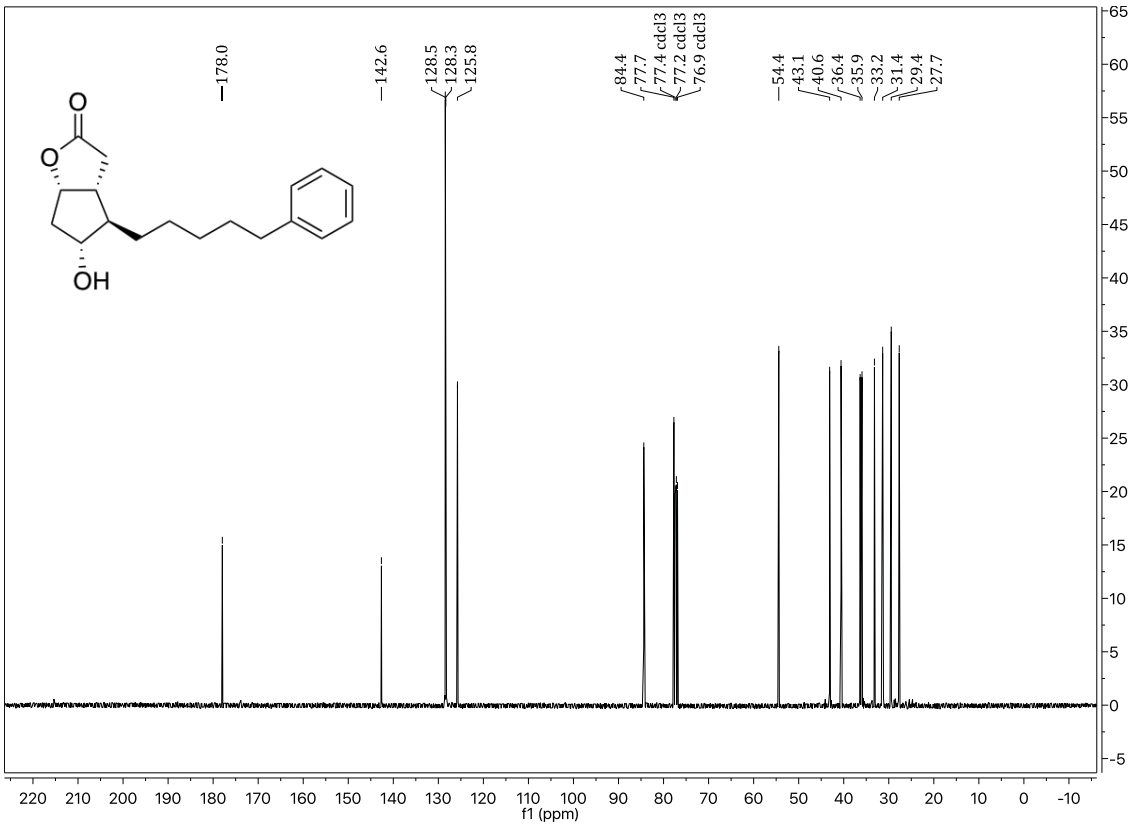
^{1}H NMR of compound **3.19**



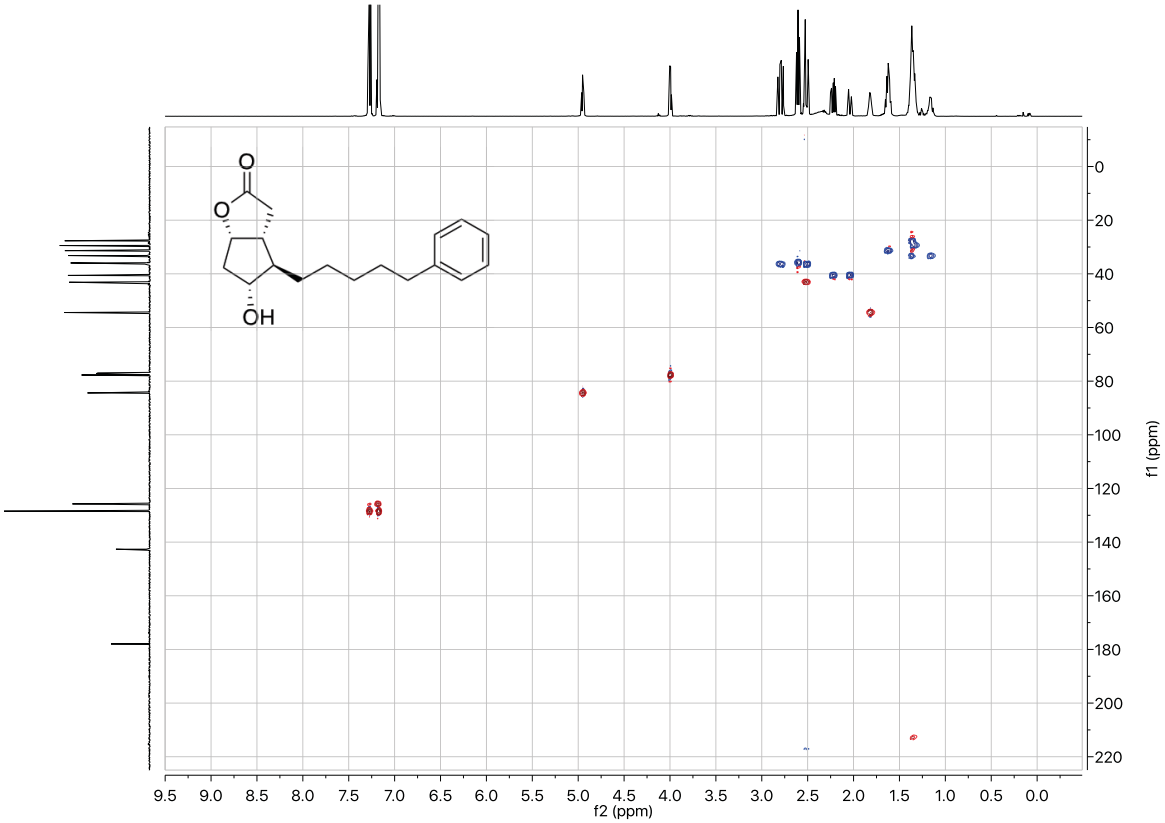
^{1}H NMR of compound **3.19**, zoom 1-3 ppm



^{13}C NMR of compound **3.19**



HSQC NMR of compound **3.19**



NOESY NMR of compound **3.19**