Exploring the effects of polymorphic variation on the stability and function of human cytochrome P450 enzymes *in silico* and *in vitro*

Lauren Beth Arendse

Dissertation submitted for the degree of Doctor of Philosophy

In the Department of Clinical Laboratory Sciences

University of Cape Town

August 2014

Abstract

Cytochrome P450s are highly polymorphic enzymes responsible for the Phase I metabolism of over 80% of pharmaceutical drugs. Polymorphic variation can result in altered drug efficacy as well as adverse drug reactions so the lack of understanding of the effects of single amino acid substitutions on cytochrome P450 drug metabolism is a major problem for drug development.

In order to begin to address this problem, this thesis describes an *in silico* analysis of over 300 non-synonymous single nucleotide polymorphisms found across nine of the major human drug metabolising cytochrome P450 isoforms. Information from functional studies - in which regions of the cytochrome P450 structure important for substrate recognition, substrate and product access and egress and interaction with the cytochrome P450 reductase were delineated - was combined with *in silico* calculations on the effect of mutations on protein stability in order to establish the likely causes of altered drug metabolism observed for cytochrome P450 variants in functional assays carried out to date. This study revealed that 75% of all cytochrome P450 mutations showing altered activity *in vitro* are either predicted to be damaging to protein structure or are found within regions predicted to be important for catalytic activity. Furthermore, this study showed that 70% of the mutations that showed similar activity to the wild-type enzyme in *in vitro* studies lie outside of functional regions important for catalytic activity and are predicted to have no effect on protein stability. Based on these results, a cytochrome P450 polymorphic variant map was created that should find utility in predicting the functional effect of uncharacterised variants on drug metabolism.

To further test the accuracy of the *in silico* predictions, *in vitro* assays were performed on a panel of CYP3A4 and CYP2C9 variants heterogeneously expressed in *E.coli.* All mutations predicted to alter protein function by stabilising or destabilising the apo-protein structure *in silico* were found to significantly alter the thermostability of the holo-protein in solution. Thermostability assays also suggest that other mutations may affect stability by disrupting haem binding, changing protein conformation or altering oligomer formation.

The utility of a fluorescence-based functional P450 protein microarray platform, previously developed in our laboratory, for generating kinetic data for multiple CYP450 variants in parallel was also examined. Since the microarray platform in its current stage of development was found to be unsuitable for this purpose, kinetic data for the full panel of CYP3A4 and CYP2C9 variants was generated using solution phase assays, revealing several variants with altered catalytic turnover and/or binding affinity for fluorescent substrates.

Declaration

I, Lauren Arendse, declare that this thesis is my own work (except where acknowledgements indicate otherwise). Neither the whole work nor part thereof has been, is being, or is to be submitted for any degree or examination at any other university.

I empower the University of Cape Town to reproduce for the purposes of research either the whole or any part of the contents of this thesis, in any manner whatsoever.

Signature of candidate: _____

Signed on the ____ day of _____, 2014

Dedication

Optimus Parentibus, David & René

Acknowledgements

I would like to express my deepest appreciation to my supervisor, Professor Jonathan Blackburn. I have benefited greatly from your remarkable scientific mind, your ability to inspire and your tendency to see 'the cup half full rather than half empty'. I truly appreciate all the incredible opportunities you have afforded me during my PhD. Thank you for your guidance, support and patience over the years.

I would also like to express my gratitude to Professor Tom Blundell for welcoming me into his Biocomputing group: thank you for your support during my time at the University of Cambridge.

To Dr Alexander Zawaira, thank you for introducing me to the world of Cytochrome P450 structures and for your contribution to the *in silico* work described in this thesis. To Dr Will Pitt, thank you for being ever willing to assist me during my time in the Blundell Group. To Dr Aubrey Shoko from the Centre for Proteomics and Genomics Research, Cape Town, thank you for your assistance with the protein microarray assays.

I also express my sincerest gratitude to the many past and present members of the Blackburn group who have assisted me in numerous ways along this journey, thank you for sharing the good times and encouraging me through the bad times.

Thank you to the Harry Crossley Research Foundation, the National Research Foundation and the Commonwealth Scholarship Commission for your financial support during this project.

To my parents, René and David Coulson, thank you for the endless sacrifices you have made for me. Your steadfast belief in me has made all the difference. Thank you for instilling perseverance and diligence in me, without which none of this would have been possible.

To my husband, Justin, thank you for being both my springboard and my soundboard throughout this PhD. Your love and support has been a much needed reminder of the truly important things in life.

Table of Contents

Abbreviations

AFM	Atomic Force Microscopy
Amp	Ampicillin
anti-His HRP	anti-6xHistidine horseradish peroxidase
BCCP	Biotin carboxyl carrier protein
BSA	Bovine serum albumin
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1- propanesulfonate
CO	Carbon Monoxide
CPR	Cytochrome P450 reductase
CuOOH	Cumene hydroperoxide
Cy5	Cyanine 5 dye
CYP450	Cytochrome P450
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinucleotide
FDA	Food and Drug Administration (USA)
FMN	Flavin mononucleotide
His-tag	polyhistidine-tag
IPTG	Isopropyl-β-D-galactoside
kDa	Kilo Daltons
LB	Luria Broth
LC	Liquid chromatography
MD	Molecular Dynamics
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry

NADP+	Nicotinamide dinucleotide phosphate
NADPH	Reduced nicotinamide dinucleotide phosphate
NHS	N-hydroxysuccinimidyl ester
NMR	nuclear magnetic resonance
NTA	Nickel Nitrilotriacetic Acid
OD	Optical density
PBS	Phosphate-buffered saline
PCA	principal component analysis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PMT	Photomultiplier tube
pw	Pathway
REMD	Random expulsion molecular dynamics
RFU	Relative fluorescence units
RMSD	Root mean square distances
rpm	Revolutions per minute
SDM	Site Directed Mutator
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SMD	steered molecular dynamics
SNP	Single nucleotide polymorphism
SRS	Substrate recognition site
Strep-HRP	Streptavidin conjugated to horse radish peroxidase
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
WT	Wild type

Chapter 1 Introduction

1.1 Nature's outstanding chemists

Found across all lineages of life, from bacteria to humans, Cytochrome P450 (CYP450) enzymes are arguably nature's most versatile catalysts, participating in key biosynthetic pathways and metabolising many diverse endogenous and exogenous substrates [1]. Their role in the synthesis of sterols, a key constituent of the plasma membrane, enabled the evolution of eukaryote organisms and their contribution to the synthesis of novel chemical compounds, exemplified by the on-going "chemical warfare" between plants and animals, has supported the diversification of eukaryotes [2]. In humans, CYP450s play a central role in steroid, fatty acid and xenobiotic metabolism.

First identified in 1962 and named for the unique absorbance soret peak observed at 450 nm for the carbon monoxide bound species [3-5], this superfamily of haem-thiolate mixed monooxygenases has been the subject of extensive research for over half a century. Research areas are vast and include the use of CYP450s in environmental bioremediation and as biocatalyst for a variety of biotechnology application including drug and antibiotic synthesis [6]; however, the majority of CYP450 research has focussed on their role in human drug metabolism. Amongst others, these enzymes are notorious in the pharmaceutical industry as they are often implicated in adverse drug reactions and therefore play a vital role in the drug development process [7].

1.2 Evolution of Cytochrome P450s

Not surprisingly, CYP450s are among the most rapidly evolving genes, consistent with organisms' need to metabolise new chemical toxins in an ever changing environment [8]. Functional diversity can be generated through a variety of genetic mechanisms, the most widely recognised being gene duplication [9]. Although CYP450s are most well-known for xenobiotic metabolism, cytochrome P450s are an ancient superfamily and it is likely that this function arose as an extension of their primary role in the metabolism of endogenous substrates [8]. Co-evolution between animals and plants is likely to have driven the evolution of CYP450s detoxifying function: as plants evolved CYP450s producing toxins that made them less desirable food sources, animals required new CYP450s capable of metabolising these new toxins [10].

1.3 Cytochrome P450 nomenclature

The cytochrome P450 superfamily consists of mainly microsomal and mitochondrial haem-thiolate proteins that catalyse oxidation, peroxidation or reduction reactions. There are now over 21 000 named CYP450 sequences found in animals, plants, fungi, protozoa, bacteria, archaea and more recently even viruses [11]. The number of genes per organism is highly variable and multiple cytochrome P450 genes can be expressed simultaneously.

The P450 Nomenclature Committee is responsible for the naming of CYP450 genes (Figure 1.1) and CYP450 genes are assigned into families and sub-families using a standard nomenclature system based on sequence identity, phylogenetic association and gene organisation [12]. The same nomenclature is used for DNA, mRNA and protein sequences; however the gene name is italicised while the mRNA and protein sequence names are shown in plain font. Defective genes that do not produce functional proteins, referred to as pseudogenes, are designated with an additional letter "P" at the end of the gene name e.g. *CYP1A8P*.

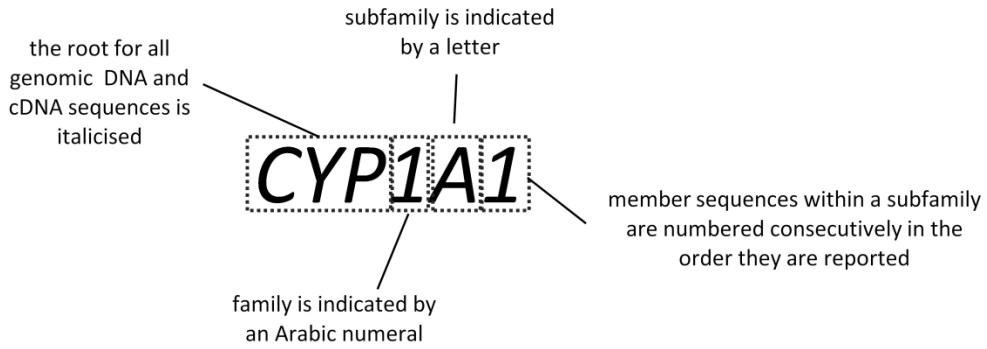


Figure 1.1 Naming CYP450 genes

In general a CYP family is made up of sequences with amino acid sequence identities greater than 40% and a subfamily with amino acid sequence identities of greater than 55%; sequences that are more than 97% identical are considered as alleles of the same gene unless there is sufficient evidence suggesting otherwise. Alleles are indicated by a star followed by a number according to the order in which they were identified. Sub-alleles with genetic variation in the non-coding regions of the genes are designated with an additional letter (e.g. CYP2C9*1C has the same coding sequence as the wild-type protein but has a mutation in the non-coding region of the gene.)

As CYP450s are ubiquitous, many more sequences will be identified as more organisms are sequenced. Blocks of numbers for CYP450 families have therefore been reserved for new families, allocated along taxonomic lines (rather than the phylogenetic relatedness); for example CYP1- CYP49 and CYP301-CYP499 are for CYP450 families found in animals.

1.4 Cytochrome P450 Chemistry

The primary role of CYP450 enzymes is to activate inert C-H bonds in organic molecules, creating functional groups suitable for conjugation to facilitate biotransformation. In general CYP450s achieved this by activating molecular oxygen, which in turn yields a highly reactive iron-oxo species that can attack the unreactive chemical sites, introducing a hydroxyl group into the molecule [8]. In xenobiotic metabolism this serves to convert unreactive lipophilic substrates into more reactive hydrophilic products, facilitated the elimination of these foreign compounds from the body. The haem group, bound within the hydrophobic centre of the protein and co-ordinated to a conserved cysteine residue, serves as the catalytic centre and consists of a large heterocyclic aromatic ring made up of four pyrrole rings linked together by methane bridges, collectively known as a porphryrin ring (Figure 1.2). The porphryrin ring has an iron atom in the centre that is responsible for the redox chemistry, cycling between the ferrous and ferric oxidation states during catalysis.

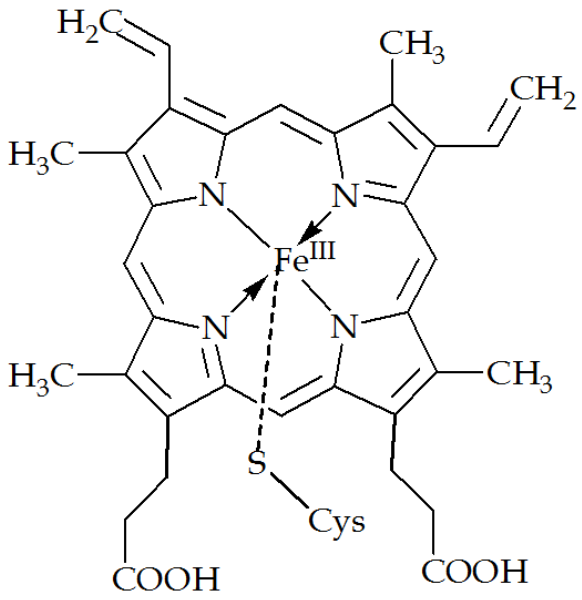


Figure 1.2 Chemical structure of the haem group coordinated to a catalytic cysteine residue.

The CYP450-catalysed hydroxylation reactions can be represented by the following equation

$$RH + O_2 + 2e^- + 2H^+ \xrightarrow{\textbf{CYP450}} ROH + H_2O$$

where RH is the substrate and ROH is the product. A redox partner protein, cytochrome P450 reductase (CPR) in the case of microsomal CYP450s, transfers two electrons sequentially from the electron donor NAD(P)H to the haem iron during the catalytic cycle:

$$NADPH \xrightarrow{\textbf{Redox partner}} NADP^+ + 2e^- + H^+$$

Other enzymes such as flavin-dependent monooxygenases and non-haem iron oxygenases can catalyse similar oxidation reactions. It is however CYP450s remarkable versatility in terms of substrate and reaction type, combined with their elegant ability to catalyse region- and stereo-selective oxidations, that makes them so important in physiology as well as attractive for biotechnological applications [13, 14]. While the above equation gives the stoichiometry of the reaction, elucidation of CYP450 catalytic cycle is required to understand the mechanistic details of the reaction.

1.4.1 The Catalytic Cycle

There has been much work, speculation and debate over the past 50 years surrounding the precise nature of the highly reactive intermediates responsible for the demanding chemistry catalysed by CYP450 enzymes. Spectroscopic techniques together with quantum and molecular mechanics approaches have given useful insight into the nature of the reactive intermediates [15]. The most highly sought after reactive intermediate in the CYP450 catalytic cycle, Compound I, was physically observed in 2006 [16] and in 2010 its capture and spectroscopic and kinetic characterization was finally reported [17].

Figure 1.3 depicts the catalytic cycle for CYP450-catalysed hydroxylations as it is currently understood [18, 19]. The cycle begins with the CYP450 in the low spin ferric (Fe^{3+}) state, with a water molecule bound to the 6th axial position of the haem iron **(1)**. Substrate binding causes structural changes that usually, but not always, lead to dissociation of the co-ordinated water

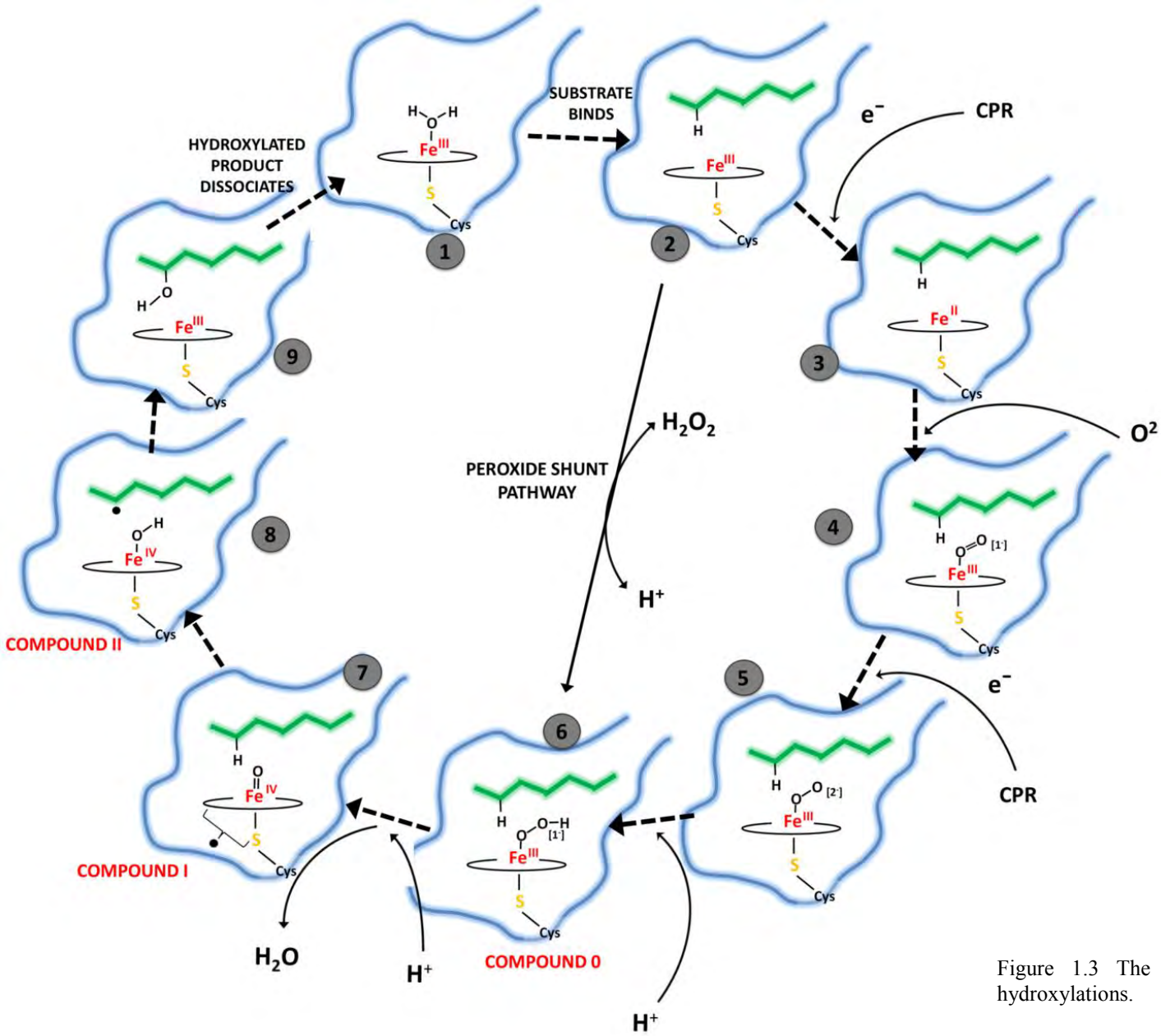


Figure 1.3 The catalytic cycle for CYP450-catalysed hydroxylations.

molecule accompanied by a high spin shift **(2)**. The substrate bound complex has a greater reduction potential and thus can now accept the first electron from its NAD(P)H dependent redox partner leading to a ferrous (Fe^{2+}) substrate-bound species **(3)**. The ferrous haem iron then binds dioxygen, forming a ferric superoxide complex (also referred to as an oxy-haem complex) **(4).** This complex accepts the second electron from its redox partner and is reduced to a ferric peroxo species **(5).** This species is then protonated at the distal oxygen generating a ferric hydroperoxo species **(6)**, which is a highly reactive intermediate referred to as Compound 0 (P450-0). Compound 0 can be artificially generated from the substrate bound ferric form of the enzyme **(2)** via the peroxide shunt pathway using a strong oxidising agent such as peroxide, circumventing the need for an electron donor to complete substrate metabolism. A second protonation at the distal oxygen cleaves the O-O bond yielding an iron(IV)-oxo radical species known as Compound I and a water molecule **(7)**. Compound I then abstracts a hydrogen atom from the substrate yielding Compound II and a substrate radical **(8)**. This is followed by a rapid rearrangement of the hydroxyl to the substrate radical to yield a hydroxylated product and a ferric enzyme **(9)**. The product then leaves the active site and a water molecule coordinates the 6^{th} axial position of the haem iron again leading to a resting ferric enzyme ready to start a new cycle **(1)**. The capture and characterisation of Compound I in 2010 was significant in confirming the "rebound mechanism" and the key role of compound II in C-H bond activation [17, 18].

1.4.2 Spectral properties

The haem group is a strong chromophore and consequently spectroscopic methods are very useful for studying CYP450s. Carbon monoxide (CO) coordinates strongly to the reduced (ferrous) iron of haem proteins and typically gives an absorption peak at 420 nm, as observed for proteins such as myoglobin [8]. CYP450 proteins were first identified as "novel microsomal carbon monoxide-binding pigments" in 1958 [20] and later named cytochrome P450s [3-5] owing to the unique absorption spectrum of the ferrous carbon monoxide-bound form, with a absorption maximum at 450 nm caused by the presence of the deprotonated sulphur on the cysteine axial ligand. When solubilised from microsomal membranes the CYP450 is relatively unstable and if this cysteine residue is protonated or replaced by a neighbouring amino acid such as a histidine the absorption maximum shifts to 420 nm [21]. This denatured or inactive form of the haem bound CYP450 is typically referred to as P420.

The spin state of the haem iron is the fundamental property affecting the haem spectrum. Cytochrome P450s exist as a mixture of low spin and high spin forms corresponding to the S = 1/2 and S = 5/2 spin states of the haem iron both in the presence and absence of substrates [22].

The spin state and hence the absorption peak of the ferric CYP450 is influenced by the ligation state of the haem iron. In the low spin state the haem iron is hexacoordinated and gives rise to an absorption peak at ~ 417 nm. In this substrate-free low spin state water usually serves as the 6^{th} distal ligand. Removal of the distal ligand leads to a high spin pentacoordinate state with an absorption peak at ~ 390 nm. In the hexacoordinated species the iron atom sits within the haem plane while in the pentacoordinated species the iron atom sits about 0.5 Å out of plane.

Substrate binding can be measured by monitoring the difference spectrum between the bound and unbound protein. In the absence of substrate the low spin state is dominant [22]. A type I difference spectrum, with a peak at 385-390nm and a trough at 420 nm, is observed when substrate binding displaces the distal water molecule causing a shift in the equilibrium towards the high spin state accompanied by an increase in redox potential facilitating reduction of the ferric iron by the redox partner [8]. In contrast, a type II difference spectrum, with a peak at 425-435nm and a trough at 390-405nm, is observed when a substrate (or inhibitor) coordinates to the iron more strongly than the water molecule it displaced. This binding spectrum is typical of molecules containing nitrogen heterocycles, where a lone pair of electrons is free to coordinate to the iron (e.g. nicotine). While substrate binding usually causes a spectral shift, there are instances where substrates bind within the active site without affecting the spin state of the haem iron [19]; this occurs when substrates bind at a periphery site or in an orientation that does not displace the distal water molecule.

The ferrous CYP450, with or without substrate, exists almost exclusively in the high spin state. Other factors such as temperature [22], ionic strength [23] and hydrostatic pressure [24] also affect the spin-state equilibrium and therefore serve as useful methods for studying CYP450s.

1.4.3 CYP450-catalysed reactions

CYP450 catalyse the metabolism of an enormous number and diverse range of substrates; however most CYP450 oxidation reactions can be broadly categorised as hydroxylations,

epoxidations or heteroatom oxidations [8] (Figure 1.4). CYP450s also show reduction activity under anaerobic conditions; reductions by the ferrous CYP450 complex frequently lead to the formation of harmful free radicals [25].

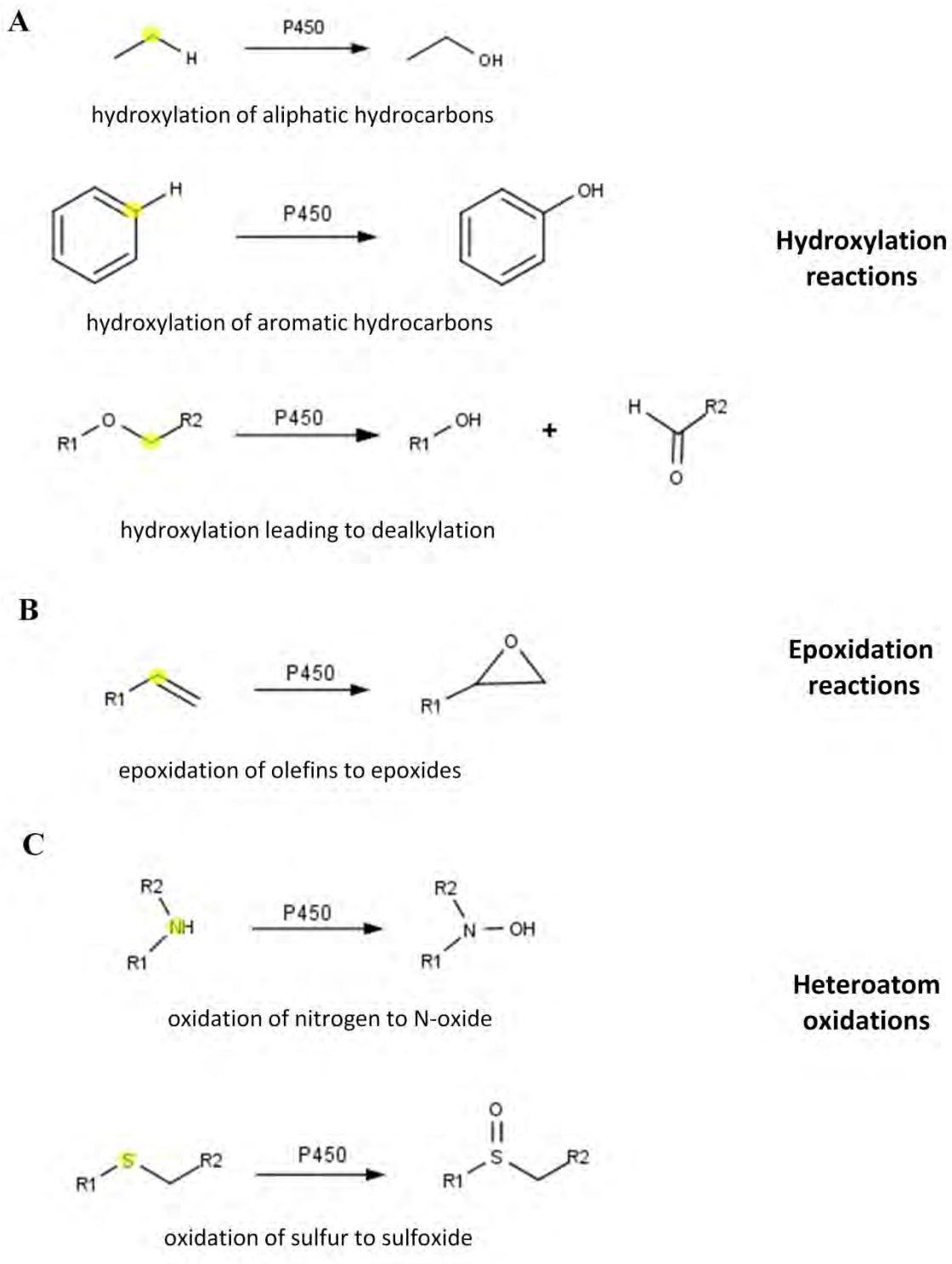


Figure 1.4 Common cytochrome P450-catalysed reactions. A) Hydroxylation, B) Epoxidation and C) Heteroatom oxidation reactions. The site of oxidation is indicated in yellow.

Hydroxylation

The hydroxylation of carbon atoms in aliphatic and polycyclic aromatic hydrocarbons is by far the most common and best understood CYP450-catalysed reaction and is central to CYP450-mediated drug metabolism [8] (Figure 1.4 A). The site of oxygen insertion is determined by the energy required to break the C-H bond as well as steric effects and the position of the substrate within the active site [26]. Tertiary C-H bonds are easier to break than primary C-H bonds and as a result hydroxylation of tertiary Carbons is far more common than hydroxylation of terminal methyl groups. However if the C-H bond is weakened by an adjacent heteroatom or halogen then hydroxylation is highly favoured, as seen for a methyl group attached to an aromatic ring. Hydroxylation of a carbon adjacent to a heteratom such as oxygen, nitrogen or sulphur leads to heteroatom dealkylation. Harmful reactive metabolites capable of acetylating proteins can however be produced if hydroxylation occurs at a carbon atom attached to more than one halogen atom [8, 27]. Bio-activation of thiophenes and sulphur containing polycyclic aromatic rings can also have toxicological implications, due to the formation of reactive metabolites capable of forming DNA abducts [8].

Epoxidation

CYP450 are capable of oxidising olefins to their corresponding epoxides (Figure 1.4 B). CYP450-mediated epoxidations of some endogenous substrates are important for the synthesis of physiologically important molecules such as arachidonic acid. However epoxidation of foreign compounds can also lead to toxic, mutagenic and teragenic metabolites [28]. Aberrant epoxidation reactions can also lead to mechanism-based inactivation on the CYP450 enzyme caused by a terminal olefin carbon associating with a pyrrole nitrogen on the CYP450 haem group [8].

Heteroatom oxidations

Cytochromes P450s can also oxidise heteroatoms within a substrate; oxidation of heteroatoms relies on the availability of lone electron pairs to which oxygen can be transferred. More electronegative atoms, such as oxygen, are less susceptible to oxidation while heteroatoms such as nitrogen and sulfur are readily oxidised to N-oxide and sulfoxide metabolites (Figure 1.4 C).

While these categories give a simplified overview of the main CYP450-catalysed reaction-types, CYP450s catalyse many variations of these reactions as well as many other more complex and unusual reactions. Guengerich and Munro recently reviewed unusual CYP450 reactions, illustrating the incredible diversity and versatility of these enzymes [29].

1.5 Cytochrome P450 structure

1.5.1 The highly conserved CYP450 fold

Camphor-metabolising CYP101 (P450cam) from *Pseudomonas putida* was the first CYP450 to have its crystal structure solved [30] and extensive studies on this enzyme, combined with other subsequently solved bacterial CYP450 structures, provided the initial framework for our understanding of structure-function relationships in CYP450 enzymes. Today there are over 500 CYP450 structures in the Protein Data Bank (PDB), including nearly 100 mammalian structures [31]. As predicted by Hasemann and co-workers [32], despite very low sequence identities (even below 20%), the CYP450 fold is highly conserved across prokaryote and eukaryote structures. CYP450s have a globular three-dimensional fold typically made up of 4 to 5 β-sheets and thirteen α-helices surrounding a large buried hydrophobic active site (Figure 1.5). The protein comprises a relatively flexible domain on the distal side of the protein, primarily responsible for substrate recognition and binding; a more rigid haem binding core; and a domain with intermediate flexibility on the proximal side of the protein that provides a binding site for the redox partner- responsible for transferring electrons to the haem iron during the catalytic cycle- in close proximately to the catalytic centre [33]. The haem binding regions are generally conserved between CYP450s while the substrate recognition regions are more variable [34]. There are a number of key conserved features found in all CYP450s, namely the I-helix catalytic groove, the K-helix core stabilising motif, the meander region and the Cys-pocket.

The *I-helix catalytic groove* is the conserved consensus motif (G/A)GX(D/E)T, located in the middle of the I-helix [30]. This region plays an important role in proton transport [35, 36] and is responsible for the formation of the oxygen binding pocket [37]. The *K-helix core stabilising motif* is the invariant EXXR motif within the K-helix that interacts with the meander region [32]. The *meander* is found between the K'-helix and the L-helix and is highly conserved in structure despite the lack of any secondary structure formation. The absolutely conserved Asp and Arg

residues in the core stabilising motif interact with an Arg/His residue in the meander region, forming a set of salt bridges known as the *ERR triad*. The *Cys-pocket* refers to the sequence motif FXXGXXXCXG that forms a β-bulge around the absolutely conserved catalytic Cys residue that coordinates to the haem iron.

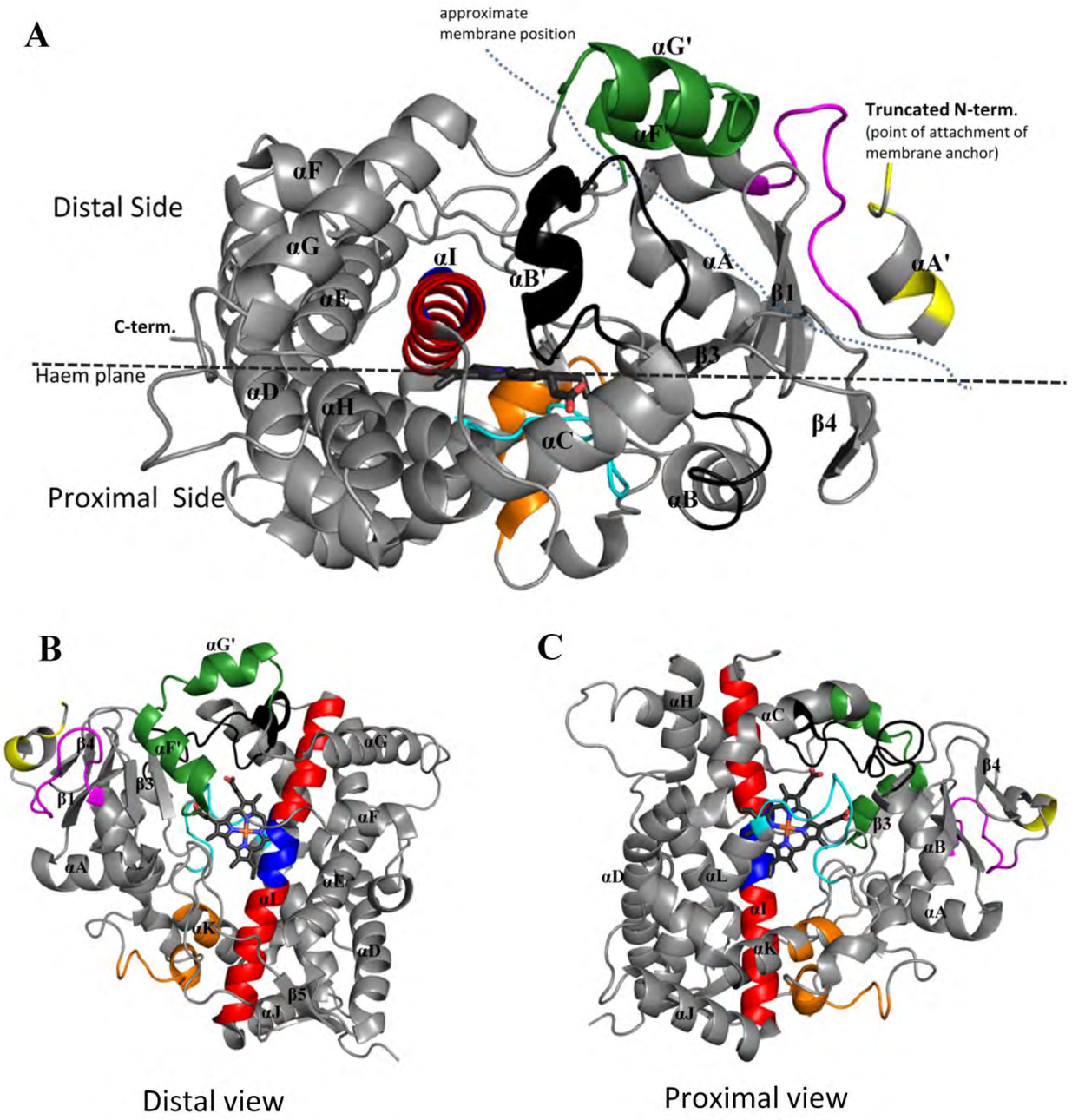


Figure 1.5 The CYP450 three dimensional structure
The CYP450 structure is shown in grey (CYP3A4, PDB structure 1TQN), is shown in grey with the haem group in black. Secondary structure elements, α-helices and β-strands are labelled using the nomenclature of Poulos et al [37]. Panel A, B and C show a side view, distal view and proximal view respectively. The I-helix is shown in red, the B-C loop in black and the F-G loop in green to aid orientation. The positions of conserve regions are indicated as follows: the halting signal (yellow); the proline-rich motif (magenta); the I-helix catalytic groove (blue); the core stabilising motif consisting of the meander stabilised by the ERR triad (orange); the cys-pocket (cyan).

1.5.2 Membrane binding

Whilst all prokaryotic CYP450s are soluble and located in the cytosol, CYP450s in eukaryotes are almost exclusively membrane bound, embedded in either the endoplasmic reticulum or the mitochondria membrane. Consequently eukaryote CYP450s have an additional membrane anchoring region that is absent in the soluble bacterial CYP450s.

CYP450s embedded in the endoplasmic reticulum, referred to as microsomal CYP450s, are the most common CYP450s in mammals and are found predominantly in the liver, the primary site of xenobiotic metabolism. The hydrophobic signal-anchor sequence is a 20 to 25 amino acids stretch near the N-terminus that is embedded in the membrane. It is usually preceded by an acidic amino acid residue that targets the protein to the endoplasmic reticulum [38]. The signal-anchor sequence is flanked on the C-terminal end by a series of basic residues referred to as the halting-signal that ensure the protein is orientated correctly in the membrane [38]. The halting-signal is followed by a proline rich motif, PPGPXPXPXXGN, which is important for protein folding rather than catalytic activity [39-41].

Determining the structure of membrane bound proteins in notoriously difficult; the full-length CYP450 is highly unstable when removed from its native membrane environment and consequently the N-terminal membrane binding region is truncated in CYP450 crystal structures and in many cases mutations are introduced to enhance the protein solubility [42]. As a result, the position of CYP450s in the membrane and the effect of the membrane on protein conformation, flexibility and substrate access remain largely unexplained. Experimental studies indicate that the membrane anchor domain forms a single transmembrane helix and that hydrophobic regions within the CYP450 fold are also embedded within the membrane [43-45]. Residues interacting with the membrane have been identified by site-specific antibody [46, 47], fluorescence quenching [48] and atomic force microscopy [49] experiments. Studies using molecular dynamics to model the orientation of microsomal human CYP2C9 within the membrane have revealed that the F-G loop region is partially buried and that the membrane stabilises an open tunnel leading to the buried active site that is likely to facilitate the access of liposoluble substrates via the membrane; the water soluble products are thought to egress through another tunnel pointing towards the water phase [50, 51] (Figure 1.6). The orientation of CYP2C9 in these models is also predicted to be favourable for electron transport from the redox

partner protein (NADPH dependent cytochrome P450 reductase (CPR)) and these models are largely consistent with data from experimental studies.

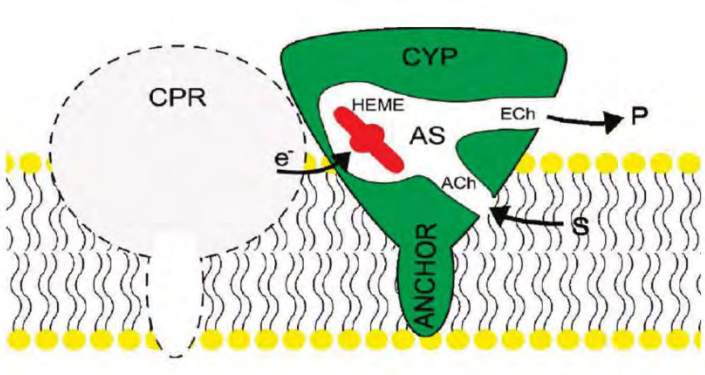


Figure 1.6 This diagram shows CYP450 and CPR embedded in a membrane. The CYP450 anchor region and part of the catalytic domain are buried in the membrane and the CYP450 is orientated in a way that would allow CPR to bind to the proximal face of the protein, facilitating electron transport to the haem iron. The hydrophobic substrate (S) accumulates in the membrane and enters the CYP450 active site (AS) via an access channel (ACh). It is oxidised to a more hydrophilic product (P); the product then leaves the active site via an egress channel (ECh) and enters the cytosol. This figure is from Berka *et al*, 2011 [50].

Recently, solid-state nuclear magnetic resonance (NMR) was used to determine a high-resolution structure of microsomal rabbit CYP2B4 embedded in an artificial membrane of temperature resistant bicelles [52]. This NMR model revealed a helical transmembrane region tilted at ~17° from the lipid bilayer normal. This study did not however give insight into associations between the catalytic domain and the bilayer or conformational changes associated with membrane binding as the experiment was set up to selectively detect signal from the rigid N-terminal transmembrane region and did not provide details on the catalytic domain; instead the N-terminal truncated CYP2B4 crystal structure was used to model the catalytic domain.

1.5.3 Substrate access and recognition regions

The substrate binding regions are found on the distal side of the protein and correspond to the less conserved or more variable regions of the protein. There are 6 recognised substrate recognition sites called named SRS 1 to 6 [53]. The variable regions of the protein that make up these SRS regions include helices A, B, B', F and G and adjacent loops: the B-C loop forms SRS 1; the F and G helices and the F-G loop forms a lid over the active site cavity forming part of a substrate access channel and SRS2 and SRS3; the loop region N-terminal of strand 1 of β-sheet 3 lines the active site and forms SRS5; and the loop region (or β-turn) between strands 1 and 2 in

β-sheet 5 protrudes into the active site forming SRS6 [34]. The centre of the I-helix, which is the only highly conserved substrate recognition region in CYP450s, makes up SRS4.

While some CYP450s are remarkably substrate specific, metabolising only a small set of very similar molecules, other CYP450s, particularly those responsible for drug metabolism, have broad substrate specificities metabolising a vast range of molecules with very different geometries and chemical properties. Although substrate specificity is generally thought of as a property inherently linked to the amino acid composition and geometry of the active site [33], structures of mammalian CYP450 reveal a large active site capable of binding more than one substrate or inhibitor molecule at once [54-57]. Additionally there are structures with a single substrate molecule bound in a non-productive position within the active site [58, 59] further suggesting that substrate binding does not conform to the traditional "lock and key" model, but is rather a multi-step process involving a flexible active site. CYP450s ability to adopt different structural conformations enables an individual CYP450 to accommodate molecules of different shapes and sizes and as a result, static structures of the active site provide limited insights into substrate specificity [33]. There are two schools of thought pertaining to conformational changes and substrate binding: 1) "induced-fit" where the substrate binding regions adjust their shape to accommodate a substrate and 2) "conformation selection" where the enzyme pre-exists in multiple conformations. Atypical kinetics has repeatedly been observed for a number of CYP450 isoform and there is an on-going debate as to whether it is a result of multiple molecules binding to the enzyme and/or the presence of kinetically distinct conformers [60-63], but either way it can likely be attributed to the high flexibility of the active site.

The role of CYP450 flexibility in substrate access and binding has been examined in numerous studies. Temperature β factors from crystallographic data have been used to infer regions of flexibility in CYP450 structures [42]. Comparing multiple crystal structure of the same or related proteins can also provide useful information on the flexibility of CYP450s [54, 64-66]. Lee *et al* compared 30 structures of bacterial P450cam bound to range of different substrates; this study revealed three distinct conformational clusters that could be attributed to the multi-step closure of the enzyme around the substrate that proceeds through a well-defined intermediate [65]. There are now a number of tools available for rapidly comparing large numbers of crystal structures enabling the inference of protein dynamics from large ensembles of static structures (e.g. ProDy [67], Polyphony (Pitt, unpublished) and as more mammalian CYP450 structures are

solved these approaches will become increasingly useful for studying CYP450 flexibility. Spectroscopic methods such as NMR [68-71], mass spectrometry [72, 73], high pressure absorption spectroscopy [74-77] and Raman spectroscopy [76, 78, 79] as well as molecular dynamic simulations [51, 76, 80] have been used extensively to study the role of CYP450 flexibility in substrate access and binding.

Classical molecular dynamics simulations typically use a crystal structure as a starting point. To prepare the structure for modelling in solution, hydrogen atoms are added, unresolved (missing) regions are modelled, protonation states of titratable amino acids are defined and the structure is subjected to energy minimisation using empirical force fields. MD simulations normally attempt to mimic physiological or *in vitro* conditions using explicit solvent, counterions and a defined temperature and pressure (1 atm). Classical modular dynamics simulations typically generate a series of conformations over a timescale of tens-of-nanoseconds, with femtosecond resolution. Large conformational changes cannot normally be detected using classical MD due to the limited timescales; consequently a number other MD techniques have been developed to observed larger motions including simulations at elevated temperature, replica exchange MD and adaptive acceleration MD. In addition MD techniques such as steered MD [81] and random expulsion MD [82] have been developed that simulate molecules exiting the active site to gain insight into substrate access and product egress routes. MD simulations for CYP450s are not straight forward due to the redox and spin states exhibited by the haem bound cofactor that cannot easily be represented by classical force field models. This has posed a challenge for modelling CYP450s and much work has gone into developing appropriate force field parameters, both for the pentacoordinate [83-85] and hexacoordinate haem states [76, 80, 86].

The most flexible parts of the protein are typically in loop regions located on the distal side of the protein, which is relatively flexible compared to the haem core. This is fitting since the haem core ensures the high regio- and stereo-specificity of metabolism while the flexible distal regions responsible for substrate recognition allow for broad substrate specificity [33, 42]. Not surprisingly, the broader a CYP450s substrate specificity, the more flexible the distal regions [76, 80]. MD simulations performed by Hendrychova *et al* on human CYP2A6, CYP2C9, CYP2D6 and CYP3A4 and rabbit CYP2B4 in the absence of substrate showed that the active site volumes fluctuate by more than 50%, predominantly due to the movement of main chains with only a small number of side chains moving independently during the simulation [87]. These

simulations also show that the bottle necks of CYP450 access/egress channels fluctuate around 1.5 Å suggesting that these channels never fully open in the absence of substrate but rather that substrates induce conformational changes as they move along a pathway into the active site. Hydration of the active site also differed between isoforms; CYP2E1 was the least hydrated, consistent with its preference for hydrophobic substrates.

Human CYP3A4, the most abundant CYP450 in the human liver, has the broadest substrate specificity and is responsible for the metabolism of around 50% of all drugs. It is the most flexible CYP450 [76, 80] and substrate binding to CYP3A4 has been extensively studied yet is still poorly understood. Sevrioukova *et al* recently reviewed the latest insights into CYP3A4s molecular mechanism [88]. In short, based on extensive studies using a range of techniques, it is now believed that substrates are not fixed in one position and may dissociate and rebind at different stages in catalysis. Following this, the binding affinity and metabolism of one substrate can be affected by another molecule (either the same or different) present in the active site. Substrate binding is thus viewed as a multistep process with substrates potentially binding at a periphery site, important for substrate recognition, before moving into the active site. Zharkova *et al* have identified subfamily specific patterns of residues located at the entrance of the binding site of drug metabolising CYP450 and suggest that these residues form a periphery binding responsible for substrate recognition [89].

1.6 Cytochrome P450's redox partners

During the catalytic cycle CYP450s source electrons from their redox partners via NAD(P)H to drive catalysis: small soluble iron-sulfur proteins such as adrenodoxin and putidaredoxin, generally serve as redox partners for bacterial and mitochrondria CYP450s, while the membrane bound NADPH dependant Cytochrome P450 reductase (CPR) serves as a redox partner for microsomal CYP450s [90]. CPR is co-immobilised with CYP450s in the endoplasmic reticulum membrane and is responsible for electron transfer to all known microsomal CYP450s.

CPR is made up of four domains: the FMN-binding, connecting, FAD-binding and NADPH-binding domains [91] (Figure 1.7). NADPH binds to the NADPH binding domain of CPR and transfers electrons to the flavin adenine dinucleotide (FAD) bound to the adjacent FAD-binding domain; FAD then transfers the electrons to the flavin mononucleotide (FMN) bound to the

FMN-binding domain and finally the elections are transferred sequentially from FMN to the haem iron of a substrate-bound CYP450. The connecting domain orientates the other two domains, bringing the flavins together for electron transfer. The flexible hinge region enables the FMN-binding domain to re-orientate itself so that the region containing the bound FMN can interact with the CYP450.

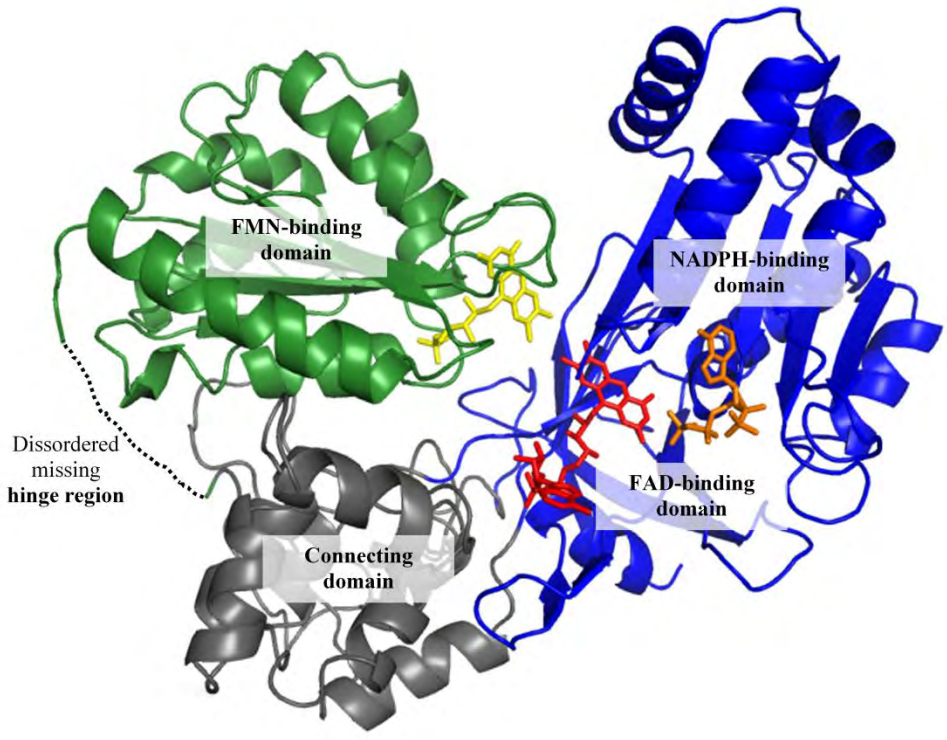


Figure 1.7 Structure of microsomal NADPH cytochrome P450 reductase (PDB code 1AMO). The NADPH- and FAD-binding domains are shown in blue, the connecting domain in grey and the FMN-binding domain in green. The bound co-factors are shown as sticks with $NADP^+$ in orange, FAD in red and FMN in yellow. The missing hinge region connecting the FMN domain to the connecting domain is indicated by a dotted line.

The interaction between CPR and CYP450 is transient and unstable, with $K_d \sim 0.05$ μM [92]. A number of studies have contributed to elucidating the CYP450-CPR interface and have identified clusters of basic residues on the proximal surface of CYP450s that interact with exposed acidic clusters on the CPR FMN domain surface [93-98]. The bacterial redox system cytochrome P450-BM-3 - a 119kDa self-sufficient enzyme from *Bacillus megaterium* - consists of a CYP450 haem domain linked to a FMN/FAD-containing CYP450 reductase domain via a single polypeptide chain, closely resembling a "fused" mammalian CYP450 redox system [90, 99]. The structure of P450-BM3 has been used as a template to model the interaction between a microsomal CYP2B4 and CPR, confirming previously identified interfacial residues and identifying additional ionic-charge clusters that mediate the interaction [98].

In 1992, Voznesensky and Schenkman reported that increased ionic strength increased the rate of first electron transfer to CYP2B4; this appeared inconsistent with reports suggesting that charged pairing stabilises the CPR-CYP450 complex [100]. Davydov *et al* however later clarified this apparent discrepancy by showing that while the affinity of the complex decreased with increased ionic strength, consistent with charge pairing, there was an increase in the rate of electron transfer through the complexes that were formed, leading to overall elevation of CYP2B4 reduction [96], as was observed by Voznesensky and Schenkman.

Early studies suggested that hydrophobic interactions involving the hydrophobic N-terminal peptides are also important for complex formation and showed that the N-terminal truncated CPR was incapable of transferring electrons to the CYP450 [101]. Subsequent studies using truncated CPR and CYP2E1 reconstituted with phospholipids, however, showed that removal of the N-terminal did not prevent CPR-CYP450 complex formation, CYP450 reduction or affect the systems sensitivity to changes in ionic strength [102]. Overall the picture that emerges is that due to the transient, weak nature of the interaction, CYP450 and CPR need to be co-immobilised for efficient CYP450 activity in order to increase the probability of a productive electron transfer encounter occurring at the appropriate points in the CYP450 catalytic cycle. CPR mediated CYP450 activity of recombinant proteins is only observed in solution in the presence of a phospholipid component or when very high protein concentrations are used (> 4μM) [103].

Phospholipids make up a large part of the microsomal membrane, acting as a matrix for the interaction between CPR and CYP450. The phospholipid composition can affect CPR-CYP450 interactions and electron transport [104-107] and can also act as effector molecules, modulating CYP450 activity directly [104].

1.7 Membrane organisation of multiple CYP450s and CPR

CYP450 and CPR, in the presence of phospholipids, form the primary components of the microsomal CYP450 monooxygenase system [108], however other proteins including cytochrome b5 and NADH-cytochrome b5 also interact with the system and can affect CYP450 activity. Although CPR and CYP450 form a 1:1 complex [92, 109], CYP450 levels exceed CPR levels by 10 to 20 fold, depending on their induction status [110]. Thus there is competition between CYP450s for CPR binding and isoforms with higher binding affinity for CPR could out-

compete isoforms with lower affinity. This suggests that other mechanisms must be in place to ensure low affinity isoforms are not rendered inactive as a result [111]. There is now substantial evidence suggesting that both homomeric and heteromeric CYP450-CYP450 interactions within membranes play a complex yet important role in modulating drug metabolism, both by affecting interactions with CPR as well as directly impacting CYP450 function [111, 112].

Early on it was postulated that several CYP450 molecules cluster around each reductase molecule and that lateral motion is required for some CYP450 molecules to interact with a reductase molecule [113]. Rotational diffusion studies in both microsomes and reconstituted systems have shown that a proportion of the CYP450 protein is immobile [114-116] as a result of CYP450 aggregation and that increased lipid content increases the protein mobility. In addition, CPR-CYP450 complex formation disrupts CYP450 aggregates and increases protein mobility [115, 117].

Both full length CYP450 proteins and N-terminal truncated CYP450 proteins tend to aggregate in solution to varying degrees, depending on the construct. Studies show that CYP2B4 exists as hexamers in solution [118-120] and CYP1A2 is reported to exist as hexamers or heptamers [119, 121, 122] or as a mixture of tetramers and 40-mers [123]. More recent studies have identified multiple, kinetically distinct CYP3A4 and CYP2B6 conformers in solution that are believed to be a result of CYP450 oligomers [124-128].

Reed *et al* recently reviewed the effects of heteromeric CYP450-CYP450 complex formation on CYP450 function [112]. Based on experimental evidence from studies that compare substrate metabolism by an individual CYP450 with substrate metabolism by a mixture of different CYP450 isoforms, they describe three interaction models: 1) *Simple competition between two CYP450s for CPR binding*. In this model the CYP450s have different binding affinities for CPR but there are no interactions between the different CYP450 that have a functional effect e.g. [129-131]. 2) *Formation of a CYP450-CYP450 complex that affects reductase binding*. For example Backes and co-workers showed that CYP1A2 and CYP2B4 form a ternary complex in both reconstituted systems and microsomes and that the CYP1A2 in complex with CYP2B4 has a higher affinity for the reductase than CYP1A2 alone [132, 133]. 3) *CYP450:CYP450 complexes that alter the rate of catalysis*. There is plenty of evidence suggesting that CYP450-CYP450 interactions can directly alter the rate of substrate turnover, with the presence of one

isoform inducing or inhibiting the activity of another isoform [134-139]. As with homo-oligomers, hetero-oligomer formation is likely to affect the conformational flexibility of individual CYP450 molecules within the oligomer, in turn affecting catalytic activity in a substrate specific manner. Apart from evidence from functional assays, there is also direct physical experimental evidence from cross-linking, fluorescent studies, fluorescence resonance energy transfer (FRET) and mutagenesis experiments for the formation of specific homomeric and heteromeric CYP450-CYP450 complexes in the membrane [112].

In summary, CYP450 exists in multienzyme complexes involving specific and nonspecific interactions that modulate CYP450 metabolism. Currently, however, little is known about the cellular conditions that control complex formation, the regions of the protein important for specific CYP450-CYP450 interactions or the mechanistic effects of these interactions [112].

1.8 Factors regulating CYP450 expression and function

Cytochrome P450 regulation occurs predominantly at the transcriptional level and CYP450 expression can be induced by both endogenous and exogenous compounds. Expression levels vary according to tissue-type, developmental stage, age, sex and disease state and while the liver is the primary site of drug metabolism, many CYP450s are also expressed in other tissue types such as intestine, skin, brain, lung and placenta, with endocrine factors (steroids) being largely responsible for the tissue-specific and sex-specific expression of CYP450s.

The expression of many drug metabolising enzymes is regulated by transcription factors that are activated by exogenous compounds. The aromatic hydrocarbon receptor (AhR) protein is a well-known example: when a polycyclic aromatic hydrocarbon ligand binds to AhR, its chaperones dissociate and it translocates to the nucleus, dimerising with the AhR nuclear translocator and leading to increased transcription of CYP450 genes [140]. Consequently, CYP450 expression is generally increased rapidly in response to the presence of xenobiotics, enabling their efficient breakdown.

1.9 Cytochrome P450s in human drug metabolism

1.9.1 Overview of drug metabolism

Xenobiotic metabolism refers to the metabolic pathways that modify the chemical structure of foreign compounds, including drugs and toxins, usually converting lipid-soluble substances to more water-soluble products that can readily be excreted. While this process generally serves to detoxify foreign compounds, it can also lead to the formation of reactive intermediates that can exhibit toxic effects. In some cases, drugs are deliberately administrated as inactive prodrugs that are activated by enzymatic modification.

Physical barriers and drug transport proteins such as P-glycoprotein serve as the body's first line of defence against foreign compounds. The cell membrane acts as a hydrophobic permeability barrier and, while useful polar molecules enter the cell through selective transport proteins, the hydrophobic nature of the membrane prevents unwanted hydrophilic molecules from entering the cell. Thus hydrophilic drugs can only enter a cell if they are recognised by a transport protein [141]. In contrast hydrophobic compounds can readily diffuse across the cell membrane and consequently most clinical drugs are lipophilic and need to be metabolised before they can be excreted.

P-glycoprotein is an ATP-dependent efflux pump with broad substrate specificity that pumps foreign compounds out of cells. P-glycoproteins are expressed predominately in tissues with excretory functions, as well as in important tissue-blood barriers. In the gastrointestinal tract it reduces drug absorption by transporting passively absorbed drugs back into the lumen; in liver cells and the proximal tubular of the kidney it pumps drugs into the bile ducts and urine-conducting ducts respectively; and in the blood–brain barrier it prevents the accumulation of harmful substances in the brain [142]. Drug transport proteins thus in some ways improve the efficacy of drug metabolism by preventing the cellular drug metabolising enzymes being overwhelmed with high drug concentrations [143].

Drug metabolism can be divided into 3 phases [144]: Phase I) reactive or polar groups are introduced into the substrate, providing sites for conjugation; Phase II) polar species such as glutathione, sulphate, acetate or glucuronic acid are conjugated to the reactive metabolites - these

anionic groups prepare the molecule for active transport out of the cell; Phase III) conjugates may be metabolised further and are excreted into the extracellular media via a variety of efflux pumps belonging to the multidrug resistance protein family that recognise the anionic groups on the conjugates [145]. Cytochrome P450s, flavin monooxygenases, dehydrogenases and esterases are responsible for phase I metabolism, while sulfotransferases, UDP-glucuronosyl transferases, N-acetyltransferases, glutathione S-transferases and methyltransferases are the main enzymes responsible for phase II metabolism [146].

1.9.2 Human CYP450 drug metabolising enzymes

In humans there are 57 functional CYP450 genes and 58 pseudogenes, divided into 18 families and 44 subfamilies [12]. While the CYP450s in most families are primarily involved in biosynthesis and metabolism of endogenous compounds, CYP450s in families 1-3 are responsible for the phase I metabolism of 70-80% of all pharmaceutical drugs, as well as the detoxification of many other xenobiotic compounds that enter the body [147]. Figure 1.8 shows the proportion of pharmaceutical drugs metabolised by each of the major human drug metabolising isoforms found in the liver. CYP450s expressed in extra-hepatic tissues only play a minor role in the total drug clearance, however they are very important for the localised metabolism of drugs in peripheral tissues and can contribute significantly to localised drug efficacy, detoxification and adverse effects [148]. The major human drug metabolising enzymes in CYP families 1-3 are listed in Table 1.1 and discussed in more detail below.

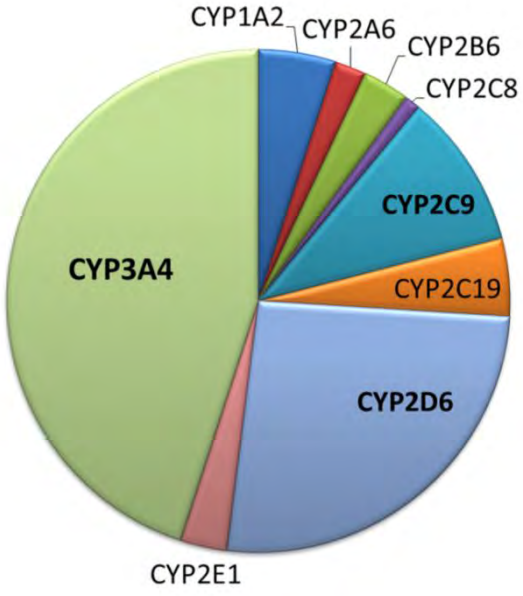


Figure 1.8 The proportion of pharmaceutical drugs metabolised by the major human drug metabolising CYP450 enzymes in the liver: CYP1A2 (5%), CYP2A6 (2%), CYP2B6 (2-4%), CYP2C8 (1%), CYP2C9 (10%), CYP2C19 (5%), CYP2D6 (20-30%), CYP2E1 (2-4%) and CYP3A4 (40-45%) [149].

Table 1.1 The major human hepatic drug metabolising CYP450s

CYP450	Main role in xenobiotic metabolism	Substrate specificity
1A2	Drugs, carcinogens & oestrogens	Planar, aromatic, polyaromatic and heterocyclic amides and amines
2A6	Drugs, carcinogens & nicotine	Non-polar low MW molecules- usually with 2 hydrogen bond acceptors (Including ketones and nitroamines)
2B6	Drugs & environmental chemicals	Neutral or weakly basic, mainly lipophilic non-planar molecules with 1 or 2 hydrogen bond acceptors (including anaesthetics, herbicides and insecticides)
2C8	Some drugs	Large and weakly acidic molecules (including antimalarials and antidiabetics)
2C9	Drugs	Weakly acidic molecules with a hydrogen bond acceptor (including most NSAIDs)
2C19	Drugs	Neutral or weakly basic molecules or amides with 2 or 3 hydrogen bond acceptors (including most proton pump inhibitors (PPIs))
2D6	Drugs	Basic molecules with a basic nitrogen 4-7 Å from the site of metabolism (includes plant alkaloids and antidepressants)
2E1	Carcinogens, solvents, few drugs	Small neutral, hydrophilic planar molecules as well as long chain fatty acids (includes aliphatic alcohols and halogenated alkanes)
3A4	Drugs & carcinogens	Very board - Large lipophilic molecules with very diverse structures (includes > 50% of clinical drugs)

Table adapted from [147, 150]

CYP1 family

There are three functional genes in the CYP1 family: CYP1A1, CYP1A2 and CYP1B1. While CYP1A2 is constitutively expressed at high levels in the liver, representing 4-16% of the total CYP450 liver content [147], CYP1A1 and CYP1B1 are primarily expressed extra-hepatically [151-156]. CYP1 expression is primarily regulated by AhR that upon activation by polycyclic aromatic hydrocarbons binds to enhancer elements flanking the CYP1A1, CYP1A2 and CYP1B1 genes, driving transcription [157].

CYP1 enzymes show overlapping catalytic activities, oxidising many polycyclic aromatic hydrocarbons and other aromatic compounds. CYP1A1 and CYP1A2 have a sequence identity of over 70% but still show distinct substrate preferences: CYP1A1 prefers planar aromatic hydrocarbons while CYP1A2 prefers aromatic amines and heterocyclic compounds [147]. Endogenous substrates for CYP1 enzymes include oestrogens, retinoic acid, arachidonic acid, melatonin, prostaglandins and uroporphyrinogen [158].

The CYP1 proteins are notorious for their role in bioactivating carcinogens including benzo(a)pyrene, arylarenes, nitroarenes, arylamines found in charred food, cigarette smoke and products from industrial combustion [8, 147]. The resultant highly reactive metabolites are capable of forming DNA abducts and have been implicated in cancer formation.

CYP1A2

CYP1A2 also plays an important role in the hepatic metabolism of numerous pharmaceutical drugs including analgesics, antipyretics, cardiovascular drugs, antipsychotics, antidepressants and anti-inflammatory drugs [147]. CYP1A2 is subject to reversible and irreversible inhibition by a number of drugs, natural substances and other compounds that bind within the active site of the enzyme [159]; as a result, drug treatments involving CYP1A2 substrates are sensitive to drug-drug interactions that can lead to adverse drug effects in patients. α-Naphthoflavone and fluvoxamine are two of the most potent and well known CYP1A2 inhibitors [160-162] but other CYP1A2 inhibitors include oral contraceptives, cimetidine and furafylline [147]. In addition to this, exposure to CYP1A2 inducers can lead to increased CYP1A2 expression and increased drug clearance, resulting in reduced drug efficacy or toxicity in the case of pro-drugs.

A crystal structure of CYP1A2 with inhibitor α-naphthoflavone bound in the active site reveals a compact, relatively narrow and planar active site that is highly suited to the shape and size of its substrates [163]. The shape of the active site is similar to that observed in the more recently solved CYP1B1 [164] and CYP1A1 [165] crystal structures but is distinct from the active site architecture observed in drug metabolising CYP2 and CYP3 structures.

CYP2 family

The CYP2 family consists of 16 functional genes (CYP2A6, CYP2A7, CYP2A13, CYP2B6, CYP2C8, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, CYP2F1, CYP2J2, CYP2R1, CYP2S1, CYP2U1 and CYP2W1), divided into 13 subfamilies [12]. Isoforms CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6 and CYP2E1 are most important in terms of hepatic clinical drug metabolism, while the other members are predominately expressed in extra-hepatic tissues, often in a sex-specific manner, and are primarily involved in the metabolism of steroids and other endogenous substrates.

CYP2A6

CYP2A6 is mainly expressed in the liver, making up ~4% of the total hepatic CYP450 content [166, 167]. Data suggests that the constitutive androstane receptor (CAR), pregnane X receptor (PXR) and glucocorticoid receptor (GR) are involved in CYP2A6 regulation, however regulation of CYP2A6 expression is still relatively poorly understood [168]. CYP2A6 has a preference for low molecular weight, non-planar molecules, usually with 2 hydrogen bond acceptors [147]. CYP2A6 crystal structures show a relatively small, compact, hydrophobic active site [169]. Active site residue Asn297 serves as a hydrogen bond donor playing an important role in the positioning of substrates for regio-selective oxidation.

CYP2A6 oxidises several endogenous compounds including steroids, retinoic acid and bilirubin [168, 170] as well as numerous xenobiotics including pharmaceutical drugs, carcinogens, constituents of tobacco smoke and environmental toxins. CYP2A6 is responsible for the 7-hydroxylation of coumarin, the C-oxidation of nicotine and the 3-hydroxylation of cotinine in the liver. It is also the main isoform responsible for the bioactivation of the cancer prodrug tegafur

[171] and the metabolism of aromatase inhibitor letrozole to its inactive carbinol metabolite [172]. CYP2A6 inhibitors include the selective competitive inhibitors pilocarpine and tranylcypromine and mechanism based inhibitors selegiline, methoxsalen, (R)-(+) menthofuran and decursinol angelate [168].

CYP2B6

CYP2B6 is a minor hepatic P450 making up ~2 – 5% of the total hepatic CYP450 content. Expression levels vary greatly both between individual and within individuals at different time points and inter-individual variation of up to 300 fold has been reported [147]. The major regulators of CYP2B6 gene expression are nuclear receptors CAR *via* a phenobarbital-response enhancer module [173] and PXR *via* a distal xenobiotics-response enhancer module within the CYP2B6 promoter [174]. These receptors act as xenobiotic sensors and compounds such as barbiturates, rifampicin, hyperforin and other CYP2B6 substrates act as ligands for these receptors, inducing CYP2B6 transcription [147].

CYP2B6 substrates tend to be neutral or slightly basic and are typically lipophilic non-planar molecules. Crystal structures of CYP2B6 bound to various ligands [56, 175-177] reveal that the enzyme can adopt very different conformations. Conformational changes of several flexible regions in the protein drastically alter the shape of the active site, enabling the enzyme to efficiently bind and metabolise molecules of different shapes and sizes without disrupting the overall CYP450 fold [178].

CYP2B6 metabolises pharmaceutical drugs, pesticides and other environmental chemicals and pollutants [8, 179]. CYP2B6 is generally thought to play a supportive role in drug metabolism but is dominant in the metabolism of a number of drugs including cyclophosphamide, efavirenz, artemisinin, ketamine and methodone [180]. CYP2B6 also N-demethylates the recreational drug "ecstasy" leading to the formation of reactive metabolites that are potentially neurotoxic [181].

CYP2C8, CYP2C9, CYP2C19

CYP2C8, CYP2C9 and CYP2C19 make up ~20% of the hepatic CYP450 content [8]: CYP2C9 is the most abundant, followed by CYP2C8 and CYP2C19 which have expression levels ~2 and

~10 fold lower than CYP2C9 respectively [147, 182]. CYP2C enzymes are also expressed in extra-hepatic tissue such as the heart [183] and small intestine [184, 185]. CYP2C genes are regulated by PXR/CAR, GR and vitamin D receptor pathways *via* response elements in their promoter regions and are induced to varying degrees by xenobiotics [186].

Despite having an amino acid sequence identity > 80% they each have distinct substrate specificity and play unique roles in drug metabolism. CYP2C8 prefers large, weakly acidic molecules whilst CYP2C9 substrates are also generally weakly acid molecules and usually have a hydrogen bond acceptor; in contrast CYP2C19 usually binds neutral or weakly basic molecules or amides with 2 or 3 hydrogen bond acceptors [147]. Comparison of the available crystal structures of the three enzymes reveals that CYP2C8 has the largest active site [187]. While the backbone of CYP2C8 is most similar to CYP2C19, the size of their active sites differs substantially as they are lined by different amino acid side chains. The active sites of CYP2C9 and CYP2C19 are much more similar, apart from the distal B-C loop region which adopts different conformation in the two structures and it has been suggested that sequence differences within and adjacent to the B-C loop affect conformational flexibility and substrate specificity. A crystal structure of CYP2C8 with two molecules of 9-cis retinoic acid bound in the active site demonstrates conclusively that more than one substrate molecule can bind simultaneously within its active site [55].

CYP2C enzymes have a number of endogenous substrates, including arachidonic acid and some steroids [8, 147]. CYP2C9 contributes most significantly to drug metabolism in humans, metabolising most non-steroidal anti-inflammatory drugs, the anticoagulant warfarin, the anticonvulsants phenytoin and valproic acid, antidiabetics like tolbutamide and glibenclamide and many other drugs. CYP2C8 shows some overlap in substrate specificity with CYP2C9, as well as distinct catalytic activities. CYP2C8 is the main enzyme responsible for the metabolism of anticancer drug paclitaxel, antidiabetics rosiglitazone and pioglitazone and the antimalarial drug amodiaquine and also plays a major role in the metabolism of other antimalarial drugs including chloroquine and dapsone. CYP2C19 is the isoform primarily responsible for the metabolism of anticonvulsant mephenytoin, proton pump inhibitors such as omeprazole, the anticoagulant prodrug clopidogrel, the antimalarial drug proguanil as well as a number of antidepressants [8, 147].

CYP2D6

CYP2D6 makes up about 2% of the total hepatic CYP450 content [166], and is also expressed at low levels in the brain and the small intestine [185, 188]. CYP2D6 is typically considered a non-inducible gene [189, 190] and as a result its transcriptional regulation is poorly understood. Studies however suggest that hepatocyte nuclear factor 4 alpha (HNF4α) and antagonist COUP Transcription Factor 1 (COUP-TF-1) play a role in regulating CYP2D6 expression via the positive DR-1 element [191] and rifampicin has been shown to induce CYP2D6 *in vivo* [192]. CYP2D6 down regulation has also been observed in response to nitric oxide produced during inflammation [193].

Typically CYP2D6 substrates contain a basic nitrogen, 4 to 7 Å from the site of metabolism, and a planar aromatic ring. Crystal structures of CYP2D6 show that the active site adopts different conformations in the presence and absence of ligands [194, 195]. In a prinomastat bound CYP2D6 structure, the active site adopts a closed conformation and its shape closely conforms to the shape of the bound substrate [194]. The thioridazine bound structure adopts a substantially different conformation, accommodating two substrate molecules in its active site (to be published, PDB code 3TBG). This structural flexibility is likely to contribute to the catalytic versatility of the CYP2D6 enzyme. Phe120 within the B-C loop controls the orientation of the substrates aromatic ring with respect to the haem and mutations at this position can affect the enantioselectivity of oxidation [194, 196].

CYP2D6 O-demethylates endogenous substrate 5-methoxytryptamine back to serotonin and as well as two related 5-methoxyindolethylamines derived form 5-hydroxytryptamine [197, 198]. CYP2D6 also metabolises a surprising large number of drugs considering its relatively low expression levels in the liver, metabolising over 20% of all clinically used drugs [149] including antidepressants, antipsychotics, antiarrhythmics, narcotic analgesics, opioid analgesics, β-blockers and anti-cancer drugs [147]. Highly selective drugs used for CYP2D6 phenotyping include debrisoquine, dextromethorphan, metoprolol, sparteine, and tramadol [199]. CYP2D6 is highly prone to inhibition as many molecules that are not necessarily substrates bind with high affinity to its active site and as a result drug treatments involving CYP2D6 substrates are particularly sensitive to drug-drug interactions [147].

CYP2E1

CYP2E1 is expressed at high levels in the liver, making up on average ~7% of the hepatic CYP450 content [166]. It is also expressed at low levels in range of other tissues including the brain, lung, testis, ovaries and cardiac tissue. Expression levels in the liver are subject to high variability [200-202] and CYP2E1 is inducible by hormones as well as by many of its substrates *via* complex regulatory mechanisms. Transcript levels do not tend to correlate with protein expression levels and pre-translational, translational and post-translational mechanisms appear to play a significant role in regulating CYP2E1, including both mRNA and protein stabilising mechanisms [203, 204]. CYP2E1 substrates are generally small, planar molecules including organic solvents and low molecular weight xenobiotics [147]. Consistent with its preference for small substrates, the CYP2E1 active site is the smallest of the human drug metabolising isoforms. CYP2E1 however also metabolises much larger endogenous C9–C20 fatty acid signalling molecules. Fatty acid bound crystal structures reveal that the rotation of a single phenylalanine side chain merges the active site with an adjacent void forming part of the substrate access channel, significantly increasing the size of the active site and thereby allowing it to accommodate long chain fatty acid molecules [205]. This is consistent with molecular dynamic studies which reveal that the volume of the CYP2E1 active site fluctuates between 220 and 1310 Å during simulation due to the motion of side chains acting as gates to two adjacent cavities [87].

CYP2E1 metabolises solvents such as ethanol and acetone, as well as many industrial and environmental toxins including carcinogens [206, 207]. CYP2E1 is particularly prone to the production of reactive oxygen species caused by the uncoupling of oxygen consumption with NADPH [208]. This, as well as CYP2E1 tendency to metabolise xenobiotics into reactive metabolites, has implemented CYP2E1 in liver damage. CYP2E1 plays a role in the metabolism of a few clinically important drugs including chlorzoxazone, paracetamol, enflurane and general anesthetics such as halothane and sevoflurane [147].

CYP3 family

The CYP3 family consists of functional genes CYP3A4, CYP3A5, CYP3A7 and CYP3A43. The substrate specificities of these proteins overlap but they show different tissue expression patterns

[209]. CYP3A4 is the most abundant CYP450 isoform in the liver. Expression levels vary more than 100 fold between individuals and CYP3A4 can make up as much as 60% of the hepatic CYP450 content [210, 211], but averages are between 14-24% [147]. It is also expressed in the small intestine where it is the dominant CYP450 involved in first-phase metabolism of orally administrated drugs. In contrast CYP3A5 is only expressed in the liver in ~20% of individuals but is the dominant CYP3 isoform in several extra-hepatic tissues [147, 209]. CYP3A5 and CYP3A4 have an amino acid identity of over 85% and metabolise many of the same substrates, but CYP3A5 is generally more limited in its metabolic capabilities and catalyses reactions at a slower rate [212, 213]. CYP3A7 is the dominant CYP3 isoform in fetal liver [214, 215] and it is estimated that only ~20% of adult livers express active CYP3A7 [209]. CYP3A43 is expressed at low levels in the liver but at relatively high levels in the prostate and testis, where it is thought to play a role in steroid metabolism [209]; little is known about its substrate specificity but it is not expected to contribute significantly to human drug metabolism.

CYP3A protein expression is highly inducible by wide range of xenobiotics and endogenous molecules, with multiple signalling pathways regulating CYP3A genes. The major cis-acting modules responsible for inducible transcriptional regulation include the proximal PXR responsive element prPXRE, the distal xenobiotic-responsive enhancer module XREM, and the constitutive liver enhancer module CLEM4 [147]. Nuclear receptors proteins involved in CYP3A transcription regulation include xenobiotic sensors PXR and CAR [216], GR [217], the bile acid receptor FXR [218], the oxysterole receptor LXR [219], the vitamin D receptor [220] and the lipid regulator PPARα [221, 222]. Inducers of CYP3A4/5 expression include glucocorticoids, statins, baribiturates rifampicin, ritonavir, imatinib and many other drugs [147].

CYP3A4

CYP3A4 is the most important CYP450 isoform with regard to human drug metabolism. It preferentially binds large, lipophilic molecules of very diverse shapes, sizes and chemical properties but also metabolises many smaller molecules [147]. Crystal structures reveal that CYP3A4 has a large active site capable of adopting markedly different conformations and binding multiple drug molecules simultaneously [54, 57, 58, 223-227]. Molecular dynamic studies predict that CYP3A4 is the most flexible of all CYP450s, consistent with its ability to metabolise the largest and most diverse range of substrates [76, 80].

CYP3A4 substrates include drugs, carcinogens, macrocyclic natural products and endogenous steroids such as testosterone progesterone and androstenedione. CYP3A4 plays a major role in metabolism of about 30% of clinically used drugs but metabolises well over half of drugs to some degree [147]. Drugs metabolised by CYP3A4 include macrolide antibiotics (e.g. erythromycin), anti-arrhythmics (e.g. quinidine), HIV-antiretrovirals (e.g. ritonavir), calcium array blockers (e.g. nifedipine) and immunosuppressants (e.g. cyclosporine) [8]. While several probe drugs (e.g. midazolam, erythromycin and dextromethorphan) are used for CYP3A4 phenotyping, allosteric regulation and overlapping substrate binding regions result in poor correlation between data for different substrates [228-230].

1.9.3 CYP450 drug-drug interactions

Co-administration of drugs can results in CYP450 drug-drug interactions that reduce drug efficacy and/or cause severe side effects for patients. As discussed above, many drugs act as CYP450 inducers or inhibitors. In addition, more than one drug may compete for metabolism by the same CYP450 isoform or bind simultaneously to the enzyme leading to altered rates of metabolism or even altered metabolite formation, the net result being substantial changes in drug exposure. It therefore follows that the safety and efficacy profile of a drug may be drastically affected by the presence of other drugs or xenobiotics, which is particularly problematic for drugs with narrow therapeutic windows [231].

There are numerous examples of CYP450 related drug-drug interactions that have serious clinical implications. For example, the co-administration of the antihistamine prodrug terfenadine with antimicrobial agents such as ketoconazole [232, 233] or macrolide antibiotics such as erythromycin [234] leads to a type of cardiac dysrhythmia known as ‘torsades de pointes’. Terfenadine is a potassium channel blocker and is cardiotoxic at high doses – normally it is rapidly metabolised to non-toxic metabolites by CYP3A4 and as a result plasma levels of terfenadine are very low. The co-administration of CYP3A4 inhibitors however greatly reduces CYP3A4 mediated metabolism leading to high plasma levels of terfenadine that in turn disrupts the heart’s electrical cycle by causing prolonged QT intervals. Terfenadine was used by over 100 million patients prior to 1990 [235] but was later superseded by fexofenadine, the non-cardiotoxic active metabolite of terfenadine [236], due to these health risks.

A problem particularly pertinent to health care in Africa is the concurrent treatment of Tuberculosis (TB) and human immunodeficiency virus (HIV) [237]. TB and HIV co-infection is very common and co-administration of anti-tuberculosis and antiretroviral drugs has become the standard care for patients. Rifamycin antibiotics such as rifampin are an essential part of multidrug regimens for TB treatment, but they are strong inducers of drug transporters, CYP450s and Phase II drug metabolising enzymes and therefore greatly reduce the plasma concentrations of co-administrated antiretrovirals (e.g. nevirapine) that are metabolised by these enzymes, potentially compromising HIV treatment [238-241].

To further complicate drug treatment, unregulated and non-prescribed herbal remedies, traditional medicines and even foodstuffs (e.g. St John's wort, the African potato, ginseng, garlic and grapefruit) can alter CYP450 metabolism and cause severe drug-drug interactions when ingested with prescription drugs [242]. For example, the African potato is widely used in Africa for the treatment of HIV, yet compounds in the African potato can inhibit CYP3A4 activity by up to 90% so if this traditional remedy is used together with prescription drugs that are CYP3A4 substrates, those drugs may accumulate to toxic levels in the body [243].

In addition to small molecule drugs, therapeutic proteins such as cytokines used in cancer treatment (or proteins that modulate cytokine activity such as monoclonal antibodies) can also cause drug-drug interactions by influencing the expression of CYP450 enzymes and drug transporters proteins [244].

1.9.4 CYP450 pharmacogenetics: inter-individual differences in drug response

As a consequence of both intrinsic and extrinsic factors, drug response - both in terms of drug efficacy and the occurrence and severity of side effects - can differ dramatically from one patient to another [245]. Differences in drug response can often be attributed to differences in CYP450 mediated drug metabolism, which is highly variable between patients. Although non-genetic factors such as age, sex, diet, disease state and co-administration of drugs affects CYP450 expression and function, genetic variation underlies many of the inter-individual differences in drug metabolism and associated adverse drug reactions [147].

While CYP450s have very broad and often overlapping substrate specificities, many drugs are only or mainly metabolised by one CYP450 isoform and as a result changes in the expression or function of even a single isoform can significantly alter the pharmacokinetic profile of a drug. If a drug is metabolised too fast, the bioavailability may be too low making the drug ineffective; on the contrary, if the drug is metabolised too slowly, drug levels may be too high for too long causing toxicity (in the case of a prodrug, where the CYP450 metabolism activates the drug, the opposite is true). Differences in the nature and relative quantities of reactive metabolites formed can also lead to varied side effects.

CYP450 drug metabolising enzymes are highly polymorphic in the population and the occurrence and frequency of variant alleles varies between ethnic groups. Variant alleles include deletions, insertions, copy number variants and single nucleotide polymorphisms (SNPs) both in the coding and non-coding regions of the genes. This variation can lead to aberrant splicing, frameshifts, premature termination of translation, can affect transcription, alter protein expression levels or influence protein function. Consequently, aberrant CYP450 alleles can diminish, enhance or alter CYP450-mediated drug metabolism. Table 1.2 gives examples of major aberrant CYP450 alleles that influence clinical drug metabolism and patient outcomes - the global minor allele frequency and frequency within different ethnic groups is also shown for each allele.

The terms “extensive metaboliser”, “poor metaboliser”, “intermediate metaboliser” and “ultra-rapid metaboliser” are commonly used to describe the pharmacokinetic phenotype associated with CYP450 genotypes [149]. Extensive metabolisers have two normal functional alleles and most recommended drug doses assume patients are extensive metaboliser. Poor metabolisers are homozygous for a non-functional allele or compound heterozygous for two different non-functional alleles and therefore lack active enzyme. As a result, poor metabolisers show low drug clearance, which can lead to toxicity and adverse drug reactions due to prolonged drug exposure or alternatively they do not benefit from prodrugs that are activated by that particular CYP450. Intermediate metabolisers are heterozygous for a non-functional allele or homozygous for two partially functioning alleles and can also experience reduced drug clearance, but to a lesser degree than poor metabolisers. Intermediate metabolisers are also more susceptible to developing a poor metaboliser phenotype in response to drug inhibition or environmental factors. Ultrarapid metabolisers have an increased number of functional gene copies and as a result have

an increased rate of drug clearance, reducing drug efficacy; conversely rapid activation of prodrugs by ultrarapid metabolisers can lead to toxicity.

These phenotype-genotype associations are of course greatly over-simplified and do not hold true for all drugs or in all patients, since polymorphisms can have substrate specific effects and genetic variation in drug transporters, drug receptors, other drug metabolising enzymes and the drug target itself may also have functional implications that make it difficult to predict the clinical outcome [246, 247].

Owing to CYP1A, CYP1B and CYP2E1 proteins' role in the activation and clearance of carcinogens and the metabolism of signalling molecules, genetic variation in these enzymes has been studies extensively as risk factors for cancer development and several variants have been identified that increase patient's susceptibility to certain cancers (examples in Table 1.2).

CYP2A6 enzymes are highly polymorphic and ~ 9 % of Caucasians, ~22 % of Africans and 50% of Asians are estimated to have low-activity or loss-of-function alleles [248]. Patients that are homozygous or compound heterozygous for these defective alleles are poor metabolisers of nicotine, which may influence smoking behaviour and susceptibility to certain cancer. These CYP2A6 polymorphisms also influence the metabolism of some pharmaceutical drugs such as the anticancer prodrug tegafur.

Polymorphic variation in CYP2B6, CYP2C9, CYP2C19 and CYP2D6 has been implicated widely in altered drug response and adverse reactions (Table 1.2). Alleles causing changes in expression as well as altered protein function have been identified that lead to both poor and ultrarapid metaboliser phenotypes, with serious implications for patients on certain drug treatments. CYP2D6 polymorphic variation has the highest clinical impact since environmental and non-genetic factors have comparably little influence on CYP2D6 expression and many null alleles as well as alleles with several fold amplifications exist [147].

In contrast, CYP3A4 polymorphic variation is generally thought to have less significant clinical impact as there are no null alleles and, while variants in the coding region can affect protein function, they occur at low frequencies of less than 5% heterozygous with the wild-type allele, limiting their effect on drug clearance [249]. In addition CYP3A4 is a highly inducible enzyme

and expression levels vary greatly within and between patients in response to non-genetic factors. Importantly however single amino acid substitutions in CYP3A4 can affect the regioselectivity of oxidation [250-252]; consequently CYP3A4 polymorphic variation should not in fact be ignored since mutations could lead to the formation of toxic metabolites and may have important implications for drug-drug interactions commonly observed for CYP3A4 substrates.

Table 1.2 Examples of the major aberrant CYP450 alleles that influence clinical drug metabolism and patient outcomes

CYP450 allele	Key mutations	Location/amino acid change	Allele Frequencies	Functional effect	Clinical consequence	References
CYP1A2*1C	−3860G>A	Promoter	18.8 % (26–40% Africans, African Americans; 21–27 % Asians, Pacific; 1-8 % Caucasians; 20–30 Hispanic)	Decreased inducibility (smokers)	May influence susceptibility to certain cancers; decreased caffeine 3-demethylation in Japanese smokers	[253]
CYP1A2*1F	−163C>A	Intron 1	35 % (A>C) (note C is the reference allele even though A is more common)	Increased inducibility (smokers and coffee consumers)	Increased susceptibility to cancer; increased metabolism of olanzapine; increase susceptibility to myocardial infarction	[254-260]
CYP2A6*2	1799T>A	L160H	1.3 % (0–1 % African, African American; 0–2.5 % Asian; 4-10% Caucasian)	No activity	Decreased nicotine metabolism; increased risk of tobacco-related lung cancer; influence on smoking behaviour and nicotine dependence; decreased activation and clearance of anticancer prodrug tegafur.	[261, 262]
CYP2A6*4A to *4H	Recombination	Gene deletion	(1–2 % African American; 5–24 % Asian; 1–4% Caucasian)	Null alleles		
CYP2A6*7	6558T>C	I471T	4 % (0 % African, African American; 6 - 13% Asian; 0 % Caucasian)	Decrease activity (Substrate specific *in vitro*)		
CYP2A6*9	48T>G	Promoter, TATA-box	13 % (4–2 % African, African American; 16–27 % Asian; 4–5 % Caucasian)	Decreased activity		
CYP2A6*10	1412T>C; 1454G>T	I471T; R485L		Decreased activity (essentially inactive towards nicotine *in vivo*)		
CYP2A6*17	5065G>A	V365M	2.5 % (4 – 50% African, African American; 0 % Asian; 0-2% Cauc.)	Decreased activity (substrate specific effect *in vitro)*		

CYP2B6*4	18053A>G	K262R	2-6 % (0% African, African American; 5-12 % Asian; 4 % Caucasian)	Increased activity (substrate specific effects *in vitro*)	Increased drug clearance (bupropion, efavirenz, nicotine)	[263-265]
CYP2B6*6	15631G>T; 18053A>G	Q172H; K262R	27 % (33-50 % African, African American; 21 % Asian; 14-27 % Caucasian; 62% Pacific)	Decreased expression in the liver (substrate specific effects *in vitro*)	Decreased drug clearance leading to adverse reactions (efavirenz, nevirapine, S-methadone)	[266-272]
CYP2C9*2	3608C>T	R144C	6.9 % (0-2 % African, African American; 0–2 % Asian, Pacific; 10-17 % Caucasian; 6.5 % Hispanic)	Decreased activity	Decreased drug clearance and increased risk of adverse drug reactions (Anticoagulants warfarin, acenocoumarol and phenprocoumone; sulfonylurea hypoglycemic drugs; NSAIDS)	[273-280]
CYP2C9*3	42614A>C	I359L	4.3 % (0-1% African , African; 2 – 6 % Asian; 6% Caucasian)	Decreased activity		
CYP2C19*2	19154G>A	Splicing defect	20 % (10- 17 % African, African American; 22-32 % Asian, Pacific; 6-15 % Caucasian; 15 % Hispanic)	Null allele	Decreased clearance of PPIs leading to increased efficacy in Helicobacter pylori eradication therapy; decrease clearance and increased risk of adverse drug reactions for antidepressants (amitriptypine, citalopram, clomipramine, moclobemide), antimalarials (proguanil) and antifungals (voriconazole); decreased activation of prodrug clopidogrel resulting in decreased anticoagulation effect and increased risk of cardiovascular events.	[281-295]
CYP2C19*3	17948G>A	W212X	1.4 % (0-1 % African, Caucasian, Hispanic; 3-7 % Asian, Pacific)	Null allele		

CYP2C19*17	−806C>T	Promoter	15 % (15–27 African, African American; 0-2 % Asian; 21-25 % Caucasian)	Increased expression	Increased clearance of PPIs leading to decreased efficacy; increased activation of prodrug clopidogrel leading to increased risk of bleeding	[282, 283, 296, 297]
CYP2D6*3	2549delA	Frameshift	gMAF 0.009 ~0.01 all ethnicities	Null allele	Decreased clearance of many antiarrhythmics, antidepressants and antipsychotics leading to increased risk of adverse drug reactions; decreased activation of opioids (codeine, dihydrocodeine, oxycodone, tramadol) leading to decreased analgesic effect; decreased activation and efficacy of tamoxifen.	[298-302]
CYP2D6*4	Recombination	Deletion	3-6 % all ethnicities	Null allele		
CYP2D6*6	1707delT	Frameshift	~1% all ethnicities	Null allele		
CYP2D6*10	100C>T	P34S	26 % (8-12% African, African American; 40-70 % Asian; 2 % Caucasian)	Decreased expression & activity		
CYP2D6*17	1023C>T	T107I R296C	5 % (for 1023C>T) (14-24 % African; 0 % Asian, Caucasian)	Decreased expression & activity		
CYP2D6*41	2988G>A	Splicing defect	5.5 % (1-6 % African, Asian, Pacific, Hispanic; 9 % Caucasian)	Decreased expression & activity		
CYP2D6*Nxn	Recombination	Copy number variations	Up to 30 % African, African American; 1-9 % Caucasian	Increased expression & activity	Increased activation of opioid prodrugs leading to toxicity	[303-305]
CYP3A4*22	15389 C>T	Intron 6	2.1 % (4.3 % African American; 4.3 % Asian; 2.5-8% Caucasian)	Decreased expression & activity	Decreased metabolism of statins (decreased clearance of simvastatin leading to increased lipid-lowering response); decreased dose requirements of calcineurin inhibitors (tacrolimus) for stable renal transplant patients	[306-308]

Table adapted from [147]

Polymorphisms affecting protein function

Table 1.3 shows the number of distinct star-alleles with variant amino acid sequence or with proven functional affects for each of the major human drug metabolising isoforms ([309]; accessed 20 Jan 2014)). Many SNPs code for amino acid changes that occur in different combinations and in addition to these SNPs, haplotype variants and SNPs not yet assigned to a particular allele also exist. A list of the single amino acid substitutions resulting from polymorphic variation for each isoform can be found in Appendix B Table B1-B9.

Table 1.3 The number of distinct star-alleles for each of the major human drug metabolising isoforms

CYP450 isoform	No. of star-alleles
CYP1A2	21
CYP2A6	38
CYP2B6	38
CYP2C8	14
CYP2C9	58
CYP2C19	34
CYP2D6	105
CYP2E1	7
CYP3A4	24

While *in vivo* data provides useful information on the clinical phenotype associated with a particular allele, results are often contradictory due to other compounding genetic and non-genetic factors [147]. The presence of multiple SNPs in a single allele can also make it difficult to understand the contribution of each individual SNP to an altered phenotype and hence the molecular mechanisms underlying the changes in CYP450 drug metabolism are poorly understood based on clinical studies alone.

Recombinant expression systems for individual CYP450 enzymes have enabled site-directed mutagenesis studies to investigate the effect of SNPs on protein structure and function in controlled systems. *In vitro* studies have shown that single amino acid substitutions can affect the binding affinity of substrates, alter the turnover rates of oxidation by as much as 1000 fold, or even change the enantioselectivity or regioselectivity of oxidation [149, 252, 310-312]. Mutations may alter protein function by changing the stability or flexibility of the holoprotein,

by directly affecting substrate binding, or by altering interactions with protein partners. The effect of SNPs can also be substrate specific, affecting the metabolism of some substrates but not others [313-317]. For example, Ariyoshi *et al* found that allele *CYP2A6*7*, coding for the mutation I471T, lacked nicotine C-oxidase activity yet coumarin-7 hydroxylase activity was conserved. Similarly, Wang *et al* found that CYP2C19 E122A showed comparable *S*-mephenytoin 4-hydroxylation activity to the wild-type but significantly decreased omeprazole 5'-hydroxylase activity. Furthermore, single amino acid substitutions can affect CYP450 inhibition in an inhibitor dependant manner. Additionally the presence of more than one mutation does not necessarily have an additive effect since two mutations that show no effect on protein function in isolation may significantly disrupt protein function when they occur in combination; alternatively instability as a result of one mutation may be counter-balanced by another mutation. A screening platform has been used recently to assay 18 CYP2C19 variants against 10 different drugs [315]; however few such studies have been reported to date and as a result, the effect of non-synonymous mutations on drug metabolism remains poorly understood experimentally for all except a limited set of mutations and substrates.

1.9.5 CYP450s and drug development

The pharmaceutical industry is facing unprecedented financial challenges as despite escalating costs, research and development (R&D) productivity has been declining [318, 319]. Drug failure rates are high with only 8% of drugs entering clinical trials eventually making it to market [320]. The average drug takes over 10 years to reach the shelf [320] and costs ~$ 1.8 billion to develop - most of which represents the cumulative costs of failed drug initiatives [318]. Problems with drug efficacy, toxicology and safety are the leading causes of attrition in drug discovery. In addition, drugs are only effective in ~40% of patients to whom they are prescribed [321] and even after gaining FDA-approval, around 4% of approved drugs are later withdrawn due to unexpected side effects [320]. This highlights the need for new approaches to drug discovery and improved methods of testing the efficacy and safety of drugs prior clinical trials.

CYP450s are notorious in the pharmaceutical industry as they are regularly responsible for adverse drug reactions, poor drug efficacy and the failure of drugs in late stages of clinical trials. Consequently CYP450 drug screening has become a routine part of pre-clinical drug testing.

The drug discovery process typically begins with target identification followed by drug screening to identify promising lead compounds. Computational tools are used to predict the absorption, distribution, metabolism, and excretion (ADME) and toxicity of lead compounds based on their chemical structure and properties [322, 323]. These *in silico* studies are followed by preclinical testing to identify the primary drug elimination pathways. Key preclinical studies include *in vitro* kinetic studies using hepatocytes, liver microsomes and recombinant CYP450s to determine which CYP450 isoforms are responsible for metabolising the drug and to investigate enzyme induction and inhibition that could lead to drug-drug interactions [324]. In addition, genotoxicity and cytotoxicity studies are carried out on the parent compound and its metabolites to test for reactivity that may lead to adverse side effects. After rigorous preclinical testing, successful drug candidates are tested *in vivo* using animal models (usually in mice, rats, dogs or other primates) and finally, once a drug meets all the safety requirements, it enters clinical trials in humans.

Clinical trials can be divided into three phases [325]: Phase I trials are used to determine tolerable drug doses based on pharmacokinetic parameters as well as drug safety in 20 to 80 healthy volunteers. In Phase II trials the dosing regimen is verified and adverse reactions are monitored at higher dosages over longer periods in a larger group of individuals (several hundred) with the target disease. Phase III trials serves as a final test for safety and efficacy. These trials are performed in much larger groups of people (several thousand) and are randomised, including both disease and control groups.

1.9.6 The shift towards personalised medicine

Treatment prescriptions and dosage guidelines have traditionally been based on the general or average population rather than an individual or a particular ethnic group. Pharmacogenetic studies are now however showing an increasing number of genetic markers that can consistently be used to predict adverse drug reactions and drug efficacy. As this information is used more effectively, the medical profession seems set to be revolutionised as it begins to move away from a “one drug fits all” approach to drug treatments tailored for the individual. While the incorporation of this valuable genetic information into clinical practice has initially been slow [326], substantial progress developing standardised clinical pharmacogenetic guidelines for clinical practice has been made in in the past few years as a result of collaborative efforts by the

Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Pharmacogenetics Working Group of the Royal Dutch Pharmacists Association [327-329]. These guidelines are now freely available online [330]. In a list of over 500 well-known pharmacogenetic associations compiled by the CPIC, CYP450 genes are implicated in over 50% of cases [331]. In addition, CYP450 genotype-phenotype associations are mentioned in 56 of the 91 FDA-approved drug labels containing pharmacogenomic information [332]. The high prevalence of CYP450 genes in these lists further emphasises the clinical relevance of CYP450 polymorphic variation. Several clinically relevant genetic tests for CYP2C9, CYP2C19 and CYP2D6 genotyping now exist including Roche AmpliChip CYP450 Test, DMET Plus (Affymetrix, Inc); VeraCode ADME Core Panel (Illumina, Inc) and TaqMan Drug Metabolism Genotyping Assay Sets (Applied Biosystems, Inc) [333].

The anticoagulant drug warfarin is a well-known example of a drug with established genetic biomarkers that can guide dosing and help prevent adverse drug reactions and poor efficacy. Warfarin is the most commonly prescribed anticoagulant for the treatment and prevention of thrombotic disorders in the world [334] but it has a very narrow therapeutic window and consequences of over- or under-dosing can be devastating [335-337]. There is substantial inter-individual variation in patient responses to warfarin and consequently dosage adjustments have to be made frequently. Numerous genetic factors as well as clinical factors such as age, weight, smoking status and co-administration of other medications influence a patient response [275]. Polymorphisms in the *CYP2C9* and *VKORC1* genes have been identified as genetic markers affecting warfarin dosage requirements. *S*-warfarin, the most potent enantiomer, is primarily metabolised by CYP2C9 and *CYP2C9*2* and *CYP2C9*3* alleles encoding CYP2C9 proteins with R144C and I359L mutations respectively have been associated with poor warfarin clearance. Poor metabolisers are at risk of bleeding complications due to prolonged drug exposure and require lower warfarin doses; the mechanistic basis for this is that warfarin reduces blood-clotting by binding to the vitamin K reductase protein, thereby inhibiting the formation of vitamin-K-dependent clotting factors. Common SNPs in the non-coding region of the *VKORC1* gene can increase or decrease the expression levels of the vitamin K reductase protein and consequently affect a patient's sensitivity to warfarin. This pharmacogenetic information has been incorporated into the FDA-approved drug label under the "Dosage and Administration" and "Clinical Pharmacology" sections (Figure 1.9). Genotyping prior to use is not yet mandatory but, McWilliam *et al* estimated that the incorporating mandatory genetic testing into warfarin

treatment would prevent 85 000 serious bleeding events and 17 000 strokes and reduce healthcare expenditure by over $1 billion annually [338].

A

Dosing Recommendations with Consideration of Genotype

Table 1 displays three ranges of expected maintenance COUMADIN doses observed in subgroups of patients having different combinations of CYP2C9 and VKORC1 gene variants [see *Clinical Pharmacology (12.5)*]. If the patient's CYP2C9 and/or VKORC1 genotype are known, consider these ranges in choosing the initial dose. Patients with CYP2C9 *1/*3, *2/*2, *2/*3, and *3/*3 may require more prolonged time (>2 to 4 weeks) to achieve maximum INR effect for a given dosage regimen than patients without these CYP variants.

Table 1: Three Ranges of Expected Maintenance COUMADIN Daily Doses Based on CYP2C9 and VKORC1 Genotypes†

VKORC1	CYP2C9					
	*1/*1	*1/*2	*1/*3	*2/*2	*2/*3	*3/*3
GG	5-7 mg	5-7 mg	3-4 mg	3-4 mg	3-4 mg	0.5-2 mg
AG	5-7 mg	3-4 mg	3-4 mg	3-4 mg	0.5-2 mg	0.5-2 mg
AA	3-4 mg	3-4 mg	0.5-2 mg	0.5-2 mg	0.5-2 mg	0.5-2 mg

†Ranges are derived from multiple published clinical studies. VKORC1 –1639G>A (rs9923231) variant is used in this table. Other co-inherited VKORC1 variants may also be important determinants of warfarin dose.

B

CYP2C9 and VKORC1 Polymorphisms

The *S*-enantiomer of warfarin is mainly metabolized to 7-hydroxywarfarin by CYP2C9, a polymorphic enzyme. The variant alleles, CYP2C9*2 and CYP2C9*3, result in decreased *in vitro* CYP2C9 enzymatic 7-hydroxylation of S-warfarin. The frequencies of these alleles in Caucasians are approximately 11% and 7% for CYP2C9*2 and CYP2C9*3, respectively.

Other CYP2C9 alleles associated with reduced enzymatic activity occur at lower frequencies, including *5, *6, and *11 alleles in populations of African ancestry and *5, *9, and *11 alleles in Caucasians.

Warfarin reduces the regeneration of vitamin K from vitamin K epoxide in the vitamin K cycle through inhibition of VKOR, a multiprotein enzyme complex. Certain single nucleotide polymorphisms in the VKORC1 gene (e.g., –1639G>A) have been associated with variable warfarin dose requirements. VKORC1 and CYP2C9 gene variants generally explain the largest proportion of known variability in warfarin dose requirements.

CYP2C9 and VKORC1 genotype information, when available, can assist in selection of the initial dose of warfarin [see *Dosage and Administration (2.3)*].

Figure 1.9 Extracts from the prescription information for warfarin sold under the brand name Coumadin. A) From Dosage and Administration section. B) From Clinical Pharmacology section.

Although clinical pharmacogenetics guidelines are being established and used in clinical practice for a range of drugs, genetic testing prior to drug administration is not yet a routine procedure. In addition, there are likely to be many more clinically important CYP450 alleles affecting drug

response since to date the effect of CYP450 polymorphism on drug metabolism has only been determined for a relatively small number of SNPs and potential drugs. Thus CYP450 pharmacogenetic information is far from complete and the full benefit of the existing information is yet to be realised.

1.9.7 *In silico* approaches for predicting the effects of SNPs on protein function.

There are a number of *in silico* tools which aim in general to predict the effect of single point mutations on protein stability, based on methods such as physical effective energy functions [339-341], empirical potential energy functions [342, 343], statistical potential energy functions [344-346] and machine learning [347-350]. Of these, physical potential energy functions are likely to be the most accurate but are generally computationally expensive and time consuming whilst machine learning methods and empirical potential energy functions use experimental data to fit their function and generally suffer from over fitting the function to the training data set. In addition, although a large number of mutations do result in loss of protein function due to effects on protein stability, a significant number also affect functional sites directly [351].

A number of *in silico* tools are available that aim to rapidly distinguish between disease associated mutations and benign mutations based on the residues position relative to functional sites, protein stability calculations or a combination of both. These tools include SIFT [352, 353], SNAP [354], FoldX [355], SDM [356-358], PolyPhen [359], I-mutant [360] amongst many others. However, whilst a few of these tools have been used to predict the effect of SNPs on CYP450 function [361], none of them are specifically tailored for CYP450 enzymes and consequently, if taken into account at all, functional site predictions are by no means comprehensive and the predicted effects of mutations on functional sites are only qualitative. Quantitatively predicting the effect of mutations on ligand and co-factor binding is still a huge challenge for computational biology as accurate calculations are very computationally intensive, limiting their use for large interactions surfaces or large numbers of mutations.

1.9.8 *In vitro* methods for testing CYP450 drug metabolism

Apart from the ethical issues surrounding animal testing, animal studies are also expensive and time consuming and while there are similarities in drug metabolism between animal models and

humans, animals have different CY450 isoforms with different expression levels and catalytic properties, limiting their applicability to human drug metabolism. As a consequence, accurate *in vitro* human drug metabolism models for testing drugs prior to clinical trials would add substantial value. Current *in vitro* methods for measuring CYP450 drug metabolism include liver slices, primary hepatocytes, transformed cell lines, microsomes and recombinant enzymes [362, 363]. *In vitro* screening techniques to monitor CYP450 substrate turnover include radiometric, fluorogenic and chromatography/mass spectrometry based activity assays [364-366]. A number of mathematical models have been developed to extrapolate *in vivo* CYP450 activity from *in vitro* assays: this is simplest for more complex systems that more closely mimic the liver environment, such as liver slices and hepatocytes [367], but strategies for reconstructing the relative contribution of individual CYP450 isoform based on recombinant proteins assays have also been developed [368]. While these *in vitro–in vivo* extrapolation techniques are limited, they are likely to become more reliable as the quality of *in vivo* and pharmacokinetic data improves. Despite the challenges, *in vitro* assays are already very useful for predicting CYP450 drug metabolism and clinical important drug-drug interactions [369] .

Liver slices, primary hepatocytes and transformed cell lines

Liver slices and primary hepatocytes most closely resemble the *in vivo* system and contain the full complement of CYP450s and other drug metabolising enzymes within their cellular environments. Access to high quality human liver samples is however limiting and viability is poor because samples deteriorate rapidly. In addition cells can be damaged during high precision tissue slicing or hepatocyte extraction whilst liver samples also vary due to inter-individual differences between donors e.g. age, sex, disease state and genetic variation. Thus, while liver slices may be useful for studying inter-individual variations, sample sizes are normally too small to be of statistical significance and poor reproducibility complicates routine testing. Hepatocytes are however particularly good models for predicting enzyme induction as well as liver toxicity [363].

Primary hepatocytes can be cultured up to a few passages, but CYP450 expression tends to decline in an isoform specific manner [370]. Various culturing techniques have been explored to preserve hepatocyte activity during prolonged culturing, including: co-culturing hepatocytes with other cells including other liver cells such as hepatic Kupffer cells [371] or extracellular matrix

producing cells such as fibroblasts [372, 373]; the use of culturing matrices such as hydrogels that attempt to mimic the extra-cellular environment [374]; and the addition of hormones, specific nutrients, and inducers to the culture medium [362]. However, concerns remain about the effect of such methods on data interpretation [362].

A number of transformed human liver cell lines exist, such as HepG2 and BC2 but, while they are more stable, they are less reliable than primary hepatocyte cultures, generally showing low or partial CYP450 expression with expression patterns that do not closely resemble those in the liver [362].

While whole cell systems provide a more "*in vivo*-like" environment for the CYP450s, quantifying the contribution of individual CYP450 isoforms to drug metabolism is more difficult in samples containing multiple CYP450 isoforms. Isoform specific probes can give information on individual CYP450s activity but, due to overlapping specificities between isoforms, many probes have low specificity making accurate metabolic profiling challenging [375]. Determining the effect of polymorphic variation on drug metabolism is also difficult due to the confounding effects of other isoforms and other inter-individual differences; in addition, availability of liver samples expressing low frequency alleles is limiting.

"Organs on a chip" technologies

More recent advances in microfluidics and tissue engineering technologies have led to the development of highly miniaturized "organ on a chip" tissue models [376]. A number of these 3D systems attempt to mimic the liver environment using co-cultured hepatocytes [377, 378]. The fluid channels in the microfluidic systems enable the tightly controlled flow of medium creating organ-like perfusion. *In vitro* microfluidic chips for monitoring drug metabolism have also been developed using cell free systems, subcellular fractions (microsomes), cell lines and intact tissue [379]. Despite promising progress in this field however, these systems are not yet able to mimic hepatic drug metabolism as effectively as primary hepatocyte cultures or liver slices.

Microsomes

Microsomes are vesicle like fractions of the endoplasmic reticulum formed during tissue homogenisation and can be prepared by differential centrifugation [380]. Human liver microsomes are the most common source of CYP450s for *in vitro* metabolism studies. The proportion of membrane bound proteins in microsomes is representative of the proportions found in the liver, however drug transport is not taken into account and many phase II drug metabolising enzymes are missing from these preparations [362]. While microsomes are useful for CYP450 activity assays, they are not suitable for induction studies as the required protein expression systems are no longer intact. Microsomes are now commercially available together with prior characterisation of the CYP450 activity and variability between preparations is reduced by creating microsome pools which are representative of a particular group of people rather than just one individual. In addition, microsomes generally show higher activity than hepatocytes or liver slices as the CYP450 content is enriched and there is less competition with other enzymes.

Recombinant enzymes

While recombinant CYP450 expression systems do not accurately mimic the *in vivo* environment, they are the most useful tool for studying the activity of individual CYP450 isoforms in isolation [363]. As opposed to trying to understand drug metabolism in a complex system like the liver, recombinant enzymes provide a way to deconvolute the system and determine the role of individual enzymes in the metabolism of a given drug. What is more, this renewable enzyme source enables the production of high purity proteins in high quantity suitable for biophysical studies (e.g. X-ray crystallography and spectroscopic studies) and site-directed mutagenesis enables structure-function studies. This is particularly useful for determining the effect of non-synonymous mutations on protein stability and/or function and such information can be used both to rationalise *in vivo* data and to guide further clinical studies.

Recombinant CYP450 have been expressed in a variety of different host cells including bacteria, yeast, insect and mammalian cells and whole cells (transgenic cell lines), microsomal preparations (referred to as supersomes or baculosomes when expression is baculovirus mediated) or purified proteins can be used in downstream applications.

Despite these advances however, the pharmaceutical industry remains in need of cost-effective, high-throughput methods suitable for screening large numbers of drugs candidates against wild-type CYP450 enzymes as well as against all clinically relevant polymorphic forms. Typically high-throughput CYP450 screens are very expensive and not widely accessible, requiring large quantities of protein and banks of automated liquid chromatography/mass spectroscopy (LC/MS) machines for metabolite detection [321]. While there have been several attempts to create miniaturised high-throughput CYP450 assays that do not rely on large quantities of protein or LC/MS technology (e.g. [381, 382]), miniaturising CYP450 assays is fraught with challenges and these approaches only work on a limited scale. As such, there is still much scope for the development of high-throughput assays for monitoring CYP450 activity and screening for P450-mediated drug-drug interactions.

1.10 Objectives

The effect of non-synonymous single nucleotide polymorphisms on cytochrome P450 drug metabolism is currently poorly understood due to the large number of polymorphisms and the diversity of potential substrates. The pharmaceutical industry is thus in need of new reliable *in silico* and *in vitro* approaches to predict the metabolism of drug candidates in individual patients at early stages of drug development.

A reliable *in silico* approach for predicting CYP450 variants with altered function would serve as a useful tool for prioritising variants for *in vitro* study as well as for rationalising *in vitro* and clinical data. Furthermore a reliable cost-effective high-throughput platform for monitoring CYP450 activity enabling the rapid screening of selected variants in parallel for altered drug metabolism, inhibition and drug-drug interactions would provide valuable data for preclinical drug development and for guiding pharmacogenetic studies carried out in clinical trials. These are therefore the broad aims and objectives addressed in this thesis.

The subsequent chapters of this thesis thus describe:

1) The relationship between structure and function in human CYP450 and the delineation of regions important for substrate access and binding.

2) An *in silico* approach for predicting the effect of mutations on CYP450 stability and function and guiding the selection of variants for *in vitro* testing.
3) The *in vitro* thermostability and kinetic parameters of a panel of CYP3A4 and CYP2C9 variants.
4) The utility of a high-throughput CYP450 protein microarray platform for rapidly determining the kinetic parameters of panels of CYP450 variants in parallel.

Chapter 2 Substrate recognition sites in mammalian CYP450s

2.1 Introduction

Xenobiotic metabolising CYP450s are often termed "promiscuous" as individual isoforms can metabolise compounds with very different shapes, sizes and chemical properties. This broad substrate specificity allows relatively few enzyme catalysts to handle a large range of foreign compounds. While there is some overlap between isoforms, the substrate specificities, even between CYP450s with over 95% sequence identity, are distinguishable. For example, both CYP2C4 and CYP2C5 catalyse the 21-hydroxylation of progesterone, but the binding affinity of progesterone is 10-fold greater for CYP2C4 than for CYP2C5, despite a >95% sequence identity between these two enzymes [383]. This difference in binding affinity is caused by a single amino acid difference within a substrate binding region. Amino acid differences within substrate binding regions can also result in differences in the regioselectivity of metabolism as seen for the metabolism of aflatoxin B1 by human CYP3A4 and CYP3A5, where 3α-hydroxylation by CYP3A4 leads to detoxification of aflatoxin B1 but exo-8,9-oxidation by CYP3A5 leads to the formation of a genotoxin [384]. Determining the regions of the enzyme responsible for this vast catalytic diversity, yet distinct substrate specificity, displayed by CYP450s is crucial for predicting the effects of polymorphic variation on drug metabolism.

In 1992, Gotoh *et al* described the first substrate recognition site map for mammalian cytochrome P450s by inferring substrate interacting residues from the only substrate bound CYP450 crystal structure available at the time [53]. Using the crystal structure of *Pseudomonas putida* P450 101A (P450cam) in complex with a camphor molecule, Gotoh manually identified all residues lying within 10 Å of the oxygen atom of the bound camphor molecule as putative substrate recognition sites (SRS). SRS regions were then inferred in CYP2 enzymes by group to group alignment of mammalian CYP2 sequences with P450cam and other bacterial CYP450 sequences. From this work six putative SRS regions were identified (Figure 2.1).

By interchanging or mutating residues within SRS regions or studying chimeras of closely related CYP450s, experimental studies have validated Gotoh's SRS map, showing that residues within these SRS regions are indeed important for substrate binding and specificity in mammalian CYP450s [385-392]. There are however other experimental studies suggesting that residues lying outside of Gotoh's SRS map are also important for substrate binding in mammalian CYP450s [385, 393, 394].

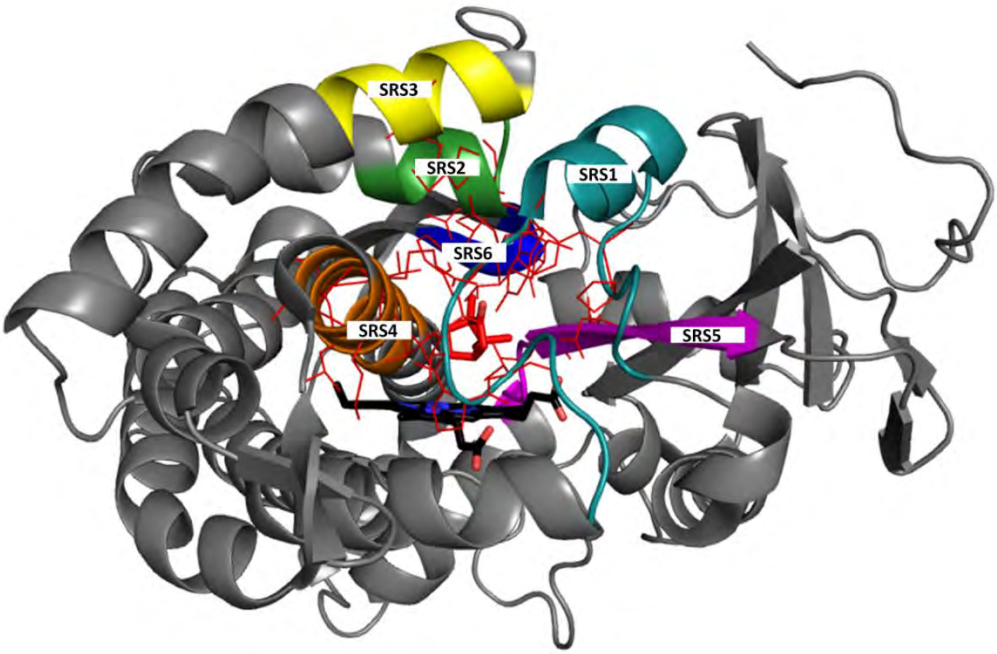


Figure 2.1 Gotoh's original substrate recognition site map
Structure of bacterial cytochrome P450cam bound to a camphor molecule (red sticks), PDB code 2CPP. The SRS regions delimited by Gotoh are coloured as follows: SRS1, cyan; SRS2, green, SRS3, yellow; SRS4, orange; SRS5, magenta; SRS6, blue. Residues that Gotoh found to be within 10Å of the camphor molecule are shown as red lines. The haem co-factor is shown in black.

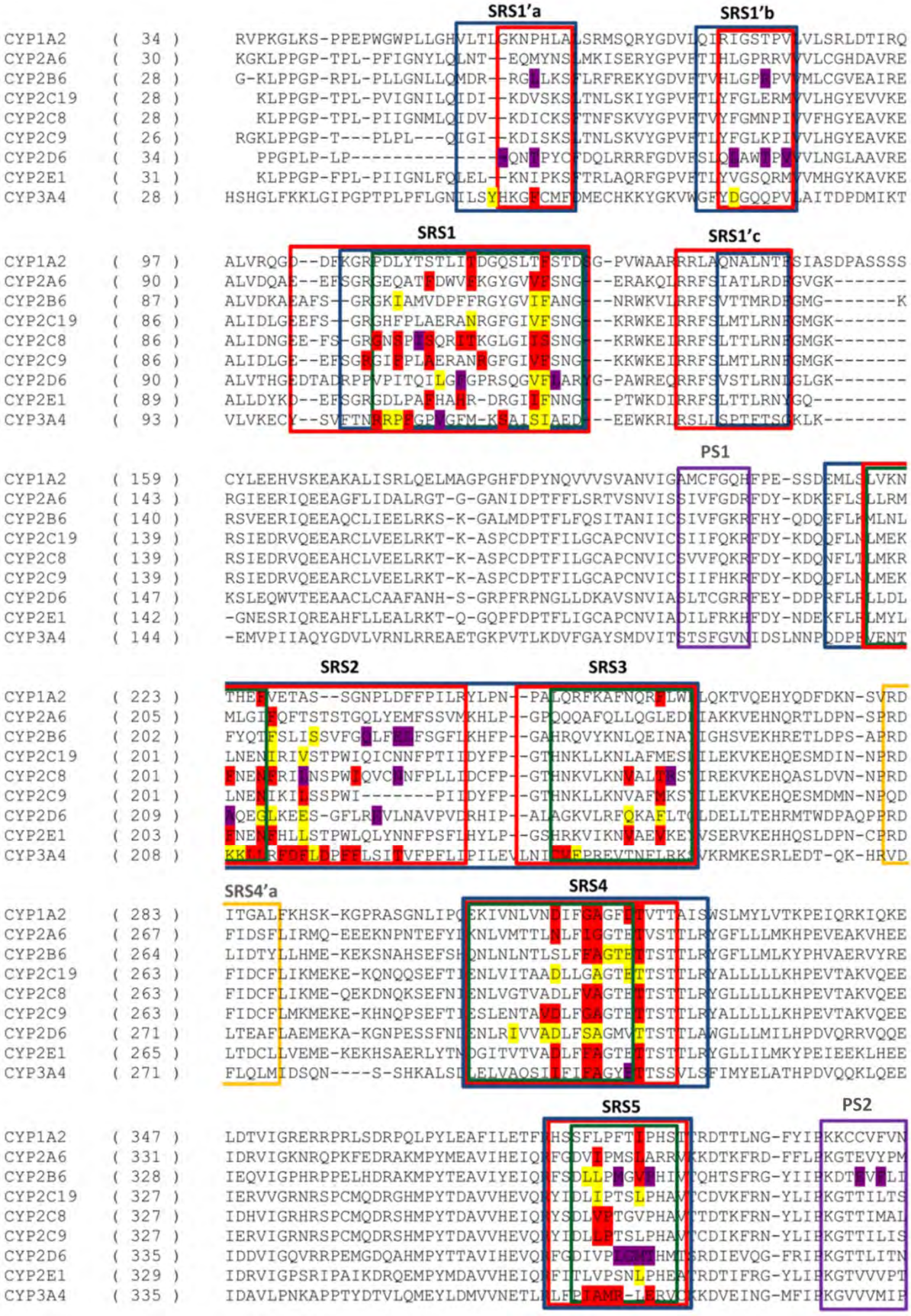
A large number of mammalian CYP450 structures in complex with substrates or inhibitors have been deposited in the RCSB Protein Data Base since Gotoh published the first CYP450 SRS map in 1991, providing direct evidence for specific protein-ligand contacts; most of these are of human CYP450s. There are also several structures of the same enzyme in complex with different ligands for a number of isoforms. For example, CYP3A4 complexes are now available for progesterone, metyraprone [58], erythromycin, ketoconazole [54], bromoergocryptine [227], ritonavir [224] and 8 different ritonavir analogues [57, 225, 226].

Furthermore, compared to Gotoh in 1992, we now have access to bioinformatic tools that rapidly and accurately determine protein-ligand interactions in protein complexes (e.g. LIGPLOT [395] and Zamora's molecular interaction fields [396]) as well as databases dedicated to storing and comparing these protein-ligand interactions (e.g. CREDO [397, 398] and PLI [399]). Furthermore, docking and molecular dynamic tools can today be used in combination with site of

metabolism predictions [400, 401] to predict the binding modes of substrates for which there is presently no structural data available.

Using a combination of new mammalian CYP450 structural data and *in silico* docking experiments this Chapter defines and describes an expanded and more complete mammalian CYP450 SRS map. This Chapter includes work done in collaboration with Prof Yap Chun Wei from the Department of Pharmacy at the University of Singapore that was published in Current Drug Metabolism in 2011 ([402], full manuscript appended at the end of Thesis). New insights into substrate recognition sites based on substrate/inhibitor mammalian CYP450 crystal structures published after the manuscript was completed are also described. My contribution to the experimental work described in this Chapter included identification of protein-ligand interactions in crystal structures of CYP450 complexes, delimitation of the "X-ray structures SRS map" and "periphery site map" as well as the analysis of new structures published after the manuscript was complete. The docking experiments were performed by Lim Yen Ching and Yap Chun Wei and analysis of the docking results and generation of the "Docking CYP450 SRS map" was done by Alexander Zawaira from the Blackburn Laboratory.

2.2 Results



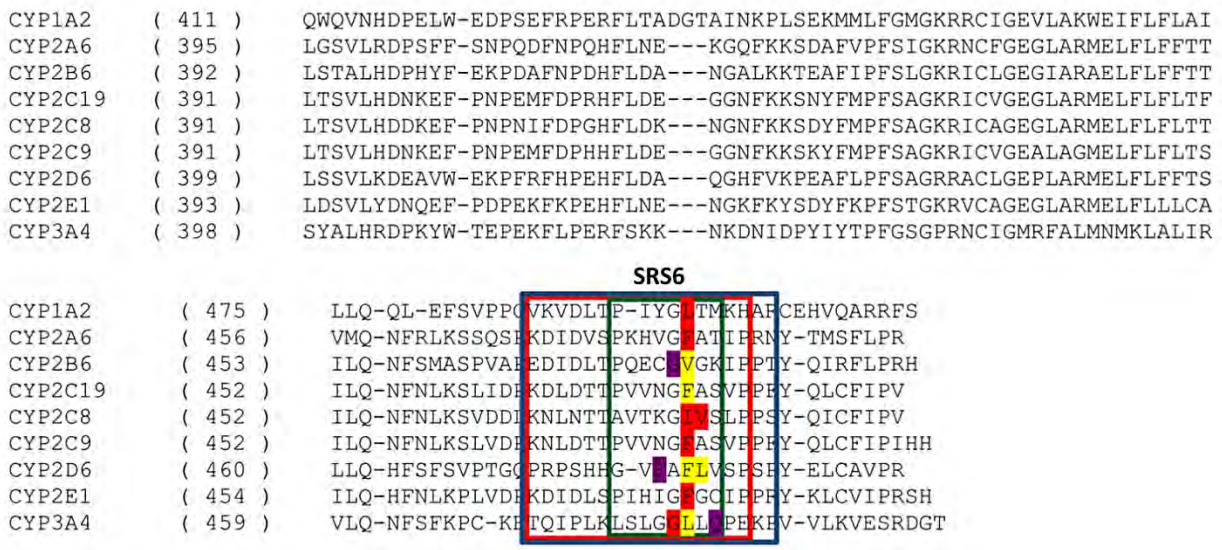


Figure 2.2 Structural alignment showing the contact residues and SRS maps for the major human drug metabolising CYP450 isoforms. Ligand contact residues in CYP450-ligand complexes available in 2011 delimited by LIGPLOT are highlighted in red. Contacts from new CYP450-ligand complexes published after 2011 are shown in yellow. Residues making contact with a substrate/inhibitor molecule bound at a periphery site are highlighted in magenta. Gotoh's SRS map is boxed in green, the X-ray structures SRS map in red and the docking SRS map in blue. The new proposed SRS 4'a region is boxed in yellow and periphery site regions (PS1 and PS2) are boxed in magenta. The demarcated SRS regions are mapped from the sequence alignment to the structural alignment based on CYP3A4 structure (1TQN) - boundaries of SRS regions will vary slightly for some isoforms depending on whether a sequence or structural alignment is used.

2.2.1 The X-ray structures CYP450 SRS map

Mammalian CYP450 crystal structures in complex with substrate or inhibitor molecules provide direct evidence for substrate binding residues. The 43 inhibitor or substrate bound mammalian structures available in 2011 were therefore used to delimit the new SRS map based on X-ray structures as follows. The majority of substrate/inhibitor molecules in these structures are bound at the catalytic site with either a known or predicted site of metabolism or a lone pair of electrons within 5 Å of the haem iron. Residues making contact with a molecule bound to a catalytic site were determine by LIGPLOT [395] and mapped onto a sequence alignment of mammalian CYP450s (Appendix A, Figure A1). Assuming that residues at the equivalent positions in each isoforms may also be substrate contact residues, the positions of interacting residues were inferred onto the other aligned structures and boundaries of SRS regions were delimited 3 residues on either side of identified contact residues. Figure 2.2 shows the contact residues identified within the major human CYP450 drug metabolising isoforms mapped onto a structural alignment and highlighted in red. The "X-ray structure SRS map" determined from the full mammalian CYP450 alignment is boxed in red and, for comparison, Gotoh's original SRS map

is boxed in green. The new map can be seen to incorporate Gotoh's map as a subset, but it has further revealed, amongst other things, two new SRS regions near the N-terminal of the enzyme, SRS1'a and SRS1'b. All SRS regions have been extended in the new map, particularly SRS2 now incorporating a large section of the F-G loop with only a few residues separating it from SRS3.

Given that it is known that CYP450s metabolise a diverse range of substrates and that isoforms responsible for drug metabolism have broad substrate specificities, it seems obvious that the ligands bound within these crystal structures only represent a small percentage of possible CYP450 ligands. The principal component analysis (PCA) plot shown in Figure 2.3 was generated using PubChem fingerprints as chemical descriptors to compare the distribution of ligands from CYP450-ligand complexes with "total known" human CYP450 substrates in chemical space. While the ligands used to delimit the X-ray structures SRS map fall within all four quadrants of the PCA plot, is its clear that the coverage of the "total known" CYP450 substrate chemical space is low.

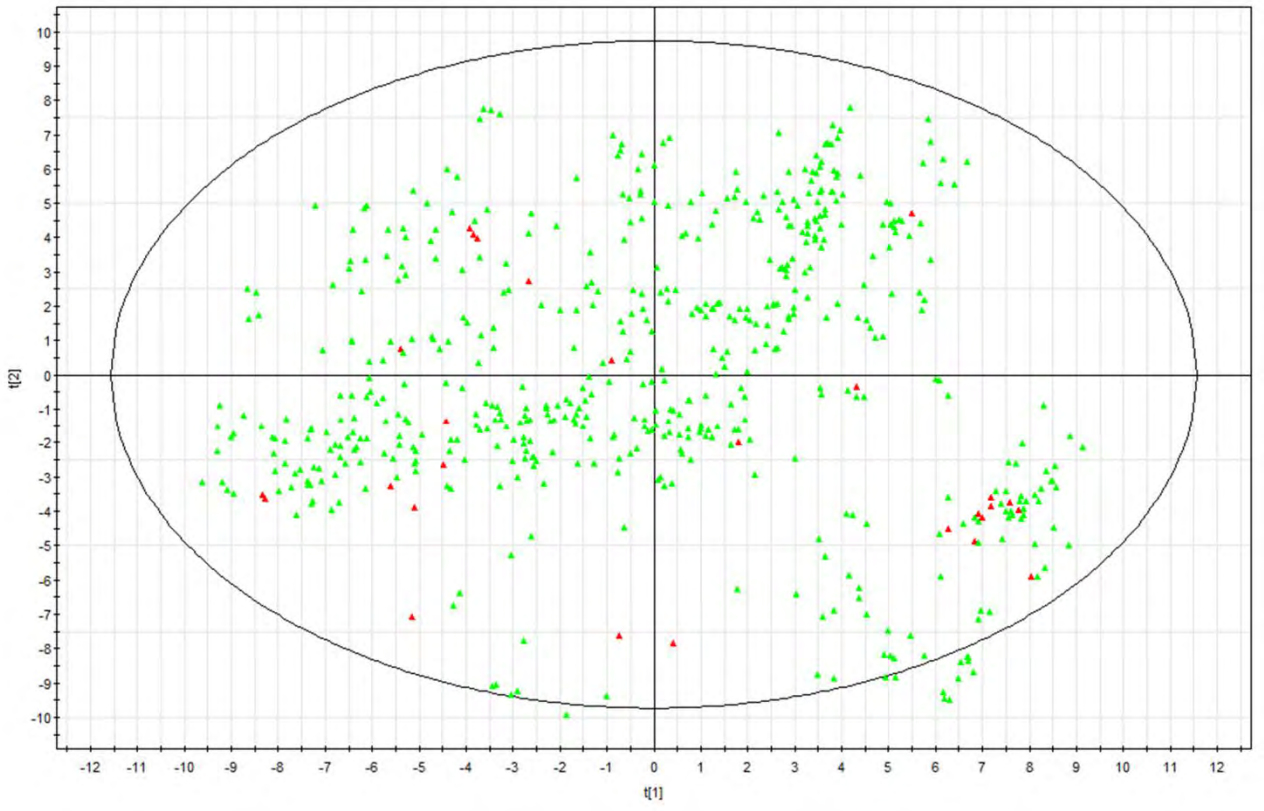


Figure 2.3 PCA plot showing the chemical diversity of CYP450 ligands used to delineate the X-ray structures SRS map. The plot shows the distribution of known CYP450 substrates in chemical space using PubChem fingerprints. Green triangles represent "all known" CYP450 substrates and red triangles represent the ligands used to delineate the X-ray structures SRS map. This PCA analysis was carried out by Prof Wei's group, University of Singapore.

2.2.2 The Docking CYP450 SRS map

In order to address the limited coverage of total known CYP450 substrate space in the PDB, a "docking CYP450 SRS map" was therefore determined by docking 868 known substrates into 10 mammalian isoforms and identifying residues making contact with the docked substrates. Collaborators Lin Yen Chen and Yap Chun Wei carried out docking experiments at the University of Singapore using AutoDock Vina software. The Singapore group used a -1kcal/mole binding affinity cut-off and a 5Å distance cut-off between the site of metabolism and the haem iron for docked complexes in an attempt to include as many substrates-enzyme complexes as possible without including too many incorrect poses. Table 2.1 shows the number of substrates used to generate the docking SRS map for each isoform. The majority of known substrates were successfully docked into the corresponding isoform within the cut-off criteria giving a very high coverage of the "known" CYP450 substrate chemical space for these isoforms.

Table 2.1 The number of known substrates docked into each CYP450 isoforms to generate SRS map

CYP450	No. of known substrates	No. of substrates used to generate docking SRS map	% of known substrates used to generate SRS map
1A2	65	61	94
46A1	11	11	100
2A6	33	26	79
2B4	12	12	100
2C5	5	5	100
2C8	28	27	96
2C9	156	152	97
2D6	212	186	88
2R1	2	2	100
3A4	401	386	96

Using the curated docking poses, contact residues, determined by LIGPLOT, were, thereafter, mapped onto a sequence alignment of the CYP450 isoforms used for the docking experiments (Appendix A, Figure A2; this map was generated by a post-doctoral fellow from the Blackburn group, Dr Alexander Zawaira). The docking SRS map delimited from these contacts is boxed in blue in Figure 2.2. Interestingly, this map closely resembles the X-ray SRS map, the most

notable difference being the merging of SRS2 and SRS3, which suggests that the F-G loop is also important for substrate binding in mammalian CYP450s.

2.2.3 Insights from new crystal structures of mammalian CYP450 complexes

Since the publication of the x-ray and docking SRS maps in 2011, the number of mammalian CYP450 crystal structure complexes available in the PDB has more than doubled enabling verification of aspects of the docking SRS map. Crystal structure complexes are now also available for the following human isoforms: CYP2C19, CYP2D6, CYP1A1, CYP11B2, CYP7A1, CYP17A1 and CYP1B1. New ligand contact residues identified for the human CYP450s in the structural alignment in Figure 2.2 are highlighted in yellow. Notable, all new contacts for these isoforms, except for one, fall within the X-ray structures SRS map. In particular, new contact residues within the newly delimited SRS1'a and SRS1'b confirm the role of these new regions in substrate binding whilst CYP3A4 Tyr 53 - which falls on the boundary of SRS1'a, outside of the X-ray SRS structures map but within the docking SRS map - provides physical evidence for the extension of SRS1a' region.

Other new mammalian structures were also examined for new binding modes that might lead to the identification of new or extended X-ray SRS regions or confirm the extended regions in the docking SRS map. For example, the crystal structure of human CYP7A1 bound to cholest-4-en-3-one (Figure 2.4) and 7-ketocholesterol (PDB codes 3SN5 and 3V8D respectively) revealed a new substrate binding mode with a contact residue falling outside both the X-ray SRS map and the docking SRS map. Contact residues Phe 129 and Ile 125 in Figure 2.4 lie within the newly delimited SRS1'c region, SRS1'c was defined based on contact residues identified in CYP2B4 (Appendix A, Figure A1), thus demonstrating that SRS1'c is used by at least one other isoform. Arg 260 however falls between SRS3 and SRS4 (Figure 2.4), in a region that has not previously been identified as a SRS region in crystal structures or docking studies; this new putative SRS region, labelled SRS4'a, has been mapped onto the alignment in Figure 2.2.

Notably though, all other new mammalian structures showed similar binding modes to those previously observed with contact residues falling within the delimited SRS regions, thus serving broadly to substantiate the x-ray and docking SRS maps.

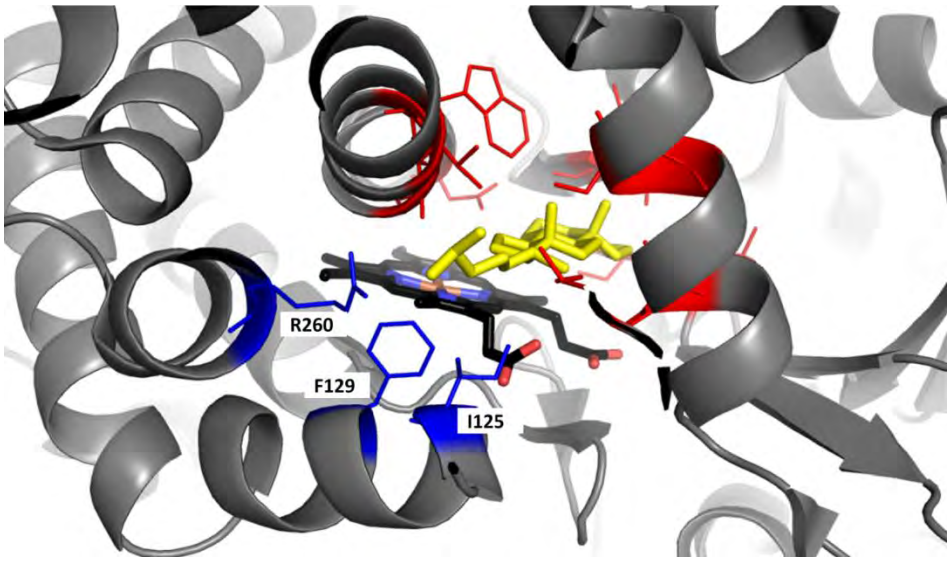


Figure 2.4 Structure of CYP7A1 in complex with cholest-4-en-3-one showing a new substrate binding site. Residues in red make contact with cholest-4-en-one (yellow) and fall within Gotoh's SRS map. Residues labelled and coloured in blue make contact with the substrate but fall outside Gotoh's SRS map. F129 and I125 fall within SRS1c' in the X-ray SRS map whereas R260 lies between SRS3 and SRS4.

2.2.4 Mammalian CYP450 periphery sites

There is a large amount of evidence, including crystallographic data, showing that multiple substrate/inhibitor molecules can bind within the active site. Binding of addition molecules within the active site can affect the enzyme conformation and substrate binding affinity and may result in atypical kinetic profiles commonly observed for drug metabolising CYP450 isoforms [61]. In addition, substrate binding has been described as a multistep process and residues on the periphery of the catalytic binding site are thought to form an initial binding site important for substrate specificity in some isoforms [88, 89]. Consequently, residues making up periphery sites within the active site are likely to play a role in substrate recognition and mutations at these positions may lead to altered drug metabolism.

Binding of substrates to periphery sites were not considered in the docking studies described above as there is no way to determine categorically which docking solutions are correct. By contrast, crystal structures with multiple ligands bound to the active site provide useful insight into the residues important for the binding of multiple molecules. In this analysis, only ligands bound within the active site or in cavities clearly linked to the active site in crystal structures were considered; molecules bound at separate sites on the surface or at dimer interfaces were not

considered. For example, Figure 2.5 shows two retinoic acid molecules bound to the active site of CYP2C8: one bound to the catalytic site and one bound to a periphery site.

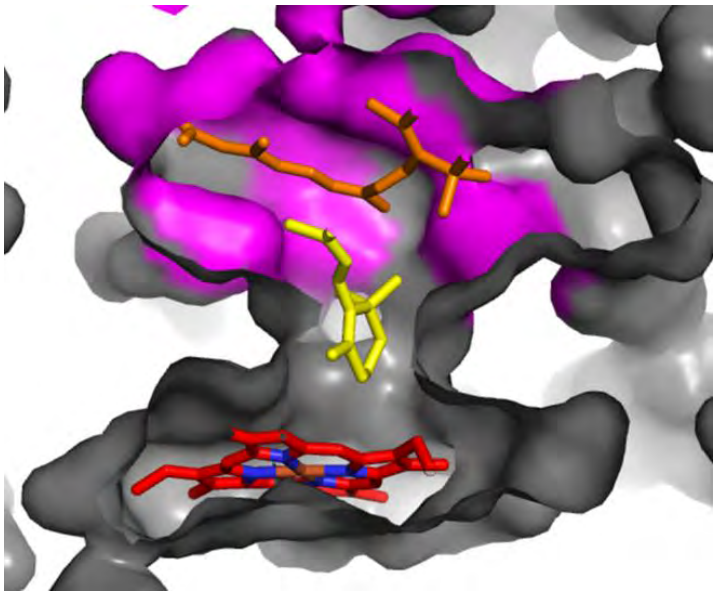


Figure 2.5 Cross section through the cavities within the structure of CYP2C8, PDB code 2NNH, showing two molecules of retinoic acid bound to the active site. The molecule bound at the catalytic site is shown in yellow and the molecule bound at the periphery site, shaded in purple, is shown in orange. The haem group is shown in red.

Residues making contact with ligands bound at a periphery site, and which were not previously identified as ligand contacts, are highlighted in purple in Figure 2.2. Interestingly, the majority of the periphery site contact residues identified by this approach fall within the delimited SRS regions. However a new periphery site region lying outside the main SRS regions was identified in the crystal structure of human CYP2B6 in complex with two amlodipine molecules, with one bound to the haem iron and one extending from the active site into the substrate access tunnel (Figure 2.6 A); here, Glu 387 and Phe 389 within strand 2 of β-sheet 3, between SRS5 and SRS6, interact with the second amlodipine molecule via hydrophobic interactions. This region has been designated periphery site 2 (PS2) and is boxed in purple in the alignment in Figure 2.2.

Periphery site contacts at structurally equivalent positions to CYP2B6 Glu 387 were also found to be present in CYP2B4 in complex with two amlodipine molecules and in human CYP8A1 with minoxidil bound at the catalytic site and β-octylglucoside bound at a periphery site (Figure 2.6 B and C).

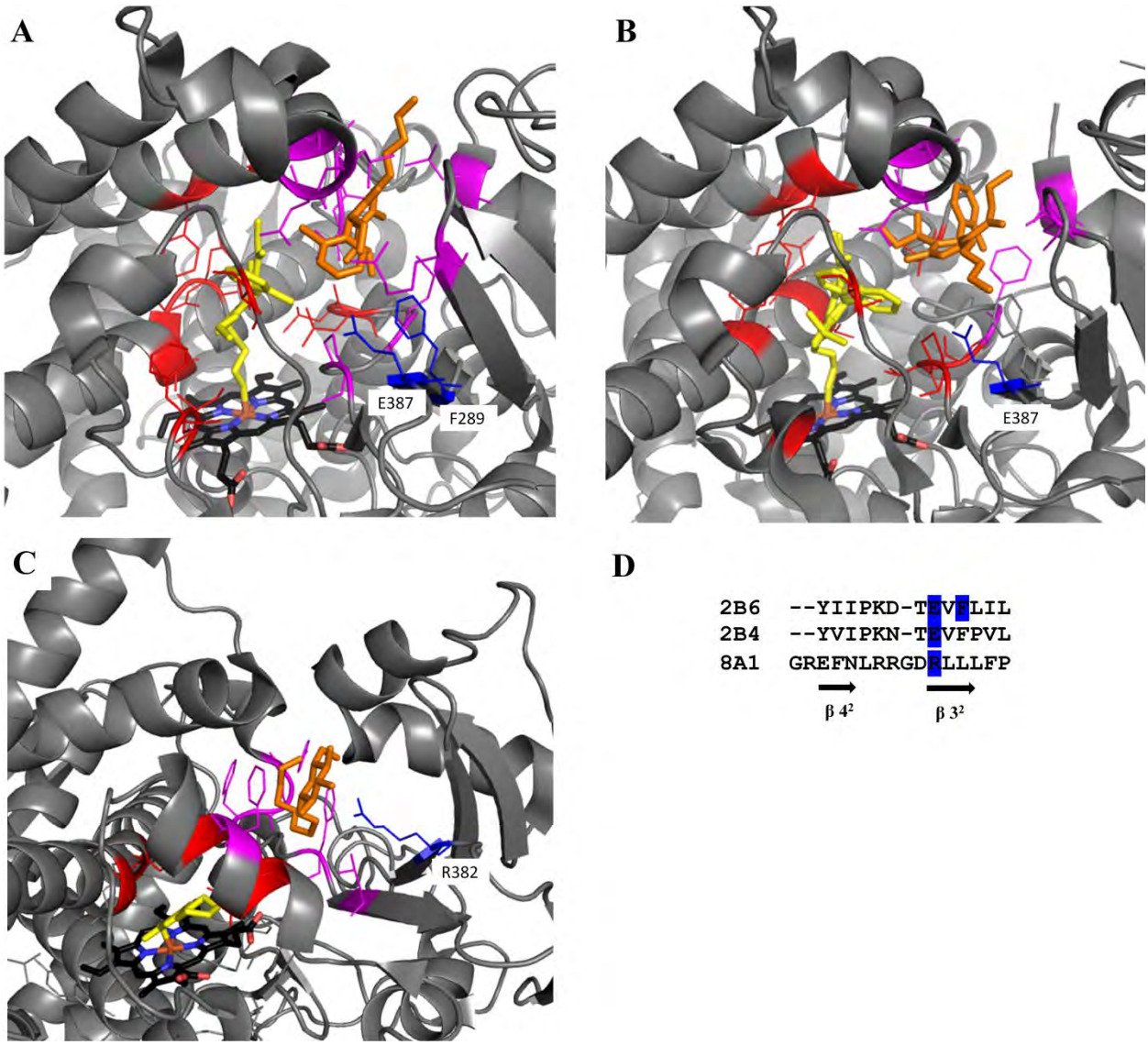


Figure 2.6 A) Structure of CYP2B6 bound to two amlodipine molecules, PDB code 3UA5. B) Structure of CYP2B4 bound to two amlodipine molecules, PDB code 3TMZ. C) Structure of CYP8A1 bound to the inhibitor minoxidil, PDB code 3B6H. Residues making contact with the molecule in yellow bound at the catalytic site are shown in red and residues in contact with the molecule in orange bound at the periphery site are shown in magenta. The amino acids labelled and coloured in blue form part of the secondary periphery site but fall outside of previously delimited SRS regions. D) A portion from a structural alignment of CYP2B6, CYP2B4 and CYP8A1 showing the position of these periphery site contact residues falling within strand 2 of β-sheet 3 in between SRS 5 and 6.

The structure of CYP2B4 in complex with three molecules of the antifungal drug bifonazole - one at the catalytic site and the other two at periphery sites on opposite ends of the active site cavity - also revealed two additional contact residues at positions lying outside the SRS map (Figure 2.7): Gln 45 lies on the boarder SRS1'a in the X-ray SRS map while Phe 184 lies at the end of α-helix between SRS1'c and SRS2; this region has been designated periphery site1 (PS1) in the alignment in Figure 2.2. A surface rendering of the CYP2B4 structure shows it to be in an

unusually wide-open conformation; despite this and the binding of two drug molecules at periphery sites over 10 Å apart, however, all other ligand contact residues fall within the main SRS map.

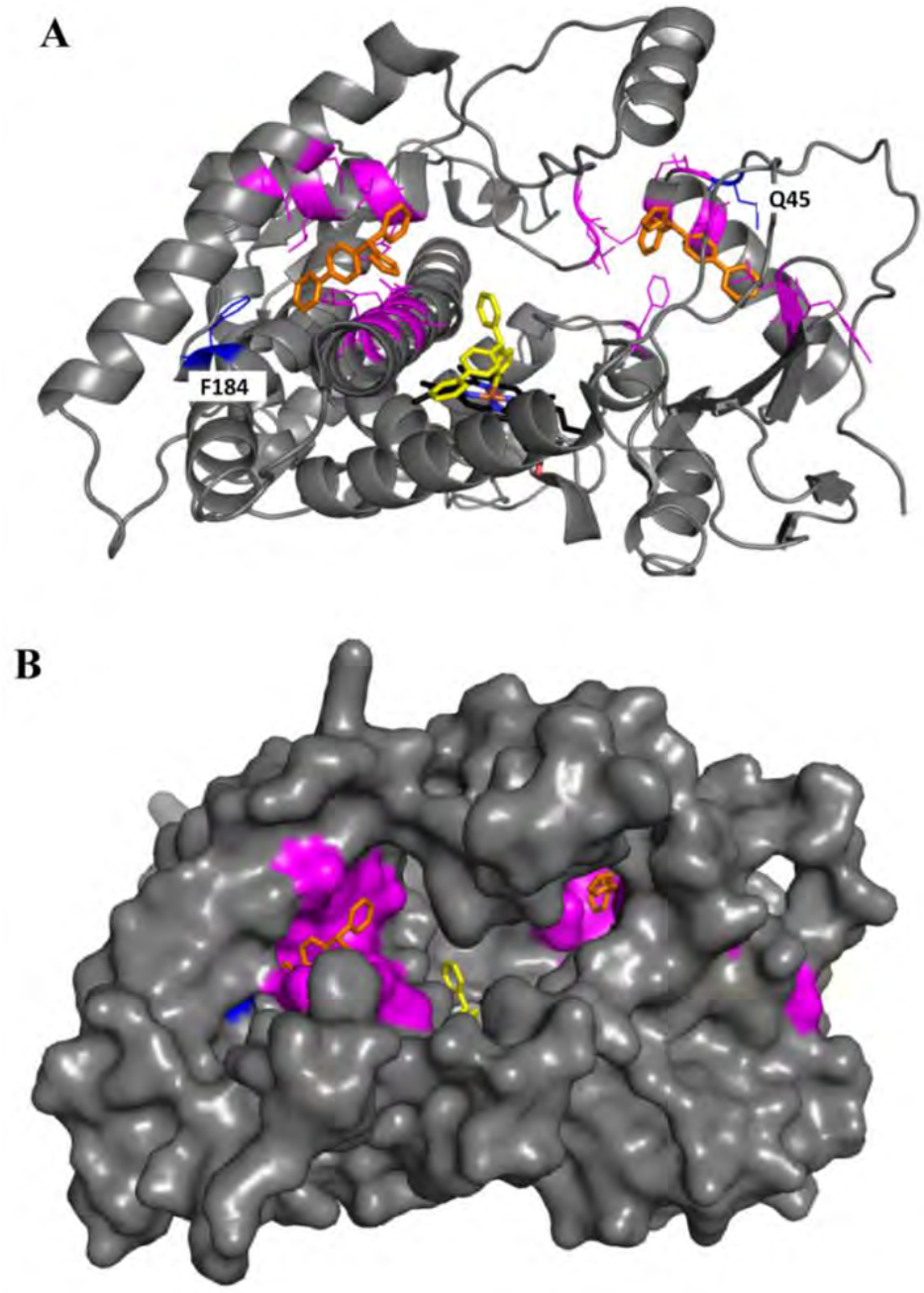


Figure 2.7 Structure of rabbit CYP2B4 in complex with three bifonazole molecules (antifungal drug), PDB code 2BDM. A) The bifonazole molecule bound at the catalytic site is shown in yellow and bifonazole molecules bound at periphery sites are shown in orange. Periphery site contact residues lying within the X-ray structures SRS map are shown in magenta while contact residues Q45 and F184, lying outside the SRS regions, are labelled and coloured in blue. B) Surface view of the CYP2B4 complex showing the unusual open conformation.

2.2.5 Structural analysis of the newly-delimited SRS regions based on the X-ray structures and docking SRS maps

A structural analysis of each of the new SRS regions was carried out to determine whether they are truly new separate SRS regions or whether they are merely extensions of existing SRS regions.

SRS1'a, SRS1'b and SRS1'c

Figure 2.8 shows the relationship between SRS1'a, SRS1'b, SRS1 and SRS1'c. SRS1'a and SRS1'b are separated by residues that are removed from the active site cavity and shielded from the active site by other parts of the protein suggesting that SRS1'a and SRS1'b are distinct SRS regions.

In contrast, the region between SRS1 and SRS1'c is not shielded from the active site. A surface rendering of the active site cavity shows that Trp126, which lies 3 position from SRS1 and 3 positions from SRS1'c, forms part of the active site cavity and could therefore also interact with a bound substrate.

To test the possibility that this region could make contact with the substrates bound to the active site, additional docking studies were performed by the Singapore group using CYP2B4 structure 2BDM, where the CYP450 shows an unusual open conformation. Nine of the 12 compounds docked into the active site of the CYP2B4 structure made contact with residues in the region between SRS1 and SRS1'c. Following this it seems clear that SRS1'c should be merged with SRS1 into one continuous SRS region.

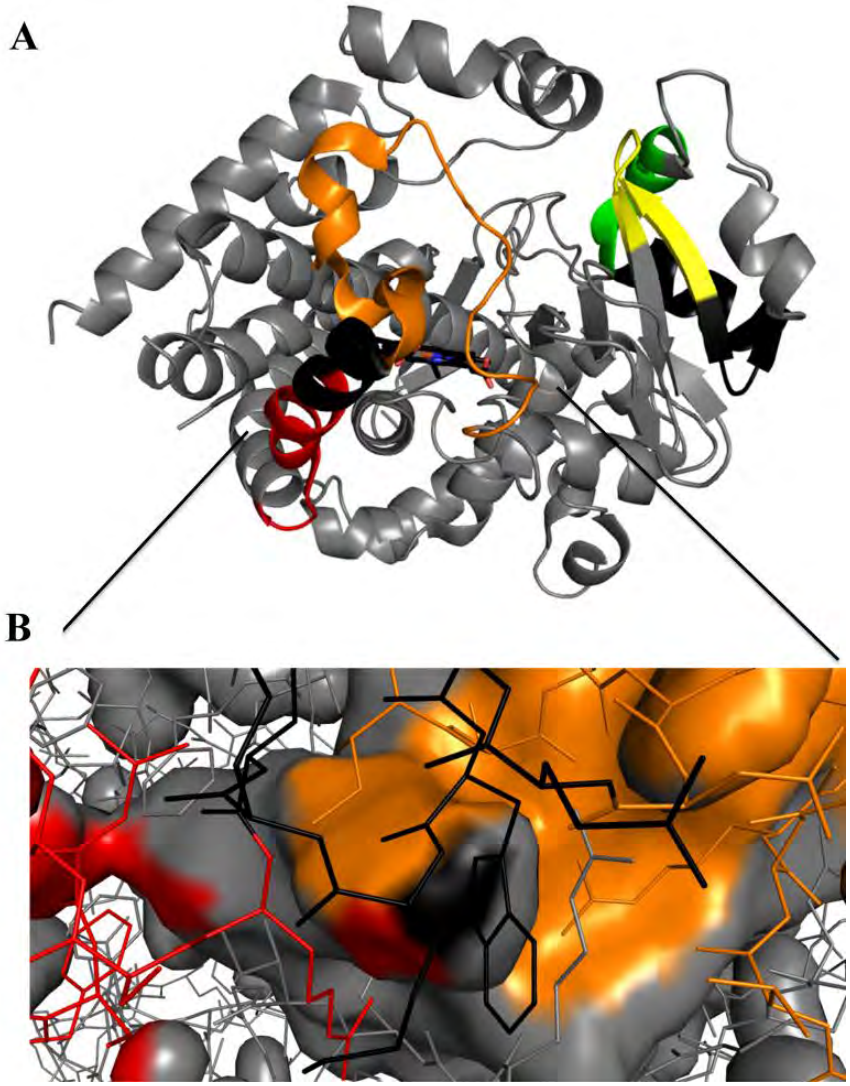


Figure 2.8 The relationship between SRS1'a, SRS1'b, SRS1 and SRS1'c.
A) The structure of CYP3A4, PDB code 2V0M is shown in grey with SRS1'a residues in green, SRS1'b residues in yellow, SRS1 residues in orange and SRS1'c residues in red. The residues that lie between SRS1'a and SRS1'b and between SRS1 and SRS1'c are shown in black. **B)** A surface rendering of the active site cavity showing that residue Trp126 between SRS1 (orange) and SRS1'c (red) forms part of the wall of the active site cavity indicating that SRS1'c is likely to be an extension of SRS1 rather than a new separate SRS region.

Merging of SRS2 and SRS3

The F, G, F' and G' helices and interlinking loop regions make up SRS2 and SRS3. According to the structural alignment in Figure 2.2 there are four to five residues separating SRS1 and SRS2 in the X-ray structures SRS map. The merging of these two regions in the docking map was due to ligands contacts Tyr 225 and His 230 identified in CYP2C9 docking complexes (Figure 2.9). Tyr 225 made contact with vardenafil and His 230 made contact with glibenclamide and vardenafil when these ligands were docked into CYP2C9 structure 1R9O. In CYP2C9

structures 1OG2 and 1OG5, the region between SRS2 and SRS3 is a helical structure bent away from the active site; in contrast this region is a more disordered loop region position closer to the active site in structure 1R9O. A comparison of CYP450 crystal structures shows that the F-G loop is very flexible and can take on different conformations; it is therefore likely that the region between SRS2 and SRS3 does indeed play a role in substrate binding as suggested by the docking studies. Following this, the merged region has been labelled SRS (2,3).

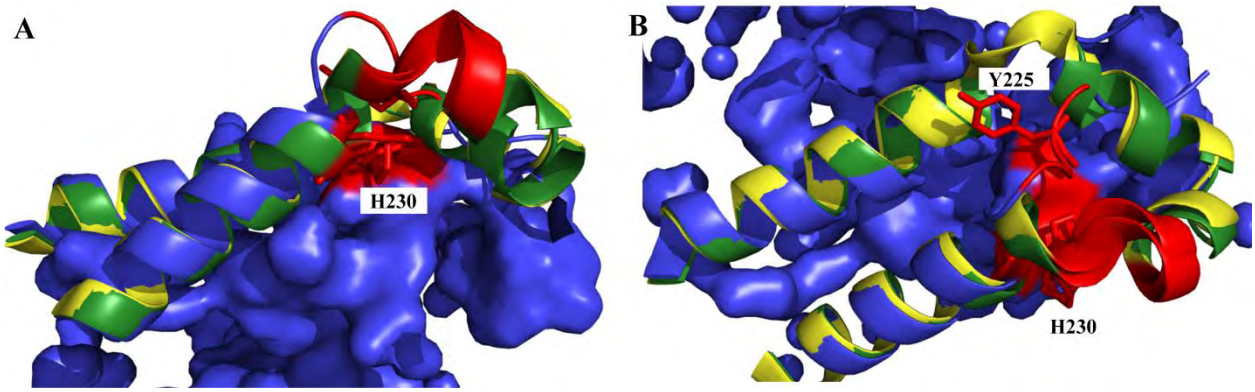


Figure 2.9 Relationship between SRS2, SRS3 and the active site cavity in three different CYP2C9 crystal structures. The SRS2 and SRS3 regions of PDB structures 1OG2, 1OG5 and 1R9O, shown in yellow, green and blue respectively, were aligned in PyMol. The region separating SRS2 and SRS3 and His 230 is shown in red. Substrate contact residues Tyr 225 and His 230 identified in the docking studies and responsible for the merging of SRS2 & 3 are shown as sticks. A surface rendering of the cavities in the enzyme shows that residues Tyr 225 and His 230 form part the active site cavity. A) A side view of the F-G loop. B) A top view of the F-G loop.

SRS4'a

SRS4'a was delimited as a result of a ligand contact within α-helix H that occurred in crystal structures of the human Cholesterol 7 alpha-hydroxylase CYP7A1, however there were no ligand contacts in this region in any of the other mammalian crystal structures or docking complexes. Interestingly, CYP7A1 ligand complexes adopt different conformations to the structures of other drug metabolising CYP450 isoforms observed to date, with the I and H helices bent upwards to accommodate the ligand (Figure 2.10). As crystal structures represent a single snap shot in time, it is difficult to predict whether other isoforms are able to adopt a similar conformation to bind ligands in a similar orientation; it is therefore difficult to predict whether SRS4'a is also used for substrate binding in other isoforms or whether it is unique to CYP7A1.

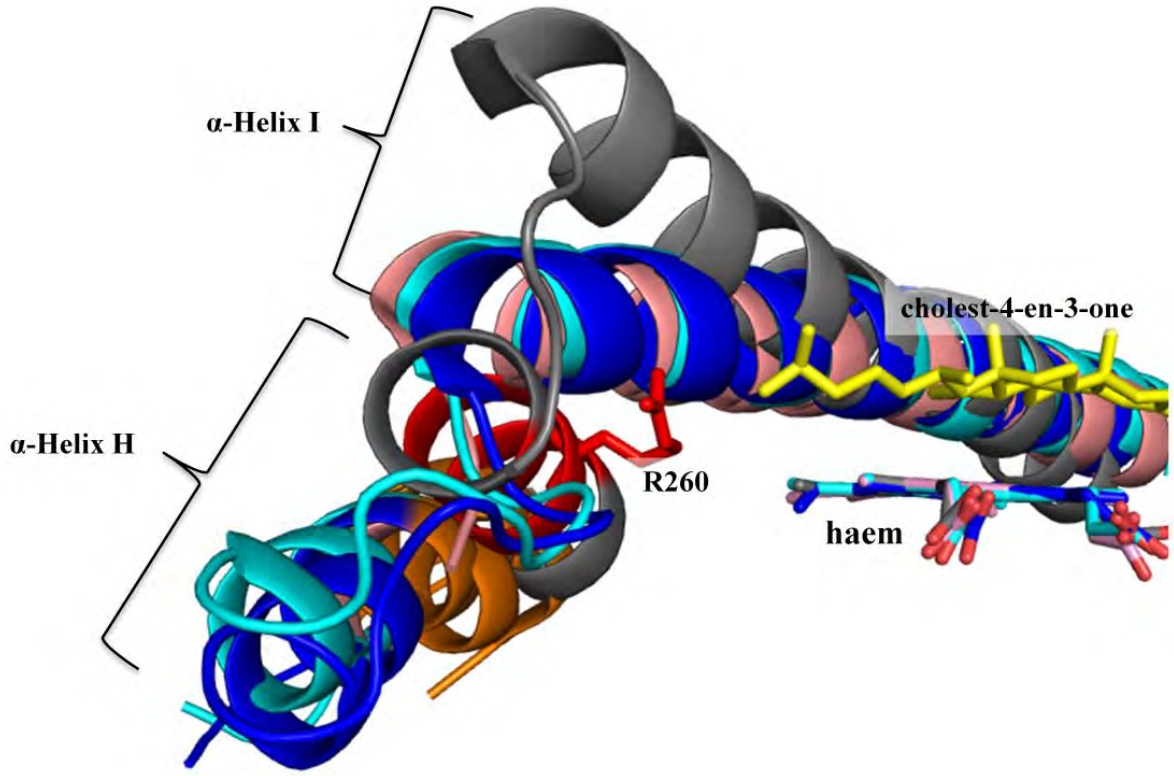


Figure 2.10 Alignment of SRS4'a in CYP7A1 with CYP2C9, CYP2D6 and CYP3A4 structures. The H and I helices of each structures were aligned in PyMOL based on the position of their haem groups. CYP7A1 (PDB code 3SN5, grey) is shown in complex with ligand cholest-4-en-3-one (yellow). CYP7A1 ligand contact Arg 260, responsible for the identification of SRS4'a, is labelled and shown as a stick model. The SRS4'a region is shown in red in CYP7A1 and orange in CYP2C9 (PDB code 1OG5, blue), CYP2D6 (PDB code 2FDQ, cyan) and CYP3A4 (PDB code 1TQN, light pink).

Final Combined SRS map

Based on the structural analysis described above, the combined SRS map is made up of seven distinct regions: SRS1'a, SRS1'b, SRS1, SRS(2,3), SRS4, SRS5 and SRS6 (Figure 2.11). SRS4'a is still considered a putative SRS region since it has only been identified in a single isoform that adopts an unusual conformation.

The combined SRS map covers ~33% of the CYP450 protein sequence. It is important to note however that many of the residues within these regions point away from the active site and are unlikely to make direct contact with substrates. Of the residues falling within SRS regions in CYP2C9, 55 % have side chains pointing away from the active site in the CYP2C9 crystal structure 1OG2. (Figure 2.11). These residues have not however been excluded from the SRS regions due to the dynamic nature of protein structures as well as differences in side chain positions that may exist between isoforms.

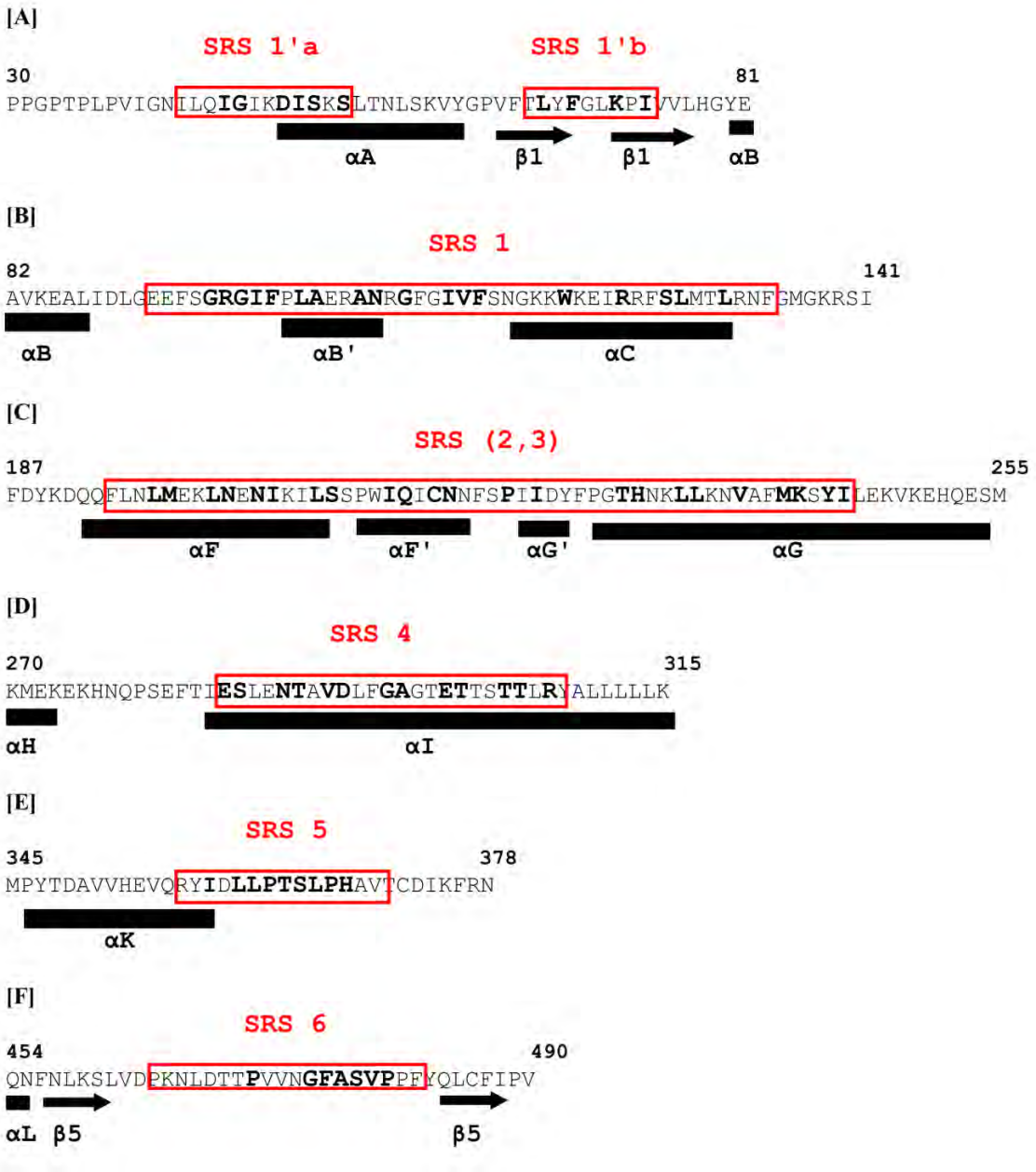


Figure 2.11 The final combined SRS map
Regions of the CYP2C9 sequence that fall within the final SRS regions 1'a to 6 are shown in panels [A] to [F]. SRS zones are boxed and labelled in red and secondary structures are shown below the sequence: black rectangular boxes represent α-helices and black arrows represent β-sheets. Based on visual inspection of the CYP2C9 1OG2 structure, SRS residues with side chains facing towards the active site are shown in bold face while those pointing away are shown in plain font. Numbering of sequence positions in according to the CYP2C9 UniProt sequence P11712.

2.3 Discussion

The work described in this Chapter has resulted in the generation of a new extended mammalian CYP450 substrate recognition map based on protein-ligand contacts delimited in mammalian CYP450 crystal structures and CYP450 docking studies. While the new map includes all regions of the SRS map described by Gotoh in 1991, it is based on a much larger experimental data set and exhibits a number of new features, covering a substantially larger region of the CYP450 sequence. As such this new map represents a significant advance over the largely speculative predictions of Gotoh that were extrapolated from the very limited experimental data set available in 1991.

The new map, based on a combination of the maps derived from crystal structures and docking complexes, cover 33% of the CYP450 sequence, compared to Gotoh's map covering only 16%. The following new and extended features account for this two-fold increase in sequence coverage: (1) Two new regions, SRS1'a and SRS1'b, appear before SRS1 in the new map. (2) SRS1 is extended by ~20 residues. (3) Extension of SRS2 and SRS3 in the new map results in the co-joining of these two regions to form the newly designated SRS(2,3), which is ~ 3 times the size of Gotoh's original SRS2 and SRS3. (4) SRS4 has an N-terminal extension of ~7 residues. (5) SRS5 is extended by ~3 residues. (5) SRS6 is extended by ~ 9 residues. A new putative SRS region between SRS(2,3) and SRS4 has also been identified and designated SRS4'a. This later region resulted from a single ligand contact observed in CYP7A1 crystal structures and consists of 7 residues which would effectively add an additional 1.5% sequence coverage to the SRS map.

Despite the large sequence coverage of the new map, a number of key residue positions were found to interact with the majority of bound substrates: this is particularly apparent in the docking SRS map where many of the residues contacting more than 80% of docked substrates for different isoforms occur at the same position in the alignment (Appendix A, Figure A2). Mutations at these key residue positions are therefore likely to affect the binding of a broad range of CYP450 substrates for a given isoform.

Two periphery site ligand contact regions, responsible for the binding of multiple ligands within the active site were also identified in crystal structure complexes. PS1 lies just N-terminal of SRS2 and PS2 lies C-terminal of SRS5. All other periphery site contacts fell on the boundaries of, or within, SRS regions.

While we can be confident of the ligand poses observed in crystal structures, docking algorithms are not very good at accurately predicting binding energies; furthermore, docking molecules into large flexible CYP450 sites is particularly challenging. This raises concerns that incorrect docking modes may lead to the incorrect assignment of ligand contact residues and an overestimation of the sequence coverage of the SRS map. The X-ray structures SRS map and the docking SRS map were however found to be very similar, despite the large increase in the range and diverse of chemical structures included in the docking complexes; importantly, the new SRS1'a, SRS1'b and SRS1'c (SRS1'c was combined with SRS1 following a structural analysis of the two regions and further docking studies) appeared in both maps, effectively cross-validating the data generated by the two approaches. The main additional insight gained from the docking study was the co-joining of SRS2 and SRS3, still separated by ~5 residues in the X-ray structures map. This suggests that a relatively small number of ligands sample most regions of the SRS map.

Looked at from a different perspective, the lack of new contact residues identified in the docking studies may be a result of docking ligands into static structures, thereby limiting protein conformations to those observed in crystal structures. Since the new combined SRS map lines the majority of the active site cavity observed in crystal structures, the CYP450 would need to adopt a different conformation for new contacts to be identified. As an example of this, the CYP7A1 structure in complex with cholest-4-en-3-one (Figure 2.4) reveals a new CYP450 conformation and binding mode that led to the identification of a new SRS region. Despite the wealth of CYP450 structure data available, it is still therefore conceivable that we have not observed all possible CYP450s ligand-binding conformations, particularly with regard to ligands binding at periphery sites or sites completely separate from the active site cavity.

Table 2.2 Examples of "non-Gotoh-SRS" residues affecting substrate binding in mammalian CYP450 enzymes.

CYP450	Residues	Position relative to SRS map	Activity affected	Reference
CYP1A2 (rabbit)	(67-78) (L(S)QQY-GDVLQIR)	N-term of SRS1	delimited using mechanism based inactivators	[403]
CYP1A2 (rat)	(175-184) FQELN-AAVGR	between SRS1 & SRS2		
CYP2B1	Leu58	N-term of SRS1	stereo- & regio-selectivity of steroid hydroxylation	[404]
CYP2B1	Ser407 Asn417 Ala419	between SRS5 & SRS6	inhibition by acetylenic compounds	[405]
CYP2B5	His120 Pro221	C-term of SRS1 between SRS2 & SRS3	progesterone hydroxylation	[393]
CYP2C11	Val4 Phe187	N-term of SRS1 between SRS1 & SRS2	testosterone hydroxylation	[394]
CYP2C19	Pro220 Thr221	between SRS2 & SRS3	omeprazole hydroxylation	[385]
CYP3A12	Thr187	between SRS1 & SRS2	steroid hydroxylation	[386]

The information in this table was collated by Dr Zawaira

Despite these caveats, the new SRS map and periphery site maps offer a more comprehensive overview of substrate binding than Gotoh's original map and, moreover, they are much more extensively rooted on experimental data sets, so they should aid in predicting the effects of mutations on substrate binding. Importantly, the new map is able to explain the substrate specific effects observed for a significant number of "non-Gotoh-SRS" residues in the literature (Table 2.2), including:

1) A study by Ibeanu *et al* showed that CYP2C19 residues Pro 220 and Thr 221were important determinants of omeprazole hydroxylase activity [385]. These residues lie within the F-G loop and SRS(2,3), in the region co-joining Gotoh's SRS2 and SRS3. SRS(2,3) acts as a "lid" over the active site cavity and is comprised of the F and G helix connected by the F-G loop. The F-G loop is much longer in mammalian CYP450s than in bacterial CYP450s (more than 2 times the length) and is highly flexible, adopting very different conformations in different structures of the same isoform. In some structures this region is comprised of two short F' and G' helixes (e.g. CYP2C9 structure 1OG2) while in other structures it is a disordered loop region (e.g. CYP2C9 1R9O). The F-G loop regions forms part of many of the major substrate access channels (see Chapter 3) and when the enzyme is in a closed conformation comes in close proximity with the active site, thus playing an important role in substrate recognition in mammalian CYP450s.

2) He *et al* showed that CYP2B5 residue Pro 221, another non-Gotoh-SRS residue position in the F-G loop and SRS(2,3), plays a role in progesterone hydroxylase activity [393].

3) Hu *et al* also identified CYP2B5 "non-Gotoh-SRS" residue His 120 as an important determinant of progesterone hydroxylase activity; this residue falls within the extended SRS1 region in the new map.

4) Yun *et al* identified rabbit CYP1A2 residues 67-78 (L(S)QQYGDVLQIR) and rat CYP1A2 residues 175-184 (FQELMAARVGR) as substrate recognition regions using mechanism-based inactivators. Rabbit CYP1A2 residues 76-78 (QIR) fall within the newly designated SRS1'b, which is consistent with these experimental results. Rat CYP1A2 residues 175 to 184 lie between SRS1 and SRS(2,3); however, while there is no crystal structure for rat CYP1A2, the homologous residues in a human CYP1A2 structure occur at the surface of the enzyme, far away from the active site (Figure 2.12). It appears that these residues were probably mis-identified, perhaps as a result of the reactive product diffusing away from the active site before reacting and thus binding randomly to a good nucleophile rather than to residues in close proximity to its binding site.

5) Separate studies identified canine CYP3A12 residue Thr 187 and rat CYP2C11 residues Phe 187 and Val 4 as important residues for steroid hydroxylation [386, 394]. CYP3A12 Thr 187 and CYP2C11 Phe 187 are not homologous residues, aligning with CYP3A4 residues Thr187 and

Phe189 respectively, however they both lie upstream of SRS(2,3) in the newly delimited periphery site 1 (PS1) region. This supports the proposition that these periphery site regions are important for catalytic activity and substrate recognition. The closely related human CYP3A4 and CYP2C9 enzymes both display atypical kinetics for some substrates and are able to bind multiple substrates within the active site [61]. Following this, it is plausible that CYP3A12 residue Thr 187 and CYP2C11 residue Phe 187 residues affect the binding of a second steroid molecule within the active site. CYP2C11 Val 4 falls within the N-terminal hydrophobic membrane-binding domain and does not form part of the catalytic domain so mutations at this position must influence the catalytic activity indirectly through another mechanism.

Figure 2.12 Structure of human CYP1A2 showing the position of residues 176-185 (LQELMAGPGH) in red. These residues are homologous to rat CYP1A2 residues 175-184 (FQELMAARVGR) which were shown to be important for substrate recognition, however they are far removed from the active site.

"Non-Gotoh-SRS" CYP2B1 residues Leu 58, Ser 407, Asn 417 and Ala 419 identified as residues important for catalytic activity [404, 405] fall outside the new SRS map. Leu 58 falls four residues downstream of SRS1'a in the multiple sequence alignment but there is no crystal structure of CYP2B1 available yet so it is not possible to tell if this residue would fall within the same structural environment as SRS1'a in a structural alignment. Von Weymer *et al* showed that the substitution of CYP2B1 residues Ser 407, Asn 417 and Ala 419 to equivalent residues in CYP2B2 prevented inactivation by two mechanism-based inhibitors. These residues occur between SRS5 and SRS6; Ser 407 lies just before the conserved meander region important for

enzyme stability and Asn 417 and Ala 419 are part of the haem binding region. It is therefore possible that the experimental observations are a result of secondary effects on catalysis such altered enzyme stability and conformation or changes to the haem environment.

Overall it is encouraging that the new SRS and periphery site maps are able to explain seven of the twelve examples of "non-Gotoh-SRS" residues or sequences affecting substrate recognition found in the literature, exemplifying the predictive power of the new maps.

2.4 Methods

2.4.1 Structures

Ligand bound CYP450 structures were retrieved from the RCSB Protein Data Bank (PDB) [31]. Structures deposited in the PDB before 2011 were as follows: **CYP1A2** (*Homo sapiens*): 2HI4.**CYP2A6** (*Homo sapiens*): 3EBS, 1Z11, 2FDU, 2FDV, 2FDW and 2FDY. **CYP2A13** (*Homo sapiens*): 2P85. **CYP2B4** (*Oryctolagus cuniculus*): 2Q6N, 2BDM, 1SUO, 3KW4 and 3ME6. **CYP2B6** (*Homo sapiens*): 3IBD. **CYP2C5** (*Oryctolagus cuniculus*): 1N6B and 1NR6. **CYP2C8** (*Homo sapiens*): 2NNJ, 2NNI, 2NNH and 2VN0. **CYP2C9** (*Homo sapiens*): 1OG5 and 1R9O. **CYP2E1** (*Homo sapiens*): 3GPH, 3KOH, 3LC4, 3E4E and 3E6I. **CYP2R1** (*Homo sapiens*): 3C6G, 3DL9 and 3CZH. **CYP3A4** (*Homo sapiens*): 1W0F, 1W0G, 2J0D, 2V0M and 3NXU. **CYP46A1** (*Homo sapiens*): 2Q9F, 3MDM, 3MDR, 3MDT and 3MDV. **CYP19A1** (*Homo sapiens*): 3EQM. **CYP8A1** (*Homo sapiens*): 3B6H. **CYP11A1** (*Homo sapiens*): 3NA0.

Structures deposited in the PDB between Jan 2011 and April 2014 ("New structures") were as follows: **CYP1A1** (*Homo sapiens*): 4I8V. **CYP1B1** (*Homo sapiens*): 3PM0. **CYP2A6** (*Homo sapiens*): 3T3R, 4EJJ. **CYP2A13** (*Homo sapiens*): 4EJG, 4EJH, 4EJI, 3T3S. **CYP2B4** (*Oryctolagus cuniculus*): 4JLT, 4H1N, 3UAS, 3TMZ, 3TK3, 3R1A, 3R1B. **CYP2B6** (*Homo sapiens*): 3UA5, 4I91, 3Q0A, 3QU8. **CYP2C19** (*Homo sapiens*): 4GQS. **CYP2D6** (*Homo sapiens*): 3QM4, 3TDA, 3TBG. **CYP2E1** (*Homo sapiens*): 3T3Z. **CYP3A4** (*Homo sapiens*): 4K9T, 4K9U, 4K9V, 4K9X, 4K9W, 3TJS, 4I4G, 4I4H, 3UA1. **CYP7A1** (*Homo sapiens*): 3V8D, 3SN5. **CYP11A1** (*Homo sapiens*): 3N9Y. **CYP11A1** (*Bos taurus*): 3MZS. **CYP11B2** (*Homo sapiens*): 4DVQ, 4FDH, 3N9Z, 3NA1. **CYP17A1** (*Homo sapiens*): 3RUK, 3SWZ. **CYP19A1** (*Homo sapiens*): 4KQ8, 4GL5, 4GL7, 3S79, 3S7S. **CYP46A1** (*Homo sapiens*): 4J14, 4FIA, 4ENH.

All CYP450 structural images in this Chapter and throughout this Thesis were generated using PyMOL [406]

2.4.2 Identifying Protein-ligand interactions in crystal structures of CYP450 complexes

Both LIGPLOT [395] and Zamora's identifies Zamora's molecular interaction fields (MIFs) [396] were tested as methods to identify residues interacting with ligands in crystal structures. LIGPLOT identifies hydrophobic interactions and hydrogen bonding between amino acids and the specified ligand while Zamora's method calculates the interaction energies between the amino acids and the ligand. Identical residues were identified using LIGPLOT (in default mode) and Zamora's MIFs (interacting residues with MIF-calculated interaction energies ≤ 1 kcal/mol). Following this, for simplicity only LIGPLOT was used (in default mode) to identify residues interacting with bound ligands in all CYP450 complexes deposited in the PDB before 2011 and CYP2A6, CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and CYP3A4 complexes deposited between Jan 2011 and April 2014. All other new mammalian structures were aligned in PyMOL and examined for new/unusual ligand binding poses, making use of regions that fall outside of the previously defined X-ray structures SRS map. Contact residues for structures showing unusual binding poses were subsequently identified using LIGPLOT.

Substrate/inhibitor molecules were considered to be bound to the catalytic site only if there was a known or predicted site of metabolism (determined by Dr Zawaira using the program MetaSite [401]) or, in the case of an inhibitor, a lone pare of electrons within 5 Å of the haem iron. Using these criteria, substrate/inhibitor molecules bound within the active site but not at the catalytic site were identified as molecules bound to a periphery site and interactions with these ligands were delimited as periphery site contact residues rather than catalytic site contact residues. CYP450 structures in complex with the following ligands fell into the "periphery site category": (CYP2B4, 2BDM, TM1 502 and TM1 503); (CYP2B6, 3UA5, OX6 502); (CYP2C8, 2NNH, REA 1502); (CYP2D6, 3TBG, RTZ 1 and 2); (CYP3A4, 2V0M, KLN 1501); (CYP3A4, 3K9T, 1RD 602); (CYP3A4, 4K9U, 5AW 503); (CYP8A1, 3B6H, BOG 701). Ligands bound to the surface of the CYP450 substrates or sites completely separate from the main active site cavity were not consider as periphery sites. Surface renderings of the cavities within the structure were visualised in PyMOL to determine whether sites fell within the active site cavity.

2.4.3 PCA of the chemical diversity of CYP450 substrates using PubChem fingerprints

The principal component analysis of CYP450 substrates chemical descriptors was carried out by collaborators Lim Yen Ching and Yap Chun Wei as follows. A list of known substrates for CYP1A2, CYP2A6, CYP2B4, CYP2C5, CYP2C8, CYP2C9, CYP2D6, CYP2R1, CYP3A4 and CYP46A1 was compiled from primary literature, Micromedex, Protein Data Bank and other online databases and resources [31, 407-417]. Chemical structures for these substrates were obtained from PubChem database and were processed using Pipeline Pilot Student Edition v6.1.5 Standardize Molecule, Strip Salts, and "Add Hydrogens" modules [418]. PubChem fingerprints [419] for each substrate were computed using PaDELDescriptor v2.4 [420], and a Principal component analysis (PCA) was performed using SIMCA-P+ v12.0.1 [421] to show the distribution of substrates in X-ray CYP450 complexes in the chemical space of "all known substrates".

2.4.4 Sequence and structural alignments

Sequences with the following accession numbers were retrieved from the UniProt Data Base and aligned using ClustalW (default parameters) [422]: CYP1A2, P05177 (*Homo sapiens*); CYP2A6, P11509 (*Homo sapiens*); CYP2A13, Q16696 (*Homo sapiens*); CYP2B4, P00178 (*Oryctolagus cuniculus*); CYP2B6, P20813 (*Homo sapiens*); CYP2C5, P00179 (*Oryctolagus cuniculus*); CYP2C8, P10632 (*Homo sapiens*); CYP2C9, P11712 (*Homo sapiens*); CYP2E1, P05181 (*Homo sapiens*); CYP2R1, Q6VVX0 (*Homo sapiens*); CYP3A4, P08684 (*Homo sapiens*); CYP46A1, Q9Y6A2 (*Homo sapiens*); CYP19A1, P11511 (*Homo sapiens*); CYP8A1, Q16647 (*Homo sapiens*) and CYP11A1, P05108 (*Homo sapiens*).

A multiple CYP450 structural alignment was generated including one representative structure for each isoform using the program Baton (D.Burke, unpublished data and [423]). The PDB codes for representative structures were as follows **CYP1A2,** 2HI4; **CYP2A6**, 2FDV; **CYP2B6**, 3IBD; **CYP2C8**, 1PQ2; **CYP2C9**, 1R9O; **CYP2C19**, 4GQS; **CYP2E1**, 3E4E; **CYP2D6**, 2F9Q; **CYP3A4**, 1TQN.

2.4.5 Delimiting the X-ray structures SRS map

The ligand contact residues identified in X-ray crystal structures of mammalian CYP450s deposited in the PDB before 2011 were annotated on the multiple sequence alignment of mammalian CYP450 to generate an SRS map. All residues at equivalent positions (*i.e.* evolutionarily equivalent residues) in other isoforms in the sequence alignment were assumed to be contact residues. SRS regions were delimited as blocks of contact residues within three amino acid positions of each other. SRS boundaries were delimited three residues on either side of the outer most contacts in each block. The use of three amino acids is arbitrary but is consistent with the way Gotoh defined the boundaries of the first CYP450 SRS map [53].

Following this, the structural alignment of the major human drug metabolising isoforms was also annotated with contact residues and the X-ray SRS map based on the boundaries of the SRS map delimited in the sequence alignment for CYP3A4. Boundaries for other isoforms may differ slightly between the sequence and structural alignments but as the SRS regions include 3 residues on either side of the outer most contact residues, all contact residues still fall within the map. The structural alignment was also annotated with new ligand contacts identified in new crystal structures, new SRS regions and periphery site contacts.

2.4.6 Delimiting the Docking SRS CYP450 map based on docking complexes

Docking studies were carried out by collaborators Lim Yen Ching and Yap Chun Wei as follows. CYP450 structures CYP1A2 (2HI4), CYP2A6 (1Z10), CYP2B4 (2Q6N & 2BDM), CYP2C5 (1NR6), CYP2C8 (2VN0), CYP2C9 (1R9O), CYP2D6 (2F9Q), CYP2R1 (3CZH), CYP3A4 (2V0M) and CYP46A1 (2Q9F) were retrieved from the RCSB Protein Data Base and prepared for docking using AutoDock Vina [424]. Structure preparation included (1) the removal of multiple protein chains, water molecules, buffer molecules and ligands from the pdb file, (2) the addition of hydrogen atoms and (3) conversion to pdbqt format using AutoDock tools [425].

The test whether the docking system, Auto Dock Vina, was suitable for docking substrates into the active sites of the CYP450s in question in default mode, substrates present in X-ray CYP450 structures were re-docked into the protein structure. The position of the ligand in the docking complex was compared to the ligand position in the crystal structure. RMSD values, representing

the difference between docking complexes and crystal structure complexes, in the re-docking tests were ≤ 2 Å. This indicated that AutoDock Vina under default parameters was suitable for docking substrates into CYP450 proteins.

Following this, PaDEL-AD v1.7 [426] was used to automate the docking of the 868 known substrates into the relevant CYP450 structures and the identification of LIGPLOT contacts in the docked complexes. A maximum of 9 docking modes were generated per substrate giving approximately 7000 docking modes in total. Several of the resultant CYP450:substrate complexes were randomly selected and visually inspected for quality control purposes. The docking modes of some substrates with high binding energies were also visualised to check for errors. LIGPLOT contacts for docking poses with a ligand binding free energy of less than -1 kcal/mol and atoms within 5 Å of the haem iron were used to generate the Docking SRS map as described for the X-ray structures SRS map (the docking SRS map was generated by Dr Zawaira). The 5 Å distance cut-off is a commonly accepted upper limit for the distance between the haem iron and a site of metabolism. A -1 kcal/mol cut-off was chosen to include as many known substrates for each isoform as possible without introducing too many incorrect binding modes.

Chapter 3 Tunnel gating in human CYP450s

3.1 Introduction

The dynamic motions of CYP450 enzymes controlling the access and egress of molecules into and out of the active site are still poorly understood. Most CYP450 crystal structures show the protein in a "closed" conformation with no access routes large enough to allow a water molecule to pass from the bulk solvent into the active site. Isolation of the active site from the bulk solvent is a requirement for catalysis as it prevents uncoupling of the reaction yet substrate and solvent access channels must by necessity exist. There are however also "open" crystal structures with pathways wide enough to allow molecules at most the size of water molecules (2.8 Å in diameter) to enter the active site [427]. "Wide-open" conformations with tunnels wide enough to accommodate molecules the size of a typical substrate have only been observed in a few cases (e.g. CYP55A1 structure 1ROM [428] and CYP450BM-3 structure 2HPD [429]). The inference from the structural data is therefore that CYP450s must undergo considerable conformational changes to allow substrate/products and water molecules in and out of the active site. Factors such as pathway width and curvature, as well as the properties of residues lining the pathway may selectively control the access of molecules into the enzyme's active site [430]. Gora *et al* recently reviewed general structural mechanisms controlling the passage of small molecules including substrates, products, ions and solvent molecules, into and out of proteins. These mechanisms, referred to as gates, are dynamic systems made of individual residues, loops or secondary structure elements or entire domains that can reversibly switch between an open and closed conformation, thereby controlling pathways within enzymes.

A number of *in silico* approaches, including thermal motion pathway analysis [431], adiabatic mapping [81], molecular dynamics [80, 87, 431-436] and computational geometry tools [66, 437-440] have been used to gain insight into CYP450 flexibility and tunnel-gating. These independent methods are in general agreement and several pathways found in distinct parts of the CYP450 fold have been identified as putative substrate access and solvent tunnels.

Thermal motion pathway (TMP) analysis of crystallographic temperature factors was the first systematic method used to predict putative substrate tunnels in closed structures [431]. Putative tunnels were identified by linking chains of residues with elevated temperature factors, corresponding to regions of high flexibility, from the active site to the enzyme surface.

Standard molecular dynamics approaches should in principle allow the visualisation of the opening and closing motions of tunnels. In practice, however the time scales on which ligand binding/unbinding occur are much longer than are computationally realistic to simulate for a large protein. Sampling techniques such as steered molecular dynamics (SMD) [81] and random expulsion molecular dynamics (REMD) [82] attempt to solve this problem. In SMD a force is applied to pull a ligand from the active site to the exterior of the protein along a tunnel and a rupture force is calculated; this information has been used to complement energy profiles obtained from adiabatic mapping, giving insight into the probability of a ligand passing through a given passage [432]. REMD reduces the timescales of spontaneous substrate exit by applying an artificial force to the substrate in a random direction in addition to the standard molecular mechanics force field.

Pathways accessible to spherical probes the size of water molecules identified using REMD closely corresponded to those identified by TMP [434], revealing three main exit routes (pw1, pw2 and pw3) defined by the secondary structures surrounding their point of exit. Pw2 was further subdivided into 5 subclasses (pw2a-e), which all share proximity with the B-C loop/B' helix, on the basis of REMD ligand egress trajectories [433]. In addition two other small channels referred to as the "solvent channel" (tunnel S) and the "water channel" (tunnel W) have been identified [427, 441, 442]. Tunnel S is thought to convey and control the access of protons to the active site, however substrates and other molecules may also use this channel to access and exit the active site [427]. Tunnel W acts as an aqueduct for water molecules to access and exit the active site during the catalytic cycle and is controlled at least in part by the binding of CPR to the proximal face of the CYP450 [442]. A description of each REMD derived tunnel is given in Table 3.1 [66, 427] and the exit point of each tunnel is illustrated in Figure 3.1. This nomenclature is now commonly used to describe the position of tunnels in CYP450 enzymes.

Gating residues have been identified for some of these pathways using molecular dynamics approaches. The tunnel-closing interactions identified are generally present in the start configuration of the simulation and only break when the ligand escapes from the protein [433, 434]. A study by Fishelovitch *et al*, analysing substrate tunnels in CYP3A4 using SMD, revealed that л-stacked phenylalanine residues were the main residues responsible for gating pw2a, 2b, 2c, 2e and pw3 [434]. This study also found that some residues gate more than one pathway and that, in some instances, the identification of gating residues was substrate dependant. Another

study by Schleinkofer *et al* identified a gating residue pair for rabbit CYP2C5 pw2c tunnel using REMD simulations with progesterone and 21-hydroxyprogesterone as substrates [433].

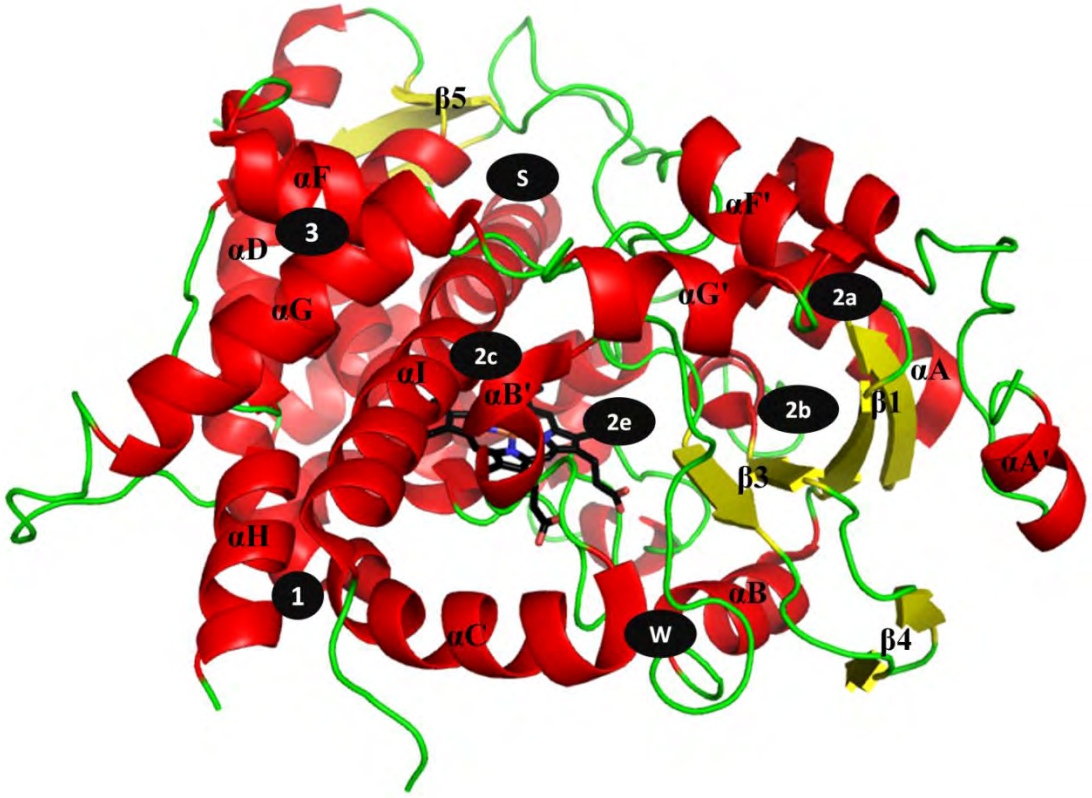


Figure 3.1 Positions of REMD tunnel egress points projected onto the CYP450 fold.
CYP3A4 structure 1TQN is shown with helical regions in red, β-sheets in yellow and loop regions in green. Secondary structure elements are labelled using the nomenclature of Poulos *et al* [37]. The position of pw2d is not clearly visible from this angle as it egresses between the N-terminus and the A' and A helices.

Table 3.1 CYP450 pathways identified in REMD studies

Pathway	Description
pw1	Passes under the helix I, exiting the protein between the C/C' and H or L helices close to the G-H loop and β-sheet 2.
pw2a	Exits between the F–G loop, B–C loop/B' helix/BB' loop and β-sheet 1.
pw2b	Exits between the B/B' loop and the β-sheet 1and β-sheet 3
pw2c	Exits between the G and I helices and the B' helix/B–C loop
pw2d	Exits between the N-terminus and the A' and A helices
pw2e	Exits through the B–C loop
pw3	Exits between the F and G helices or at the EF loop
S	Exits between the F, E and I helices and the β-sheet 5
W	Exits at the base of the B–C loop near the C-terminus of the B helix

In 2006 Cojocaru *et al* used CAVER, a computational geometry tool designed to rapidly identify pathways from protein clefts to the protein surface in static or dynamic structures [437], to determine the distribution and pattern of tunnels in static CYP450 structures [66]. The tunnels identified using this approach largely corresponded to those identified in REMD simulations and hence it is possible to assign CAVER tunnels to REMD tunnels using the standard nomenclature. In addition to previously delimited REMD tunnels, two new pathways (pw4 and pw5) were identified as well as two new subclasses of pw2 (pw2f and pw2ac). Pw4 exits the protein through the F-G loop while, pw5 exits between the K and K' helices. Pw2ac exits between the top of the B-C loop (or B' helix) and the G helix, between pw2c and pw2a/pw2b. Pw2f could be considered a variant of pw2a, exiting between the F' helix/F-G loop and β-sheet 5.

This Chapter describes the use of differential tunnel-opening patterns observed in static human CYP450 crystal structures to identify gating residue and tunnel-closing interactions for REMD equivalent pathways. The computational geometry tool, CAVER, was used to identify open tunnels in all available CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2D6 and CYP3A4 crystal structures. Based on the assumptions that amino acid side chain motions are the main mechanism controlling the opening and closing of ligand access tunnels, "tunnel-closed" structures were aligned with "tunnel-open" structures to identify side chains obstructing the pathway of interest. Following this, the intra-protein interactions made with these gating residues that are responsible for tunnel closure were identified.

A large portion of the work described in this Chapter was done in collaboration with a post-doctoral fellow in the Blackburn laboratory, Dr Alexander Zawaira, and was published in the Journal of Structural Biology in 2011 ([443], full manuscript appended at the end of Thesis). My contribution to the experimental work in this publication included the following: determination of the optimal gauge width cut-off for tunnel identification using CAVER; tunnel calculations and assignment as well as the identification of gating models in a large portion of the structures analysed; calculation of the average rate of evolution of tunnel-intersecting residues in CYP2 isoforms. I have subsequently updated and expanded the work independently to include gating mechanisms identified in two addition CYP450 isoforms (CYP2B6 and CYP2D6) as well as to incorporate further insights from new structural data.

3.2 Results

CAVER uses the Dijkstra algorithm (a graph search algorithm designed to find the shortest path from a specified point to all other points within a graph simultaneously), to calculate the shortest low-cost path from a user defined starting point within the structure to the bulk solvent. CAVER returns a list of possible routes out of the molecule, giving greatest priority to routes with the widest bottleneck (gorge width) irrespective of the length of the tunnel. There are no or very few tunnels wide enough for typical substrates or water molecules to pass through (radius > 2.8 Å) in most CYP450 crystal structures. Tunnel-opening resulting from the disruption of tunnel-gating interactions is likely to lead to larger conformational changes, allowing molecules in and out of the protein. Similarity to REMD tunnels was thus an important criterion for identifying open tunnels and the subsequent delimitation of gating models. Based on this criterion, a gorge width cut-off suitable for identifying differential tunnel-opening patterns in CYP450 structures, while restricting calculated tunnels to REMD equivalent tunnels as far as possible, was required to distinguish open tunnels from closed tunnels.

A suitable cut-off was chosen based on a preliminary thresholding exercise aimed at finding an optimal balance between the number of tunnel-open/tunnel-closed structure pairs and the resemblance of CAVER calculated tunnels to REMD tunnels. At gorge width cut-offs < 0.7 Å there were no differential tunnel-opening patterns as all REMD tunnels were open with over 35 tunnels calculated per structure. At gorge width cut-offs > 2.5 Å there were no open tunnels apart from artefacts arising from missing regions within crystal structures. At gorge width cut-offs < 1 Å, the majority of calculated tunnels did not match REMD tunnels. A gorge width cut-off of 1.4 Å (~ half a water molecule) gave the best balance between tunnel similarity and differential tunnel-opening in most CYP450 isoforms and was therefore used throughout the rest of this study.

The CAVER algorithm divides the 3-dimensional protein space into a grid and calculates a convex envelope around the protein. Grid points lying on the envelope are possible tunnel exit point. If several of these possible exit points fall within the region of a REMD tunnel exit, several variants of the same REMD tunnel may be calculated by CAVER [66]. All tunnel variants were recorded but in principle, only one of these tunnels need be considered. The tunnel

that most closely resembled the REMD tunnel with the widest gorge width was used to identify gating residues.

Tunnel-gating residues where identified by aligning structures with differential tunnel-opening patterns and manually identifying residue side chains obstructing the tunnels in the tunnel-closed structures. Where possible structures of the same isoform were aligned, enabling the delimitation of two-state gating models made up of a tunnel-intersecting residue and its interacting partners. Protein interaction calculator [444] was used to identify tunnel-closing interactions defined as intra-protein interactions involving the tunnel-intersecting residue that are present in the tunnel-closed structure but absent in the tunnel-open structure.

In the cases where a tunnel showed no differential tunnel-opening patterns between structures of the same isoform, tunnel-intersecting residues were identified by aligning structures of different isoforms. Only singlet gating models could be delimited in these cases because tunnel-closing interactions exclusive to the tunnel-closed structure could not be identified in cases where the protein sequences differed between open and closed structures.

3.2.1 Differential tunnel-opening patterns and gating models in CYP2A6 structures

Table 3.2 shows the distribution of open tunnels in CYP2A6 crystal structures. CAVER was used to identify two open tunnels in CYP2A6: the water channel (tunnel W), open in 6 out of 8 CYP2A6 structures, and a previously undefined tunnel open in all 10 structures. This new tunnel was named pw6 in keeping with the naming convention of Cojocaru *et al* [66]. Pw6, starting at the haem iron, initially follows a similar path to pw1 passing under the I helix but unlike pw1, exits on the opposite side of the C helix, between the C helix and the N-terminus of the L helix, near the C-D loop (Figure 3.2). Gating models for pw6 were not delimited here as REMD simulations have not confirmed the use of this tunnel.

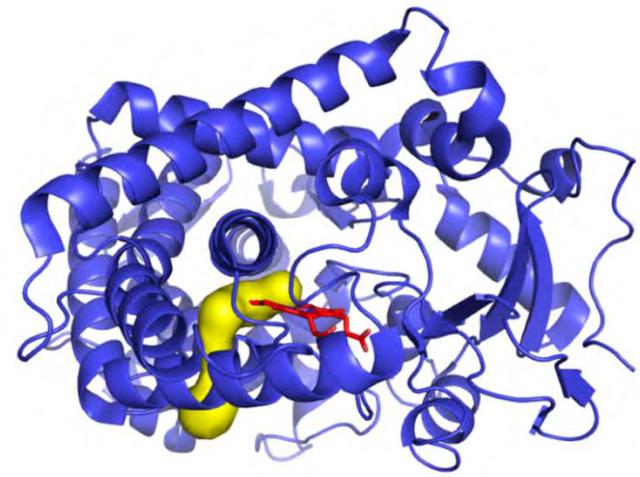


Figure 3.2 CYP2A6 structure 3T3R showing CAVER calculated tunnel pw6 with a gorge width of 1.56 Å.

Table 3.2 CAVER-calculated tunnels in CYP2A6 crystal structures

PDB	Tunnel No.	pw2a	pw2b	pw2c	pw2e	pw3	S	W	*pw2ac*	*pw6*	Width(Å)
1Z10	1									×	1.54
	2							×			1.51
1Z11	1									×	1.50
	2							×			1.48
2FDU	1							×			1.56
	2									×	1.52
2FDV	1							×			1.55
	2									×	1.53
	3							×			1.43
2FDW	1									×	1.57
	2							×			1.45
2FDY	1									×	1.47
2PG5	1									×	1.49
	2							×			1.48
	3									×	1.42
3T3R	1									×	1.56
	2							×			1.47
4EJJ	1									×	1.49
3T3Q	1									×	1.50
	2							×			1.47

Note: pw2ac and pw6 (in italics) have only been identified in static structures and have not yet been confirmed as REMD tunnels.

Singlet gating models in CYP2A6

In an attempt to identify a two-state gating model for tunnel W, tunnel-open structure 2FDV was aligned to tunnel-closed structure 2FDY. Lys 125 was identified as a tunnel-intersecting residue. The only difference in intra-protein interactions involving Lys 125 in the tunnel-open and tunnel-closed structures was the Lys 125- Glu 96 interaction present exclusively in the tunnel-open structure (Figure 3.3). Due to the lack of a tunnel-closing interaction, the gating of tunnel W can only be described by a singlet gating model (Table 3.3).

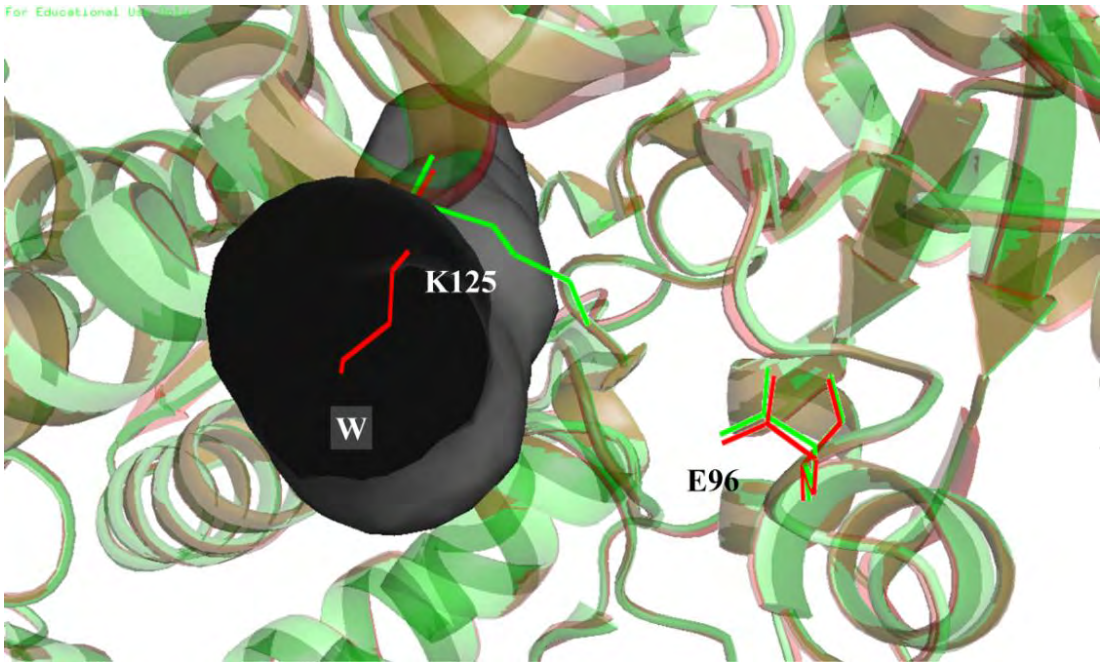


Figure 3.3 CYP2A6 structure 2FDV (green) with open tunnel W (grey) aligned to tunnel-closed structure 2FDY (red) showing tunnel-intersecting residue K125. The interaction between K125 and E96 keeps the tunnel open in structure 2FDV (tunnel-opening interaction).

Singlet gating models were also identified for pw2a, pw2c, pw2e and tunnel S by aligning CYP2A6 structures to open tunnels observed in CYP2C8 structures (Table 3.3). Figure 3.4 shows CYP2A6 residues, Gln 104, Phe 107, Phe 118, Leu 370 and Phe 480 intersecting pw2e.

Singlet gating models were not delimited for pw2b and pw3 as backbone regions of CYP2B6 rather than amino acid side chains obstruct these tunnels in structural alignments with tunnel-open structures.

Table 3.3 Singlet gating models in CYP2A6

Gated tunnel	Structures: open/closed	Tunnel-intersecting residues
W	2FDV(CYP2A6)/2FDY(CYP2A6)	**Lys 125**[a]
pw2a	2NNH(CYP2C8)/2FDV(CYP2A6)	**Phe 480**
pw2c	2NNH(CYP2C8)/2FDV(CYP2A6)	**Phe 107, Phe 111, Phe 118, Gln 242, Glu 245, Met 293**
Pw2e	2NNH(CYP2C8)/2FDV(CYP2A6)	**Gln 104, Phe 107, Phe 118, Leu 370, Phe 480**
S	2NNH(CYP2C8)/2FDV(CYP2A6)	**Phe 107, Phe 111, Phe 118, Met 205, Leu 241**

[a] No tunnel-closing interaction, only tunnel-opening interaction K125-E96.

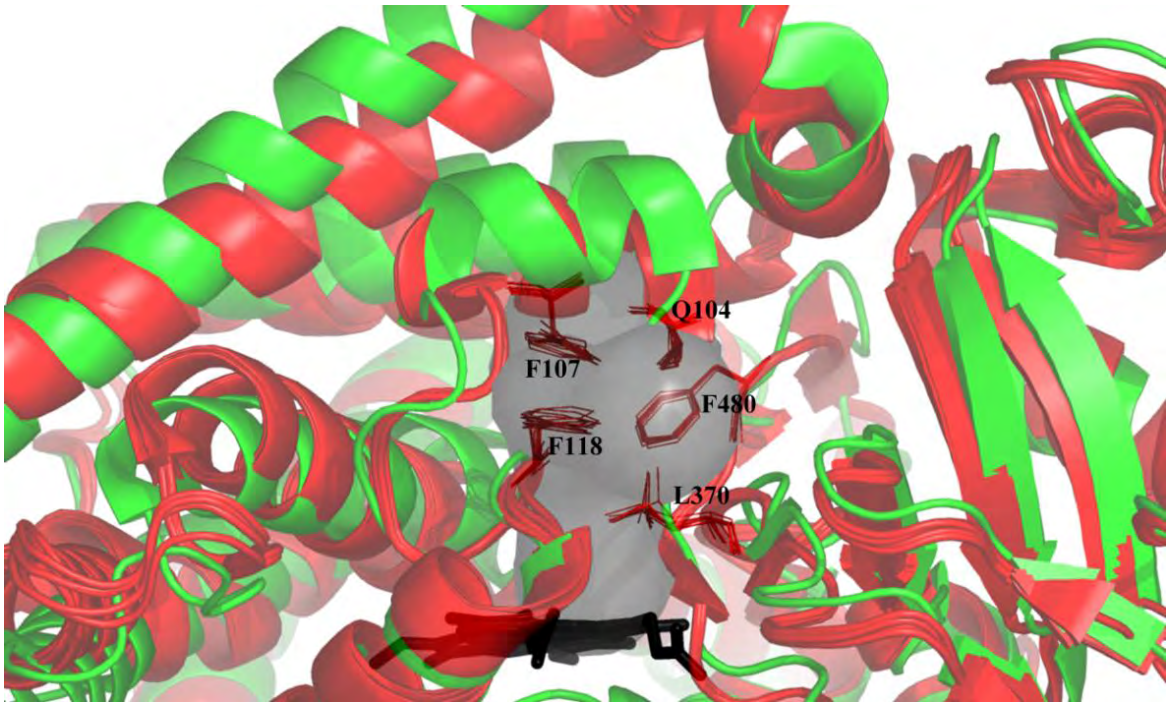


Figure 3.4 CYP2C8 structure 2NNH (green) with open tunnel pw2e (grey) aligned to all 10 CYP2A6 structures (red) showing CYP2A6 pw2e tunnel-intersecting residues Q104, F107, .F118, L370 and F480.

3.2.2 Differential tunnel-opening patterns and gating models in CYP2B6 structures

Table 3.4 shows the distribution of open tunnels in CYP2B6 crystal structures. Differential tunnel-opening patterns were observed for REMD tunnels pw2a, pw2b and tunnel S as well as the newly assigned tunnel pw6.

Table 3.4 CAVER-calculated tunnels in CYP2B6 crystal structures

PDB	Tunnel No.	pw2a	pw2b	pw2c	pw2e	pw3	S	W	*pw2ac*	*pw6*	Width(Å)
3IBD	1									×	2.11
	2[a]		×*								1.53
3UA5	1	×									2.24
	2						×				2.16
4I91	1									×	1.41
3QOA	1									×	1.40
3QU8	none										

Note: pw2ac and pw6 (in italics) have only been identified in static structures and have not yet been confirmed as REMD tunnels.
* indicates tunnel may be an artefact of regions missing from the crystal structure.
[a] Missing Glu218 atoms in the crystal structure render pw2b open in 3IBD

Two- state gating models in CYP2B6

Table 3.5 shows the two-state tunnel-gating models identified in CYP2B6. The side chain of Leu 51 intersects pw2a in all the tunnel-closed structures. Leu 51 interacts hydrophobically with Leu 219 and Met 365 and forms a side chain-main chain interaction with Gln 215 in the tunnel-closed structures. Gln 215 forms part of a small section of the F-G loop backbone that obstructs pw2a. The side chain of adjacent residue Leu216 also intersects pw2a, interacting hydrophobically with Leu 219 and Leu 43. These tunnel-closing interactions are all absent in the tunnel-open structure 3UA5.

Table 3.5 Two-state gating models in CYP2B6

Gated tunnel	Structures: open/closed	Tunnel-closing interaction	
pw2a	3UA5/(3IBD, 4I91, 3QOA, 3QU8)	**Leu 51**-Leu 219	hydrophobic
		Leu 51-Met 365	hydrophobic
		Leu 51-Gln 215	side chain-main chain H-bond
		Leu 216-Leu 219	hydrophobic
		Leu 216-Leu 43	hydrophobic
S	3UA5/(4I91, 3QOA, 3QU8)[a]	**Phe 206**-Leu 363	hydrophobic
		Phe 206-Ile 480	hydrophobic

Tunnel-intersecting residues are shown in bold face
[a] Tunnel S is also closed in 3IBD however the side chain of F206 does not intersect S in this structure. Residues Ser107, Glu474 and Ile480 appear to restrict tunnel S in this case.

Phe 206 gates tunnel S in CYP2B6, intersecting tunnel S in 3 of the 4 tunnel-closed structures and making hydrophobic tunnel-closing interactions with Leu 363 and Ile 480 (Figure 3.5). Tunnel S is also closed in structure 3IBD but is not intersected by Phe 206; rather it appears to be restricted by Ser 107, Glu 474 and Ile 480.

A two-state gating model was not identified for pw2b because the open tunnel in 3IBD is an artefact of the missing side chain atoms of Glu 218. It does however indicate that Glu 218 may play a role in pw2b gating in CYP2B6.

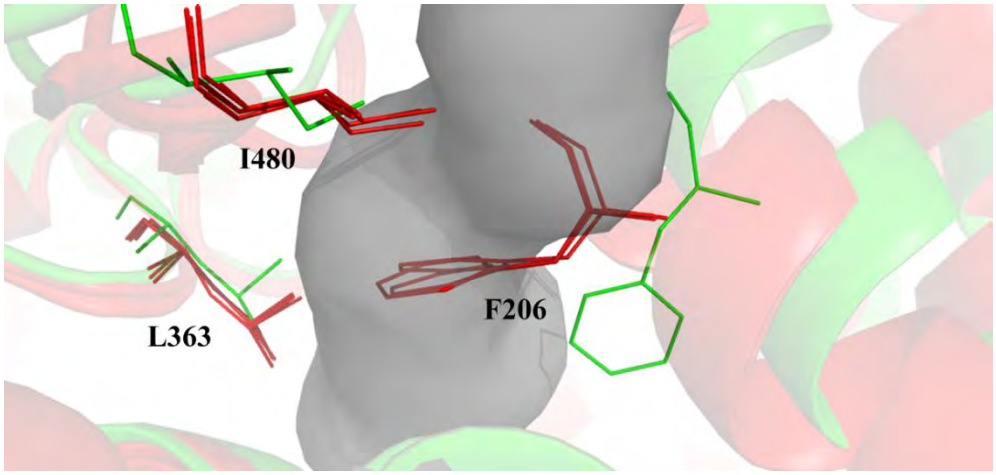


Figure 3.5 Gating model for tunnel S in CYP2B6.
CYP2B6 structure 3UA5 (green) with open tunnel S (grey) aligned to closed structures 4I91, 3QOA and 3QU8 (red) in PyMOL. F206 obstructs tunnel S in closed structures 4I91, 3QOA and 3QU8 forming hydrophobic interactions with L363 and I480.

Singlet gating models in CYP2B6

Table 3.6 shows the CYP2B6 singlet gating models identified for pw2b, pw2c, pw2e and tunnel W by aligning CYP2B6 structures to tunnel-open structures from other isoforms.

Table 3.6 Singlet gating models in CYP2B6

Gated tunnel	Structures: open/closed	Gating residue
pw2b	1TQN(CYP3A4)/3QOA(CYP2B6)	**Glu 218**
pw2c	2NNH(CYP2C8)/3QOA(CYP2B6)	**Phe 108, Leu 290**
pw2e	2NNH(CYP2C8)/3QOA(CYP2B6)	**Ile 101, Val 104, Phe 115**
W	2FDV(CYP2A6)/3QOA(CYP2B6)	**Trp 121**

Pw2b tunnel-intersecting residue Glu 218 was identified by aligning CYP2B6 structures to CYP3A4 structure 1TQN, as speculated above. Pw2b gating residues could not be identified when CYP2B6 structures were aligned to tunnel-open structures from more closely related CYP2C8 and CYP2C9 isoforms as the B-C loop backbone intersected pw2b in all these structural alignments.

Tunnel-intersecting residues were not identified for CYP2B6 pw3 as the G helix backbone obstructs pw3 in structural alignments with CYP3A4 tunnel-open structure 2V0M.

3.2.3 Differential tunnel-opening patterns and gating models in CYP2C8 structures

Table 3.7 shows the distribution of open tunnels in CYP2C8 crystal structures. Differential tunnel-opening patterns were observed for REMD tunnels pw2a, pw2c, pw2e and tunnel S as well as the newly assigned tunnel pw6. Pw2b was open in all structures while pw3 and tunnel W were closed in all structures.

Table 3.7 CAVER-calculated tunnels in CYP2C8 crystal structures

PDB	Tunnel No.	pw2a	pw2b	pw2c	pw2e	pw3	S	W	*pw2ac*	*pw6*	Width(Å)
1PQ2	1		×								2.05
	2		×								1.55
	3	×									1.40
2NNH	1		×								2.46
	2			×							1.70
	3				×						1.53
	4[a]										1.50
	5						×				1.44
2NNI	1		×								2.34
	2						×				1.55
	3			×							1.50
	4									×	1.42
	5									×	1.42
	6									×	1.42
2NNJ	1		×								2.28
	2						×				1.53
	3			×							1.50
	4									×	1.49
2VN0	1		×								2.15
	2				×						1.60
	3									×	1.55
	4									×	1.44

Note: pw2ac and pw6 (in italics) have only been identified in static structures and have not yet been confirmed as REMD tunnels.
[a] Cojocaru *et al* identified a tunnel in the same position in P450cam using CAVER and assigned it pw5 however, it has not yet been observed in REMD simulations.

Two-state gating models in CYP2C8

Table 3.8 shows the two-state gating models identified for tunnels pw2a, pw2c, pw2e and tunnel S in CYP2C8 structures. Ile 476 intersects pw2a in all tunnel-closed CYP2C9 structures with the exception of 2NNJ, which is intersected by Ile 50. Ile 476 gates pw2a in the tunnel-closed structure 2NNH by forming a network of hydrophobic interactions with Val 362, Pro 367 and Met 388 that are absent in the tunnel-open structure 1PQ2. Ile 476 also intersects pw2a in tunnel-closed structures 2NNI and 2VN0 but only the Ile 476 – Pro 367 hydrophobic tunnel-

closing interaction is present in 2NNI, 2VN0 and 2NNH suggesting that it is the primary tunnel-closing interaction.

Table 3.8 Two-state gating models in CYP2C8

Gated tunnel	Structures: open/closed	Tunnel closing interaction	
pw2a	1PQ2/2NNH	**Ile 476**-Val 362	hydrophobic
		Ile 476-Pro 367	hydrophobic
		Ile 476-Met 388	hydrophobic
pw2c	2NNH/1PQ2	**Arg 241**-Glu 285	ionic
pw2e	2NNH/1PQ2	**Asn 99**-Ser 114	side chain-main chain H-bond
S	2NNI/(1PQ2, *2VN0*[a])	**Lys 199** – Phe 168	hydrophobic

Tunnel-intersecting residues are shown in bold face.
[a] Tunnel-closing interaction Lys199-Phe168 is present in tunnel-closed structure 2VN0 shown in italics however the tunnel-intersecting residue Lys199 does not intersect the tunnel in this structure.

Pw2c is gated by the Arg 241-Glu 285 ionic interaction in CYP2C8 structure 1PQ2 (Figure 3.6 A) reducing the size of the gorge width from 1.7 Å/1.5 Å (observed in the tunnel-open structures) to 0.9 Å. This ionic interaction is absent in tunnel-open structures 2NNH, 2NNI and 2NNJ. Note that this ionic interaction is not observed in tunnel-closed structure 2VN0, however while pw2c is classified as closed in this structure, it has a larger gorge width than 1PQ2 (1.24 Å compared with 0.9 Å). Pw2c 2VN0 may therefore represent an intermediate between the tunnel-open and tunnel-closed states.

Arg 99 – Ser 114 side chain-main chain interaction, present in all three of the CYP2C8 tunnel-closed structures, gates pw2e. In these structures, the side chain of Arg 99 is orientated towards the centre of the B-C loop, intersecting tunnel pw2e, whereas in the tunnel-closed structures the Arg side chain is orientated away from pw2e and does not interact with Ser 114 (Figure 3.6 B).

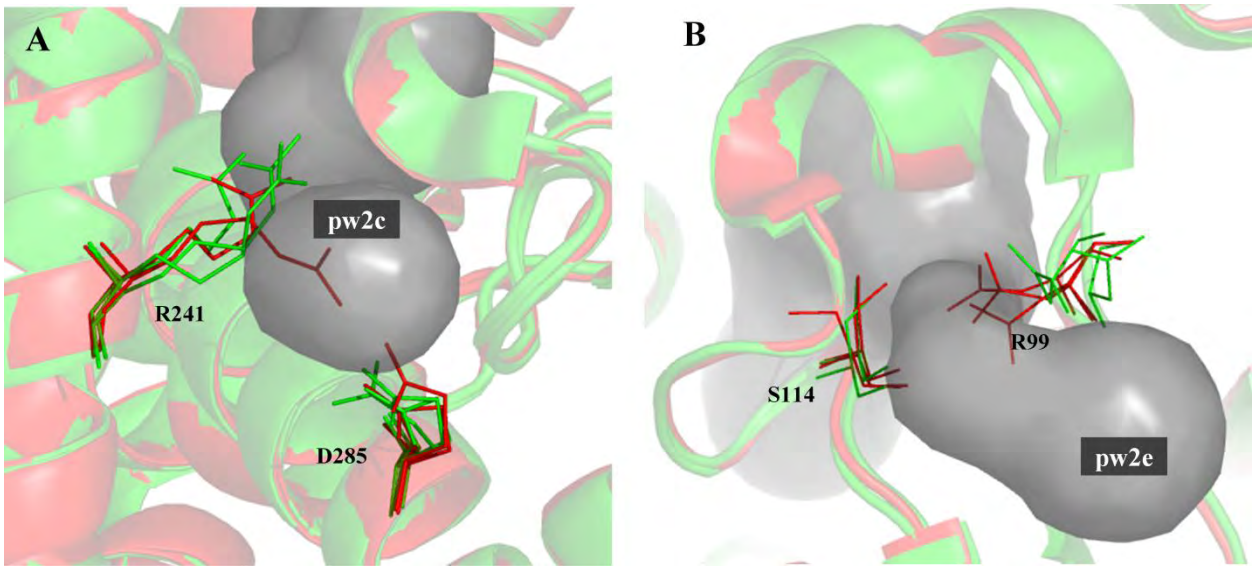


Figure 3.6. Gating model for pw2c and pw2e in CYP2C8.
A) CYP2C8 structure 2NNH, 2NNI and 2NNJ (green) with open tunnel pw2c (grey) aligned with tunnel-closed structures 1PQ2 and 2VN0 (red) in PyMOL. The side chain of R241 intersects tunnel pw2c in structure 1PQ2 and forms an ionic interaction with D285, reducing the gorge width of pw2c to 0.9 Å. The side chain of R241 in 2VN0 also intersects pw2c but to a lesser degree reducing the gorge width to 1.24 Å. **B)** CYP2C8 structures 2NNH and 2VN0 (green) with open pw2e were aligned with tunnel-closed structures 2NNI, 2NNJ and 1PQ2 (red) in PyMOL. The side chain of R99 forms a hydrogen bond with the main chain of S114 closing pw2e in all three closed structures. This R99-S114 interaction is absent in both open structures.

Figure 3.7 shows tunnel S gated by tunnel-intersecting residue Lys199. This residue interacts with Phe 169 in the tunnel-closed CYP2C8 structure 1PQ2 but not in the tunnel-open structure. Although Lys 199 does not intersect tunnel S in the other tunnel-closed structure 2VN0 the Lys 199 – Phe 169 hydrophobic tunnel-closing interaction is also present in this structure.

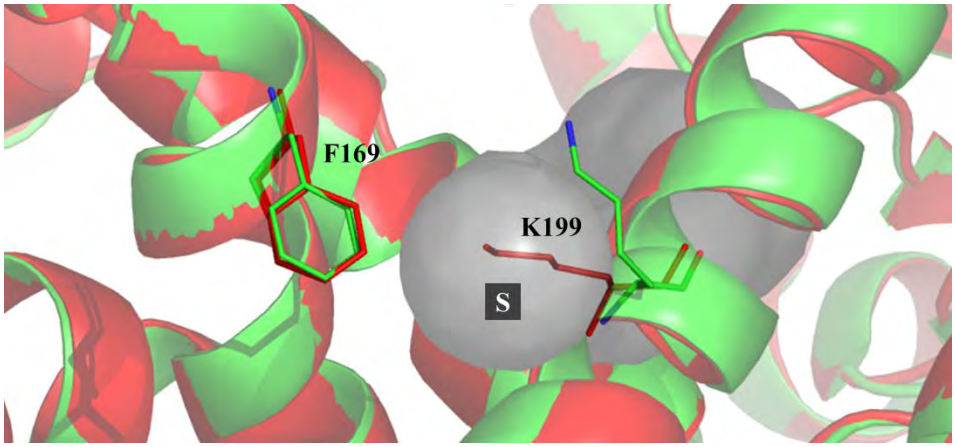


Figure 3.7 CYP2C8 structure 2NNI (green) with open tunnel S (grey) aligned to tunnel-closed structure 1PQ2 (red). The side chain of Lys 199 intersects tunnel S forming a hydrophobic interaction with Phe 169.

Singlet gating models in CYP2C8

Table 3.9 shows the CYP2C8 singlet gating models identified for pw3 and tunnel W by aligning CYP2C8 structures to tunnel-open structures from other isoforms. Gating residues could not be identified for pw2b, as pw2b was open with a gorge width > 2 Å in all CYP2C8 structures.

Table 3.9 Singlet gating models in CYP2C8

Gated tunnel	Structures: open/closed	Gating residue
pw3	2V0M (CYP3A4)/2NNH(CYP2C8)	**Ile 207, Leu 208, Thr 229, Val 233**
W	2FDV(CYP2A6)/2NNH(CYP2C8)	**Trp 120, Arg 125**

3.2.4 Differential tunnel-opening patterns and gating models in CYP2C9 structures

Table 3.10 shows the distribution of open-tunnels in CYP2C9 crystal structures. Differential tunnel-opening patterns were observed for REMD tunnels pw2c and pw2e as well as CAVER calculated tunnel pw2ac and the newly assigned tunnel pw6. Pw2a and pw2b were closed in all CYP2C9 structures and the tunnel S was open in all structures.

Table 3.10 CAVER-calculated tunnels in CYP2C9 crystal structures

PDB	Tunnel No.	pw2a	pw2b	pw2c	pw2e	pw3	S	W	*pw2ac*	*pw6*	Width(Å)
1OG2	1						×				1.75
	2			×							1.52
1OG5	1						×				1.78
	2			×							1.55
	3									×	1.45
1R9O	1						×				2.21
	2								×		1.93
	3				×						1.88
	4								×		1.68
	5				×						1.48

Note: pw2ac and pw6 (in italics) have only been identified in static structures and have not yet been confirmed as REMD tunnels.

Two-state gating models in CYP2C9

Differential tunnel-opening patterns in CYP2C9 static crystal structures revealed two-state gating models for REMD tunnels pw2c and pw2e (Table 3.11). Arg 108 gates pw2c, forming an ionic interaction with Asp 293 and main chain-main chain interactions with Phe 110 and Gly 111 (Figure 3.8 A). Phe 110 is also involved in gating pw2e; Phe 110 and Ile 99 interact hydrophobically, intersecting pw2e in the tunnel-closed structures (Figure 3.8 b). Other interactions responsible for pw2e tunnel closure are hydrophobic interactions Phe 110-Phe 114, Ile 99-Ala 103 and Ile 99-Phe 114.

Table 3.11 Two-state gating models in CYP2C9

Gated tunnel	Structures: open/closed	Tunnel closing interaction	
pw2c	(1OG5, 1OG2)/1R9O	**Arg 108**-Asp 293	ionic
		Arg 108-Phe 110	main chain-main chain H-bond
		Arg 108-Gly 111	main chain-main chain H-bond
pw2e	1R9O/(1OG5, 1OG2)	**Phe 110-Ile 99**	hydrophobic
		Phe110–Phe 114	hydrophobic
		Ile 99–Ala 103	hydrophobic
		Ile 99–Phe 114	hydrophobic

Tunnel-intersecting residues are shown in boldface.

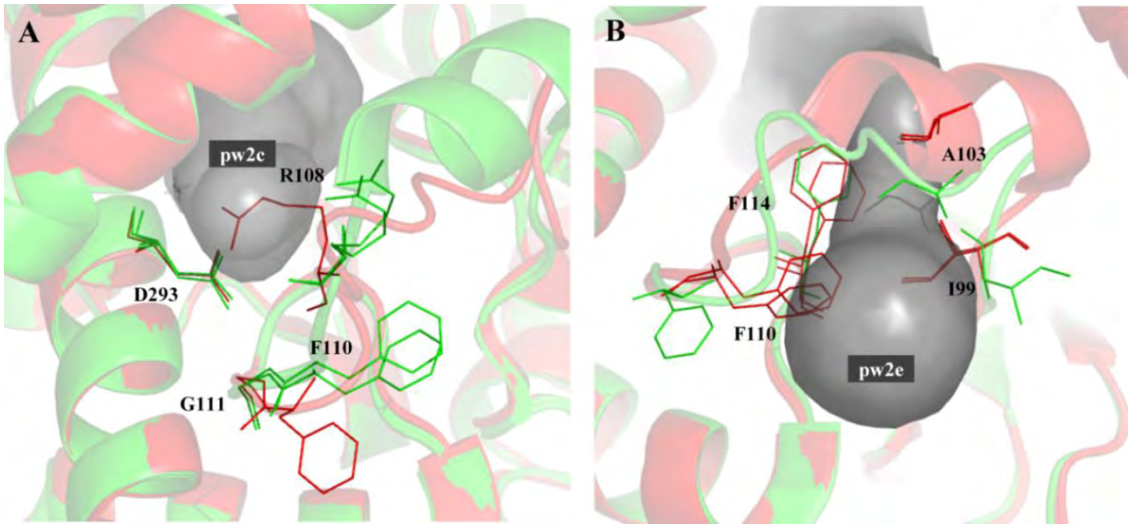


Figure 3.8 Gating model for pw2c and pw2e in CYP2C9.
A) CYP2C9 structures 1OG2 and 1OG5 (green) with open tunnel pw2c (grey) aligned to tunnel-closed structure 1R90 (red) in PyMOL. The side chain of R108 intersects tunnel pw2c in 1R9O forming an ionic interaction with D293 and main chain-main chain hydrogen bonds with F110 and G111; these interactions are only present in the tunnel-closed structure. **B)** CYP2C9 structure 1R9O (green) with open pw2e aligned to tunnel-closed structures 1OG2 and 1OG5 (red) in PyMOL. The side chain of residues F110 and I99 insect tunnel pw2e in both of the tunnel-closed structures and form the following hydrophobic tunnel closing interactions that are absent in 1R9O: F110-I99; F110-F114; I99-A103; I99-F114.

Singlet gating models in CYP2C9

Table 3.12 shows the CYP2C9 singlet gating models identified for pw2a, pw2b, pw3 and W by aligning CYP2C9 structures to tunnel-open structures from other isoforms. Gating residues could not be identified for tunnel S in CYP2C9 static structures as this tunnel was open in all three structures.

Table 3.12 Singlet gating models in CYP2C9

Gated tunnel	Structures: open/closed	Gating residue
pw2a	1PQ2(CYP2C8)/1OG5(CYP2C9)	**Phe 476**
pw2b	1PQ2(CYP2C8)/1OG5(CYP2C9)	**Phe 100, Lys 72**
pw3	2V0M(CYP3A4)/1OG5(CYP2C9)	**Ile 207, Leu 208, Thr 229**
	2V0M(CYP3A4)/1R90(CYP2C9)	**Tyr 225**
W	2FDV(CYP2A6)/1OG5(CYP2C9)	**Trp 120**

3.2.5 Differential tunnel-opening patterns and gating models in CYP2D6 structures

Table 3.13 CAVER-calculated tunnels in CYP2D6 crystal structures

PDB	Tunnel No.	pw2a	pw2b	pw2c	pw2e	pw3	S	W	*pw2ac*	*pw6*	Width(Å)
3QM4	1		×								1.45
2F9Q	1						×				2.25
3TDA[a]	1		×								1.49
	2	×									1.49
	3[b]	×									1.49
3TBG	1[b]	×									2.13
	2		×								2.13
	3						×				2.01
	4	×									1.90
	5						×				1.85
	6				×						1.44

[a] Three tunnels branch from one tunnel and therefore have an identical minimum gorge width
[b] Variant of pw2a displaced closer to the solvent channel than the canonical pw2a.

Table 3.13 shows the distribution of open tunnels in CYP2D6 crystal structures. Differential tunnel-opening patterns were observed for REMD tunnels pw2a, pw2b, pw2e and tunnel S.

Two- state gating models in CYP2D6

Two-state gating models were delimited for tunnels pw2a, pw2e and tunnel S (Table 3.14). While there is one CYP2D6 structure where pw2b is closed, a portion of the F-G loop backbone rather than an amino acid side chains obstructs this tunnel and hence a gating model based on the movement of side chains could not be delimited.

Table 3.14 Two-state gating models in CYP2D6

Gated tunnel	Structures: open/closed	Tunnel closing interaction	
pw2a	3TBG/3QM4	**Phe 219- Phe 51**	hydrophobic (aromatic)
		Phe 219-Gln 52	main chain-main chain H-bond
		Phe 219- Leu 46	hydrophobic
		Phe 51-Gly 218	main chain-main chain H-bond
		Phe 51-Thr 54	main chain-main chain H-bond
pw2e	3TBG/(3QM4, 3TDA)	**Phe 483**-Phe120	hydrophobic (aromatic)
		Phe 483-Val 374	hydrophobic
S	3TBG/(3QM4, 3TDA)	**Gln 210**-Asp 179	side chain-side chain H-bond

Tunnel-intersecting residues are shown in bold face

All four pw2a tunnels observed in CYP2D6 structures vary in position, with tunnel 1 in 3TBG and tunnel 3 in 3TDA (Table 3.13) positioned closer to tunnel S than the canonical pw2a identified by REMD. Tunnel 4 in 3TBG was used to identify pw2a tunnel-intersecting residues since it is the tunnel with the widest gorge width that most closely resembles the canonical pw2a tunnel. The side chains of Phe 219 and Phe 51 in the tunnel-closed structure 3QM4 intersect this tunnel (Figure 3.9). Table 3.14 lists the tunnel-closing interactions made by these residues. The F-G loop of tunnel-closed structure 2F9Q adopts a substantially different conformation to 3TBG and 3QM4, with the F-G loop backbone passing through the middle of pw2a. Residues 42 to 51 are also missing in this structure and thus this tunnel-closed structure was not used to identify pw2a gating residues.

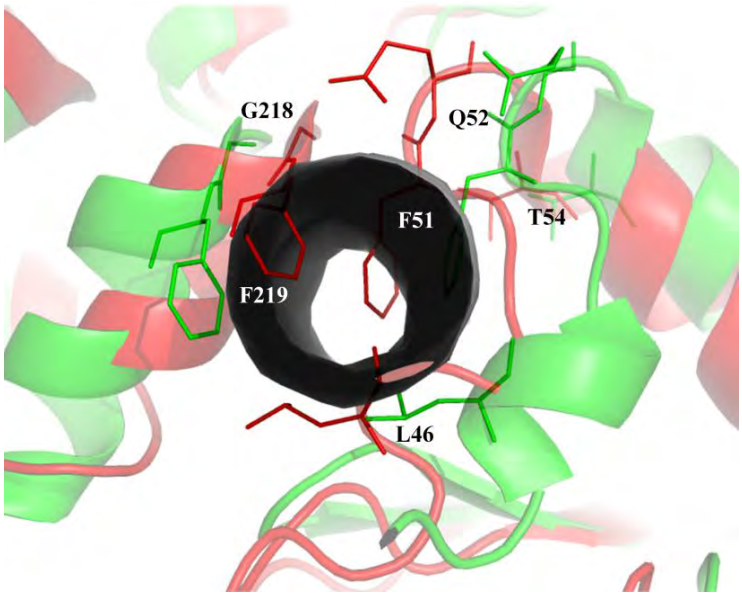


Figure 3.9 Gating model for pw2a in CYP2D6.
CYP2D6 structure 3TBG (green) with open tunnel pw2a (grey) was aligned with tunnel-closed structure 3QM4 (red) in PyMOL. The side chain of F219 and F51 as well as the main chain of L46 intersects the tunnel in 3QM4. F219 forms hydrophobic interactions with F51 and L46 and a main chain-side chain interaction with Q52 in the tunnel-closed structure 3QM4 but not in the tunnel-open structure 3TBG. F51 also forms a side chain-main chain hydrogen bond with T54 and a main chain-main chain bond with Q52 that are not present in the tunnel-open structure.

Phe 483 gates pw2e in tunnel-closed structures 3QM4 and 3TDA (Figure 3.10 A). Tunnel-closing interactions F483-V374 and F483-F120 were present in both 3QM4 and 3TDA. A different gating model is however observed for the third tunnel-closed structure 2F9Q (Figure 3.10 B), this gating model was however not considered due to the V374M mutations in this structure, which appears to alter the enzyme conformation. M374 interacts with F220 which, in contrast to F220 that faces away from the active site in the other three CYP2D6 structures, is orientated towards the active site due to a conformational change in the F-G loop region. This structure illustrates how a single amino acid substitution can affect enzyme conformation and tunnel gating.

The three solvent tunnels that are observed in CYP2D6 structures (one in 2F9Q and two in 3TBG) are not equivalent. Tunnel number 5 in 3TBG, with a gorge width of 1.8 Å (Table 3.13) most closely resembles the canonical tunnel S delimited in REMD studies and was used to determine tunnel-gating residues. The side chain of Gln 210 forms a hydrogen bond with the side chain of Asp 179, closing tunnel S in structures 3QM4 and 3TDA (Figure 3.11).

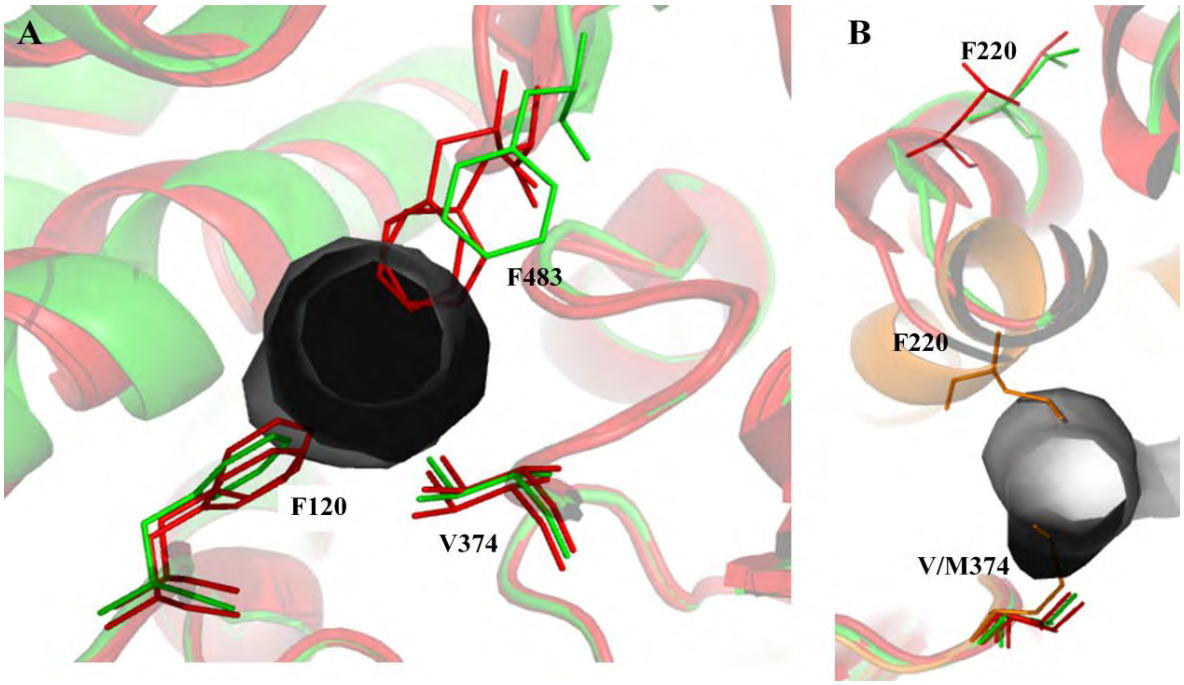


Figure 3.10 Gating model for pw2e in CYP2D6.
A) CYP2D6 structure 3TBG (green) with open tunnel pw2e (grey) aligned to tunnel-closed structures 3QM4 and 3TDA (red) in PyMOL. F483 interacts with F120 and V374 via hydrophobic interactions to close pw2e. F483 does not interact with F120 and V374 in the tunnel-open structure. B) Tunnel-closed structure 2F9Q (orange) with mutation V374M aligned to the other three CYP2D6 structures showing the gating of pw2e by the M374-F220 interaction.

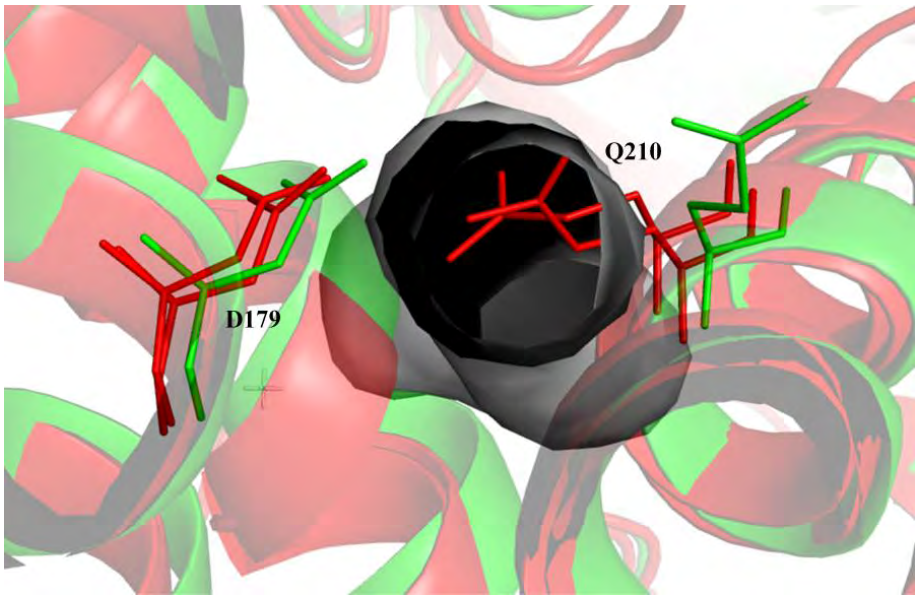


Figure 3.11 Gating model for tunnel S in CYP2D6. CYP2D6 structure 3TBG (green) with open tunnel S (tunnel number 5 in Table 3.13) aligned to tunnel-closed structures 3QM4 and 3TDA (red) in PyMOL. The side chain of Q210 forms a hydrogen bond with the side chain of D179, closing tunnel S in structures 3QM4 and 3TDA.

Singlet gating models in CYP2D6

Table 3.15 shows the CYP2D6 singlet gating models identified for pw2c and tunnel W. Gating residues could not be identified for pw3 as the protein backbone of tunnel-closed structures obstructs this tunnel.

Table 3.15 Singlet gating models in CYP2D6

Gated tunnel	Structures: open/closed	Gating residue
pw2c	1OG5(CYP2C9)/2F9Q(CYP2D6)	**Leu 110, Leu 213, Glu 216**
W	2V0M(CYP3A4)/2F9Q(CYP2D6)	**Trp 128, Arg 133**
	2V0M(CYP3A4)/3TBG(CYP2D6)	**Trp 128, Arg 129**

3.2.6 Differential tunnel-opening patterns and gating models in CYP3A4 structures

Table 3.16 shows the distribution of open tunnels in CYP3A4 crystal structures. Differential tunnel-opening patterns were observed for REMD tunnels pw2a, pw2b, pw3 and tunnel S. Pw2c and tunnel W is closed in all structures while pw2e is open in all structures.

Table 3.16 CAVER-calculated tunnels in CYP3A4 crystal structures

PDB	Tunnel No.	pw2a	pw2b	pw2c	pw2e	pw3	S	W	*pw2ac*	*pw6*	Width(Å)
1TQN	1		×								2.33
	2				×						1.86
	3						×				1.43
	4						×				1.43
1W0F	1		×								2.21
	2				×						1.87
1W0G	1		×								2.31
	2				×						1.90
2J0D	1[a]	×*									3.29
	2[b]										3.23
	3		×								2.40
	4	×									2.24
	5[b]										2.12
	6						×*				1.73
	7				×						1.69
2V0M	1		×								2.41
	2	×									1.56
	3				×						1.55

	4					×					1.53
	5	×									1.50
3NXU	1		×								2.09
	2				×						1.83
4K9T	1	×									2.36
	2				×						1.49
	3	×									1.43
4K9U	1[c]										3.92
	2[d]										2.46
	3		×								2.16
	4				×						1.69
	5					×					1.42
	6[e]						×*				1.41
4K9V	1		×								2.31
	2				×						1.84
4K9X	1[f]										2.39
	2		×								2.16
	3				×						1.62
	4[f]										1.51
4K9W	1		×								2.13
	2				×						1.68
3TJS	1[g]										2.45
	2		×								2.23
	3[h]						×*				1.89
	4				×						1.56
4I3Q	1		×								2.09
	2						×				1.98
	3				×						1.81
	4						×				1.59
4I4G	1		×								1.96
	2				×						1.68
4I4H	1		×								2.17
	2				×						1.83
	3		×								1.66
3UA1	1		×								2.35
	2				×						1.89

* indicates tunnel may be an artefact of regions missing from the crystal structure.

a This tunnel is probably an artefact of the gap created by missing residues 214-218 in 2J0D. It lies in close proximity to canonical pw2a however, others have assigned it as pw2f.

b This tunnel is probably an artefact of the gap created by missing residues 214-218 in 2J0D. It falls between canonical tunnel S and pw2a.

c This tunnel is probably an artefact of the gap created by missing residues 210-215 in 4K9U and it is not clear whether the tunnel would pass through the F-G loop (pw3) or in the region of tunnel S.
d This tunnel is probably an artefact of the gap created by missing residues 210-215 in 24K9U. It falls between canonical tunnel S and pw2a
e This is probably an artefact of the gap created by missing residues 210-215 in 4K9U.
f This is probably an artefact of the gap created by missing residues 212-214 in 4K9X. It falls between canonical tunnel S and pw2a.
g This is probably an artefact of the gap created by missing residues 212-218 in 3TJS. It falls between canonical tunnel S and pw2a
h This is probably an artefact of the gap created by missing residues 212-218 in 3TJS

Two-state gating models in CYP3A4

Differential tunnel-opening patterns in CYP3A4 static crystal structures revealed two-state gating models for tunnels pw2a, pw2b, pw3 and tunnel S (Table 3.17).

A network of aromatic interactions gate pw2a in CYP3A4. Phe 215, a residue within the F-G loop region, is the main gating residue, intersecting pw2a in all the tunnel-closed CYP3A4 structures (except for 4K9U and 3TJS where a region of the protein containing this residue is missing from the crystal structure). CYP3A4 structures can be clustered into 3 groups based on the position of Phe 215 side chain (Figure 3.12): 1) Tunnel-open structures where the side chain faces away from pw2a and the active-site cavity towards the bulk solvent. 2) Tunnel-closed structures where Phe 215 interacts with Phe 57 and Phe 53, with its side chain partially intersecting pw2a. 3) Tunnel-closed structures where Phe 215 lies in the middle of pw2a, closing the tunnel completely, with its side chain pointing towards the active site forming hydrophobic interactions with Phe 57, Phe108 and Phe 220. These three clusters provide a visual model for the closing of pw2a where Phe 215 makes a series of aromatic contacts, first interacting with Phe 57 and Phe 53 in the intermediate state, followed by interactions with Phe 108 and Phe 220, closing the tunnel further. The Phe 215 – Phe 57 interactions appears to be the most vital for tunnel-closing and is present in both the intermediate and final closed states.

Pw2b is open in 15 out of the 16 CYP3A4 crystal structures. In the closed structure, 4K9T, F108 is flipped around with the side chain intersecting pw2b instead of falling within the centre of the B-C loop bordering pw2e (Figure 3.13). F108 forms tunnel-closing interactions with I223 and R106. F108 also interacts with F220 but an aromatic interaction between F108 and F220 is also present in the tunnel-open structure so by our definition it is not a tunnel-closing interaction. It is however in a different position and orientation in the tunnel-closed structure so it is likely to break and then reform on tunnel-closure.

Table 3.17 Two-state gating models in CYP3A4

Gated tunnel	Structures: open/closed	Tunnel closing interaction	
pw2a	2V0M/(1TQN, 1W0F, 1W0G, 3NXU, 4K9V, 4K9X, 4K9W, 4I3Q, 4I4G, 3UA1)	**Phe 215**-Phe57	hydrophobic (aromatic)
	2V0M/(4K9X, 4K9W, 4I4G, 3NXU)	**Phe 215**-Tyr 53	hydrophobic (aromatic)
	2V0M/(1TQN, 1W0F, 1W0G, 4K9V, 4I3Q, 3UA1)	**Phe 215**-Phe 108 **Phe 215**-Phe220	hydrophobic (aromatic) hydrophobic (aromatic)
pw2b	1TQN/4K9T	**Phe 108**-I223 **Phe 108**-R106	hydrophobic cation-pi
pw3 (model 1)	2V0M/(1TQN, 1W0F, 1W0G, 4K9V, 4I3Q, 3UA1, 4K9T)	**Phe 213**-Phe 304 **Phe 213** –Phe 108 **Phe 213**-Phe 215 **Phe213**-Phe 241 **Phe 213**-V240	hydrophobic (aromatic) hydrophobic (aromatic) hydrophobic (aromatic) hydrophobic (aromatic) hydrophobic
	2V0M/(1TQN, 1W0F, 1W0G, 4K9V, 4I3Q, 3UA1)	**Phe 304**-Phe 241	hydrophobic (aromatic)
pw3 (model 2)	2V0M/(4K9W, 4I4G, 3NXU, 4I4H)	**Leu 211**- Phe220	hydrophobic
S	1TQN/(1W0F, 4I4G, 4K9T, 4I4H, 2V0M)	**Arg 212**-Asp 214	main chain-main chain H-bond (and ionic in some structures)
	1TQN/(1W0F, 1W0G, 3UA1, *4I4H)*	**Arg 212**-Glu 308	side chain-side chain and ionic
	1TQN/1W0G	**Arg 212**-Leu 483	main chain-side chain H-bond
	1TQN/3UA1	**Arg 212**-Ile 369	main chain-side chain H-bond
	1TQN/(2V0M, 4K9W, 4I4G, *3NXU*)	**Leu 482**-Leu 211	hydrophobic
	1TQN/(2V0M, 4K9W, *3NXU*)	**Leu 482**-Phe 304	hydrophobic
	1TQN/(2V0M, 4K9X, 4K9T)	**Leu 482**-Ile 369	hydrophobic
	1TQN/(3NXU, *4I4G, 4K9X*)	**Gln 484**-Glu 308)	side chain-side chain H-bond

Tunnel-intersecting residues are shown in bold face. Structures where the tunnel-closing interactions is present but the intersecting residue indicated in bold does not intersect the base of tunnel S (*i.e.* it is not the main residue responsible for tunnel-closure) are shown in italics.

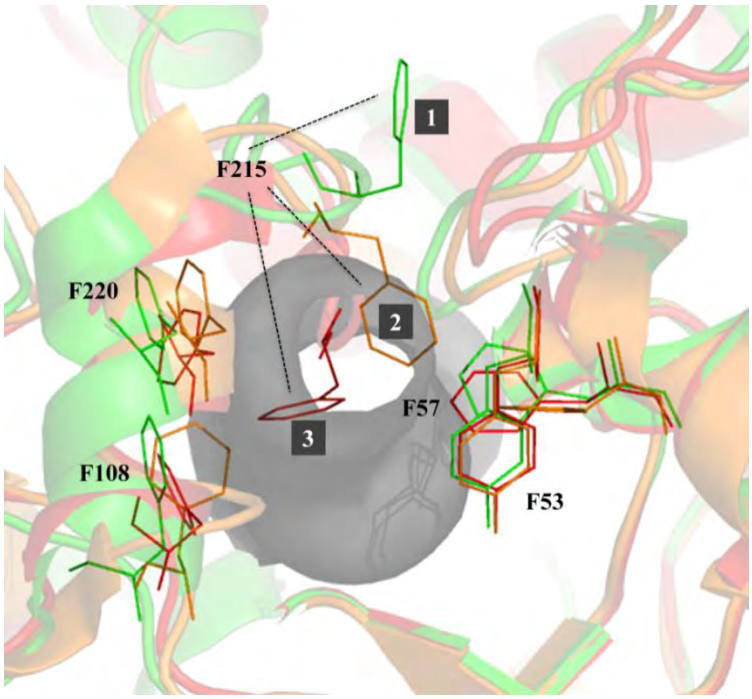


Figure 3.12 Gating model for tunnel pw2a in CYP3A4
CYP3A4 structures cluster into 3 groups based on the position of F215. A representative structure from each cluster is shown here: 1) 2V0M (green) with open tunnel pw2a aligned to tunnel-closed structures 2) 4K9X (orange) and 3) 1TQN (red). F215 forms tunnel-closing interactions with F57 and F53 in structures in cluster 2 (4K9X, 4K9W, 3NXU and 4I4G) and tunnel-closing interactions with F57, F108 and F220 in structures in cluster 3 (1TQN, 1W0F, 1W0G, 4K9V, 4I3Q, 3UA1) that are absent in the tunnel-open structures (cluster 1).

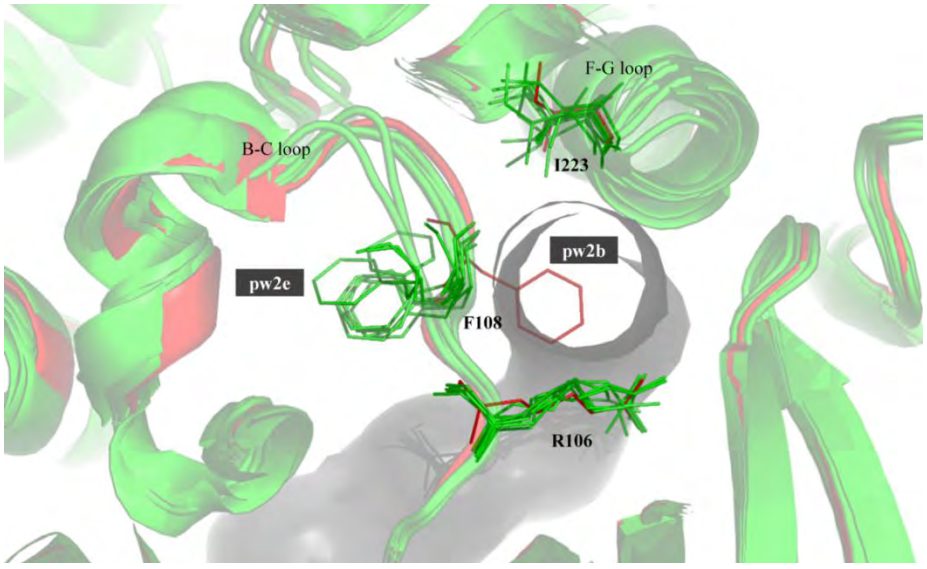


Figure 3.13 Tunnel gating model for pw2b in CYP3A4
CYP3A4 tunnel-open structures (1TQN, 1W0F, 1W0G, 2J0D, 2V0M, 3NXU, 4K9U, 4K9V, 4K9X, 4K9W, 3TJS, 4I3Q, 4I4G, 4I4H and 3UA1) (green) with open tunnel pw2b aligned to tunnel-closed structure 4K9T (red) in PyMOL. The side chain of F108 intersects pw2b in the tunnel-closed structure with tunnel-closing interactions F108-I223 and F108-R106. In the tunnel-open structures, F108 is flipped around facing towards pw2e instead of pw2a and does not make contact with R106 or I223.

Tunnel pw3 is open in 2 (2V0M and 4K9U) of the 16 CYP3A4 structures. While the main part of the tunnel is similar in the two tunnel-open structures, the exit points vary considerably. Pw3

is shorter in 2V0M and exits the molecule by a more direct route between the loop region flanked by helix F and helix F' and the loop region flanked by helix G and helix G'. Pw3 bends over helix I in 4K9U exiting between the N-terminal of helix F, the C-terminal of helix G and the E-F loop region. Structure 4K9U has a missing region between helix F and F', which boarders the tunnel so only pw3 in 2V0M was used to identify gating residues in the tunnel-closed CYP3A4 structures. Structures 2J0D, 4K9X, 3TJS also have regions missing that border pw3 so they were excluded from the analysis. In four of the eleven tunnel-closed CYP3A4 structures included in the analysis, L211 intersects pw3; in six of the structures, F213 and F304 intersect pw3 and in the remaining structure, F213 but not F304 intersects pw3. According to the definition of open/closed tunnels and gating residues used in this study, there are at least two different gating models for pw3. The tunnel-closing interactions making up each of these models are shown in Table 3.17 and illustrated in Figure 3.14 A and B.

CAVER identified a total of 4 open solvent tunnels that are not artefacts of missing residues in CYP3A4 structures. Two of these solvent tunnels are present in 1TQN and are essentially the same tunnel branching into two near the surface due the CAVER algorithms use of grid points (Figure 3.15). The other two solvent tunnels were identified in 4I3Q and are variants of the canonical tunnel S. Since the solvent tunnels in 1TQN most closely resemble the canonical tunnel S identified in REMD analysis, they were used to identify gating residues – many of the gating residues identified however also intersect the solvent tunnels in 4I3Q.

Tunnel-intersecting residues responsible for tunnel S closure were identified on the basis that they intersected both the open solvent tunnels in 1TQN, *i.e.* intersect the tunnel in the base region rather than the branch region, thereby closing both tunnels. Tunnel-closed CYP3A4 structures 2J0D, 4K9U and 3TJS have missing residues in the solvent tunnel region so they were excluded from the analysis.

Arg 212, Leu 482 or Gln 484 or a combination of these residues intersect the solvent tunnel in the tunnel-closed CYP3A4 structures (Figure 3.16). There were no tunnel-closing interactions common to all tunnel-closed CYP3A4 structures and tunnel-intersecting residues make several different interactions that result in tunnel closure. These interactions are listed in Table 3.17.

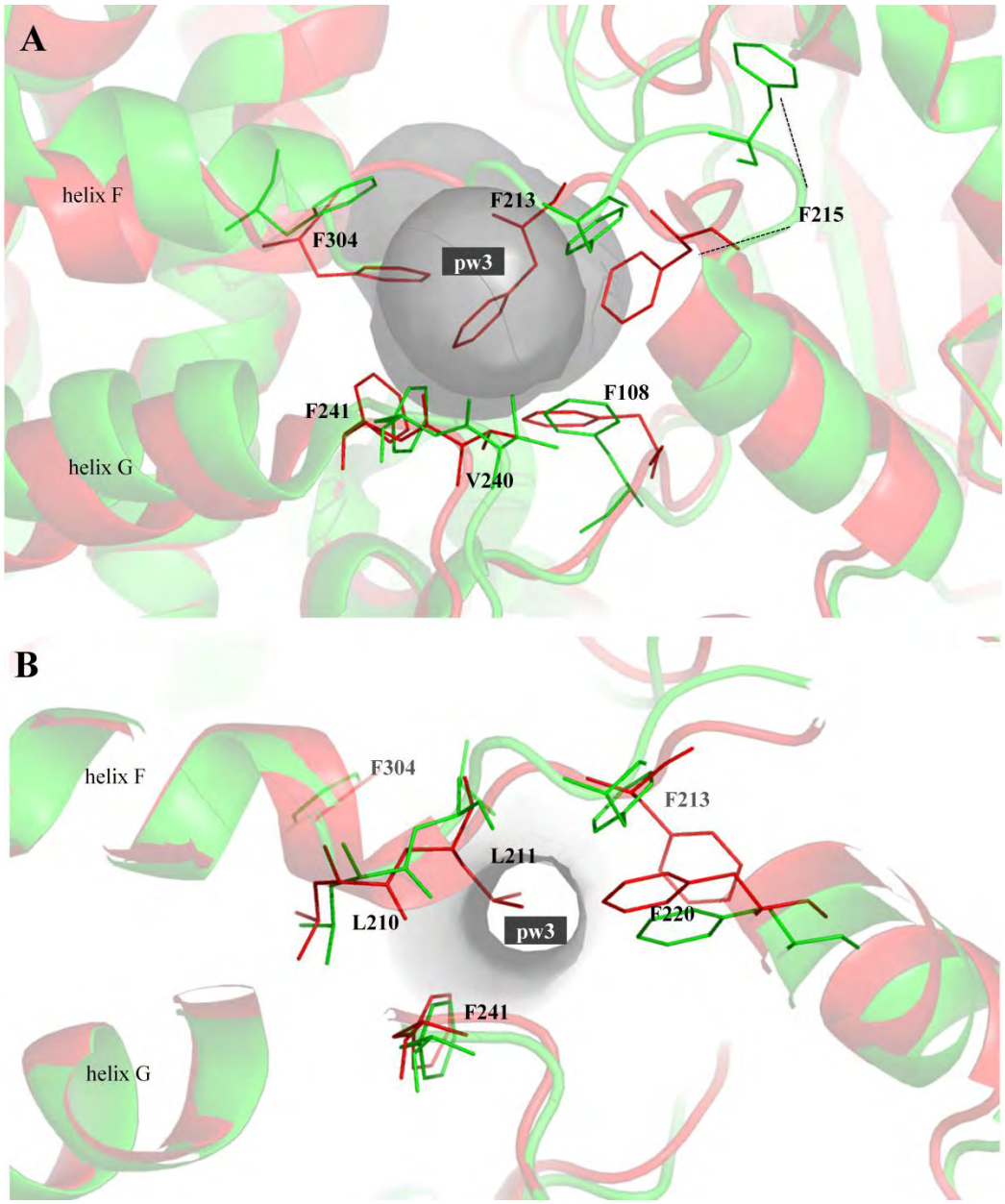


Figure 3.14 Gating models for pw3 in CYPA4
CYP3A4 structure 2V0M (green) with open tunnel pw3 (grey) aligned to tunnel-closed structure 1TQN (red) in (A), illustrating pw3 gating model 1, and aligned to 4I4H (red) in (B) illustrating pw3 gating model 2. Residues involved in tunnel-closing interactions are indicated are labelled. (A) F213 and F304 intersect tunnel pw3 in structure 1TQN and are involved in the following tunnel-closing interactions that are absent in the tunnel-open structure: F213-F304, F213-F108, F213-F215, F213-V240, F213-F241, F304-F241. F213 also interacts with F220 however this interactions is also present in the tunnel-open structure. (B) L211 intersects tunnel pw3 in structure 4I4H and is involved in the following tunnel-closing interactions that are absent in the open structure: L211-L210, L211-F220 and L211-F241. L211-F220 is however the only tunnel-closing interaction present in all four tunnel-closed structures where pw3 is gated by L211. F213 and F304 are not involved in the gating of pw3 in this structure but their positions are also labelled for comparison.

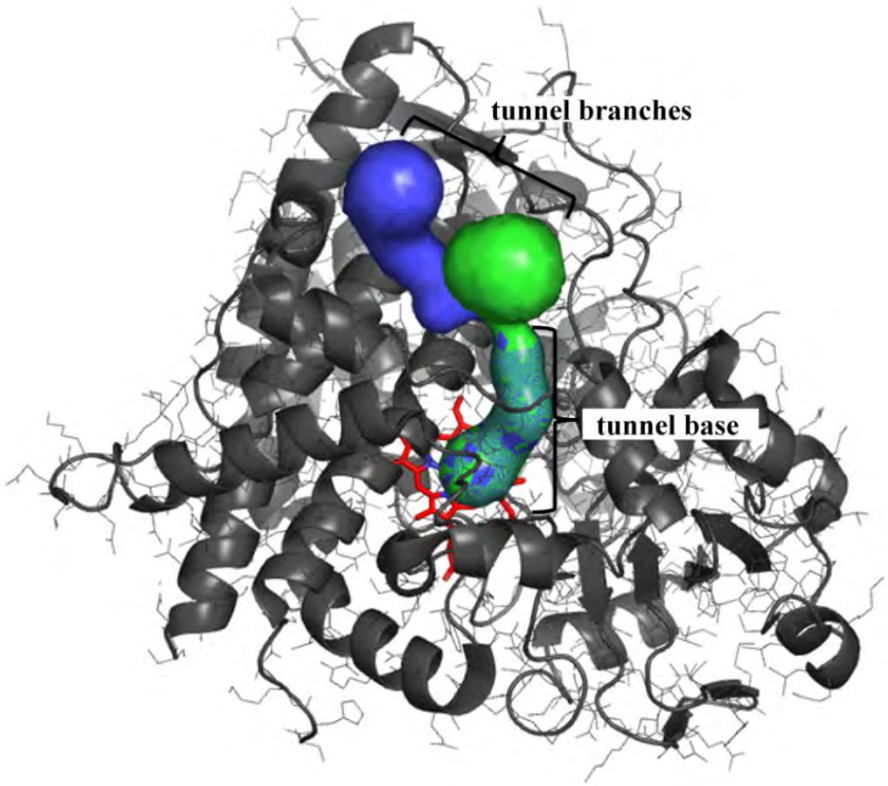


Figure 3.15 Solvent tunnels calculated by CAVER in CYP3A4 structure 1TQN.
Solvent tunnels calculated by CAVER are shown in green and blue (tunnels 3 and 4 in Table 3.16 respectively) both with a gorge width of 1.43 Å. The tunnels are identical at the base branching into the surface of the molecule. The branches are not separated by the protein structure.

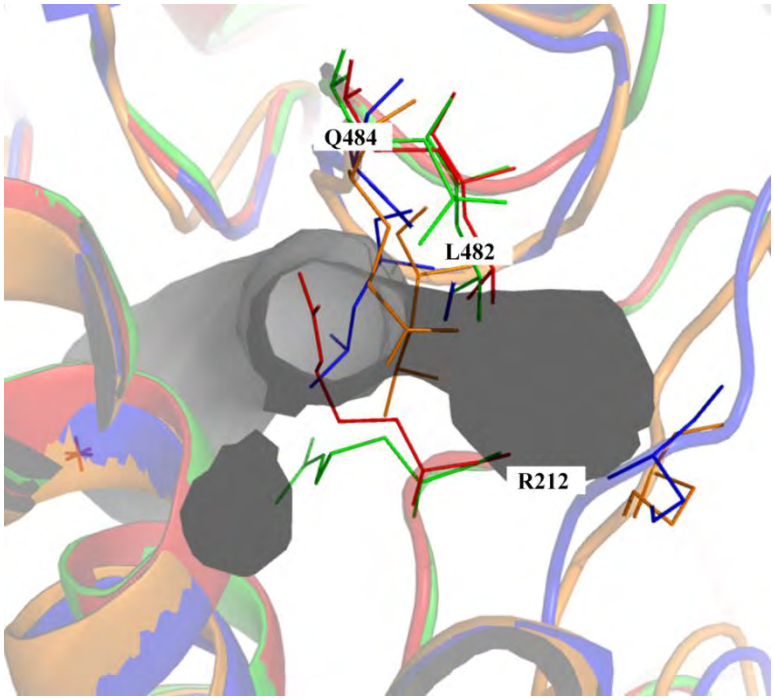


Figure 3.16 Examples of tunnel S intersecting residues in CYP3A4 structures
CYP3A4 structure 1TQN (green) with open tunnel S (grey) was aligned to tunnel-closed structures 1W0F (red) with tunnel-intersecting residue R212, 2V0M (blue) with tunnel-intersecting residue L484 and 3NXU (orange) with tunnel-intersecting residue Q484. The figure shows a cross-section through the base of the branched tunnel S.

Singlet gating models in CYP3A4

Table 3.18 shows the CYP3A4 singlet gating models identified for pw2c and tunnel W. Gating residues were not delimited for pw2e as this tunnel was open in all 16 CYP3A4 crystal structures.

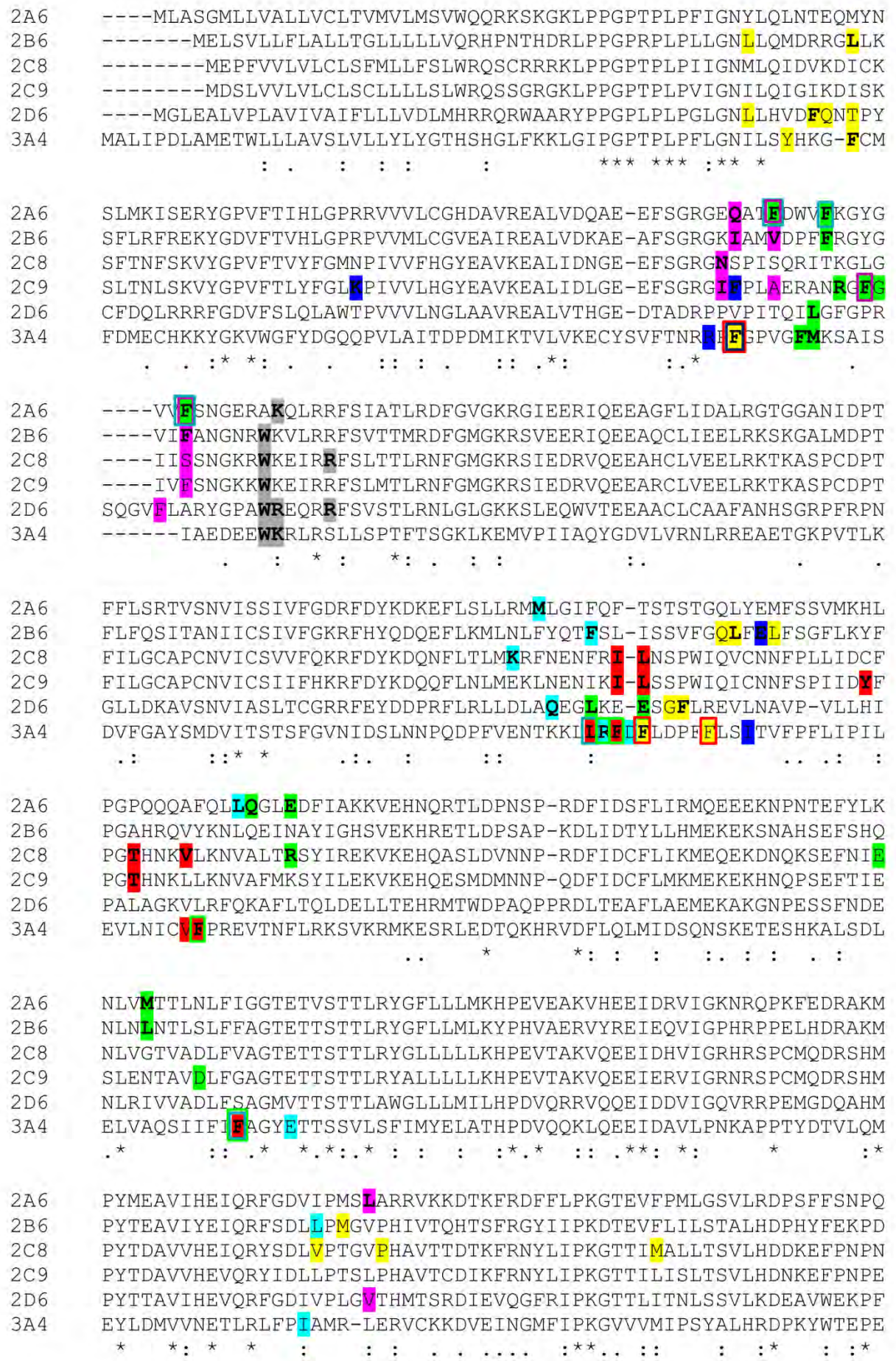
Table 3.18 Singlet gating models in CYP3A4

Gated tunnel	Structures: open/closed	Gating residue
pw2c	1OG5(CYP2C9)/1TQN(CYP3A4)	**Phe 113, Met 114, Arg 212, Phe 213, Phe 304, Phe 241**
W	2FDV(CYP2A6)/1TQN(CYP3A4)	**Trp 126, Lys 127**

3.2.7 Analysis of gating models

An annotated sequence alignment shows the relative positions of gating residues making up the singlet and two-state gating models in isoforms CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2D6 and CYP3A4 (Figure 3.17).

Tunnel-intersecting residues identified in each of the six CYP450 isoforms were classified according to their side chain chemical types (Table 3.19). 'Hydrophobic aromatic' was the most common side chain chemical type observed in tunnel gating, making up the highest percentage of tunnel-gating residues in all isoforms except CYP2B6 and CYP2C8. Only 10% of CYP2C8 gating residues have aromatic side chains, with the majority having polar or charged side chains. In contrast, half CYP2B6 gating residues are 'hydrophobic aliphatic' and only 10 % have polar or charged side chains.


2A6 ----MLASGMLLVALLVCLTVMVLMSVWQQRKSKGKLPPGPTPLPFIGNYLQLNTEQMYN
2B6 -------MELSVLLFLALLTGLLLLLVQRHPNTHDRLPPGPRPLPLLGNLLQMDRRGLLK
2C8 --------MEPFVVLVLCLSFMLLFSLWRQSCRRRKLPPGPTPLPIIGNMLQIDVKDICK
2C9 --------MDSLVVLVLCLSCLLLLSLWRQSSGRGKLPPGPTPLPVIGNILQIGIKDISK
2D6 ----MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQNTPY
3A4 MALIPDLAMETWLLLAVSLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGNILSYHKG-FCM
2A6 SLMKISERYGPVFTIHLGPRRVVVLCGHDAVREALVDQAE-EFSGRGEQATFDWVFKGYG
2B6 SFLRFREKYGDVFTVHLGPRPVVMLCGVEAIREALVDKAE-AFSGRGKIAMVDPFFRGYG
2C8 SFTNFSKVYGPVFTVYFGMNPIVVFHGYEAVKEALIDNGE-EFSGRGNSPISQRITKGLG
2C9 SLTNLSKVYGPVFTLYFGLKPIVVLHGYEAVKEALIDLGE-EFSGRGIFPLAERANRGFG
2D6 CFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGE-DTADRPPVPITQILGFGPR
3A4 FDMECHKKYGKVWGFYDGQQPVLAITDPDMIKTVLVKECYSVFTNRRPFGPVGFMKSAIS
2A6 ----VVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIDALRGTGGANIDPT
2B6 ----VIFANGNRWKVLRRFSVTTMRDFGMGKRSVEERIQEEAQCLIEELRKSKGALMDPT
2C8 ----IISSNGKRWKEIRRFSLTTLRNFGMGKRSIEDRVQEEAHCLVEELRKTKASPCDPT
2C9 ----IVFSNGKKWKEIRRFSLMTLRNFGMGKRSIEDRVQEEARCLVEELRKTKASPCDPT
2D6 SQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSGRPFRPN
3A4 ------IAEDEEWKRLRSLLSPTFTSGKLKEMVPIIAQYGDVLVRNLRREAETGKPVTLK
2A6 FFLSRTVSNVISSIVFGDRFDYKDKEFLSLLRMMLGIFQF-TSTSTGQLYEMFSSVMKHL
2B6 FLFQSITANIICSIVFGKRFHYQDQEFLKMLNLFYQTFSL-ISSVFGQLFELFSGFLKYF
2C8 FILGCAPCNVICSVVFQKRFDYKDQNFLTLMKRFNENFRI-LNSPWIQVCNNFPLLIDCF
2C9 FILGCAPCNVICSIIFHKRFDYKDQQFLNLMEKLNENIKI-LSSPWIQICNNFSPIIDYF
2D6 GLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKE-ESGFLREVLNAVP-VLLHI
3A4 DVFGAYSMDVITSTSFGVNIDSLNNPQDPFVENTKKLLRFDFLDPFFLSITVFPFLIPIL
2A6 PGPQQQAFQLLQGLEDFIAKKVEHNQRTLDPNSP-RDFIDSFLIRMQEEEKNPNTEFYLK
2B6 PGAHRQVYKNLQEINAYIGHSVEKHRETLDPSAP-KDLIDTYLLHMEKEKSNAHSEFSHQ
2C8 PGTHNKVLKNVALTRSYIREKVKEHQASLDVNNP-RDFIDCFLIKMEQEKDNQKSEFNIE
2C9 PGTHNKLLKNVAFMKSYILEKVKEHQESMDMNNP-QDFIDCFLMKMEKEKHNQPSEFTIE
2D6 PALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDE
3A4 EVLNICVFPREVTNFLRKSVKRMKESRLEDTQKHRVDFLQLMIDSQNSKETESHKALSDL
2A6 NLVMTTLNLFIGGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRAKM
2B6 NLNLNTLSLFFAGTETTSTTLRYGFLLMLKYPHVAERVYREIEQVIGPHRPPELHDRAKM
2C8 NLVGTVADLFVAGTETTSTTLRYGLLLLLKHPEVTAKVQEEIDHVIGRHRSPCMQDRSHM
2C9 SLENTAVDLFGAGTETTSTTLRYALLLLLKHPEVTAKVQEEIERVIGRNRSPCMQDRSHM
2D6 NLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHM
3A4 ELVAQSIIFIFAGYETTSSVLSFIMYELATHPDVQQKLQEEIDAVLPNKAPPTYDTVLQM
2A6 PYMEAVIHEIQRFGDVIPMSLARRVKKDTKFRDFFLPKGTEVFPMLGSVLRDPSFFSNPQ
2B6 PYTEAVIYEIQRFSDLLPMGVPHIVTQHTSFRGYIIPKDTEVFLILSTALHDPHYFEKPD
2C8 PYTDAVVHEIQRYSDLVPTGVPHAVTTDTKFRNYLIPKGTTIMALLTSVLHDDKEFPNPN
2C9 PYTDAVVHEVQRYIDLLPTSLPHAVTCDIKFRNYLIPKGTTILISLTSVLHDNKEFPNPE
2D6 PYTTAVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPF
3A4 EYLDMVVNETLRLFPIAMR-LERVCKKDVEINGMFIPKGVVVMIPSYALHRDPKYWTEPE

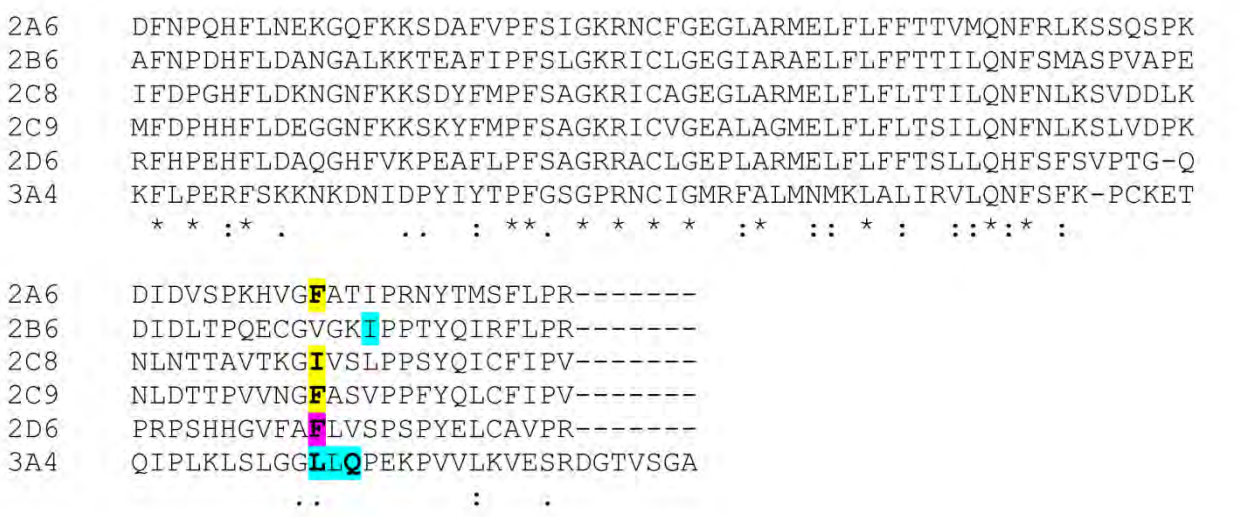


Figure 3.17 Multiple sequence alignment of human CYP450 isoforms showing residues forming part of tunnel gating models. Sequences were retrieved from the UniProt Data Base. Tunnel-intersecting residues are shown in bold while their interacting partners are shown in plain font. Residues are colour coded according to the tunnel they gate: pw2a (yellow), pw2b (blue), pw2c (green), pw2e (purple), pw3 (red), tunnel S (cyan), tunnel W (grey). Residues that gate multiple tunnels are boxed according to the colours of the addition tunnels they gate.

Table 3.19 A summary of the distribution of tunnel-intersecting residue side chain chemical types

Isoform	Total	Hydrophobic aromatic		Polar/Charged		Hydrophobic aliphatic	
CYP2A6	18	10	**(55.5 %)**	4	(22.2 %)	4	(22.2 %)
CYP2B6	10	4	(40.0 %)	1	(10.0 %)	5	**(50.0 %)**
CYP2C8	10	1	(10.0 %)	5	**(50.0 %)**	4	(40.0 %)
CYP2C9	11	5	**(45.5 %)**	3	(27.3 %)	3	(27.3 %)
CYP2D6	10	4	**(40.0 %)**	4	(40.0 %)	2	(20.0 %)
CYP3A4	16	9	**(56.3 %)**	4	(25.0 %)	3	(18.8 %)
Total	75	33	**(44.0 %)**	21	(28.0%)	21	(28.0 %)

Note: If a residue intersected two different tunnels it the same isoform it was counted twice. The chemical-type with the highest percentage of intersecting residues for each isoform is shown in bold.

The different usage of side chain chemical types in tunnel gating within the CYP2 family, particularly between CYP2C8 and CYP2C9 that have high sequence similarity, suggests that the evolution rates may be high for tunnel gating residues. To test this hypothesis, the rate of evolution of each sequence positions in the CYP2 family was determined using rate4site software as described in the Methods section. Table 3.20 shows the average relative rate of evolution of tunnel-intersecting residues positions for each isoform. Calculated rates of evolution were normalised so that the average rate of evolution of all the positions in the sequence was zero and the standard deviation was one. Negative scores therefore represent positions that are more conserved than average and positive scores indicated positions that are less conserved than

average. The average rate of evolution of tunnel-gating positions in all isoforms was found to be significantly higher than the average rate of evolution of the entire sequence.

Table 3.20 Average rate of evolution of tunnel-intersecting residues in CYP2 isoforms

Isoform	Average relative rate of evolution for tunnel-intersecting residues	Student t-test P Value
CYP2A6	1.1402	0.00006
CYP2B6	0.9881	0.00104
CYP2C8	0.9740	0.00123
CYP2C9	1.4125	0.00000
CYP2D6	0.7668	0.00850

The average rate of evolution of the entire sequence is zero and the standard deviation is 1. The null hypothesis that the average rate of evolution of the entire sequence is the same as the average rate of tunnel-intersecting residues was rejected for the alternative hypothesis that the average rate of evolution of the tunnel-intersecting residues is higher than the average rate of evolution of the entire sequence for all isoforms at the 5% level.

3.3 Discussion

Based on the assumptions that amino acid side chains play a critical role in tunnel gating and that CAVER calculated tunnels are sufficiently similar to canonical REMD tunnels, gating models for a total of 34 CYP450 tunnels in six of the major human drug metabolising isoforms were identified using *in silico* tools.

Using a 1.4 Å gorge width cut-off, 75% of the open tunnels identified by CAVER were assigned to REMD tunnels. The assignment of CAVER tunnels to REMD tunnels, pw2b, pw2c, pw2e and tunnel W was straightforward. There were however cases where the assignment of pw2a and tunnel S was less clear as a number of tunnel variants were calculated for these tunnels, some falling midway between pw2a and tunnel S. Of the remaining CAVER tunnels that did not match REMD tunnels, two thirds were assigned to the newly defined tunnel pw6, however there are no REMD studies to date confirming that substrate molecules use this pathway so gating models were not delimited for this tunnel. There were no CAVER tunnels with a gorge width greater than 1.4 Å that resembled the REMD tunnels pw1 or pw2d in the crystal structures analysed in this study.

It was not possible to identify gating models for all REMD equivalent tunnels calculated by CAVER. Neither two-state nor singlet gating models could be delimited for tunnels that were open in all structures of a given isoform, as was the case for CYP2C8 pw2b, CYP2C9 tunnel S, CYP2D6 pw2b and CYP3A4 pw2e. In other cases, gating models could not be identified because the protein backbone intersected the tunnel of interest substantially in tunnel-closed structures, as was the case for CYP2A6 pw2b, CYP2A6 pw3 CYP2B6 pw3 and CYP2D6 pw3.

Gora *et al* recently reviewed different gating mechanisms in enzymes, classifying them into six groups based on their structure: 1) “wing” gates controlled by the rotation of a single side chain; 2) “swinging doors” controlled by the rotation of the side chains of two residues distant in sequence, 3) “aperture” gates controlled by the backbone motion of 3-4 residues, distant in sequence, resulting in tunnel constriction; 4) “drawbridges” or 5) “double drawbridges” controlled by the movement of one or two fragments consisting of 3-20 residues neighbouring in

sequence and 5) "shell" gates based on the movement of entire domains separated by a hinge region [430].

The approach used in this study is most suited to identifying "wings" and "swinging" doors, with the majority of gating models identified here falling into the "wings" category, typified by the gating models observed for CYP3A4 pw2a and pw2b (Figure 3.12 and Figure 3.13). Aperture gating models were however also observed since, although backbone motions control these gates, the movement of the backbone results in side chains moving closer together and intersecting the tunnel rather than intersection by the backbone itself. Gating models for CYP2D6 pw2a and pw2e (Figure 3.9 and Figure 3.10 A) provide good examples of aperture gates as the side chains of the tunnel-intersecting residues move closer together in the closed structure without changing orientation *i.e*. the tunnel appears to constrict rather than close due to the swinging motion of the side chains. In cases where backbone regions of the protein intersect the tunnel, it is difficult to predict the gating mechanism as larger conformational changes in the protein structure occur that may involve aperture or drawbridge gating mechanisms.

Table 3.21 Human CYP450 gating models identified in molecular dynamics studies

Tunnel	Gating residues	Products used in SMD simulations
CYP3A4 pw2a	**F57-F215**	temazepan and testosterone-6OH
CYP3A4 pw2b	**F108**-F220	temazepan and testosterone-6OH
CYP3A4 pw2c	F108-**F241**	temazepan and testosterone-6OH
CYP3A4 pw2e	F108-F220	temazepan
	R105-S119	testosterone-6OH
CYP3A4 pw3	**F213-F241**	temazepan
	F213-V240	testosterone-6OH
CYP3A4 S	**R212-L482**	temazepan and testosterone-6OH

Interactions that appear in the two-state gating models in this Chapter, matched to the same tunnel, are underlined and in bold. Residue that appear in two-state or singlet gating models, matched to the same tunnel but not necessarily the same interaction partner, are shown in bold.

CYP3A4 is the only human CYP450 isoform for which gating models have been identified in molecular studies to date. Table 3.21 gives a list of these gating models, identified by Fishelovitch *et al* using steered molecular dynamic simulations [434]. Here, three of the same gating interactions in CYP3A4 were identified and matched to the same tunnel as Fishelovitch *et al.* Notably, the CYP3A4 pw2b F108-F220 interaction was identified as an interaction that would break and reform in a new position/orientation during tunnel closure, but because it was

present in both the tunnel-closed and tunnel-open structures it was not included as a tunnel-closing interaction. Table 3.21 also shows that 62.5% (10/16) of the residues that appear in the molecular dynamics gating models also appear in the gating models delimited in this Chapter, matched to the same tunnel. As noted above, gating models for CYP3A4 pw2e could not be delimited using the approach followed here; if we exclude pw2e literature gating residues from the calculation, this study identified 10/12 (83%) gating residues previously reported for studies using molecular dynamics simulations; this indicates that a large proportion of gating residues identified in molecular dynamic studies can also be identified in static crystal structure using the far simpler approach described here.

While molecular dynamic studies generally identify one interacting partner per gating residue, the present work regularly identified multiple tunnel-closing interacting partners for each tunnel-intersecting residue. This suggests that either molecular dynamic studies are more proficient at identifying a smaller number of true gating interactions or, perhaps more likely, that gating mechanisms are more complex than previously thought. Coordinated motion of side chains within aromatic residue clusters of gating residues (e.g. Figure 3.18) may provide one such more complex mechanism for coordinating the opening and closing of different tunnels.

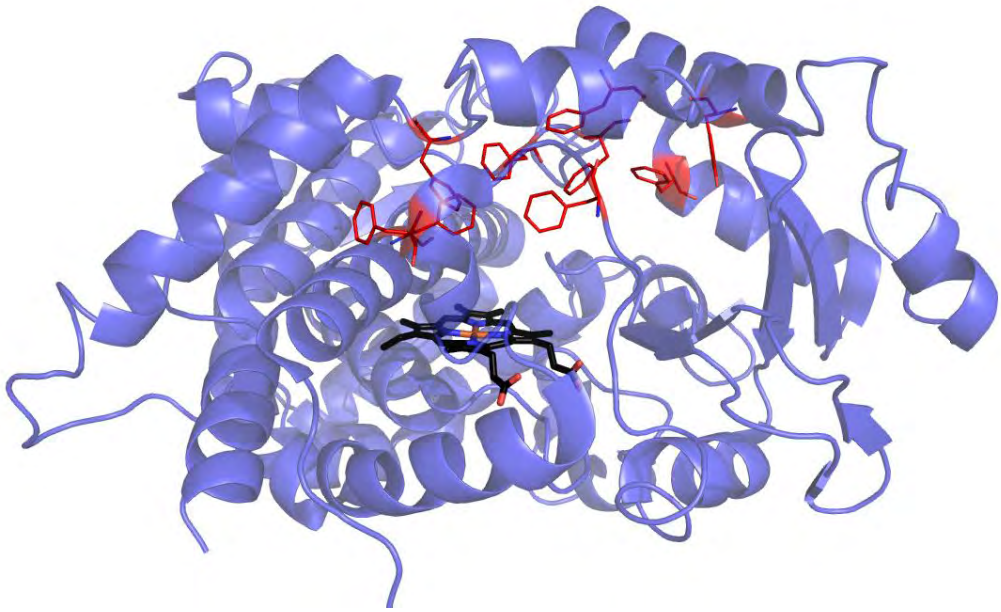


Figure 3.18 Aromatic cluster of gating residues found in CYP3A4. CYP3A4 structure 1TQN is shown in blue with the haem group in black and aromatic residues appearing in two-state and singlet gating models in this work shown in red.

An annotated sequence alignment of the CYP450 isoforms used in this study shows that gating residues regularly fall at evolutionarily equivalent positions in different isoforms, yet despite this, the amino acid types at these positions, even in closely related isoforms, often differ. The rates of evolution of gating residue positions compared to the average rate of evolution of the entire sequence within the CYP2 family suggests that gating residues are subject to fast evolutionary dynamics. Ligand tunnels can play a role in substrate specificity and the varied amino acid types used in tunnel gating in different isoforms seems likely play an important role in selective substrate entry. Tunnel gates can be found at any position along the tunnel but are most regularly found at the entrance to the tunnel, at the bottleneck of the tunnel (narrowest point) or near the end of the tunnel at the entrance to the active site cavity [430]. The fact that different viable tunnel gating combinations may evolve at any position along the tunnel may explain the fast evolutionary dynamics observed for gating residues.

The gating of tunnel W by the highly conserved tryptophan residues, present in all isoforms except CYP2A6, is the most conserved gating model observed here. Controlling the flow of water molecules in and out of the active site is vital for the CYP450 catalytic cycle. Molecular dynamic studies using CYP3A4 have also found that Arg 375, another highly conserved residue, is important for the CPR mediated control of tunnel W [442]. This further suggests that in contrast to the gating of tunnels used by substrate molecules, water channel gating is conserved across isoforms.

In conclusion, the results presented here suggest that a simple and rapid approach using static crystal structures and a combination of tools such as CAVER and Protein Interaction Calculator provides a comprehensive alternative to complex molecular dynamic studies for identify gating models for substrate access and egress pathways. This work reports the largest set of gating models for CYP450 tunnels to date, identifying 49 tunnel-closing interactions (two-state gating models) and an additional 54 tunnel-intersecting residues (singlet gating models) thereby adding valuable insight into CYP450 tunnel gating and setting a benchmark for future molecular dynamic studies.

3.4 Methods

3.4.1 Structures

Structures were retrieved from the RCSB Protein Data Bank [31] as follows. **CYP2A6**: 1Z10, 1Z11, 2FDU, 2FDV, 2FDV, 2FDW, 2FDX, 2FDY, 2PG5, 3T3R, 4EJJ, 3T3Q. **CYP2B6:** 3IBD, 3UA5, 4I41, 3QOA, 3QU8. **CYP2C8**: 1PQ2, 2NNH, 2NNI, 2NNJ, 2VN0. **CYP2C9**: 1OG2, 1OG2, 1R90. **CYP2D6**: 3QM4, 2F9Q, 3TDA, 3TBG. **CYP3A4**: 1TQN, 1W0F, 1W0G, 2J0D, 2V0M, 3NXU, 4K9T, 4K9U, 4K9V, 4K9X, 4K9W, 3TJS, 4I3Q, 4I4G, 4I4H, 3UA1.

3.4.2 Tunnel identification using CAVER

Tunnels were identified in each CYP450 structure using the computational geometry tool CAVER [437]. The haem iron coordinates were read in from the PDB file and used as the starting point for tunnel searches and water and ligand molecules and hydrogen atoms were not taken into account. CAVER returns a ranked list of possible routes out of the molecule giving greatest priority to routes with the widest gorge width irrespective of the length of the tunnel. In cases where two tunnels have identical gorge widths the shorter tunnel is prioritised.

Determining the gorge width cut-off

A preliminary study was carried out to determine 1) the variation in the number of tunnel-open/tunnel-closed structure pairs with changes in gorge width cut-off and 2) the resemblance of CAVER calculated tunnels to REMD tunnels with changes in gorge width cut-off. The number of tunnel/open tunnel-closed pairs for each isoform was recorded as the gorge width cut-off was increased from 0.7 Å to 3.3 Å in 0.1 Å increments and tunnels resembling REMD tunnels were assigned to determine the verisimilitude of the accepted tunnels at each gorge width cut-off.

Based on this preliminary study a gorge width cut-off of 1.4 Å was chosen as it gave the best balance between the verisimilitude of calculated tunnels and the number of tunnel-open/tunnel-closed structure pairs for REMD equivalent tunnels. Tunnels with a gorge width cut-off > 1.4 Å were therefore assigned as open tunnels and where possible were assigned to REMD tunnels in accordance with the standard nomenclature [66, 427].

3.4.3 Delimiting two-state gating models

Two-state gating models were identified using structures from the same isoform showing differential tunnel-opening patterns. Tunnel-open and tunnel-closed structure pairs were identified for a given tunnel and a given isoform and aligned in PyMOL with the tunnel of interest in the open structure displayed. Amino acid side chains from the tunnel-closed structure that intersect the tunnel of interest in the tunnel-open structure were manually delimited by searching along the length of the tunnel. In cases where there was more than one tunnel-closed structure for a given isoform and tunnel, all tunnel-closed structures were aligned to the open structure to determine if the same tunnel-intersecting residues were responsible for tunnel closure in all structures. In cases were a large portion of the protein backbone intersected a tunnel, gating residues were not identified.

Residues that directly interact with the tunnel-intersecting residues were identified separately for the tunnel-open and tunnel-closed structures using Protein Interaction Calculator [444]. Hydrophobic, aromatic, main chain-main chain, main chain-side chain, side chain-side chain, ionic and cation-pi intra-protein interactions were identified using default parameters. Interactions with the tunnel-intersecting residue present in the tunnel-closed structure but absent in the tunnel-open structure are assigned as tunnel-closing interactions while those present in the tunnel-open structure but absent in the tunnel-closed structure are assigned as tunnel-opening interactions. Interactions present in both structures were omitted from the two-state gating models.

3.4.4 Delimiting singlet gating models

In cases where all a given tunnel was closed in all structures of a given isoform singlet gating models were identified by aligning the tunnel-closed structures to a tunnel-open structure of a different isoform. As for two-state gating models, residues from the tunnel-closed structures intersecting the tunnel of interest in the tunnel-open structure were manually delimited. Tunnel-closing interactions cannot be determined in these cases as there is no simple way of identifying interactions that occur exclusively in the tunnel-closed structure when the protein sequences differ between structures.

3.4.5 Analysis of gating models

Sequence alignment

Sequences with the following accession numbers were retrieved from the UniProt Data Base and aligned using ClustalW (default parameters) [422]: CYP2A6, P11509 (*Homo sapiens*); CYP2B6, P20813 (*Homo sapiens*); CYP2C8, P10632 (*Homo sapiens*); CYP2C9, P11712 (*Homo sapiens*); CYP2D6, P10635; CYP3A4, P08684 (*Homo sapiens)*. Residues in two-state and singlet gating models were highlighted on the alignment according to the tunnel they gated with tunnel-intersecting residues in bold and interacting residues in plain font.

Distribution of amino acid side chain types

Tunnel-intersecting residues were classified as *hydrophobic aromatic* (Phe, Trp, Tyr), *polar/charged* (Arg, Lys, Glu, Asp, Thr, Gln, Asn, Ser, His, Cys) or *hydrophobic aliphatic* (Leu, Ile, Val, Met, Pro) according to side chain chemical type. The propensity of side chain chemical types responsible for tunnel gating in each isoform was determined by calculating the percentage of tunnel-intersecting residues falling into each group. If the same residue acted as a gating residue for two different tunnels it was counted twice.

Evolutionary dynamics of gating residues in CYP2 family

The following CYP2 sequences were aligned using ClustalW (default parameters) [422]: CYP2A6, P11509 (*Homo sapiens*); CYP2B6, P20813 (*Homo sapiens*); CYP2C8, P10632 (*Homo sapiens*); CYP2C9, P11712 (*Homo sapiens*) and CYP2D6, P10635. PROML (Protein Maximum Likelihood program) in the PHYLIP suit of programs (version 3.695, default parameters, [445]) was used to generate a phylogenetic tree using the ClustalW alignment in PHYLIP format as input. Rate4site (version 2.01 for Windows, default parameters, [446]) was then used to calculate the relative rate of evolution of each sequence position in the CYP2 alignment; the ClustalW alignment and phylogenetic tree were used as input. The rates were normalised so that the average rate of evolution of all positions in the sequence was zero and the standard deviation was one. The average rate of tunnel-intersecting sequence positions for each isoform was calculated and a one-tailed student t-test was performed in Excel to determine if the rate of

evolution of tunnel-gating residues in each isoform was significantly higher than the average rate of evolution of the entire sequence.

Chapter 4 *In silico* analysis of the effect of polymorphic variation on protein stability and function

4.1 Introduction

The sheer number of non-synonymous CYP450 polymorphisms, combined with the observed substrate specific phenotypic effects and the many diverse potential drugs, highlights the need for improved *in silico* prediction tools for determining the effects of single amino acid substitutions in CYP450 enzymes on drug metabolism. Changes in the amino acid sequence of CYP450s can directly affect structural stability, haem binding, substrate access and recognition, as well as interactions with CPR.

Chapters 2 and 3 described the delimitation of substrate recognition sites and tunnel gating residues. In addition to these functional studies, a number of other studies from the Blackburn group and elsewhere have contributed to elucidating the CYP450-CPR interface and have identified clusters of basic residues on the proximal surface of CYP450s that interact with exposed acidic clusters on the CPR FMN domain surface [93-98].

In order to then analyse the impact of polymorphic variation on enzyme stability and function in human CYP450 drug metabolising enzymes, representative *in silico* modelling tools were used to predict the qualitative effect of mutations on protein stability; the resulting data was combined with previous *in silico* analyses of conserved structural features, SRS regions, CPR contacts and tunnel gating residues to create a unified CYP450 SNP map, which was used to correlate the location of individual mutations with their experimentally observed impact on CYP450 function. This work described in this Chapter was published in Current Drug Metabolism in 2013 ([447], full manuscript appended at the end of Thesis).

4.2 Results

This study focused on a set of 9 human CYP450 isoforms responsible for the majority of known drug metabolism: CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and CYP3A4. An *in silico* structural analysis of 348 non-synonymous single nucleotide polymorphisms found across these isoforms was carried out to determine the effects of mutations on enzyme stability and function.

4.2.1 Experimental Data

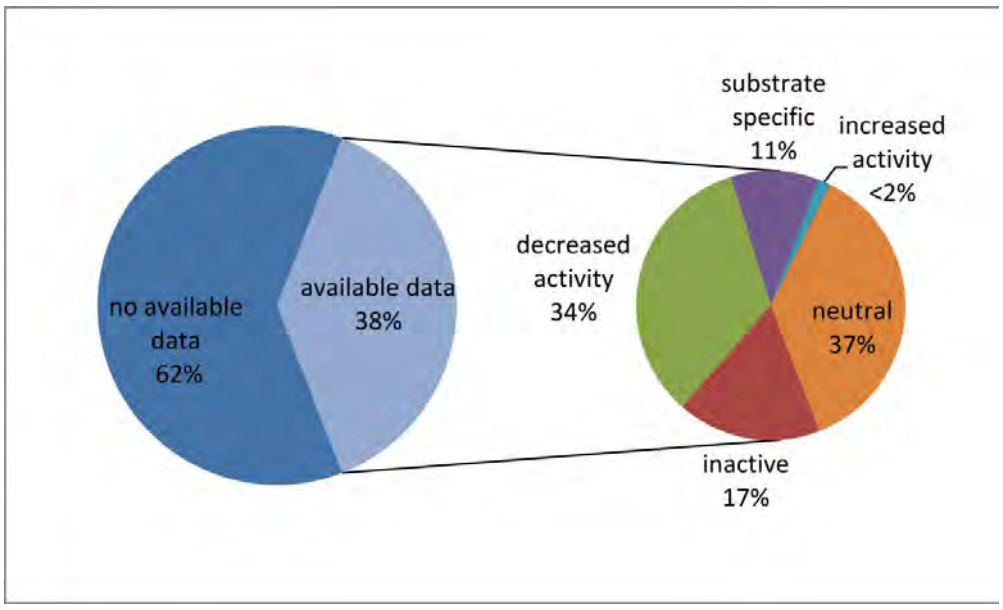


Figure 4.1 SNPs categorised according to available experimental data.

In order to enable the correlation of stability prediction data for individual CYP450 SNPs with known effects on function, experimental data was first extracted and compiled from the literature. Experimental activity data was available for 38% of the SNPs analysed in this study (Figure 4.1); these data are collated in Appendix B, Tables B1-B9. Analysis of the data set showed that 17% of these SNPs were categorised as inactive, 34% of SNPs led to decreased activity and 11% displayed substrate specific effects on activity. It must be noted though that only a small percentage of SNPs have been tested on more than one substrate in the same *in vitro* experiment, so it is possible that some SNPs that currently fall within the neutral or decreased activity category, based on available data, would show substrate specific effects if tested on a

broader range of different substrates. CYP3A4 L293P and CYP2C9 R150H were the only SNPs reported to give increased activity *in vitro*. The remaining 37% of SNPs tested showed no significant effect on enzyme expression or activity.

The number of SNPs per isoform, classified into 6 categories on the basis of available experimental data, is shown in Table 4.1. To date, CYP2B6 and CYP2C19 polymorphic variants have been the most thoroughly characterised, the largest contribution for CYP2B6 variant data coming from a study by Watanabe *et al*, who investigated the effect of 26 variant CYP2B6 alleles, expressed in COS-7 cells, on 7-ethoxy-4-trifluoromethylcoumarin (7-EFC) and selegiline metabolism [448]. There is *in vitro* data for 24 out of the 39 CYP2C19 SNPs and 20 of these were tested by Wang *et al* who reported expression levels, P420:P450 holoprotein levels and kinetic data for 3 different substrates and 10 inhibitors using high through-put activity assays [315, 316]. The functional data available for individual CYP2D6 SNPs is the most incomplete and as a result, only 12% of the SNPs found in CYP2D6 could be classified. CYP2D6 shows the highest levels of polymorphic variation and although there is a large amount of clinical data available, as well as data from several *in vitro* studies, most CYP2D6 alleles contain multiple SNPs in the coding and/or non-coding regions of the gene, which makes it difficult to determine the contribution of individual mutations to the observed phenotype.

Table 4.1 Summary of SNP classification based on available experimental data for each of the 9 major drug metabolising CYP450 isoforms

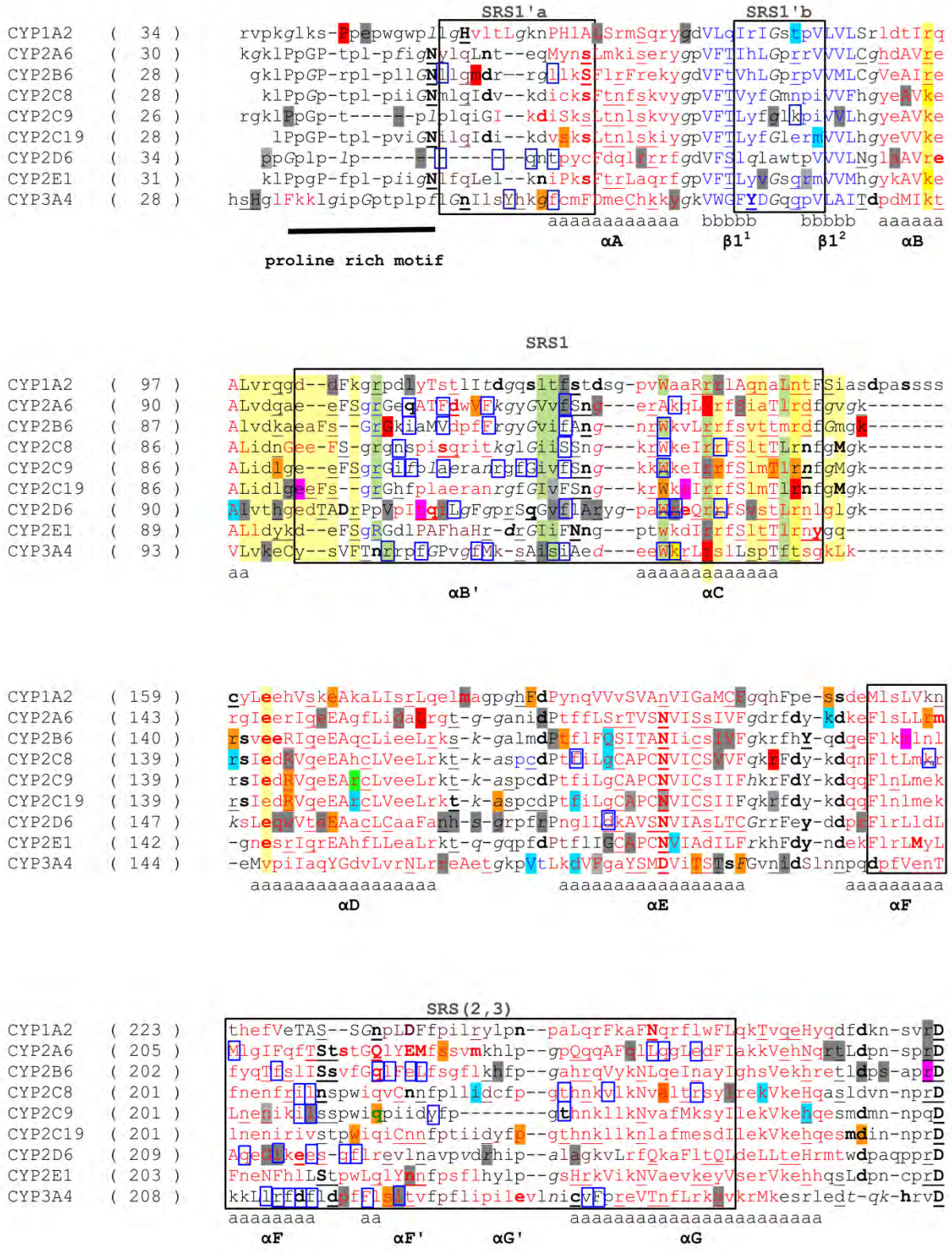
Experimental Data	1A2	2A6	2B6	2C8	2C9	2C19	2D6	2E1	3A4	Total
inactive	4	3	7	1	2	2	2	0	2	23
decreased activity	3	5	3	5	12	9	1	1	6	45
substrate specific	3	3	2	0	0	2	4	0	0	14
increased activity	0	0	0	0	1	0	0	0	1	2
neutral	4	4	9	5	9	9	3	2	4	49
unknown	24	26	10	8	21	17	71	15	23	215
Total SNPs	38	41	31	19	45	39	81	18	36	348

Table shows the number of SNPs classified into each of the 6 categories based on experimental data as described in Methods, Section 4.4.3.

4.2.2 The CYP450 SNP map

In order to visualise the positions of individual CYP450 SNPs relative to key structural and functional regions of the enzymes, a multiple structural alignment, generated based on a representative structure for each of the nine isoforms analysed in this study, is shown in Figure

4.2. This alignment has been annotated with structural features using the program JOY [449].



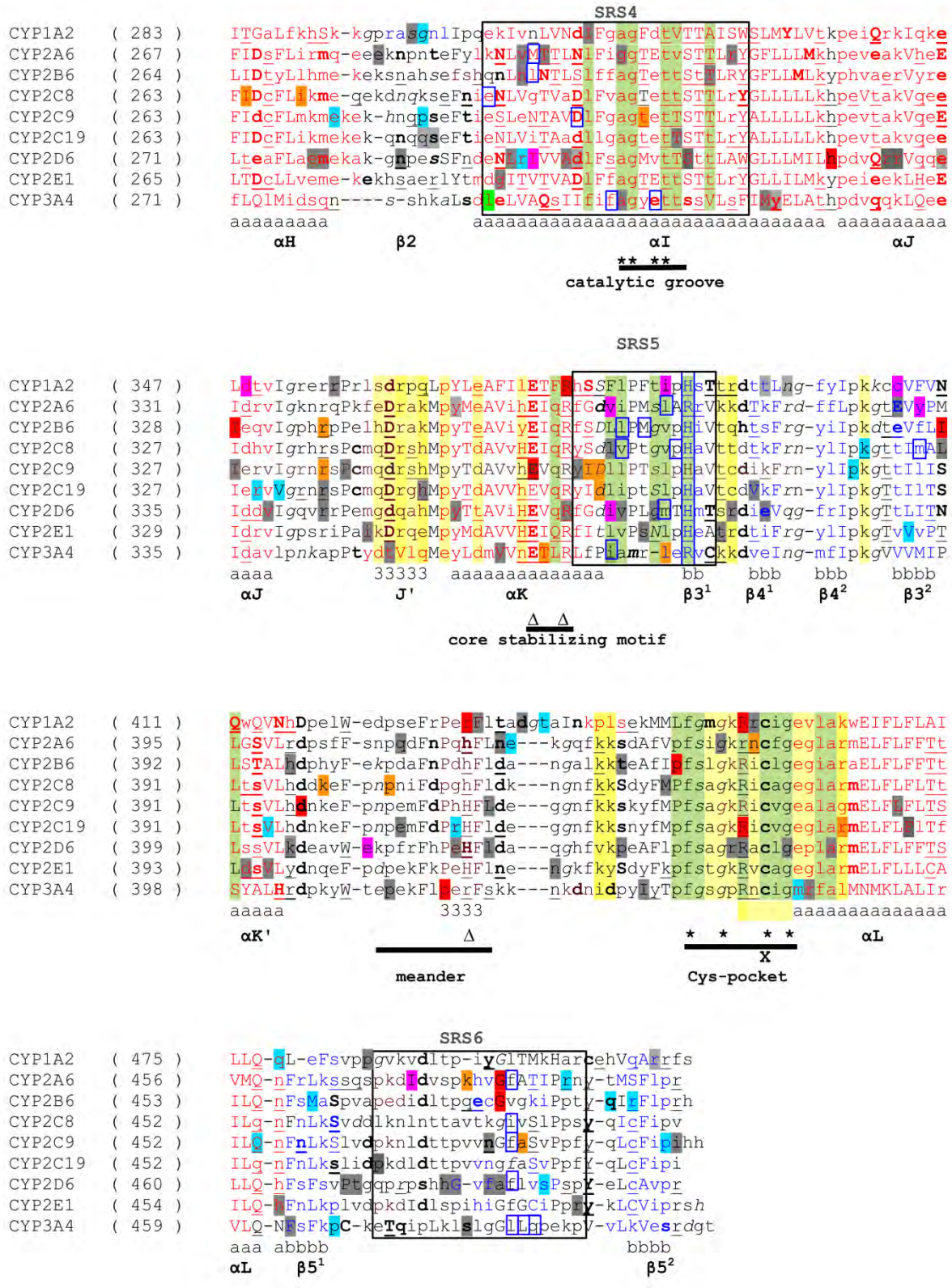


Figure 4.2 The CYP450 SNP map consists of a multiple structural alignment, generated based on a representative structure for each of the 9 drug metabolising CYP450 isoforms.

The alignment is annotated with the following structural features according to the JOY format: Solvent inaccessible/buried residues (upper case), Solvent accessible (lower case), Positive Φ (italics), Hydrogen bond to

main chain amide (bold), Hydrogen bond to main chain carbonyl (underlined), α-helix (red), β-sheet (blue), 3^{10}-helix (dark red).

Secondary structures are named according to the nomenclature of Poulos *et al* [37]. The positions of conserved motifs are also indicated below the alignment.

Functional regions as described in Methods are indicated as follows: SRS regions (boxed in grey), haem contacts (shaded in green), CPR contacts (shaded in yellow- light yellow for residues within 6Å of interface and bright yellow for charge cluster residues), gating residues (boxed in blue) and aqueduct gating residues (boxed in dashed-blue).

The positions of known SNPs are highlighted according to their classifications based on experimental data available in the literature. Inactive (red), decreased activity (orange), substrate specific (magenta), increased activity (bright green), neutral (blue) and unknown (grey). If more than one SNP can occur at the same residue position then the position is highlighted according to the SNP that shows the most drastic effect on enzyme function.

Secondary structure assignment and regions important for the enzymes' structural integrity and function are also indicated, as described in Methods. The positions of SNPs are highlighted according to their assignment based on available experimental data. The alignment does not give information on the mutant residue type, but this information can be found in Appendix B, Tables B1-B9. If more than one SNP occurs at the same position, the wild-type residue is highlighted according to the mutant residue type that has the greatest known effect on enzyme function.

Approximately half of SNPs analysed fell within ordered secondary structure regions (helices A-L or β-sheets 1-5), while the other half occurred in loop/disordered regions. Ordered helices and β-strands make up about 50% of the CYP450 structure, indicating that the SNPs are evenly spread across the protein structure (Table 4.2). A visual representation of the positions of CYP3A4 SNPs relative to secondary structure elements, conserved motifs and functional regions within the 3-D protein structure is shown in Figure 4.3.

Table 4.2 Distribution of SNPs across secondary structure regions

Secondary structure	1A2	2A6	2B6	2C8	2C9	2C19	2D6	2E1	3A4	Total
ordered	16	22	11	9	25	19	41	14	23	180 (52%)
loop/disordered	22	19	20	10	20	20	40	4	13	168 (48%)
Total SNPs	38	41	31	19	45	39	81	18	36	348

Table show the number of SNPs in each isoform falling within ordered and disordered regions of the protein structure.

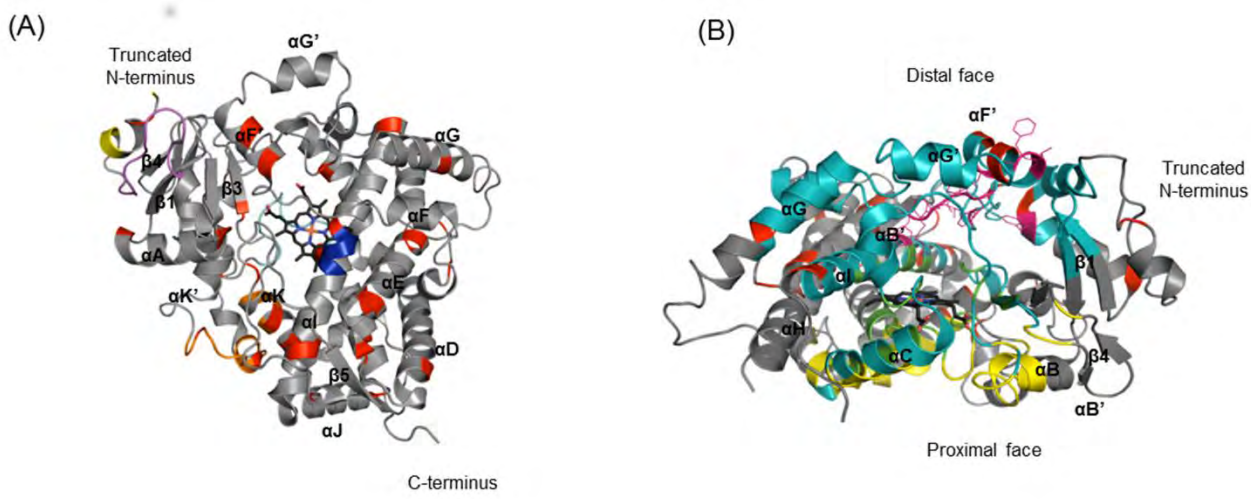


Figure 4.3 CYP3A4 Structural overview
CYP3A4, PDB structure 1TQN, is shown in grey with the haem group in black. Secondary structure elements, α-helices and β-strands are labelled using the nomenclature of Poulos and co-workers [37]. Positions of polymorphic variants found in CYP3A4 are indicated in red. (A) Distal face view showing the position of the halting signal (yellow), the proline-rich motif (magenta), the I-helix catalytic groove (blue), the core stabilising motif, the meander stabilised by the ERR triad (orange) and the cys-pocket (cyan). (B) Side view showing the position of the SRS regions (cyan), the gating residues (magenta), the haem contacts (green) and the CPR interface (yellow) on the proximal face of the protein.

4.2.3 Effect of SNPs on protein stability

Site Directed Mutator (SDM) [356], a statistical potential energy function first developed by Topham *et al* [344], was used as a representative *in silico* tool to predict the effect of each individual CYP450 SNP on protein stability as described in detail in Methods, section 4.4.7 and 4.4.8. SDM uses environment-specific substitution frequencies within homologous protein families to calculate a stability score. The environment specific substitution tables (ESSTs) provide information on the amino acids local structural environment and the probability of it being substituted by another amino acid. The algorithm uses these ESSTs to calculate a stability score, which predicts the difference in free energy between a wild-type and mutant protein, by analogy to a reversible folding-unfolding thermodynamic cycle. Based on this pseudo ΔΔG stability score, SDM classifies mutations as stabilising, destabilising or neutral.

Table 4.3 gives SDM results for three CYP2C9 SNPs, providing an example of a mutation predicted to be neutral, a mutation predicted to be damaging due to stabilising effects, and a mutation predicted to be damaging due to destabilising effects; SDM results for the full data set can be found in Appendix B, Table B1–B9. Tables in Appendix B include pseudo ΔΔG scores - together with predicted effect on stability – and structural information, including the relative

solvent accessibility (RSA) of the wild-type and mutant residue as well as the secondary structure location and the functional region within which the mutation falls and a summary of any available experimental *in vitro* data for each of the 348 CYP450 SNPs in the data set.

Table 4.3 SDM results for three different CYP2C9 SNPs

CYP2C9 SNP	**L208V**	**Q214L**	**L413P**
Secondary Structure	α-helix F	α-helix F' (F-G loop)	between 3_{10} helix & helix L
Relative solvent accessibility of WT residue	51.6 %	6.6 %	2.6 %
Relative solvent accessibility of mutant residue	47.9 %	14.9 %	7.2 %
pseudo ΔΔG scores (kcal.mol^{-1})	- 0.6	3.2	- 4.1
SDM Prediction	Neutral	Damaging - stabilising	Damaging- destabilising

SDM predicted 28% to be damaging mutations that are likely to result in a functionally impaired enzyme (Figure 4.4). The majority of these mutations are predicted as damaging due to destabilising effects (-ΔΔG scores) however 6% of the SNPs are predicted as damaging due to a stabilising effect (+ΔΔG scores); the literature rationalisation on positive ΔΔG scores as damaging is based on the premise that mutations that are highly stabilising may reduce conformational flexibility that is important for function [450-453]. 64% of the SNPs were predicted to be neutral by SDM. SDM could not be used to predict the effect of the remaining 8% of the SNPs on stability as these mutations lie within the N-terminal domain, which is truncated or disordered in crystal structures. These mutations were however considered to be neutral since they are unlikely to affect the stability of the catalytic domain and since *in vitro* studies have shown that the N-terminal membrane anchor region is not important for enzyme activity [454]. A summary of the SDM predictions for SNPs in each isoform is shown in Table 4.4.

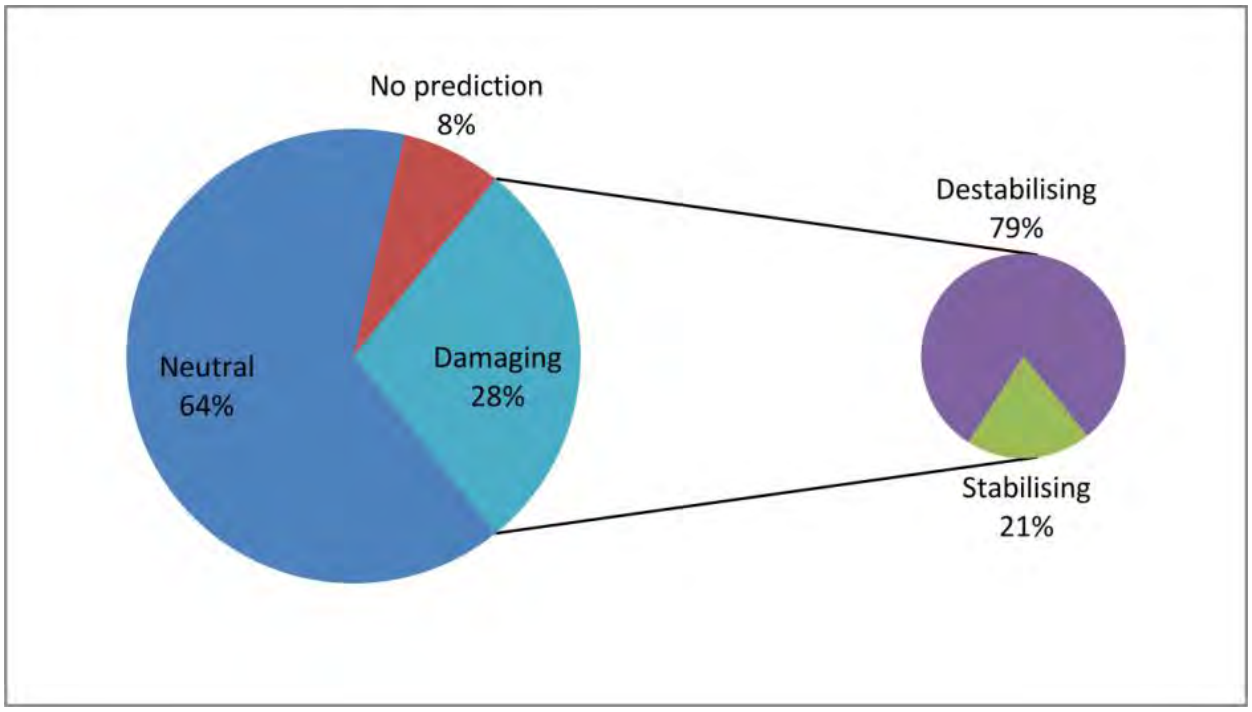


Figure 4.4 Proportion of SNPs predicted to be damaging to protein structure

Table 4.4 Summary of SDM predictions

Effect of SNPs on stability	1A2	2A6	2B6	2C8	2C9	2C19	2D6	2E1	3A4	Total
Neutral	27	28	19	12	26	28	50	15	19	**224**
damaging – destabilising	7	4	5	6	14	7	20	2	12	**77**
damaging – stabilising	1	6	2	0	1	1	5	1	4	**21**
No prediction	3	3	5	1	4	3	6	0	1	**26**
Total	**38**	**41**	**31**	**19**	**45**	**39**	**81**	**18**	**36**	**348**

Table shows the number of SNPs in each isoform predicted to be neutral ($-2 < \Delta\Delta G < 2$ kcal.mol^{-1}), damaging due to destabilising effects ($\Delta\Delta G < -2$ kcal.mol^{-1}) and damaging due to stabilising effects ($\Delta\Delta G < 2$ kcal.mol^{-1}) as well as the number of SNPs in each isoform for which a prediction could not be obtained.

4.2.4 Correlation between stability predictions and experimental data

Table 4.5 lists the SNPs that could be classified based on experimental data, together with their secondary structure and functional region location, their predicted effect on protein stability, and their classification based on experimental data.

Table 4.5 Structural information and protein stability predictions for SNPs that could be classified according to their effect on enzyme activity based on experimental data collected from the literature.

SNP	STRUCTURAL INFORMATION		PROTEIN STABILITY		Classification based on experimental data
	2° structure location	Functional region location	SDM Score kcal.mol^{-1}	Prediction	
CYP1A2					
P42R	loop btwn N-term & helix A	proline rich motif	-0.3	neutral	inactive
T83M	B-sheet1^{2}	SRS1'b	0.1	neutral	neutral
E168Q	α-helix D	undefined	0.0	neutral	decreased activity
F186L	loop btwn Helix D & E	undefined	-1.0	neutral	decreased activity
S212C	loop btwn helix E & F	undefined	-1.9	neutral	decreased activity
G299A	B-sheet 2	undefined	-5.2	damaging	neutral
D348N	α-helix J	undefined	0.8	neutral	substrate specific
R377Q	α-helix K	ERR triad/core stabilising motif	-1.1	neutral	inactive
I386F	loop btwn helix K & B-sheet 3^{1}	haem contact/SRS5	0.1	neutral	substrate specific
C406Y	B-sheet 3^{2}	undefined	0.0	neutral	substrate specific
R431W	3_{10} helix	ERR triad/meander	2.3	damaging	inactive
T438I	loop btwn 3_{10} helix & helix L	undefined	-0.8	neutral	neutral
R456H	loop btwn 3_{10} helix & helix L	Cys pocket/haem contact/CPR contact (charge cluster residue)	-1.5	neutral	inactive
Q478H	end of helix L	undefined	0.3	neutral	neutral
CYP2A6					
V110L	α-helix B'	SRS1	0.3	neutral	decreased activity
R128Q	α-helix C	haem contact/CPR contact	-0.9	neutral	inactive
L160H	α-helix D	undefined	-1.9	neutral	inactive
K194E	loop btwn helix E & F	undefined	-0.4	neutral	neutral
R203S	α-helix F	SRS(2,3)	-1.2	neutral	neutral
R203C	α-helix F	SRS(2,3)	-7.2	damaging	decreased activity
S224P	α-helix G'	SRS(2,3)	1.6	neutral	decreased activity
V365M	loop btwn helix K & B-sheet 3^{1}	SRS5	1.5	neutral	substrate specific
Y392F	B-sheet 3^{2}	undefined	1.3	neutral	substrate specific
E419D	loop btwn B-sheet 3^{2} & helix K'	undefined	0.2	neutral	neutral
N438Y	loop btwn 3_{10} helix & helix L	Cys-pocket/CPR contact	2.1	damaging	decreased activity
I471T	btwn B-sheet 5^{1} & 5^{2}	SRS6	-2.3	damaging	substrate specific
K476R	btwn B-sheet 5^{1} & 5^{2}	SRS6	1.3	neutral	decreased activity
G479V	btwn B-sheet 5^{1} & 5^{2}	SRS6	2.1	damaging	inactive
R485L	btwn B-sheet 5^{1} & 5^{2}	SRS6	1.3	neutral	neutral
CYP2B6					
R22C	Truncated N-terminal	halting signal	n/a	n/a	neutral
T26S	Truncated N-terminal	undefined	n/a	n/a	neutral
D28G	Truncated N-terminal	undefined	n/a	n/a	neutral
R29T	Truncated N-terminal	halting signal	n/a	n/a	neutral

M46V	loop btwn N-term & helix A	SRS1'a	-1.8	neutral	inactive
G99E	B-C loop	SRS1	-1.0	neutral	inactive
K139E	loop btwn helix C & D	undefined	0.6	neutral	inactive
R140Q	loop btwn helix C & D	undefined	0.4	neutral	decreased activity
T168I	start of helix E	undefined	3.4	damaging	decreased activity
Q172H	α-helix E	undefined	2.4	damaging	neutral
M198T	α-helix E	SRS(2, 3)	-2.7	damaging	substrate specific
S259R	loop btwn helix G & H	undefined	0.0	neutral	neutral
K262R	loop btwn helix G & H	undefined	-0.4	neutral	substrate specific
I328T	α-helix J	undefined	-3.9	damaging	inactive
R336C	loop btwn helix J & J'	undefined	-3.0	damaging	decreased activity
I391N	loop btwn B-sheet 3^2& helix K'	undefined	-4.6	damaging	inactive
P428T	loop btwn 3_{10} helix & helix L	haem contact	-1.2	neutral	inactive
M459V	B-sheet 5^1	undefined	-0.5	neutral	neutral
G476D	loop bwn B-sheet 5^1 & 5^2	SRS6	0.5	neutral	inactive
Q485L	loop bwn B-sheet 5^1 & 5^2	undefined	-1.7	neutral	neutral
R487C	Start of B-sheet 5^2	undefined	-3.6	damaging	neutral
CYP2C8					
R139K	loop btwn helix C & D	undefined	-1.4	neutral	neutral
G171S	α-helix E	undefined	1.5	neutral	neutral
R186G	loop btwn helix E & F	undefined	-4.6	damaging	inactive
I223M	F-G loop	SRS(2,3)	0.2	neutral	neutral
A238P	α-helix G	SRS(2,3)	-2.2	damaging	decreased activity
K247R	α-helix G	undefined	0.5	neutral	neutral
I264M	α-helix H	undefined	-0.4	neutral	decreased activity
I269F	α-helix H	undefined	0.2	neutral	decreased activity
K383N	loop btwn B-sheet 4^2 & 3^2	CPR contact	-1.1	neutral	neutral
K399R	loop btwn helix K' & 3_{10} helix	undefined	-0.4	neutral	decreased activity
P404A	loop btwn helix K' & 3_{10} helix	meander	1.2	neutral	decreased activity
CYP2C9					
L19I	truncated N-terminal	signal-anchor sequence	n/a	n/a	neutral
L90P	loop btwn helix B & C	CPR contact	-2.5	damaging	decreased activity
R125H	α-helix C	CPR contact/SRS1	-0.6	neutral	decreased activity
T130R	α-helix C	SRS1	-2.1	damaging	decreased activity
R132Q	loop btwn helix C & D	CPR contact (charge cluster residue)/SRS2	-0.2	neutral	decreased activity
R144C	α-helix D	undefined	-11.5	damaging	decreased activity
R150H	α-helix D	undefined	-0.6	neutral	increased activity
R150L	α-helix D	undefined	-0.2	neutral	neutral
Q214L	α-helix F' (F-G loop)	SRS(2,3)	3.2	damaging	decreased activity
H251R	α-helix G	undefined	-0.1	neutral	neutral
E272G	α-helix H	undefined	-4.0	damaging	neutral
P279T	loop btwn helix H & I	undefined	1.1	neutral	neutral
T299A	centre of α-helix I	catalytic groove/SRS4	1.9	neutral	decreased activity
R335Q	loop btwn helix J & J'	undefined	-0.3	neutral	neutral

R335W	loop btwn helix J & J'	undefined	-2.2	damaging	decreased activity
E354K	α-helix K	ERR triad / core stabilizing motif	-4.0	damaging	inactive
I359L	end α-helix K	SRS5	0.2	neutral	decreased activity
I359T	end α-helix K	SRS-5	-3.9	damaging	decreased activity
D360E	loop btwn helix K & B-sheet 3^1	SRS-5	-1.3	neutral	decreased activity
P382S	loop btwn B-sheet 4^2 & 3^2	SRS5	-0.9	neutral	neutral
D397A	loop btwn helix K' & 3_{10} helix	undefined	1.2	neutral	inactive
Q454H	α-helix L	undefined	1.4	neutral	neutral
A477T	loop btwn B-sheet 5^1 & 5^2	SRS-6	0.6	neutral	decreased activity
P489S	end of B-sheet 5^2	undefined	1.7	neutral	neutral
CYP2C19					
L17P	truncated N-terminal	signal-anchor sequence	n/a	n/a	neutral
I19L	truncated N-terminal	signal-anchor sequence	n/a	n/a	neutral
S51G	α-helix A	SRS1'a	-2.0	neutral	decreased activity
M74T	B-sheet 1^2	SRS1'b	-2.7	damaging	neutral
E92D	loop btwn helix B & B'	CPR contact/SRS1	0.0	neutral	substrate specific
W120R	α-helix C	haem contact/SRS1	-2.7	damaging	decreased activity
E122A	α-helix C	SRS1	1.4	neutral	substrate specific
R132Q	loop btwn helix C & D	CPR contact (charge cluster residue)/SRS1	0.4	neutral	inactive
R144H	α-helix D	undefined	0.2	neutral	decreased activity
R150H	α-helix D	undefined	-0.6	neutral	neutral
A161P	loop btwn helix D & E	undefined	-10.5	damaging	decreased activity
F168L	α-helix E	undefined	1.0	neutral	neutral
W212C	α-helix F'	SRS(2,3)	-2.1	damaging	decreased activity
P227L	F-G loop	SRS(2,3)	0.6	neutral	decreased activity
D256N	loop btwn helix G & H	undefined	-0.3	neutral	decreased activity
R329H	α-helix J	undefined	-0.6	neutral	neutral
$I331V^1$ (WT 1.B)	loop btwn helix J & J'	undefined	-0.4	neutral	neutral
D360N	loop btwn helix K & B-sheet 3^1	SRS5	1.5	neutral	decreased activity
V394M	α-helix K'	undefined	-0.1	neutral	neutral
R410C	3_{10} helix	meander	-0.3	neutral	neutral
R433W	loop btwn 3_{10} helix & helix L	Cys-pocket/haem contact/CPR contact (charge cluster residue)	0.4	neutral	inactive
R442C	α-helix L	CPR contact (charge cluster residue)	-10.0	damaging	decreased activity
CYP2D6					
A90V	α-helix B	undefined	-0.2	neutral	neutral
T107I	α-helix B'	SRS1	4.2	damaging	substrate specific
E156A	α-helix D	undefined	0.42	neutral	decreased activity
R296C	α-helix I	SRS1	-8.8	damaging	neutral
I297L	α-helix I	SRS1	0.2	neutral	substrate specific
H324P	loop btwn helix H & I	undefined	-4.6	damaging	inactive
I369T	loop btwn helix K & B-sheet 3^1	Haem contact/SRS5	-3.2	damaging	substrate specific

E410K	loop btwn α-helix K' & 3_{10} helix	undefined	0.4	neutral	substrate specific
R441C	loop btwn 3_{10} helix & helix L	Cys-pocket/haem contact/CPR contact (charge cluster residue)	-10.4	damaging	inactive
S486T	B-sheet 5^2	SRS6	0.7	neutral	neutral
CYP2E1					
R76H	B-sheet 1^2	SRS1'b	0.8	neutral	decreased activity
V179I	α-helix E	undefined	-0.1	neutral	neutral
V389I	B-sheet 3^2	undefined	-0.3	neutral	neutral
CYP3A4					
G56D	btwn N-term & α-helix A	SRS1'a	-2.3	damaging	decreased activity
R130Q	α-helix D	haem contact/CPR contact/SRS1	-1.4	neutral	inactive
V170I	loop btwn helix D & E	undefined	-0.3	neutral	neutral
D174H	α-helix E	undefined	0.2	neutral	neutral
T185S	α-helix E	undefined	-1.4	neutral	decreased activity
F189S	loop btwn helix E & F	undefined	-5.2	damaging	decreased activity
S222P	α-helix F'	SRS(2,3)	-0.9	neutral	decreased activity
L293P	Start of Helix I	start of SRS4	-0.5	neutral	increased activity
T363M	α-helix K	core-stabilising motif	2.1	damaging	decreased activity
L373F	start of B-sheet 3^1	haem contact/SRS5	0.0	neutral	decreased activity
P416L	start of 3_{10} helix	meander	3.9	damaging	inactive
M445T	start of α-helix L	CPR contact	-2.6	damaging	neutral
P467S	loop btwn B-sheet 5^1 & 5^2	undefined	-1.4	neutral	neutral

The correlation between SDM results and experimental data is shown in Figure 4.5, from which it can be seen that of the 38 CYP450 SNPs for which there was experimental data and which were predicted to be damaging by SDM, 31 showed altered activity in the experimental data, giving a positive predictor value (PPV) (*i.e.* the percentage of test positives that are true positives) of 82% for SDM. Conversely, of the 84 SNPs in the data set that showed altered activity experimentally, SDM predicted that 31 would be damaging, meaning that, used on its own, SDM has a sensitivity (*i.e.* the percentage of correctly classified true positives) of 37% in analysis of CYP450 SNPs. Finally, of the 42 SNPs classified as neutral based on experimental data and which could be analysed by SDM, 35 were predicted to have no significant effect on protein stability, meaning that used on its own, SDM has a specificity (*i.e.* the percentage of correctly classified true negatives) of 83% in analysis of CYP450 SNPs.

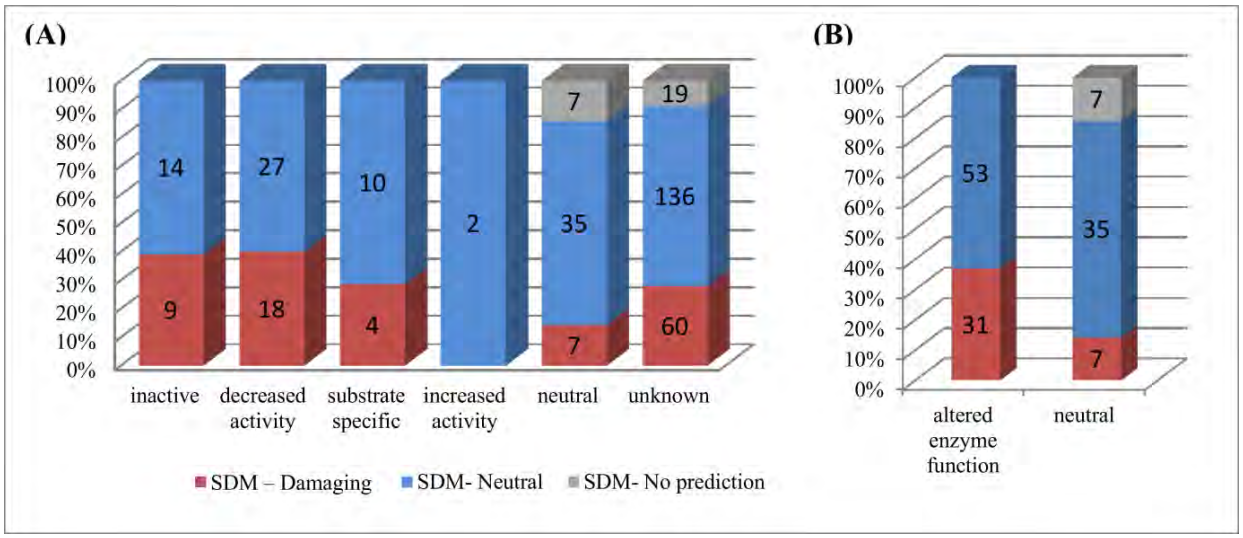


Figure 4.5 Correlation between SDM stability predictions and experimental data.

The y-axis shows the percentages of SNPs predicted to be damaging to protein structure (red columns), those predicted to be neutral (blue columns) and those that fell within truncated regions of the structure that could not be analysed by SDM (grey columns) for (A) each experimentally classified group and (B) for a combination of all groups showing altered enzyme function vs. those classified as neutral, as shown along the x-axis. The numbers of SNPs accounting for these percentages are shown within the columns.

4.2.5 Effect of SNPs on function

According to the analysis performed here, 54% of the CYP450 sequence space forms part of one or more functional regions (Figure 4.2 & Figure 4.6): The seven conserved motifs cover ~13% of the CYP450 sequence space; just over 5% of CYP450 residues form haem contacts; ~8% form the CPR interface; & SRS's are the largest functional regions, covering one third of the sequence space (Chapter 2). Gating residues have additionally been identified in six of the nine isoforms (Chapter 3) and currently make up an average of 2% of the CYP450 sequence.

Figure 4.6 also shows the percentage of SNPs falling within each functional region. Based on the percentage sequence space covered by each functional region, SNPs were found to be evenly spread throughout all functional regions, including the conserved motifs, with 52% of SNPs falling within one or more defined regions and the other 48% falling within undefined functional regions. A breakdown of the number of SNPs falling within each functional region for each isoform is given in Table 4.6.

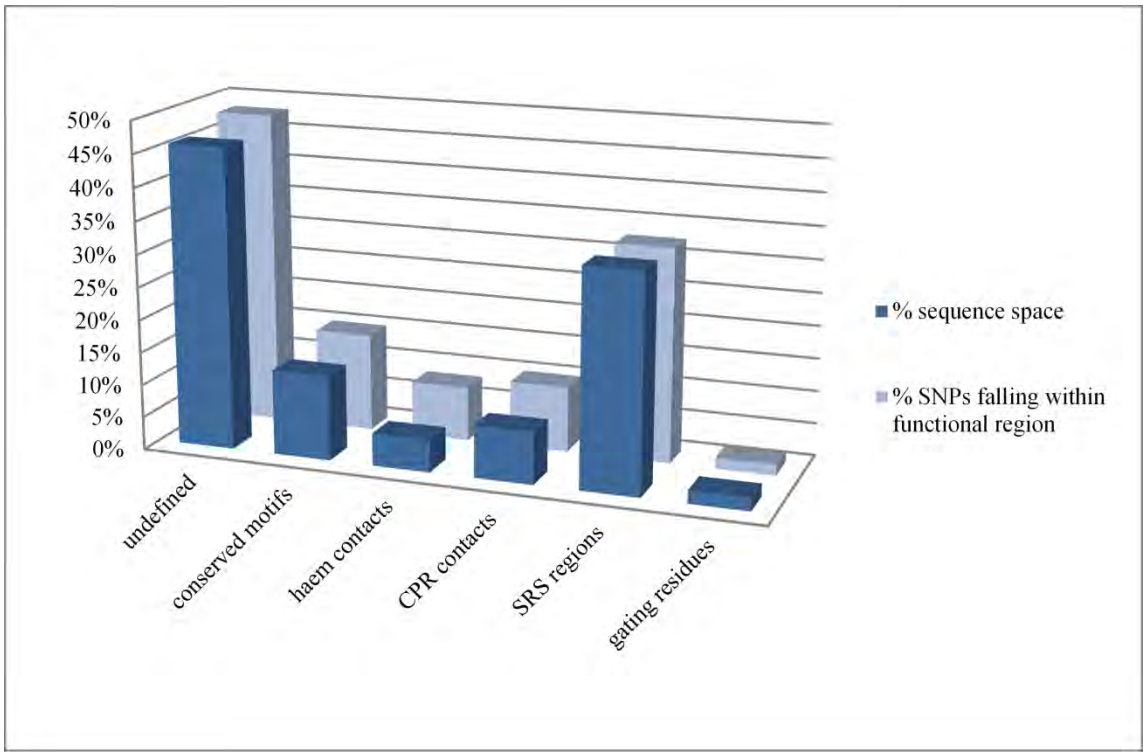


Figure 4.6 Percentage of SNPs falling within each functional region and the percentage sequence space covered by each functional region, illustrating that SNPs are spread evenly across functional regions.

Table 4.6 Summary of SNPs locations within functional regions for each isoform

Functional Region	1A2	2A6	2B6	2C8	2C9	2C19	2D6	2E1	3A4	Total
undefined	**22**	**17**	**20**	**12**	**21**	**19**	**31**	**8**	**18**	**168**
SNPs in functional region	**16**	**24**	**11**	**7**	**24**	**20**	**50**	**10**	**18**	**180**
conserved motifs	9	6	4	2	6	6	15	0	5	**53**
haem contacts	3	3	2	1	1	3	9	3	4	**29**
CPR contacts	3	4	0	1	5	8	9	3	5	**38**
SRS regions	7	17	7	4	18	13	30	8	11	**114**
gating residues	**0**	1	0	0	1	0	3	0	1	**6**

The numbers in this table represent the number of SNPs falling within each functional region for each isoform. Note that because functional regions overlap many SNPs fall within more than one functional region.

4.2.6 Correlation between functional region analysis and experimental data

The correlation between the functional region location of SNPs and experimental data is shown in Figure 4.7, from which it can be seen that of the 64 CYP450 SNPs for which there is experimental data and which fall in functional regions important for catalytic activity (and which are therefore potentially damaging), 53 showed altered activity in the experimental data, giving a

positive predictor value (PPV) of 83% for this functional region analysis. Conversely, of the 84 SNPs in the data set that showed altered activity experimentally, 53 fall in functional regions important for catalytic activity (and are therefore potentially damaging), giving a sensitivity of 63%. Finally, of the 49 SNPs classified as neutral based on experimental data, 38 were found to fall within undefined regions or regions that are not important for catalytic activity (and are therefore less likely to be damaging), giving a specificity of 78%.

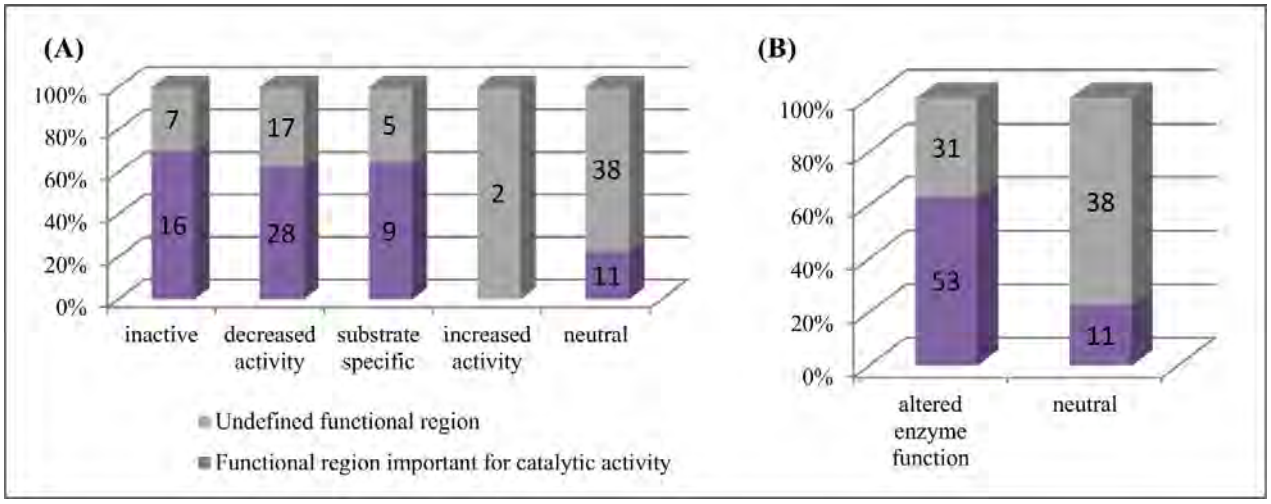


Figure 4.7 Correlation between functional analysis and experimental data
The y-axis shows the percentages of SNPs falling within functional regions important for catalytic activity (purple) vs. those falling within undefined functional regions or regions that do not affect catalytic activity (grey) for **(A)** each experimentally classified group and for **(B)** a combination of all groups showing altered enzyme function vs. those classified as neutral, as shown along the x-axis. The numbers of SNPs accounting for these percentages are shown within the columns. Note that functional regions important for catalytic activity include all defined regions excluding the N-terminal signal- anchor motif and halting signal.

Table 4.7 shows the number of SNPs occurring in each functional region, broken down according to their experimentally determined effect on function. Where a SNP occurred in more than one region, it was assigned to the functional region that corresponds best with the observed experimental data. For example, if a mutation occurred at the position of a haem contact and it fell within a SRS region, but the experimental data shows that no holoprotein could be detected for this variant, then this SNP was assigned to the haem contact region instead of the SRS region. If it was unclear which functional region best fitted the data then the SNP was assigned to both regions, thus explaining the differences in the absolute numbers in Figure 4.7 and Table 4.7. The numbers in brackets indicate the number of these SNPs predicted to be damaging by SDM.

Five SNPs within the N-terminal signal anchor sequence and halting signal have been functionally characterised *in vitro* and all of them were classified as neutral according to experimental data. N-terminal regions play a role in membrane binding but mutations in these regions are not expected to have an effect on protein folding or catalytic activity. The majority of the other neutral SNPs (67%) occurred within undefined regions, while 16% occurred within SRS regions.

Of the 13 mutations falling within the other conserved motifs - namely the proline rich motif, the I-helix catalytic groove, the Cys-pocket, the core stabilising motif and the meander region – 12 resulted in an inactive protein or a protein with decreased activity. Three of these inactive SNPs fell within the ERR triad.

Approximately 82% of SNPs falling within SRS regions had some effect on protein function, with the majority causing decreased activity. Experimental data showed that 31 SNPs falling within undefined functional regions had an effect on enzyme activity, of which 10 were predicted to be damaging by SDM.

Table 4.7 Correlation between experimental data and functional region localisation of SNPs

Functional Region	Inactive	Decreased activity	Substrate specific	Increased activity	Neutral
signal anchor sequence	0 (0)	0 (0)	0 (0)	0 (0)	3 (n/a)
halting signal	0 (0)	0 (0)	0 (0)	0 (0)	2 (n/a)
proline rich motif	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)
I-helix catalytic groove	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)
cys-pocket	3 (1)	1 (1)	0 (0)	0 (0)	0 (0)
core stabilizing motif	2 (1)	1 (1)	0 (0)	0 (0)	0 (0)
meander	2 (2)	1 (0)	0 (0)	0 (0)	1 (0)
ERR-Triad	3 (2)	0 (0)	0 (0)	0 (0)	0 (0)
haem contact	6 (1)	2 (1)	0 (0)	0 (0)	0 (0)
CPR contact	1 (0)	5 (3)	0 (0)	0 (0)	2 (1)
SRS	5 (1)	23 (8)	9 (4)	0 (0)	8 (2)
gating residue	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
undefined	7 (4)	17 (6)	5 (0)	2 (0)	33 (4)

4.2.7 Combining stability predictions and functional region analysis

The correlation between the experimental data and the combined SDM stability and functional analysis predictions is shown in Figure 4.8, from which it can be seen that of the 78 CYP450 SNPs for which there is experimental data and which fall in functional regions important for catalytic activity and/or are predicted to be damaging by SDM, 63 showed altered activity in the experimental data, giving a positive predictor value of 81% for this combined analysis. Conversely, of the 84 SNPs in the data set that showed altered activity experimentally, 63 fall in functional regions important for catalytic activity and/or are predicted to be damaging by SDM, giving a sensitivity of 75% for this combined analysis. Finally, of the 49 SNPs classified as neutral based on experimental data, 34 were found to fall within undefined regions or regions that are not important for catalytic activity and are predicted to have no effect on stability, giving a specificity of 69% for this combined analysis.

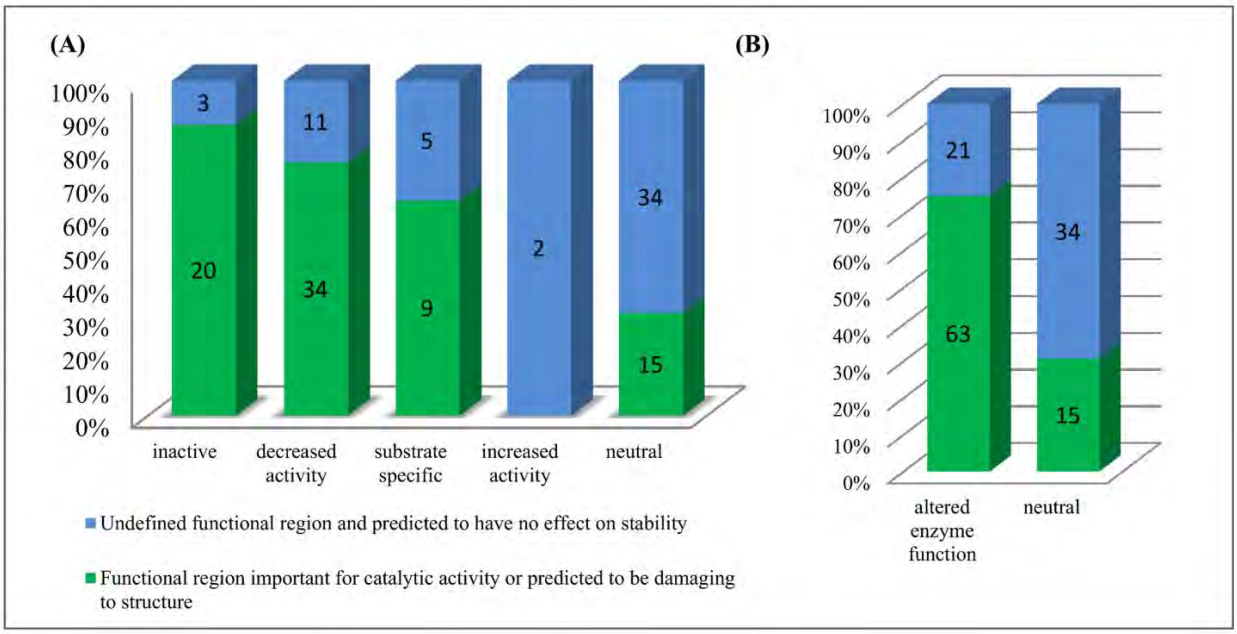


Figure 4.8 Correlation between experimental data and the combined SDM stability and functional analysis predictions.

The y-axis shows the percentage of SNPs found in functional regions important for catalytic activity or predicted by SDM to be damaging to the structure vs. those found in undefined functional region and predicted to have no effect on stability **(A)** for each experimentally classified group and **(B)** for combination of all groups showing altered enzyme function vs. those classified as neutral. The numbers of SNPs accounting for these percentages are shown within the columns. Note that functional regions important for catalytic activity include all defined regions excluding the N-terminal signal- anchor motif and halting signal.

4.3 Discussion

Previous functional studies have delineated regions of the cytochrome P450 structure important for substrate recognition (Chapter 2), substrate and product access and egress from the active site (Chapter 3) and interaction with the cytochrome P450 reductase [98]. In this Chapter, this functional information has been combined with *in silico* calculations on the effect of polymorphic variations in the major human drug metabolising CYP450 enzymes on protein stability. These results were compared to experimental data in order to establish the likely causes of altered drug metabolism observed for cytochrome P450 variants in functional assays to date. In the process, a cytochrome P450 polymorphic variant map was created (Figure 4.2), which gives an overview of the location of known CYP450 polymorphic variation in relation to important structural and functional regions, whilst detailed SNP tables (Appendix B, Tables B1-B9) provide information on the structural environments of each variant residue, the predicted effect of the mutation on protein stability and a summary of the available *in vitro* data for each SNP.

A number of *in silico* tools are available that aim to distinguish between disease associated mutations and benign mutations based on the residues position relative to functional sites, protein stability calculations or a combination of both. Using the tools PolyPhen [359] and SIFT [353, 455], Wang *et al* previously predicted that 39-43% of the human CYP450 SNPs they analysed would have an impact on protein function [361] and correlation with known experimental data suggested a PPV of 70%. Other CYP450 studies have used a combination of docking and molecular dynamics to assess changes to the binding pocket and ligand interactions for small sets of SNPs [456-458]. However, whilst tools such as PolyPhen take the positions of some functional sites into account, quantitatively predicting the effect of mutations on ligand and co-factor binding is still a huge challenge for computational biology as accurate calculations are very computationally intensive, limiting their use for large interaction surfaces.

The aim the work described in this Chapter was to break down the analysis of possible functional effects of SNPs into different components in order to distinguish between mutations affecting stability and those lying within functional regions of the enzyme, as well as to more precisely predict the molecular mechanisms underlying changes in enzyme activity caused by different

CYP450 SNPs. Site Directed Mutator (SDM) was therefore chosen as a representative tool to predict the effect of mutations on the stability of the apoprotein, enabling the differentiation between mutations affecting the inherent stability of the protein fold as opposed to those at functional sites important for haem or substrate binding. SDM - which is a statistical potential energy function - provides a quick, high throughput method for calculating the effect of single point mutations on protein stability and performs favourably compared to a number of other published methods [357]. SDM is not based on prior knowledge of the mutant's thermodynamic measurements and has been found to demonstrate better sensitivity in predicting stabilizing mutations than other methods [356]. Importantly, for the purpose of this study, SDM focuses solely on predicting the energetic effects of mutations on protein stability ($\Delta\Delta G$), ignoring effects on functional regions such a ligand contacts. Interestingly, when SDM was used in isolation here, a PPV of 82% was observed, apparently immediately exceeding the performance of PolyPhen or SIFT for CYP450 enzymes.

While it may have been interesting in principle to compare results from other tools that specifically predict effects of SNPs on protein stability, many of these tools have previously been compared extensively in other studies and have been found to be comparable [356, 357]; furthermore, the shortage of quantitative experimental data on the effect of SNPs on CYP450 protein stability would make it difficult to be sure which tool was the most accurate for a given mutation in any such comparative study. It therefore seemed unlikely that a comparative study of *in silico* tools would add much value to the analysis. Related to this point, although initially specific $\Delta\Delta G$ values were calculated for each SNP using SDM, this data was subsequently only used in a qualitative manner, serving to ameliorate any deficiency in the absolute accuracy of individual SDM-computed $\Delta\Delta G$ values across the large panel of variants.

Reliable *in silico* prediction of the effect of SNP variants on CYP450 protein function remains a challenging endeavour today. In this Chapter, using a representative, computationally-light *in silico* tool in isolation to predict the effects of individual SNPs on the stability of the CYP450 apoproteins, an encouraging specificity of 83% and a positive predictor value of 82% was obtained when predictions were compared to available experimental activity data, but a sensitivity of only 37% was found for CYP450 SNPs (Figure 4.5, Table 4.8), suggesting that altered CYP450 activity is only partially explained by changes in protein stability. This is in agreement with studies by Blundell *et al* who have shown that although changes in protein

stability can be relatively accurately predicted *in silico*, these predictions alone are in general not sufficient to enable accurate large-scale classification of benign and disruptive single amino acid substitutions [356, 357, 459]. However, by combining the SDM predictions on effect of each SNP on protein stability with an analysis of whether or not the relevant SNP fell within a known functional region important for catalysis, the sensitivity of the predictions was significantly improved to 75%, and very good positive predictor and specificity values were still retained (Table 4.8). In more detail, the combined 'SDM plus functional region analysis' approach enabled us to account for 87% of the SNPs that result in very low or inactive enzyme activity; 76% of the SNPs that result in decreased activity and 64% of the SNPs that result in substrate specific affects *in vitro* (Figure 4.8).

Table 4.8 Summary performance of SDM and functional region analysis on CYP450 SNPs

	Sensitivity (%)	Specificity (%)	PPV (%)
SDM alone	37	83	82
Functional region analysis alone	63	78	83
SDM plus functional region analysis	75	69	81

By way of illustration of the approach used in this study, CYP2C19*5, corresponding to the single amino acid substitution R433W, has been identified as a defective allele that contributes to the s-mephenytoin poor metaboliser phenotype in Caucasian and Oriental populations [460]. *In vitro* functional assays have shown that expression levels in yeast are similar to the wild-type but that P450 holoprotein is not detected [315]. Figure 4.9 shows a model of the CYP2C19 polymorphic variant R433W aligned to the wild-type structure. Arg 443 is a buried residue with a relative solvent accessibility of 5.3%. When this residue is mutated to a tryptophan, the side chain volume increases by 54.4 $Å^3$ while the relative solvent accessibility changes from 2% to 17%. The ΔΔG SDM score for this mutation is 0.4 $kcal.mol^{-1}$, which indicates that this mutation is neutral and does not have a significant effect on the stability of the apoprotein. However, Arg 433 is part of the Cys-pocket, or β-bulge, which is a highly conserved region important for haem binding. Based on the protein-ligand interaction data for CYP2C19, Arg 433 is a haem contact residue that interacts through its 5-guanidino group with the haem propionate via a salt bridge. Mutating this residue to a tryptophan disrupts this interaction and the increase in volume and the position of the side chain may plausibly result in a steric clash between tryptophan and the haem group, providing a possible explanation for the lack of holoprotein observed *in vitro*. Importantly, if the functional region and stability analyses had not been segregated, this level of mechanistic interpretation would not have been possible.

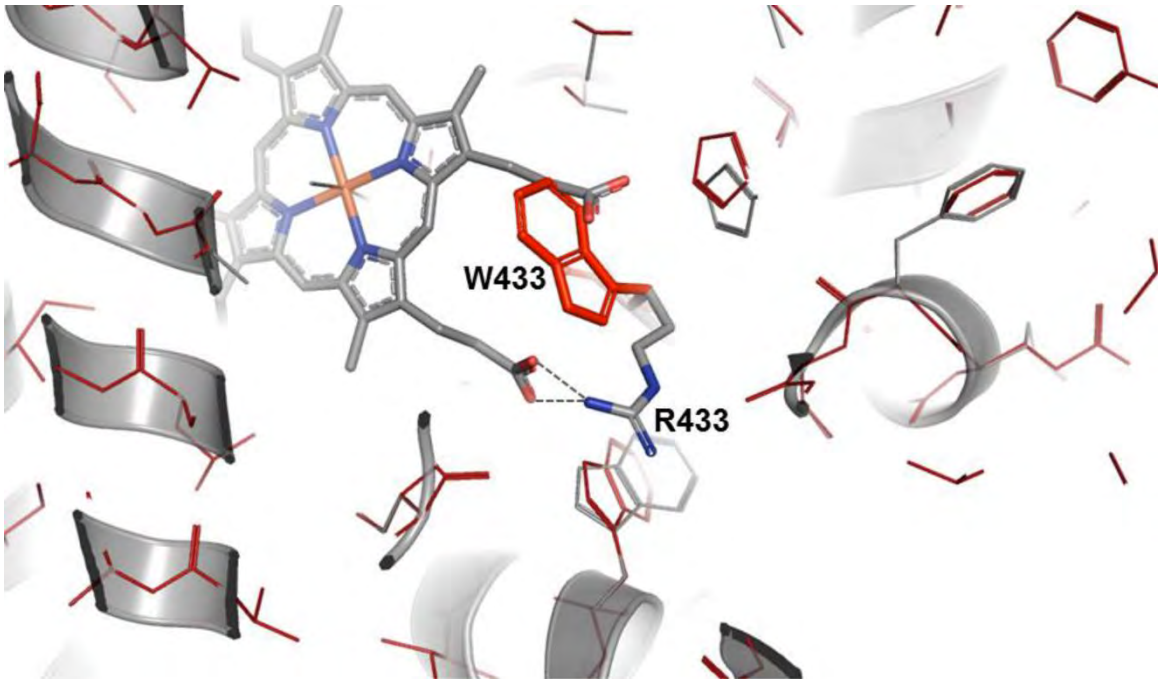


Figure 4.9 CYP2C19 R433W model aligned to the wild-type structure
The CYP2C19 R433W model (red) generated by ANDANTE is aligned to the wild-type CYP2C19 pdb structure 4GQS (grey). Wild-type residue Arg 433 is shown interacting with the haem group.

Although the alignment in Figure 4.2 has been annotated with the known CYP450 functional regions, CYP450 function is by no means completely understood. Information about residues important for conformational change, oligomerisation and interactions with effector molecules is still limited. Investigating further mutations that show altered *in vitro* activity but do not affect protein stability and lie outside known functional regions may shed light on hinge regions and peripheral or allosteric binding sites important for conformational change and substrate access in CYP450s. In this context, it is interesting that the analyses carried out here reveal that 5 out of the 14 SNPs that show altered substrate specific *in vitro* activity, together with one SNP that showed increased *in vitro* activity, lie within currently undefined functional regions of CYP450 structure (Table 4.6). For example, CYP2C19 variant D256N shows similar expression and holoprotein levels to the wild-type enzyme but shows substrate specific changes in K_m and reduced V_{max} [461]. Asp 256 lies in the loop region between helices G and H (Figure 4.10) and this G/H loop has shown high conformation variability in molecular dynamics studies on mammalian CYP450 structures [76, 80] so Asp 256 may therefore form part of a hinge region important for conformational changes within the enzyme. While the putative periphery site regions and SRS 4'a region identified in Chapter 2 were not included as functional regions in

this study, it is interesting to note that of the 5 SNPs that show altered substrate specific *in vitro* activity but which do not affect protein stability and which lie outside known functional regions, 2 fall within PS2 (CYP1A2 C406Y and CYP2A6 Y392F align with CYP2B6 E387 and F389 that made contact with amlodipine molecules bound at periphery sites within the active site – Chapter 2, Figure 2.6) and 1 falls within the putative SRS 4'a region (CYP2B6 K262R). Furthermore, one of the SNPs that showed deceased *in vitro* activity and does not affect protein stability falls within putative SRS4'a.

A significant limitation in the analysis carried out here on damaging SNPs is the quality, availability and variability of experimental data found in the literature. In particular, heterogonous expression system, total protein quantification, methods for generating kinetic data, as well as kinetic models for data analysis, vary between studies and can all profoundly influence results, as recently reviewed by Hiratsuka [462]. For example, most CYP450 substrate turnover is still analysed today using Michaelis-Menten kinetics, yet it is well documented that some CYP450 isoforms such as CYP3A4 show atypical kinetics due to multiple binding sites, substrate inhibition, allosteric regulation or multiple active forms of the enzyme. Depending on the model to which the experimental data was fitted, substantially different kinetic parameters might be reported for any given system, complicating quantitative comparisons between different studies on mammalian CYP450 SNPs.

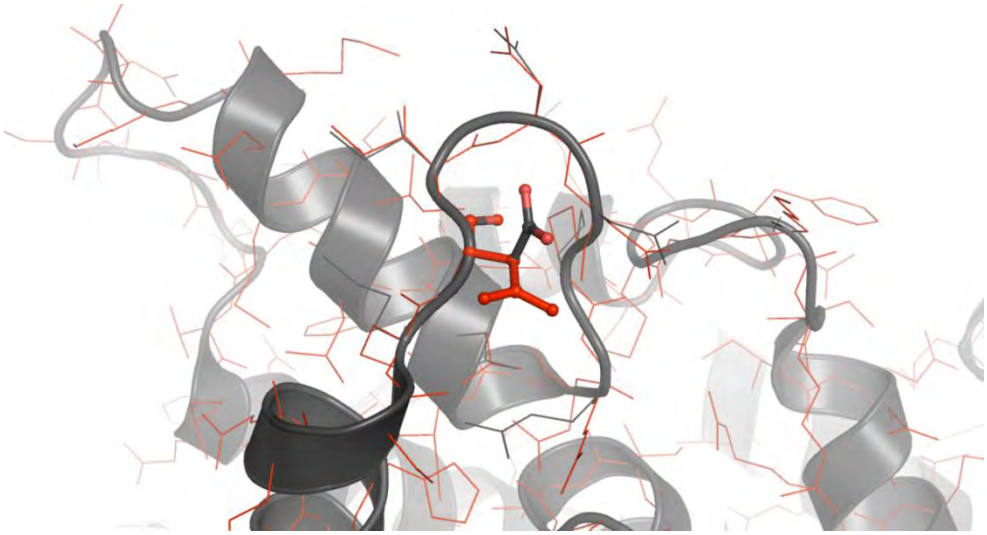


Figure 4.10 CYP2C19 D256N model aligned to the wild-type structure
The CYP2C19 D256N model (red) generated by ANDANTE is aligned to the wild-type CYP2C19 pdb structure 4GQS (grey). Asp256 and Asn256 are shown as sticks.

The sum of the defined functional regions drawn together in the CYP450 SNP map (Figure 4.2) make up over half of the protein, whilst the substrate recognition regions cover ~33% of the

sequence space, reflecting the broad substrate specificities of mammalian CYP450s. It is important to recognise therefore that not all mutations in these functional regions will affect enzyme activity for all possible substrates. It is also important to note that certain residues within SRS regions have side chains facing away from the active site (see for example Chapter 2, Figure 2.11); some of these residues make contact with the substrate *via* main chain interactions rather than side chain interactions and mutation of such residues might directly affect ligand interactions (and hence enzyme function) as a result either of altered positioning of main chain atoms relative to the bound ligand within the active site, or alternatively by directly affecting protein stability. In other instances, such residues with side chains facing away from the active site do not obviously make main chain contacts to the ligand; here mutational effects on enzyme function are more likely to be manifested through either direct effects on protein stability or through co-operative conformational changes that indirectly perturb the active site. By way of example of this latter type, CYP2A6 variant R203S, lying within the F helix and SRS(2,3), shows similar activity to the wild-type enzyme in *in vitro* studies; the side chain of Arg 203 faces away from the active site (Figure 4.11) and SDM predicts that mutation of Arginine 203 to Serine will not be damaging, consistent with experimental data. By contrast, variant R203C does show reduced activity *in vitro;* this mutation is predicted to be highly destabilising by SDM and is likely to affect substrate binding by disrupting the F helix.

It could be argued that residues in SRS regions that make neither side chain nor main chain contacts to the ligand in known CYP450 structures should be excluded from the annotated functional regions in the CYP450 SNP Map. However, as further CYP450 structures are determined, the precise positioning of equivalent residues relative to bound ligands may change, arguing for their retention as part of the SRS regions at this point; indeed there are already examples of SRS residues that face towards the active site in one structure and away from the active site in other structures as a result of conformational changes, particularly in regions that can convert between ordered helices and disordered loop regions such as the F-G loop and B-C loop (e.g. in CYP2C9 structures 1OG2 and 1OG5 the F-G loop and B-C loop are predominantly helices but in the 1R9O structure these regions are disordered loops). In addition, mutations can also change the conformation of a protein and as a result the orientation of SRS residue side chains, as seen for the V374M mutation in the CYP2D6 structure 2F9Q (Chapter 3, Figure 3.10 B).

Inclusion of SDM data in the CYP450 SNP map has proved useful in extending predictions of the likely consequence of mutation on CYP450 function, including for those variants that fall within known functional regions. Importantly, most SNPs analysed in this study have only been tested to date for one or a small subset of substrates, so it is plausible that more SNPs lying within SRS regions will show altered metabolism or substrate specific effects when tested on a larger panel of drugs. The work described in this Chapter therefore provides a useful rationalisation of CYP450 structure-function relationships by combining both *in silico* and experimental data into a unified CYP450 SNP Map that has surprisingly good predictive capability (judged by the sensitivity and specificity data presented here) and which should therefore find utility in predicting the likely effects of uncharacterised CYP450 variants on drug metabolism.

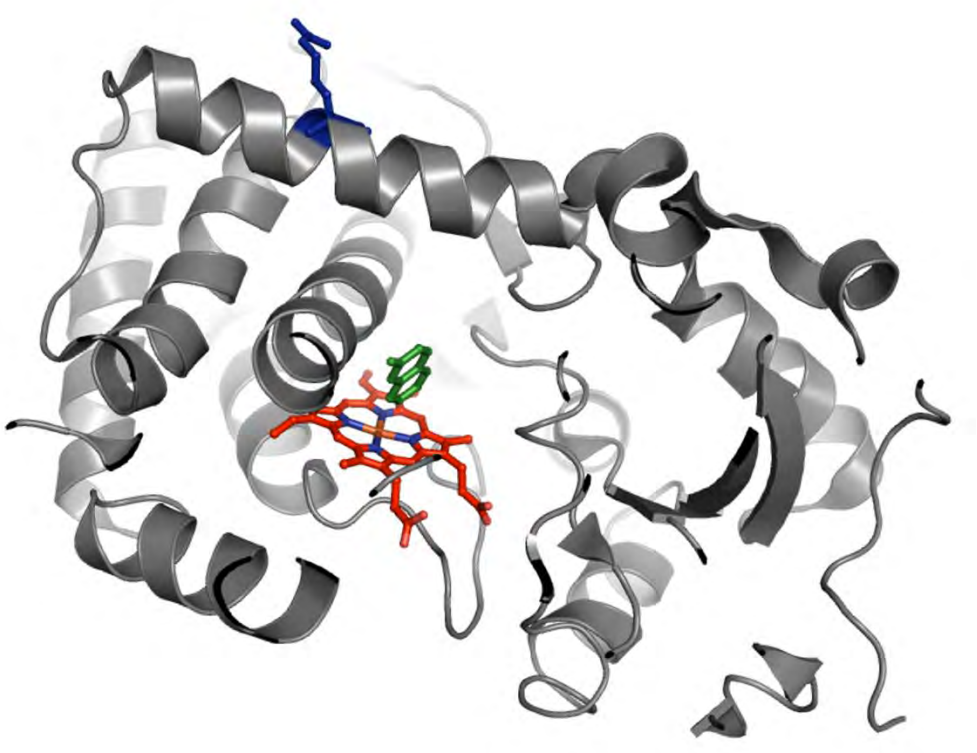


Figure 4.11 CYP2A6 model showing the position of R203.

CYP2A6, PDB structure 1Z10 chain A , is shown in grey, the haem group in red and a coumarin substrate molecule bound within the active site shown in green. R203 is shown in blue.

4.4 Methods

4.4.1 Non-synonymous single nucleotide polymorphisms

Non-synonymous single nucleotide polymorphisms (SNPs) were sourced from the Human Cytochrome P450 Allele Nomenclature Committee home page ([309]; accessed 10 Oct 2012) and NCBI dbSNP [463] for isoforms CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and CYP3A4.

4.4.2 Experimental Data Collection

Available experimental data on protein expression, holoprotein levels, enzyme activity and kinetics (K_m, V_{max} and V_{max}/K_m) was extracted from the literature and tabulated for each SNP (Appendix B, Table B1-B9). Due to the wide range of data types and the great diversity in methodology all data was expressed as a fraction (variant data/ wild-type data) measured in the same experiment. If the difference between wild-type and variant data was reported to be statistically significant, it is indicated in Tables B1-B9 with an asterisks (some studies provide quantitative data without statistical evidence). If no quantitative data was available then data was expressed qualitatively as similar to wild-type "similar", significantly lower than wild-type "decreased", significantly higher than wild-type "increased" or not detectable "ND". All *in vivo* data was recorded qualitatively. If more than one non-synonymous SNP coexist in one allele and have therefore been tested as a double or triple mutant (as is common for CYP2D6), data was included only if it is clear which SNP is dominant in the resulting phenotype.

4.4.3 Classifying SNPs based on Experimental data

After careful analysis of all available data, SNPs were classified into six groups based primarily on the *in vitro* experimental evidence. *In vivo* data was however, considered if it was the only data available. "Inactive" refers to mutations that resulted in protein/holoprotein levels that are not detectable or activity that is less than 2% of wild-type activity for all expression systems and substrates tested. "Decreased activity" refers to SNPs that resulted in significantly decreased expression levels, haem incorporation or altered kinetic parameters, leading to impaired enzyme function. "Increased activity" refers to mutants that show significant increases in activity.

"Substrate specific" refers to mutants that show more than 50% difference in mutant/wild-type activity ratios for different substrates. "Neutral" refers to mutants that show similar holoprotein levels and activity to that of the wild-type enzyme. "Unknown" refers to SNPs that have not yet been tested *in vitro* or for which the experimental data is inadequate or contradictory.

4.4.4 Crystal Structures

Structures of human CYP450 proteins and rat CYP2B4 were retrieved from the RCSB Protein Data Bank [31]. The PDB codes for the structures used in this study are as follows. **CYP1A2**: 2HI4. **CYP2A6**: 1Z10, 1Z11, 2FDU, 2FDV, 2FDW, 2FDY, 2PG5, 2PG6, 2PG7, 3EBS. **CYP2B6**: 3IBD. **CYP2C8**: 1PQ2, 2NNH, 2NNI, 2NNJ, 2VN0. **CYP2C9:** 1OG2, 1OG5, 1R9O. **CYP2C19:** 4GQS. **CYP2D6:** 2F9Q. **CYP2E1:** 3E6I, 3E4E, 3GPH, 3KOH, 3LC4. **CYP3A4**: 1TQN, 1W0E, 1W0F, 1W0G, 2J0D, 2V0M, 3NXU. **CYP2B4** (rat):2BDM. Only chain A was used for structures that have more than one CYP450 chain in the asymmetric unit.

4.4.5 Generating an annotated structural alignment

A multiple CYP450 structural alignment was generated including one representative structure for each isoform using the program Baton (D.Burke, unpublished data and [423]). The PDB codes for representative structures were as follows CYP1A2: 2HI4, CYP2A6: 2FDV, CYP2B6: 3IBD, CYP2C8: 1PQ2, CYP2C9**:** 1R9O, CYP2C19: 4GQS, CYP2E1: 3E4E, CYP2D6: 2F9Q, CYP3A4: 1TQN.

The program JOY [449] was used to annotate the protein sequence alignment with 3D structural features. Here is a brief description of the structural features represented by JOY.

Solvent accessibility is measured using the program PSA which implements the algorithm developed by Lee and Richards (1971). A relative solvent accessibility (RSA) is calculated by comparing an extended conformation of a residue to the residue conformation in question. The default cut-off setting of 7% RSA was used. This classifies residues with less than 7% RSA as inaccessible and residues of >7% as accessible. In the Joy format, buried residues are displayed in upper case and accessible residues in lower case.

Secondary structure assignment is calculated using the program SSTRUC (D.K.Smith, unpublished). In the JOY format, residues forming α, 3_{10} and π helices, and β strands are displayed in different colours. The default consensus secondary structure definition was used which means a consensus secondary structure is assigned, underneath the alignment, if more than 70% of the residues at that position are in a particular conformational state. α-helices A – L and β-sheets 1-5 are labelled below the alignment according to CYP450 nomenclature of Poulos and co-workers [37].

Main chain dihedral angles are also calculated by SSTRUC. Residues with a positive main chain φ angle are shown in italics.

Hydrogen bonding within a structure is calculated using the program HBOND (J.Overington, unpublished), which calculates the associated energy and angles for interactions between heavy atom donor and acceptor pairs. JOY uses a lenient distant cut-off of 3.5Å for hydrogen bonds. The hydrogen bond data is split into three classes: 1) side chain to main chain amide; 2) side chain to main chain carbonyl; 3) side chain to other side chain. Residues whose side chain is bonded to a main chain amide are denoted in bold face. Residues that are underlined indicate that their side chain is hydrogen bonded to a main chain carbonyl. Residues whose side chain is hydrogen bonded to another side chain are not displayed in the JOY format used here.

4.4.6 Defining CYP450 functional regions

The structural alignment was further annotated with known functional regions. If residues did not fall within any of these regions then their functional region was classified as undefined. Functional regions can overlap so one residue may fall within more than one functional region. The definitions of functional regions are as follows.

Conserved motifs

The hydrophobic *signal-anchor sequence* and the *halting-signal* are usually truncated in crystal structures and therefore do not appear in the structural alignment. The *proline rich motif* (N-terminal sequence motif PPGPXPXPXXGN), the *I-helix catalytic groove* (conserved consensus motif (G/A)GX(D/E)T), the *K-helix core stabilising motif* (invariant EXXR motif within the K-

helix) and *meander,* and the *Cys-pocket* (sequence motif FXXGXXXCXG containing the absolutely conserved catalytic Cys residue) have been labelled on the alignment.

Haem contacts

Haem-residue interactions were identified using the protein-ligand interaction database CREDO [397]. CREDO assigns each contact residues to an interaction type based on the geometry and distance thresholds using criteria adapted from Marcou and Rognan [464]. Residues interacting with the haem group were identified in all the available PDB structures for each isoform, combined and mapped onto the representative structure in the sequence alignment. This was done to account for any differences in the haem binding regions for structures of the same isoform due to conformational changes.

SRS regions

SRS regions were assigned based on the SRS map delimited in Chapter 2.

Cytochrome P450 reductase (CPR) contacts

CPR contacts were assigned based on the recent model for interactions between CYP2B4 and FMN domain of rat CYP450 reductase [98]. This model was built using the bacterial redox complex structure (CYP102: 1BVY, chain A) as a template and was used to identify the key residues involved in ionic interactions at the interface. All CYP2B4–P450-oxidoreductase interfacial residues calculated from the 1BVY.pdb-derived structural model at 6Å distance cut-off were mapped onto the human CYP450 representative structures *via* structural alignment. Residues forming part of ionic charge clusters where further categorised as *ionic charge cluster residues* to distinguish them from the other CPR contact residues at the interface.

Gating residues

Gating residues were assigned as follows: 1) Residues involved in the putative gating models identified in Chapter 3 were assigned as *gating residues* (**CYP2A6**: Q104, F107, F111, F118, K125, M205, L241, Q242, E245, M293, L370, F480; **CYP2B6**: L43, L51, I101, V104, F108, F115, W121, F206, Q215, L216, E218, L219, L290, L363, M365, I480; **CYP2C8**: N99, S114,

W120, R125, F168, K199, I207, L208, T229, V233, R241, E285, V362, P367, M388, I476; **CYP2C9**: K72, I99, F100, A103, R108, F110, G111, F114, W120, I207, L208, Y225, T229, D293, F476; **CYP2D6**: L46, F51, Q52, T54, L110, F120, W128, R129, R133, D179, Q210, L213, E216, G218, F219, V374, F483, **CYP3A4**: Y53, F57, R106, F108, F113, M114, W126, K127, L211, R212, F213, D214, F215, F220, I223, V240, F241, F304, E308, I369, L482, L483, Q484.). 2) Gating residues identified in the literature using steered molecular dynamics simulations [434] (CYP3A4: F57, R105, F108, S119, I120, R212, F213, F215, F220, V240, F241, L482.). 3) Residues involved in an aqueduct gating mechanism regulated by binding of the CYP450 reductase identified using solvent explicit molecular dynamics [442]. (**CYP3A4**: R375) As R375 is highly conserved in CYP450 is has been suggested that this could be a general aqueduct gating mechanism. All residues at this position in the alignment have been defined as *aqueduct gating residues*. (**CYP1A2:** H388**, CYP2A6**: R372, **CYP2B6**: H369, **CYP2C8:** H368, **CYP2C9:** H368**, CYP2C19:** H368**, CYP2D6:** H376**, CYP2E1:** H370**, CYP3A4:** R375)

4.4.7 Comparative modeling of polymorphisms using ANDANTE

Polymorphisms were modelled using ANDANTE, a program that makes use of environmental substitution tables (ESSTs) to reduce the side chain rotamer search space during comparative modeling of single amino acid mutations [465]. A Python script was written to run ANDANTE in batch mode. Mutant structures were generated for each SNP using all the available PDB crystal structures for the respective isoform. ANDANTE was run using the default parameters and the "- local" option which attempts to minimise all rotamer clashes in the local area that result from the mutation. Note that ANDANTE does not take co-factors or ligands into account when modeling side chain positions. All mutant structures for each SNP were visualised and compared to the original template crystal structures using PyMOL [406].

4.4.8 Predicting the effect of polymorphisms on protein stability using SDM

SDM input: SDM requires a wild-type and mutant structure (which has only one single amino acid substitution) in PDB format as input. The mutant structures built using ANDANTE, together with their template structures, were used as input for the mutant and wild-type structures respectively.

SDM output: SDM output gives details of the local structural environment of the wild-type and mutant residue which includes secondary structure prediction and percentage solvent accessibility. SDM also outputs a pseudo ΔΔG score or stability score and a prediction of disease association. Negative scores suggest that the mutation is destabilising and positive scores indicate a stabilising effect. A cut-off of 2 $kcal.mol^{-1}$ had previously been used to indicate that a mutation is damaging to the protein structure. Both stabilising and destabilising mutations that reach this cut-off were therefore categorised as damaging mutations whilst mutation with scores between -2 and 2 $kcal.mol^{-1}$ are classified as neutral. It must be noted that like ANDANTE, SDM does not consider co-factors or ligands when calculating solvent accessibility or stability scores.

SDM was run on all ANDANTE derived mutant structures. If ANDANTE generated more than one structure for the same SNP, an average SDM score and standard deviation were calculated. In subsequent analysis, the output from SDM was used in a qualitative manner, mutations were described as damaging or neutral and no further specific use of the absolute numerical SDM score was made.

Chapter 5 Cloning and Expression of CYP450 proteins

5.1 Introduction

Heterogeneously expressed recombinant CYP450 proteins are useful for studying CYP450 function in simplified *in vitro* systems. The ease of site directed mutagenesis in these systems enables the study of non-synonymous mutation - this is particularly useful for low frequency alleles where clinical samples are limited. The first heterologous mammalian CYP450 was expressed in 1991 in *E.coli* [466]; recombinant CYP450s are now routinely expressed in bacteria, yeast, insect and mammalian cells.

E.coli expression of recombinant proteins is in general fast, easy and inexpensive, producing relatively high yields of protein compared with other expression systems. Expression of mammalian CYP450s in *E.coli* is, however, challenging due to hydrophobic membrane binding regions, differences in codon usage between *E.coli* and mammalian cells and problems with protein folding in the absence of endogenous chaperones [454]. Post-translational modification of recombinant proteins seldom occurs in *E.coli*, but this does not generally pose a problem for recombinant CYP450 expression as haem incorporation is the only post-translational modification thought to be important for catalytic activity and occurs in *E.coli*, provided the protein folds correctly [467]. Co-expression of molecular chaperones can improve protein folding [468].

N-terminal sequence alterations within the hydrophobic membrane-binding region are normally required for adequate recombinant expression in *E.coli* due to the lack of 5' features necessary for high protein expression in bacteria and to differences in codon usage between bacteria and mammals. These alterations are not thought to affect enzyme activity because the N-terminal region does not play a role in catalysis [469, 470] and, where tested, heterologously expressed CYP450 proteins show comparable activity to native proteins purified from liver samples [471, 472]. The N-terminal membrane-binding region, however, does influence sub-cellular localisation and CYP450s can be designed to localise in the inner membrane, cytoplasm or periplasm; to aggregate into inclusion bodies; or to be secreted into the extra cellular medium. CYP450s are however most commonly localised in the cytoplasm with the N-terminal anchor embedded in the plasma membrane [454]. Truncating the N-terminal anchor region can increase expression and solubility and simplify purification although some truncated isoforms still

associate with the membrane *via* peripheral membrane interaction; these interactions can be disrupted with high salt buffers or detergents post cell lysis.

Strategies to combine heterologously expressed CPR and CYP450 redox partners to reconstitute the full active monooxygenase system include the following: 1) combining enzymes after expression and purification; 2) co-expressing enzymes in the same host cell using two different vectors; 3) co-expressing proteins using a bicistronic vector or a dual promoter vector; 4) expressing CPR and CYP450 as a single fused polypeptide.

The aim of this Chapter was to generate a panel of CYP3A4 and CYP2C9 variants for *in vitro* characterisation and kinetic assays. A heterologous expression system enabling easy site-directed mutagenesis, time and cost effective protein expression and simple protein purification was required. Previous i*n vitro* studies on CYP3A4 wild-type protein in the Blackburn laboratory provided the ground work for this study and a suitable CYP450 and CPR expression system was already established [473]. The DH5α *E.coli* strain, commonly used for CYP450 expression, was used for CYP3A4 and CPR expression and offered the advantage of high insert stability and good plasma yield and quality. Various full length and truncated CYP3A4 and CPR constructs, with tags on either the N-terminus or the C-terminus, were tested for expression and activity. N-terminal truncated CYP3A4 and CPR protein constructs with C-terminal tags were found to give the highest levels of active protein.

The BCCP-tag, an 83 amino acid biotin carboxyl carrier protein domain from the *E.coli* acetyl-coenzyme A carboxylase, is a compact structure dominated by β-strands (Figure 5.1) that is recognised and biotinylated at a specific lysine residue by the host cell's native biotin ligase [474]. The biotin molecule can then act as a point of attachment for the protein, binding specifically to streptavidin. The BCCP tag was thus included in the CYP450 and CPR constructs to allow orientated immobilisation of proteins to derivatised microarray surfaces *via* the high affinity streptavidin-biotin interaction. The BCCP tag has also previously been reported to improve folding and soluble expression and can additionally serve as a useful protein-folding marker: presuming the fusion protein folds co-translationally, if the N-terminal domain mis-folds, this will cause aberrant exposure of hydrophobic regions which will either disrupt the correct folding of the BCCP-tag or drive aggregation, thus by either route preventing biotinylation [475].

The polyhistidine-tag (His-tag), a protein motif consisting of 6 histidine residues, was fused to the C-terminus of the BCCP-tag *via* a glycine-serine linker (G-S linker). It was included to allow easy purification as it binds reversibly to Ni^{2+}, allowing one-step non-denaturing purification *via* Ni^{2+}- affinity chromatography.

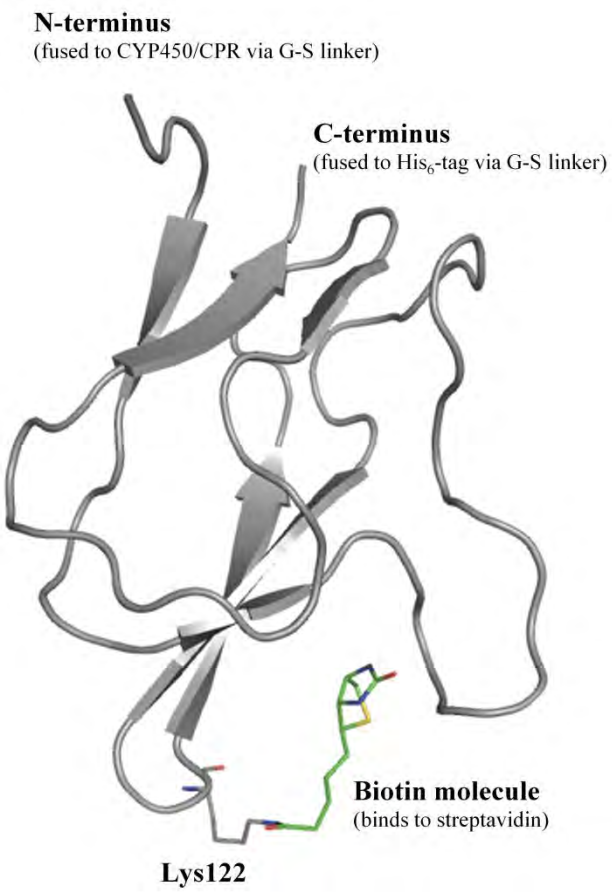


Figure 5.1 Structure of the biotinylated biotin carboxyl carrier protein domain from the *E.coli* acetyl-coenzyme A carboxylase (BCCP tag). The protein, PDB structure 1BDO, is shown in grey and the biotin molecule attached to the side chain of Lysine 122 is shown in green.

Using this CYP450 expression system, the objectives of this Chapter were as follows:

1. Select a panel of CYP3A4 and CYP2C9 polymorphic variants for study
2. Generate expression vectors for each variant *via* site-directed mutagenesis
3. Express and purify proteins
4. Quantify CYP450 holoprotein levels

5.2 Results

5.2.1 Generating CYP3A4 and CYP2C9 variant constructs

Selecting polymorphic variants for testing

To test the effect of clinically relevant mutations on CYP450 stability and structure *in vitro,* eight CYP3A4 and seven CYP2C9 polymorphic variants were selected for expression in *E.coli* (Figure 5.2). The selected variants included mutations predicted from the work described in Chapter 4 to be neutral, stabilising or destabilising, falling within both functional and undefined regions of the protein structure (Table 5.1). There are reports in the literature showing altered drug metabolism for some of these variants, whilst others have not previously been tested *in vitro.*

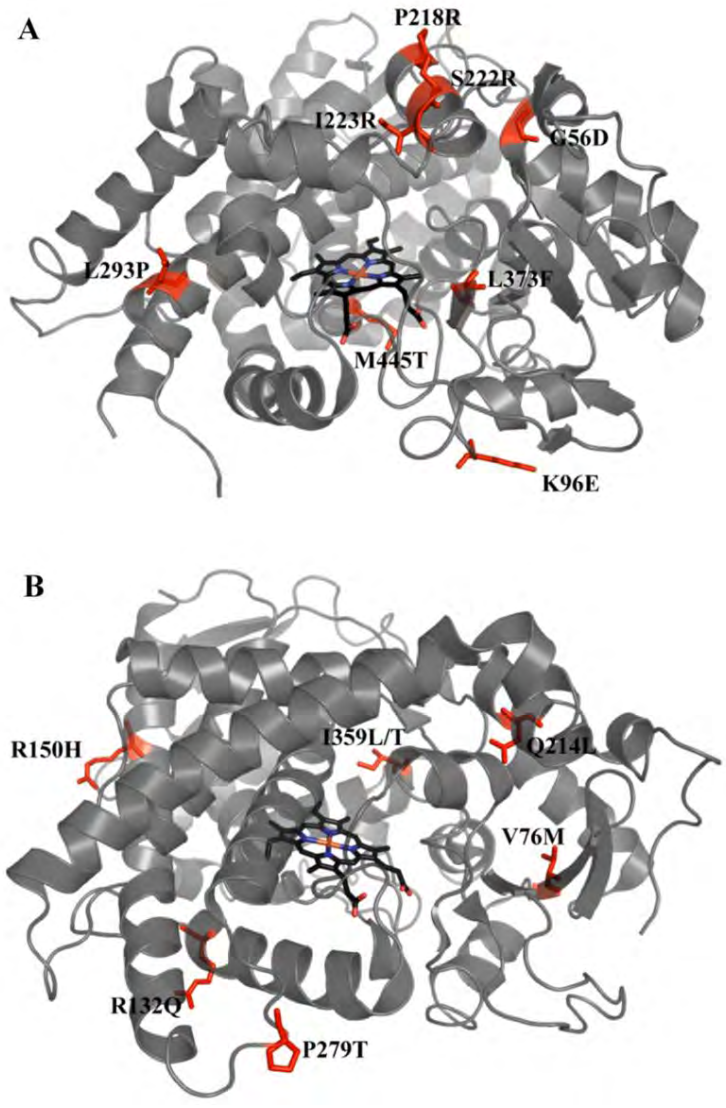


Figure 5.2 CYP3A4 and CYP2C9 structures showing the positions of amino acid substitutions in polymorphic variants selected for study. **(A)** CYP3A4, PDB structure 1TQN, and **(B)** CYP2C9, PDB structure 1OG2, are shown in grey with haem groups in black. The positions of mutations are labelled and indicated in red.

Table 5.1. Summary of information from *in silico* analysis and classifications based on experimental data for selected CYP450 variants.

P450 variant	Functional region location*	Predicted effect on protein stability†	Classification based on experimental data from literature[Ψ]
CYP3A4			
G56D	SRS1'a	destabilising	decreased activity
K96E	CPR contact	neutral	unknown
P218R	SRS(2,3)	neutral	unknown
S222P	SRS(2,3)	neutral	decreased activity
I223R	SRS(2,3)	destabilising	unknown
L293P	Start of SRS 4	neutral	increased activity
L373F	haem contact/SRS5	neutral	decreased activity
M445T	CPR contact	destabilising	neutral
CYP2C9			
V76M	undefined	neutral	unknown
R132Q	CPR contact/SRS2	neutral	decreased activity – decreased CPR affinity
R150H	undefined	neutral	increased activity - increased CPR affinity
Q214L	SRS(2,3)	stabilising	decreased activity
P279T	undefined	neutral	similar
I359L	SRS5	neutral	decreased activity
I359T	SRS5	destabilising	decreased activity

*Functional region location based on CYP450 SNP map (see Chapter 4, Figure 4.2)
† Based on Site Directed Mutator (SDM) scores: destabilising, $\Delta\Delta G < -2.0$ kcal.mol $^{-1}$; neutral, $-2.0 < \Delta\Delta G < 2.0$ kcal.mol^{-1}; destabilising $\Delta\Delta G > 2.0$ kcal.mol^{-1}
Ψ Variants were classified based on *in vitro* data as follows: "Inactive" - undetectable protein/holoprotein levels or activity < 2% of wild-type for all expression systems and substrates tested; "Decreased activity"- decreased expression levels, holoprotein levels or activity compared to wild-type; "Increased activity" – increased activity compared to wild-type; "Neutral" - similar holoprotein levels and activity to wild-type enzyme; "Unknown"- absent, inadequate or contradictory experimental data.

Recombinant CYP3A4, CYP2C9 and CPR wild-type constructs

Figure 5.3 A shows the plasmid expression vector pBJW102.2 used for cloning and expressing CYP450 proteins and CPR. This vector has a T5 promoter under the control of two lac-operator sequences allowing tightly regulated, graded, high-level protein production that can be induced with isopropyl-β-D-thiogalactopyranoside (IPTG). The lac repressor protein, coded for by the lac I gene, controls the basal activity of the lac-promotor; the β-lactamase coding sequence confers ampicillin resistance. The translated recombinant proteins have a BCCP-tag (~9 kDa) and a His6-tag fused to the C-terminus of the CYP450 or CPR proteins by G-S linkers (Figure 5.3 B). The BCCP-tag is quantitatively biotinylated by endogenous biotin ligase following the addition of free biotin to the *E.coli* cultures. CYP3A4, CYP2C9 and CPR were all expressed as N-terminal truncated proteins and the truncated sequences are shown in Figure 5.4.

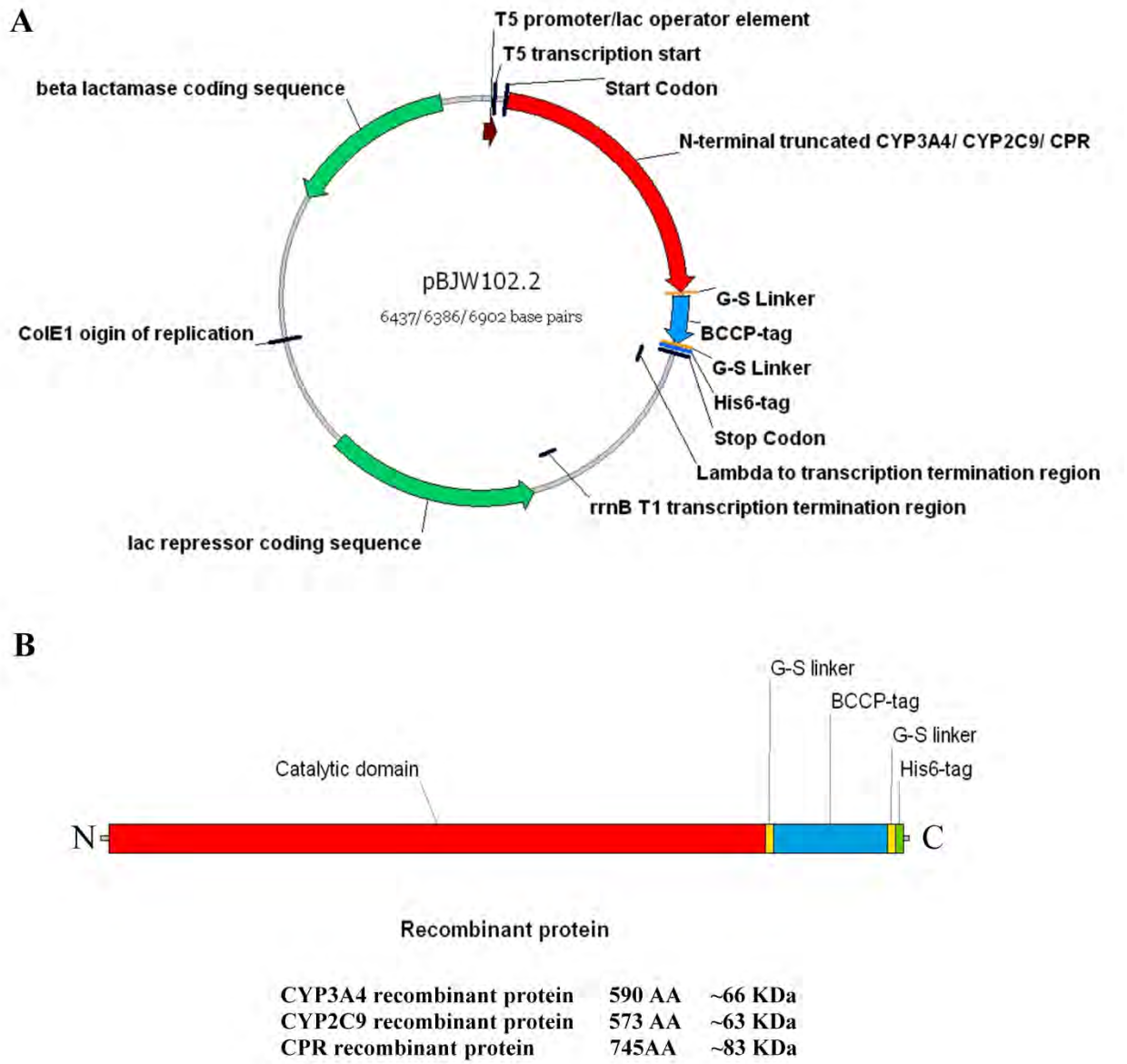


Figure 5.3 (A) Plasmid map of pBJW102.2 vector used for expressing CYP3A4, CYP2C9 and CPR proteins. The red arrow indicates the position of the CYP3A4, CYP2C9 or CPR gene inserted into the pBJW102.2 backbone for expression. The size of each of the resulting expression vectors is shown in the centre of the map. (B) A schematic representation of the recombinant protein constructs expressed using the pBJW102.2 plasmid show in A. The N-terminal truncated proteins are fused to a BCCP-tag followed by a polyhistidine-tag on the C-terminus. The length and size of the full recombinant protein constructs are given for CYP3A4, CYP2C9 and CPR.

Mutations corresponding to the selected variants were introduced into the CYP3A4 and CYP2C9 genes within the pBJW102.2 vector using inverse PCR, as described in Methods. DNA sequences for pBJW102.2 + CYP3A4 wild-type, pBJW102.2 + CYP2C9 wild-type and pBJW102.2 + CPR together with the translated protein sequences are given in Appendix C. The

point mutations introduced by inverse PCR and the resulting amino acid substitutions are also indicated within the DNA and protein sequences in Appendix C.

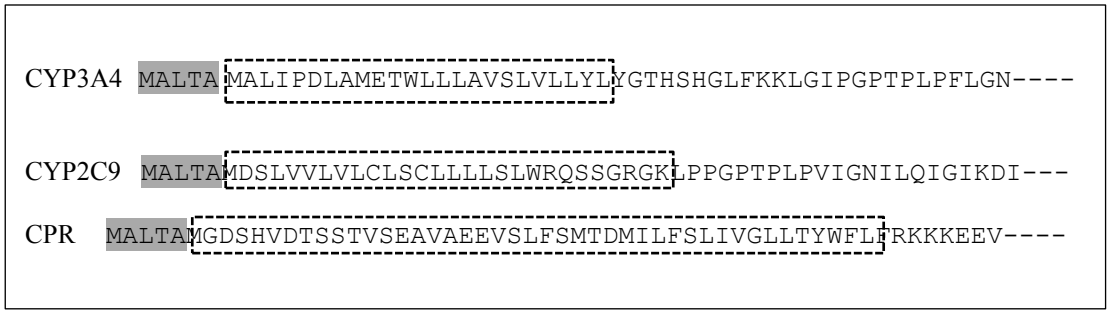


Figure 5.4 N-terminal truncated sequences for CYP3A4, CYP2C9 and CPR recombinant proteins. Boxes indicate sequences that were removed. Each construct started with the N-terminal sequence highlight in grey.

5.2.2 Protein expression and purification

All CYP3A4 and CYP2C9 wild-type and variant proteins were successfully expressed in *E.coli* DH5α. Soluble crude lysates were extracted and separated for visualization by electrophoresis on denaturing polyacrylamide gels (Figure 5.5 A & B). The recombinant Proteins expressed at low levels and bands of the correct size were not clearly visible on stained protein gels, but all variant proteins were detected by western blot analysis using both streptavidin-horseradish peroxidase (strep-HRP) (Figure 5.5 C & D) and anti-6xHistidine horseradish peroxidase (anti-His HRP) (Figure 5.5 E & F) indicating that both the biotinylated BCCP-tag and the His-tag were present.

All of CYP3A4 and CYP2C9 proteins showed some degree of proteolysis in western blots. A greater degree of proteolysis was visible for the CYP3A4 samples, which showed multiple bands in the 35 - 50 kDa and the 10 - 20 kDa regions. CYP2C9 proteins generally showed fewer and less distinct bands in these regions. The bands at 18 kDa, only visible in the Strep-HRP blots, correspond to the native biotinylated *E.coli* protein [476].

Recombinant proteins were partially purified using Ni-TED His-tag purification columns in the presence of 20% glycerol; crude lysates were kept on ice and all purifications were carried out on the same day as cell lysis. Partially purified samples were immediately concentrated and

buffer exchanged into a buffer suitable for protein storage to minimize protein unfolding and degradation. Figure 5.5 G & H show denaturing protein gels for partially purified CYP3A4 and CYP2C9 samples, respectively. The bands around 65 kDa, indicated by an arrow, correspond to the CYP450 proteins. Another dominant band appears at 40 kDa but, whilst there were bands at 40 kDa in western blot analysis of purified protein (data not shown) those bands were comparatively fainter relative to the 65 kDa band than the dominant band seen in Figure 5.5 G & H, indicating that the band at 40 kDa is likely to be a contaminating *E.coli* protein that co-migrates with the degradation products. After partial purification, the CYP3A4 proteins again show varying degrees of degradation while CYP2C9 proteins appear to be less susceptible to degradation.

Figure 5.6 shows protein gel and western blot analysis of CPR protein samples. Panel A shows a denaturing protein gel of CYP3A4 crude lysate and partially purified protein for easy comparison. A strong band at the expected molecular weight for CPR was present in the crude lysate and partially purified protein and is visible in western blots detected with both anti- His HRP and strep-HRP (Figure 5.6 B, C and D). Once again, there was a strong band at 40 kDa in the partially purified samples, which is not visible in the western blots, confirming the presence of a co-purifying *E.coli* protein. Increasing imidazole concentrations were included in the wash buffer in an attempt to reduce the levels of contaminating proteins in the partially purified protein samples; however, this did not substantially improve the purity but did reduce the protein yield (data not shown).

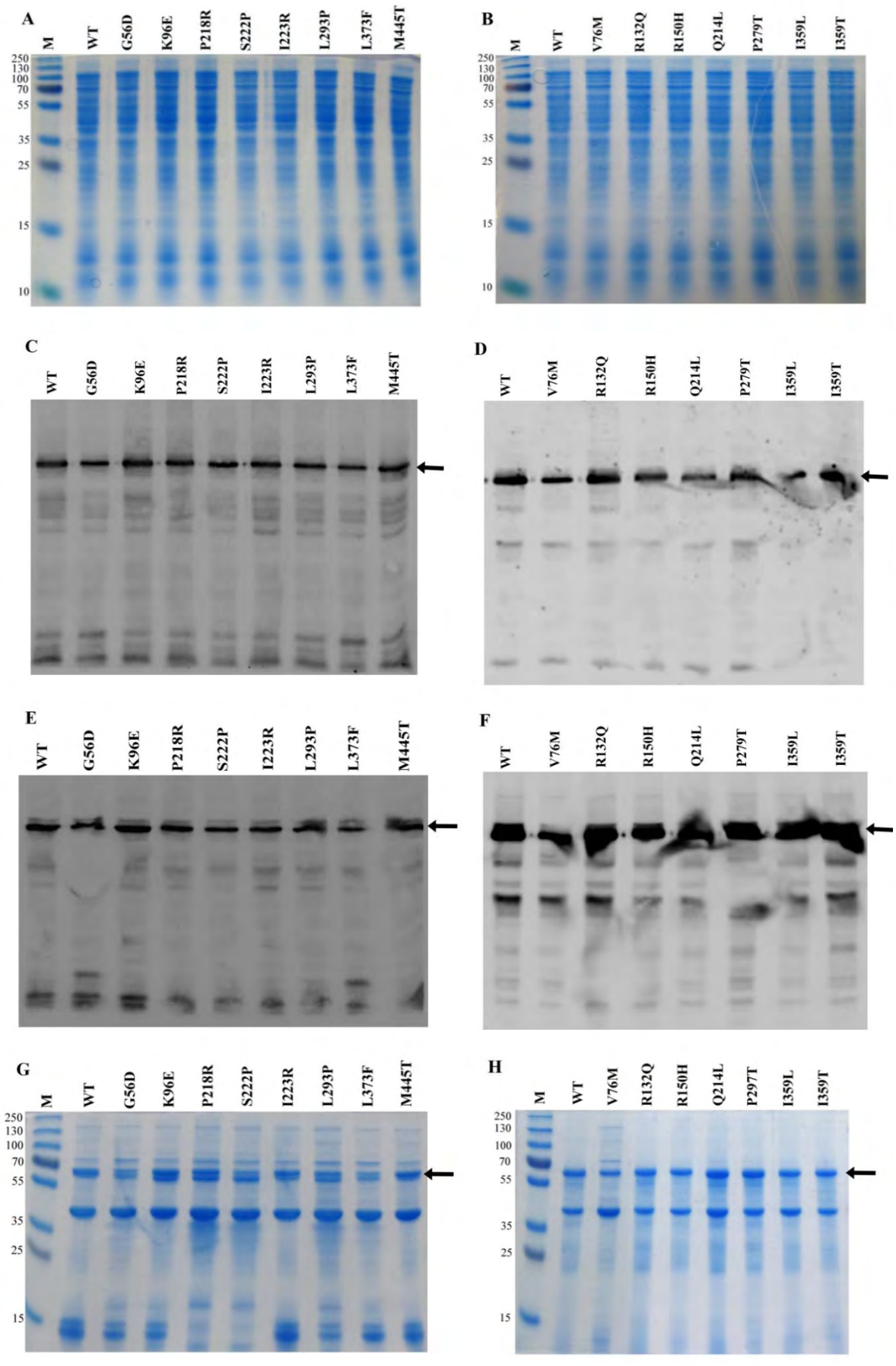

A
M
WT
G56D
K96E
P218R
S222P
I223R
L293P
L373F
M445T
250
130
100
70
55
35
25
15
10
B
M
WT
V76M
R132Q
R150H
Q214L
P279T
I359L
I359T
250
130
100
70
55
35
25
15
10
C
WT
G56D
K96E
P218R
S222P
I223R
L293P
L373F
M445T
D
WT
V76M
R132Q
R150H
Q214L
P279T
I359L
I359T
E
WT
G56D
K96E
P218R
S222P
I223R
L293P
L373F
M445T
F
WT
V76M
R132Q
R150H
Q214L
P279T
I359L
I359T
G
M
WT
G56D
K96E
P218R
S222P
I223R
L293P
L373F
M445T
250
130
100
70
55
35
25
15
H
M
WT
V76M
R132Q
R150H
Q214L
P297T
I359L
I359T
250
130
100
70
55
35
25
15

Figure 5.5 SDS-PAGE and western blots analysis of recombinant CYP3A4 and CYP2C9 wild-type and variant proteins. (A) and (B) show SDS-PAGE of soluble crude lysates containing CYP3A4 and CYP2C9 proteins respectively. ~10 µg of protein lysate was loaded per lane. (C) and (D) show western blots of soluble crude lysates containing CYP3A4 and CYP2C9 proteins detected with strep-HRP and (E) and (F) show western blots detected with anti-His HRP antibody. 5 µg of protein lysate was loaded per lane for western blot analysis. (G) and (H) show SDS-PAGE analysis for partially purified CYP3A4 and CYP2C9 proteins, respectively. Lanes containing molecular weight markers are indicated by "M" and labelled in kDa. An arrow indicates CYP450 proteins at expected molecular weights.

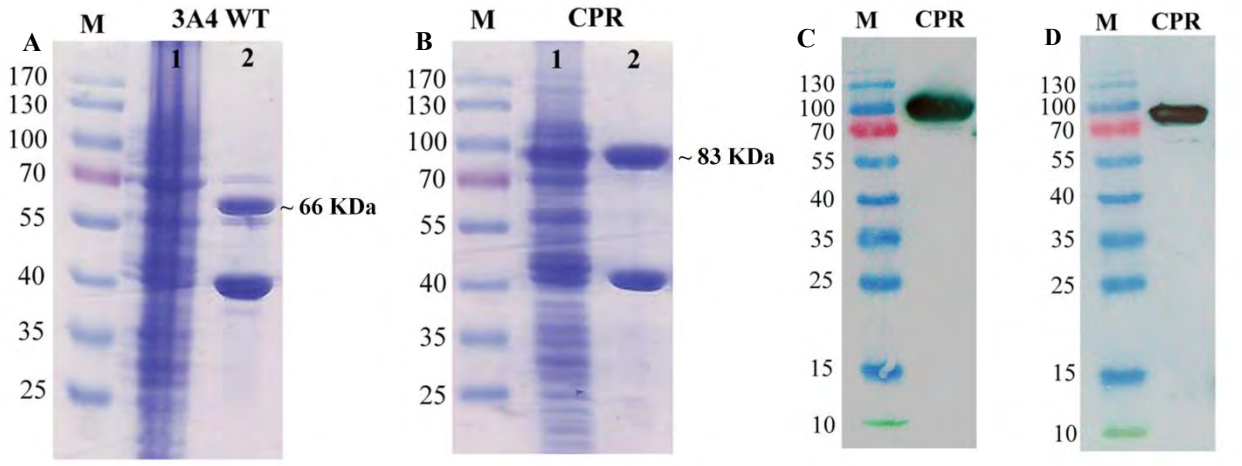


Figure 5.6 SDS-PAGE and western blot analysis of recombinant CPR protein. SDS-PAGE analysis of soluble crude lysate (lane 1) and partially purified protein (lane 2) for (A) CYP3A4 wild-type protein preparations (included for comparison) and (B) CPR protein preparations. (C) and (D) show western blots of soluble crude lysates containing CPR proteins detected with anti-His HRP antibody and Strep-HRP, respectively. Lanes containing molecular weight markers are indicated by "M" and labelled in kDa.

Specific protein quantification

The holoprotein content in crude lysates and partially purified samples was determined using carbon monoxide (CO) P450 spectral assays [5, 477]. This assay distinguishes between two forms of the holoprotein: the correctly folded P450 form and the inactive P420 form. The reduced ferrous from of the haem ion binds CO, yielding an absorption peak at 450nm, provided the haem group is correctly incorporated into the CYP450 (Figure 5.7). The inactive form (P420) has an absorption maximum at 420 nm and forms due to disruptions to the haem environment.

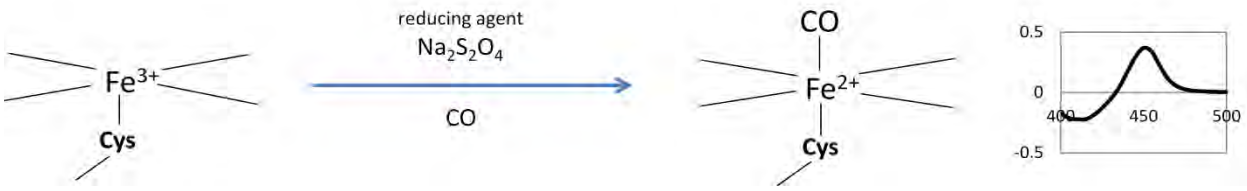
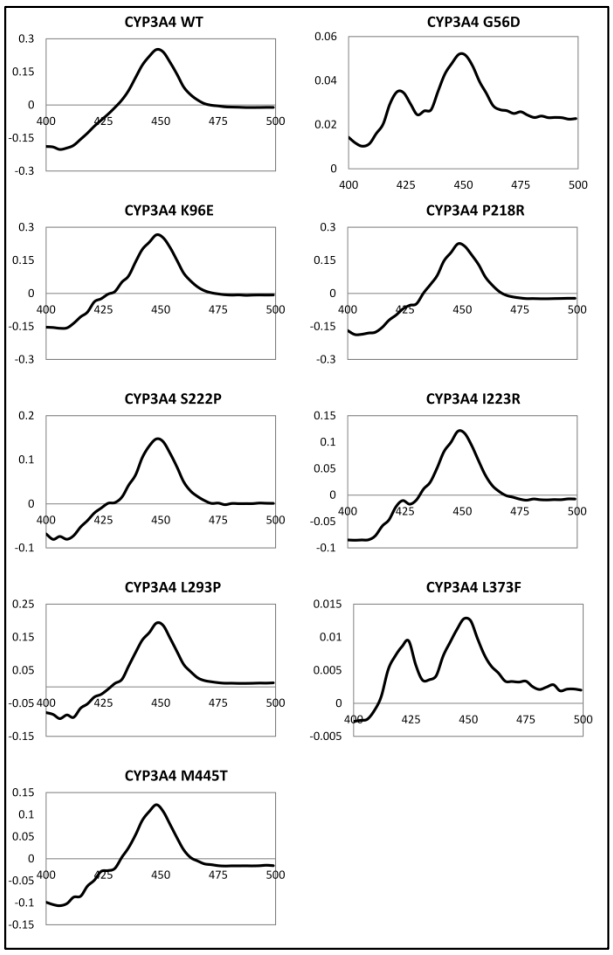


Figure 5.7 Schematic showing the reduction of ferric haem ion to ferrous ion in the presence of a reducing agent, followed by the co-ordination of CO at the axial position leading to an absorption peak at 450nm.



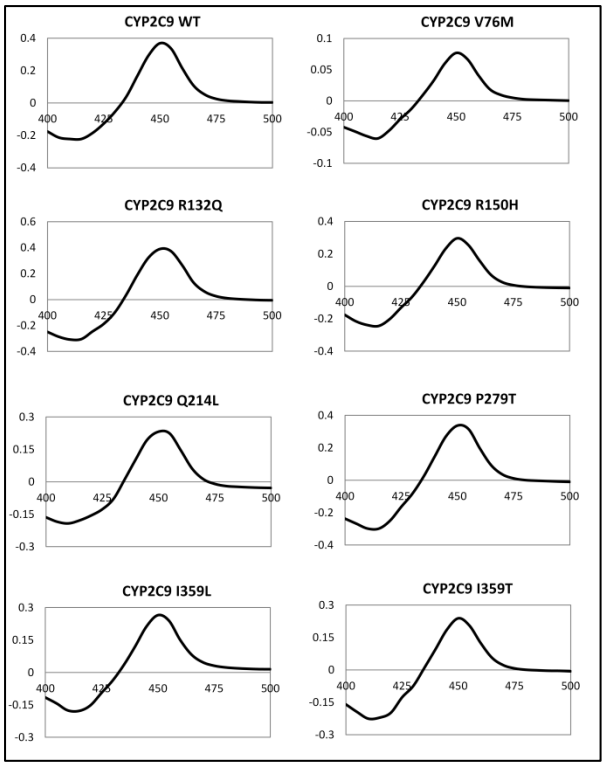


Figure 5.8 Ferrous CO vs. ferrous difference spectra for (A) CYP3A4 and (B) CYP2C9 wild-type and variant partially purified protein samples. Wavelength is shown on the x-axis and difference in absorbance is shown on the y-axis.

Figure 5.8 shows the ferrous CO vs. ferrous difference spectra for CYP3A4 and CYP2C9 wild-type and variant partially purified protein samples. The P450 content, P420 content and P420 to P450 ratio was calculated from these spectra (Figure 5.9).

The CYP3A4 protein samples all contained P450 and P420 protein at varying levels. All CYP3A4 variants showed elevated P420 to P450 ratios compared with the wild-type protein. G56D and L373F variants had the lowest P450 content – only 12% and 4% of wild-type - and the highest P420 to P450 ratios, suggesting that these mutations decrease protein stability and/or haem incorporation. While reducing the expression temperature from 30 °C to 25 °C for G56D and L373F variants reduced the P420 to P450 ratio, total P450 yield was also lower (data not shown).

CYP2C9 wild-type and variant protein samples showed only the P450 form of the protein and the P450 content was generally higher than for CYP3A4 protein samples. The V76M variant had the lowest P450 content of the CYP2C9 variants (20% of wild-type).

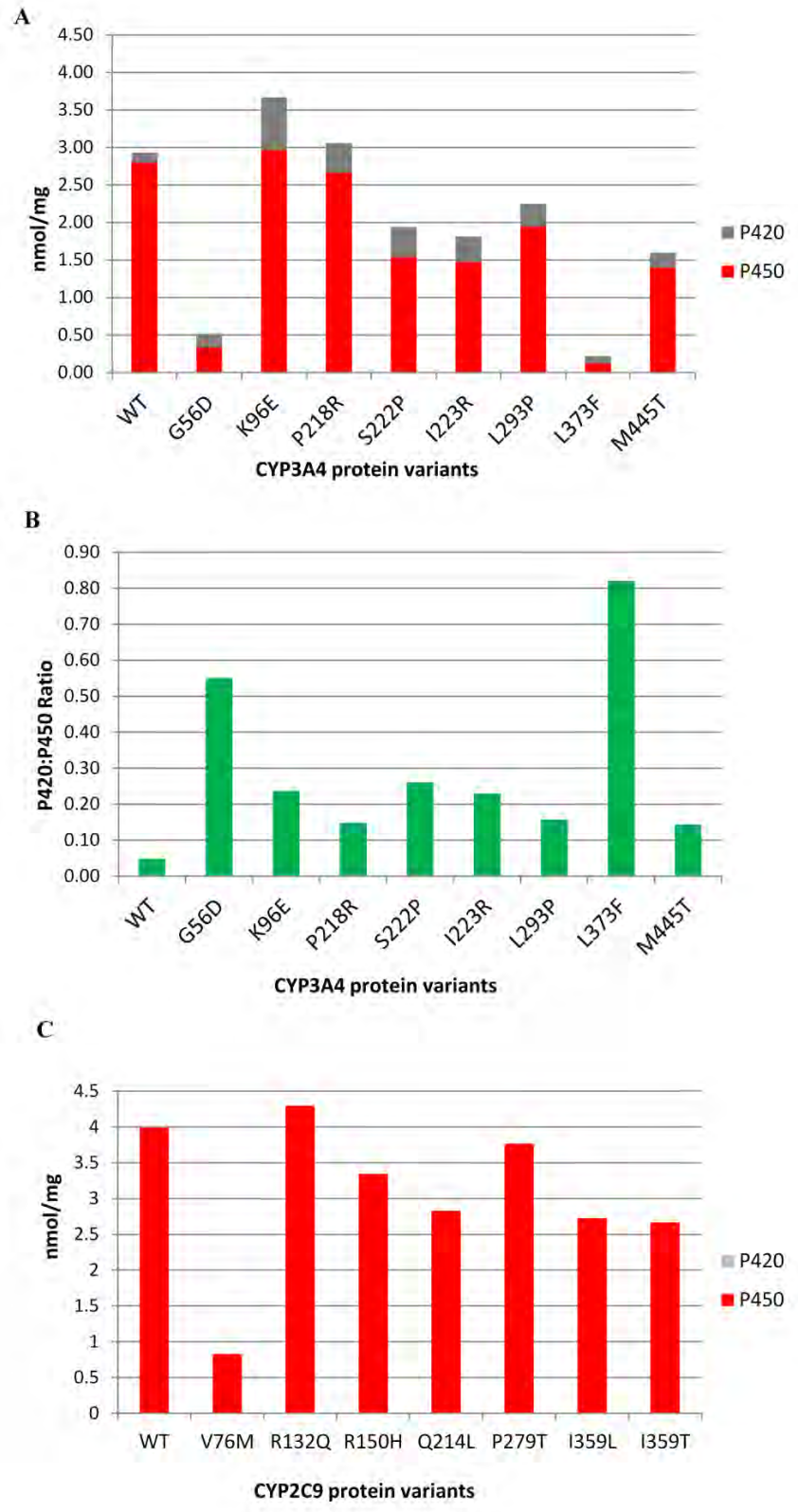


Figure 5.9 P450 and P420 protein content in CYP3A4 and CYP2C9 partially purified protein samples measured by CO P450 spectral assays. (A) nmols P450 and P420 per mg of total protein for CYP3A4 wild-type and variant samples. (B) P420:P450 ratios for CYP3A4 wild-type and variant samples. (C) nmols P450 per mg of total protein for CYP2C9 wild-type and variant samples. Note: P420 protein was not detected in any of the CYP2C9 protein preparations.

Table 5.2 Protein enrichment table for CYP3A4 recombinant proteins expressed in *E.coli*

3A4	Crude Lysates			Partially purified protein				
Variant	**Total Protein (mg)***	**Specific P450 protein (nmol/mg) Ψ**	**Total P450 (mg)**	**Total recovered Protein† (mg)**	**Specific P450 protein (nmol/mg) Ψ**	**Total recovered P450 (mg)**	**Fold enrichment**	**% recovery**
WT	560	0.17	6.3	3.2	2.8	0.6	16	9
G56D	448	0.02	0.5	2.5	0.3	0.1	18	10
K96E	592	0.25	9.8	2.5	3.0	0.5	12	5
P218R	528	0.14	4.7	1.7	2.7	0.3	20	6
S222P	416	0.1	2.8	2.0	1.5	0.2	15	7
I223R	320	0.08	1.7	2.9	1.5	0.3	19	17
L293P	480	0.1	3.2	2.2	2.0	0.3	20	9
L373F	384	0.002	0.1	2.4	0.1	0.02	60	37
M445T	544	0.11	4.0	3.6	1.4	0.3	13	8

* Total protein extracted from 1 litre *E.coli* culture (protein expressed for 18 h at 30°C). Ψ Specific P450 protein content measured by carbon monoxide P450 spectral assays. † Total protein recovered from a 1 litre *E.coli* culture (protein expressed for 18 h at 30°C) after His-tag purification.

Table 5.3 Protein enrichment table for CYP2C9 recombinant proteins expressed in *E.coli*

2C9	Crude Lysates			Partially purified protein				
Variant	**Total Protein* (mg)**	**Specific P450 protein Ψ (nmol/mg)**	**Total P450 (mg)**	**Total recovered Protein† (mg)**	**Specific P450 protein Ψ (nmol/mg)**	**Total recovered P450 (mg)**	**Fold enrichment**	**% recovery**
WT	368	0.34	7.9	5.6	4.0	1.4	12	18
V76M	432	0.06	1.6	4.4	0.8	0.2	14	15
R132Q	384	0.42	10.2	8.0	4.3	2.2	10	21
R150H	496	0.32	10.0	9.5	3.3	2.0	10	20
Q214L	352	0.24	5.3	5.4	2.8	1.0	12	18
P297T	464	0.40	11.6	8.7	3.8	2.1	9	18
I359L	272	0.21	3.5	9.0	2.7	1.5	13	44
I359T	416	0.31	8.1	15.0	2.7	2.5	9	31

* Total protein extracted from 1 litre *E.coli* culture (protein expressed for 18 h at 30°C). Ψ Specific P450 protein content measured by carbon monoxide P450 spectral assays. † Total protein recovered from a 1 litre *E.coli* culture (protein expressed for 18 h at 30°C) after His-tag purification.

Purification enriched the P450 content relative to total protein by 12 to 20 fold in all CYP3A4 protein samples apart from variant L373F, which was enriched 60 fold (Table 5.2). The total P450 protein recovered after His-tag purification was between 5 and 17 % for all CYP3A4

proteins except for L373F, which showed a 37% recovery. P420 was similarly enriched in these samples as His-tag purification does not selectively purify P450 from P420.

Purification enriched the P450 content in CYP2C9 protein samples by 9 to 14 fold and the recovery after purification was between 15 and 44 % (Table 5.3). The yield of CYP2C9 P450 protein was generally higher than the yield of CYP3A4 P450 protein, particularly after purification. CYP3A4 wild-type total P450 yield per litre of culture was 20% less than CYP2C9 before purification (6.3 vs 7.9 mg) and 60% less than CYP2C9 after purification (0.6 vs 1.4 mg) indicating that more CYP3A4 P450 protein is lost during the purification process, potentially through conversion to apoprotein or P420.

His-tag purification of CPR from crude lysate gave a total protein yield of ~ 1.5 mg per litre of *E.coli* culture. Based on a protein gel, which indicated that CPR made up ~50 % of the protein in the partially purified CPR protein sample, the CPR concentration after His-tag purification was ~ 6 nmol per mg of partially purified protein and the total CPR yield was ~ 9nmol per litre of *E.coli* culture.

5.3 Discussion

Recombinant P450 holoprotein was successfully expressed and purified for all the CYP3A4 and CYP2C9 variant proteins cloned in *E.coli*; however, expression levels varied and the ratio of active to inactive protein differed between isoforms as well as variant and wild-type proteins.

Protein yield before purification was ~ 95 nmol P450 per litre culture for CYP3A4 wild-type and 125 nmol P450 per litre of culture for CYP2C9 wild-type protein. Other studies have obtained expression levels of 370 nmol per litre culture for CYP3A4 [478] and 9-19 nmol per litre culture for CYP2C9 [471] in DH5α where proteins were expressed with modified N-terminal domains as opposed to truncated ones.

The differences between expression levels and protein recovery observed between isoforms and variant proteins may result from varied translational efficiencies, differences in protein targeting or varying levels of haem incorporation. Previously, western blot analysis showed that while truncating the N-terminal membrane-binding domain of CYP3A4 increases soluble protein expression, a large portion of the protein still expresses in the insoluble fraction [473]. Protein in the insoluble fraction may result from aberrant mis-folded protein that aggregates into inclusion bodies or from proteins associated with the membrane. CYP450s can associate with the membrane through peripheral or ionic interactions in the absence of the membrane-binding domain. Here, detergents were added to cell lysates to disrupt these interactions as previous work showed that the addition of detergents increased the levels of the CYP3A4 in the soluble fraction [473]. It is notable in this regard that the data shown in Table 5.2 and Table 5.3 reflects CYP450 concentrations determined by CO P450 spectral assays in which any apoprotein - whether mis-folded or not - would be invisible.

The other likely explanation for differences in P450 yields between proteins is that while proteins may initially express as active P450, more inherently unstable isoforms or variants may destabilise during cell lysis and purification, converting into P420 or losing their haem group altogether. This would be consistent with the observation that the difference in CYP3A4 wild-type and CYP2C9 wild-type P450 yields is greater in partially purified samples than in the crude lysates and that P420 was detected in CYP3A4 samples but not in CYP2C9 samples.

While the BCCP tag can act as a folding marker (the BCCP domain is only biotinylated if it folds correctly and it is only likely to fold correctly if the N-terminal protein folds correctly), folded apoprotein and protein that initially folds correctly and subsequently unfolds or converts to P420 will be biotinylated but inactive. This may be problematic for downstream assays, as the P450 active form is unlikely to bind preferentially to a Ni^{2+} or streptavidin surface, making it difficult to separate the active and inactive CYP450 protein.

The bands observed at the molecular weight corresponding to CYP450 proteins (~65 kDa) on the denaturing protein gels represent all forms of the CYP450 protein present: P450, P420 and apoprotein. While CO P450 spectral assays only quantify CYP450 holoprotein, this data combined with a rough estimate of total CYP450 from protein gels suggests that CYP3A4 wild-type protein in partially purified protein samples is made up of ~ 25 % apoprotein, < 5% P420 and ~ 70% P450 and that CYP2C9 wild-type protein in partially purified protein samples is made up of ~ 50% apoprotein, no P420 and ~ 50% P450.

All the CYP3A4 variants showed a somewhat higher P420 to P450 ratio than the wild-type protein after purification, but the differences were most striking for variants G56D and L373F. These variants also showed lower levels of total CYP3A4 in protein gels. Single point mutations can affect the solubility of a protein on overexpression and a study by Idicula-Thomas *et al* showed a positive correlation between thermostability and solubility of heterologous proteins expressed in *E.coli* in general [479]. Here, G56D was predicted to destabilise protein structure (Table 5.1), and while L373F is not predicted to be destabilising, Leu 373 is a haem contact residue so it is not surprising that this mutation alters the haem environment, leading to an increase in P420 levels.

P420 was not detected in any of the CYP2C9 protein samples, suggesting that CYP2C9 is more stable than CYP3A4. Specific P450 protein levels in partially purified samples ranged between 65 and 110 % of wild-type levels for all CYP2C9 variants except for V76M, which showed much lower P450 levels.

While His-tag purification enriched P450 proteins adequately for downstream applications, a large amount of P450 protein was lost during purification and the purity of samples was low. Larger purification tags such as glutathione S-transferase (GST) and maltose binding protein

(MBP) are known to give better protein yields but for the purposes of this work the His-tag was chosen as it is much smaller and therefore less likely to affect CYP450 conformation and function downstream [480].

Other affinity resins for His-tag protein purification were previously tested in the Blackburn laboratory in an attempt to improve the yield and purity of CYP450 samples. In a comparison between the cobalt-based resin TALON and the nickel-based resins HisLink, Ni-IDA and Ni-TED, Ni-TED gave the best balance between yield and purity [473]. In other work, MagReSyn NTA (ReSyn Biosciences, South Africa) gave much better purity than Ni-TED resin, but only the inactive P420 form was detected in those purified samples (personal communication, Omesan Nair). MagReSyn NTA beads also resulted in inactive protein when they were used to purify the haem enzyme catalase (personal communication, Andrew Nel), perhaps suggesting that the magnetic centres or the polyethylene imine coating of these beads interact with haem groups, dislodging or altering their position in the active sites.

While other purification methods in conjunction with nickel affinity chromatography could be used to improve purity, additional purification steps would be time consuming, decrease yield and potentially increase the ratio of P420 to P450. *E.coli* does not have any native CYP450 genes so it seemed unlikely that contaminating proteins would interfere with downstream assays; as such, efforts to further purify each of the 17 CYP450 proteins were not pursued.

In conclusion, while further optimisation may improve expression, purification and total P450 yields, P450 protein was obtained for all CYP3A4 and CYP2C9 variants selected for *in vitro* studies and levels were deemed suitable for downstream assays. It is clear from the results in this Chapter that single amino acid substitutions affect soluble expression levels and haem incorporation. P420:P450 ratios immediately suggest that some mutations have a substantial effect on the stability of the holoprotein structure *in vitro.* The effect of mutations on protein stability will be tested and discussed in the next Chapter.

5.4 Methods

5.4.1 Recombinant cloning of CYP3A4, CYP2C9 and CPR constructs

Original parent vectors

N-terminal truncated CYP3A4 and CYP2C9 constructs (both cloned into pBJW102.2 parent vectors) and a full length CPR construct (cloned into pMD004 vector) were a gift from Procognia Ltd, Maidenhead, UK. The N-terminal truncated version of CPR was then cloned into pBJW102.2 by a former post-doctoral fellow in our laboratory, Dr Siddharth Sharma. pBJW102.2 + CYP3A4, pBJW102.2 + CYP2C9 and pBJW102.2 + CPR vector sequences can be found in Appendix C. The pBJW102.2 vector was originally created by cloning the BCCP domain of the *E. coli* AccB enzyme (amino acids 74 - 156) and a glycine-serine linker sequence in frame, 3' to the His-6 tag of the commercially available pQE-80L *E.coli* expression vector (Qiagen, USA).

The translated CYP3A4 protein has a 24 amino acid deletion at the N-terminus and shows mutations V175I and V253I when compared to human CYP3A4 UniProt sequence P08684**.** These mutations have not previously been characterised as polymorphisms and are not predicted to have an effect on protein stability. V175I does not fall within a functional region and although V253I is the last residue of SRS (2, 3) it faces away from the active site and is not expected to affect enzyme function. For the purposes of this study, this sequence is considered the wild-type sequence. The translated CYP2C9 protein has a 28 amino acid deletion at the N-terminus and contains no mutations when compared to human CYP2C9 UniProt sequence P11712. The translated CPR protein has a 43 amino acid N-terminal deletion and has one mutation, V226L, when compared to the human CPR UniProt sequence P16435. This mutation is on the opposite side of the protein to the NADPH, FAD and FMN binding pockets and is not expected to affect protein function. CYP3A4, CYP2C9 and CPR recombinant protein sequences can be found in Appendix C.

Inverse PCR to introduce single point mutations into CYP3A4 and CYP2C9 genes

Inverse PCR, using the pBJW102.2 + CYP3A4 vector or the pBJW102.2 + CYP2C9 vector as a template, was used to generate variant expression vectors. Primers were designed using Vector

NTI Advance 11.0 (Invitrogen, USA) and synthesized in 5'phosphorylated form by Inqaba Biotechnical Industries, South Africa or by Integrated DNA Technologies, USA. (Table 5.4)

Table 5.4 Primers used to introduced single point mutations into CYP3A4 and CYP2C9 genes within the pBJW102.2 vector by inverse PCR

Mutation	Change	Forward Primer (5'-3')	Reverse Primer (5'-3')	T_A (°C)
3A4				
G56D	224 G>A	ACTTTTGTATGTTTGACATGGAATG*	CCTTATGGTAGGACAAAATATTTCC*	51
K96E	343 A>G	GAAGAATGTTATTCTGTCTTCAC	CACTAGCACTGTTTTGATCATGTC	54
P218R	710 C >G	GATTCTTTCTCTCAATAACAGTCT*	GATCCAAAAAATCAAATCTTAAAAG*	47
S222P	721 T>C	CCAATAACAGTCTTTCCATTCCTC*	GAGAAAGAATGGATCCAAAAAATC*	51
I223R	725 T>G	GAACAGTCTTTCCATTCCTCATC	TTGAGAGAAAGAATGGATCCAAA	54
L293P	935 T>C	CGGAGCTCGTGGCCCAATCAATTA	GATCGGACAGAGCTTTGTGGGACT	62
L373F	1174 C>T	TTTGAGAGGGTCTGCAAAAAAGATG*	TCTCATAGCAATTGGGAATAATCTG*	55
M445T	1391 T>C	CGAGGTTTGCTCTCATGAACATG	TGCCAATGCAGTTTCTGGGTCCAC	70
2C9				
V76M	271 G>A	ATGCTGCATGGATATGAAGCAGTG	CACTATGGGTTTCAGGCCAAAATA	57
R132Q	440 G>A	AGAATTTTGGGATGGGGAAGAGG	GCAGCGTCATGAGGGAGAAACG	62
R150H	494 G>A	ACTGCCTTGTGGAGGAGTTGAGA	GGGCTTCCTCTTGAACACGGTC	62
Q214L	686 A>T	TGATCTGCAATAATTTTTCTCCTA	GGATCCAGGGGCTGCTCAAAAT	57
P279T	880 C>A	ACATCTGAATTTACTATTGAAAGC	TTGGTTGTGCTTTTCCTTCTCCA	57
I359L	1120 A>C	CTTGACCTTCTCCCCACCAGCCT	GTATCTCTGGACCTCGTGCACCA	64
I359T	1121 T>C	CTGACCTTCTCCCCACCAGCCT	TGTATCTCTGGACCTCGTGCACC	64

* indicates primers synthesized by Integrated DNA Technologies; all other primers were synthesized by Inqaba Biotechnical Industries. All primers were 5'phosphorylated. Annealing temperatures (T_A) used for each set of primers are shown in the last column.

DNA amplification was performed using the KAPA HiFi PCR Kit (Kapa Biosystems, South Africa). The reaction set up and PCR cycling protocol used are shown in Table 5.5 and Table 5.6 respectively. A negative control containing no template DNA was included. After PCR optimization (25 µl reactions) and product verification by electrophoresis, bulk PCRs (2 × 50 µl for each variant) were performed.

Table 5.5 PCR reaction setup

Component	Final concentration
KAPA HiFi Buffer	1X (includes 2 mM $MgCl_2$)
dNTP Mix	0.3 mM
Forward Primer	0.3 mM
Reverse Primer	0.3 mM
Template DNA	50 ng per 50 µl reaction
KAPA HiFi DNA Polymerase	1U per 50 µl reaction
PCR-grade water	make up to total volume

Table 5.6 PCR cycling protocol

Step	Temperature	Duration	Cycles
Initial denaturation	95°C	4 min	1
Denaturation	98°C	20 sec	25
Annealing (T_A)	See Table 5.4	15 sec	
Extension	72°C	3.5 min	
Final extension	72°C	5 min	1

Agarose gel electrophoresis

The linear PCR products were verified electrophoretically on 1% agarose gels. Agarose gels were prepared by melting agarose (w/v) in Tris Acetate-EDTA buffer (TAE: 40 mM Tris base, 20 mM acetic acid, 1 mM EDTA, pH 8.0), adding 1:10 000 SYBR® Green (Life Technologies) to melted agarose for DNA visualisation, pouring it into a gel mould and allowing it to set. 1 × TAE was used as the running buffer and PCR products were mixed with 1 × DNA loading dye (Fermentas) and loaded into the gel wells. Mass Ruler^{TM} DNA ladder (Fermentas) was also loaded to estimate the size of the products and the original DNA templates were loaded as controls. Electrophoresis was carried out at 120V for ~40 min and the gels were viewed using the SynGene ChemiGenius Bio Imaging System (Synoptics Ltd, UK) and the SynGene GeneSnap software version 7.07.01 was used to acquire and process the gel images.

DNA-cleanup and Dpn1 digestion

The bulk PCR products were purified using GeneJET PCR Purification Kit (Fermentas) and treated with FastDigest *Dpn*1 (Fermentas) to restrict the parental template DNA, according to manufacturer's instructions. The PCR products were then separated by electrophoresis on a 0.8% agarose gel as described above. The target PCR bands were excised from the gel under Ultraviolet (UV) illumination with a sterile scalpel. The PCR products were then purified from the gel using the GeneJET Gel Extraction Kit (Fermentas) according to manufacturer's instructions. This was done to remove any non-specific PCR products from the sample. The purified DNA samples were quantified using the NanoDrop ND-100 spectrophotometer (NanoDrop Technologies, USA)

DNA ligation

The ends of the PCR product contained the 5'phospho group from the original primers, aiding the blunt end ligation of the linear DNA back into a circular plasmid form. DNA products were ligated using T4 DNA ligase (Fermentas) according to manufacturer's instructions for self-circularisation of blunt ended linear DNA.

Competent cell preparation

The rubidium chloride method was used to prepare competent DH5α cells. 5ml of Luria Broth (LB) was inoculated with freshly plated DH5α cells and incubated overnight at 37°C with shaking. The 5ml culture was then used to inoculate a 500ml culture and incubated at 37°C with shaking until OD_{600}~0.5. The culture was cooled on ice for 15min and cells were pelleted by centrifugation for 10min at 4500 rpm, 4°C. The cell pellets were resuspended in 30ml TfbI (100 mM RbCl, 50 mM $MnCl_2$, 30 mM KOAc, 10 mM $CaCl_2$, 15%(v/v) glycerol, pH5.8), left on ice for 30min and cells were again pelleted by centrifugation for 5min at 4000 rpm, 4°C. The cell pellet was resuspended in 6ml TfbII (10 mM MOPS, 75 mM CaCl2, 10 mM RbCl and 15% (v/v) glycerol, pH 6.5) on ice and stored in aliquots at -80°C.

DNA transformations

DH5α competent cells were thawed on ice for 10min. 10 µl of each ligation reaction (~10ng DNA) was added to a separate tube containing 100 µl of competent cells. An extra tube of competent cell, serving as a negative control, was included to which no DNA was added. Tubes

were swirled gently to mix and placed on ice for 30 min. The cells were heat shocked in a water bath set at 42°C for 90 s and then cooled instantly on ice for 2 min. 900 µl of LB, pre-warmed at 37°C, was added to each tube and tubes were incubated at 37°C for 2 h with gentle shaking. 20 µl of each transformation mix was plated onto separate Luria Agar plates supplemented with 100µg/ml ampicillin (LB-Amp agar plates) and plates were incubated at 37°C overnight. No colonies - or very few - grew on the negative control plates, while a large number of colonies formed on the sample plates, indicating that plasmid DNA containing the ampicillin resistant gene was successfully incorporated.

Plasmid DNA extraction for DNA sequencing

3 to 5 individual colonies were selected from each plate and used to inoculate 5ml LB supplemented with 100µg/ml ampicillin (LB-Amp) and incubated for 12-16 h at 37°C with shaking. Cells were harvested from 2 ml of the 5 ml cultures by centrifugation at 14000rpm for 10min. Plasmid DNA was isolated from these cells using GeneJET Plasmid Miniprep Kit (Fermentas) according to manufacturer's instructions, quantified by Nano Drop and sent for sequencing to confirm that point mutations were correctly incorporated.

Sequencing

Sequencing was carried out by Inqaba Biotechnical Industries or the Central DNA Sequencing Facility at the University of Stellenbosch using the dye terminator method. At least one correct clone was obtained for each variant.

Glycerol stocks

Sequence verified plasmid DNA was then used to transform DH5α cells to ampicillin resistance and recombinant colonies where grown in 5ml LB-Amp cultures as before. Glycerol stocks were prepared by diluting cultures 1:2 with 30% glycerol and aliquots were stored at -80°C for future use.

5.4.2 Protein Expression and Enrichment

CYP3A4, CYP2C9 and CPR protein expression

Glycerol stock were streaked out onto LB –Amp agar plates and incubated at 37°C overnight. Single colonies were inoculated into 5-10ml LB-Amp starter cultures and incubated at 37°C with shaking overnight. 5ml of each starter culture was then added to 500 ml of Terrific Broth (12g yeast extract powder, 6 g Tryptone powder, 0.4% glycerol in 89 mM potassium phosphate buffer pH 7.4) supplemented with trace elements (Final concentration: 250µM $FeCl_3$, 24µM $ZnCl_3$, 21µM $CoCl_2$, 21µM Na_2MoO_4, 0.15µM $CaCl_2$, 20µM H_3BO_3, 10µM $CuCl_2$, 0.03% HCl), 1 mM thiamine and 100µg/ml ampicillin. Two 500 ml cultures (total of 1 litre) were grown per protein. The 500 ml cultures were incubated at 37°C with shaking at 160 rpm until the cell density reached OD_{600} 0.4 - 0.6 (~4 h). 0.042g (0.5 mM) of δ-aminolevulinic acid (Sigma-Aldrich, Germany) was then added to CYP3A4 and CYP2C9 cultures to facilitate haem synthesis and cultures were incubated at 30°C with shaking at 160 rpm for a further 30 min. Protein expression was induced with 1 mM IPTG (Melford, UK) and 50 µM Biotin (Sigma-Aldrich, USA) was added to facilitate biotinylation of the BCCP-tag. Expression was carried out at 30°C with shaking at 120 rpm for ~18 h. Cells were harvested by centrifugation at 6000 rpm at 4°C for 15 min. Pellets from the two 500 ml cultures were washed in 1 x phosphate-buffer saline (PBS pH7.4, Sigma-Aldrich, Germany), combined, centrifuged as before and stored at -20°C.

Protein extraction

Cell pellets were thawed on ice and 14 ml of lysis buffer (20 mM phosphate buffer, pH 7.4; 20% glycerol (v/v); 10 mM β-mercaptoethanol) was used to resuspend a cell pellet harvested from 1 litre of culture giving a total volume of ~20 ml. Following this, 1.5 mg/ml lysozyme (Fluka, Belgium), 80 U/ml DNase I (Roche, Germany), 0.5 mM $CaCl_2$ and 2.5 mM $MgCl_2$ was added and the cell suspension was left on ice with gentle shaking for 30min. The following detergents were then added to disrupt aggregates and solubilise membranes: 0.15% IgePal CA-630, 0.5% CHAPS and 0.1% Triton X-100 (Sigma-Aldrich, USA). The suspension was left on ice with gentle shaking for a further 30 min. Crude soluble lysate was then separated from the cell debris by centrifugation at 10 000 ×g at 4° C for 30 min.

Protein purification

His-tag purification was carried out on the same day as protein extraction using Protino Ni-TED packet columns (Macherey-Nagel, Germany), according to manufacturer's instructions. These columns are packed with dry silica based resin that is pre-charged with Ni^{2+} ions. The Ni^{2+} ions are chelated to the resin by TED-pentadentate metal chelator which coordinates 5 of the 6 coordination sites on nickel. This leaves one site open to bind the polyhistidine tag on the recombinant protein, allowing purification. The polyhistidine- tagged proteins were eluted from the column with a buffer containing 300 mM imidazole. Imidazole, which has a high affinity for Ni^{2+}, displaces the protein from the nickel coordination site, allowing rapid elution of the protein of interest. The eluted samples were concentrated to < 700 µl using Amicon Ultra 15 ml centrifugal filters (10K MWCO; Merk Millipore Ltd, Ireland) by centrifugation at 4000 ×g at 4°C for ~ 45min. Protein samples were then buffer exchanged from the Ni-TED elution buffer into P450 storage buffer (20 mM potassium phosphate, pH 7.4; 20% glycerol (v/v); 0.2 mM EDTA; 1 mM DTT) using Zeba desalting columns (7K MWCO; Pierce, USA) according to manufacturer's instructions and aliquots were stored at -20°C. CYP3A4, CYP2C9 and CPR proteins stored in this way retained activity for over 8 months.

Total protein quantification

Total protein was quantified by Bradford assay [481] using the Bio-Rad Protein Assay Dye Reagent (Bio-Rad, USA) according to manufacturer's instructions. Crude lysates were quantified using the standard assay and partially purified samples were quantified using the microassay. Standard curves were generated using bovine serum albumin (BSA, Roche, Germany). Protein concentrations were measured in triplicate in a 96 well plate format using a Bio-Rad iMark microplate reader.

Gel electrophoresis

Protein samples were visualised using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Samples were suspended in 1× sample application buffer (12 mM Tris-HCl pH 6.8, 1% (v/v) β-mercaptoethanol, 0.4% (w/v) SDS, 5% glycerol, 0.02% (w/v) bromophenol blue) with vortexing and then boiled at 95°C for 5 min. The samples were then

loaded onto a two phase polyacrylamide gel immersed in 1x running buffer (2.5 mM Tris base, 0.19 M glycine, 0.1% (w/v) SDS, pH 8.3). The polyacrylamide gel generally consisted of a 10% separating gel (10% (w/v) acrylamide/bis-acrylamide; 375 mM Tris-HCl, pH 8.8; 0.1% SDS; 0.03% (w/v) ammonium persulfate; 0.1% (v/v) TEMED) overlaid with stacking gel (4% (w/v) acrylamide/bis-acrylamide; 125 mM Tris-HCl, pH 6.8; 0.1% SDS; 0.03% (w/v) ammonium persulfate; 0.3% (v/v) TEMED) to insure even migration of the protein. A prestained protein ladder (either PageRuler or PageRuler Plus from Fermentas) was also loaded onto the gel to estimate the size of the protein bands. Samples were separated by electrophoresis at 120 V for ~1 h using the Bio-Rad Mini Format 1-D Electrophoresis Systems (Bio-Rad, USA). The gel was immersed in Coomassie blue staining solution (0.05% (w/v) Coomassie brilliant blue, 50% (v/v) methanol, 7% (v/v) acetic acid) overnight followed by destaining (5% (v/v) methanol, 7% (v/v) acetic acid) or stained using Acqua Stain (Acquascience, UK) according to manufacturer's instructions.

Western Blot analysis

Protein samples were separated by SDS-PAGE as described above, followed by electroblotting at 100 volts for 1 h onto a BioTrace NT nitrocellulose blotting membrane (Pall Corporation, Mexico) using a wet blotting system (Bio-Rad, USA) and 1X transfer buffer (30 mM Tris, 0.19 M glycine, 20% (v/v) isopropanol). The membrane was immersed in Ponceau S reversible stain (Sigma Aldrich) to check protein transfer was successful and was then blocked in 5% fat free milk powder (w/v) in 1 x TBS-Tween (0.1 % Tween (v/v); 50 mM Tris base, pH7.5; 0.15 M NaCl) for 1 hour to prevent non-specific binding of the antibody to the membrane. The membrane was then probed with 1:2000 Streptavidin Peroxidase labelled conjugate (strep-HRP) (Kirkegaard and Perry Laboratories, USA) or 1:15 000 Anti-His HRP conjugated (Sigma-Aldrich) in 5% fat free milk powder (w/v) overnight at 4°C with gentle shaking. After three 5 min washes in 1x TBS-Tween, the SuperSignal® West Pico Chemiluminescent Substrate (Thermo Scientific, USA) was used for the detection of HRP according to manufacturer's instructions. The membrane was subsequently visualised using the SynGene ChemiGenius Bio Imaging System (Synoptics Ltd, UK) or exposed to X-ray film that was then developed.

5.4.3 Specific protein quantification: Carbon monoxide P450 spectral Assays

CYP450 protein samples were diluted in P450 storage buffer as shown in Table 5.7 and divided into two 1 ml plastic cuvettes (one reference and one sample cuvette). The cuvettes were placed into a Varian Cary 50 UV Visible spectrophotometer (Varian, Australia) and a baseline reading between 400 and 500 nm was recorded. Carbon monoxide (Speciality Gases, SA) was bubbled into the bottom of the sample cuvette (~60 bubbles at a rate of 1 bubble per second). A small spatula tip (~1 mg) of sodium dithionite ($Na_2S_2O_4$, Sigma-Aldrich, Germany) was added to each cuvette. Cuvettes were covered with parafilm and inverted several times to dissolve the sodium dithionite. The cuvettes were then placed back inside the spectrophotometer and spectra between 400 and 500 nm were measured several times until the peak at 450 nm stopped increasing.

Table 5.7 Protein concentrations used in carbon-monoxide P450 spectral assays for CYP3A4 and CYP2C9 protein samples.

Protein sample	Total protein concentration
CYP3A4 crude lysates	~ 10 mg/ml
CYP2C9 crude lysates	~ 5 mg/ml
CYP3A4 partially purified samples	1 mg/ml
CYP2C9 partially purified samples	1 mg/ml

Readings from the final spectra were used to calculate the amount of P450 and P420 protein present in the sample using the following equations:

$$[(\Delta A_{450} - \Delta A_{420})_{\text{observed}} - (\Delta A_{450} - \Delta A_{420})_{\text{baseline}}]/0.091 = \text{nmol P450 per ml} \quad [4.1]$$

$$[(\Delta A_{420} - \Delta A_{490})_{\text{observed}} - (\Delta A_{420} - \Delta A_{490})_{\text{theoretical}} - [(\Delta A_{420} - \Delta A_{490})_{\text{baseline}}]/0.110 = \text{nmol of P420 per ml} \quad [4.2]$$

$$\text{where } (\Delta A_{420} - \Delta A_{490})_{\text{theoretical}} = (\text{nmol P450 per ml from equation [4.1]}) \times (-0.041) \quad [4.3]$$

Chapter 6 Testing the effect of polymorphic variation on the thermostability of CYP450 proteins

6.1 Introduction

The folded structure of a protein is stabilised by an intricate network of weak interactions and changes to this interaction network can lead to conformational changes and altered protein stability. Protein stability can also be altered by chemical changes affecting covalent bonds within the protein structure [482]. While covalent changes are usually irreversible, the folded (native) and unfolded (denatured) state of a protein are in constant equilibrium due to the flexible and dynamic nature of proteins. In general, for wild-type proteins under physiological conditions, the folded state is much more stable than the less ordered unfolded state so proteins quickly resume their native structures making this dynamic equilibrium of little consequence to the function of proteins [483]. Changes in the protein environment as well as mutations can however shift the equilibrium towards the unfolded state. Combinatorial mutagenesis on thermostable CYP-119 protein revealed that salt bridges, charge-charge interactions, aromatic stacking and side chain volume of hydrophobic residues contribute to enhanced thermostability in the CYP450s [484].

Mutations are generally thought to affect the equilibrium between the folded and unfolded states of proteins by changing the free energy of the folded state. The unfolded or denatured state is typically however not a random coil but rather a poorly defined partially folded structure. As a result, the effects of a mutation on the stability of the denatured protein can also shift the equilibrium [483]. Consequently, a mutation increasing the conformational entropy of the unfolded state by destabilising the partially folded state can stabilise the native protein. While we can attempt to predict and rationalise the effects of mutations on the folded protein by analysing crystal structures, effects on the undefined denatured state are far more difficult to comprehend.

The conformational stability of a protein can be determined by calculating the equilibrium constant and free energy change for the conversion between the folded and unfolded states under a chosen condition [482]. Changes to the protein environment such as increased temperature, high salt concentrations, chemical denaturants, changes in pH and increased hydrostatic pressure can be used to drive protein unfolding by shifting the equilibrium towards the unfolded state. Techniques used to monitor protein unfolding include UV difference spectroscopy, circular dichroism, fluorescent-based assays, activity measurements, optical rotatory dispersion and nuclear magnetic resonance (NMR) [482].

Protein unfolding curves, comprising of pre-transition, transition and post-transition phases, can be generated by monitoring changes to the protein state as the severity of the denaturing condition is increased. It is essential that the unfolding reaction is reversible and reaches equilibrium for accurate thermodynamic measurements.

Choosing denaturation conditions and a monitoring technique depends largely on the equipment and quantity of protein available, as well as the structural information required. Fluorescent assays generally require less protein than other methods, however pre- and post-transition baselines are steep and highly sensitive to temperature, making fluorescence inappropriate for monitoring thermal unfolding [482]. While NMR requires large amounts of pure protein, it yields a much greater level of structural detail than other techniques. Circular dichroism monitors changes to secondary structures within the protein, while UV difference spectroscopy gives information on the environments of aromatic residues, giving insight into changes in tertiary structure [482].

CO P450 spectral assays offer a convenient method for monitoring changes between the correctly folded active state of the protein (P450) and the inactive state (P420). The switch between the P450 and P420 is caused by alterations to the cysteine co-ordinated haem environment. The P420 form still has a haem group attached within the hydrophobic cavity and it is generally accepted that the cysteine thiolate coordinated to the haem iron at the proximal position is either protonated, resulting in a weak thiol ligand [485-487] or replaced by another amino acid [21, 488]. The P420 form is not a uniform, well define-structure but appears to differ structurally under different denaturing conditions and between isoforms [489-492]. Several studies have suggested that the axial cysteine ligand is replaced by a histidine residue also located within the β-bulge or within the haem environment, however not all isoforms have a histidine residue in close proximity to the haem group. Lu *et al* showed that the thermal denatured CO bound ferrous CYP2C8 P420 has a protonated cysteine thiolate co-ordinating the haem iron and that the P450 state could not be regenerated by reducing the temperature [492]; this is in agreement with other studies showing that the thermal inactivation of CYP450s is irreversible [493, 494].

The aim of the work described in this Chapter was to determine the effect of polymorphic variation on CYP450 stability for the panel of CYP3A4 and CYP2C9 variant proteins that were cloned, expressed and purified as described in Chapter 5. Thermodynamic protein unfolding curves could not be accurately used to test the thermostability of the CYP450 due to the irreversible nature of the thermally induced P450 to P420 conversion. Generating unfolding curves using other denaturing conditions that show reversible denaturation, such as hydrostatic pressure, requires specialised equipment not available locally and can be protein intensive. However, since the goal here was to determine the relative stability of the variant proteins compared to the wild-type, it was deemed that reversible unfolding curves were not necessary. CO P450 spectral thermostability assays were therefore carried out at one constant temperature and the decrease in P450 was monitored over time to determine the half-life of each protein.

6.2 Results

To test the thermostability of the 17 CYP3A4 and CYP2C9 proteins produced in Chapter 5, CO P450 difference spectra were used to measure the decrease in P450 content at an elevated temperature as a function of time. When the CYP450 protein samples were incubated at an elevated temperature, the peak began to shift from 450 nm to 420 nm indicating the irreversible thermal conversion of active P450 into inactive P420 (Figure 6.1). Assays were performed using partially purified recombinant CYP3A4 and CYP2C9 proteins, previously expressed in *E.coli* (Chapter 5). Different temperatures were tested to ensure the decrease in P450 could be monitored at a constant temperature over a 30 minute period: 34 °C was chosen for CYP3A4 proteins and 48 °C for the more stable CYP2C9 proteins.

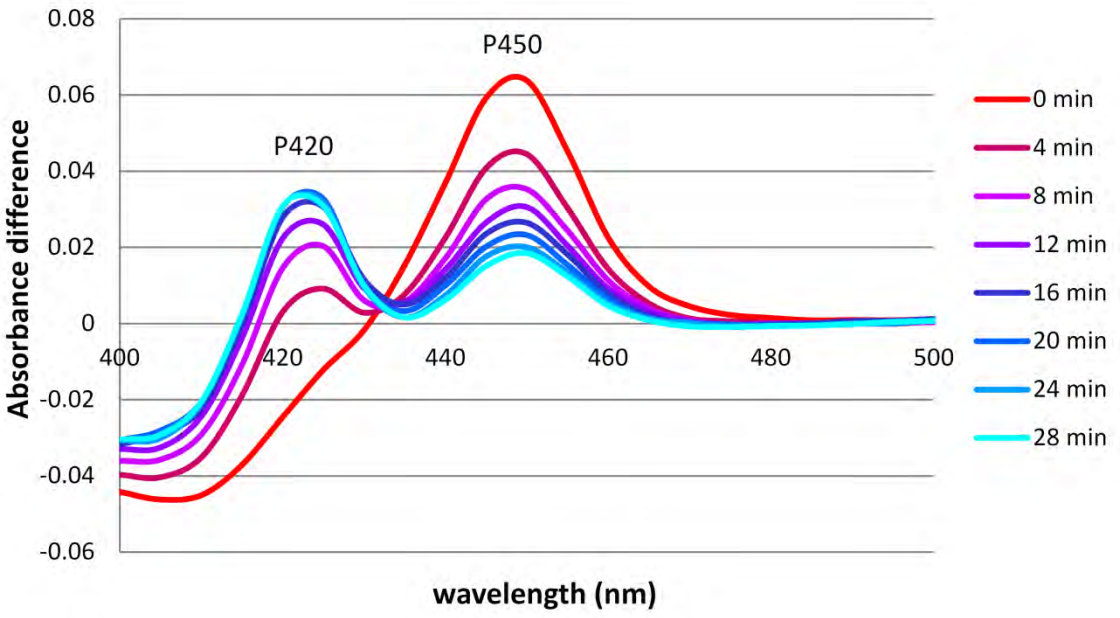


Figure 6.1 Ferrous CO vs. ferrous difference spectra for CYP23A4 wild-type thermostability assay. The difference spectra show a decrease in P450 and an increase in P420 with time. Assay was carried out at 34 °C and absorbance spectra were measured every 4 min are shown here.

CYP3A4 and CYP2C9 wild-type protein samples both showed an exponential decrease in P450 content, although the graphs did not tend to zero, suggesting either the presence of a stable P450 sub-population or background signal. When the temperature was elevated further, up to ~60°C, in both cases the peak at 450 nm disappeared rapidly followed by a decrease in P420 until no

peaks were visible (data not shown). This data implies that the asymptote - or 'plateau' - was probably a result of a more stable P450 sub-population rather than background signal.

6.2.1 Comparing the thermostability of CYP3A4 polymorphic variants

Figure 6.2 shows the results from the CYP3A4 thermostability assays. Both a one phase and two phase exponential decay model was fitted to the data sets using Graph-Pad Prism software. While the one phase model fitted all the data reasonably well provided the plateau was not constrained to zero, the two phase model was the preferred fit for wild-type and variants K96E, S222P, I223R and L293P. The more complex two phase model gave rate constants with very large confidence intervals that were often inconsistent between replicates. All data sets were therefore analysed using the one phase model, assuming that the contribution of the slower phase to the decrease in P450 content is very small due to the fact that it appears to have a much longer half-life than the fast phase, resulting in a graph that appears to reach a plateau.

To determine the effect of polymorphic variation on CYP3A4 stability, the effect of mutations on the half-life of the unstable sub-population and the effect of mutations on the plateau - the proportion of P450 making up the stable sub-population – were considered. The wild-type protein had a half-life of 6.3 minutes at 34°C and reached a plateau with 29 % P450 remaining, indicating that approximately one third of the P450 was in the more stable form.

The most notable differences in stability compared to wild-type were observed for variants G56D, L373F, P218R, S222P and M445T (Figure 6.2). G56D and L373F plateaued at ~ zero, indicating that these mutations abolish the stable P450 sub-population. They also had the lowest half-lives of all the variants, ~ 3 and 2 minutes respectively. P218R, S222P and M445T also had a very significant effect on the plateau, with only 7-8 % of P450 making up the stable sub-population; however, they only a moderate effect on the half-life (15 – 30% decrease) of the unstable sub-population.

I223R and K96E mutation had much smaller effects on stability while L293P showed no significant differences compared to wild-type (Figure 6.2).

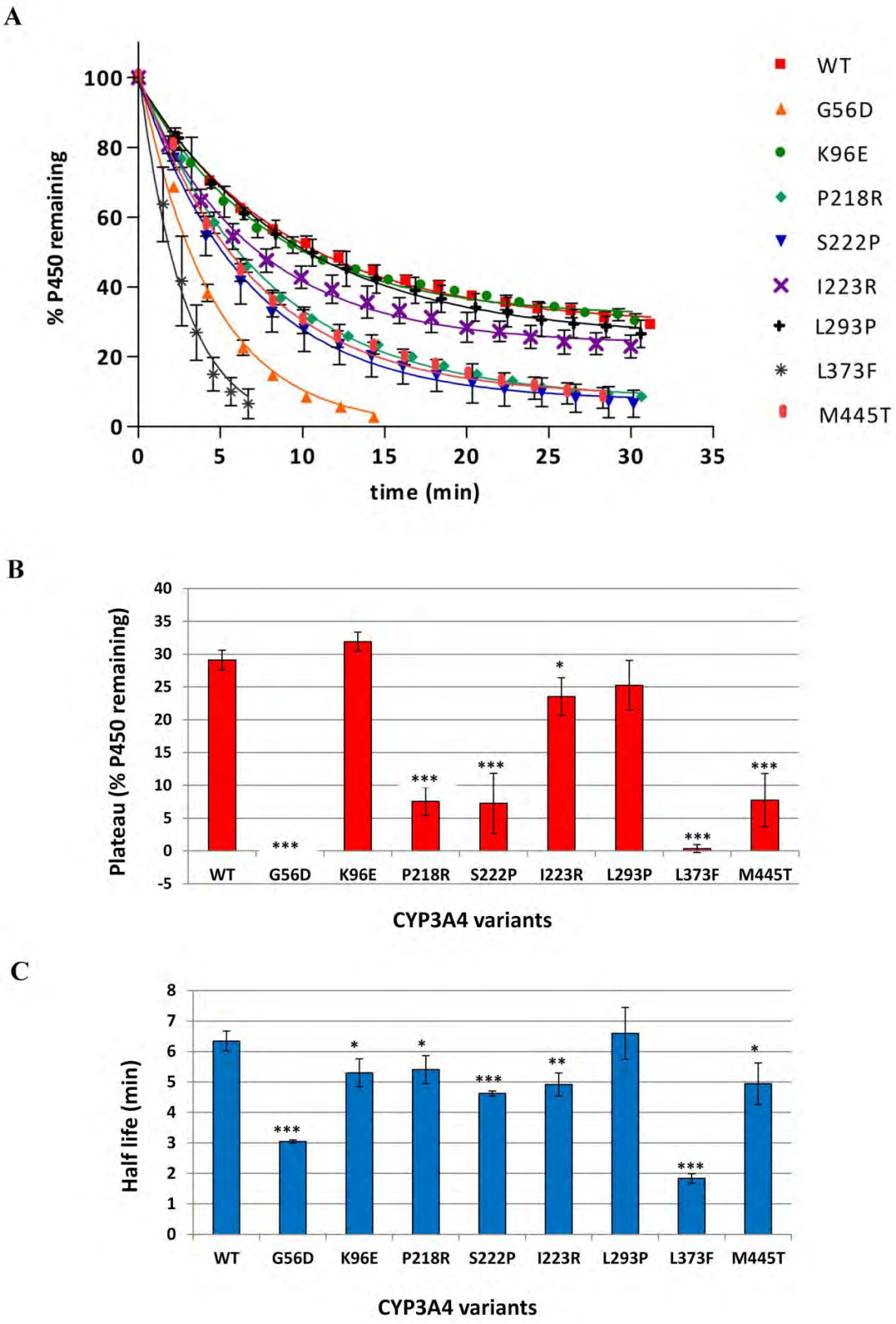


Figure 6.2 Thermostability of CYP3A4 polymorphic variants at 34°C. A) Graphs showing the decrease in P450 content at 34 °C over 30 minutes. A one phase exponential decay model was fitted to the data using non-linear regression as described in Methods. B) Comparison of % P450 remaining when the graph reached a plateau for each variant. C) Comparison of the half-life of each variant at 34 °C. Error bars represent the standard deviation of three replicates. Student t-test was used to determine values significantly different from wild-type: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$.

6.2.2 Comparing the thermostability of CYP2C9 polymorphic variants

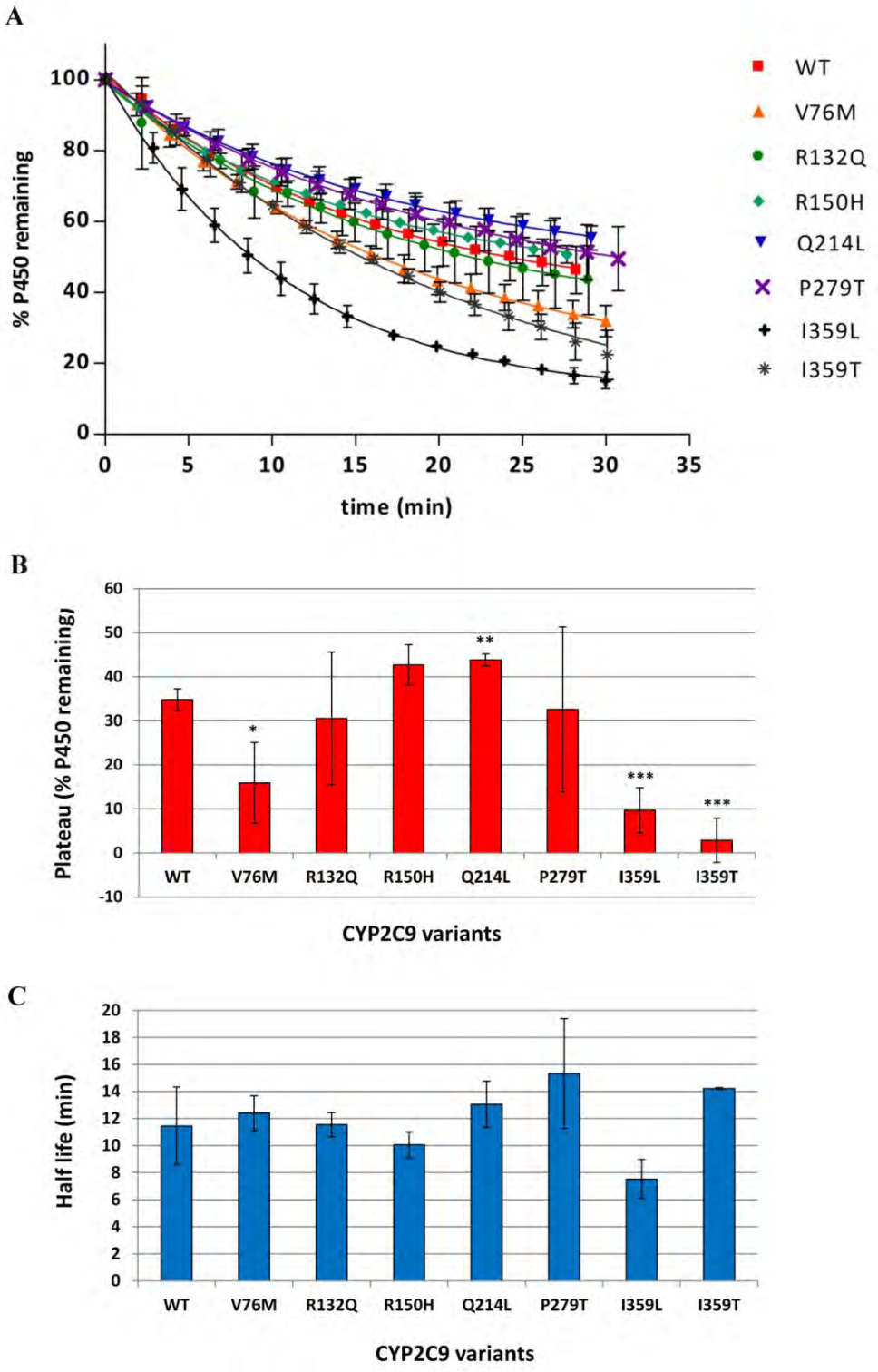


Figure 6.3 Thermostability of CYP2C9 polymorphic variants at 48°C. A) Graphs showing the decrease in P450 content at 48°C over 30 minutes. A one phase exponential decay model was fitted to the data using non-linear regression as described in Methods. B) Comparison of % P450 remaining when the graph reaches a plateau for each variant. C) Comparison of the half-life of each variant at 48 °C. Error bars represent the standard deviation of three replicates. Student t-test was used to determine values significantly different from wild-type: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$.

Figure 6.3 shows the results from the CYP2C9 thermostability assays. The one phase exponential decay model was the preferred model for all the CYP2C9 proteins. The wild-type protein had a half-life of 11.5 min at 48°C and the graph reached a plateau with 35% P450 remaining.

P279T had a half-life of 15.3 minutes (33% higher than wild-type) and I359L had a half-life of 7.5 minutes (35% lower than wild-type); however, differences in the half-lives were not significantly different for any of the CYP2C9 variants compared to wild-type.

Q214L, V76M, I359L and I359T did however have a significant effect on the plateau. Q214L had a higher proportion of P450 making up the stable sub-population compared to the wild-type protein, plateauing with 44% P450 remaining. V76M, I359L and I359T all had adverse effects on the stable P450 populations, with graphs plateauing at 15 %, 10 % and 3 % P450 remaining respectively.

R132Q and P279T had particularly large standard deviations, with one of the three replicates in each case giving curves with much lower plateau's than the other two replicates, making it difficult to determine whether these mutations have an effect on the stable P450 population.

6.3 Discussion

Thermostability of wild-type CYP3A4 and CYP2C9 proteins

CYP2C9 proteins were much more stable than CYP3A4 proteins, to the extent that CYP3A4 had a shorter half-life at 34°C than CYP2C9 had at 48°C. The observed differences in stability are consistent with the differences in expression levels and P450:P420 ratios determined in Chapter 5, where total P450 yield for CYP3A4 wild-type was 20% lower than CYP2C9 wild-type prior to His-tag purification and 60% lower after purification. Notably, the P420 form of the protein was not detected in any of the CYP2C9 partially purified samples, while the P420 form was present in all CYP3A4 samples. Proteins are often observed to be more stable in crude lysates than in partly purified protein preparations and can denature during purification. In this work, and important observation is that expression was carried out at 30°C for both CYP3A4 and CYP2C9 proteins, but this temperature is only 4°C lower than that subsequently used for the CYP3A4 thermostability assays, possibly explaining why a portion of the CYP3A4 protein was in the P420 form even before purification. This suggests that a lower expression temperature might improve P450:P420 ratios for CYP3A4 proteins; however, while expression of CYP3A4 proteins at lower temperatures did moderately improve P450:P420 ratios, this was at the cost of overall P450 yield.

Exponential decay curves for both CYP3A4 and CYP2C9 indicated energetically distinct P450 sub-populations that did not interconvert between each other during the course of the assay. In support of this idea, previous published studies examining the binding kinetics of CO to CYP3A4 using flash photolysis have argued that there are two kinetically distinct pools of P450, both in the membrane and in solution [128, 495, 496].

Other studies have also observed conformational heterogeneity for CYP2B4 [497, 498] and CYP3A4 proteins [124] using hydrostatic pressure spectroscopy. Elevated pressure typically causes a high-spin to low-spin shift of the haem iron [499-502], followed by the conversion of P450 to P420; these two pressure-induced processes can occur one after the other [503] or within overlapping pressure ranges [498, 502, 504]. Davydov *et al* observed that only 65 to 70% of CO bound ferrous CYP2B4 in solution was susceptible to a pressure-induced P450 to P420 transition in the absence of detergents, however all P450 could be converted to P420 in the presence of

detergent indicating that the observed heterogeneity may be a result of oligomer formation [497]. Similar heterogeneity was also seen for ferric CYP2B4 and CYP2B4 within proteoliposomal membranes [498].

In another study it was shown that 70% of CYP3A4 in solution and 50% in microsomes was susceptible to substrate-induced spin shifts and rapid, reversible pressure-induced P450 to P420 conversions [124]. The other fraction of the protein however remained predominately in the low-spin state. In solution, this remaining fraction was subject to slow, irreversible inactivation at high pressures indicating that the pressure induced P450 to P420 conversion of CYP3A4 was biphasic consisting of a fast reversible phase and a slow irreversible phase. In microsomes however only the fast reversible phase was observed. A pressure induce high-spin to low-spin shift was also detected in solution phase but not in microsomes, suggesting that the high-spin state is stabilised by interactions with other proteins and lipids within the membrane. Stabilisation of the high-spin state is thought to reflect decreased water accessibility of the haem [77].

Similarly to CYP2B4, the addition of detergents known to break up oligomers has been reported to reduce the heterogeneity of CYP3A4 observed in solution. This led the authors to suggest that the 1:2 conformer distribution in solution is a result of hexameric organisation, where 2 subunits have a different orientation to the other four. Interestingly both CYP2B4 and CYP1A2 have been shown to exist as hexamers in solution [119-121, 505].

In this Chapter, similar conformational heterogeneity for both CYP3A4 and CYP2C9 in solution was observed; only 65 – 70% of the P450 was susceptible to rapid heat-induced P450 to P420 conversion at a constant elevated temperature while the remaining fraction was much more stable, consistent with the literature. As suggested by Davydov *et al,* this 1:2 conformer ratio may indicate the formation of hexamers in solution for both CYP2C9 and CYP3A4, but gel filtration experiments and further thermostability assays done in the presence of detergents would be required to confirm this.

Effect of mutations on thermostability

A number of the single amino acid substitutions studied here affected the proportion of protein that was sensitive to the rapid heat-induced P450 to P420 conversion and/or affected the half-life of this conversion. Table 6.1 provides a summary of the results, comparing the predicted effects of mutations on the stability of the apo monomeric P450 structure (as determined *in silico* by SDM in Chapter 4) with the effect of mutations on the thermostability of the P450 proteins determined in solution in this Chapter.

Table 6.1 Predicted effects of mutations on protein stability versus experimentally determined effects on protein stability.

CYP450 variant	% RSA $^{\Psi}$ of wild-type residue	Predicted effect on stability by SDM†	Experimentally determined effect on thermostability
CYP3A4			
G56D	13.4	**destabilising**	**decrease**
K96E	60.7	neutral	none*
P218R	85.8	neutral	decrease
S222P	52.1	neutral	decrease
I223R	0.0	**destabilising**	**decrease**
L293P	54.1	neutral	none
L373F	13.5	neutral	decrease
M445T	39.6	**destabilising**	**decrease**
CYP2C9			
V76M	0.1	neutral	decrease
R132Q	63.2	neutral	none
R150H	68.3	neutral	none
Q214L	6.6	**stabilising**	**increase**
P279T	90.2	neutral	none
I359L	1.7	neutral	decrease
I359T	1.7	**destabilising**	**decrease**

$^{\Psi}$ RSA refers to the relative solvent accessibility as calculated by (Site Directed Mutator) SDM .
† Based on SDM scores: destabilising , $\Delta\Delta G < -2.0$ kcal.mol $^{-1}$; neutral, $-2.0 < \Delta\Delta G < 2.0$ kcal.mol^{-1}; stabilising $\Delta\Delta G > 2.0$ kcal.mol^{-1}
* K96E showed a small decrease in half-life however this may have been accounted for by the small (although insignificant) increase in the plateau and was therefore not classified as a destabilising mutation.

Not surprisingly mutations in close proximity to the haem binding region, such as L373F (haem contact) and M445T (predicted to be destabilising by SDM) in CYP3A4, appear to destabilise the haem significantly in both conformers, increasing the proportion of protein sensitive to rapid

heat-induced conversion to P420 and decreasing the half-life of this conversion. The most significant effect was observed for L373F, which is in direct contact with the haem (Figure 6.4): the slow phase was abolished completely and monophasic conversion to P420 occurred at a much more rapid rate. CYP3A4 mutation G56D had a similar effect to the L373F mutation; while Gly 56 is not in close proximity to the haem binding region, it was predicted to be a destabilising mutation by SDM.

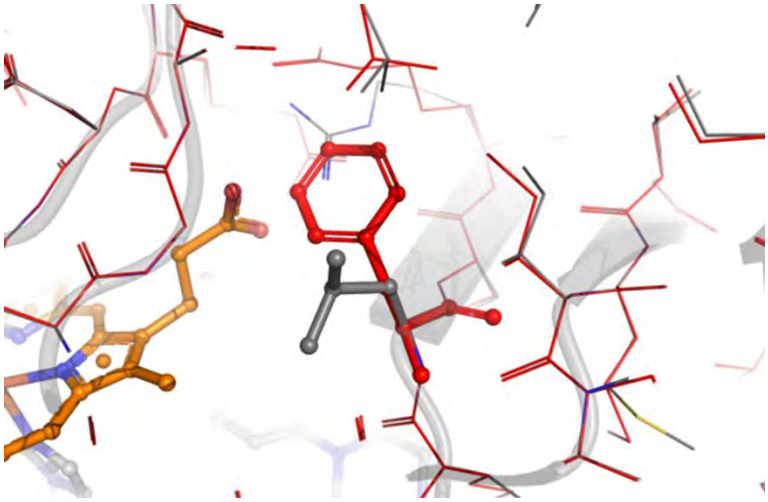


Figure 6.4 CYP3A4 L373F model aligned to the wild-type structure
The CYP3A4 L373F model (red) generated by ANDANTE aligned to the wild-type CYP3A4 pdb structure 1TQN (grey) with wild-type residue Leu373 and variant residue Phe373 shown in bold. The haem group, which interacts with the residue 373, is shown in yellow

For reasons less immediately apparent, P218R and S222P had a significant effect on the thermostability of CYP3A4; P218R and S222P were both predicted to be neutral by SDM and fall within the flexible F-G loop, well away from the haem binding region. However, it is interesting to note that the F-G loop has been implemented in the dimerization of mammalian P450s, both in solution and membranes [506, 507]. In membrane bound P450s this loop lies in close proximity to the membrane or is partially embedded within the membrane. Furthermore, CYP2B4 crystallises as a dimer along a two-fold symmetrical axis, made up of two open structures with the extended F-G loop from one molecule fitting into the open cleft of the second molecule (Figure 6.5 A) [506]. Close hydrophobic contacts between residues 213-230 and the binding partner were observed, whilst H226 co-ordinated to the haem iron, providing a spectral reporter for dimerization in solution. H226-Fe CYP2B4 dimers were also detected in solution and dimerization could be reversed to form catalytically active monomers.

The F-G loop also forms part of the dimer interface in CYP2C8 crystal structures [508] (Figure 6.5 B) and cross-linking experiments confirmed that this region also forms part of the dimer interface between CYP2C8 protein partners in membranes [509].

CYP3A4 mutations P218R, S222P and I223R and CYP2C9 mutation Q214L fall within the flexible F-G loop. While the quaternary structure of CYP3A4 in solution is unknown, based on evidence from other isoforms, it seems reasonable to speculate that if CYP3A4 is forming oligomers in solution as predicted, mutations in this region may affect interactions between binding partners.

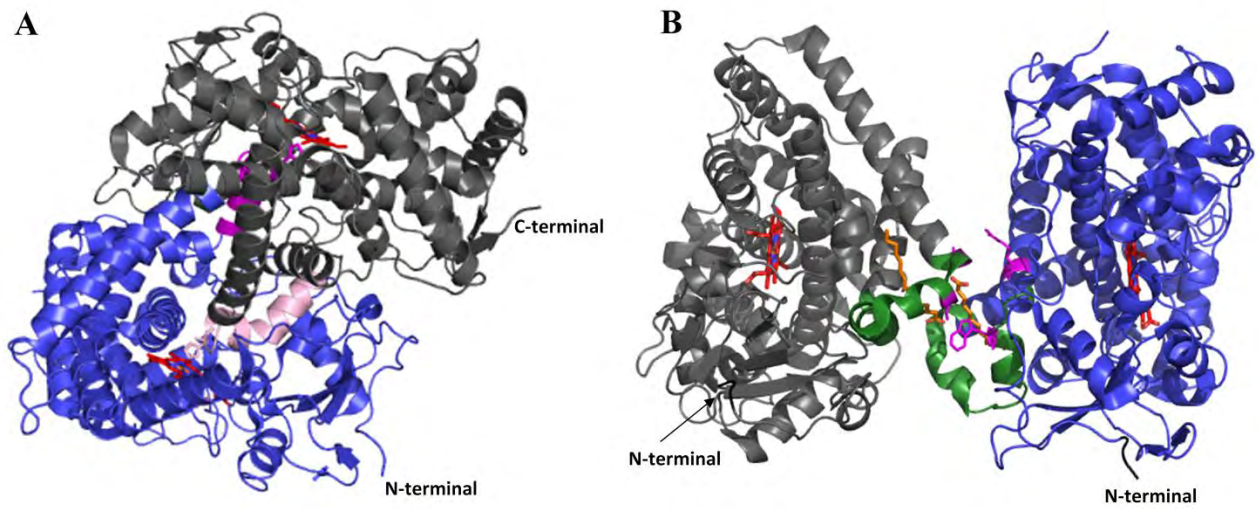


Figure 6.5 CYP450 dimerization observed in crystal structures. A) CYP2B4 crystal structure, PDB code 1PO5, rendered using PyMOL showing dimerization of two P450s in an open conformation related by two-fold symmetry. One molecule (blue) interacts with a second molecule (grey) by coordination of H226 to the haem iron and close hydrophobic interaction between residues 213-230 shown in purple and the open cleft of the binding partner. B) CYP2C8 crystal structure, PDB code 1PQ2, rendered using PyMOL showing dimerization of two asymmetric CYP2C8 molecules *via* the F-G loop (green). Residues making the closest contact between molecules are shown in purple and the two palmitic acid molecules bound at the interface of the dimer are shown in yellow. The orientation of the dimer is compatible with membrane binding as the N-terminals of both molecule lie on the same side of the dimer.

The dominant effect of P218R and S222P mutations on thermostability is a shift in towards the unstable, heat sensitive conformer, which could be attributed to the disruption of oligomer formation. I223R causes a significant ($P < 0.05$) but more moderate shift towards the unstable conformer and a moderate decrease in the half-life of the fast phase. Unlike Pro218 and Ser222, which are both solvent exposed residues with side chains orientated towards the bulk solvent, Ile 223 is buried (Table 6.1) indicating that it is unlikely to interact with other CYP3A4 molecules

directly without a substantial change in protein conformation (Figure 6.6). I223R is however predicted to be damaging to protein structure by SDM so the observed effects on thermostability may be as a result of destabilisation of the secondary/tertiary structure rather than a direct effect on the quaternary structure.

CYP2C9 variant Q214L, a buried residue within the F' helix, was the only variant predicted to be stabilising by SDM and the only variant that showed a significant increase in thermostability, showing a significant increase in the proportion of stable conformer in solution, but no significant difference in the half-life of the fast phase.

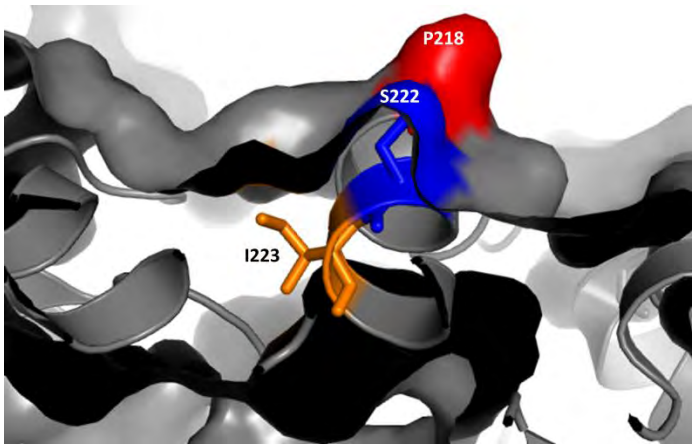


Figure 6.6 CYP3A4, PDB structure 1TQN rendered using PyMOL, illustrating the surface accessibility of residues P218 (red), S222 (blue) and I223 (yellow) within the F' helix. P218 and S222 side chains form part of the surface of the molecule whereas I223 is buried beneath the surface.

I359L and I359T mutations in CYP2C9 decrease the proportion of protein making up the stable P450 sub-population. Ile358 is the last residue of α-helix K and is buried within the hydrophobic core of the protein. SDM correctly predicts that a mutation to threonine at this position will destabilise the protein fold but incorrectly predicts the effect of the conserved mutation to leucine; while isoleucine and leucine are both hydrophobic residues with the same volume and similar properties, the subtle differences between these residues at this position appear to have a significant effect on the conformation of the CYP2C9 fold.

CYP2C9 variant V76M also showed significantly less P450 in the stable sub-population than observed for the wild-type protein but was not predicted to be a destabilising mutation by SDM. V76 is a buried residue within strand 2 of β-sheet 1 and does not fall within any know functional region of the protein.

In conclusion, 10 out of the 15 variants proteins tested had significantly altered thermostability compared to the corresponding wild-type protein. The results suggest that the observed effects of mutations on protein thermostability in solution may result from changes in the haem binding environment, alterations in oligomer formation or destabilisation of the protein fold. Based on these *in vitro* results, *in silico* predictions by SDM had a specificity of 100% (5/5) and a sensitivity of 50% (5/10). SDM's low sensitivity is expected since effects on haem binding and quaternary structure are not accounted for by SDM: of the 5 false negatives, as predicted by SDM, one (L373F) is a haem contact and two may alter oligomer formation (P218R, S222P); the other two variants were either incorrectly predicted by SDM or have long range effects on oligomer formation, P450 conformation or haem binding, without significantly disrupting the stability of the protein fold.

Based on the evidence, it seems reasonable to speculate that CYP3A4 adopts a hexameric structure in solution, in which two monomers are intertwined, as seen for the CYP2B4 crystal structure where the F-G loop from one monomer fits into the active site of the other monomer, by inference producing a more stable P450 form that can be disrupted by mutations in the F-G loop. In support of this hypothesis, kinetic data for CYP3A4 described in Chapter 7 suggests that two kinetically distinct CYP3A4 conformers exist in solution.

If oligomer formation does indeed play a key role in CYP450 regulation in membranes [111, 112], the effect of mutations on P450:P450 interactions may be an important consideration when determining the effect of polymorphic variation on drug metabolism, but accounting for such quaternary effects of polymorphic variation is beyond the scope of this work.

6.4 Methods

6.4.1 Thermostability assays

Partially purified CYP450 protein samples were diluted in P450 storage buffer to a concentration of ~ 1 µM P450, estimated based on CO P450 spectral assays performed on similar protein preparations (Section 5.4.3). 250 µl of the diluted sample was place into 2 quartz cuvettes: one reference and one sample cuvette (quartz cuvettes, holding smaller volumes, were used instead of 1 ml plastic cuvettes used previously in order to save on protein sample). The cuvettes were placed into a Varian Cary 50 UV Visible spectrophotometer and a baseline reading between 400 and 500 nm was recorded. CO was bubbled into the bottom of the sample cuvette (~30 bubbles at a rate of 1 bubble per second). A small spatula tip (< 1 mg) of sodium dithionite was added to each cuvette. Cuvettes were covered with parafilm and inverted several times to dissolve the sodium dithionite and incubated at room temperature for 5 min to allow time for the P450 peak to reach a maximum prior to starting the thermostability assay. The cuvettes were then placed back into the spectrophotometer maintained at a constant temperature of 34°C or 48°C by circulating water from a heated water bath through the cell holder. Several readings were taken over the first 30 seconds to ensure the P450 peak had stopped increasing and to allow time for the cuvette to equilibrate to the new temperature. The reading with the highest P450 peak was taken as time zero and following this readings were taken at 2 minute intervals for 30 minutes. Assays were carried out in triplicate. The amount of P450 in each sample at each time point was calculated using the equations described in Section 5.4.3.

6.4.2 Data analysis

Time vs. % P450 remaining was plotted for each replicate. The % remaining P450 was calculated by comparing the concentration of P450 in the sample at each time point to the initial P450 concentration at time zero. Readings on the spectrophotometer had to be taken manually at each time point so time points varied slightly between replicates. Figure 6.2 A and Figure 6.3 A shows average time vs. average % P450 remaining for visualisation purposes, however average plateau's, half-lives and standard deviations were calculated by fitting a curve through each replicate data set separately to account for differences in time points. Whether the data was

averaged prior to curve fitting or afterwards however made no significant difference to the results.

Graph-Pad Prism software (San Diego, USA) was used to fit both one phase (equation 6.1) and two phase (equation 6.2) exponential decay models to the data sets using non-linear regression. The plateau was constrained to values greater than zero. A comparison of the two models was performed in Prism using the F-test. The model that fitted better to 2 or more of the 3 replicate data sets was considered the preferred model for the given variant protein. Reported parameters are however all based on the one phase model.

$$Y = (Y_0 - Plateau)e^{-Kx} + Plateau \qquad [6.1]$$

Where Y_0 is the Y value when X (time) is zero (*i.e.* 100% P450), Plateau is the Y value at infinite X (time), K is the rate constant (min^{-1}) and half-life = ln(2)/K.

$$Y = Plateau + SpanFast^{*}e^{-Kfast^{*}X} + SpanSlow^{*}e^{-Kslow^{*}X} \qquad [6.2]$$

Where SpanFast + SpanSlow + Plateau = Y_0, Kfast and Kslow are the two rate constants (min^{-1}), half-life(fast) = ln(2)/Kfast and half-life(slow) = ln(2)/Kslow.

Chapter 7 Testing the catalytic activity of CYP3A4 and CYP2C9 proteins

7.1 Introduction

In vitro CYP450 kinetic data has long been used to predict *in vivo* CYP450 drug metabolism as well as drug-drug interactions. Many CYP450 kinetic profiles show hyperbolic saturation kinetics *in vitro* and classic Michaelis-Menten kinetics can be used to determine kinetic parameters. With an increase in available *in vitro* data, it is however becoming evident that atypical kinetic profiles observed for CYP450s, first thought to be an exception to rule, may in fact be the norm for many CYP450 isoforms. These unusual kinetics have most commonly been associated with CYP3A4, but have also been reported for CYPs 1A2, 2B4, 2B6, 2C8, 2C9, 2D6, 2E1 and 3A5 [61, 63]. CYP450 atypical kinetics are substrate dependant, with different substrates potentially showing different kinetic profiles when metabolised by the same CYP450 isoform. Michaelis-Menten models are therefore often erroneously forced though data sets that are actually atypical, or the data sets are simply truncated removing the atypical portion of the curve (usually occurring at high substrate concentrations) in both cases leading to inaccurate estimations of V_{max} and K_m. The correlation between atypical *in vitro* kinetic data and *in vivo* data is still poorly understood [510], however the allosteric effects causing atypical kinetic profiles may play an important role in regulation of substrate metabolism *in vivo*. Even if the atypical kinetics observed for a substrate is an artefact of the *in vitro* system and does not occur *in vivo*, it will affect the scaling of data to an *in vitro* system and therefore should not be ignored [510, 511].

A number of theories, with supporting experimental evidence and kinetic models, attempt to explain the atypical kinetic profiles observed for these enzymes. These include multiple substrate/effector binding sites [61, 512, 513], multiple kinetically distinct enzyme conformations [62] and CYP450-CYP450 allosteric interactions [112]. While each theory may explain some individual data sets, the combined evidence suggest that in reality a combination of all these factors may be at play [60].

Multiple binding sites

The failure of substrates to conform to Michaelis-Menten kinetics and the lack of mutually competitive inhibition between substrates metabolised by the same enzyme first indicated the

possibility of several distinct binding sites within the CYP450 active site. This was later supported by crystallographic data giving direct evidence of more than one substrate molecule bound within the active site e.g. PDB structure 2V0M shows two molecules of ketoconazole bound within the active site of CYP3A4. Evidence from mutagenesis studies [251, 514], spectral binding studies [515, 516], studies using FRET and Job's titration to determine the stoichiometry of the CYP450−substrate complexes [517], as well as studies using fluorescent probes, also support the presence of multiple binding sites

The binding of a second molecule within the active site may result in steric hindrance or a conformational change that leads to a change in the binding affinity or the rate of turnover of the first substrate, thereby either activating or partially or fully inhibiting substrate turnover. Two molecules bound in two distinct positions, both with access to the reactive oxygen (Figure 7.1 A) or able to attain a productive orientation (Figure 7.1 B) with distinct binding constants and oxidation rates, may compete for the reactive oxygen when bound simultaneously leading to partial substrate inhibition [61, 518]. Furthermore, different binding orientations/sites can also yield different products, as seen for the oxidation of triazolam to both 1'-hydroxytriazolam and 4'-hydroxytriazolam by CYP3A4 [518].

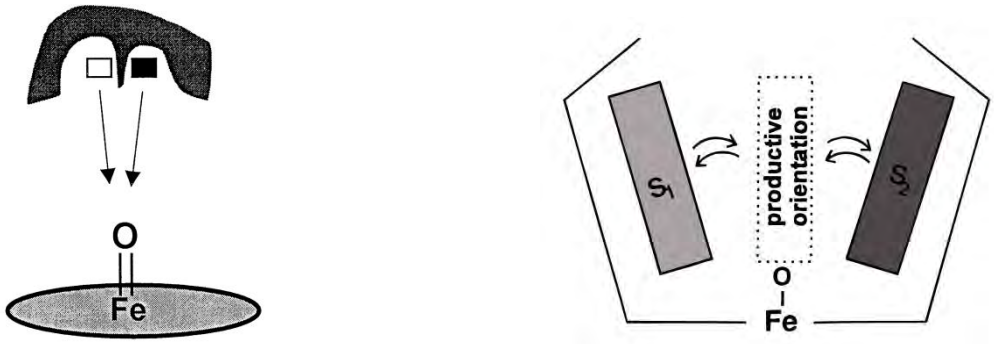


Figure 7.1 A) Two substrates bound at separate sites, competing for the reactive oxygen. Figure adapted from [518]. B) Two substrates bound at separate sites, both able to attain a productive orientation. Figure adapted from [518].

Effector binding sites have also been implicated in CYP450 kinetics, where a molecule other than the substrate binds to the enzyme, altering its activity by changing the enzyme conformation, sterically hindering the bound substrate or by altering the haem environment. This can lead to unexpected drug-drug interactions and an atypical kinetic profile.

A number of generic models have been developed to take into account the possibility of multiple substrate binding sites, as well as the presence of effector sites [519-521]. While these models account for many possible scenarios, they still make assumptions about the nature by which the binding sites interact and, even so, lead to complex kinetic equations with many unknown variables making the calculation of kinetic parameters by curve fitting very difficult if not impossible for many data sets.

Kinetically distinct CYP450 conformers

As discussed in more detail previously in Chapter 6, multiple conformations of the enzyme can exist in solution. Instead of two binding sites with different kinetic properties accounting for the atypical kinetic profiles, it is plausible that two enzyme conformers, each with different binding affinities and rate constants, account for the atypical behaviour observed. Although considering more than one enzyme conformer further complicates the kinetic models, Atkin *et al* argue that based on the biophysical data on conformational relaxation kinetics and the dynamic nature of CYP450 proteins, the simplest working kinetic model should include more than one P450 conformer [62].

CYP450-CYP450 interactions

While CYP450s were originally thought to be monomeric proteins, there is now evidence that CYP450s form homomeric and heteromeric complexes in the endoplasmic reticulum and that these interaction can modulate CYP450 function [112]. Both full length and N-terminal truncated recombinant CYP450s have been shown to form aggregates in solution to varying degrees, through a combination of hydrophobic and electrostatic interactions [112, 470, 522-525]. CYP450 complex formation is likely to alter the enzyme conformation and access to binding sites and may be the reason for the presence of non-equilibrating CYP450 conformers [60]. Figure 7.2, taken from the review by Davydov and Halpert *et al*, illustrates how CYP450 cooperativity may combine multiple substrate binding sites, effector modulated conformational changes and CYP450-CYP450 interactions.

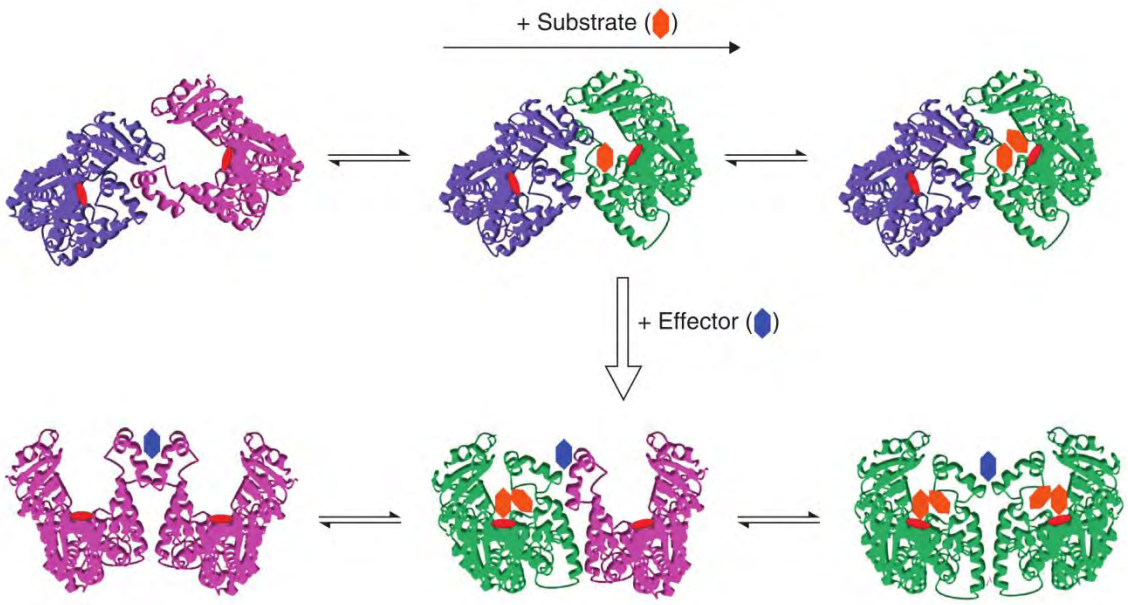


Figure 7.2 A model of CYP450 cooperativity based on the combination of multiple substrate binding, effector modulated conformational changes and CYP450-CYP450 interactions. The CYP450 oligomer consists of two enzymes in two different conformational states, one in a functional state able to bind substrate and one in a non-functional state. The binding of an effecter molecule causes a rearrangement of the P450 oligomer converting the non-functional enzyme into a functional state. Figure from [60].

Atypical kinetic curves

The allosteric effects discussed above result in substrate concentration *versus* velocity curves that deviate from the typical hyperbolic curve described by the Michaelis-Menten model. The most common atypical curves and corresponding Eadie- Hofstee plots are shown in Figure 7.3 [63]. A biphasic curve indicates the presence of a second low affinity binding site or enzyme conformer. At low substrate concentrations, this curve appears hyperbolic but at high substrate concentration the curve increases linearly. A sigmoidal curve is typically a result of autoactivation: a substrate activates its own metabolism, either by binding to a second site within the active site or by binding to a separate allosteric site. Substrate inhibition or partial substrate inhibition, occurring when the velocity decreases at higher substrate concentrations, results in a

convex shaped substrate concentration *versus* velocity curve and a Eadie-Hofstee plot with a hook in the upper quadrant.

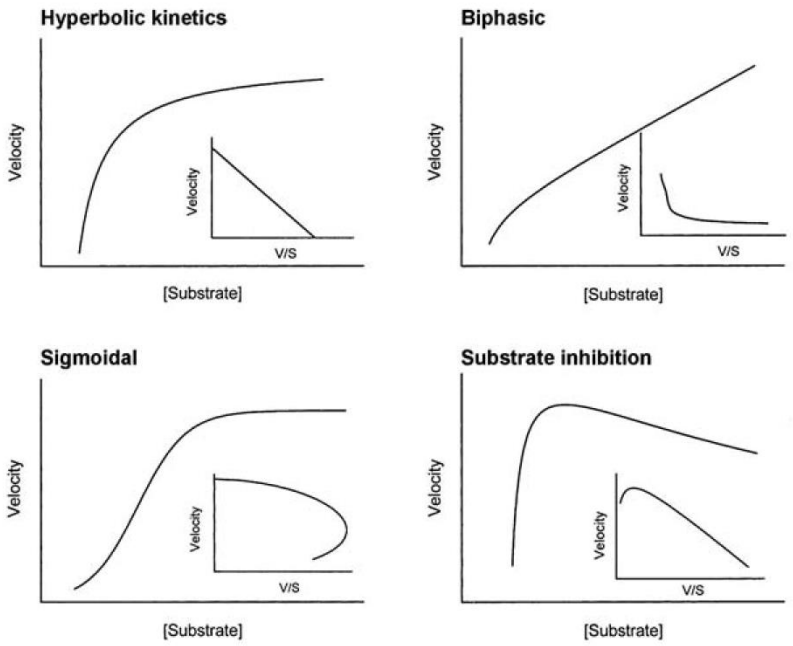


Figure 7.3 [Substrate] versus velocity graphs and their corresponding Eadie-Hofstee plots characteristic of hyperbolic, biphasic, sigmoidal and substrate inhibition kinetics. Figure from [63].

There are numerous examples of CYP3A4 and CYP2C9 substrates showing each of these atypical kinetic profiles *in vitro*: CYP3A4 metabolism of aflatoxin B1 (sigmoidal autoactivation) [526], naphthalene (biphasic) [61] and testosterone (substrate inhibition) [513]; CYP2C9 metabolism of dapsone (sigmoidal autoactivation) [61], naproxen (biphasic) [61] and celecoxib (substrate inhibition) [513]. Many reports on atypical kinetics are however incongruent, suggesting that differences in experimental setup, including enzyme source and the presence or absence of cytochrome B5, can greatly affect the kinetic profile of a drug [511] e.g. piroxicam 5′-hydroxylation shows substrate inhibition in purified baculovirus-CYP2C9 and liver microsomes but not in expressed CYP2C9 microsomes.

Effects of SNPs on CYP450 kinetics

Single amino acid substitutions can affect the K_m and V_{max} of CYP450s in a substrate dependent manner, as discussed for polymorphic variants in Chapter 4 (examples of the effects of polymorphic variation on kinetic parameters from the literature can be found in Appendix B

Tables B1 - B9). There are also a number of studies showing that amino acid changes in CYP450s can affect the kinetic profile of a substrate [527].

Aims and Objectives

The aim of this Chapter was to test the kinetic activity of the panel of CYP3A4 and CYP2C9 variant proteins in simple fluorescent-based assays, making use of the P450 shunt-pathway and thereby circumventing the need for membrane bound CYP450s or the reconstitution of CPR mediated CYP450 activity.

The objectives were as follows:

1. Test the activity of CYP3A4 and CYP2C9 protein samples using fluorescent substrates.
2. Test the effect of glycerol on CYP3A4 and CYP2C9 kinetics.
3. Determine suitable kinetic models to calculate the kinetic parameters for CYP3A4 and CYP2C9 turnover of fluorescent substrates.
4. Determine the kinetic parameters for the panel of CYP3A4 and CYP2C9 variants to establish the effect of individual mutations on enzyme activity.

7.2 Results

7.2.1 Testing CYP450 catalytic activity using fluorescent substrates

Fluorescent based activity assays were used to test the catalytic activity of recombinant CYP3A4 and CYP2C9 proteins previously expressed in *E.coli* and partially purified by His-tag purification (Chapter 5). These N-terminal truncated proteins are referred to as 'solution-phase CYP450s' in this Chapter to distinguish them from membrane bound baculosomal CYP450s.

Commercially available CYP450 Vivid® substrates can be oxidised at either one of two potential oxidation sites, yielding a highly fluorescent product and thus enabling the direct monitoring of substrate turnover (Figure 7.4). The fluorescein- and resorufin-based substrates used to test CYP3A4 and CYP2C9 activity are shown in Figure 7.5 together with their fluorescent products.

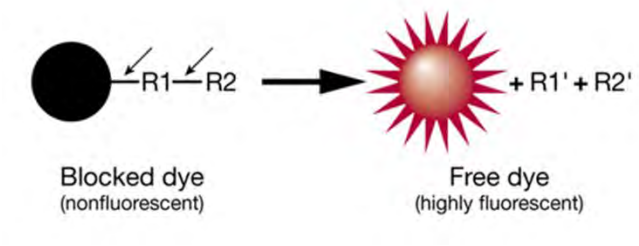


Figure 7.4 Schematic showing the cleavage of a Vivid substrate leading to a highly fluorescent product. (figure from www.b2b.invitrogen.com)

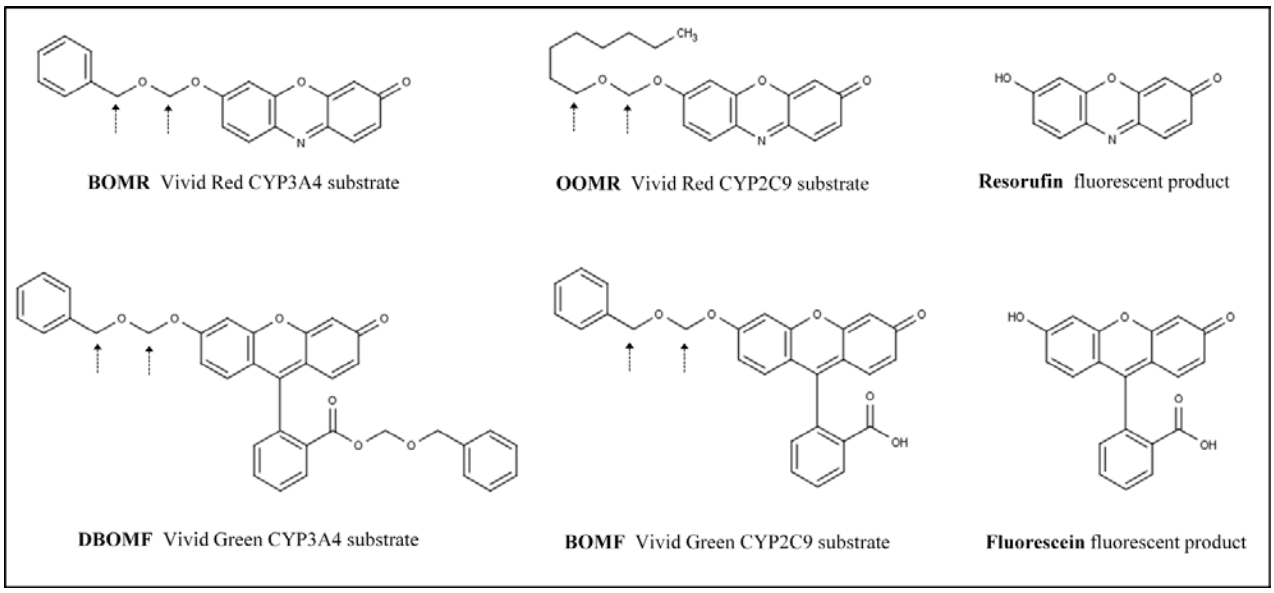
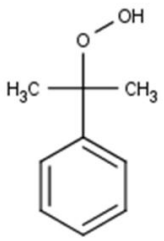


Figure 7.5 Structures of resorufin- and fluorescein-based CYP3A4 and CYP2C9 Vivid substrates and their fluorescent products. Sites of potential oxidation are indicated by an arrow.

To avoid the difficulties associated with reconstituting CYP450 activity from recombinant CPR and CYP3A4 proteins in solution, cumene hydroperoxide (CuOOH) (Figure 7.6) was used to mediate CYP3A4 activity *via* the shunt pathway, circumventing the need for an electron donor.



Cumene hydroperoxide (CuOOH)

Figure 7.6 Structure of cumene hydroperoxide used to mediate CYP450 activity via the shunt pathway.

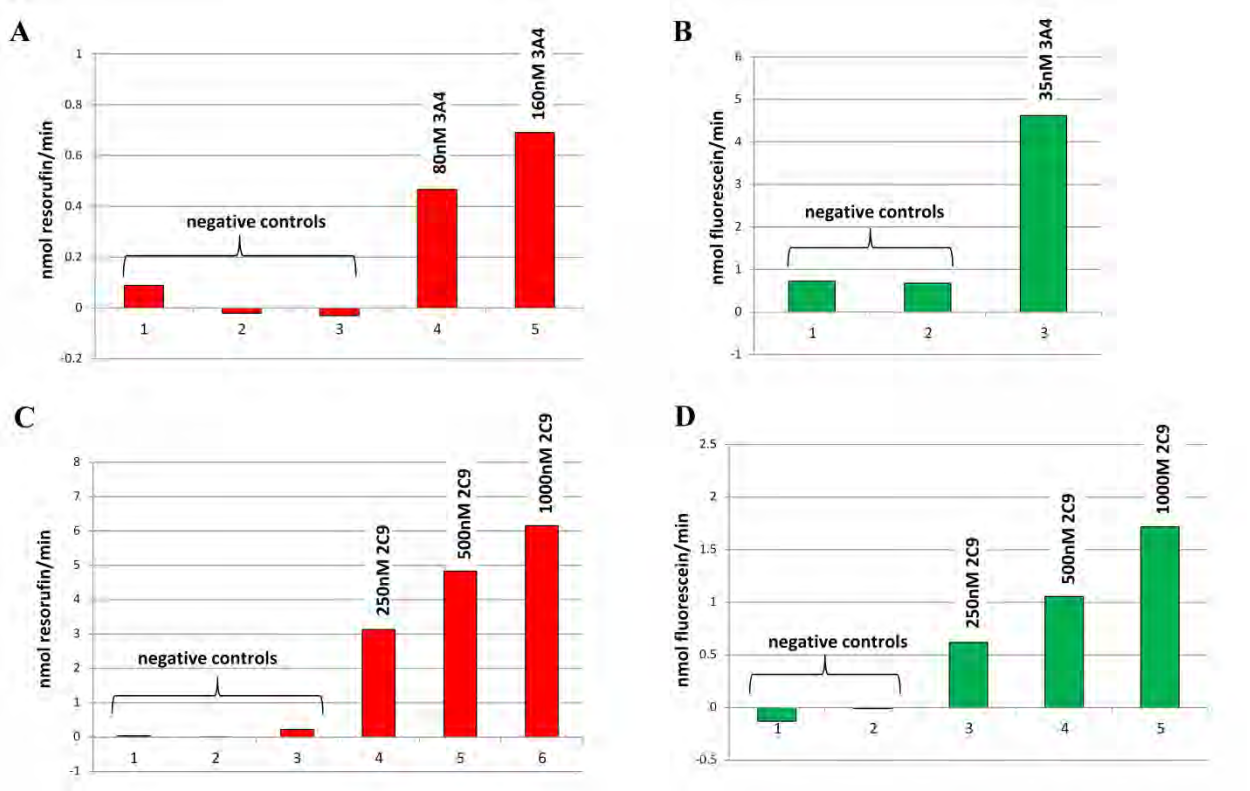


Figure 7.7 Testing Vivid Red substrate turnover by solution-phase CYP3A4 and CYP2C9.
100µl reactions consisted of 4µM Vivid substrate, varying concentrations of P450 (based on P450 concentration in His-tagged enriched samples measured by CO binding spectral assays) and 0.5mM CuOOH in 100mM potassium phosphate buffer, pH 7.4. The change in fluorescent units (RFU) was measured over time for the linear range of the reaction after the addition of CuOOH. **A)** Turnover of BOMR by CYP3A4. Negative controls consisted of 1) substrate and CuOOH, no 3A4; 2) CuOOH and 80nM 3A4, no substrate; 3) substrate and 80nM 3A4, no CuOOH. **B)** Turnover of DBOMF by CYP3A4. Negative controls consisted of 1) substrate and CuOOH, no 3A4; 2) substrate and 35nM 3A4, no CuOOH. **C)** Turnover of OOMR by CYP2C9. Negative controls consisted of 1) substrate and CuOOH, no 3A4; 2) 500nM 2C9 and CuOOH, no substrate; 3) substrate and 500nM 2C9, no CuOOH. **D)** Turnover of BOMF by CYP2C9. Negative controls consisted of 1) substrate and CuOOH, no 2C9; 2) 500nM 2C9 and CuOOH, no substrate; 3) substrate and 500nM 2C9, no CuOOH.

Various concentrations of solution-phase P450 were tested for each substrate and negative controls were included to ensure activity was P450 dependent (Figure 7.7). The following negative controls were included: 1) substrate and CuOOH, no protein; 2) P450 and CuOOH, no substrate and 3) P450 and substrate, no CuOOH. While low levels of substrate turnover (< 16% of activity) was seen for some of the controls, particularly CYP3A4 substrate DBOMF, there was clear CYP450-depend substrate turnover for all four substrates. CYP3A4 showed more than 20 fold higher CYP450-dependent turnover per nmol P450 for the fluorescein-based substrate DBOMF than for the resorufin-based substrate BOMR. CYP2C9 showed more than 4 fold higher CYP450-dependent turnover per nmol P450 for the resorufin-based substrate OOMR than for the fluorescein-based substrate BOMF.

7.2.2 CYP3A4 kinetics

CYP3A4 kinetic assays were carried out for both BOMR and DBOMF substrates using recombinant solution-phase CYP3A4 wild-type protein samples and CuOOH. Atypical CYP3A4 kinetic profiles were observed for both substrates (Figure 7.8). The presence of 15% glycerol in the reaction mixture altered the kinetic profile and increased the activity at all substrate concentrations for both substrates. Note that the scales in Figure 7.7 and Figure 7.8 are different because Figure 7.7 shows the total velocity of the reaction whereas Figure 7.8 shows velocity per nmol P450.

The presence of glycerol results in a more complex kinetic profile for the BOMR substrate, showing a decrease in velocity between 5 and 8 μM followed by a more gradual increase with increasing substrate concentrations (Figure 7.8 A). This kinetic profile closely resembles the non-hyperbolic profile observed previously observed in our laboratory for NADPH-CPR mediated turnover of BOMR by commercial baculosomal CYP3A4 proteins in the absence of glycerol (Figure 7.9) [473].

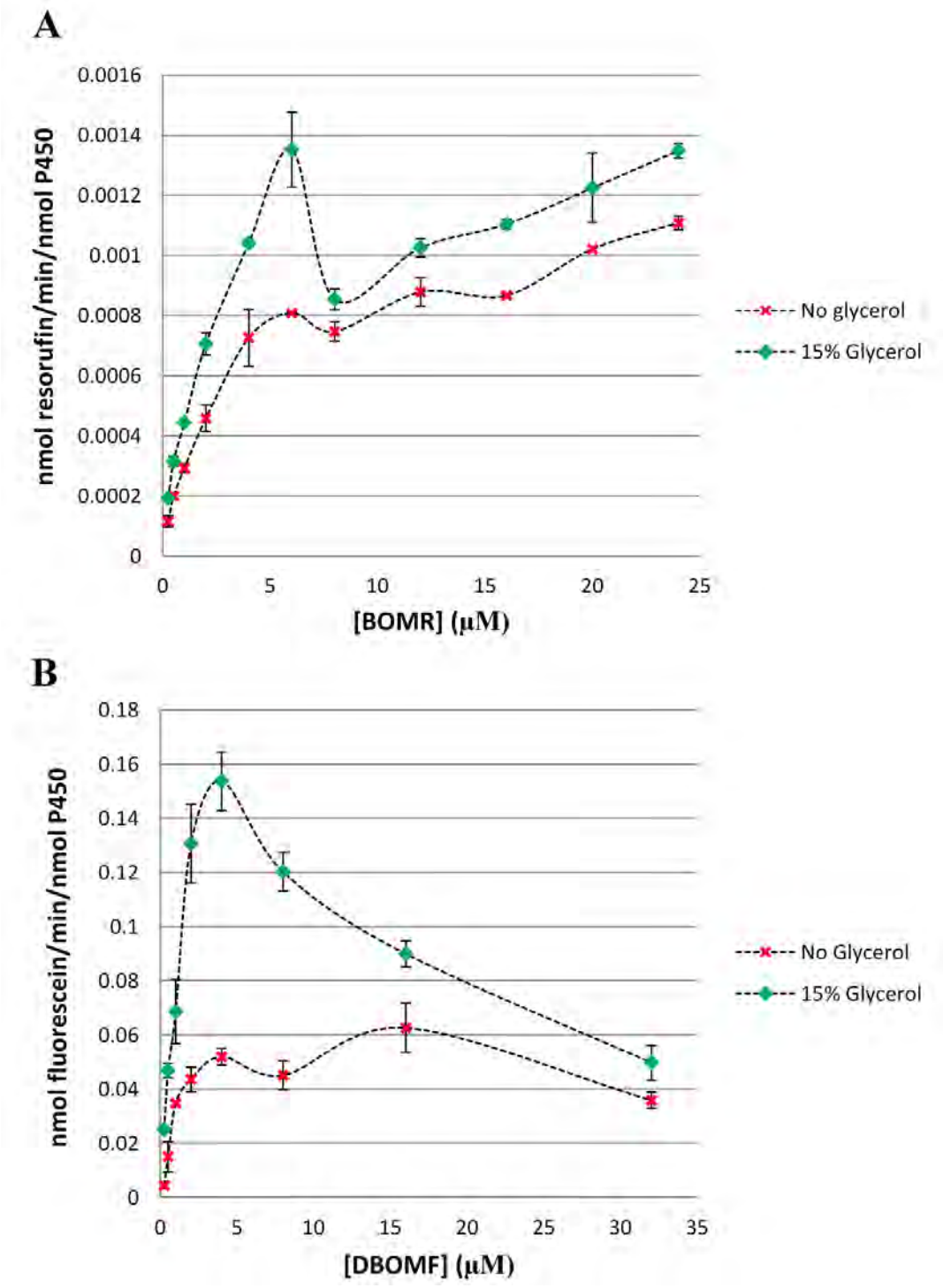


Figure 7.8 Solution-phase CYP3A4 kinetics for the turnover of Vivid substrates in the presence and absence of 15% glycerol.

Reactions were carried out in 100mM potassium phosphate buffer with and without 15% (v/v) glycerol and 0.5mM CuOOH was used to initiate the reaction, mediating CYP3A4 activity *via* the shunt pathway. The initial velocity (nmol fluorescent product formed/min/nmol P450) was plotted for each substrate concentration. Error bars represent standard deviations of 2 replicas. **A)** CYP3A4 kinetics for the turnover of BOMR Vivid red substrate (0.25 - 24 µM). A final concentration of 200nM CYP3A4 was used in each reaction. **B)** CYP3A4 kinetics for the turnover of DBOMF Vivid green substrate (0.25 - 32 µM). A final concentration of 35nM CYP3A4 was used in each reaction.

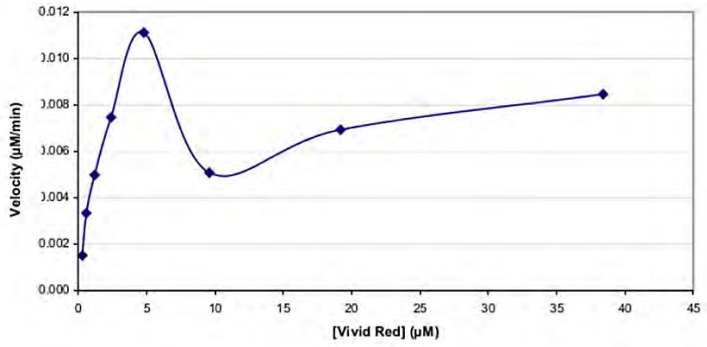


Figure 7.9 Baculosomal CYP3A4 turnover of BOMR Vivid Red substrate displaying a non-hyperbolic kinetic profile. (figure from [473])

In contrast, the presence of glycerol simplified the kinetics for the DBOMF substrate, resulting in a profile that resembles substrate inhibition (Figure 7.8 B). The profile in the absence of glycerol is complex, showing two different peaks. This kinetic assay was repeated with more data points to confirm the presence of two peaks and to ensure that it was reproducible (Figure 7.10). Once again two distinct peaks were visible at 4 µM and 16 µM DBOMF confirming the atypical kinetics observed in Figure 7.8 B.

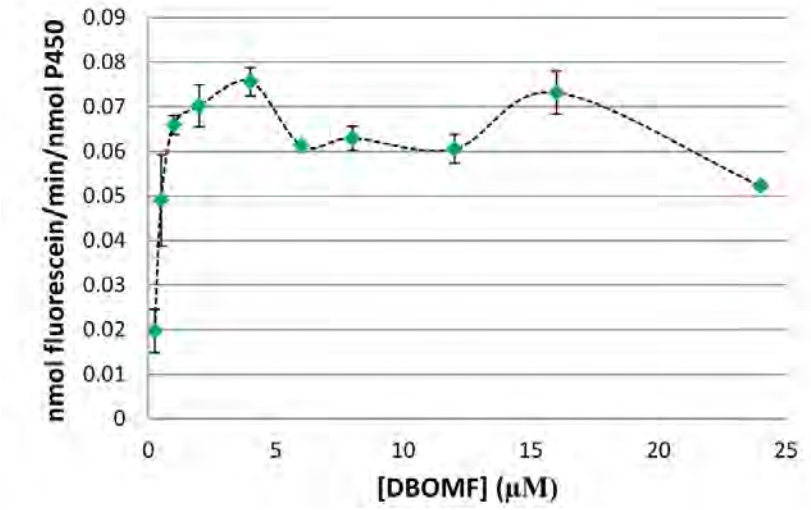


Figure 7.10 Solution-phase CYP3A4 kinetics for the turnover of DBOMF Vivid Green substrate in the absence of glycerol, repeated with addition data points. A final concentration of 35nM P450 in 100mM potassium phosphate was used for each reaction. 0.5mM CuOOH was used to mediate CYP3A4 activity *via* the shunt pathway and the initial velocity (nmol resorufin formed/min/nmol P450) was plotted for each substrate concentration. Error bars represent standard deviations of 2 replicas.

NADPH-CPR mediated baculosomal CYP3A4 kinetics for the turnover of DBOMF was carried out in the presence and absence of 15% glycerol for comparison (Figure 7.11). In contrast to solution-phase CYP3A4 wild-type protein, 15% glycerol did not appear to alter the kinetic profile of baculosomal CYP3A4 and decreased the velocity at all DBOMF substrate concentrations. CuOOH mediated baculosomal CYP3A4 kinetics in the absence of glycerol was also carried out; the kinetic profile and rate of substrate turnover was comparable to NADPH-CPR mediated kinetics in the absence of glycerol (data not shown).

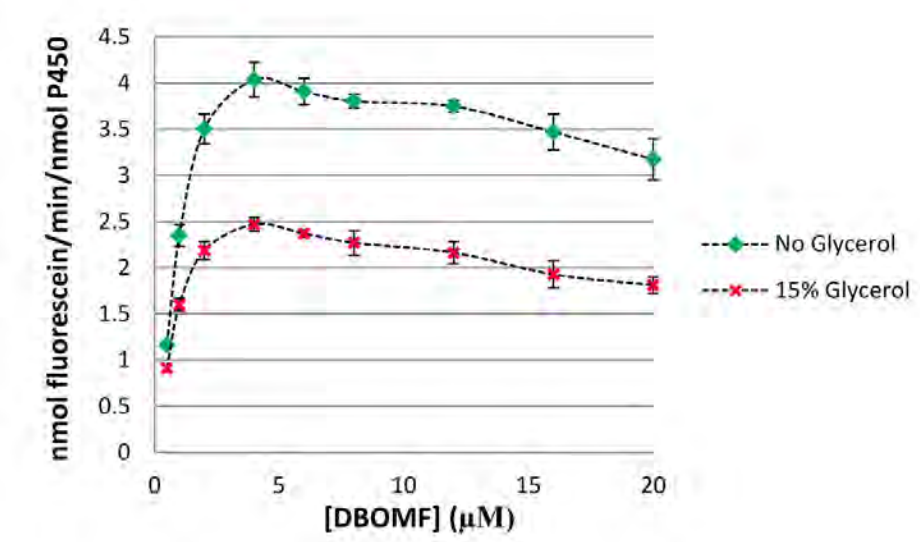


Figure 7.11 Baculosomal CYP3A4 kinetics for DBOMF (0.5-20 μM) in the presence and absence of 15% glycerol. The CYP3A4 baculosomal reagent also contained the reductase partner CPR, however the concentration is not known. A final concentration of 2nM CYP3A4 and 1mM NADPH in 100mM potassium phosphate buffer was used in each reaction. Error bars represent standard deviation of 3 replicas.

Fitting kinetic models to CYP3A4 data sets to determine kinetic parameters

While the kinetic profiles for CYP3A4 turnover of BOMR and DBOMF substrates in the presence and absence of glycerol resemble Michaelis-Menten kinetics in the lower substrate concentration range, more complex models are required to describe the kinetics observed at higher substrate concentrations. Figure 7.12 shows a schematic of the Michaelis-Menten model and Figure 7.13 shows this model fitted to truncated CYP3A4 kinetic data sets (lower substrate

concentration ranges only). Table 7.1 gives kinetic constants, K_m and V_{max} (= k_{cat}) derived from the Michaelis-Menten model, as well as R^2 values for curve fitting.

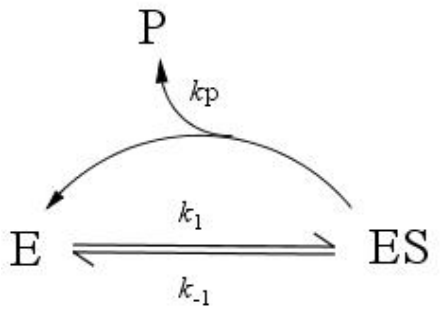


Figure 7.12 Schematic of Michaelis-Menten kinetic model where E is an enzyme molecule, S is a substrate molecule, P is the product and k_p is the rate constant.

The V_{max} for baculosomal CYP3A4 was much higher than for solution-phase CYP3A4 for both substrates (over 600 fold higher for BOMR and nearly 80 fold higher for DBOMF when comparing the kinetic constants in the absence of glycerol). The presence of 15% glycerol increased the V_{max} by 2- and 3-fold for solution-phase CYP3A4 turnover of BOMR and DBOMF respectively but decreased the V_{max} by 2-fold for baculosomal CYP3A4 turnover of DBOMF. The K_m values were similar for solution-phase and baculosomal CYP3A4 samples and the presence of 15% glycerol did not have a significant effect on K_m for either substrate.

Table 7.1 A comparison of kinetics constants for truncated CYP3A4 kinetic data sets in the presence and absence of 15% glycerol calculated using the Michaelis-Menten kinetic model.

Michaelis-Menten Kinetics	V_{max} (nmol/min/nmol P450)	K_m (μM)	R^2
BOMR Vivid Red			
Solution-phase CYP3A4 - No Glycerol	0.0056 ± 0.0005	3.1 ± 0.5	0.979
Solution-phase CYP3A4- 15% Glycerol	0.0096 ± 0.0009	3.8 ± 0.7	0.978
Baculosomal CYP3A4 – no glycerol*	3.6	3.2	
DBOMF Vivid Green			
Solution-phase CYP3A4 - No Glycerol	0.075 ± 0.009	1.6 ± 0.4	0.929
Solution-phase CYP3A4 - 15% Glycerol	0.24 ± 0.02	2.1 ± 0.4	0.959
Baculosomal CYP3A4 – No glycerol	5.8 ± 0.4	1.5 ± 0.2	0.961
Baculosomal CYP3A4 – 15% Glycerol	3.3 ± 0.1	1.2 ± 0.1	0.980

± standard error. * Data for 'Baculosomal CYP3A4 – no glycerol' is from [473].

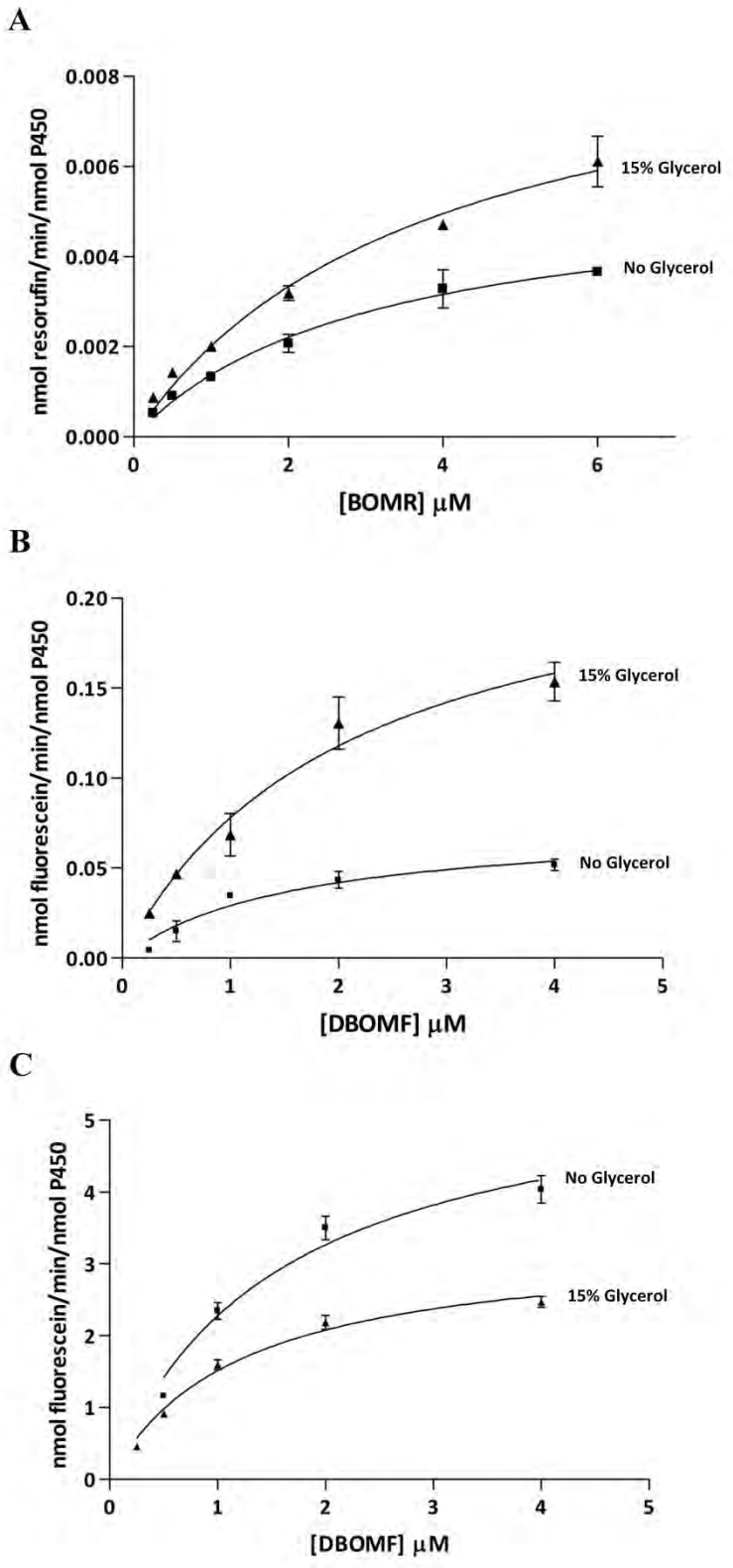


Figure 7.13 Michaelis-Menten model fitted to the truncated CYP3A4 kinetic data sets for the turnover of Vivid substrates in the presence and absence of 15% glycerol using non-linear regression.
A) Solution-phase CYP3A4 BOMR Vivid Red and (B) DBOMF Vivid Green kinetics, error bars represent standard deviations of 2 replicas. (C) Baculosomal CYP3A4 DBOMF Vivid Green, error bars represent standard deviations of 3 replicas.

The next step was to find a kinetic model that could adequately explain the full kinetic data sets. This was however only possible for the turnover of BOMR in the absence of glycerol and the turnover of DBOMF in the presence of 15% glycerol as the other profiles were to complex.

Biphasic CYP3A4 turnover of BOMR in the absence of glycerol

The multisite biphasic kinetics model shown in Figure 7.14 has previously been used to describe CYP3A4 kinetic data and was fitted to the full data set for CYP3A4 turnover of BOMR in the absence of glycerol; the Michaelis-Menten model was used for comparison (Figure 7.15). It was difficult to determine the correct model from the Eadie-Hofstee plot of the data (Figure 7.15 B), however a comparison of the two fits in Prism using the F-test showed that the biphasic model was the preferred model for this data set.

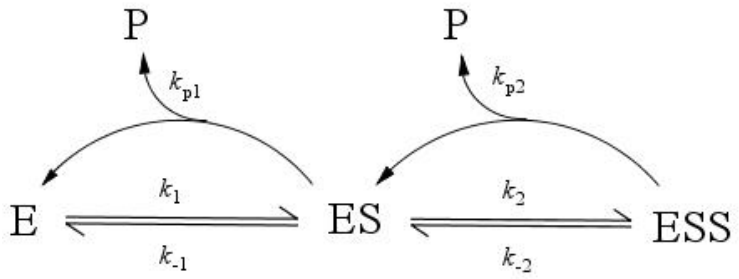


Figure 7.14 Scheme of biphasic kinetic model where E is an enzyme molecule, S is a substrate molecule and P is the product. The different rate constants are represented by *k*.

The kinetic constants derived from each model and the R^2 for the curve fitting are shown in Table 7.2. In the biphasic model, K_{m1} and V_{max1} are observed at low substrate concentrations and result from substrate binding to the low K_m, low V_{max} binding site. The substrate turnover by the second low affinity site (high K_m), shown by the linear portion of the graph, is described by intrinsic clearance parameter (Cl_{int}), which is the ratio of V_{max2}/K_{m2}. It is difficult to estimate the actual values of V_{max} and K_m for the second binding site as saturation is not achieved and the linear part of the graph is formed by a combination of substrate turnover by both the high and low affinity site [528]. The equation for the biphasic model can be found in the Methods, Section 7.4.4.

The single site Michaelis-Menten model gives a higher K_m and V_{max} than the multisite biphasic model for the high affinity binding site, however the student t-test showed that the difference

was not significant. The biphasic model converges back to the Michaelis-Menten model if there is no second binding site or if the second binding site contributes very little to substrate turnover.

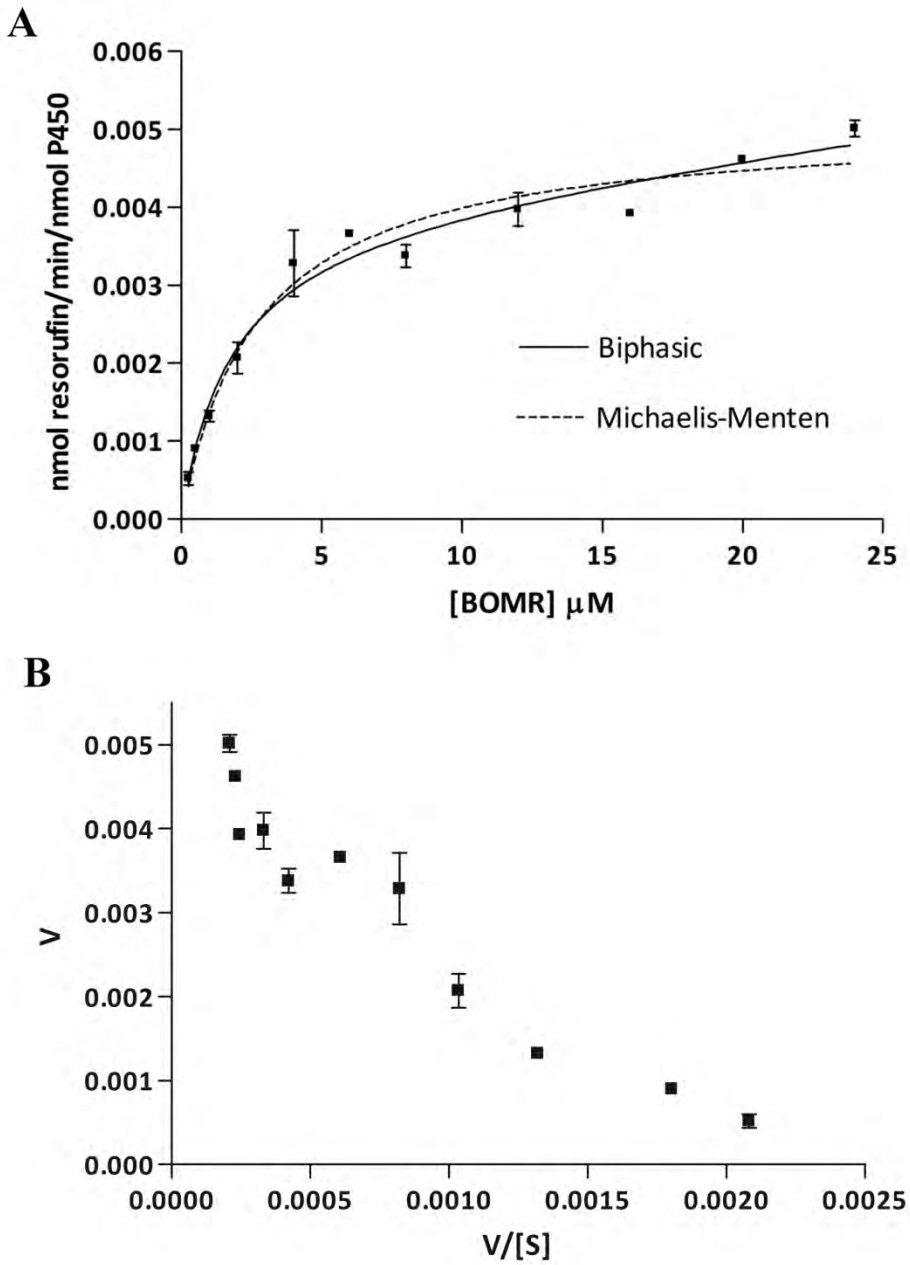


Figure 7.15 Graphs fitted to solution-phase CYP3A4 kinetic data for the turnover of BOMR Vivid Red in the absence of glycerol using non-linear regression. A) The biphasic model is shown by the solid line and the Michealis-Menten model is shown by the dotted line. A comparison of the two fits in Prism using the F-test showed that the biphasic model was the preferred model for this data set. B) Eadie-Hofstee plot of data showing V (nmol resorufin.min^{-1}.nmol P450^{-1}) *versus* V/[S] (nmol resorufin.min^{-1}.nmol P450^{-1}.µM BOMR^{-1}). Error bars represent standard deviations of 2 replicas.

Table 7.2 A comparison of kinetic constants calculated using Michaelis-Menten and biphasic kinetic models for CYP3A4 turnover of BOMR in the absence of glycerol

Michaelis-Menten Model		**Biphasic Model**	
V_{max}	0.0051 ± 0.0002 nmol/min/nmol P450	V_{max1}	0.0040 ± 0.0004 nmol/min/nmol P450
K_m **(μM)**	2.7 ± 0.4 μM	K_{m1} **(μM)**	1.8 ± 0.4 μM
R^2	0.963	CL_{int} (V_{max2}/K_{m2})	0.00005 ± 0.00002 $min^{-1}\mu M^{-1}$
		R^2	0.971

± standard error. A comparison of the two fits, in Prism using the F-test, showed that the biphasic model was the preferred model for this data set. The null hypothesis that the Michaelis-Menten kinetic model was the best fit was rejected, P value = 0.027.

Substrate inhibition for CYP3A4 turnover of DBOMF in the presence of 15% glycerol

The rate of DBOMF substrate turnover by solution-phase CYP3A4 decreased at higher substrate concentration suggesting that substrate inhibition was occurring. CYP450 substrate inhibition has been observed for a number of different substrates and various substrate inhibition models have been developed to describe the kinetic curves obtained [513, 518]. Three different models (Figure 7.16, Figure 7.17 and Figure 7.18), of varying complexity, were fitted to the kinetic data set to determine the model that best described the kinetics observed. The equations derived from these models, based on both the steady-state and rapid equilibrium assumptions that the rate of formation and decomposition of enzyme-substrate complexes are equal, can be found in the Methods, Section 7.4.4, Table 7.9.

The first model, referred to here as the 'substrate inhibition model' (Figure 7.16), is a general model commonly used to describe the substrate inhibition observed for a range of enzymes [529]. This model assumes that at high substrate concentrations a second substrate molecule binds to the enzyme substrate complex rendering it unproductive and thereby decreasing substrate turnover.

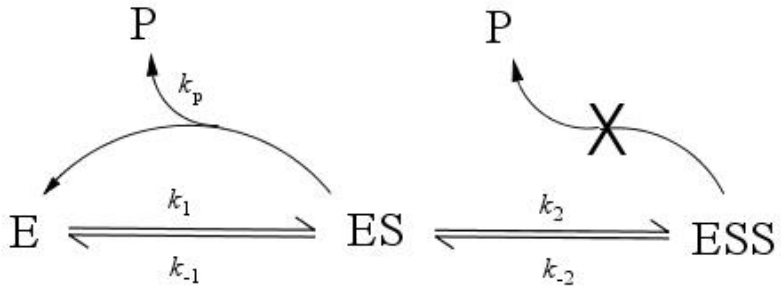


Figure 7.16 Substrate inhibition model, where E is an enzyme molecule, S is a substrate molecule and P is the product. The different rate constants are represented by *k*.

The second model, derived by Lin *et al* in 2001 and referred to here as the 'P450 substrate inhibition model 1' (scheme shown in Figure 7.17), assumes that there are two binding site simultaneously available to bind substrate - one catalytic and one inhibitory - and that the sites interact in the presence of substrate. This causes a change in the binding constants for the catalytic site (K_s) and the inhibition site (K_I) as well as the V_{max} when a second substrate is bound. α and β are the factors by which the dissociation constants and V_{max} change in the presence of a second bound substrate.

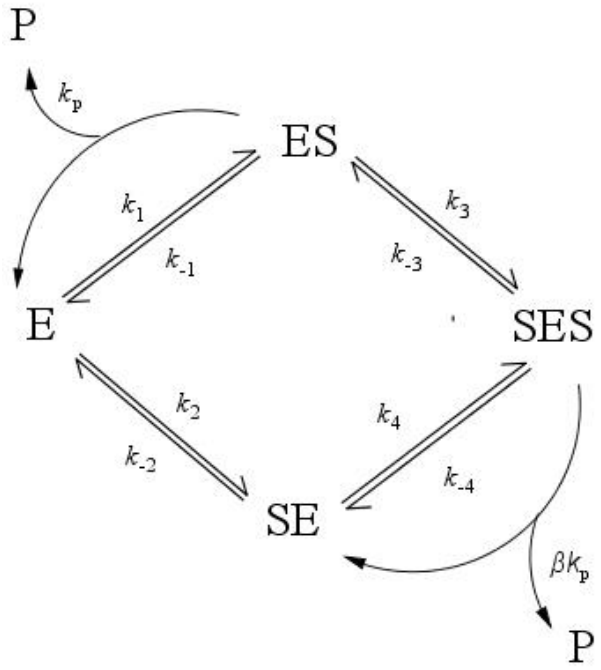


Figure 7.17 P450 substrate inhibition kinetic model 1.
E is an enzyme molecule, ES is an enzyme molecule with a substrate molecule bound to the catalytic binding site and SE is an enzyme molecule with a substrate molecule bound to the non-productive inhibitory site. SES is an enzyme with substrate molecule bound to both sites. k_p is the rate at which product, P, is formed from the productive complex ES (ES → E + P). β is the factor by which the rate k_p changes when there is also a substrate molecule bound at the inhibitory site (SES → SE + P). Dissociation constants are defined as follows $Ks = k_{-1}/k_{-1}$; $K_I = k_2/k_{-2}$; $\alpha K_I = k_3/k_{-3}$; $\alpha K_s = k_4/k_{-4}$.

The third model, derived by Schrag *et al* in 2002 and refered to here as 'P450 substrate inhibition model 2' (Figure 7.18), also has two binding sites similtaneously available to bind substrate but differs from the previous model in that the two sites do not interact, so the presence of one substrate molecule does not affect the binding constant of the second substrate molecule. Substrate inhibition results from the substrate molecules at the two different sites competing for the reactive oxygen; thus, while the binding constants are unaffected, the presence of a second

substrate changes the rate constant and hence V_{max} by a factor of α. The second site is also shown to be catalytically active in this model but results in the formation of a different product, P_2.

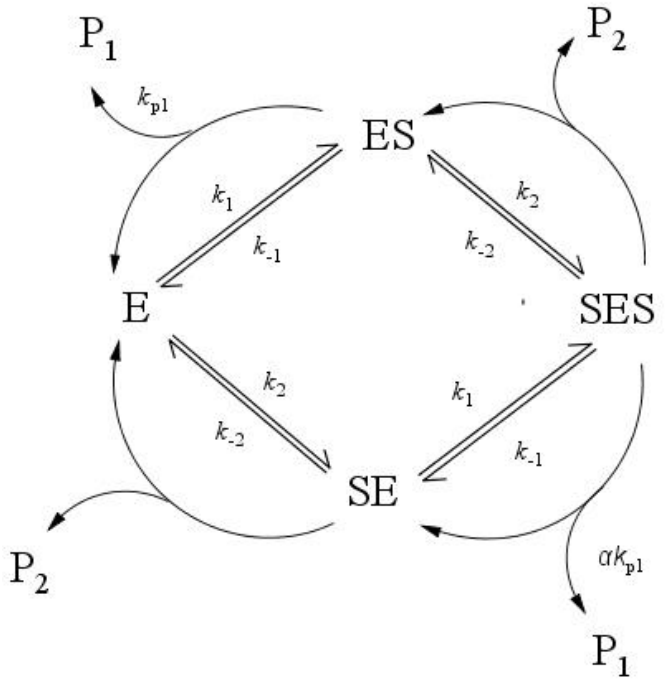


Figure 7.18 P450 substrate inhibition model 2
E is an enzyme molecule, ES is an enzyme molecule with a substrate molecule bound to the catalytic binding site 1 and SE is an enzyme molecule with a substrate molecule bound to the catalytic binding site 2. SES is an enzyme with a substrate molecule bound to both sites. k_{p1} is the rate at which the product of interest, P_1, is formed from the productive complex ES (ES → E + P). α is the factor by which the rate k_{p1} changes when there is a substrate molecule bound at the other catalytic site (SES → SE + P_1). Dissociation constants are defined as follows $K_{m1} = k_1/k_{-1}$; $K_{m2} = k_2/k_{-2}$.

Figure 7.19 shows the substrate inhibition model and the P450 substrate inhibition model fitted to the kinetic data set for the turnover of DBOMF by solution-phase CYP3A4 in the presence of 15% glycerol. Figure 7.20 shows the same models fitted to the kinetic data set for the turnover of DBOMF by baculosomal CYP3A4 in the presence and absence of 15% glycerol. P450 substrate inhibition model 2 was also fitted to the data but cannot clearly be seen in this graph as it followed the same curve as the P450 substrate inhibition model 1. The Eadie-Hofstee plots for each data set are also included, showing a curve characteristic of substrate inhibition. A comparison of the fits in prism using the F-test showed that the substrate inhibition model was the preferred model for the turnover of DBOMF by solution-phase CYP3A4 in the presence of 15 % glycerol but that the P450 substrate inhibition model was the preferred model for the turnover of DBOMF by baculosomal CYP3A4 in the presence and absence of 15 % glycerol.

Both models fit reasonably to all data sets and, while the more complex P450 substrate inhibition model is likely to describe CYP3A4 substrate inhibition more adequately, many combinations of solutions to the equation lead to curves that fit the data equally well. This is due to the high number of variables (V_{max}, K_s, K_i, α and β) in the equation and leads to an ambiguous fit in Prism where kinetic parameters cannot be determined with any confidence.

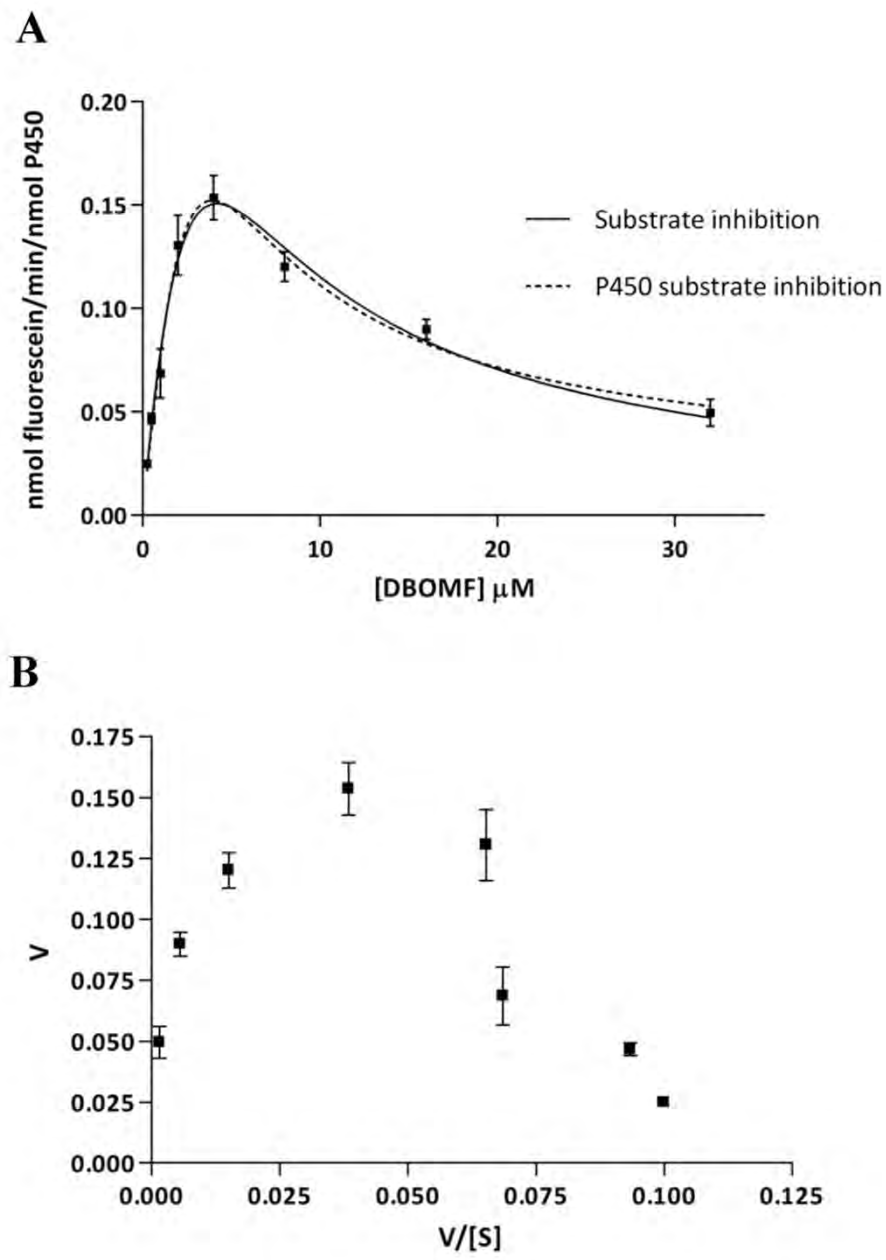


Figure 7.19 Graphs fitted to solution-phase CYP3A4 kinetic data for the turnover of DBOMF Vivid Green in the presence of 15% glycerol using non-linear regression. A) The substrate inhibition model is shown by the solid line and the P450 substrate inhibition model 1 is shown by the dotted line. A comparison of the two fits in Prism using the F-test showed that the substrate inhibition model was the preferred model for this data set. B) Eadie-Hofstee plot of data showing V (nmol fluorescein.min^{-1}.nmol P450^{-1}) *vs.* V/[S] (nmol fluorescein.min^{-1}.nmol P450^{-1}.μM BOMR^{-1}). Error bars represent standard deviations of 2 replicas.

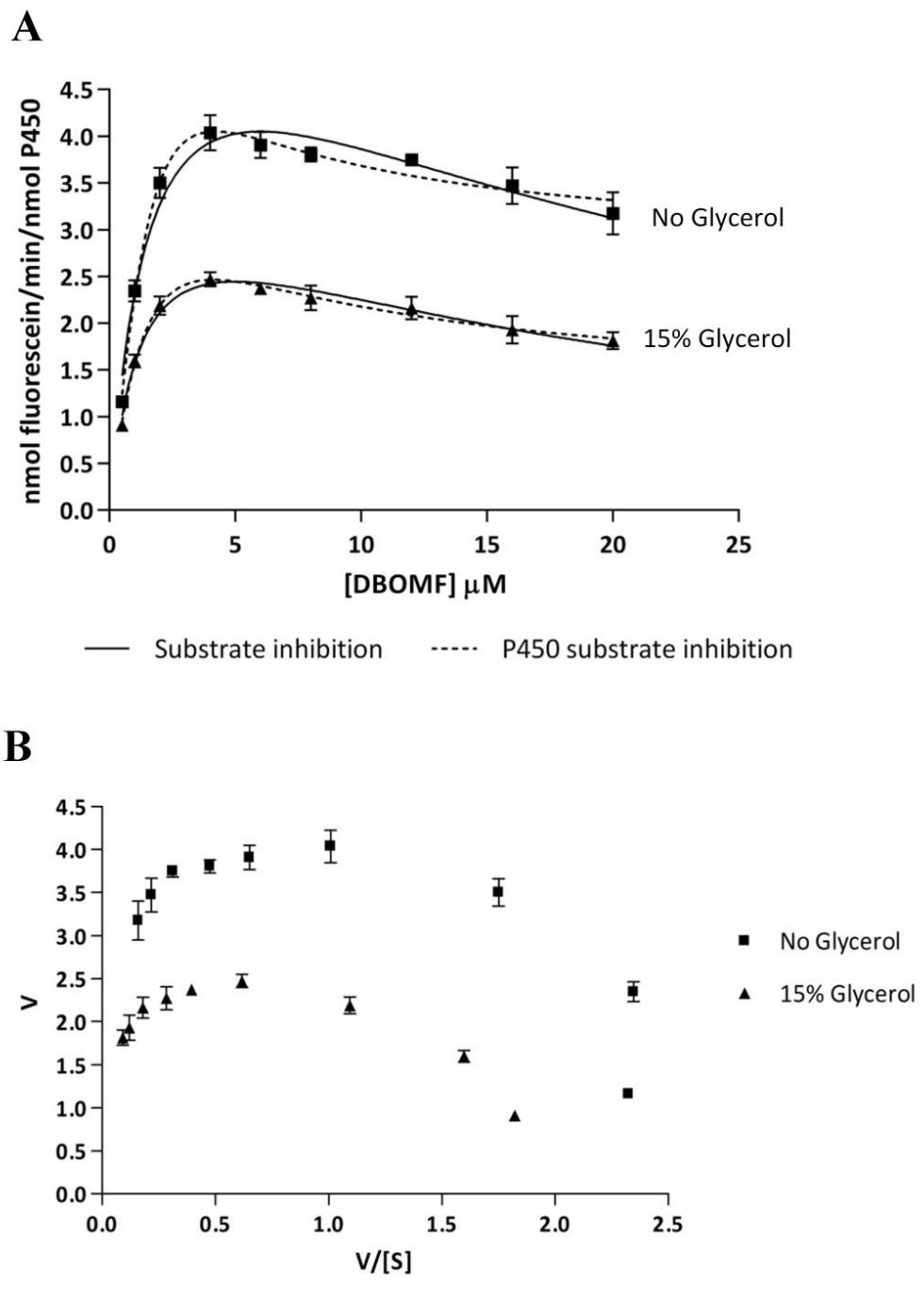


Figure 7.20 Graphs fitted to baculosomal CYP3A4 kinetic data for the turnover of DBOMF Vivid Green in the presence and absence of 15% glycerol using non-linear regression. A) The substrate inhibition model is shown by the solid line and the P450 substrate inhibition model 1 is shown by the dotted line. A comparison of the two fits in Prism using the F-test showed that the P450 substrate inhibition model 1 was the preferred model for this data set. B) Eadie-Hofstee plot of data showing V (nmol fluorescein.min^{-1}.nmol P450^{-1}) *vs.* V/[S] (nmol fluorescein.min^{-1}.nmol P450^{-1}.μM DBOMF^{-1}). Error bars represent standard deviations of 2 replicas.

The kinetic parameters obtained by fitting the substrate inhibition model to the three data sets are shown in Table 7.3. The V_{max} (nmol fluorescein formed per min per nmol P450) was significantly lower (6 fold decrease) for solution-phase CYP3A4 turnover than for baculosomal CYP3A4 turnover in the presence of 15% glycerol. The presence of 15% glycerol decreased the V_{max} for baculosomal CYP3A4 by ~ 2 fold. The K_m for solution-phase CYP3A4 was higher than for baculosomal CYP3A4, however the difference between K_m values was not statistically significant.

Table 7.3 Kinetic constants calculated using the substrate inhibition model for CYP3A4 turnover of DBOMF

Substrate Inhibition Model	V_{max} (nmol/min/nmol P450)	K_m (μM)	K_i (μM)	R^2
Solution-phase 3A4 - 15 % Glycerol	0.6 ± 0.2	6 ± 3	3 ± 1	0.954
Baculosomal 3A4 – No Glycerol	6.3 ± 0.4	1.7 ± 0.2	21 ± 4	0.952
Baculosomal 3A4 – 15% Glycerol	3.8 ± 0.2	1.3 ± 0.2	19 ± 2	0.954

A student t-test showed that V_{max} values for the three data sets were significantly different (P value < 0.05) while K_m values did not differ significantly from each other (P value > 0.05)

7.2.3 CYP2C9 kinetics

CYP2C9 kinetic assays were carried out for OOMR and BOMF substrates using solution-phase CYP2C9 wild-type protein samples and CuOOH. The CYP2C9 kinetic profiles observed for each substrate in the presence and absence of 15% glycerol are shown in Figure 7.21. The presence of glycerol in the reaction mixture increased the turnover rate of OOMR at all substrate concentrations and does not alter the shape of the kinetic profile. The presence of glycerol does not appear to affect the rate of turnover of BOMF at lower substrate concentrations but at high substrate concentrations the turnover rate is lower in the presence of glycerol.

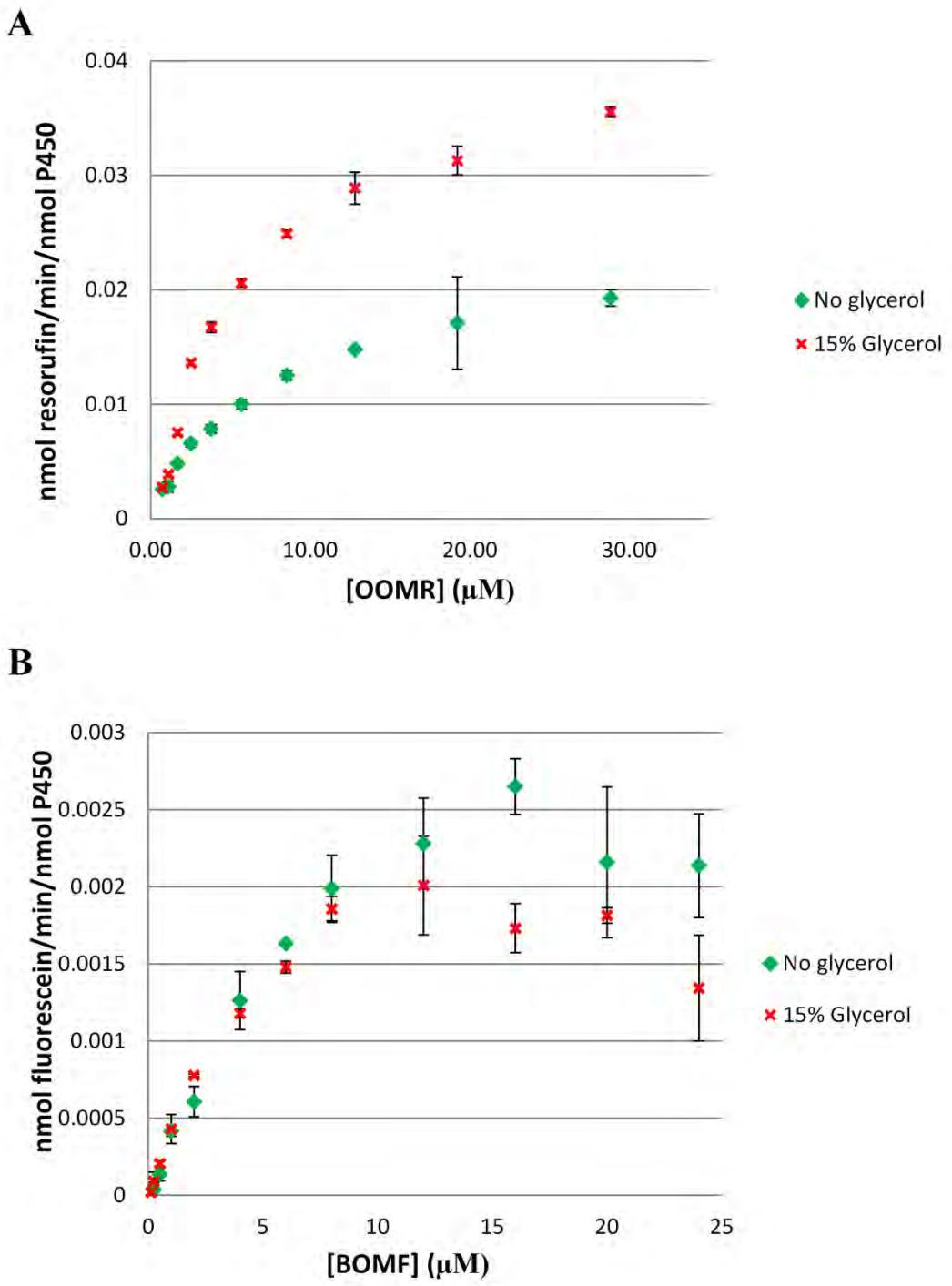


Figure 7.21 Solution-phase CYP2C9 kinetics for the turnover of Vivid substrates in the presence and absence of 15% glycerol.

Reactions were carried out in 100mM potassium phosphate buffer with and without 15% (v/v) glycerol and 0.5mM CuOOH was used to initiate the reaction, mediating CYP3A4 activity *via* the shunt pathway. The initial velocity (nmol fluorescent product formed/min/nmol P450) was plotted for each substrate concentration. A) CYP2C9 kinetics for the turnover of OOMR Vivid Red substrate (0.75 – 28.8 µM). A final concentration of 300 nM CYP3A4 was used in each reaction. Error bars represent standard deviations of 3 replicas. B) CYP2C9 kinetics for the turnover of BOMF Vivid Green substrate (0.25 - 32 µM). A final concentration of 500 nM CYP2C9 was used in each reaction. Error bars represent standard deviations of 2 replicas.

Fitting kinetic models to CYP2C9 data sets to determine kinetic parameters

Figure 7.22 shows the Michaelis-Menten model and the biphasic model fitted to solution-phase CYP2C9 kinetic data set for the turnover of OOMR substrate. A comparison of the two fits in Prism using the F-test shows that the Michaelis-Menten model is the preferred model. The Eadie-Hofstee plot is linear – indicating Michaelis-Menten kinetics – except at very low substrate concentrations. This suggests that there is only one binding site for OOMR contributing to substrate turnover. The kinetic parameters determined using the Michaelis-Menten model are shown in Table 7.4.

For solution-phase CYP2C9 turnover of BOMF, substrate inhibition appears to be occurring at higher BOMF substrate concentrations, however the substrate inhibition models discussed previously fit poorly to the full data set. For this reason the Michaelis-Menten model was fitted to the truncated data set (Figure 7.23) from which kinetic parameters were calculated (Table 7.4). V_{max} and K_m values were similar to that for the OOMR substrate. The presence of glycerol did not have a statistically significant effect on the V_{max} or K_m for the turnover of BOMF.

Table 7.4 A comparison of kinetics constants for solution-phase CYP2C9 kinetic data sets in the presence and absence of 15% glycerol calculated using the Michaelis-Menten kinetic model.

Michaelis-Menten Kinetics	**V_{max} (nmol/min/nmol P450)**	**K_m (μM)**	**R^2**
OOMR Vivid Red – full data sets			
2C9 - No Glycerol	0.024 ± 0.001	7.5 ± 0.9	0.9604
2C9 - 15% Glycerol	0.044 ± 0.001*	6.7 ± 0.5	0.9849
BOMF Vivid Green – truncated data set			
2C9 - No Glycerol	0.0043 ± 0.0005	10 ± 2	0.9707
2C9 - 15% Glycerol	0.0031 ± 0.0002	6 ± 1	0.9743

± standard error. The student t-test showed that V_{max} for the turnover of OOMR by CYP2C9 – 15% glycerol was significantly higher than CYP2C9 – no glycerol, $P < 0.005$ but there was no significant difference between the K_m for the two data sets. There is no significant difference between V_{max} and K_m parameters for the turnover of BOMF in the presence and absence of 15% glycerol.

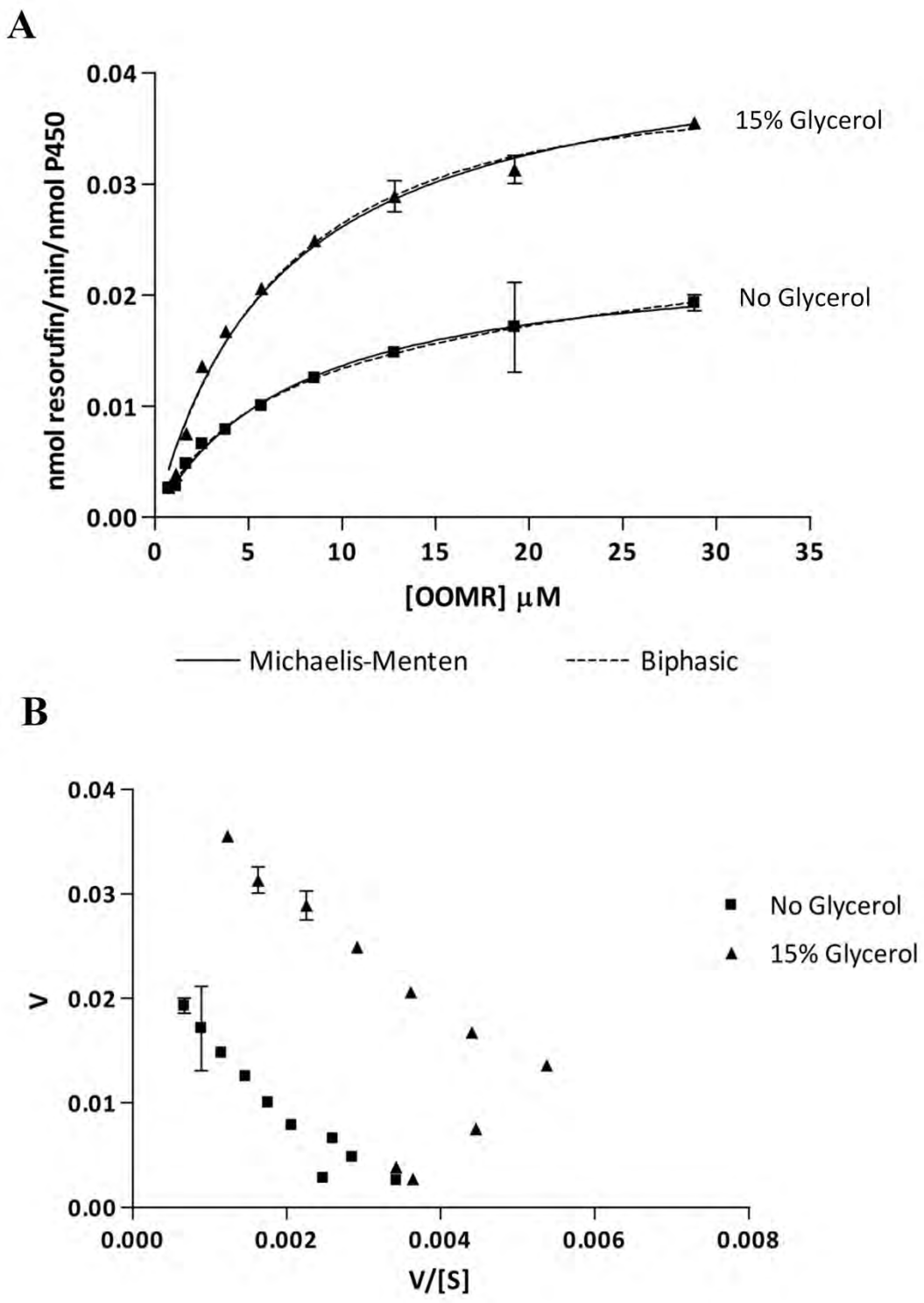


Figure 7.22 Graphs fitted to CYP2C9 kinetic data for the turnover of OOMR Vivid Red in the presence and absence of 15% glycerol using non-linear regression. A) The Michaelis-Menten model is shown by the solid line and biphasic model is shown by the dotted line. B) Eadie-Hofstee plot of data showing V (nmol resorufin.min^{-1}.nmol P450^{-1}) *vs.* V/[S] (nmol resorufin.min^{-1}.nmol P450^{-1}.µM OOMR^{-1}). Error bars represent standard deviations of 3 replicas.

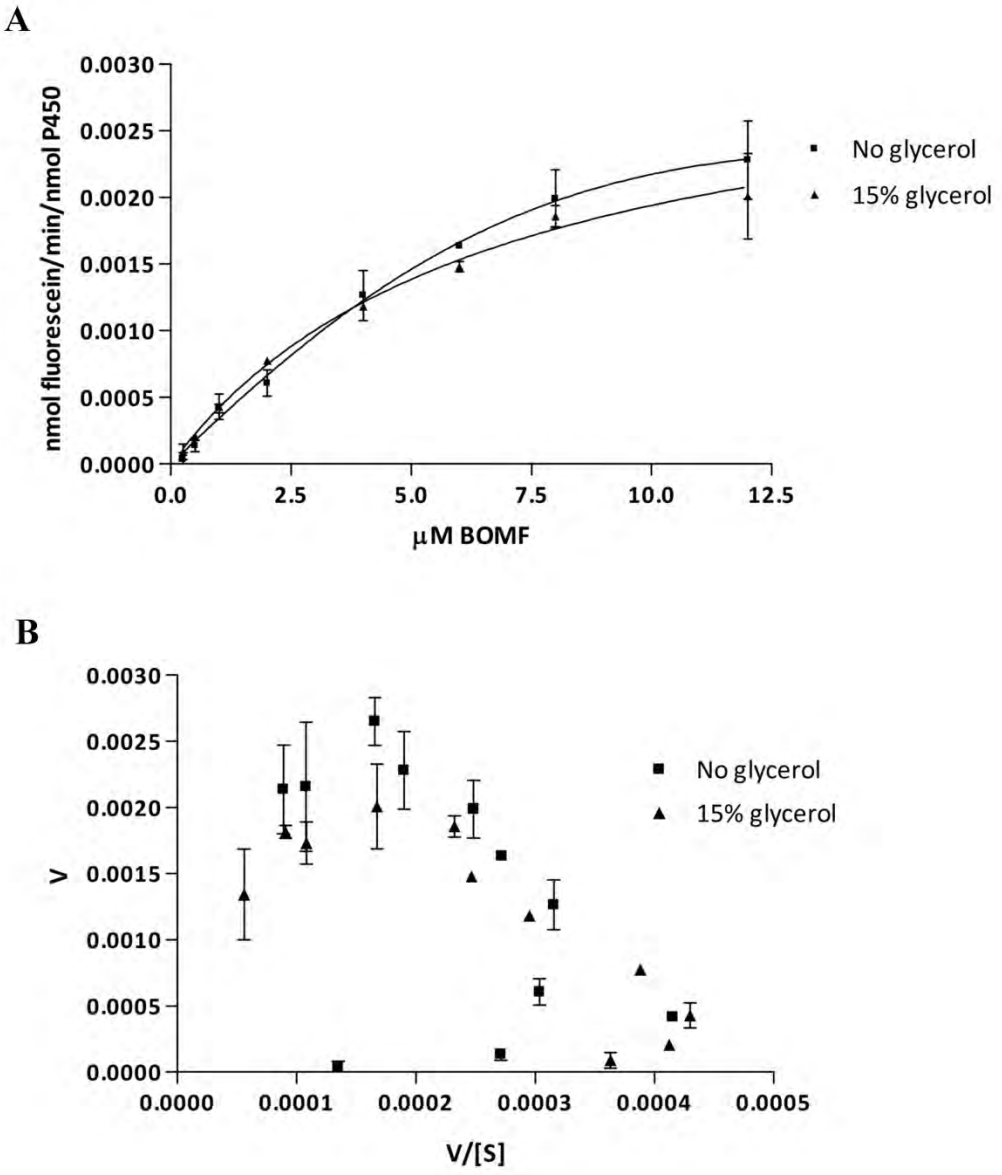


Figure 7.23 Graphs fitted to solution-phase CYP2C9 kinetic data for the turnover of BOMF Vivid Green in the presence and absence of 15% glycerol using non-linear regression. A) The Michaelis-Menten model fitted to the truncated data set. B) Eadie-Hofstee plot of the full data set showing V (nmol fluorescein.min^{-1}.nmol P450^{-1}) *vs.* V/[S] (nmol fluorescein.min^{-1}.nmol P450^{-1}.µM BOMF^{-1}). Error bars represent standard deviations of 2 replicas.

7.2.4 Activity of CYP3A4 variants

CYP3A4 kinetic assays were carried out for nine CYP3A4 variants, including the wild-type enzyme, using solution-phase CYP3A4 protein samples. The reactions were initiated with CuOOH as before. DBOMF Vivid green substrate was chosen to test the activity of CYP3A4 variants because wild-type CYP3A4 showed high turnover levels for DBOMF compared to the other Vivid substrates. 15% glycerol was included in all reactions as this gave a kinetic profile similar to that seen for baculosomal CYP3A4, increased enzyme activity and simplified the kinetic profile enabling the calculation of kinetic parameters based on the full data set using the substrate inhibition model.

Figure 7.24 A shows the Michaelis-Menten model fitted to the truncated kinetic data set (0.75 – 5.6 μM DBOMF) and Figure 7.24 B shows the substrate inhibition model fitted to the full kinetic data set (0.75 - 43.3 μM DBOMF) for each variant. The residual plots are also included to illustrate how well each model fits the data. The residual is the distance of a data point from the fitted curve, with a point above the curve having a positive residual and a point below the curve having a negative residual. If the model fits the data well the residual data points should be randomly distributed close to the x-axis with no apparent trend in their distribution above and below the axis. The residual plot in panel A shows that, even for the truncated data set, the Michaelis-Menten model does not adequately explain the data as a clear trend is visible. Although there is still a trend in the residual plot in panel B, the points are more randomly distributed suggesting the substrate inhibition model is a better fit for the data. The Eadie-Hofstee plots for the full data sets are shown in panel C**.**

The kinetic parameters V_{max}, K_m, V_{max}/K_m and K_i calculated using each model are shown in Table 7.5 together with the R^2 value for curve fitting. A visual comparison of the V_{max}, K_m, K_i and V_{max}/K_m for the different variants and models is shown in Figure 7.25.

CYP3A4 variants I223R and M445T showed a significant decrease in V_{max} compared to wild-type when calculated using either model. S222P also showed a decrease in V_{max} but the decrease was only significant for the Michaelis-Menten model fitted to the truncated data set. L373F and L293P both showed an increase in V_{max} but this increase was once again only significant for the Michaelis-Menten model fitted to the truncated data set.

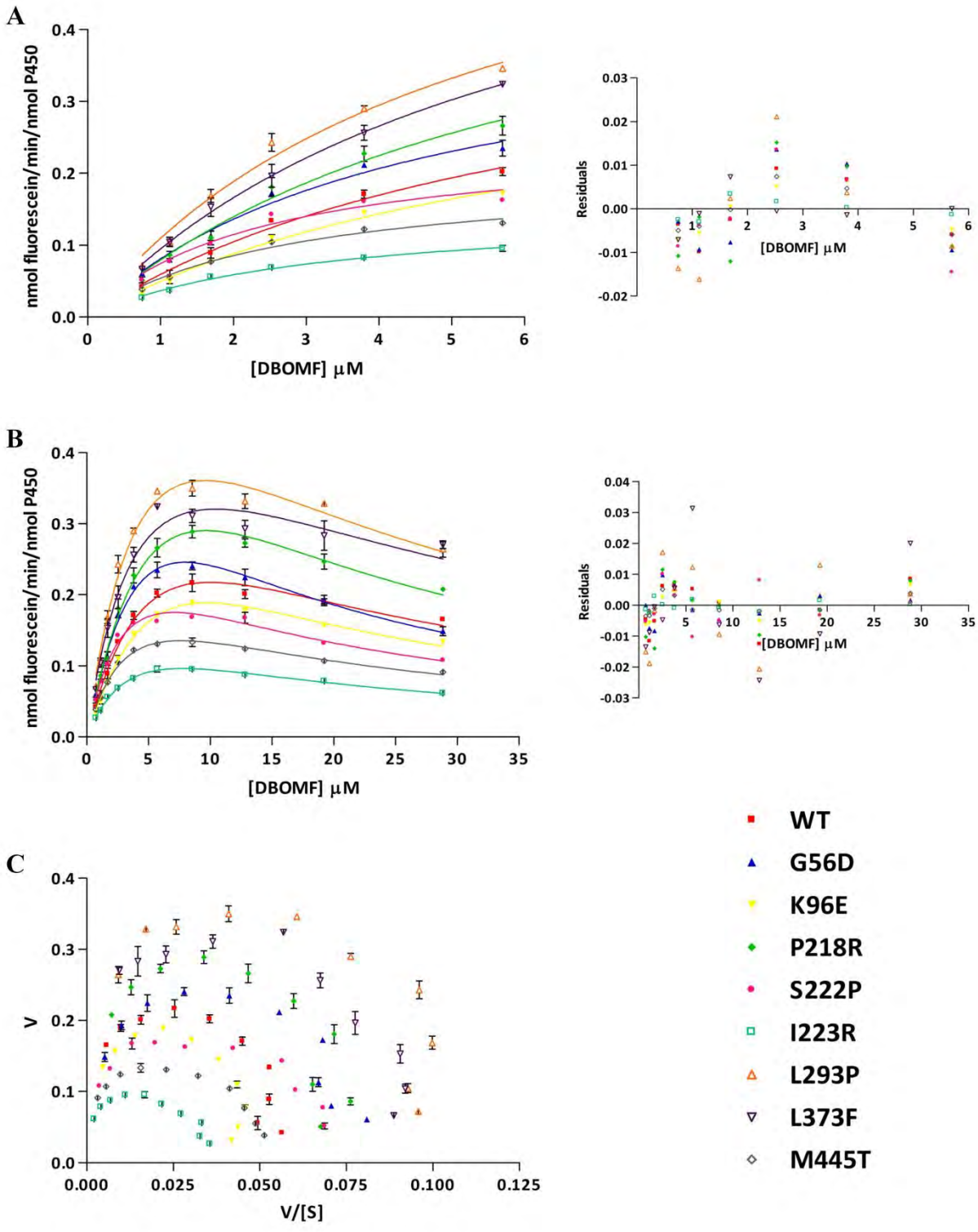


Figure 7.24 Turnover of DBOMF Vivid Green by CYP3A4 polymorphic variants
A final concentration of 50 nM P450, 0.5mM CuOOH and 15% glycerol in 100mM potassium phosphate buffer pH7.4 was used in each reaction. The error bars represent the standard deviation of 3 replicas. A) Michaelis-Menten model fitted to the truncated CYP3A4 kinetic data sets; B) Substrate inhibition model fitted to the full CYP3A4 kinetic data set. The residual plot for each fitted curve is also shown. C) Eadie-Hofstee plots for the full kinetic data set showing V (nmol fluorescein.min^{-1}.nmol P450^{-1}) *vs.* V/[S] (nmol fluorescein.min^{-1}.nmol P450^{-1}.µM DBOMF^{-1})

The substrate inhibition model showed no significant difference in K_m or K_i values for variants compared to wild-type. S222P, I223R and M445T had significantly lower K_m values when the Michaelis-Menten model was used on the truncated data set. S222P and I223R lie within substrate recognition sites so it is plausible that they may affect substrate binding. The catalytic efficiency of the CYP3A4 variants was not significantly different to the wild-type protein when either of the kinetic models were used (Figure 7.25 D).

Table 7.5 Kinetic constants for the turnover of DBOMF Vivid Green by CYP3A4 variants calculated using the Michaelis-Menten and substrate inhibition models

CYP3A4	WT	G56D	K96E	P218R	S222P	I223R	L293P	L373F	M445T
Michaelis-Menten									
V_{max} **(nmol/min/nmol P450)**	0.45 ± 0.04	0.43 ± 0.03	0.39 ± 0.03	0.58 ± 0.07	0.25 ± 0.02*	0.15 ± 0.01***	0.68 ± 0.05*	0.67 ± 0.04*	0.20 ± 0.01**
K_m **(μM)**	7 ± 1	4.3 ± 0.6	6.9 ± 0.9	6 ± 1	2.4 ± 0.3*	3.1 ± 0.3***	5.2 ± 0.7	6.1 ± 0.6	2.8 ± 0.3*
V_{max}/K_m	0.07 ± 0.01	0.10 ± 0.01	0.057 ± 0.009	0.09 ± 0.02	0.11 ± 0.02	0.048 ± 0.005	0.13 ± 0.02	0.11 ± 0.01	0.07 ± 0.009
R^2	0.980	0.975	0.987	0.971	0.952	0.985	0.979	0.989	0.976
Substrate inhibition									
V_{max} **(nmol/min/nmol P450)**	0.60 ± 0.08	0.8 ± 0.1	0.69 ±0.08	0.9 ± 0.1	0.39 ± 0.03	0.23 ± 0.01*	0.86 ± 0.09	0.64 ± 0.08	0.29 ± 0.02*
K_m **(μM)**	9 ± 2	9 ± 2	13 ± 2	11 ± 2	4.3 ± 0.6	5.2 ± 0.5	7 ± 1	5 ± 1	4.4 ± 0.4
V_{max}/K_m **($min^{-1}\mu M^{-1}$)**	0.07 ± 0.02	0.09 ± 0.02	0.05 ± 0.01	0.09 ± 0.02	0.09 ± 0.01	0.043 ± 0.005	0.13 ± 0.03	0.12 ± 0.03	0.066 ± 0.007
K_i **(μM)**	11 ± 2	7 ± 1	7 ± 1	9 ± 2	12 ± 2	11 ± 1	14 ± 3	21 ± 5	13 ± 1
R^2	0.978	0.985	0.992	0.981	0.973	0.988	0.977	0.957	0.986

Notes: ± Standard error. Student t-test was used to determine values significantly different from wild-type: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$.

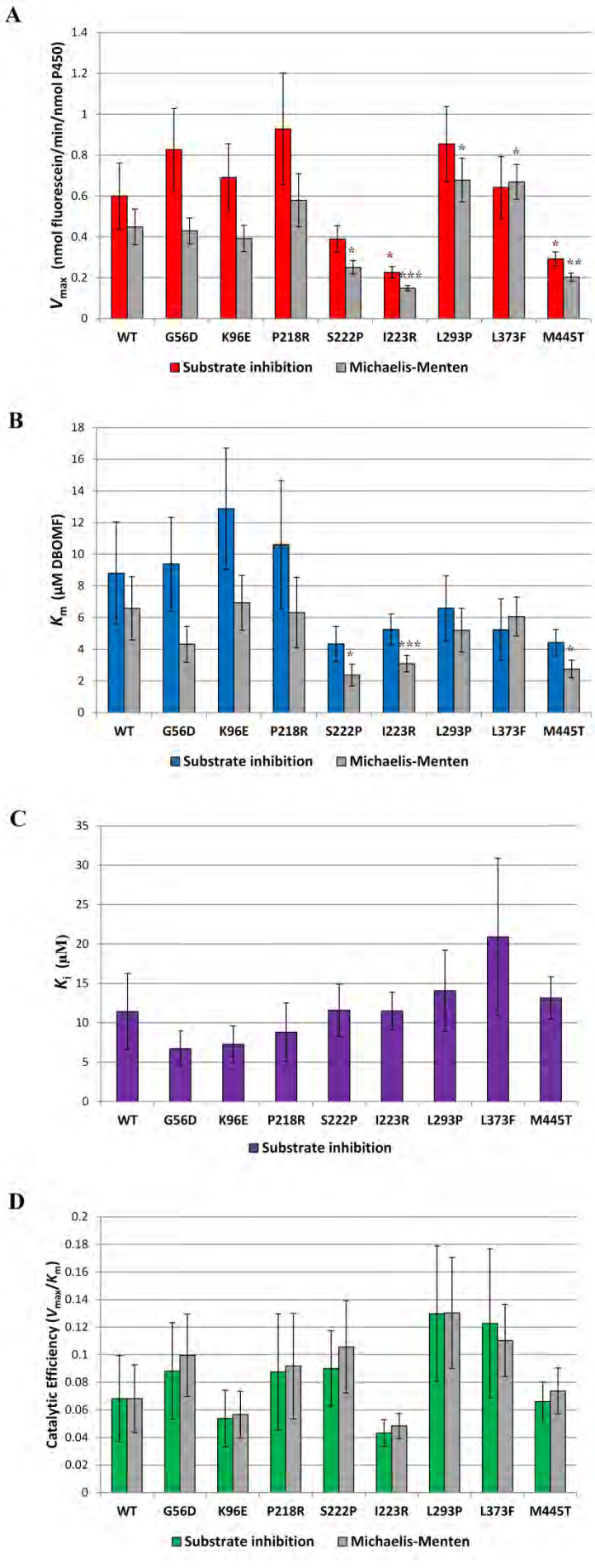


Figure 7.25 A visual comparison of the A) V_{max} , B) K_m, C) K_i and D) catalytic efficiency of different CYP3A4 variants for the turnover of DBOMF Vivid Green calculated using substrate inhibition model on the full data set and Michaelis-Menten model on the truncated data set. The error bars represent the 95% confidence intervals of the mean of 3 replicas. Student t-test was used to determine values significantly different from wild-type: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$

7.2.5 Activity of CYP2C9 variants

Kinetic assays were carried out for eight CYP2C9 polymorphic variant proteins, including the wild-type, using solution-phase CYP2C9 protein samples, OOMR Vivid red substrate and CuOOH to initiate the reactions as described above. OOMR was the substrate of choice because it showed typical Michaelis-Menten kinetics and high turnover when tested using wild-type CYP2C9. 15% glycerol was included in the reactions because it was shown to increase the rate of OOMR turnover.

Figure 7.26 shows the Michaelis-Menten model fitted to the CYP2C9 variant kinetic data sets, together with the residual and Eadie-Hofstee plots for the data. When compared to the biphasic model (data not shown), the Michaelis-Menten model was the preferred model for all variants and residual plot shows a random distribution of data points around the x-axis indicating that the model adequately explains the data. This is also confirmed by the linear nature of the Eadie-Hofstee plots.

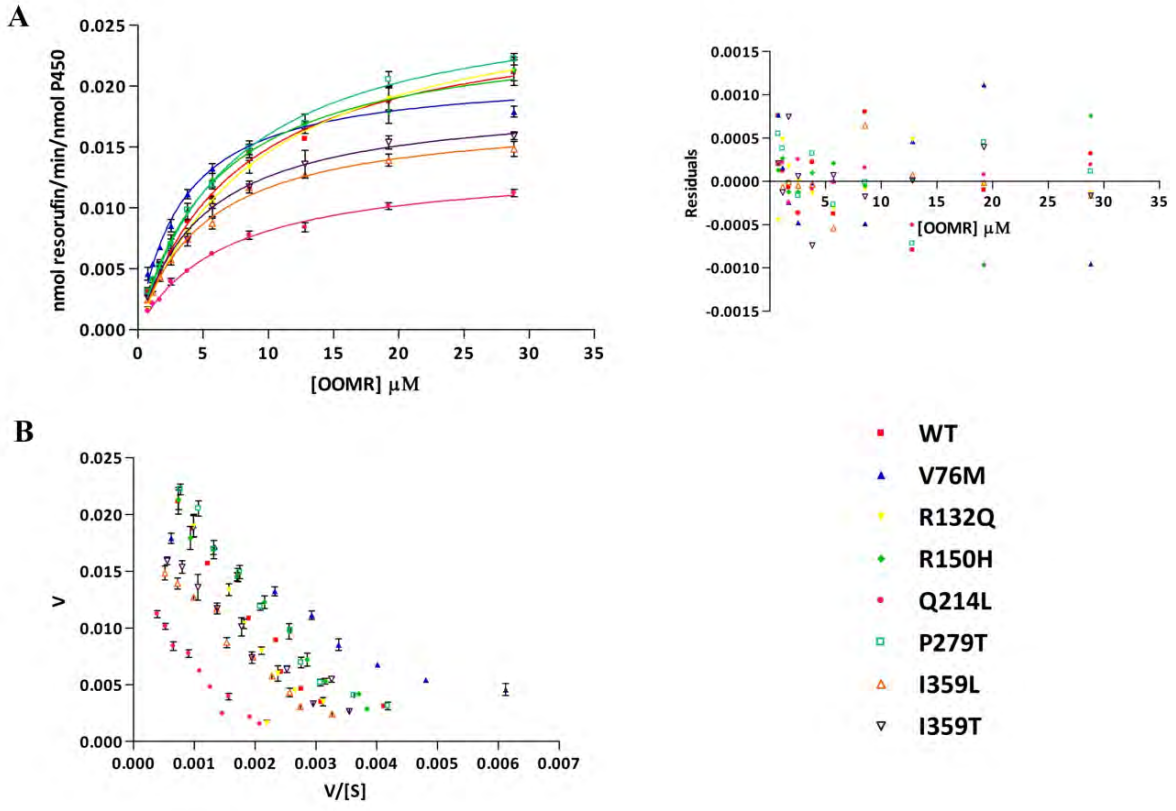


Figure 7.26 Turnover of OOMR Vivid Red by CYP2C9 variants. 250 nM P450, 0.5mM CuOOH and 15% glycerol (v/v) in 100mM potassium phosphate buffer pH7.4 was used in each reaction. The error bars represent the standard deviation of 3 replicas **A**) Michaelis-Menten model fitted to the CYP2C9 kinetic data set for each variant. The residual plot of the data is shown on the right. **B**) Eadie-Hofstee plot of data showing V (nmol resorufin.min^{-1}.nmol P450^{-1}) *versus* V/[S] (nmol resorufin.min^{-1}.nmol P450^{-1}.µM OOMR^{-1}). The error bars represent the standard deviation of 3 replicas.

Table 7.6 Kinetic constants for the turnover of OOMR Vivid green by CYP2C9 variants

CYP2C9	WT	V76M	R132Q	R150H	Q214L	P279T	I359L	I359T
Michaelis-Menten								
V_{max} (nmol/min/nmol P450)	0.0264 ± 0.0005	0.0211 ± 0.0005***	0.0283 ± 0.0007	0.0249 ± 0.0005	0.0136 ± 0.0003***	0.0277 ± 0.0005	0.0177 ± 0.0003***	0.0189 ± 0.0004***
K_m (μM)	7.7 ± 0.4	3.4 ± 0.2***	9.4 ± 0.5	6.0 ± 0.3*	6.8 ± 0.3	7.3 ± 0.3	5.2 ± 0.2***	5.0 ± 0.3***
V_{max}/K_m	0.0034 ± 0.0002	0.0062 ± 0.0005**	0.0030 ± 0.0002	0.0041 ± 0.0002	0.0020 ± 0.0001***	0.0038 ± 0.0002	0.0034 ± 0.0002	0.0037 ± 0.0003
R^2	0.993	0.983	0.992	0.989	0.991	0.994	0.992	0.984

± Standard error. Student t-test was used to determine values significantly different from wild-type: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$.

The kinetic parameters V_{max}, K_m and V_{max}/K_m for each variant are shown in Table 7.6 together with the R^2 values for curve fitting. A visual comparison of the V_{max}, K_m and V_{max}/K_m values are shown in Figure 7.27.

Variants V76M, Q214L, I359L and I359T had significantly lower V_{max} values than the wild-type. All four of these mutation were also found to have a significant effect on the thermal stability of the P450 form of the protein, with V76M, I359L and I359T found to be destabilising and Q214L found to be stabilising (Chapter 6 Results). V76M, I359L and I359T, as well as R150H, also showed a significant decrease in K_m compared to wild-type. V76M showed ~ 2 fold increase in catalytic efficiency due to the increase in *K*m and Q214 showed a ~ 40% decrease in catalytic efficiency due to the decrease in velocity. All other variants showed comparable catalytic efficiency to the wild-type enzyme.

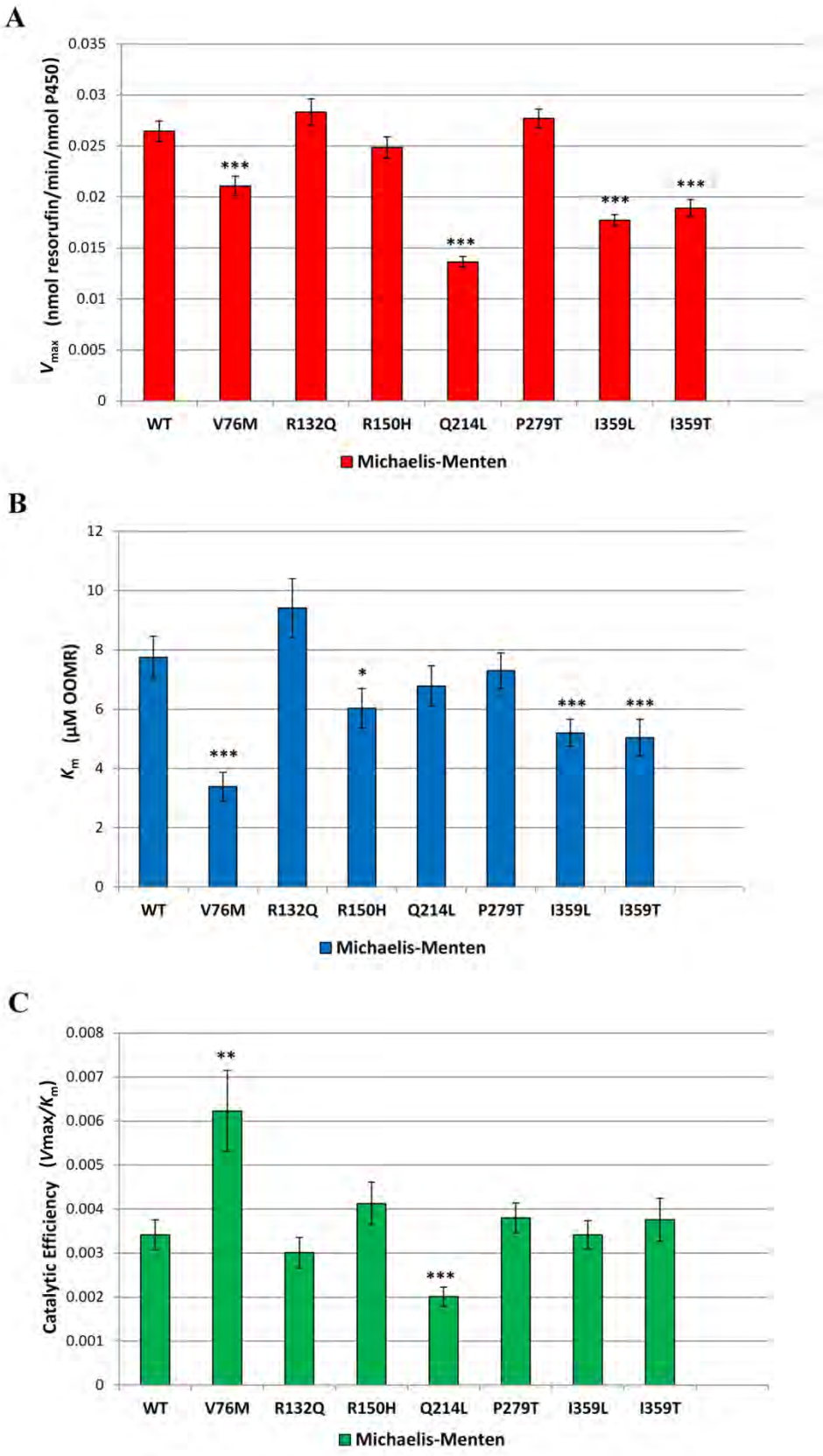


Figure 7.27 A visual comparison of the A) V_{max}, B) K_m, and C) catalytic efficiency of CYP2C9 variants for the turnover of OOMR Vivid Red calculated using the Michaelis-Menten model. The error bars represent the 95% confidence intervals of the mean of 3 replicas. Student t-test was used to determine values significantly different from wild-type: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$.

7.3 Discussion

7.3.1 The effect of glycerol on Cytochrome P450s

The inclusion of glycerol in the CYP450 assay mixes has a substantial effect on both the magnitude and shape of the kinetic profiles. Glycerol is well known for its protein stabilising properties and is commonly used in CYP450 storage buffers to preserve protein structure. The addition of sugars such as glycerol, sucrose and glucose to an aqueous solution containing proteins leads to the preferential hydration of the proteins' and hence an increase in the proteins chemical potential [530-532]. This in turn makes any increase in protein surface area in contact with the solvent or the exposure of additional hydrophobic residues to the solvent even more energetically unfavourable than in water, thereby stabilising the folded native state of globular proteins.

Glycerol also has other more specific effects on CYP450 proteins, with a number of studies showing that glycerol can affect the P450 spin state and inhibit P450 activity. Headlam *et al* found that glycerol shifted the spin state equilibrium of CYP11A1 towards the low-spin state and inhibited CYP11A1 activity in a substrate dependent manner [533]. Based on this study the authors concluded: 1) osmotic stress exerted by glycerol at concentrations up to 20% caused water molecules to dissociate from the active site, decreasing the molecular volume of CYP11A1 and causing a high-spin to low-spin shift; and 2) glycerol inhibited CYP450 activity by preventing the association of critical water molecules on substrate binding, the extent of inhibition depending on the size of the substrate and number of water molecules that usually associate with the CYP450 for the given substrate. In contrast to CYP11A1, cytochrome P450cam showed a shift from low-spin to high-spin as a result of the dissociation of 19 water molecules in the presence of sucrose or stachyose [501]. The effect of osmotic pressure on the spin equilibrium thus appears to be isoform specific and probably depends on the hydration state of the active site for the low- and high- spin conformers.

The addition of 20% glycerol increases the viscosity of water by 50% (West, 1978), however the viscosity of a liver microsomal membrane is about 2-fold higher than water (Feinstein 1978) so it seems unlikely that glycerol would have a significant influence on the movement of proteins

within the membrane [534]. It has however been suggested that glycerol inhibits electron transfer between CPR and the CYP450 by decreasing the dielectric constant of the medium and promoting inhibitory electrostatic interactions between the redox partners [534].

Perhaps more relevant to the work described here, glycerol can also affect CYP450 aggregate formation: a study showed that purified cytochrome CYP2B4 ($P\text{-}450_{LM4}$) monodispersed into hexameric oligomers in 20% glycerol but had a tendency to from larger non-specific aggregates at lower glycerol concentration [119]. This hexameric form of CYP2B4 is thought to be the specific, naturally occurring quaternary structure. Based on the thermostability assays on solution-phase CYP3A4 and CYP2C9 proteins described in Chapter 6, it seems reasonable to speculate that CYP3A4 and CYP2C9 may also form hexamers in solution: those assays were carried out in P450 storage buffer containing 20% glycerol and it is distinctly possible therefore that CYP3A4 and CYP2C9 form larger non-specific aggregates in the absence of glycerol, as seen for CYP2B4. However further studies would be required to confirm this and are beyond the scope of the present Thesis.

Based on the evidence from previous studies, it is clear that glycerol has a number of confounding effects of CYP450s that may vary according to the protein source and environment. This highlights the importance of considering the effects of glycerol when ascribing a kinetic model to the reactions as these effects may shed light on the molecular mechanisms underlying the observed atypical kinetics.

7.3.2 CYP3A4 and CYP2C9 wild-type kinetics

Enzyme activity was observed for CYP3A4 and CYP2C9 wild-type recombinant protein in solution using resorufin and fluorescein based fluorescent substrates in a simple 96 well plate assay format making use of the peroxide shunt pathway. Generally the K_m values for all four substrates were comparable to literature values for kinetics assays using baculosomal CYP450s and an NADPH regenerating system. The k_{cat} values for solution-phase CYP450 turnover of Vivid substrates however were substantially lower than those reported in the literature, regardless of the presence or absence of glycerol in the assays or the kinetic model fitted to the data (Table 7.7).

Table 7.7 Comparing kinetic parameters for the turnover of Vivid substrates by CYP450 with those reported in the literature

CYP450: Vivid substrate	This work				Literature	
	Solution * (+ 15% glycerol)		Baculosomes * (no glycerol)		Baculosomes †	
	K_m (μM)	k_{cat} (min^{-1})	K_m (μM)	k_{cat} (min^{-1})	K_m (μM)	k_{cat} (min^{-1})
CYP3A4:BOMR	3.8	0.01	3.2 Ψ	3.6 Ψ	2.3 / 1.4	6.3 / 1.1
CYP3A4:DBOMF	6	0.6	1.7	6.3	7.8 / na	8.6 / na
CYP2C9:OOMR	6.7	0.04	-	-	2.0 / 1.4	0.1 / 0.66
CYP2C9:BOMF	6	0.003	-	-	2.1 / 15	0.4/ 9

* Best fit model to the full data set were used where possible. Kinetic parameters for BOMR and BOMF in the presence of 15% glycerol were determined using the Michaelis-Menten model on the truncated data sets as models did not fit well to full data sets.
Ψ Baculosome kinetics for BOMR turnover were performed previously in the Blackburn laboratory where kinetic constants were obtained by fitting Michaelis-Menten model to truncated data set. [473]
† Kinetic parameters reported by [535]/ [536]

Analogous kinetic assays using the DBOMF substrate and commercially available baculosomal CYP3A4 preparations containing CPR gave much higher k_{cat}'s, comparable to the literature, and the kinetic parameters were similar regardless of whether NADPH or CuOOH was used to drive the reaction. This suggests that the differences between the CuOOH driven peroxide shunt pathway and the NADPH-CPR dependent catalytic cycle do not account for the differences in activity between solution-phase CYP450s and membrane bound CYP450s. Differences in protein environment must therefore be responsible for the variation in activity and it may therefore be relevant that interaction with lipids and accessory proteins within the membrane stabilise the protein have been shown to modulate CYP450 activity [124].

One possible explanation here is that the formation of oligomers or aggregates in solution – not formed in the membrane - may trap proteins in a closed, inactive conformation unable to bind larger substrates but still susceptible to reduction by dithionate and CO binding. If this were the case, the number of CYP450 molecules available to turnover substrate would be much less than predicted by CO P450 spectral assays and would lead to an underestimation of $k_{cat.}$

Solution-phase CYP3A4 displayed atypical kinetics profiles for both resorufin- and fluorescein-based Vivid substrates that differed from the atypical profiles observed for baculosomal CYP3A4. Interestingly the inclusion of 15% glycerol in the reaction mixture gave kinetic profiles that closely resembled those obtained for baculosomal CYP3A4 and increased the rate of

substrate turnover per molecule P450 at all substrate concentrations. It appears therefore that glycerol may mimic the membrane environment to some extent; this is likely a result of glycerol's stabilising properties and its ability to affect the spin state of CYP450s *via* osmotic pressure. Davydov *et al* have shown that a pressure-sensitive conformational heterogeneity exists in both CYP3A4 oligomers and monomers, revealing that a conformational equilibrium is inherent to the monomeric structure. The authors found that allosteric substrates [77] and the membrane environment [124] could shift the conformational equilibrium, stabilising the high-spin state of the haem iron by reducing the accessibility of water into the active site. It therefore seems reasonable to propose here that, in a similar way to the membrane or an effector molecule, glycerol is shifting the conformational equilibrium towards the high-spin state by displacing water molecules from the active site, leading to kinetic profiles that resemble those seen for baculosomal CYP450s.

Despite obtaining a similar shape kinetic curve and an increase in the rate of substrate turnover for solution-phase CYP3A4 in the presence of 15% glycerol, the k_{cat} was still ~ 400 fold lower for CYP3A4 turnover of BOMR (calculated using the Michaelis-Menten model) and 10 fold lower for DBOMF (calculated using the substrate inhibition model) than that observed for baculosomal CYP3A4 in the absence of glycerol. While the 10 fold difference for DBOMF may be attributed to effects of the membrane environment and/or the formation of oligomers/aggregates in solution (even in the presence of glycerol) it is not immediately apparent why there should be a 400 fold difference for the turnover of BOMR by CYP3A4; this apparent anomaly will be discussed further in Chapter 8.

In contrast to CYP3A4 in solution, 15% glycerol inhibited the NADPH-CPR mediated baculosomal CYP3A4 turnover of DBOMF; reducing the k_{cat} by 40%. Here, glycerol may inhibit CYP450 activity by inhibiting electron transfer by the CPR as described by Voznesensky *et al* but the effect of glycerol on CuOOH driven baculosomal CYP3A4 would need to be tested to determine whether glycerol primarily inhibits NADPH-CPR mediated baculosomal activity and this was deemed to be outside the scope of this Thesis.

Proposed Kinetic model for CYP3A4 turnover of DBOMF

The kinetic profile for the turnover of DBOMF in the absence of glycerol consists of two peaks resembling two overlapping substrate inhibition curves; one with a low K_m and one with a higher K_m (Figure 7.8 B). In the presence of 15% glycerol the kinetic profile has a single peak to which a single substrate inhibition curve with a low K_m could be fitted (Figure 7.19 A). The model proposed here for the turnover of DBOMF by CYP3A4 thus assumes two enzyme conformers: one with a high binding affinity and one with a low binding affinity, each able to bind two molecules of DBOMF and with the binding of the second molecule resulting in enzyme inhibition (Figure 7.28 A). In the proposed model, the presence of glycerol stabilises the high affinity conformer, shifting the equilibrium and thereby increasing the velocity in the low concentrations range (Figure 7.28 B). By contrast, the high affinity conformer is dominant in the baculosomal protein preparations as it is stabilised by membrane interactions even in the absence of glycerol. The classic substrate inhibition model is shown for each conformer in Figure 7.28 for simplicity but it is likely that the substrate inhibition may be more complex in reality, resembling something more like the P450 substrate inhibition model (Figure 7.17).

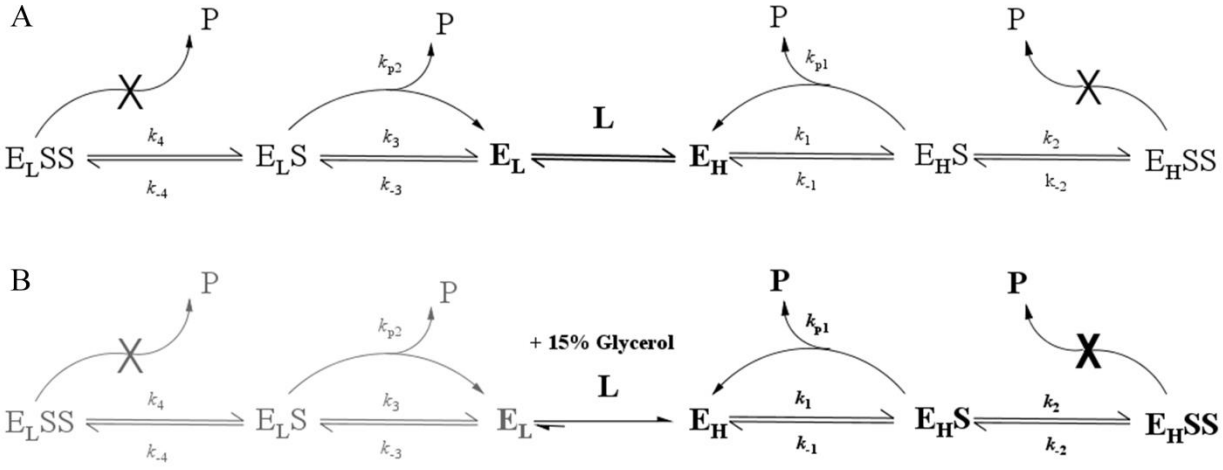


Figure 7.28 Kinetic model for the turnover of DBOMF by CYP3A4 in the absence (A) and presence of 15% glycerol (B). The high affinity CYP3A4 conformer (E_H) and low affinity conformer (E_L) are in equilibrium (equilibrium constant L) and can both bind two substrate molecules (S) with the binding of the second substrate leading to a non-productive enzyme species. The rates of product formation by E_H and E_L are given by rate constants k_{p1} and k_{p2} respectively.

The K_m values for both substrates were comparable to those observed for the NADPH-CPR mediated baculosomal CYP3A4 system and were not significantly affected by the presence of 15% glycerol in the reactions mixtures using the Michaelis-Menten model fitted to the truncated data sets or the biphasic or substrate inhibition models fitted to the full data sets where applicable. This shows that glycerol is not affecting the binding affinity of the substrate to high affinity enzyme - as would be expected for the model described here - and supports the theory that baculosomal CYP3A4 is predominantly in a state comparable to the high affinity state observed for the solution-phase CYP3A4.

Proposed Kinetic model for CYP3A4 turnover of BOMR

An analogous explanation for the curves observed for the turnover of BOMR in the absence and presence of 15% glycerol is more complicated. The kinetic curve in the absence of glycerol shows biphasic kinetics (Figure 7.8 A). One again the presence of 15% glycerol increases the height of the initial hyperbolic part of the profile (Figure 7.8 A) and the K_m values remain comparable for the Michaelis-Menten model fitted to the truncated data sets. The presence of glycerol causes a decrease in velocity after the initial hyperbolic part of the profile – resembling substrate inhibition - troughing around 8µM followed by a linear increase as seen for the baculosomal preparations.

A curve described by the biphasic model (Figure 7.14) with two binding site is mathematically equivalent to a model consisting of two conformers: one with a high affinity binding site and one with a low affinity binding site (Figure 7.29). In this case, the hyperbolic section results from turnover from the high affinity site and the linear portion of the graph arises from a combination of turnover from both conformers.

Similarly to the CYP3A4 model for DBOMF, the presence of glycerol shifts the equilibrium towards the high affinity conformer, which is then subject to substrate inhibition, making the presence of substrate inhibition more evident in the profile. The low affinity conformer does not show substrate inhibition or reach saturation in the substrate concentration range tested and hence the velocity increases linearly with BOMR concentration after the trough (Figure 7.8 A).

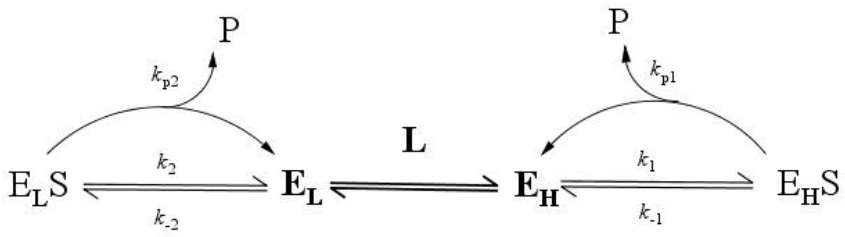


Figure 7.29 Biphasic model based on the presence of two enzyme conformers. High affinity CYP3A4 conformer (E_H) and low affinity conformer (E_L) are in equilibrium (equilibrium constant L) and can both bind substrate molecule (S). The rates of product formation by E_H and E_L are given by rate constants k_{p1} and k_{p2} respectively.

CYP2C9 turnover of OOMR and BOMF

The turnover of OOMR by CYP2C9 followed Michaelis-Menten kinetics in the absence and presence of 15% glycerol. Glycerol increased the rate of substrate turnover while the Km remained unchanged at ~ 7μM.

The rate of turnover of BOMF by CYP2C9 was very low and followed a profile that resembles the start of a substrate inhibition, but none of the models tested fitted adequately to the data. Glycerol did not have a significant effect on K_m or V_{max} when the Michaelis-Menten model was fitted to the truncated data set. The error bars were large owing to the very low changes in fluorescence with time; this assay would need to be repeated with a higher number of replicates and a larger range of substrate concentrations to determine whether substrate inhibition is truly occurring and to find a suitable model to fit the data.

Glycerol had a less prominent effect on CYP2C9 activity than on CYP3A4 activity; CYP2C9 was much more stable in solution than CYP3A4 and, while thermostability studies suggest that there are energetically distinct sub-populations of CYP2C9 in solution that may result from oligomers/aggregates (Chapter 6), there is less evidence in the literature for catalytically distinct CYP2C9 conformers in solution.

7.3.3 Effect of CYP3A4 and CYP2C9 polymorphic variation on Vivid substrate turnover

Kinetic assays for CYP3A4 and CYP2C9 polymorphic variants were carried out in the presence of glycerol because glycerol increased the rate of substrate turnover and gave kinetic profiles that

more closely resembled profiles obtained for membrane bound CYP450s. V_{max} in nmol product formed per min per nmol P450 (=k_{cat}) was calculated based on the concentration of specific P450 in the protein sample, as determined by carbon monoxide binding assays, ruling out any differences in activity resulting from different levels of expression or P450:P420 ratios.

The kinetic profiles for all CYP3A4 variants showed substrate inhibition at higher substrate concentrations, as seen for the wild-type. Since most literature reports comparing the kinetic parameters of CYP3A4 variants simply assume Michaelis-Menten kinetics, both the substrate inhibition model and the Michaelis-Menten models were fitted to the full and truncated data sets respectively in order to see how the choice of model affects the kinetic parameters and conclusions drawn from the data.

CYP3A4 variants

The substrate inhibition model gave the best fit to the data sets, but kinetic parameters calculated using this more complex model had larger confidence interval and hence fewer variants had a significantly different V_{max} compared to the wild-type and none of the variants had a significantly different K_m.

CYP3A4 variants I223R maximum turnover rate of DBOMF was only 30% of the wild-type. I223R was found in breast cancer patients in 2009 [537] but there are no studies reported to date that have tested the effect of this mutation on CYP450 drug metabolism. The side chain of this residue - falling within the F-G loop and SRS (2,3) - is orientated towards the active site and has been implemented in the gating of pw2b. Both *in silico* predictions by SDM and thermostability assays indicate that this mutation is destabilising. These results suggest that the I223R mutation may be linked to a poor metaboliser phenotype.

The V_{max} for the CYP3A4 variant M445T was 50% lower than the wild-type. M445T corresponds to allele CYP3A4*3 and was first identified in a Chinese population at a frequency of < 1% [538] and later found in Caucasian's with a frequency of 4%, with Eastern Europeans having the highest frequency (10%) [539]. Met445 is positioned at the start of α-helix on the border of the cys pocket, just two residues away from the haem bound cysteine residue (Figure 7.30). *In silico* predictions using SDM and thermostability assays both indicate that this mutation

is destabilising. Furthermore this residue is solvent accessible and is predicted to form part of the P450:CPR interface, although the assay setup used here however excludes any changes in activity resulting from a mutational effect on CPR binding. It is interesting to note though that, since CPR binds to the proximal face of CYP450, this mutation alters interactions in this region (as seen in Figure 6.30) so it is likely that it will alter interactions between the redox partners in addition to destabilising CYP3A4.

By contrast, other studies where CYP3A4 M445T was expressed in *E.coli*, quantified using CO spectral binding assays and reconstituted with CPR and lipids, reported no significant difference in turnover numbers for testosterone [539, 540] or chlorpyrifos at the single substrate concentrations tested [539]. Met445 does not lie within an SRS region so it would not automatically be expected to have substrate specific effects. One explanation for this apparent conflict in data may therefore be that the presence of lipids and CPR in their reconstituted reaction mixture had a stabilising effect on the protein, minimising the observed effect of the mutation. Notable though, the studies on testosterone and chlorpyrifos were not full kinetic studies so a comparison of their data with the effects of the M445T on catalytic efficiency rather than k_{cat} may be more appropriate; this would render the results consistent as the catalytic efficiency of M445T determined was similar to the wild-type enzyme.

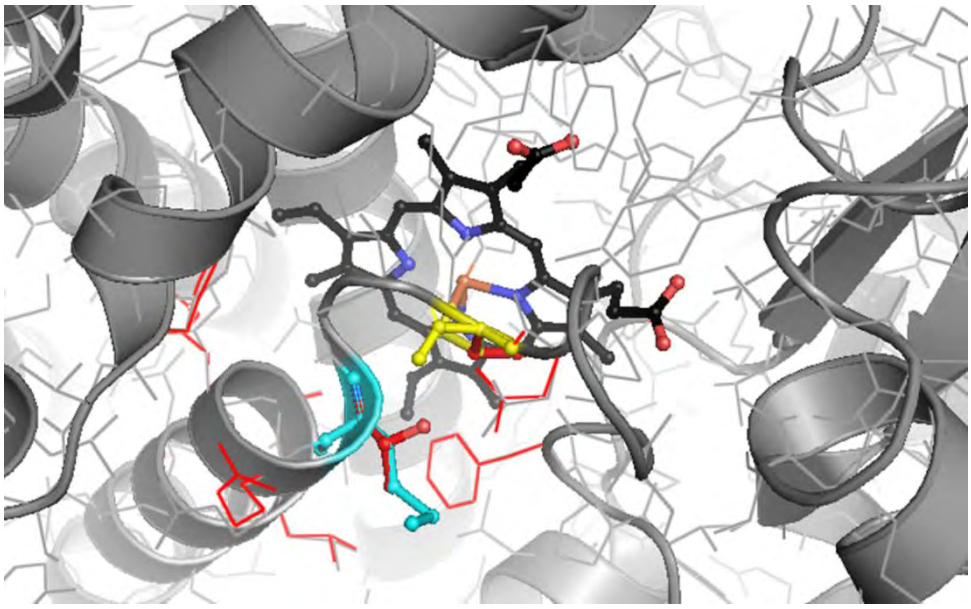


Figure 7.30 Proximal view of CYP3A4 M445T model aligned to the wild-type structure
The wild-type CYP3A4 pdb structure 2V0M is shown in grey. The M445T model was generated using the program, ANDANTE. Wild-type residue Met445 is shown in cyan ball and sticks and the side chain of the variant residue Thr445 is shown in red ball and sticks. Other side chain positions in the M445T model that deviate from the wild-type structure can be seen in red. The haem group is coordinated to the cysteine residue, in yellow.

S222P, L293P and L373F only showed significantly altered kinetic parameters when the Michaelis-Menten model was fitted to the truncated data set. All three variants had a higher affinity for the DBOMF substrate than the wild-type, reflected by significantly lower K_m values. S222P also had a lower k_{cat} whereas L293P and L373F had higher turnover numbers.

Previous *in vitro* studies have shown that the S222P mutation causes a 15 – 70% decrease in K_{cat} values for the turnover of testosterone, midazolam and nifedipine [538, 541]. The effect of S222P on substrate affinity was reported to be substrate specific and resulted in the following changes to K_m: 3-4 fold higher for nifedipine, 2 fold higher for midazolam and similar for testosterone. Ser222 falls within the F' helix and SRS (2, 3) and, although the side chain is not facing towards the active site in the crystal structure, this mutation effects substrate binding in a substrate specific manner. A proline residue is usually found at the end of an alpha helix and is known as a helix breaker; this S222P mutation is therefore likely to alter the conformation of the flexible F-G loop region. In Chapter 6 it was proposed that this mutation may affect CYP450 oligomer formation and, as discussed here, this could also lead to changes in kinetic parameters.

L293P is a solvent exposed residue found near the start of the helix I, on the border of SRS 4. This mutation was predicted to be neutral by SDM and did not affect the thermostability of the enzyme. Previous *in vitro* studies either found that this variant had similar kinetic parameters to the wild-type enzyme [541, 542] or that it gave a 2 fold increase in turnover number compared to wild-type for the substrates tested [539]. This is similar to the results seen here for DBOMF turnover where the V_{max} was about 50% higher but the difference was only significant for the Michaelis-Menten model fitted to the truncated data sets.

L373F gave a four-fold increase in K_m for midazolam metabolism and displayed an altered testosterone metabolite profile in an *in vitro* study using a bacterial expression system [540]. Leu373 is a haem contact and thermostability assays confirmed that mutating this residue to the larger phenalaline residue destabilises the P450 form of the protein (see Chapter 6, Figure 6.4 for a model of L373F aligned to the wild-type structure). Surprisingly, despite being destabilising, this mutation did not have a significant effect on substrate binding or catalytic efficiency and showed a moderate but significant increase in K_{cat} for DBOMF turnover when the Michaelis-Menten model was fitted to the truncated data set.

CYP2C9 variants

CYP2C9 variants all followed Michealis-Menten kinetics for the turnover of OOMR, making data analysis more straight forward and giving much smaller confidence intervals for kinetic parameters. V76M, Q214L, I359L/T and R150H variants all showed altered CYP2C9 metabolism.

Despite decreasing the thermostability of CYP2C9 and leading to a moderate decrease in K_{cat}, V76M had an increased catalytic efficiency due to a two fold increase in binding affinity for the OOMR substrate. This was surprising as V76M does not fall within an SRS region. While this mutation was predicted to be neutral by SDM, it decreased the proportion of stable P450 in thermostability assays (Chapter 6). Binding of a substrate to a less stable CYP450 may stabilise the structure, causing a larger decrease in free energy than binding to a more stable CYP450; consequently, the substrate would have a higher binding affinity for the less stable mutant enzyme. The CYP2C9*23 allele responsible for the in V76M mutation was first identified in an European American population in 2005 [543] but there are no reports to date of the effects of this mutation on enzyme activity.

Q214L did not alter the K_m of CYP2C9 for OOMR but did decrease the k_{cat} by two fold, leading to a significant decrease in catalytic efficiency. Q214L is a stabilising mutation - confirmed by both SDM predictions and thermostability assays - and despite falling within SRS (2, 3), it did not have a significant effect on substrate affinity. This confirms the premise that stabilising mutations can also be damaging: increased stability can lead to reduced conformational flexibility that is important for function. CYP2C9*28 containing the Q214L mutation was found in Japanese subjects with an allele frequency of 0.002 [544]. *In vitro* studies using mammalian or insect cell expression systems found that Q214L decreases catalytic efficiency by between 45 to 85% [544, 545] and a significant decrease in binding affinity was reported for two of the substrates tested, indicating that this residue plays a more important role in the binding of some substrates than others.

I359L and I359T are the non-synonymous amino acid substitutions associated with CYP2C9*3 and CYP2C9*4 alleles, respectively. I359L has long been associated with poor warfarin metabolism and patients with this polymorphism require lower warfarin doses as they are at

higher risk of bleeding complications when using this anticoagulant [546]. I359L and I359T showed a similar decrease in K_{cat} and K_m but no significant change to the catalytic efficiency for OOMR turnover. Ile359 falls within SRS 5 and a mutation at this position - to either leucine or threonine - greatly decreases the thermostability of the wild-type enzyme in solution (Chapter 6). Several *in vitro* studies, most of which are reports on I359L, show that mutations at this position leads to a large decrease in catalytic activity and an increase in K_m for some but not all substrates tested [544, 547-553]. Figure 7.31 shows models of variants I359L and I359T aligned to the wild-type CYP2C9 structure.

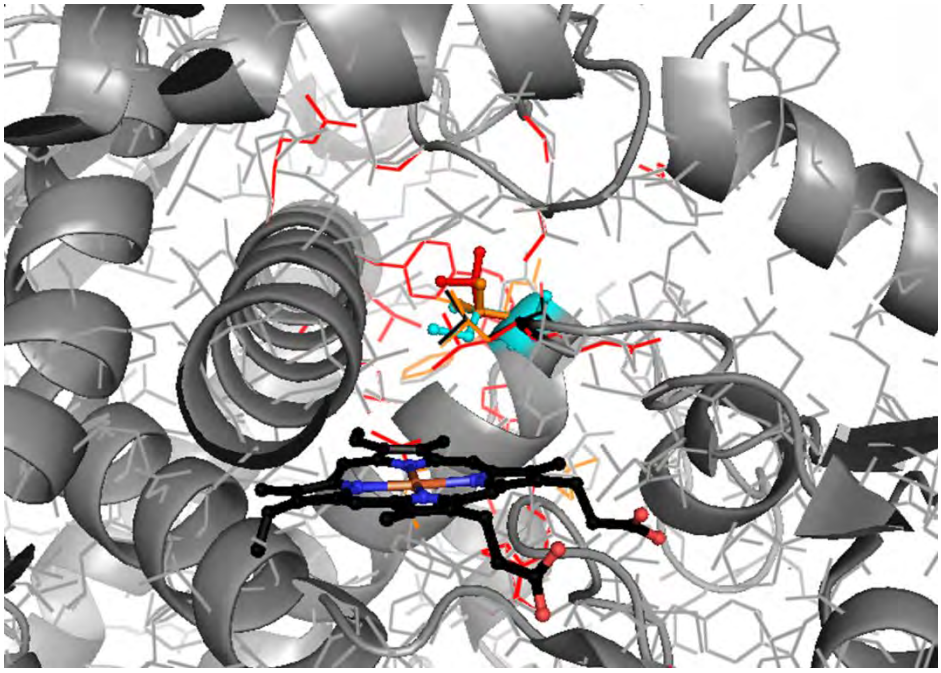


Figure 7.31 CYP2C9 I359L and I359T models aligned to the wild-type structure
The wild-type CYP2C9 pdb structure 1R9O is shown in grey. The I359L and I359T models were generated using the modelling program, ANDANTE. Wild-type residue Ile359 is shown in cyan balls and sticks and the side chains of the variant residues Leu359 and Thr359 are shown in orange and red ball and sticks, respectively. Other side chain positions that deviate from the wild-type structure can be seen in orange (I359L) and red (I359T). The haem group is shown in black.

The CYP2C9*8 R150H allele commonly occurs in African Americans and has been associated with decreased warfarin metabolism in *in vivo* studies [554]. An *in vitro* study also showed decrease phenytoin metabolism by this variant compared to the wild-type enzyme [555]. The only previous *in vitro* study that determined the full kinetic parameters of R150H found that this variant had improved affinity for CPR and a two fold lower K_m, resulting in a two fold increase in catalytic efficiency for tolbumide metabolism [556]. While the kinetic assays in this Chapter did not account for any changes in enzyme activity resulting from changes in CPR affinity, a

significant decrease in K_m was also observed here for R150H. The k_{cat} and catalytic efficiency were also higher compared to the wild-type enzyme, but the difference was not statistically significant. Interestingly R150H does not fall within any of the known functional regions of CYP2C9, was not predicted to be damaging by SDM and did not have an effect on the thermostability of the enzyme (Figure 7.32). Despite this, it appears to affect CYP2C9 function *in vivo* and *in vitro*.

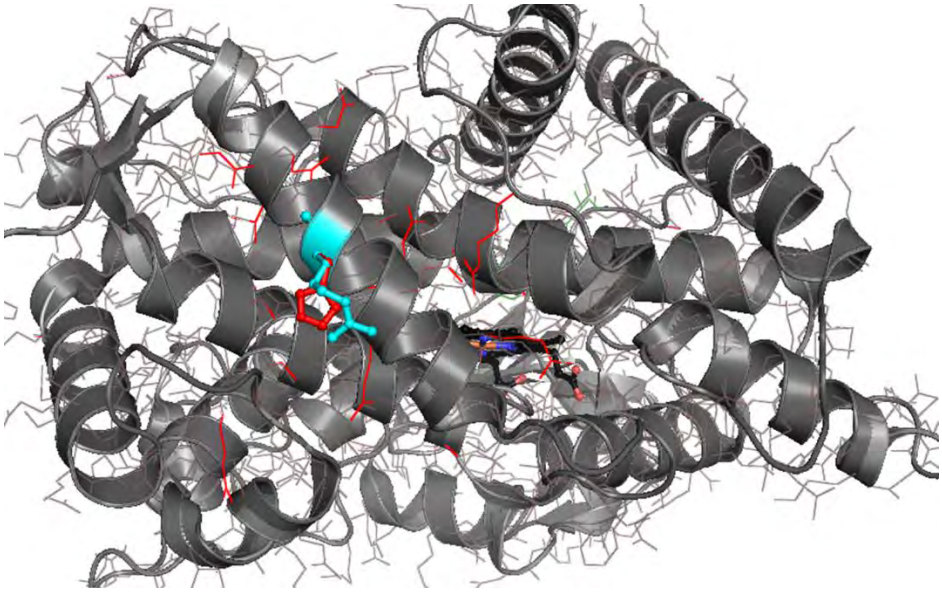


Figure 7.32 CYP2C9 R150H model aligned to the wild-type structure
The wild-type CYP2C9 pdb structure 1R9O is shown in grey. The R150H variant model was generated using the modelling program, ANDANTE. Wild-type residue Arg150 is shown in cyan ball and sticks and the side chain of the variant residue His150 is shown in red ball and sticks. Other side chain positions that deviate from the wild-type structure, as a result of the mutation, can be seen in red. The haem group is shown in black.

7.3.4 Conclusion

Solution phase P450 assays, making use of fluorescent substrates and CuOOH to drive catalysis, were used to test the effect of polymorphic variation on the kinetic activity of CYP450s and have provided new insights into the effect of mutations on CYP450 function. This solution-phase assay setup does however have the following disadvantages: 1) Human CYP450s in solution do not always behave in the same way as CYP450 immobilised within a membrane. This led to very low levels of metabolism for solution-phase CYP450s compared with baculosomal CYP450s; 2) The effect of mutations on CPR: P450 interactions cannot be accounted for in this assay format; 3) The assays are labour intensive and require large amounts of reagents and expensive

fluorescent substrates, making it difficult to test large panels of variants in parallel; 4) Vivid substrates, particularly DBOMF and OOMR, were unstable in aqueous solutions with significant auto-oxidation occurring over a 6 hour time period, making it essentially impossible to use the same working substrate solution to test all the variants, even if all the assays were carried out consecutively on the same day. These limitations once again highlight the need for a robust, cost-effective, high-throughput platform to monitor CPR mediated CYP450 drug metabolism.

7.4 Methods

7.4.1 Testing the activity of CYP3A4 and CYP2C9 using Vivid substrates

CYP3A4 and CYP2C9 wild-type proteins were expressed in *E.coli* and purified using by His Ni-TED columns and P450 levels in the partially purified samples were measured using CO P450 spectral assays as described in Chapter 5. The CYP3A4 and CYP2C9 Vivid substrates (Invitrogen) listed in Table 7.8 were dissolved in 100% acetonitrile and the 2 mM stock solutions were stored at -20°C. Fresh working stocks of Vivid substrates were made up in 100 mM potassium phosphate buffer pH 7.4 before each assay. Enzymes-substrate premixes were prepared on ice, aliquoted into Nunc 96-well polypropylene microwell plates (Thermo Scientific, USA) and equilibrated at room temperature. Reactions were initiated with cumene hydroperoxide (CuOOH, Sigma-Aldrich) and the plate was shaken on a plate vortex for ~10 s prior to starting readings. 100 µl reaction mixtures contained final concentrations of 4µM Vivid substrate, varying concentrations of P450 and 0.5 mM CuOOH in 100 mM potassium phosphate buffer, pH 7.4. The change in fluorescent units (RFU) was monitored continuously for 20 minutes using a Varian Cary Eclipse fluorescence plate reader. Instrument setting for the different substrates are shown in Table 7.8. Negative controls consisted of the following: substrate and CuOOH, no P450; CuOOH and P450, no substrate; substrate and P450, no CuOOH. The change in fluorescent units (ΔRFU) was plotted vs. time (min) and the initial velocity (ΔRFU/min) was calculated for the linear part of the reaction (usually over the first 5 or 10 minutes). Velocity was converted from ΔRFU/min to nM product formed/min using a standard curve generated from a fluorescein sodium salt (Sigma-Aldrich) or a resorufin sodium salt (Sigma-Aldrich) dilution series measured using the same instrument settings used in the assay.

Table 7.8 Fluorescent plate reader settings used for CYP450 activity assays

Vivid Substrate	Product	Ex/Em (nm)	Slit width Ex/Em (nm)
BOMR 3A4 red	resorufin	530/585	10/10
DBOMF 3A4 green	fluorescein	485/530	10/10
OOMR 2C9 red	resorufin	530/585	10/10
BOMF 2C9 green	fluorescein	485/530	10/5

7.4.2 CYP3A4 and CYP2C9 wild-type kinetic assays

CYP3A4 and CYP2C9 kinetics assays for the turnover of Vivid substrates in the presence and absence of 15% glycerol were carried as above with the following amendments:

Solution-phase CYP3A4 kinetics for the turnover of BOMR Vivid Red substrate
Each 100µl reaction contained final concentrations of 0.25-24 µM BOMR substrate, 200 nM CYP3A4, 0.5mM CuOOH and 0% or 15% (v/v) glycerol in 100 mM potassium phosphate buffer pH 7.4. Assays were carried out in duplicate for each substrate concentrations and the average initial velocity and standard deviation was calculated over the first 5 min of the reactions.

Solution-phase CYP3A4 kinetics for the turnover of DBOMF Vivid Green substrate
Each 100µl reaction contained a final concentration of 0.25 - 32µM DBOMF substrate, 35 nM CYP3A4, 0.5 mM CuOOH and 0% or 15% (v/v) glycerol in 100 mM potassium phosphate buffer pH 7.4. Assays were carried out in duplicate and the average initial velocity and standard deviation was calculated over the first 5 min of the reactions.

Baculosomal CYP3A4 kinetics for the turnover of DBOMF Vivid Green substrate
NADPH-CPR mediated CYP3A4 kinetic assays were carried out using a commercial baculosomal reagent co-expressing CYP3A4 and CPR (Invitrogen). Fresh 30 mM NADPH (Sigma-Aldrich) stocks in 100 mM potassium phosphate buffer were made up and stored at 4°C for up to two weeks. Reactions were initiated by the adding substrate-NADPH premixes to the baculosomal regent. Each 100 ul reaction contained a final concentration of 2 nM CYP3A4 (the concentration of CPR in the baculosomal reagent was not supplied by Invitrogen), 0.5 – 20 µM DBOMF, 1 mM NADPH and 0% or 15% (v/v) glycerol in 100mM potassium phosphate buffer. Assays were carried out in triplicate and the average initial velocity and standard deviation was calculated over the first 5 min of the reactions.

Solution-phase CYP2C9 kinetics for the turnover of OOMR Vivid Red substrate
Each 100 µl reaction contained final concentrations of 0.75-28.8 µM OOMR substrate, 250 nM CYP2C9, 0.5 mM CuOOH and 0% or 15% (v/v) glycerol in 100 mM potassium phosphate buffer pH 7.4. Assays were carried out in triplicate for each substrate concentrations and the average initial velocity and standard deviation was calculated over the first 5 min of the reactions.

Solution-phase CYP2C9 kinetics for the turnover of BOMF Vivid Green substrate

Each 100 µl reaction contained final concentrations of 0.25 - 24 µM BOMF substrate, 250 nM CYP2C9, 0.5 mM CuOOH and 0% or 15% (v/v) glycerol in 100mM potassium phosphate buffer pH 7.4. Assays were carried out in duplicate for each substrate concentrations and the average initial velocity and standard deviation was calculated over the first 8 min of the reactions.

7.4.3 CYP3A4 and CYP2C9 variant kinetic assays

Variant proteins were expressed in *E.coli* and purified using by His Ni-TED columns and P450 levels in the partially purified samples were measured using CO P450 spectral assays as described in Chapter 5. Variants and wild-type kinetic assays were all carried out on the same day, however separate 100 µM Vivid working solutions were made up from the same 2 mM Vivid stock in acetonitrile immediately before assaying each variant because both DBOMF and OOMR substrates were found to be unstable in aqueous solution, either as a result of autoxidation or precipitation. Kinetics assays for the turnover of Vivid substrates were carried out as described in section 7.4.1 with the following amendments:

CYP3A4 variant kinetic assays

Each 100 µl reaction contained final concentrations of 0.75 -28.8 µM DBOMF substrate, 50 nM CYP3A4, 0.5 mM CuOOH and 15% (v/v) glycerol in 100 mM potassium phosphate buffer pH 7.4. Assays were carried out in triplicate for each substrate concentrations and the average initial velocity and standard deviation was calculated over the first 5 min of the reactions.

CYP2C9 variant kinetic assays

Each 100 µl reaction contained final concentrations of 0.75 -28.8 µM OOMR substrate, 250 nM CYP2C9, 0.5 mM CuOOH and 15% (v/v) glycerol in 100mM potassium phosphate buffer pH 7.4. Assays were carried out in triplicate for each substrate concentrations and the average initial velocity and standard deviation was calculated over the first 10 min of the reactions.

7.4.4 Kinetic data analysis

Graph-Pad Prism software (San Diego, USA) was used to fit various kinetic models to the data using non-linear regression. Equations used for curve fitting to determine kinetic parameters for the different kinetic models are shown in Table 7.9. Eadie-Hofstee graphs (velocity vs.

velocity/substrate) were also plotted to give insight into which kinetic model was likely to fit the data best.

Table 7.9 Equations representing the different kinetic models fitted to the kinetic data using nonlinear regression.

Kinetic Model	Equation	Reference
Michaelis-Menten	$v = \frac{Vmax \times [S]}{Km + X}$	[557]
Biphasic	$v = \frac{(Vmax_1 \times [S]) + (CLint \times [S]^2)}{Km_1 + [S]}$	[61]
Substrate Inhibition	$v = \frac{Vmax \times [S]}{Km + [S] + \frac{[S]}{Ki}}$	[529]
P450 Substrate Inhibition	$v = \frac{Vmax\,(\frac{1}{Ks} + \frac{\beta[S]}{\alpha KiKs})}{\frac{1}{[S]} + \frac{1}{Ks} + \frac{1}{Ki} + \frac{[S]}{\alpha KsKi}}$	[513]
P450 Substrate Inhibition 2	$v = \frac{\frac{Vmax[S]}{Km_1} + \frac{\alpha Vmax[S]}{Km_1 Km_2}}{1 + \frac{[S]}{Km_1} + \frac{[S]}{Km_2} + \frac{[S]^2}{Km_1 Km_2}}$	[512]

Chapter 8 The CYP450 protein microarray platform

8.1 Introduction

8.1.1 Overview of protein microarray technology

Protein microarrays are miniaturised assays for the high-throughput analysis of protein interactions and function. Based on the technology originally developed for DNA microarrays, protein microarrays are typically fabricated by depositing tiny droplets of protein sample, with volumes in the picolitre to nanolitre range, onto a chemically derivatised glass slide using a contact spotter or a non-contact microarrayer. The high-density in format principle allows thousands of different biomolecular interactions to be assayed in parallel, eradicating variation introduced from one experiment to the next and ensuring that all assays are essentially carried out under the same experimental conditions [558]. The low volumes and high throughput nature of these assays also make them both time and cost-effective, saving precious samples and expensive reagents.

Protein microarrays can be categorised into two groups: analytical microarrays and functional microarrays [559]. Analytical protein microarrays are routinely used to detect the presence and abundance of a target protein in a complex sample. Resembling miniaturised ELISA or sandwich assays, the capture molecules, typically antibodies, are arrayed onto the surface of a glass slide and used to probe a complex mixture of proteins (e.g. a cell lysate) [560]. Reverse-phase arrays, where the complex mixture of proteins is printed onto the slide surface and then probed using antibodies, are also used [561, 562].

A pivotal study in 2000 showed that proteins could be immobilised onto a glass slide while retaining their native structures and biological functions [563]. This provided the foundation for the development of a diverse range of functional protein microarrays used to study protein-protein, protein-lipid, protein-nucleic acid, antibody-antigen, protein-carbohydrate, receptor-ligand and enzyme-substrate interactions [564].

Ensuring the protein remains correctly folded and functional after immobilisation is challenging and is affected by the choice of surface chemistry. Non-specific absorption onto the surface can denature proteins and interfere with protein structure and function. The surface chemistry will also affect protein binding, the density and orientation of the protein and non-specific adsorption

of macromolecules onto the surface. There are a variety of immobilisation strategies that make use of different slide surfaces. These include the following: random covalent attachment onto epoxy, aldehyde, NHS or carboxylic ester surfaces; adsorption onto polyvinylidene fluoride (PVDF), nitrocellulose or polystyrene surfaces; embedding of proteins within gel-coated surfaces; and affinity attachment of His-tagged proteins to nickel-coated surfaces or biotinylated proteins to streptavidin-coated surfaces [559, 565, 566]. Affinity attachment is advantageous as the orientation of the protein can be controlled (directed immobilisation), giving a homogeneous protein layer whilst potentially ensuring functional regions are accessible rather than buried in the surface [567]. Keeping the proteins hydrated on the array surface is also important for retaining protein function whilst choice of binding buffer, slide wash and block steps as well as storage method for fabricated arrays will also affect the 3-D structure and functional-state of immobilised proteins [568]. Jonkheijm *et al* have reviewed microarray surface chemistry, protein attachment and the retention of protein function on surfaces in more detail [569].

Protein microarray detection strategies include both label free and label-dependent methods [570]. Of these, fluorescence is the most common detection method used due to its stability and sensitivity, as well as the availability of fluorescent scanners and dyes commonly used for DNA microarrays. Other label-dependent methods include chemiluminescence, colorimetry and radioactivity. Label-free methods include surface plasmon resonance imaging, mass spectrometry and atomic force microscopy. While fluorescent-based assays are simple, conducive to multiplexing and have a high signal-to-noise ratio, label-free methods eradicate the possibility of labelling interfering with protein interactions as well as quantitation difficulties arising from differences in labelling efficiency between proteins.

8.1.2 Enzyme activity arrays

Protein microarrays capable of detecting and characterising enzyme activity generally make use of fluorescently-labelled activity-based probes (including mechanism-based suicide inhibitors and other low molecular weight fluorescent affinity labels (FALs)) to selectively label enzymes based on their activity and screen for suitable inhibitors [571-573]. By including time points and varying the concentrations of activity based probe and inhibitor applied to different sub-arrays, kinetic parameters and inhibition constants can be determined [572, 574]. The analyte solutions are typically applied to individual replica sub-arrays that are separated by teflon coatings or by a

gasket with different chambers. However, while this fluorescent-based inhibitor-based approach is able to screen a large number of protein samples for activity, it does not monitor substrate turnover directly, but is rather an activity-based binding assay.

Drop-to-drop arrays, where one drop containing the enzyme is overlaid by drops containing the other reaction components (or *vice versa*) create an array where a completely separate reaction, with different analyte concentrations, can take place in each spot. Evaporation from the surface however becomes problematic at sub-nanolitre reaction volumes. Hydrophilic-in-hydrophobic micropatterned surfaces have been used to generate larger spots (30- 150nL) with high surface tension, reducing the effects of evaporation on kinetics [575]. This was, however, at the cost of the number of reactions per slide, reducing the high-throughput nature of the array. In another study, Sakakihara *et al* used a hydrophilic-in-hydrophobic micropatterned surface covered in a layer of oil to generate femtolitre droplets for single molecule enzymatic arrays [576].

Microfluidic chips, made up of large arrays of nanolitre to femtolitre chambers etched into a surface (e.g. glass, polymer of silica) can be used to monitor enzyme-dependent substrate turnover. [577]. Here, fluid flow is tightly controlled, allowing the sequential addition of reaction components whilst the chambers reduce problems of evaporation as well as cross contamination between reactions. These arrays are extremely sensitive; for example, a recent study generated single molecule kinetics for the turnover of a resorufin-based fluorescent substrate by horseradish peroxidase using femtoliter-sized fused silica chambers [578]. The disadvantage is that this technology requires special apparatus for etching chambers into surfaces and loading samples.

8.1.3 Human CYP450 enzyme microarrays

Miniaturising CYP450 assays is challenging due to the instability of human CYP450 enzymes and the instability of the CYP450:CPR complex. This complex has a K_d = 0.05 μM [92] and electron transfer from the reductase has a K_m = 0.2 μM [100] . Early efforts to miniaturise CYP450 assays involved immobilising CYPs to electrodes made from graphite, glassy-carbon or gold, with the electrode serving as an electron source, eliminating the need for co-immobilisation of a reductase. This approach has been used to develop CYP450 biosensors to study the redox properties of CYP450s and to improve the electrochemical performance of CYP450s for

technological and commercial applications [579, 580]. These approaches are however not readily amenable to high-throughput multiplexed assays and drug screening and there are some concerns about the timing of electron transfer from the electrode to the CYP450 affecting the reaction cycle.

Specific and highly sensitive fluorescent CYP450 substrates have facilitated medium-throughput assays, with reaction volumes of 5 to 250 µl, in 96- 384- and 1536- well plates e.g. [581]. Lee *et al* described one of the first noteworthy CYP450 protein microarrays in which baculosomal CYP450 and CPR protein preparations, together with a NADPH regenerating system, were encapsulated in an alginate matrix in a nanolitre-scale microarray format and used for the cell-based screening of drugs metabolised by CYP450s and their toxic metabolites [582]. In an extension of this work, alginate CYP450 microarrays were used for fluorescence-based kinetic and inhibition studies [381]. While these arrays were quantitative, they were not truly amenable to truly high-throughput studies due to the long gelation and aging times for enzyme encapsulation- required to stabilise proteins on the surface. A 30% to 60% decrease in enzyme activity compared with solution phase assays was observed in these studies.

In a more recent study, a highly miniaturised solution phase platform termed 'luminometic sub-nanoliter droplet-to-droplet array (LUMDA)', with greatly reduced reaction volumes (picolitre range) and no requirement for enzyme encapsulation has been developed [382]. In a proof of principle study, this approach was used to monitor the inhibition of liposomes expressing CYP3A4 by the prototypical CYP3A4 inhibitors, erythromycin and ketoconazole, as well as two new structurally related inhibitors. The assay was luciferase-based, using a 'pro-luciferin' CYP3A4 substrate to couple enzyme activity to light emission. Droplets containing different reaction components were sequentially overlaid onto a silicon dioxide surface using piezoelectric inkjet printing followed by incubation and luminometric detection. All reaction components were printed in buffers containing 30% (v/v) glycerol; the inclusion of 30% glycerol was reported to eradicate evaporation and stabilise droplets during the assay procedure.

Notably, the enzyme microarray approaches described above aim simply to miniaturise microsome or liposome-based CYP450 assays and do not immobilise the CYP or CPR proteins onto the surface; thus they are not true protein microarrays, as conventionally understood.

8.1.4 The P450 Biochip

The P450 Biochip, previously developed in the Blackburn laboratory by Dr Beeton-Kempen, is a fluorescence-based protein microarray platform developed to provide quantitative kinetic data for CYP450 drug metabolism [473]. This platform relies on the oriented co-immobilization of soluble recombinant CYP450 and it redox partner CPR onto a surface allowing functional interactions between these two partners in a membrane-, alginate- and lipid-free environment. The robotically printed arrays - with sub-nanolitre reaction volumes – in principle enable the high-throughput study of CYP450 metabolism since, with up to 1100 reactions per slide, many assays can be run in parallel.

Figure 8.1 shows a schematic of BCCP-tagged CYP3A4 immobilised to a streptavidin-coated microarray surface. Commercially available hydrogel-coated glass slides (Schott H-slides) were used for array fabrication. The hydrogel is made up of long chain polyethylene glycol (PEG) polymers covalently attached to the glass slide and functionalised with N-hydroxysuccinimidyl ester (NHS) groups [583]. This "non-stick" PEG layer prevents non-specific macromolecule adsorption onto the glass surface, whilst providing an aqueous-like surface environment that should help to stabilise specifically-immobilised proteins on the microarray.

The slides were coated with streptavidin, allowing the oriented immobilisation of biotinylated recombinant proteins onto the surface *via* non-covalent biotin-streptavidin interactions. Due to the highly specific, essentially irreversible interaction between biotin and streptavidin ($K_d = 10^{-15}$ M), unbound proteins can be washed off the slide without disrupting the bound proteins [584]. This allows the binding and purification of biotinylated CYP450 and CPR proteins from partially pure protein preparations or crude lysates in a single step [585, 586]. This directed immobilisation ensures that the CYP450 and CPR proteins are orientated away from the surface, whilst the PEG coating and flexible BCCP-linker allows the immobilised proteins some degree of movement, in principle facilitating interactions between neighbouring protein partners.

In this P450 Biochip format, fluorescent CYP450 substrates together with the cofactor NADPH are overlaid over the spots of immobilised proteins and the increase in fluorescent signal is monitored with time. An advantage of this platform is that array printing and detection can be

done using conventional DNA microarray technology (a contact printer with solid pins and a fluorescent laser scanner).

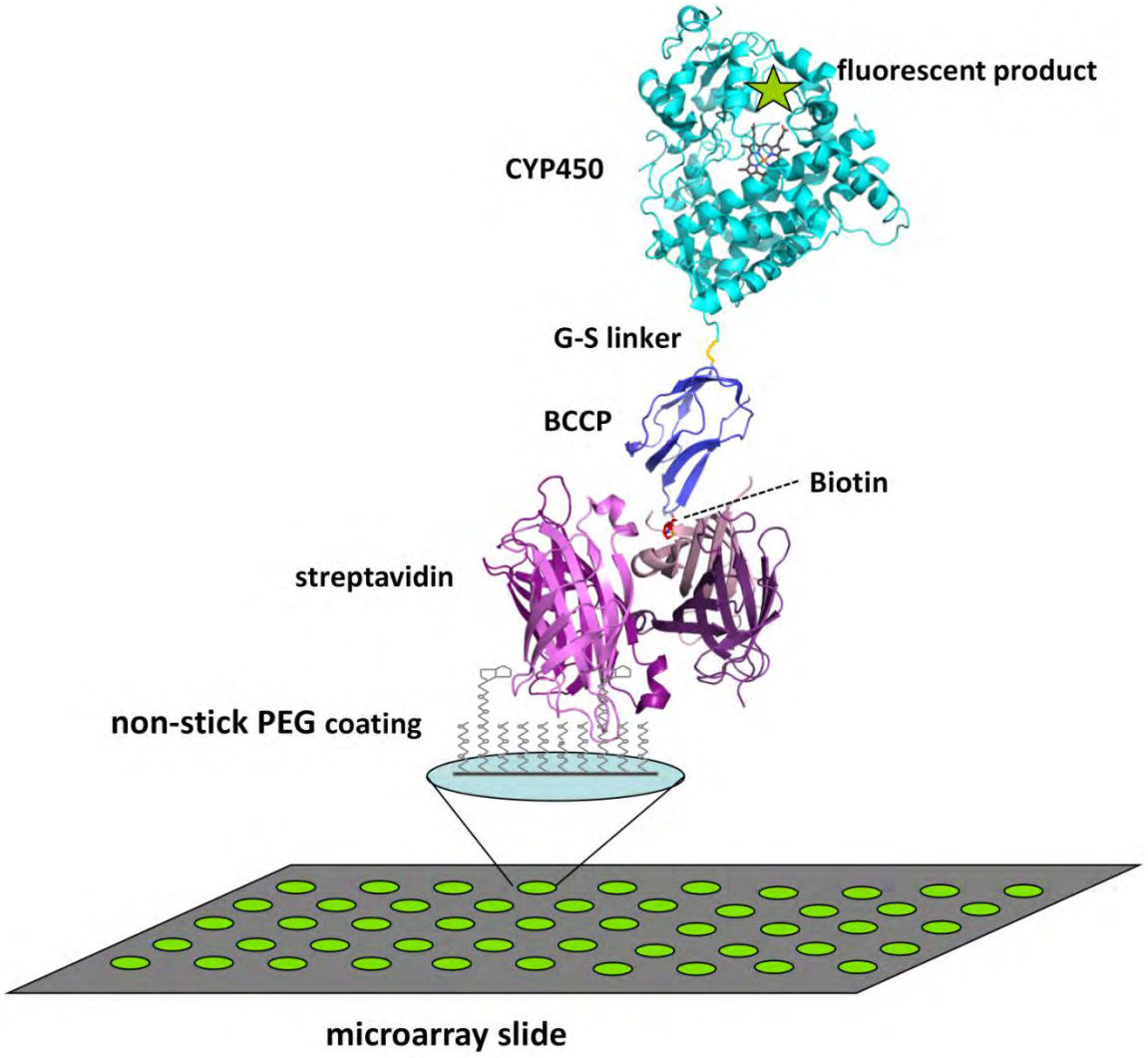


Figure 8.1 Schematic of a biotinylated BCCP-tagged CYP450 protein immobilised to a streptavidin-coated slide surface.

This platform was developed and optimised, by a previous student in the Blackburn laboratory, Dr Beeton-Kempen, using the same CYP3A4 wild-type and CPR constructs used in this Thesis and Biphasic kinetics, K_{m1} = 1.6 μM and V_{max1} = 0.110 nM/min, were observed for the metabolism of BOMR (substrate concentrations 0.75 - 43.25 μM) by CYP3A4 wild-type protein. These initial results were promising as the K_m determined was comparable to literature reports for the turnover of BOMR as well as to solution-phase CuOOH assays performed in the same study (K_m = 1.4 μM for substrate range 1 – 6 μM using the Michaelis-Menten model). The K_m reported for the array platform is also in agreement with the more recent results in Chapter 7

of this Thesis where biphasic kinetics for CuOOH mediated CYP3A4 turnover of BOMR with K_m = 1.8 μM were observed for substrate range 0.75 - 43.25 μM.

Problems with the P450 Biochip

Despite the encouraging data obtained for turnover of the BOMR substrate by CYP3A4 on the prototype array surface, inhibition studies using the potent CYP3A4 inhibitor ketoconazole on this platform gave unexpected data [473]: while the reported K_i values for ketoconazole usually fall between 4 nM and 180 nM, depending on the substrate and assay condition [587], ketoconazole concentrations between 0.1 nM and 100 μM apparently did not inhibit BOMR turnover on the P450 Biochip array surface at all; inhibition was only eventually observed at ketoconazole concentrations greater than 100 μM. The mechanism of inhibition of CYP3A4 by ketoconazole is still unclear, with different studies reporting competitive, non-competitive and mixed-type CYP3A4 inhibition [587, 588]; even so, the inhibitory profile obtained on the array surface did not conform to any of the inhibitory models previously reported in the literature. 15% (v/v) glycerol had been included in the array buffers to slow down evaporation and had been observed to be essential for substrate turnover on the array surface. Dr Beeton-Kempen thus concluded that glycerol may be interfering with ketoconazole inhibition (noting that glycerol is known to be able to bind the haem iron), raising questions as to whether glycerol was a suitable anti-evaporating agent [473].

One shortfall of the P450 Biochip platform is the inability to control evaporation from the array surface: while glycerol prevents spots drying out completely and retains some water due to its hygroscopic nature, rapid evaporation must still occur from the ~ 10 nanolitre reaction droplets. Notably the array printing was performed within a printing chamber where the humidity and hence evaporation could be controlled but the slide was then by necessity removed from the chamber for scanning and scanned every 5 min over 30 to 60 min during which time the humidity was not controlled. Thus, the influence of evaporation on substrate concentration and hence on reaction rates was not accounted for in the prototype P450 Biochip development work.

Interestingly, a study analysing single glycerol droplets in humid air streams showed that the glycerol droplets, with an average starting diameter of 10 μm, have ~ 20% (v/v) water content at 50% humidity after they reach equilibrium [589]; the time for a glycerol droplet to reach

equilibrium was found to be in the order of 1 min. This is in agreement with the graphs of % relative humidity *versus* % glycerol at equilibrium reported in [590] showing that at 50% relative humidity the glycerol content is 80% (v/v) and a relative humidity greater than 90% is required to maintain a water content of over 75% (v/v) at equilibrium.
However, evaporation from the PEG-coated P450 array surface is obviously much more complicated than this since the 3-D surface itself is hydrated and water molecules, as well as reagent components can diffuse from the droplet into the bulk surface and *vice versa* until an equilibrium is reached. Furthermore, proteins and other reagents within the spots will also affect the water content at equilibrium, making it very difficult overall to estimate the true effective concentration of reaction components on this platform.

Regardless of the true extent of evaporation, it seems reasonable to assume that printed droplets will have reached equilibrium on the array surface by the time they are place in the scanner and the first scan is performed. Consequently, the spot volume is likely to remain constant during the scanning period when the kinetic data is collected, but the droplet volume is likely to be significantly reduced, leading to an increase in substrate, glycerol, NADPH and salt concentrations.

As evaporation was not prevented on the P450 biochip and the effects of evaporation on substrate concentration were not taken into account, it appears serendipitous that the K_m for the turnover of BOMR Vivid substrate was comparable to the K_m observed in the literature and CuOOH mediated solution-phase assays. This Chapter aims to explore the anomalous ketoconazole inhibition results and the effect of glycerol and evaporation on substrate turnover on the array surface to ascertain whether the CYP450 Biochip is in fact suitable for testing the kinetic activity of multiple variant CYP450 proteins in parallel.

8.1.5 Aims and objectives

1. Test CPR activity in solution
2. Test the effect of increasing glycerol and sucrose concentration on activity and ketoconazole inhibition

3. Determine whether reliable kinetic parameters for NADPH CPR-dependent turnover of Vivid substrates by CYP3A4 and CYP2C9 polymorphic variants can be obtained on the P450 Biochip array surface.

8.2 Results

8.2.1 CPR activity

Prior to any array work, the activity of the solution-phase N-terminal truncated CPR protein expressed in *E.coli* (Chapter 5) was tested by monitoring the NADPH-dependant reduction of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) to blue-formazan by CPR (Figure 8.2). The formation of blue-formazan was measured spectrophotometrically by the increase in absorbance at 610nm.

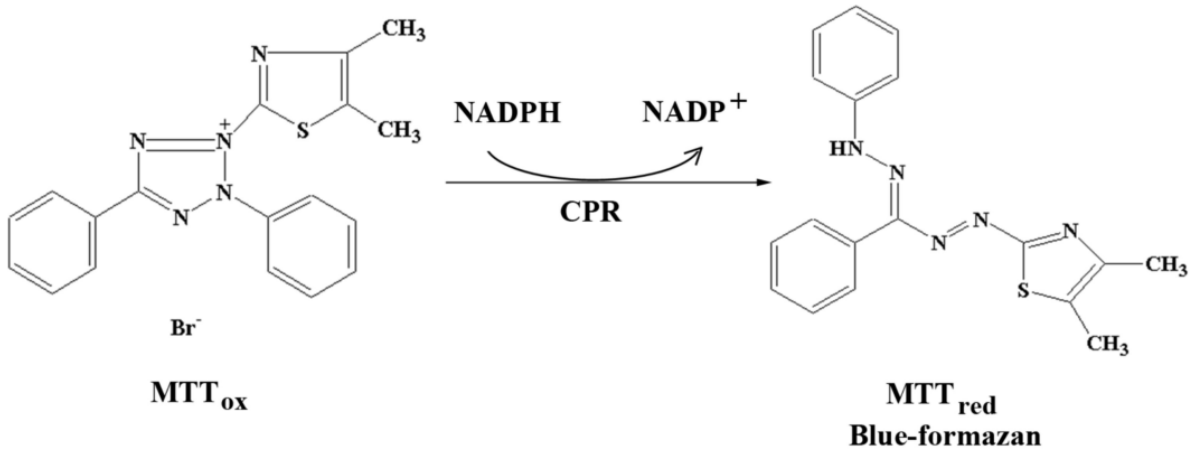


Figure 8.2 Reduction of MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) to blue formazan by CPR in the presence of NADPH (Figure from [591])

The His-tag purified recombinant CPR showed satisfactory activity, suitable for downstream CYP450 assays. MTT reduction by CPR followed Michaelis-Menten kinetics (Figure 8.3) with K_m = 109 ± 15 µM and k_{cat} = 858 ± 43 min^{-1}. The k_{cat} was comparable to results obtained previously for the same protein construct, as well as for the rat full length CPR, while the K_m was 2 and 5 fold higher respectively (Table 8.1).

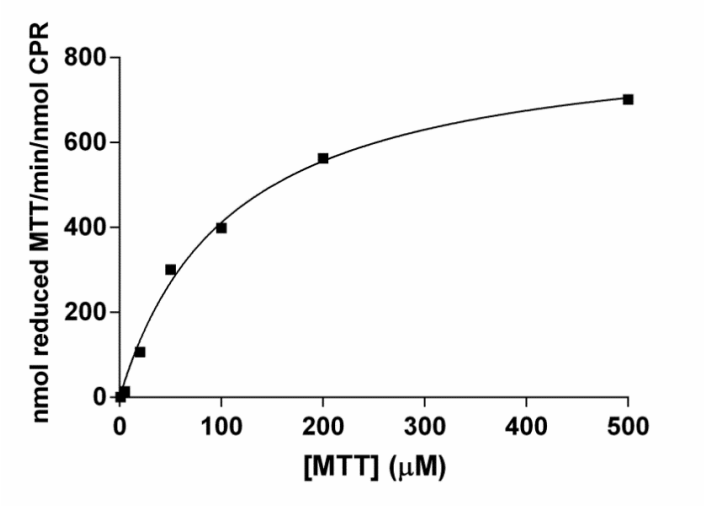


Figure 8.3 Michaelis-Menten model fitted to CPR kinetic data for the reduction of MTT in the presence of NADPH using non-linear regression. A concentration of ~20 nM CPR (partially purified protein, concentration estimated from protein gels) in 100mM potassium phosphate buffer pH 7.4 was used in each reaction and reactions were initiated by the addition of a NADPH regenerating system. The formation of blue-formazan was monitored at 610 nm for 30 s at varying MTT concentrations (0.1 - 500μM).

Table 8.1 Kinetic constants for the reduction of MTT by CPR

Protein construct	k_{cat} (min^{-1})	K_m (μM)	Reference
N-term truncated Human CPR	858 ± 43	109 ± 15	This work
N-term truncated Human CPR	766	54 ± 5	[473]
Rat full length CPR	1190 ± 40	20 ± 2	[591]

Specific activity observed in this work for partially protein = 4.5 U/mg

8.2.2 Testing the effect of glycerol on CYP450 activity and inhibition

To explore the reason for the very weak CYP3A4 inhibition by ketoconazole in the expected concentration ranges on the P450 Biochip array surface observed in previous work, activity assays systematically testing the effect of increasing glycerol concentrations on CYP3A4 activity and ketoconazole inhibition in solution were carried out using CuOOH to drive the reaction. Sucrose was also tested as an alternative anti-evaporation agent. Assays were carried out using 8μM DBOMF Vivid Green substrate and 100 nM ketoconazole.

Solution-phase CYP3A4 activity increased with glycerol concentration, peaking at 24% glycerol with activity more than 2 fold higher than in the absence of glycerol, but then began to decrease again as glycerol concentrations increased further (Figure 8.4 A). Activity was still higher in the presence of 60% glycerol than in the absence of glycerol; glycerol concentrations higher than

60% were not tested due to the difficulty of accurately pipetting viscous liquids but if the observed trend continued, higher glycerol concentrations would presumably further reduce CYP3A4 activity. By contrast, CYP3A4 activity decreased gradually with increasing sucrose concentrations (Figure 8.4 B): activity in the presence of 60% sucrose was 52% lower than in the absence of sucrose.

100 nM ketoconazole inhibited CYP3A4 activity by 70-80 % at lower glycerol concentrations, however ketoconazole's potency as an inhibitor decreased at higher glycerol concentrations: 100 nM ketoconazole showed only 40% inhibition in the presence of 60% glycerol. Increasing sucrose concentration did not have an effect on ketoconazole inhibition: 100 nM ketoconazole inhibited CYP3A4 activity by 70-82 % at all sucrose concentration tested.

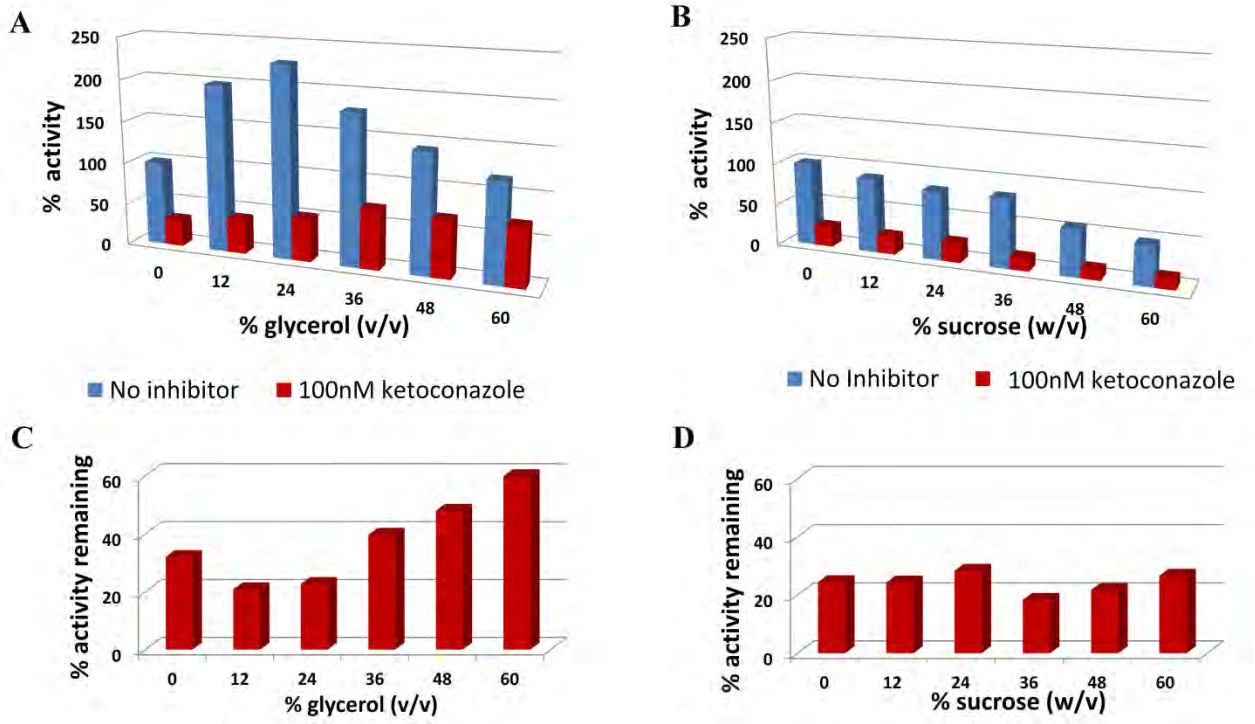


Figure 8.4 The effect of increasing glycerol and sucrose concentrations on CYP3A4 activity and ketoconazole inhibition.

75 nM CYP3A4 and 8 μM DBOMF in 100 mM potassium phosphate buffer was used in each reaction and the reactions were initiated with 0.5 mM CuOOH. The increase in fluorescence was monitored over the first 5 min of the reaction, for which the reaction was linear. **A** and **B** show the rate of CYP3A4 turnover of DBOMF at each glycerol or sucrose concentration, in the absence and presence of 100 nM ketoconazole, as a percentage of CYP3A4 turnover of DBOMF in the absence of glycerol/sucrose and inhibitor. **C** and **D** show the remaining CYP3A4 activity in the presence of 100nM ketoconazole as a percentage of the activity in the absence of ketoconazole at each glycerol or sucrose concentration.

Based on these results, glycerol is more suitable than sucrose for preserving CYP3A4 enzyme activity whilst reducing evaporation rates and even at concentrations exceeding 60% glycerol (v/v) activity levels should be similar to the activity observed in the absence of glycerol. Glycerol is however problematic for inhibition studies as it clearly interferes with ketoconazole inhibition. By comparison, sucrose did not alter ketoconazole inhibition, but high concentrations resulted in much lower levels of CYP3A4 activity.

8.2.3 CYP450 microarray assays: The importance of negative controls

A microarray assay was set up to test the kinetic activity of the 7 CYP3A4 polymorphic variant proteins including the wild-type using the original protocol established by Dr Beeton-Kempen. Briefly, a solution containing a mixture of partially purified biotinylated CYP3A4 and CPR proteins was printed onto a streptavidin-coated array surface and left in the humidified printing chamber for an additional 30 minutes to allow proteins to bind. The slide was then blocked with BSA plus free biotin and washed to remove unbound proteins. The spots containing bound protein were overlaid with a mixture of BOMR substrate and NADPH in a second print round and the increase in fluorescence was monitored in each spot over time using a fluorescent scanner. 15% glycerol was included in the printing buffer for the first and second print rounds. Two controls, not previously performed, were included at each substrate concentration: a "CPR only control" (CYP3A4 was excluded from the first round of printing) and a "no protein control" (both CYP3A4 and CPR were excluded from the first round of printing). The print layout is shown in the Methods, Section 8.4.3.

The kinetic profiles for the turnover of BOMR on the array surface by wild-type CYP3A4 and the negative controls are shown in Figure 8.5 A. Unexpectedly, both the "CPR only" and "no protein" control showed kinetic profiles comparable to the kinetic profile observed for "CPR + CYP3A4" suggesting that the substrate turnover observed was protein independent. Figure 8.5 B shows the results from identical assay repeated a few days later using fresh reagents. The magnitude and shape of the kinetic curves differed between the two assays, particularly at higher substrate concentrations, but again the kinetic profiles for the negative controls were comparable to the kinetics for CPR + CYP3A4. The data for the CYP3A4 variants was therefore not analysed as it was clear that the observed substrate turnover in these assays was not P450 dependent.

Additional controls at a single substrate concentration (5.7 µM) were also included on a separate section of the same slide on both occasions (Figure 8.5 C & D). "No NADPH", "no substrate" and "NADPH only" controls showed no significant increase in fluorescent signal with time, however the "BOMR + NADPH - no protein" control again showed a similar rate of substrate turnover to the positive control (CPR + 3A4 + BOMR + NADPH).

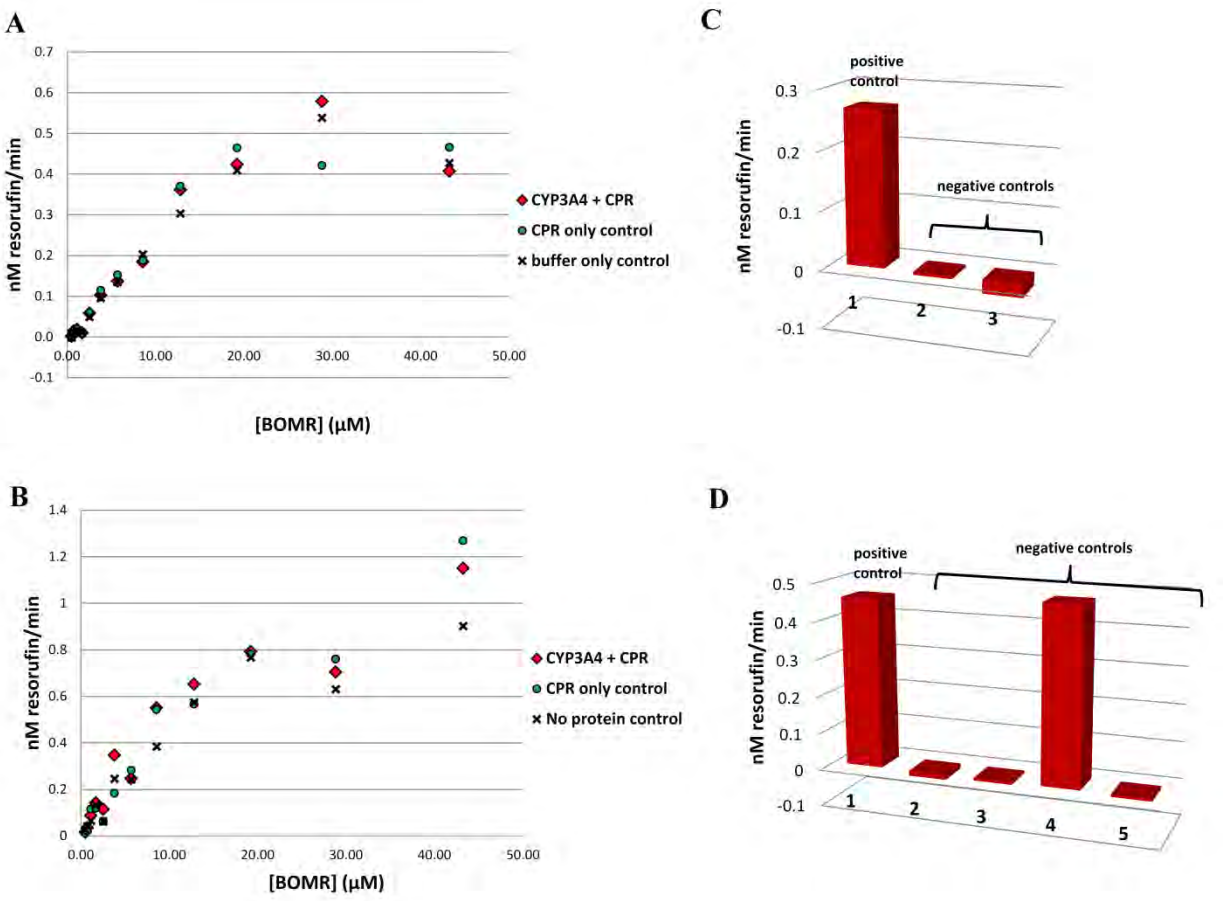


Figure 8.5 Turnover of BOMR substrate on microarray surface.
A) & **B)** Kinetic profiles for the turnover of BOMR (0.5 – 43.3 µM) on the array surface for reactions containing 3A4 and CPR and two negative controls: "CPR only" and "no protein". Graphs A and B show results from two identical experiments carried out on different days. Reactions were carried out in 100mM potassium phosphate buffer pH 7.4, 15% glycerol and initiated with BOMR and 1mM NADPH. The linear reactions were monitored for 40 min and velocities were calculated based on the average of 3 to 7 replicates. **C)** & **D)** Additional controls included on a different section of the same slide as experiments in A and B, respectively: 1) 3A4 + CPR + 5.7 µM BOMR + 1mM NADPH; 2) 3A4 + CPR + 5.7 µM BOMR, no NADPH; 3) 3A4 + CPR + NADPH, no substrate; 4) 5.7 µM BOMR + 1mM NADPH, no protein; 5) 1mM NADPH only.

Due to the high volumes of data produced in a single assay, signals from replicate spots were averaged at each time point and the averaged signal was used to plot RFU versus time graphs, rather than generating individual RFU versus time graphs for each spot and then averaging the

rates. Consequently, error bars are not shown on the graph but rather the percentage median coefficient of variation (CV) calculated across all time points for each sample is used to show the experimental error associated with the data (Table 8.2 & Table 8.3).

Table 8.2 Percentage median CVs calculated for the data shown in Fig 7.5 A and B

	Assay A			Assay B		
µM BOMR	**3A4 + CPR**	**CPR only**	**No protein**	**3A4 + CPR**	**CPR only**	**No protein**
0.50	7.3	7.9	38.6	11.5	12.1	17.0
0.75	5.6	3.8	22.4	20.5	20.2	47.0
1.13	4.7	6.0	38.2	23.1	12.2	17.8
1.69	8.1	6.7	38.4	12.3	4.0	22.5
2.53	4.3	4.1	7.7	18.2	16.5	21.0
3.80	3.4	2.5	6.3	15.8	6.9	22.0
5.70	4.0	2.2	6.4	16.1	14.6	14.2
8.54	3.3	3.3	5.0	17.0	9.8	21.6
12.81	2.9	3.3	10.3	6.6	2.6	13.4
19.22	3.1	4.0	6.2	3.7	11.5	11.6
28.83	3.6	24.0	4.0	7.8	8.5	21.0
43.25	2.8	5.9	6.8	15.0	3.9	31.0
Average	**4.4**	**6.1**	**15.9**	**14.0**	**10.2**	**21.7**

Table 8.3 Percentage median CVs calculated for the data shown in Figure 8.5 C and D

5.7 µM BOMR		**Assay A**	**Assay B**
Control 1	positive	2.8	9.5
Control 2	no substrate	4.8	7.4
Control 3	no NADPH	19.2	27.9
Control 4	no protein	na	13.6
Control 5	NADPH only	na	36.8

The variations between replica spots in the first assay (Assay A) was very low for samples at all substrate concentrations (CV < 10%) with the exception of the "no protein" control in the low substrate concentration range where CV values were as high as 38%. The variation between replica spots in the second assay (Assay B) was generally higher than the first assay and the "No protein" spots showed higher inter-spot variation than the other samples. Control samples, where very low or no increase in fluorescence was observed, also had higher CVs but this is expected due to a low signal to noise ratio.

The CVs in both assays were calculated based on the following number of replicate spots: "3A4 + CPR", 7 spots; "CPR only", 3 spots; "No protein"; 6 spots. The "3A4 + CPR" and "CPR only" were printed consecutively in the same area of the sub-array with no wash steps in-between. The "No protein" spots were scattered throughout the sub-array with other samples and pin wash steps were carried out in between them (see Methods, Figure 8.15 for print layout). This suggests that the high CV values for "No protein spots" may result from the varied position/time of printing of the replicate spots within the sub-array.

To confirm and investigate this apparent protein independent turnover on the array surface, another microarray assay, testing different combinations of reaction components, was carried out. Reaction components were printed onto two slides in parallel: Slide 1 was processed as usual with BSA in the blocking solution; whereas for Slide 2, BSA was excluded from blocking solution. The results from this assay confirmed that the BOMR substrate turnover on the slide - was NADPH dependent and protein independent (Figure 8.6). All combinations of reaction components containing both NADPH and substrate showed similar rates of substrate turnover that increased with increasing substrate concentration. While the trends were similar on both slides, the rate of substrate turnover was lower on slide 2.

The median CVs for replicate reactions on slide 1 and 2 are shown in Table 8.4. The variation between replicate spots for slide 1 was satisfactory: all median CVs < 20% even at low substrate concentrations, the average median CV was ~7% and all median CVs > 10% were for samples that had ≥ 42 replicate spots. The variation between spots on slide 2 (no BSA in blocking step) was higher than for slide 1 (blocked with BSA): all median CVs were ≤ 35% and the average median CV was 13.4%. The spots on slide 2 were printed immediately after the corresponding spot on slide 1 using the same pins and samples.

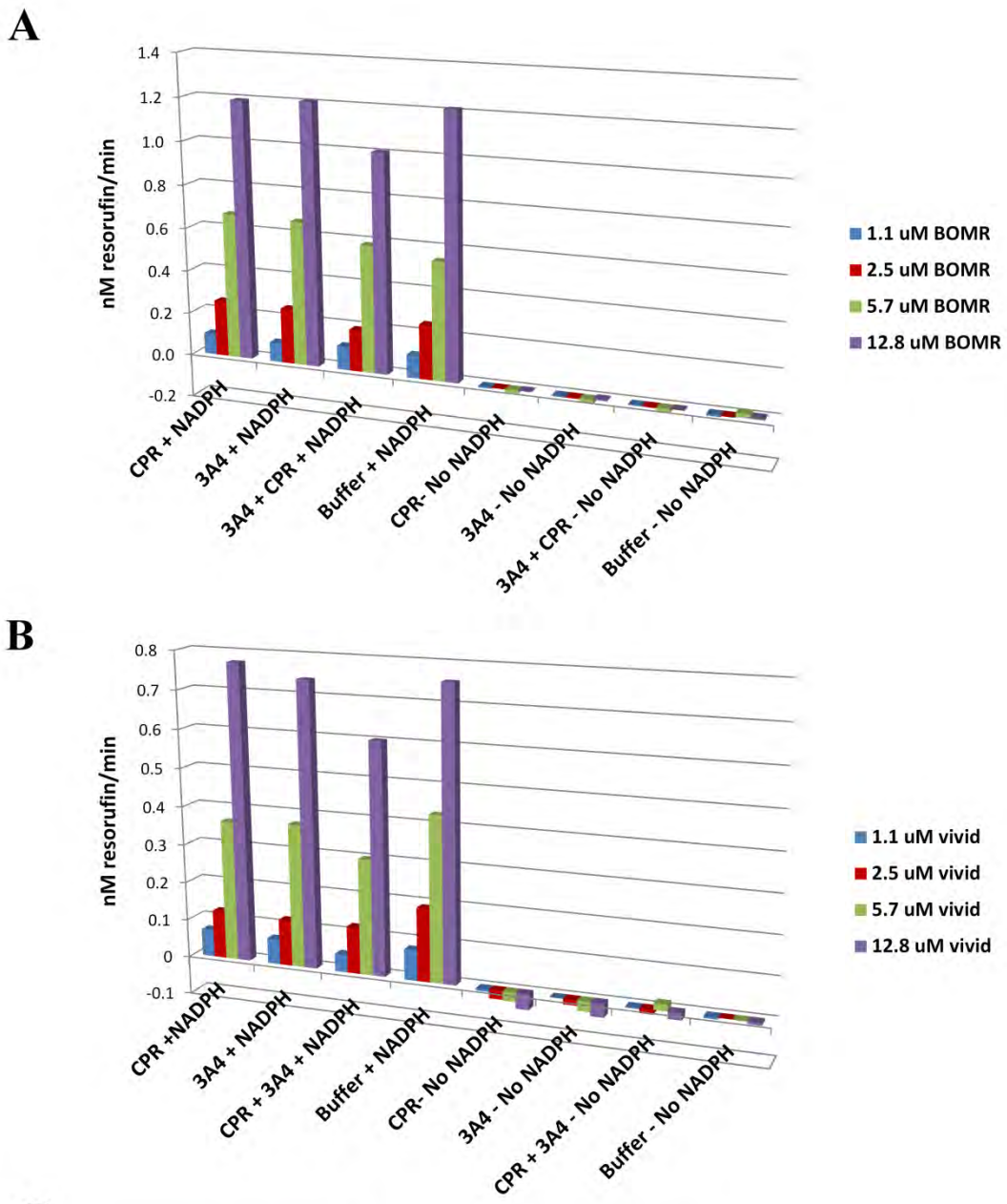


Figure 8.6 NADPH dependent BOMR turnover on the microarray surface. The graphs shows the rate of resorufin formation on the array surface for different combinations of the reaction components at varying BOMR substrate concentrations (1.1 – 12.8 µM) on **A)** slide 1, with BSA in blocking solution and **B)** slide 2, without BSA in blocking solution. The reactions were carried out in 100mM potassium phosphate buffer pH 7.4, 15% glycerol and the linear reactions were monitored for 45min. Each bar represents the average of at least 6 replicate spots. **C)** Slide 1 scanned at time 0 and after 45 min.

Table 8.4 Percentage median CVs calculated for the data shown in Figure 8.6 A and B

Reaction components	No of replicates	% Median CV				
		1.1 µM BOMR	2.5 µM BOMR	5.7 µM BOMR	12.8 µM BOMR	Average % median CV
SLIDE 1 - with BSA blocking						
CPR + NADPH	6	7.6	3.6	3.9	5.8	**5.2**
3A4 + NADPH	6	6.2	2.7	3.6	4.1	**4.1**
3A4 + CPR + NADPH	42	7.7	16.7	11.9	14.1	**12.6**
Buffer + NADPH	60	6.0	18.2	6.3	4.5	**8.7**
CPR - No NADPH	6	5.9	2.9	4.0	2.4	**3.8**
3A4 - No NADPH	6	6.5	4.1	3.6	3.3	**4.4**
3A4 + CPR - No NADPH	42	10.7	7.2	6.7	5.0	**7.4**
Buffer - No NADPH	60	9.5	7.4	7.8	5.3	**7.5**
SLIDE 2 - no BSA blocking						
CPR + NADPH	6	18.2	7.6	7.2	4.7	**9.4**
3A4 + NADPH	6	16.5	6.4	3.6	4.8	**7.8**
3A4 + CPR + NADPH	42	12.7	13.3	13.3	14.7	**13.5**
Buffer + NADPH	60	18.1	17.2	13.6	35.1	**21.0**
CPR - No NADPH	6	8.2	4.3	6.5	4.7	**5.9**
3A4 - No NADPH	6	10.8	4.5	8.2	4.9	**7.1**
3A4 + CPR - No NADPH	42	16.9	16.9	15.0	16.0	**16.2**
Buffer - No NADPH	60	31.6	23.8	30.1	20.8	**26.6**

Variation between sub-arrays

To determine the reproducibility of the array data across the slide surface, replicate spots at the same position in different sub-arrays printed by different pins at the same time were examined. The data for the "Buffer + NADPH" sample at each substrate concentration in Figure 8.6 above was an average of replicate spots from the same sub-array (sub-arrays 5 – 8). Six addition "Buffer + NADPH, no protein" spots were also printed in sub-arrays 1- 4. (See Methods, Figure 8.16 for slide layout). Figure 8.7 compares the rate of substrate turnover of "Buffer + NADPH, no protein" spots at varying substrate concentrations calculated from spots in row 10, sub-arrays 1 to 4 and spots in row 10, sub-array 5 – 8 for slide 1 and slide 2.

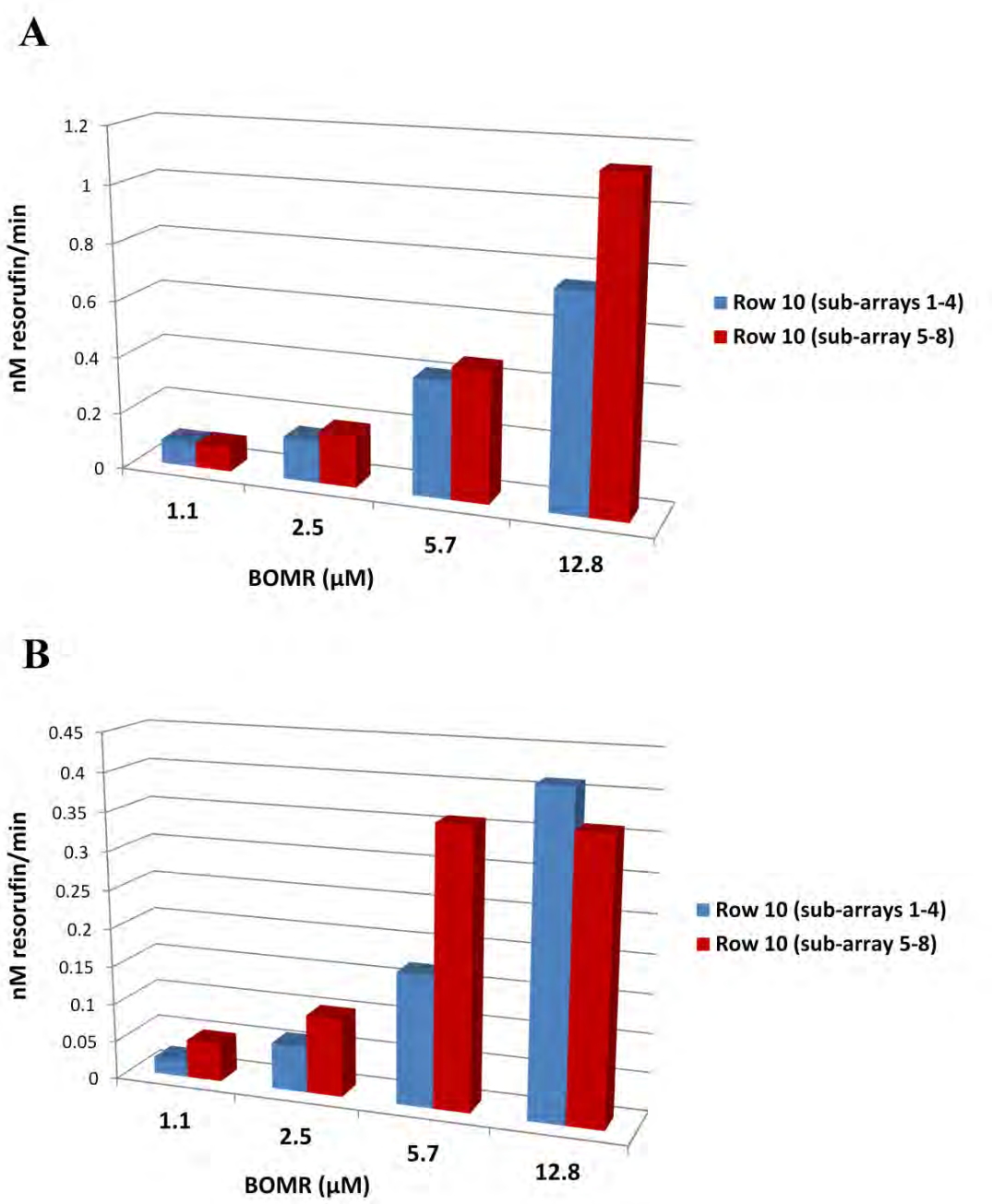


Figure 8.7 Inter-sub-array variability for the turnover of BOMR on the slide surface. P450 storage buffer only spots were printed in the first print round; spots were overlaid by 1mM NADPH and 1.1 -12.8 µM BOMR Vivid substrate in the second print round. Each bar represents the average of 6 replicate spots printed within the same row by the same pin. Panel A is data from slide 1; Panel B is data from slide 2.

The variability between data in different sub-arrays was between 3 and 56 %. Spots in sub-arrays 5 to 8 showed higher turnover at almost all substrates concentrations on both slides and as seen previously, slide 2 generally showed higher variability than slide 1; other than this, there were no obvious trends.

Variation within sub-arrays

Variation between replicate spots within a sub-array was also examined (Figure 8.8). All spots in a sub-array are printed from the same well by the same pin, one after the other. There was a 1% to 45% difference in the rate of BOMR turnover between the rows within sub-arrays on slide 1; the largest variation was seen at the lower substrate concentrations while data at the highest substrate concentration showed very little variation. There was a 3% to 54% difference in the rate of BOMR turnover between rows within sub-arrays on slide 2 and high variation was seen at both high and low substrate concentrations. The variability on both slides was generally the highest for rows furthest apart.

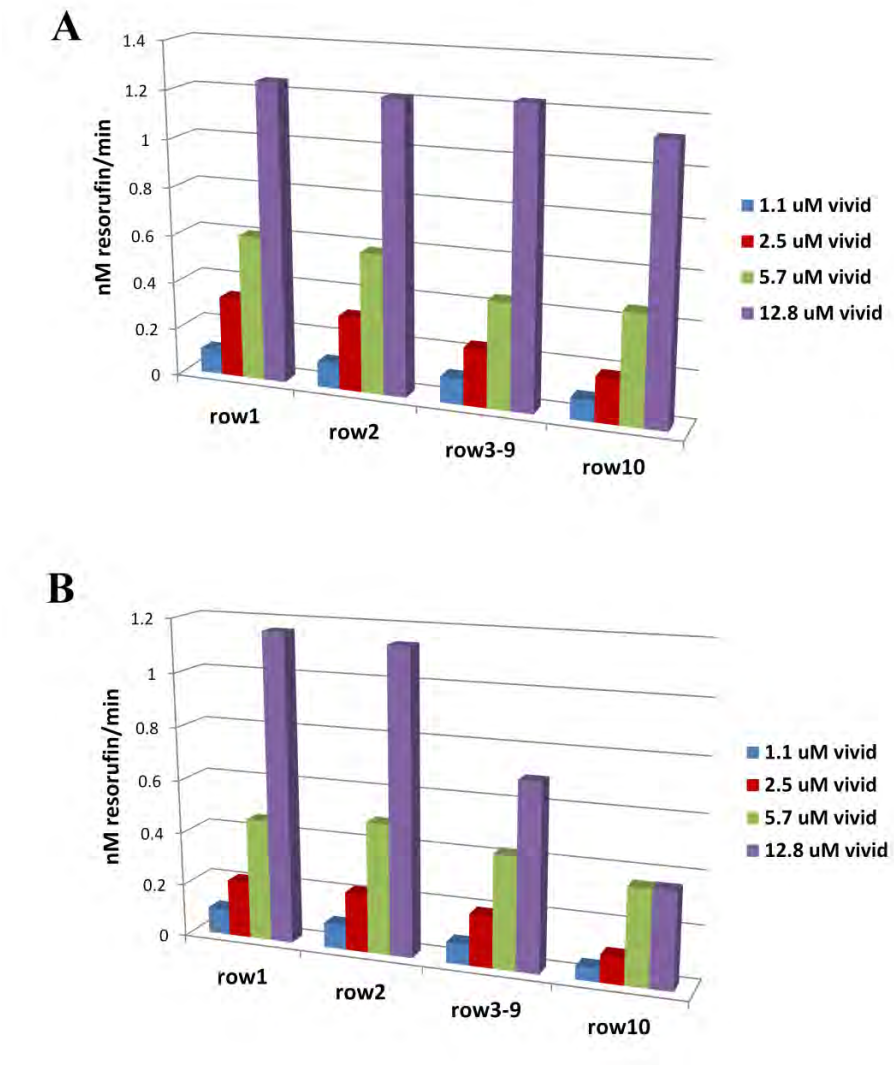


Figure 8.8 Intra- sub-array variability for the turnover of BOMR on the array surface. P450 storage buffer only spots were printed in the first print round; spots were overlaid by 1mM NADPH and 1.1 -12.8 µM BOMR Vivid substrate in the second print round. Each row contained 6 replicate spots at each substrate concentration. All rows were printed by the same pin one after the other, starting with row 10. No wash steps were included between replicate spots. Panel A is data from slide 1; Panel B is data from slide 2.

8.2.4 Further testing of fluorescent substrates in solution

Additional solution phase assays were performed in an attempt to gain insight into the protein-independent turnover observed on the array surface and to test whether any NADPH-CPR mediated CYP450 activity could be observed in solution. The resorufin- and fluorescein-based CYP3A4 and CYP2C9 Vivid substrates were tested (Figure 8.9 A-D). A positive control showing CuOOH mediated CYP3A4 or CYP2C9 activity was included for each substrate, as well as the following negative controls with and without NADPH: 1) No protein, 2) CPR only, 3) CYP3A4 or CYP2C9 only.

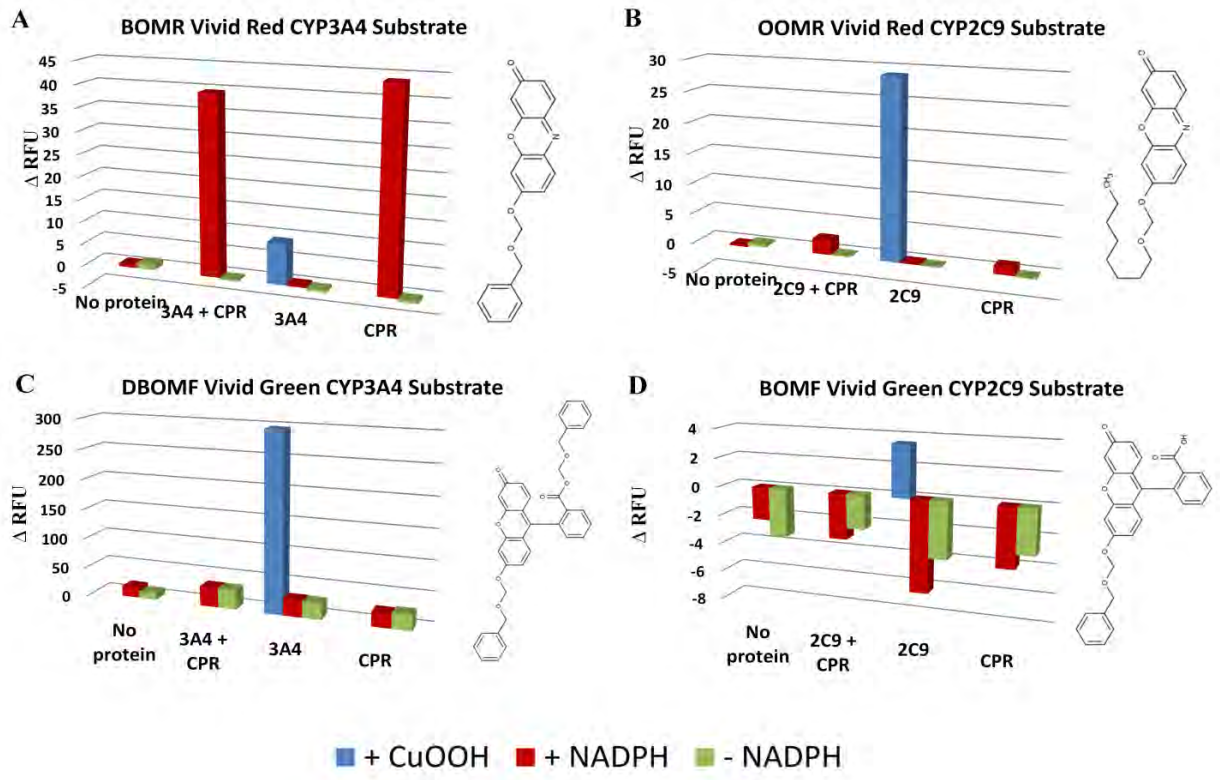


Figure 8.9 Testing the turnover of CYP3A4 and CYP2C9 Vivid fluorescent substrates in solution

Turnover of resorufin-based Vivid substrates BOMR (A) and OOMR (B) and the fluorescein-based Vivid substrates DBOMF (C) and BOMF (D) by different components in the CYP3A4 and CYP2C9 reaction setups. The blue bars in each graph represent "P450 + substrate + CuOOH", serving as the positive control. The red bars in each graph represent reactions with NADPH as follows: "No protein + substrate + NADPH", "P450 + CPR + substrate + NADPH", "P450 only + substrate + NADPH" and "CPR only + substrate + NADPH''. The green bars in each graph represent reaction without NADPH as follows: "No protein + substrate", "P450 + CPR + substrate", "P450 only + substrate" and "CPR only + substrate''.5.7 µM Vivid substrate was used for each reaction and the total increase in fluorescence after 5 min was plotted. The concentrations of other components in the reactions, where applicable, were as follows: 250nM CYP3A4 (total protein concentration), 250nM CYP2C9 (total protein concentration), 250nM CPR (total protein concentration), 0.5mM CuOOH and 1mM NADPH. The reactions were carried out in 100mM potassium phosphate buffer, pH 7.4, 15% glycerol. The structure of each Vivid substrate is shown on the right of the graph.

No protein independent substrate turnover above background was observed for any of the substrates. However, NADPH dependent CPR turnover of resorufin-based substrate BOMR, observed in the absence of CYP3A4 (CPR control), was four times higher than that observed for CuOOH mediated CYP3A4 turnover of BOMR (positive control). All negative controls for the other three substrates showed no or very low increases in fluorescence compared with the positive control.

These results confirmed that the protein-independent turnover was specific to the microarray platform. Several other attempts were made to gain further insight into the substrate independent turnover occurring on the microarray surface. Analogous assays were performed in Nunc immobilizer streptavidin-coated 96 well plates with a PEG surface similar to the microarray slides. Elevated NADPH, glycerol and potassium phosphate concentrations (10 nM, ~60% and 1M respectively) were tested in an attempt to mimic the effect of evaporation on the slide surface. The streptavidin-coated plates were also tested as an alternative platform for obtaining CPR mediated CYP3A4 activity using the less problematic DBOMF substrate. No protein independent turnover or CPR dependent CYP3A4 turnover was observed in any of these assays (results not shown).

8.3 Discussion

The utility of a P450 Biochip for generating CYP450 kinetic data was previously exemplified using recombinant human wild-type CYP3A4 and CPR protein [473]. On the grounds that previous kinetic results looked promising despite concerns that glycerol may be interfering with enzyme inhibition, a series of experiments were carried out here aimed at improving the platform, with the goal of using it to obtain kinetic data for the panel of CYP3A4 and CYP2C9 polymorphic variants. Based on the spectrophotometric assay monitoring the reduction of MTT by CPR and CuOOH mediated solution-phase CYP450 assays (Chapter 7), biotinylated BCCP-tagged CPR and CYP3A4 proteins were shown to be catalytically active and thus deemed suitable for protein microarray work.

Glycerol's effect on CYP3A4 activity and inhibition: implications for the microarray platform

Glycerol is routinely used in protein microarrays due to its hygroscopic nature, low volatility and protein stabilising properties. In particular, it is thought to maintain a solvated environment for proteins/peptides in small droplets on array surfaces and to act as a water mimic, capable of forming multiple hydrogen bonds.

To gain insight into the possible effects of glycerol on the CYP450 Biochip, the effect of increasing glycerol concentrations on CYP3A4 activity and ketoconazole inhibition was tested using CuOOH mediated solution-phase assays in microwell plates - with 100 µl reaction volumes and a 5 min assay duration, evaporation was assumed to be negligible. Increasing glycerol concentrations between 0 and 24 % (v/v) lead to increased CYP3A4 activity and glycerol concentration in this lower range did not affect ketoconazole inhibition. At glycerol concentrations ≥ 36% (v/v), however, CYP3A4 activity began to decrease again and the potency of ketoconazole inhibition declined with increasing glycerol concentrations. The increase in CYP3A4 activity observed at lower substrate concentrations is consistent with results in Chapter 7.

Interestingly, a report in the literature showed that glycerol causes CYP11A1 to shift towards the low spin state [533]. In the presence of sub-saturating glycerol concentrations (20% and 35%), a

high-spin inducing substrate shifted the protein to a high-spin state but in the presence of 65% glycerol, the addition of high-spin inducing substrate did not cause a spin-shift. High glycerol concentration (> 35%) may reduce activity by affecting the structural flexibility of the enzyme and preventing water molecules crucial for catalytic activity from entering the active site. High viscosity may also have an effect on the diffusion rates.

High glycerol concentrations have also been reported to reduce the binding affinity of ketoconazole in other related studies. Arrabito *et al* observed that ketoconazole had an IC_{50} more than 4 fold higher in the liposome-based LUMDA CYP3A4 array, where printing buffers contained 30% glycerol, than in microwell assays with no glycerol [382]. The authors claimed that 30% (v/v) glycerol keeps the water content in the spots constant and resolves evaporation issues. Using a stroboscopic imaging tool present in the inkjet printer they showed that final spots do not change shape or size for more than 8 hours. However while the spots in that study were stable after a glycerol:water equilibrium is reached, it seems inevitable that evaporation must have occurred extremely rapidly as the picolitre spots were printed; so while the glycerol content remained constant during the experiment, it is likely to be higher than expected (> 30%). If this is true, based on the results from this Chapter, the elevated glycerol concentration (> 30%) explains the reduced inhibitory effect of ketoconazole observed on the LUMDA-chip.

The inclusion of sucrose as an alternative anti-evaporation agent did not have the same effect as glycerol on the activity of CYP3A4: increasing sucrose concentrations reduced CYP3A4 activity and had no effect on the potency of ketoconazole inhibition. Different osmolytes interact differently with proteins and water molecules due to varying properties such as size and polarity [592, 593]. Perhaps importantly here, while sucrose generally stabilises proteins more than glycerol, glycerol is a much smaller molecule and is able to access regions surrounding the protein that are inaccessible to sucrose.

It appears that, in the case of the P450 Biochip, if evaporation was effectively controlled to ensure that glycerol concentration remains constant at < 30% (v/v), inclusion of glycerol in the array spotting buffer would be beneficial, stabilising the proteins and increasing activity without interfering with inhibition. Controlling or accounting for evaporation from the array surface, while keeping reaction volumes low to maintain a truly high-throughput assay format, is however by no means trivial. Overlaying immobilised proteins with substrate-NADPH mixes

encapsulated in water-in-oil emulsions or using a hydrophilic-in-hydrophobic micropatterned surface covered in a layer of oil as described by Sakakihara *et al* may be possible solutions to this problem.

NADPH-dependent, CYP450 and CPR-independent turnover on the microarray surface: Are Vivid CYP450 substrates suitable for use on the CYP450 microarray platform?

Inclusion of a "no protein control" revealed that the resorufin-based CYP3A4 substrate, Vivid Red BOMR, was oxidised to its fluorescent product resorufin in the absence of CYP3A4 and CPR on the microarray surface. This reaction was found to be NADPH-dependent and specific to the microarray platform; there was no protein-independent turnover in the corresponding reactions in solution, even when reactions were carried out in streptavidin-derivatised PEG-coated 96 well plates with elevated NADPH, glycerol and substrate concentrations.

Subsequent microarray experiments, carried out by another student in the Blackburn lab, have also shown NADPH–dependent protein-independent turnover of resorufin-based CYP2C9 substrate, OOMR [594], suggesting that resorufin-based substrates are incompatible with the microarray format. Unfortunately, the microarray scanner, optimised for DNA microarrays, only has two standard lasers λex = 532 nm and λex = 635 nm and a range of emission filters optimised for conventional dyes, Cy3 and Cy5. Consequently, the turnover of fluorescein-based substrates, DBOMF and BOMF, could not be monitored using the microarray scanner. Confocal laser scanning microscopy has a much wider range of excitation and emission wavelengths and offers an alternative method for monitoring substrate turnover. This method of detection was however only conducive to endpoint assays due to practical limitations such as long scanning times (~20 min per scan) and the proximity of the confocal microscope to the microarray printing equipment. Preliminary hand-spotted arrays, monitored using confocal microscopy, tested the CPR mediated CYP3A4 turnover of DBOMF, CYP2C9 turnover of BOMF and CYP2D6 turnover of coumarin-based Vivid substrate EOMCC [594]. CPR mediated CYP3A4 turnover of DBOMF showed promising results in this assay, however kinetic curves for CPR mediated CYP2C9 turnover of BOMF and CYP2D6 turnover of EOMCC had a similar or lower magnitude to the negative controls included in those assays [594].

The instability of Vivid substrates in aqueous solution, mentioned in Chapter 7, may also be problematic for microarray assays. The time from dilution of substrates to the end of the assays

is typically over 2 hours since it includes preparing and aliquoting samples into the 384 well sample plate for second round printing, printing the substrates (can take 20 min or longer depending on the complexity of the print pattern and number of pin wash steps) and scanning. Autoxidation of substrates over the duration of the assay may decrease the substrate concentration resulting in apparent elevated K_m values.

NADPH CPR-dependent turnover of BOMR

Further testing of Vivid substrates in solution also revealed the NADPH-dependent, CYP3A4-independent turnover of BOMR by CPR. This CPR dependent turnover was ~ 5 fold higher than the CuOOH mediated CYP3A4 turnover of BOMR and probably immediately explains the very high relative k_{cat} observed for baculosomal NADPH-CPR mediated CYP3A4 turnover of BOMR (compared with CuOOH mediated CYP3A4 turnover of BOMR in solution; 400 fold difference between baculosomal and solution phase assays for BOMR) but not for the DBOMF substrate (only a 10 fold difference in k_{cat}s for DBOMF using the same enzyme preparations) observed in Chapter 7 (see discussion in Section 7.3.2). The concentration of CPR in the commercial baculosomal CYP3A4 reagent is not specified, but it is likely that a large proportion of the observed BOMR activity is in fact CPR-dependent rather than CYP3A4 dependent. Notably, there are no other reports of NADPH-CPR dependent turnover of BOMR to date, most likely because CPR and CYP3A4 are co-expressed in microsomal and baculosomal protein preparations making it difficult to perform a "CPR only" control.

There are however studies reporting the NADPH dependent reduction of resorufin and 7-alkoxyresorufins to colourless non-fluorescent compounds by CPR [595, 596] which revealed the redox cycling of resorufin in the presence of CPR: CPR catalyses the one-electron reduction of resorufin followed by either a second one-electron reduction or a one-electron oxidation by molecular oxygen.

The results in this Chapter highlight the importance of the "CPR only" control and clearly show that BOMR is unsuitable for determining CPR-mediated CYP450 activity. However, CPR did not turn over either of the fluorescein-based Vivid substrates, DBOMF or BOMF, or the other resorufin-based substrate OOMR.

Possible mechanisms for CYP450-independent turnover of Vivid substrates

An initial hypothesis to account for the NADPH-dependant, CYP450- and CPR-independent turnover of resorufin-based Vivid substrates on the microarray surface was that it was due either to catalysis by BSA (present on the surface as a blocking agent) [597, 598] or to iron(II)-mediated GIF- or Fenton-type biomimetic chemistry (the possible iron source being from corrosion of the stainless steel pins used for printing); however, neither of these possibilities were supported by the subsequent controls that were run. A more likely hypothesis therefore is that transient, reversible reduction of BOMR by NADPH - generating a charged intermediate - enhances electrostatic interactions with the positively charged surface, thus promoting surface-mediated general acid catalysis of the acetal group (Figure 8.10). The PEG-coated array surface is likely to be positively charged due to the known ability of PEG to coordinate divalent ions. In support of this latter hypothesis, subsequent work from the Blackburn laboratory showed that reactions carried out on more strongly charged amine surfaces gave significantly higher rates of surface-dependent BOMR turnover than the PEG-coated surfaces [594]. However, further work is required to elucidate the precise mechanism of NADPH-dependent surface mediated substrate turnover.

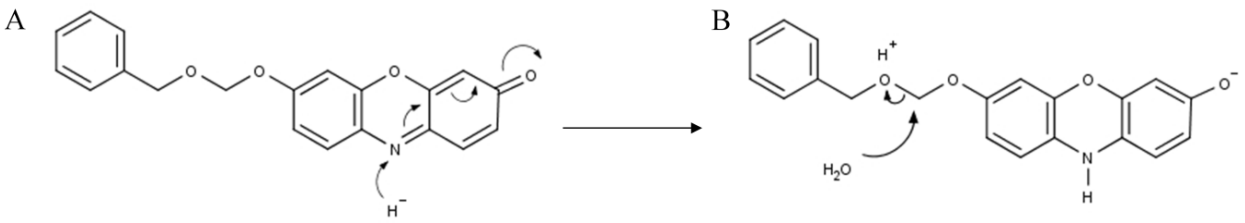


Figure 8.10 Proposed mechanism for NADPH-dependent CYP450 and CPR-independent turnover of BOMR on the array surface. **A)** NADPH transiently reduces the resorufin ring of BOMR yielding a negatively charged species. **B)** The negatively charged species associates with the positively charged surface promoting hydrolysis of the acetal linkage by general acid based catalysis leading to the formation (after re-oxidisation) of resorufin.

To account for the NADPH-dependent turnover of BOMR by CPR observed in solution, it seems possible that either (1) BOMR binds into a vacant FAD binding site in CPR due to molecular mimicry of the aromatic heterocycles of the flavin moiety by BOMR (Figure 8.11), or (2) BOMR binds non-specifically in the cleft between the NADPH and FAD binding domains in CPR. Either way, it then seems plausible that substrate binding to CPR leads to reduction of the resorufin moiety by NADPH, followed by oxidation of the substrate side chain by a bound

flavin, causing side chain cleavage and the formation (after re-oxidation) of resorufin. To account for why CPR metabolises BOMR but not the closely related OOMR substrate, at least two possibilities exist: (i) This may reflect fundamental chemical reactivity differences, since the benzylic position on BOMR is easier to oxidise than the corresponding position on OOMR (Figure 8.11); or (ii) This may reflect enzyme substrate specificity (perhaps BOMR has a higher affinity than OOMR for the FAD pocket or the cleft between the FAD and NADPH binding domains). Of these two possibilities, the explanation based on differences in chemical reactivity seems the more likely *a priori*, but further experiments would be needed to elucidate the precise mechanism of CPR-dependent substrate turnover observed here.

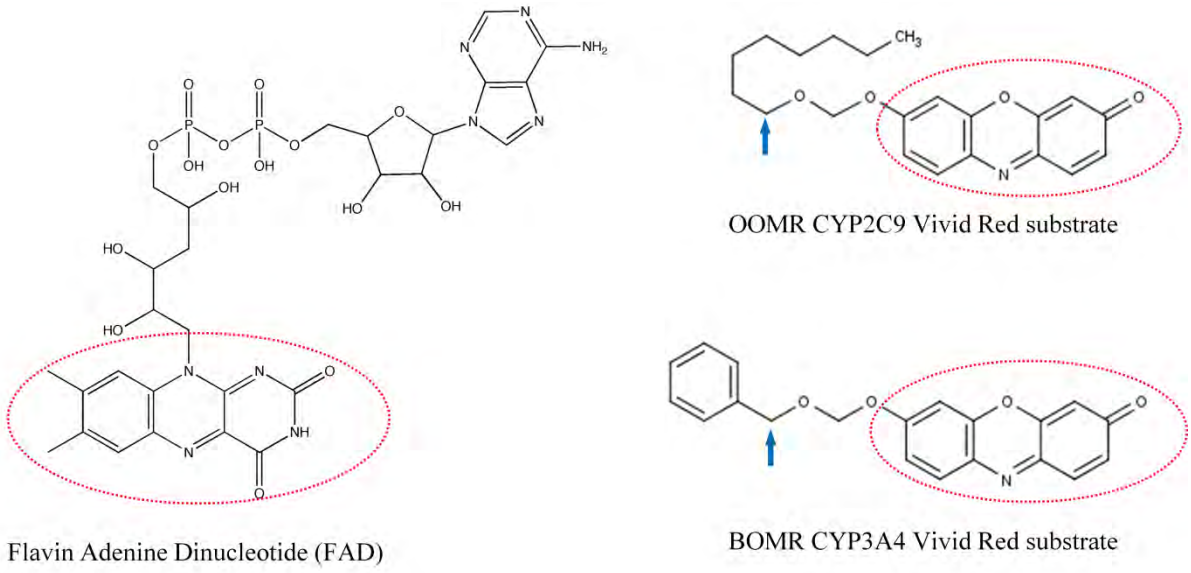


Figure 8.11 Structures of FAD, OOMR and BOMR illustrating the similarity between the Flavin and resorufin rings. The flavin and resorufin rings are circled in red. The benzylic position in BOMR and the corresponding site of oxidation in OOMR are indicated by a blue arrow.

Reproducibility across the microarray surface

To maximise the number of reactions per slide while minimising printing times, replicate reactions are usually printed one after the other in the same area of the slide. The reproducibility between replica reactions printed one after the other, in the same row by the same pin with no pin washing steps in between them, was satisfactory and generally showed median CVs < 15%. These median CVs were similar to CVs observed for replica spots in previous work on the P450 Biochip [473]. Much higher variation occurred between replicate reactions further apart on the slide surface including both high intra- and inter-sub-array variation. This poor reproducibility

would be problematic for testing CYP450 variants as it may lead to apparent differences in activity, which are in fact an artefact of printing.

Previously, spot reproducibility across the entire slide surface was tested using biotinylated-Cy5-BSA: after the fluorescent protein was printed across the entire slide surface, the slide was washed and scanned. CVs across the slide surface were satisfactory; however, importantly if spots are saturated with biotinylated protein during printing, this assay is in fact testing the variation in protein binding capacity across the surface and spot morphology rather than printing reproducibility, which becomes much more critical when substrates are printed onto the slide for kinetic studies. The CVs of a fluorescent standard printed across the whole slide, with the same pin and the same wash steps as the assay, would serve as a better measure of printing reproducibility in the future.

Is the detection of CPR mediated CYP450 activity from co-immobilised proteins on array surface actually plausible?

Surface characterisation of biotinylated CYP3A4 and CPR co-immobilised onto the streptavidin-derivatised PEG-coated slide, carried out by a Master's student in our laboratory, provided valuable information on the density of streptavidin on the surface and hence the protein binding capacity within each microarray spot [594]. Atomic force microscopy (AFM) topographical images revealed white spots 23 to 25 nm in diameter, corresponding to the expected size of several biotinylated CYP450/CPR proteins bound to a streptavidin tetramer. There were an average of 200 streptavidin molecules per μm^2 of slide surface and the distance between molecules was ~50 nm. Each droplet printed onto the slide had a volume of ~ 10 nl and a spot area of ~ 70 000 μm^2. Consequently, there were about 14 million streptavidin molecules per reaction, providing ~ 56 million binding sites for biotinylated proteins to bind to (4 binding sites per streptavidin molecule). In the best case theoretical scenario, where 2 active P450 molecules and 2 active CPR molecules bind to each streptavidin molecule, the effective P450 concentration in each 10 nl spot would therefore be ~ 4 nM; in reality it is likely to be lower than this as catalytically inactive apo-protein and P420 protein present in the CYP3A4 protein samples will also bind to streptavidin.

Based on an estimate of total CYP3A4 protein within the partly purified sample from a protein gel and P420:P450 ratios from CO spectral binding assays, P420 made up ~ 4% and apo protein

~ 25 % of the total CYP3A4 protein in the sample and apo protein made up ~ 50 % of the CYP2C9 protein (Chapter 5). Taking this into account, 2- 3 nM is probably a more accurate estimate of the effective P450 concentration in each 10 nl spot. If CYP3A4 forms oligomers in solution, as discussed in Chapters 6 and 7, it is likely that P450 also binds as oligomers to the streptavidin surface. This would serve to increase the protein concentration within each spot but would only increase substrate turnover if more than one P450 in each oligomer was in an active conformation and able to interact with CPR.

Based on the CuOOH mediated solution-phase CYP3A4 assays for the turnover of DBOMF, a 2 nM protein concentration on the array surface should result in a V_{max} of between 0.5 and 1.2 nM/min. This is well within the range of detection of the microarray scanner and should be above the background signal, provided there is no P450 independent substrate turnover. Indeed, work done in parallel has shown CuOOH mediated CYP3A4 and CYP2C9 activity for robotically printed arrays [594]. Based on k_{cat} values from solution phase assays, the observed activity in these assays would require ~3.5 nM CYP3A4 and ~0.5 nM CYP2C9 protein in each reaction spot on the surface. Keeping in mind that there is no CPR in these assays, this is well within the predicted binding capacity of the surface.

Promising preliminary results have also shown CPR mediated CYP3A4 activity in hand spotted arrays, using the DBOMF substrate and confocal microscopy for detection [594]. The spot area and spot volume for hand spotted arrays is ~ 50 times that of robotically printed arrays, giving an effective working concentration similar to that of a printed array (hand-spotted droplets had a volume of 500 nl and yielded spots ~ 2 mm in diameter). Based on the corresponding solution phase activity assay, the observed activity in the hand-spotted array would require a ~ 10 nM CYP3A4 concentration in each reaction - while this is higher than the predicted 2 to 3 nM binding capacity it is still within an order of magnitude.

Based on these experiments and calculations, it seems reasonable to conclude that the streptavidin-coated microarray surface has sufficient binding capacity to enable the detection of CPR-mediated CYP450 turnover in both hand-spotted and robotically printed arrays. The next question is therefore whether the proximity and relative orientation of CYP450 and CPR on the array surface allows for productive interactions, enabling electron transport between the redox partners.

The cut-off distance generally used to identify interacting residues at the interface of two proteins is 4.5 Å [599-601]. This suggests that the CPR and CYP450 protein need to move to within at least 4.5 Å of each other to allow for the formation of the transient CPR:CYP450 complex. A crude *in silico* model of the CYP3A4 and CPR constructs immobilised to a surface by the C-terminal BCCP-tags gave an optimal distance of 87 Å between biotinylated Lys122 residues on each protein (Figure 8.12) [473]. However, the average distance between the edges of spots consisting of immobilised streptavidin molecules and bound CPR/CYP450 protein was 500 Å according to AFM images, so it seems unlikely that electron transfer between proteins attached to separate streptavidin tetramers could occur without substantial changes in the orientation of the PEG polymers.

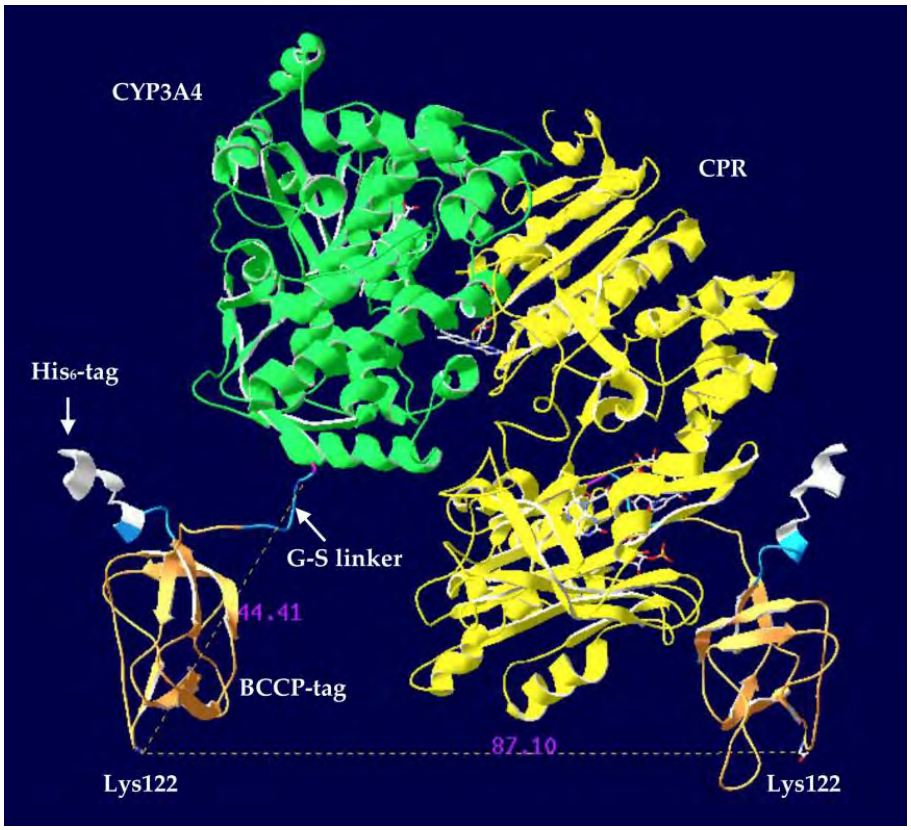


Figure 8.12 *In silico* model of CYP3A4 and CPR immobilised to a surface *via* a C-terminal BCCP tag (Figure from [473])

Following on from this, at the surface densities of streptavidin complexes measured by AFM, the transient CPR:CYP450 complex required for electron transport is only likely to form from proteins bound to the same streptavidin molecule. Figure 8.13 shows the tetrameric structure of streptavidin, made up of two dimers, with a biotin molecule bound to each of the four subunits.

The distance between the carboxyl groups of the biotin molecules bound to adjacent subunits is ~ 20 Å for biotin molecules bound to different dimers (Figure 8.13 B) and ~ 30 Å for biotin molecules bound within the same dimer (Figure 8.13 C). The carboxyl groups of biotin molecules bound to adjacent subunits in different dimers are facing towards each other, facilitating the interaction between proteins attached at these points. In contrast, the carboxyl groups of biotin molecules within the same dimer are facing away from each other and obstruction by the subunits may sterically hinder any interaction between proteins attached at these positions.

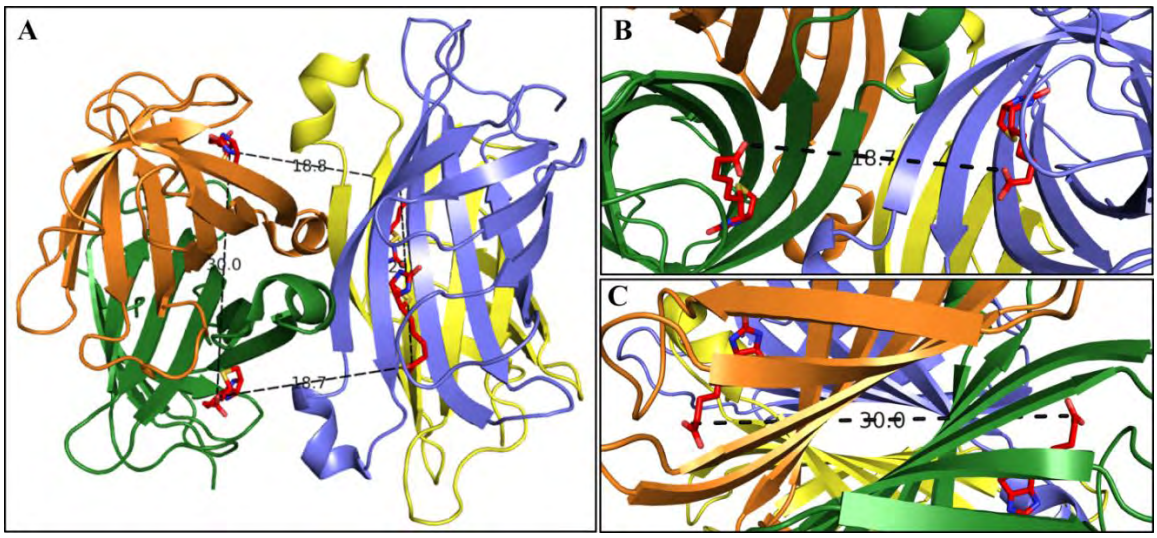


Figure 8.13 **A)** Structure of a streptavidin tetramer with a biotin molecule bound to each of the four subunits, PDB structure 3RY2. The distance between the carboxyl groups of biotin molecules bound to adjacent sub-units is shown by a dashed line. **B)** Close up view of biotin molecules bound to subunits on separate dimers with carboxyl groups facing towards each other separated by a distance of 18.7 Å. **C)** Close up view of the distance between biotin molecules bound to subunits within a dimer with carboxyl groups facing away from each other separated by a distance of 30 Å.

There is still debate around whether biotin binds to streptavidin cooperatively, with some studies suggesting cooperative binding and others suggesting non-cooperative binding [602-604]. A study by Gonzalez *et al* suggests that biotin binding causes structural cooperativity between subunits, without substantial changes to the binding affinity of the remaining binding sites [584]. There is also evidence that steric hindrance between biotinylated molecules can affect binding, leading to a decrease in the occupancy of biotin binding sites [605, 606]; this has led to the design of biotin homologs with longer spacer arms e.g. sulfo-NHS-LC-LC-biotin and NHS-PEG12-biotin. The BCCP-tag has a diameter of 25 - 30 Å, so the binding of BCCP-tags with bulky proteins attached at the two adjacent biotin sites, 20 Å apart, may be unfavourable due to steric hindrance. If this is the case, interaction between a CPR and CYP450 molecule bound to

the same streptavidin molecule may not occur unless oligomer formation allows proteins bound to subunits within the same dimer to interact. It is interesting in this regard that the data of Nair, 2014 suggests that CYP3A4, but not CYP2C9 or CYP2D6, can reconstitute activity with CPR on the array surface, possibly implying different oligomeric states of these CYP450s.

Further experimental work is therefore required to determine the quaternary structure of CPR and CYP3A4 in solution as well as the ratio of bound biotinylated proteins to streptavidin. In addition, hand-spotted microarrays need to be repeated with more data points and further negative controls to confirm that the apparent CPR-dependent CYP3A4 turnover of DBOMF on the array surface is reproducible.

Conclusion

In summary, while the P450 microarray platform in principle offers an attractive alternative to time, reagent and labour intensive solution phase assays, there are in practice a number of issues that need to be addressed before this platform can be used to reliably test the effect of polymorphic variation on drug metabolism. In particular, before the CYP450 microarray platform can be used as a quantitative tool to obtain kinetic data for CPR-mediated CYP450 activity a method is needed to prevent evaporation from the array surface, the relative orientation of CPR and CYP450 on the array surface need to be optimised, printing reproducibility needs to be improved and suitable CYP450 substrates and detection methods compatible with continuous assay measurements need to be established.

8.4 Methods

8.4.1 Testing CPR activity

The activity of the solution-phase N-terminal truncated CPR protein was tested using a continuous spectrophotometric assay based on the reduction of 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium Bromide (MTT) to blue-formazan [591]. 1 ml reaction mixtures consisted of 20 nM CPR (concentration estimated from Bradford assays together with a protein gel of the solution-phase CPR protein preparations) and varying concentrations of MTT (0.1 -500 nM) in 100mM potassium phosphate buffer pH 7.4. The reactions were carried out in disposable plastic cuvettes and initiated by the addition of a NADPH regenerating system (0.5 mM $NADP^+$, 5 mM glucose 6-phosphate and 0.5 U/ml glucose-6-phosphate dehydrogenase final concentrations). The cuvettes were inverted several times, placed in a Varian Cary 50 UV Visible spectrophotometer (Varian, Australia) and the absorbance at 610 nm was monitored for 1 min to determine the rate of MTT reduction.

8.4.2 Testing the effect of glycerol and sucrose concentrations on CYP3A4 DBOMF turnover and ketoconazole inhibition

A 10 mM ketoconazole stock solution was prepared in 100% methanol and a serial dilution in 100 mM potassium phosphate solution, pH 7.4 was performed to obtain a 1µM ketoconazole working stock solution. The ketoconazole solution became cloudy when it was diluted with buffer so the absorbance of the 10 µM stock in buffer was compared to a 10 µM stock in 100% methanol to account for any precipitation that occurred on dilution into an aqueous solution. The precipitation in the aqueous solution only decreased the expected concentration by 20%.

Activity assays were carried out as described in Chapter 7 section 7.4.1 with the following amendments: Each 100 µl reaction contained final concentrations of 8µM DBOMF substrate, 75nM CYP3A4, 0.5 mM CuOOH and 0% to 60% (v/v) glycerol or sucrose in 100 mM potassium phosphate buffer pH 7.4 and was carried out in the presence and absence of 100 nM ketoconazole. The linear reaction was monitored for 5 min.

8.4.3 Preparing streptavidin-coated microarray slides

Preparation of streptavidin solution

A vial of streptavidin lyophilisate (Prospec, Israel) was equilibrated to RT and dissolved in 1 X spotting buffer (100 mM KCl, 0.01% Triton X-100, 25 mM HEPES pH 8.5) to a stock concentration of 10 mg/ml. The solution was mixed by vortexing for 30 min. This stock was diluted further in 1X spotting buffer to a working stock solution of 1 mg/ml. The working stock can be stored at -20°C and reused numerous times until the solution runs out. The working stock solution was centrifuged before use to remove insoluble particles.

Preparation of streptavidin-coated slides

A sealed container of Nexterion H slides (Schott, Germany) was removed from -20°C storage and equilibrated at room temperature for 1 hour to minimise condensation. The following steps were performed with gentle shaking (20-40 rpm) on an orbital shaker to derivatise each slide:

1) 1 hour derivatisation step in 1 mg/ml streptavidin solution
2) 1 hour blocking step in ethanolamine solution (50 mM ethanolamine in 50 mM potassium phosphate buffer, pH 8.0)
3) 3 × 5min wash step in 100 mM potassium phosphate buffer, pH 8.0
4) 5 min wash step in dH_2O

Each slide was placed in a 50ml tube dried by centrifugation at 1600 rpm for 5 min at room temperature. Slides were stored at -20°C in a sealed container.

Preparation of Cy5-biotin-BSA

A 10 mg/ml stock solution of BSA (Roche) in PBS pH 7.4 and a 35 mg/ml stock solution of (+)-Biotin N-hydroxysuccinimide (NHS) ester in DMSO were prepared. Biotin-NHS was added to a tube of BSA stock solution to give a final concentration of 1 mg/ml (*i.e.* 10 mg BSA to 1 mg Biotin) and the tube was wrapped in foil and incubated at RT for 4 h with gentle shaking. ~ 2 mg of biotinylated BSA (200 µl) was added to one vial of Cy5 mono-reactive dye (Amersham). The vial was vortexed thoroughly to dissolve the dye, wrapped in foil at incubated at RT for 1 h with

vortexing every 10 min. The sample was added to SigmaSpin Post-Reaction Purification Columns (Sigma Aldrich) to remove free Cy5 and biotin-NHS esters. The total protein in the sample was determined by a Bradford assay and the concentration of Cy5 in the sample was measured spectrophotometrically (A_{650}, extinction coefficient 0.15 $\mu M^{-1}.cm^{-1}$). The dye to protein (D/P) molar ratio was determined. A D/P ratio ≥ 2 indicated that the Cy5 labelling was sufficient for downstream applications.

Testing the homogeneity of streptavidin coating across the slide surface

Quality control tests were performed on the first and last slide from each batch of streptavidin-coated slides. The slides were equilibrated for 1 h at RT and then immersed in 10 ng/ml Cy5-biotin-BSA in 1 × spotting buffer for 30 min at RT with gentle shaking. The slides were then washed for 3 × 5 min in PBST and 1× 5 min in dH_2O, dried by centrifugation at 1600rpm for 5 min at RT and scanned using the Tecan LS Series Laser Scanner (Tecan Group Ltd., Switzerland), λ_{ex} = 532 nm. This scanner has two lasers; λ_{ex} = 532 nm ('green') and λ_{ex} = 635 nm ('red') optimised for Cy3 and Cy5 dyes respectively. There are a range of emission filters specific to each laser: ~ 560 – 590 nm for the "green" dyes and 655 – 695 nm for the 'red dyes'. Using GenePix Pro measuring tools the average fluorescence intensity and the standard deviation across the printing area of interest were calculated and used to determine the CV for each slide. A CV < 5% within the printing area on the quality control slides was required for downstream assays.

8.4.4 CYP450 microarray assay

Sample preparation for first and second print rounds

All protein and substrate premixes were prepared on the day of printing; protein samples for the first round of printing were prepare in P450 storage buffer and BOMR substrate + NADPH premixes and resorufin standards for the second round of printing were prepared in 100mM potassium phosphate buffer pH 7.4, 15% (v/v) glycerol. Partially purified CPR and CYP3A4 protein samples, previously expressed in *E.coli* (Chapter 5) were used for the microarray assays and protein concentrations within the samples were based on the total protein determined by Bradford assays rather than P450 content, following the protocol previously established by Dr

Beeton-Kempen. 20 µl aliquots of each sample for the first and second print rounds were pipetted into the relevant wells of a 384-well Genetix source plate (Genetix Ltd., UK).

Microarray assay testing CYP3A4 variant activity

The following samples were prepared for the first print round: 10 ng/µl Cy5-biotin BSA; 250 nM CPR and a 250 nM 3A4 + 250 nM CPR protein mix for each CYP3A4 protein (WT, K96E, P218R, S222P, I223R, L293P and M445T). The following samples were prepared for the second print round: a resorufin dilution (1-500 nM) series consisting of 10 standards; 1mM NADPH only and 5.7 µM BOMR only for the negative controls; and 12 NADPH-substrate premixes consisting of 0.50 – 43.2 µM BOMR and 1mM NADPH.

Microarray assay testing different combination of reaction components to determine the components essential for BOMR turnover on the array surface

The following protein samples were prepared for the first print round: 500 nM CPR; 500nM CYP3A4 wt; 250 nM 3A4 wt + 250 nM CPR. The following samples were prepared for the second print round: a resorufin dilution (1-500 nM) series consisting of 10 standards, 4 substrate premixes with and without NADPH (1.1 µM; 2.5 µM; 5.7 µM and 12.8 µM BOMR with and without 1mM NADPH)

Print layouts

16 sub-arrays (4 x 4) were printed onto each slide using 16, solid, flat-tipped, 300 µm stainless steel Telechem pins. The position of the pins in the print head is shown in Figure 8.14 A. Each sub-array consisted of 6 x 12 spots and all spots in each sub-array were printed by the same pin, one after the other. This means that the first spot in each sub-array are printed simultaneously. The 384-well plate must be set up accordingly to the position of the pins in the pin head. The visit numbers for the plate are shown in red and the visit number that is input into the printing pattern is based on the wells visited by the first row of pins (pins 1 to 4) e.g. the wells visited by pins for visit number 25 are shown in green in Figure 8.14 B. For identical sub-arrays across the slide, the same sample should be placed in all 16 wells visited simultaneously by the pin head. The slide layout for each assay is shown in Figure 8.15 and Figure 8.16.

Printing parameters

The slides were printed with a QArray2 robotic microarrayer (Genetix Ltd.). Print runs were carried out at 25°C and the humidity in the printing chamber was set to 50%. The inking time and stamp time were set to 500 ms and 0 ms respectively and the maximum stamps per ink was 1. Three stamps per spot were used in the first print round and one stamps per spot for the second print round. The pins were washed for 60 s in dH_2O, 10 s in ethanol and dried for 11 s between unique samples.

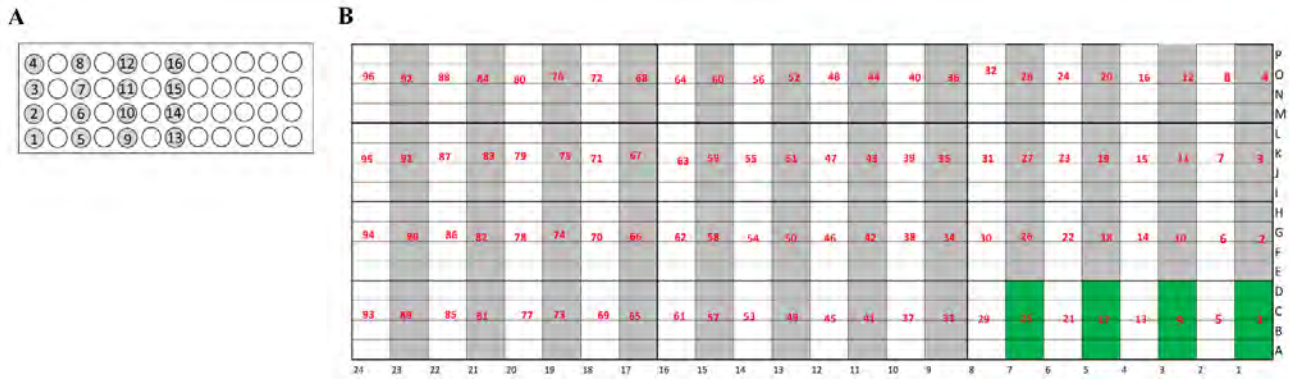


Figure 8.14 **A)** Position of pin in the pin head **B)** 384 well plate template where each block represents a well. Visit numbers are shown in red. The well visited by the 16 pins for visit number 25 are shown in green.

Assay procedure

A pack of slides were removed from -20°C storage and equilibrated at RT for 1 h before printing. Slides were then removed from the sealed package and dried by centrifugation at 1600 rpm for 5 min at RT. The 384-well sample plate was loaded and the slide lined up on the print bed for the first print run. After the proteins were printed onto the slide, the slide was left on the print bed in the printing chamber for 30min to allow the proteins to bind to the surface. The slide was then placed in P450 blocking buffer (20 mM potassium phosphate pH 7.4, 0.2 mM EDTA, 5% (v/v) glycerol, 100 µM biotin, 0.1 mg/ml BSA) on ice with gentle shaking for 15 min. The slide was washed for 3 x 5 min in P450 slide wash buffer (20 mM potassium phosphate pH 7.4, 0.2mM EDTA, 5% (v/v) glycerol,), 5 min in dH_2O and dried by centrifugation. The slide was then placed back in the same position on the print bed for the second print run where substrate premixes were overlaid over the protein spots. Immediately after the print run ended, the slides were removed from the printing chamber placed in the scanner and scanned every 5 to 10 min for 45min at 175 PMT. Slides were scanned using the 532 nm excitation laser. Since resorufin has an emission wavelength of 585 nm, falling within the range of the emission filters supplied with the laser (560 -590 nm).

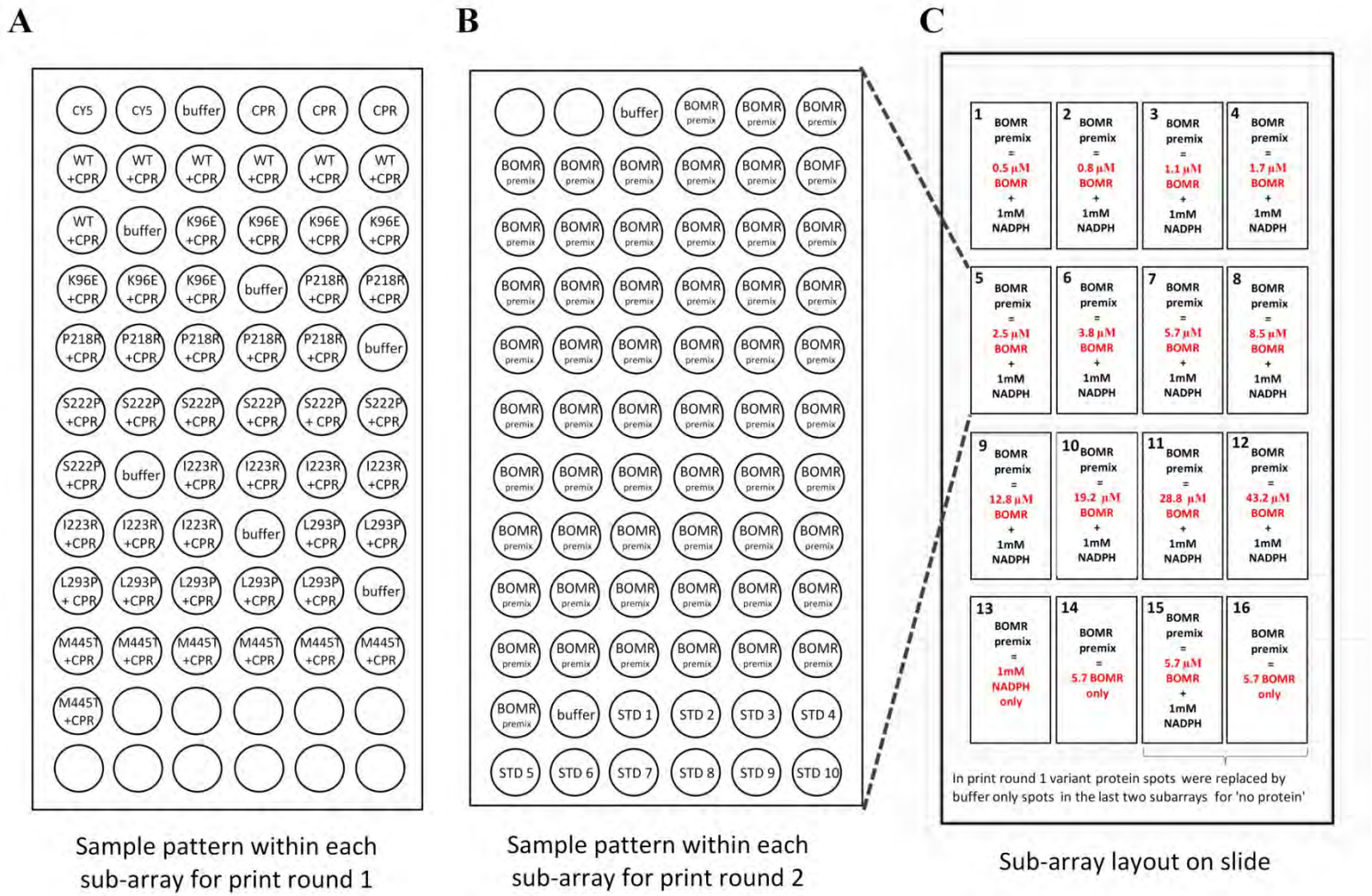


Figure 8.15 Print layout for microarray assay testing CYP3A4 variant activity A) The 6x12 layout of protein samples within each sub-array for print round 1 (except sub-arrays 15 and 16 where variant CYP3A4 spots are replace by P450 storage buffer only spots for 'no protein' negative controls). P450 storage buffer spots between each protein sample were included as no "protein controls" at each substrate concentration and as controls for protein carry over. B) The layout of substrate premixes and resorufin standards for print round 2. A different substrate concentration or control (BOMR premix) was used in each sub-array. C) The 4x4 sub-array layout on the slide; BOMR concentrations 0.5 – 43.2 µM in sub-arrays 1 to 12 with additional controls in sub-arrays 13 to 16.

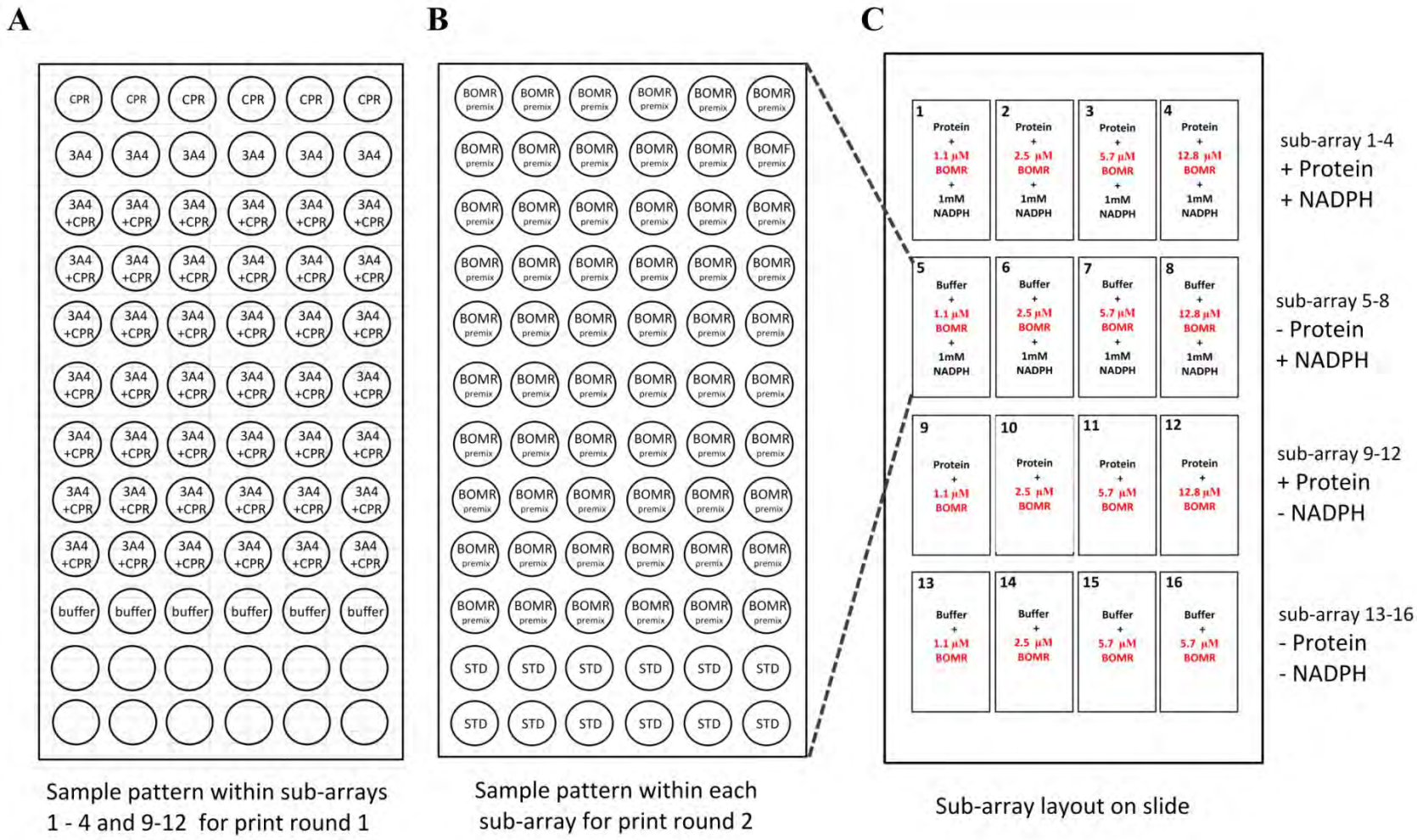


Figure 8.16 Print layout for microarray assay testing CYP3A4 variant activity A) The 6x12 layout of protein samples within each sub-arrays 1 to 4 and 9 to12 for print round 1. All spots in sub-arrays 5 to 8 and 13 to16 were P450 storage buffer only. B) The layout of substrate premixes and resorufin standards for print round 2. Resorufin standards 1 to 10 were printed in the last two rows of sub-arrays 1 to 10 with 1nM resorufin in sub-array 1 and 500 nM resorufin in sub-array 10. C) The 4x4 sub-array layout on the slide; BOMR + NADPH were printed sub-arrays 1 to 8 and BOMR only on sub-arrays 9 to 16.

Data analysis

The fluorescent signal from the slides was captured using GenePix Pro 6.1.0.4 microarray software (Axon Instruments Inc., US). A circular feature finder with a diameter of 150 μm (for the first microarray assay) or 170 μm (for the second microarray assay) was used to define the spot sizes. The median background intensity for each spot was measured in the local area around the spot within 3 feature diameters from the centre of the spot excluding the first 6 pixels surrounding the spot. The median background intensity was subtracted from the median fluorescent signal for each spot. The corrected median fluorescent signal was averaged across replicate spots at each time point and the standard deviations were calculated. Velocity graphs were plotted using the averaged values and RFU/min was converted to nM/min using the resorufin standard curve (the standard curve was stable over the 45 min period of the assay). The median % CV across all time points was calculated for each reaction (median average standard deviation across all time points divided by the median average fluorescent signal across all time points).

8.4.5 Testing fluorescent substrates in solution for NADPH-CPR mediated CYP450 turnover

The turnover of CYP3A4 Vivid substrates BOMR and DBOMF and CYP2C9 Vivid substrates BOMF and OOMR by different combination of reaction components was tested in 96 well plates Activity assays were carried out as described in Chapter 7 section 7.4.1 using partially purified protein samples. Protein concentrations were based on total protein as measured by Bradford assay. The following reactions were carried out for each substrate: 1) 5.7 μM Vivid substrate, 1mM NADPH; 2) 5.7 μM Vivid substrate; 3) 250 nM CYP3A4 or CYP2C9, 250 nM CPR, 5.7 μM Vivid substrate, 1mM NADPH; 4) 250 nM CYP3A4 or CYP2C9, 250 nM CPR, 5.7 μM Vivid substrate; 5) 250 nM CYP3A4 or CYP2C9, 5.7 μM Vivid substrate, 0.5 mM CuOOH; 6) 250 nM CYP3A4 or CYP2C9, 5.7 μM Vivid substrate, 1mM NADPH; 7) 250 nM CYP3A4 or CYP2C9, 5.7 μM Vivid substrate; 8) 250nM CPR, 5.7 μM Vivid substrate, 1mM NADPH; 9) 250nM CPR, 5.7 μM Vivid substrate. All reactions were carried out in 100 mM potassium phosphate buffer pH 7.4, 15% (v/v) glycerol.

Chapter 9 Conclusions

The aims of this Thesis were to develop *in silico* and *in vitro* approaches for determining the effect of non-synonymous mutations on CYP450 stability and function, with the intention of providing new cost effective ways to studying the effect of polymorphic variation on CYP450-mediated drug metabolism in the pre-clinical stages of drug testing.

An initial goal was to study CYP450 structure-functions relationships *in silico,* defining regions of the protein structure important for different aspects of CYP450 function. In collaboration with the group of Prof Chun Wei (University of Singapore), a new expanded mammalian CYP450 substrate recognition site map was generated based on molecular interactions identified in ligand bound crystal structures and docking complexes. In addition, using static crystal structures and the computational tools CAVER and Protein Interaction Calculator, this Thesis describes the identification of the largest set of tunnel-gating residues reported to date.

Following this, a structural alignment of the 9 major human drug metabolising CYP450s was annotated with substrate recognition sites, gating residues, CPR contact residues (based on previous study from the Blackburn group), haem contact residues and conserved motifs, as well as the positions of 348 SNPs found in these enzymes. While this alignment gives a good overview of the structure-function relationship in CYP450s, CYP450s are complex enzymes and the molecular mechanisms underlying oligomer formation and protein flexibility are still poorly understood and thus it is difficult to predict the effects of mutations on these latter aspects of enzyme function. In addition, substrate recognition site and gating residues are not necessarily important for the access and binding of all potential substrates and consequently mutations at these positions are likely to display substrate specific effects.

The representative computational tool SDM was used to calculate the effect of each SNP on protein stability. While there is a selection of tools available for predicting deleterious mutations, they are not able to quantitatively predict the effect of mutations on ligand binding; as such, SDM was used to determine the effect of mutations on the stability of the apo protein structure only. These stability calculations were then combined with the functional analyses to take the effects of mutations on haem incorporation, ligand binding and CYP450-CPR complex formation into account. The resultant CYP450 SNP Map was then compared with experimental data to establish the main causes of altered drug metabolism observed for CYP450 variants in functional assays to date. Despite its apparent simplicity, this approach was shown to have

surprisingly good predictive capability, highlighting the value of combining detailed structure-function analysis with protein stability data to predict the effect of mutations on enzyme activity. The data obtained should be useful to researchers and clinicians a like since it serves as a useful guide for selecting CYP450 variants for *in vitro* testing, for rationalising new *in vitro* and clinical data, and for predicting the effect of uncharacterised variants on drug metabolism.

Following the *in silico* analysis, nine CYP3A4 and eight CYP2C9 variants including the wild-types, were expressed in *E.coli* for *in vitro* testing. The N-terminal membrane binding region was removed allowing for soluble expression in the cytosolic fraction and a His-tag and BCCP-tag were included on the C-terminus of each construct for downstream purification and surface immobilisation. While soluble expression levels and P450:P420 ratios varied substantially between variant proteins, sufficient holoprotein P450 protein was obtained for downstream application for all 17 constructs.

The data from thermostability assays suggests that CYP3A4 and CYP2C9 proteins exist as heterogeneous populations of distinct conformers in solutions and that single point mutations can shift the equilibrium between these conformers. In addition, these assays suggest that mutations may affect protein stability by disrupting haem binding or altering oligomer formation. Further studies are, however, required to confirm that the observed conformation heterogeneity is a result of oligomer formation and to determine the quaternary structure of the CYP3A4 and CYP2C9 constructs. Two thirds of the variants showed significantly altered thermostability compared to the corresponding wild-type protein and all variants predicted to alter protein function by stabilising or destabilising the apo-protein structure *in silico* were found to significantly alter the thermostability of the holoprotein in solution.

CuOOH driven fluorescent-based kinetic assays revealed atypical kinetic profiles for both of the CYP3A4 substrates and one of the two CYP2C9 substrates tested. Various kinetic models involving a combination of kinetically distinct conformers and the binding of more than one substrate in the active site leading to substrate inhibition were proposed, but despite this, fitting curves to some of the data was not possible. Inclusion of 15% glycerol in the reaction mixtures generally increased the rate of substrate turnover and gave kinetic profiles that more closely resembled those observed for membrane bound proteins. The kinetic data obtained for CYP3A4 and CYP2C9 variants showed that several variants have significantly altered catalytic turnover

and/or binding affinity for the substrates tested compared to the wild-type proteins and revealed two previously uncharacterised variants with altered function.

One of the original aims of this work was to explore the use of a CYP450 protein microarray platform, previously developed in the Blackburn laboratory, to test the catalytic activity of the panel of CYP3A4 and CYP2C9 variants in parallel. Unfortunately, the results described in this Thesis epitomise the numerous challenges surrounding miniaturised high-throughput CYP450 platforms and suggest that the platform, in its current form, cannot provide accurate data on CYP450 turnover. While this Thesis appears to shed a more pessimistic light on the utility of the microarray platform than previously reported [473, 594], it raises important concerns about the reliability of the data generated that should be taken into account going forward. The results reported here raise questions as to the theoretical feasibility of CPR-mediated CYP450 activity on the array surface and highlight the need to explore enzyme encapsulation techniques to prevent evaporation, as well as alternative substrates and/or metabolite detections techniques.

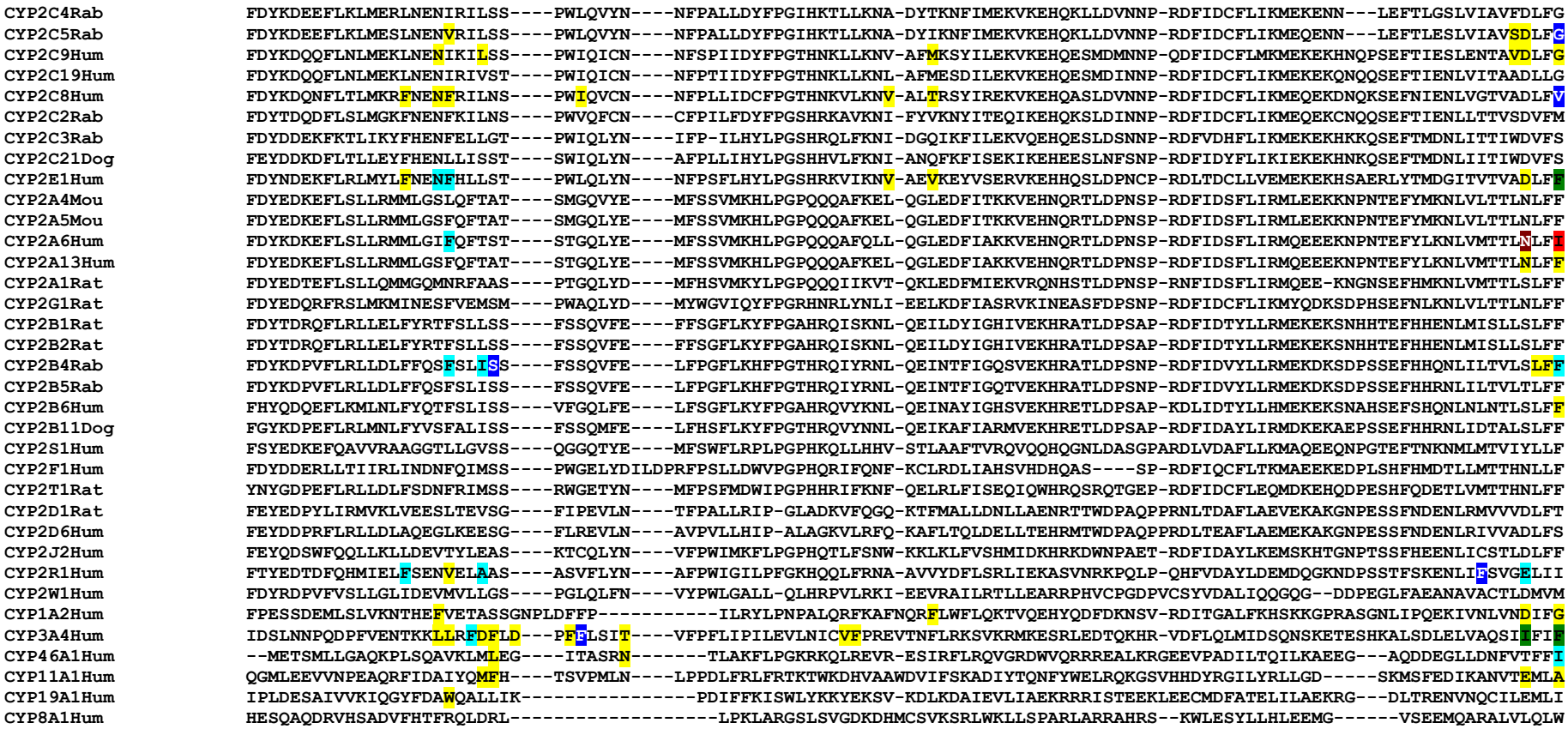
Finally, this Thesis uncovers a previously unrecognised catalytic activity of CPR in efficiently turning over the commonly used BOMR Vivid Red CYP3A4 substrate and testable mechanisms to account for this apparently paradoxical redox activity (*i.e.* CPR is a reductase, yet product formation from BOMR is designed to be an oxidative reaction) have been proposed.

Appendices

Appendix A. **Sequence alignments annotated with CYP450 ligand contacts**

```
CYP2C4Rab          ---------------MDPVAGLVLGLCCLLLLS---------LWKQNSGRGKLPPGPTPFPIIG--------NILQ------IDVKDISKSLTKFSERYGPVFTVYLGM-KPTVVLHGYK
CYP2C5Rab          ---------------MDPVVVLVLGLCCLLLLS---------IWKQNSGRGKLPPGPTPFPIIG--------NILQ------IDAKDISKSLTKFSECYGPVFTVYLGM-KPTVVLHGYE
CYP2C9Hum          ---------------MDSLVVLVLCLSCLLLLS---------LWRQSSGRGKLPPGPTPLPVIG--------NILQ------IGIKDISKSLTNLSKVYGPVFTLYFGL-KPIVVLHGYE
CYP2C19Hum         ---------------MDPFVVLVLCLSCLLLLS---------IWRQSSGRGKLPPGPTPLPVIG--------NILQ------IDIKDVSKSLTNLSKIYGPVFTLYFGL-ERMVVLHGYE
CYP2C8Hum          ---------------MEPFVVLVLCLSFMLLFS---------LWRQSCRRRKLPPGPTPLPIIG--------NMLQ------IDVKDICKSFTNFSKVYGPVFTVYFGM-NPIVVFHGYE
CYP2C2Rab          ---------------MDLVVVLGLCLSCLLLPS---------LWKQSHGGGKLPPGPTPFPILG--------NVLQ------LDFKDLSKSLTNLSKVYGPVFTVYLGM-KPTVVVHGYE
CYP2C3Rab          ---------------MDLLIILGICLSCVVLLS---------LWKKTHGKGKLPPGPTPLPVVG--------NLLQ------LETKNINKSLSMLAKEYGSIFTLYFGM-KPAVVLYGYE
CYP2C21Dog         -----------------FIVLVICLSCLISFF---------LWNQNRAKGKLPPGPTPLPIIG--------NILQ------INTKNVSKSLSKLAENYGPVFTVYFGM-KPTVVLYGYE
CYP2E1Hum          ------------MSALGVTVALLVWAAFLLLVS---------MWRQVHSSWNLPPGPFPLPIIG--------NLFQ------LELKNIPKSFTRLAQRFGPVFTLYVGS-QRMVVMHGYK
CYP2A4Mou          -----------MLTSGLLLVAAVAFLSVLVLMS---------VWKQRKLSGKLPPGPTPLPFVG--------NFLQ------LNTEQMYNSLMKISQRYGPVFTIYLGS-RRIVVLCGQE
CYP2A5Mou          -----------MLTSGLLLVAAVAFLSVLVLMS---------VWKQRKLSGKLPPGPTPLPFIG--------NFLQ------LNTEQMYNSLMKISQRYGPVFTIYLGP-RRIVVLCGQE
CYP2A6Hum          -----------MLASGMLLVALLVCLTVMVLMS---------VWQQRKSKGKLPPGPTPLPFIG--------NYLQ------LNTEQMYNSLMKISERYGPVFTIHLGP-RRVVVLCGHD
CYP2A13Hum         -----------MLASGLLLVTLLACLTVMVLMS---------VWRQRKSRGKLPPGPTPLPFIG--------NYLQ------LNTEQMYNSLMKISERYGPVFTIHLGP-RRVVVLCGHD
CYP2A1Rat          -----------MLDTGLLLVVILASLSVMLLVS---------LWQQ-KIRGRLPPGPTPLPFIG--------NYLQ------LNTKDVYSSITQLSERYGPVFTIHLGP-RRVVVLYGYD
CYP2G1Rat          -----------MALGGAFSIFMTLCLSCLLILI---------AWKRTSRGGKLPPGPTPIPFLG--------NLLQ------VRIDATFQSFLKLQKKYGSVFTVYFGP-RPVVILCGHE
CYP2B1Rat          -----------MEPT---ILLLLALLVGFLLLL---------VRGHPKSRGNFPPGPRPLPLLG--------NLLQ------LDRGGLLNSFMQLREKYGDVFTVHLGP-RPVVMLCGTD
CYP2B2Rat          -----------MEPS---ILLLLALLVGFLLLL---------VRGHPKSRGNFPPGPRPLPLLG--------NLLQ------LDRGGLLNSFMQLREKYGDVFTVHLGP-RPVVMLCGTD
CYP2B4Rab          -----------MEFS---LLLLLAFLAGLLLLL---------FRGHPKAHGRLPPGPSPLPVLG--------NLLQ------MDRKGLLRSFLRLREKYGDVFTVYLGS-RPVVVLCGTD
CYP2B5Rab          -----------MEFS---LLLLLAFLAGLLLLL---------FRGHPKAHGRLPPGPPPLPVLG--------NLLQ------MDRKGLLRSFLRLREKYGDVFTVYLGS-RPVVVLCGTD
CYP2B6Hum          -----------MELS---VLLFLALLTGLLLLL---------VQRHPNTHDRLPPGPRPLPLLG--------NLLQ------MDRRGLLKSFLRFREKYGDVFTVHLGP-RPVVMLCGVE
CYP2B11Dog         -----------MELS---VLLLLALLTGLLLLM---------ARGHPKAYGHLPPGPRPLPILG--------NFLQ------MDRKGLLKSFLRLQEKYGDVFTVYLGP-RRTVMLCGID
CYP2S1Hum          -----------MEATG-TWALLLALALLLLLTL---------ALSGTRARGHLPPGPTPLPLLG--------NLLQ------LRPGALYSGLMRLSKKYGPVFTIYLGPWRPVVVLVGQE
CYP2F1Hum          -----------MDSISTAILLLLLALVCLLLTL------------SSRDKGKLPPGPRPLSILG--------NLLL------LCSQDMLTSLTKLSKEYGSMYTVHLGP-RRVVVLSGYQ
CYP2T1Rat          ------------MVTCEATLLLLLILTLMLMSWGWL-----AHQARARMQKDLPPGPAPLPLLG--------NLLQ------LQSGHLDRVLMELSSRWGPVFTVWLGP-RPAVVLSGYA
CYP2D1Rat          --------MELLNGTGLWSMAIFTVIFILLVDLMHR---------RHRWTSRYPPGPVPWPVLG--------NLLQ------VDLSNMPYSLYKLQHRYGDVFSLQKGW-KPMVIVNRLK
CYP2D6Hum          -----------MGLEALVPLAVIVAIFLLLVDLMHR---------RQRWAARYPPGPLPLPGLG--------NLLH------VDFQNTPYCFDQLRRRFGDVFSLQLAW-TPVVVLNGLA
CYP2J2Hum          ----------MLAAMGSLAAALWAVVHPRTLLLGTVAFLLAADFLKRRRPKNYPPGPWRLPFLG--------NFF--L----VDFEQSHLEVQLFVKKYGNLFSLELGD-ISAVLITGLP
CYP2R1Hum          --------MWKLWRAEEGAAALGGALFLLLFALG------VRQLLKQRRPMGFPPGPPGLPFIG--------NIYSLA----ASSELPHVYMRKQSQVYGEIFSLDLGG-ISTVVLNGYD
CYP2W1Hum          --------MALLLLLFLGLLGLWGLLCACAQDPS--------------PAARWPPGPRPLPLVG--------NLHL------LRLSQQDRSLMELSERYGPVFTVHLGR-QKTVVLTGFE
CYP1A2Hum          ------MALSQSVPFSATELLLASAIFCLVFWVLKG------LRPRVPKGLKSPPEPWGWPLLG--------HVLT-------LGKNPHLALSRMSQRYGDVLQIRIGS-TPVLVLSRLD
CYP3A4Hum          ---------MALIPDLAMETWLLLAVSLVLLYLYG-------THSHGLFKKLGIPGPTPLPFLG--------NILS-------YHKGFCMFDMECHKKYGKVWGFYDGQ-QPVLAITDPD
CYP46A1Hum         --------MSPGLLLLGSAVLLAFGLCCTFVHR------------ARSRYEHIPGPPRPSFLLG--------HLPCFWKKDEVGGRVLQDVFLDWAKKYGPVVRVNVFH-KTSVIVTSPE
CYP11A1Hum         ---------MLAKGLPPRSVLVKGCQTFLSAPREGLGRLRVPTGEGAGISTRSPRPFNEIPSPG---DNGWLNLYHFWR--ETGTHKVHLHHVQNFQKYGPIYREKLGN-VESVYVIDPE
CYP19A1Hum         MVLEMLNPIHYNITSIVPEAMPAATMPVLLLTG-------LFLLVWNYEGTSSIPGPGYCMGIGPLI---SHG--------RFLWMGIGSACNYYNRVYGEFMRVWISG-EETLIISKSS
CYP8A1Hum          -------------MAWAALLGLLAALLLLLLS----------RRRTRRPGEPPLDLGSIPWLG--------YALD-------FGKDAASFLTRMKEKHGDIFTILVGGRYVTVLLDPHS
```

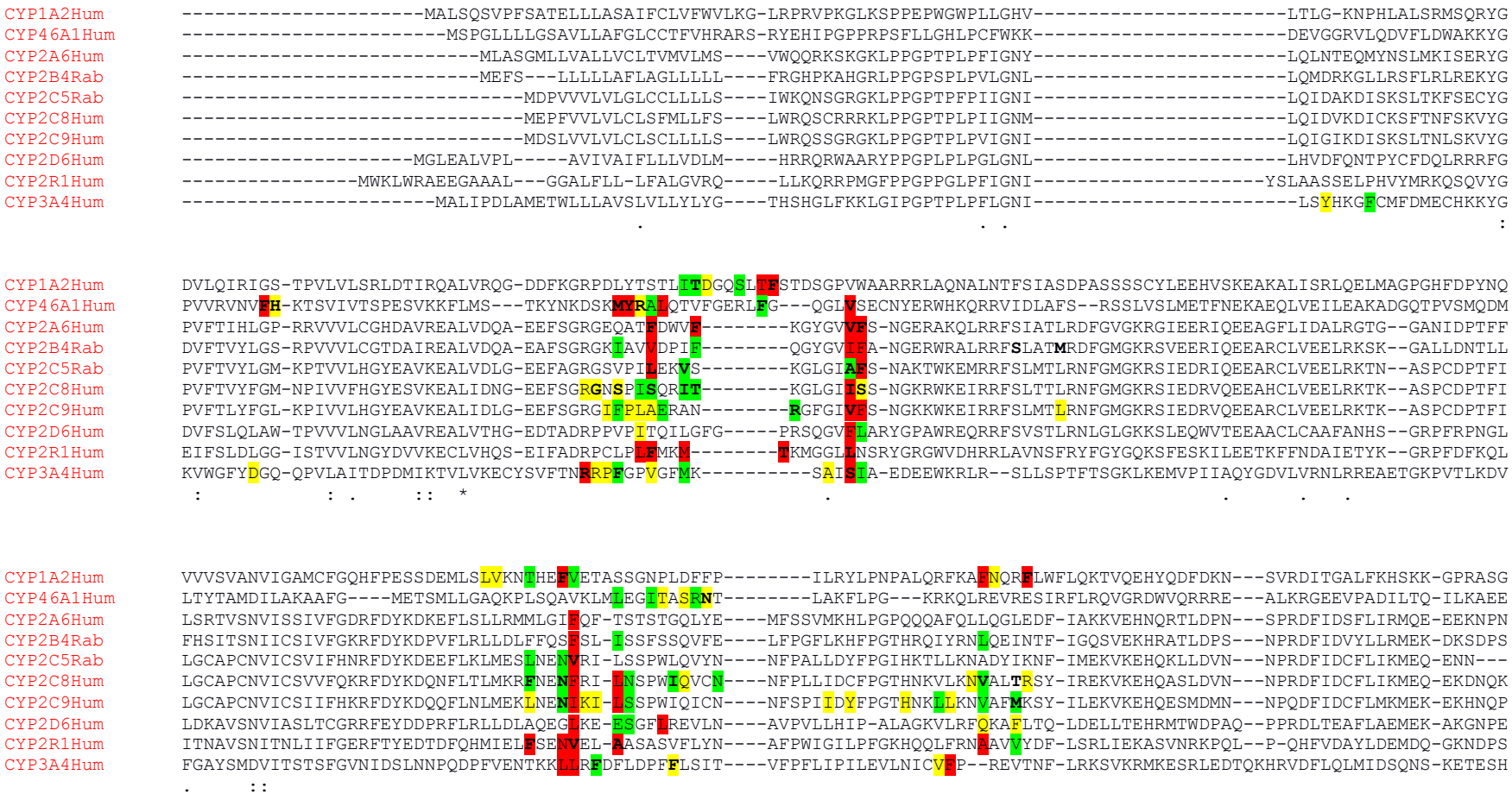
```
CYP2C4Rab          AVKEALVDLGE-EFAGR---GHFPIAEKVNKGLGIVFT--NANTWKEMRRFSLMTLRNFGMGKRSI-------EDRVQEEARCLVEELRKTNA--LPCDPTFILGCAPCNVICSVILHNR
CYP2C5Rab          AVKEALVDLGE-EFAGR---GSVPIDEKVSKGLGIAFS--NAKTWKEMRRFSLMTLRNFGMGKRSI-------EDRIQEEARCLVEELRKTNA--SPCDPTFILGCAPCNVICSVIFHNR
CYP2C9Hum          AVKEALIDLGE-EFSGR---GIFPLAERANRGFGIVFS--NGKKWKEIRRFSLMTLRNFGMGKRSI-------EDRVQEEARCLVEELRKTKA--SPCDPTFILGCAPCNVICSIIFHKR
CYP2C19Hum         VVKEALIDLGE-EFSGR---GHFPLAERANRGFGIVFS--NGKRWKEIRRFSLMTLRNFGMGKRSI-------EDRVQEEARCLVEELRKTKA--SPCDPTFILGCAPCNVICSIIFQKR
CYP2C8Hum          SVKEALIDNGE-EFSGR---GNSPISQRITKGLGIISS--NGKRWKEIRRFSLTTLRNFGMGKRSI-------EDRVQEEAHCLVEELRKTKA--SPCDPTFILGCAPCNVICSVVFQKR
CYP2C2Rab          AVKEALVDLGH-ELSGR---SRFLVTAKLNKGFGVIFS--NGKRWTETRRFSLMTLRNFGMGKRSI-------EERVQEEAHCLVEELRKTNA--SPCDPTFILGAAPCNVICSVIFQNR
CYP2C3Rab          TVKEALIDRGE-EFSGR---GIFPVFDRVTKGLGIVFS--SGEKWKETRRFSLTVLRNLGMGKKTI-------EERIQEEALCLIQALRKTNA--SPCDPTFLLFCVPCNVMCSVIFQNR
CYP2C21Dog         AVKEALIDRSE-EFSGR---GHFPLLDWTIQGLGIVFS--NGEKWKQTRRFSLTVLRNMGMGKKTV-------EDRIQEEALYLVEALKKTNA--SPCDPTFLLGCAPCNVICSIIFQNR
CYP2E1Hum          AVKEALLDYKD-EFSGR---GDLPAFH-AHRDRGIIFN--NGPTWKDIRRFSLTTLRNYGMGKQGN-------ESRIQREAHFLLEALRKTQG--QPFDPTFLIGCAPCNVIADILFRKH
CYP2A4Mou          AVKEALVDQAE-EFSGR---GEQATFDWLFKGYGIAFS--SGERAKQLRSFSIATLRDFGVGKRGI-------EERIQEEAGFLIDSFRKTNG--AFIDPTFYLSRTVSNVISSIVFGDR
CYP2A5Mou          AVKEALVDQAE-EFSGR---GEQATFDWLFKGYGVVFS--SGERAKQLRRFSIATLRDFGVGKRGI-------EERIQEEAGFLIDSFRKTNG--AFIDPTFYLSRTVSNVISSIVFGDR
CYP2A6Hum          AVREALVDQAE-EFSGR---GEQATFDWVFKGYGVVFS--NGERAKQLRRFSIATLRDFGVGKRGI-------EERIQEEAGFLIDALRGTGG--ANIDPTFFLSRTVSNVISSIVFGDR
CYP2A13Hum         AVKEALVDQAE-EFSGR---GEQATFDWLFKGYGVAFS--NGERAKQLRRFSIATLRGFGVGKRGI-------EERIQEEAGFLIDALRGTHG--ANIDPTFFLSRTVSNVISSIVFGDR
CYP2A1Rat          AVKEALVDQAE-EFSGR---GEQATYNTLFKGYGVAFS--SGERAKQLRRLSIATLRDFGVGKRGV-------EERILEEAGYLIKMLQGTCG--APIDPTIYLSKTVSNVISSIVFGER
CYP2G1Rat          AVKEALVDQAD-DFSGR---GEMPTLEKNFQGYGLALS--NGERWKILRRFSLTVLRNFGMGKRSI-------EERIQEEAGYLLEELHKVKG--APIDPTFYLSRTVSNVICSVVFGKR
CYP2B1Rat          TIKEALVGQAE-DFSGR---GTIAVIEPIFKEYGVIFA--NGERWKALRRFSLATMRDFGMGKRSV-------EERIQEEAQCLVEELRKSQG--APLDPTFLFQCITANIICSIVFGER
CYP2B2Rat          TIKEALVGQAE-DFSGR---GTIAVIEPIFKEYGVIFA--NGERWKALRRFSLATMRDFGMGKRSV-------EERIQEEAQCLVEELRKSQG--APLDPTFLFQCITANIICSIVFGER
CYP2B4Rab          AIREALVDQAE-AFSGR---GKIAVVDPIFQGYGVIFA--NGERWRALRRFSLATMRDFGMGKRSV-------EERIQEEARCLVEELRKSKG--ALLDNTLLFHSITSNIICSIVFGKR
CYP2B5Rab          AIREALVDQAE-AFSGR---GKIAVVDPIFQGYGVFFA--NGEHWRALRRFSLATMRDFGMGKRSV-------EERIQEEARCLVEELRKSKG--ALLDNTLLFHSVTSNIICSIVFGKR
CYP2B6Hum          AIREALVDKAE-AFSGR---GKIAMVDPFFRGYGVIFA--NGNRWKVLRRFSVTTMRDFGMGKRSV-------EERIQEEAQCLIEELRKSKG--ALMDPTFLFQSITANIICSIVFGKR
CYP2B11Dog         AIREALVDNAE-AFSGR---GKIAVVEPVFQGYGVVFA--NGERWKTLRRFSLATMRDFGMGKRSV-------EERIQEEAQCLVEELRKTEG--VLQDPTFFFHSMTANIICSIVFGKR
CYP2S1Hum          AVREALGGQAE-EFSGR---GTVAMLEGTFDGHGVFFS--NGERWRQLRKFTMLALRDLGMGKREG-------EELIQAEARCLVETFQGTEG--RPFDPSLLLAQATSNVVCSLLFGLR
CYP2F1Hum          AVKEALVDQGE-EFSGR---GDYPAFFNFTKGNGIAFS--SGDRWKVLRQFSIQILRNFGMGKRSI-------EERILEEGSFLLADVRKTEG--EPFDPTFVLSRSVSNIICSVLFGSR
CYP2T1Rat          ALRDALVLQAD-AFSGP---RVHGCLRTLTHGNGIVFS--NGPALATLRTLHLNTEELR--GSSTI-------EERILEETACVLDEFQATMG--APFDPRRLLDNAVSNVICTVVFGKR
CYP2D1Rat          AVQEVLVTHGE-DTADRPPVPIFKCLGVKPRSQGVILAS-YGPEWREQRRFSVSTLRTFGMGKKSL-------EEWVTKEAGHLCDAFTAQAG--QSINPKAMLNKALCNVIASLIFARR
CYP2D6Hum          AVREALVTHGE-DTADRPPVPITQILGFGPRSQGVFLAR-YGPAWREQRRFSVSTLRNLGLGKKSL-------EQWVTEEAACLCAAFANHSG--RPFRPNGLLDKAVSNVIASLTCGRR
CYP2J2Hum          LIKEALIHMDQ-NFGNRPVTPMREHI---FKKNGLIMS--SGQAWKEQRRFTLTALRNFGLGKKSL-------EERIQEEAQHLTEAIKEENG--QPFDPHFKINNAVSNIICSITFGER
CYP2R1Hum          VVKECLVHQSE-IFADRPCLPLFMKM---TKMGGLLNSR-YGRGWVDHRRLAVNSFRYFGYGQKSF-------ESKILEETKFFNDAIETYKG--RPFDFKQLITNAVSNITNLIIFGER
CYP2W1Hum          AVKEALAGPGQ-ELADRPPIAIFQLIQ---RGGGIFFS--SGARWRAARQFTVRALHSLGVGREPV-------ADKILQELKCLSGQLDGYRG--RPF-PLALLGWAPSNITFALLFGRR
CYP1A2Hum          TIRQALVRQGD-DFKGR---PDLYTSTLITDGQSLTFSTDSGPVWAARRRLAQNALNTFSIASDPASSSSCYLEEHVSKEAKALISRLQELMAGPGHFDPYNQVVVSVANVIGAMCFGQH
CYP3A4Hum          MIKTVLVKECYSVFTNR---RPFGPVGFMKSAISIAED----EEWKRLRSLLSPTFTSGKLKEMVP---------IIAQYGDVLVRNLRREAETGKPVTLKDVFGAYSMDVITSTSFGVN
CYP46A1Hum         SVKKFLMSTKY-NKDSKMYRALQTVFGERLFGQGLVSEC-NYERWHKQRRVIDLAFSRSSLVSLMET---------FNEKAEQLVEILEAKADGQTPVSMQDMLTYTAMDILAKAAFG--
CYP11A1Hum         DVALLFKSEGP-NPERFLIPPWVAYHQYYQRPIGVLLK--KSAAWKKDRVALNQEVMAPEATKNFLP----LLDAVSRDFVSVLHRRIKKAGSGNYSGDISDDLFRFAFESITNVIFGER
CYP19A1Hum         SMFHIMKHNHYSSRFGS----KLGLQCIGMHEKGIIFNN-NPELWKTTRPFFMKALSGPGLVRMVT---------VCAESLKTHLDRLEEVTNESGYVDVLTLLRRVMLDTSNTLFLR--
CYP8A1Hum          YDAVVWEPRTRLDFHAY----AIFLMERIFDVQLPHYSP-SDEKARMKLTLLHRELQALTEAMYTN------LHAVLLGDATEAGSGWHEMGL--LDFSYSFLLRAGYLTLYGIEALPRT
                                                                                                                                                .
```


CYP2C4Rab FDYKDEEFLKLMERLNENIRILSS----PWLQVYN----NFPALLDYFPGIHKTLLKNA-DYTKNFIMEKVKEHQKLLDVNNP-RDFIDCFLIKMEKENN---LEFTLGSLVIAVFDLFG
CYP2C5Rab FDYKDEEFLKLMESLNENVRILSS----PWLQVYN----NFPALLDYFPGIHKTLLKNA-DYIKNFIMEKVKEHQKLLDVNNP-RDFIDCFLIKMEQENN---LEFTLESLVIAVSDLFG
CYP2C9Hum FDYKDQQFLNLMEKLNENIKILSS----PWIQICN----NFSPIIDYFPGTHNKLLKNV-AFMKSYILEKVKEHQESMDMNNP-QDFIDCFLMKMEKEKHNQPSEFTIESLENTAVDLFG
CYP2C19Hum FDYKDQQFLNLMEKLNENIRIVST----PWIQICN----NFPTIIDYFPGTHNKLLKNL-AFMESDILEKVKEHQESMDINNP-RDFIDCFLIKMEKEKQNQQSEFTIENLVITAADLLG
CYP2C8Hum FDYKDQNFLTLMKRFNENFRILNS----PWIQVCN----NFPLLIDCFPGTHNKVLKNV-ALTRSYIREKVKEHQASLDVNNP-RDFIDCFLIKMEQEKDNQKSEFNIENLVGTVADLFV
CYP2C2Rab FDYTDQDFLSLMGKFNENFKILNS----PWVQFCN----CFPILFDYFPGSHRKAVKNI-FYVKNYITEQIKEHQKSLDINNP-RDFIDCFLIKMEQEKCNQQSEFTIENLLTTVSDVFM
CYP2C3Rab FDYDDEKFKTLIKYFHENFELLGT----PWIQLYN----IFP-ILHYLPGSHRQLFKNI-DGQIKFILEKVQEHQESLDSNNP-RDFVDHFLIKMEKEKHKKQSEFTMDNLITTIWDVFS
CYP2C21Dog FEYDDKDFLTLLEYFHENLLISST----SWIQLYN----AFPLLIHYLPGSHHVLFKNI-ANQFKFISEKIKEHEESLNFSNP-RDFIDYFLIKIEKEKHNKQSEFTMDNLIITIWDVFS
CYP2E1Hum FDYNDEKFLRLMYLFNENFHLLST----PWLQLYN----NFPSFLHYLPGSHRKVIKNV-AEVKEYVSERVKEHHQSLDPNCP-RDLTDCLLVEMEKEKHSAERLYTMDGITVTVADLFF
CYP2A4Mou FDYEDKEFLSLLRMMLGSLQFTAT----SMGQVYE----MFSSVMKHLPGPQQQAFKEL-QGLEDFITKKVEHNQRTLDPNSP-RDFIDSFLIRMLEEKKNPNTEFYMKNLVLTTLNLFF
CYP2A5Mou FDYEDKEFLSLLRMMLGSFQFTAT----SMGQLYE----MFSSVMKHLPGPQQQAFKEL-QGLEDFITKKVEHNQRTLDPNSP-RDFIDSFLIRMLEEKKNPNTEFYMKNLVLTTLNLFF
CYP2A6Hum FDYKDKEFLSLLRMMLGIFQFTST----STGQLYE----MFSSVMKHLPGPQQQAFQLL-QGLEDFIAKKVEHNQRTLDPNSP-RDFIDSFLIRMQEEEKNPNTEFYLKNLVMTTLNLFI
CYP2A13Hum FDYEDKEFLSLLRMMLGSFQFTAT----STGQLYE----MFSSVMKHLPGPQQQAFKEL-QGLEDFIAKKVEHNQRTLDPNSP-RDFIDSFLIRMQEEEKNPNTEFYLKNLVMTTLNLFF
CYP2A1Rat FDYEDTEFLSLLQMMGQMNRFAAS----PTGQLYD----MFHSVMKYLPGPQQQIIKVT-QKLEDFMIEKVRQNHSTLDPNSP-RNFIDSFLIRMQEE-KNGNSEFHMKNLVMTTLSLFF
CYP2G1Rat FDYEDQRFRSLMKMINESFVEMSM----PWAQLYD----MYWGVIQYFPGRHNRLYNLI-EELKDFIASRVKINEASFDPSNP-RDFIDCFLIKMYQDKSDPHSEFNLKNLVLTTLNLFF
CYP2B1Rat FDYTDRQFLRLLELFYRTFSLLSS----FSSQVFE----FFSGFLKYFPGAHRQISKNL-QEILDYIGHIVEKHRATLDPSAP-RDFIDTYLLRMEKEKSNHHTEFHHENLMISLLSLFF
CYP2B2Rat FDYTDRQFLRLLELFYRTFSLLSS----FSSQVFE----FFSGFLKYFPGAHRQISKNL-QEILDYIGHIVEKHRATLDPSAP-RDFIDTYLLRMEKEKSNHHTEFHHENLMISLLSLFF
CYP2B4Rab FDYKDPVFLRLLDLFFQSFSLISS----FSSQVFE----LFPGFLKHFPGTHRQIYRNL-QEINTFIGQSVEKHRATLDPSNP-RDFIDVYLLRMEKDKSDPSSEFHHQNLILTVLSLFF
CYP2B5Rab FDYKDPVFLRLLDLFFQSFSLISS----FSSQVFE----LFPGFLKHFPGTHRQIYRNL-QEINTFIGQTVEKHRATLDPSNP-RDFIDVYLLRMEKDKSDPSSEFHHRNLILTVLTLFF
CYP2B6Hum FHYQDQEFLKMLNLFYQTFSLISS----VFGQLFE----LFSGFLKYFPGAHRQVYKNL-QEINAYIGHSVEKHRETLDPSAP-KDLIDTYLLHMEKEKSNAHSEFSHQNLNLNTLSLFF
CYP2B11Dog FGYKDPEFLRLMNLFYVSFALISS----FSSQMFE----LFHSFLKYFPGTHRQVYNNL-QEIKAFIARMVEKHRETLDPSAP-RDFIDAYLIRMDKEKAEPSSEFHHRNLIDTALSLFF
CYP2S1Hum FSYEDKEFQAVVRAAGGTLLGVSS----QGGQTYE----MFSWFLRPLPGPHKQLLHHV-STLAAFTVRQVQQHQGNLDASGPARDLVDAFLLKMAQEEQNPGTEFTNKNMLMTVIYLLF
CYP2F1Hum FDYDDERLLTIIRLINDNFQIMSS----PWGELYDILDPRFPSLLDWVPGPHQRIFQNF-KCLRDLIAHSVHDHQAS----SP-RDFIQCFLTKMAEEKEDPLSHFHMDTLLMTTHNLLF
CYP2T1Rat YNYGDPEFLRLLDLFSDNFRIMSS----RWGETYN----MFPSFMDWIPGPHHRIFKNF-QELRLFISEQIQWHRQSRQTGEP-RDFIDCFLEQMDKEHQDPESHFQDETLVMTTHNLFF
CYP2D1Rat FEYEDPYLIRMVKLVEESLTEVSG----FIPEVLN----TFPALLRIP-GLADKVFQGQ-KTFMALLDNLLAENRTTWDPAQPPRNLTDAFLAEVEKAKGNPESSFNDENLRMVVVDLFT
CYP2D6Hum FEYDDPRFLRLLDLAQEGLKEESG----FLREVLN----AVPVLLHIP-ALAGKVLRFQ-KAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFS
CYP2J2Hum FEYQDSWFQQLLKLLDEVTYLEAS----KTCQLYN----VFPWIMKFLPGPHQTLFSNW-KKLKLFVSHMIDKHRKDWNPAET-RDFIDAYLKEMSKHTGNPTSSFHEENLICSTLDLFF
CYP2R1Hum FTYEDTDFQHMIELFSENVELAAS----ASVFLYN----AFPWIGILPFGKHQQLFRNA-AVVYDFLSRLIEKASVNRKPQLP-QHFVDAYLDEMDQGKNDPSSTFSKENLIFSVGELII
CYP2W1Hum FDYRDPVFVSLLGLIDEVMVLLGS----PGLQLFN----VYPWLGALL-QLHRPVLRKI-EEVRAILRTLLEARRPHVCPGDPVCSYVDALIQQGQG--DDPEGLFAEANAVACTLDMVM
CYP1A2Hum FPESSDEMLSLVKNTHEFVETASSGNPLDFFP----------ILRYLPNPALQRFKAFNQRFLWFLQKTVQEHYQDFDKNSV-RDITGALFKHSKKGPRASGNLIPQEKIVNLVNDIFG
CYP3A4Hum IDSLNNPQDPFVENTKKLLRFDFLD---PFFLSIT----VFPFLIPILEVLNICVFPREVTNFLRKSVKRMKESRLEDTQKHR-VDFLQLMIDSQNSKETESHKALSDLELVAQSIIFIF
CYP46A1Hum --METSMLLGAQKPLSQAVKLMLEG----ITASRN-------TLAKFLPGKRKQLREVR-ESIRFLRQVGRDWVQRRREALKRGEEVPADILTQILKAEEG---AQDDEGLLDNFVTFFI
CYP11A1Hum QGMLEEVVNPEAQRFIDAIYQMFH----TSVPMLN----LPPDLFRLFRTKTWKDHVAAWDVIFSKADIYTQNFYWELRQKGSVHHDYRGILYRLLGD-----SKMSFEDIKANVTEMLA
CYP19A1Hum IPLDESAIVVKIQGYFDAWQALLIK----------PDIFFKISWLYKKYEKSV-KDLKDAIEVLIAEKRRRISTEEKLEECMDFATELILAEKRG---DLTRENVNQCILEMLI
CYP8A1Hum HESQAQDRVHSADVFHTFRQLDRL----------LPKLARGSLSVGDKDHMCSVKSRLWKLLSPARLARRAHRS--KWLESYLLHLEEMG------VSEEMQARALVLQLW

```
CYP2C4Rab          AGTETTSTTLRYSLLLLLKHPEVAARVQEEIERVIGRHRSPCMQDRSH-------MPYTDAVIHEIQRFIDLLPTNLPHAVTRDVKFRN---YFIPKG-TDIITSLTSVLHDEKAFPNPK
CYP2C5Rab          AGTETTSTTLRYSLLLLLKHPEVAARVQEEIERVIGRHRSPCMQDRSR-------MPYTDAVIHEIQRFIDLLPTNLPHAVTRDVRFRN---YFIPKG-TDIITSLTSVLHDEKAFPNPK
CYP2C9Hum          AGTETTSTTLRYALLLLLKHPEVTAKVQEEIERVIGRNRSPCMQDRSH-------MPYTDAVVHEVQRYIDLLPTSLPHAVTCDIKFRN---YLIPKG-TTILISLTSVLHDNKEFPNPE
CYP2C19Hum         AGTETTSTTLRYALLLLLKHPEVTAKVQEEIERVIGRNRSPCMQDRGH-------MPYTDAVVHEVQRYIDLIPTSLPHAVTCDVKFRN---YLIPKG-TTILTSLTSVLHDNKEFPNPE
CYP2C8Hum          AGTETTSTTLRYGLLLLLKHPEVTAKVQEEIDHVIGRHRSPCMQDRSH-------MPYTDAVVHEIQRYSDLVPTGVPHAVTTDTKFRN---YLIPKG-TTIMALLTSVLHDDKEFPNPN
CYP2C2Rab          AGTETTSTTLRYGLLLLMKHPEVIAKVQEEIERVIGRHRSPCMQDRSR-------MPYTDATVHEIQRYINLIPNNVPHTTICNLKFRN---YLIPKG-TDVLTSLSSVLHDDKEFPNPD
CYP2C3Rab          AGTDTTSNTLKFALLLLLKHPEITAKVQEEIEHVIGRHRSPCMQDRTR-------MPYTDAVMHEIQRYVDLVPTTLPHAVTQDIEFNG---YLIPKG-TDIIPSLTSVLYDDKEFPNPE
CYP2C21Dog         AGTETTSTTLRYGLLVLLKHPDVTAKVQEEIHRVVGRHRSPCMQDRSC-------MPYTDAVVHEIQRYIDLVPNNLPHSVTQDIKFRE---YLIPKG-TTILTSLTSVLHDEKGFPNPD
CYP2E1Hum          AGTETTSTTLRYGLLILMKYPEIEEKLHEEIDRVIGPSRIPAIKDRQE-------MPYMDAVVHEIQRFITLVPSNLPHEATRDTIFRG---YLIPKG-TVVVPTLDSVLYDNQEFPDPE
CYP2A4Mou          AGTETVSTTLRYGFLLLMKYPDIEAKVHEEIDRVIGRNRQPKYEDRMK-------MPYTEAVIHEIQRFADLIPMGLARRVTKDTKFRD---FLLPKG-TEVFPMLGSVLKDPKFFSNPK
CYP2A5Mou          AGTETVSTTLRYGFLLLMKHPDIEAKVHEEIDRVIGRNRQPKYEDRMK-------MPYTEAVIHEIQRFADMIPMGLARRVTKDTKFRD---FLLPKG-TEVFPMLGSVLKDPKFFSNPK
CYP2A6Hum          GGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRAK-------MPYMEAVIHEIQRFGDVIPMSLARRVKKDTKFRD---FFLPKG-TEVYPMLGSVLRDPSFFSNPQ
CYP2A13Hum         AGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRAK-------MPYTEAVIHEIQRFGDMLPMGLAHRVNKDTKFRD---FFLPKG-TEVFPMLGSVLRDPRFFSNPR
CYP2A1Rat          AGSETVSSTLRYGFLLLMKHPDVEAKVHEEIEQVIGRNRQPQYEDHMK-------MPYTQAVINEIQRFSNLAPLGIPRRIIKNTTFRG---FFLPKA-TDVFPILGSLMTDPKFFPSPK
CYP2G1Rat          AGTETVSSTLRYGFLLLMKYPEVEAKIHEEINQVIGTHRTPRVDDRAK-------MPYTDAVIHEIQRLTDIVPLGVPHNVIRDTHFRG---YFLPKG-TDVYPLIGSVLKDPKYFRYPE
CYP2B1Rat          AGTETSSTTLRYGFLLMLKYPHVAEKVQKEIDQVIGSHRLPTLDDRSK-------MPYTDAVIHEIQRFSDLVPIGVPHRVTKDTMFRG---YLLPKN-TEVYPILSSALHDPQYFDHPD
CYP2B2Rat          AGTETGSTTLRYGFLLMLKYPHVTEKVQKEIDQVIGSHRPPSLDDRTK-------MPYTDAVIHEIQRFADLAPIGLPHRVTKDTMFRG---YLLPKN-TEVYPILSSALHDPQYFDHPD
CYP2B4Rab          AGTETTTSTTLRYGFLLMLKYPHVTERVQKEIEQVIGSHRPPALDDRAK-------MPYTDAVIHEIQRLGDLIPFGVPHTVTKDTQFRG---YVIPKN-TEVFPVLSSALHDPRYFETPN
CYP2B5Rab          AGTETTSTTLRYGFLLMLKYPHVTERVQKEIEQVIGSHRPPALDDRAK-------MPYTDAVIHEIQRLGDLVPFGAPHMVTKDTQFRG---YVIPKN-TEVFPVLSSALHDPRYFETPN
CYP2B6Hum          AGTETTSTTLRYGFLLMLKYPHVAERVYREIEQVIGPHRPPELHDRAK-------MPYTEAVIYEIQRFSDLLPMGVPHIVTQHTSFRG---YIIPKD-TEVFLILSTALHDPHYFEKPD
CYP2B11Dog         AGTETTSTTLRYGFLLMLKYPHIAERIYKEIDQVIGPHRLPSLDDRAK-------MPYTDAVIHEIQRFGDLLPIGVPHMVTKDICFRG---YIIPKG-TEVFPILHSALNDPHYFEKPD
CYP2S1Hum          AGTMTVSTTVGYTLLLLMKYPHVQKWVREELNRELGAGQAPSLGDRTR-------LPYTDAVLHEAQRLLALVPMGIPRTLMRTTFRFG---YTLPQG-TEVFPLLGSILHDPNIFKHPE
CYP2F1Hum          GGTKTVSTTLHHAFLALMKYPKVQARVQEEIDLVVGRARLPALKDRAA-------MPYTDAVIHEVQRFADIIPMNLPHRVTRDTAFRG---FLIPKG-TDVITLLNTVHYDPSQFLTPQ
CYP2T1Rat          GGTETTSTTLRYGLLIMLKYPEVAAKVQEELDATVGRTRAPSLADRAH-------LPYTNAVLHEIQRFISVLPLGLPRALIRDVNLRN---HFLHKG-TFVIPLLVSAHRDPTQFKDPD
CYP2D1Rat          AGMVTTATTLTWALLLMILYPDVQRRVQQEIDEVIGQVRCPEMTDQAH-------MPYTNAVIHEVQRFGDIAPLNLPRFTSCDIEVQD---FVIPKG-TTLIINLSSVLKDETVWEKPH
CYP2D6Hum          AGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAH-------MPYTTAVIHEVQRFGDIVPLGVTHMTSRDIEVQG---FRIPKG-TTLITNLSSVLKDEAVWEKPF
CYP2J2Hum          AGTETTSTTLRWALLYMALYPEIQEKVQAEIDRVIGQGQQPSTAARES-------MPYTNAVIHEVQRMGNIIPLNVPREVTVDTTLAG---YHLPKG-TMILTNLTALHRDPTEWATPD
CYP2R1Hum          AGTETTTNVLRWAILFMALYPNIQGQVQKEIDLIMGPNGKPSWDDKCK-------MPYTEAVLHEVLRFCNIVPLGIFHATSEDAVVRG---YSIPKG-TTVITNLYSVHFDEKYWRDPE
CYP2W1Hum          AGTETTSATLQWAALLMGRHPDVQGRVQEELDRVLGPGRTPRLEDQQA-------LPYTSAVLHEVQRFITLLPH-VPRCTAADTQLGG---FLLPKG-TPVIPLLTSVLLDETQWQTPG
CYP1A2Hum          AGFDTVTTAISWSLMYLVTKPEIQRKIQKELDTVIGRERRPRLSDRPQ-------LPYLEAFILETFRHSSFLPFTIPHSTTRDTTLNG---FYIPKK-CCVFVNQWQVNHDPELWEDPS
CYP3A4Hum          AGYETTSSVLSFIMYELATHPDVQQKLQEEIDAVLPNKAPPTYDTVLQ-------MEYLDMVVNETLRLFP-IAMRLERVCKKDVEING---MFIPKG-VVVMIPSYALHRDPKYWTEPE
CYP46A1Hum         AGHETSANHLAFTVMELSRQPEIVARLQAEVDEVIGSKRYLDFEDLGR-------LQYLSQVLKESLRLYP-PAWGTFRLLEEETLIDG---VRVPGN-TPLLFSTYVMGRMDTYFEDPL
CYP11A1Hum         GGVDTTSMTLQWHLYEMARNLKVQDMLRAEVLAARHQAQGDMATMLQL-------VPLLKASIKETLRLHP-ISVTLQRYLVNDLVLRD---YMIPAK-TLVQVAIYALGREPTFFFDPE
CYP19A1Hum         AAPDTMSVSLFFMLFLIAKHPNVEEAIIKEIQTVIG-ERDIKIDDIQK-------LKVMENFIYESMRYQPVVDLVMRKALEDD-VIDG---YPVKKG-TNIILNIGRMHR-LEFFPKPN
CYP8A1Hum          ATQGNMGPAAFWLLLFLLKNPEALAAVRGELESILWQAEQPVSQTTTLPQKVLDSTPVLDSVLSESLRLTAAPFITREVVVDLAMPMADGREFNLRRGDRLLLFPFLSPQRDPEIYTDPE
                   .   .          :    .    :  *:                              : *  *                .      :     :            :  *
```

Figure A1: Alignment of mammalian CYP450 isoforms showing ligands contacts identified in X-ray crystal structures of CYP450-ligand complexes.

Ligand contacts highlighted here were identified in 43 different CYP450-ligand complexes deposited in the PDB before 2011. Residues are colour-coded based on the number of times they are involved in ligand contacts in different structures of the same isoform as follows: Yellow (1), Blue (2), Turquoise (3), Green (4), Red (5) and Dark red (6).


CYP1A2Hum
CYP46A1Hum
CYP2A6Hum
CYP2B4Rab
CYP2C5Rab
CYP2C8Hum
CYP2C9Hum
CYP2D6Hum
CYP2R1Hum
CYP3A4Hum

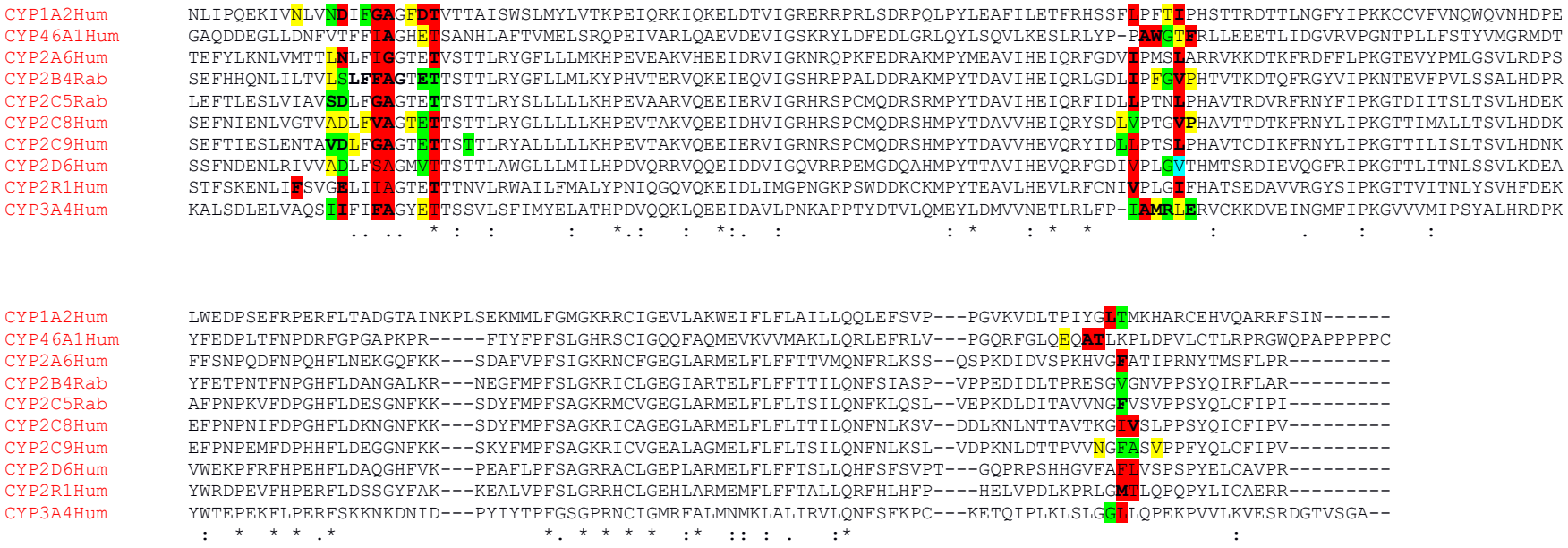


Figure A2: Alignment of mammalian CYP450 isoforms showing ligands contacts identified in docking CYP450-ligand complexes.

Residues are colour-coded based on the frequency with which they interact with ligands in docking complexes. Yellow (less than 20 % of contacts), Green (20-80 %) and Red (> 80 %).

Appendix B. **CYP450 SNP tables**

Table B1: CYP1A2 SNP Table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY		EXPERIMENTAL DATA FROM THE LITERATURE									
	Δ Volume (Å^3)	RSA[2] (%)		2° structure location	Functional region location	SDM Score (kcal. mol^{-1})	Effect	Expression system	Relative expression levels	Relative holoprotein levels [3]	Substrate	Relative activity [4]	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
		wt	mut														
S18C[1]	19.5	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a										unknown
F21L[1]	-23.2	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a										unknown
P42R	60.7	2.6	0.0	loop btwn N-term & helix A	proline rich motif	-0.3	neutral	Chinese hamster V79/Sf21	decreased (0.66)	ND	7-ERF		2.88	0.02*	0.01*	[607]	inactive
											Phenacetin	>0.04					
E44K	30.2	30.2	45.6	loop btwn N-term & helix A	proline rich motif	-2.0	damaging										unknown
L65P	-54	0.4	3.6	α-helix A	undefined	-4.8	damaging										unknown
G73R	113.3	61.1	46.6	end of helix A	undefined	1.8	neutral										unknown
T83M[1]	46.8	41.5	62.2	B-sheet1[2]	SRS1'b	0.1	neutral	Chinese hamster V79	similar		7-ERF	similar	0.90	0.72	0.76	[608]	neutral
											Phenacetin	similar					
D104N	3.0	14.2	32.8	loop btwn helix B & B'	SRS1	-2.2	damaging										unknown
L111F	23.2	17.4	11.7	loop btwn helix B &	SRS1	-0.1	neutral										unknown

				B'													
F125I	-23.2	11.4	4.5	B-C loop	SRS1	-0.9	neutral										unknown
R138H	-20.2	79.0	68.1	α-helix C	CPR contact /SRS1	-0.6	neutral										unknown
E168Q[1]	5.4	10.1	10.7	α-helix D	undefined	0.0	neutral	Chinese hamster V79	similar		7-ERF	similar	0.53*	0.31	0.59	[608]	decreased activity
											Phenacetin	similar					
M180V	-22.9	20.5	25.1	α-helix D	undefined	-0.9	neutral										unknown
F186L[1]	-23.2	6.7	26.8	loop btwn helix D & E	undefined	-1.0	neutral	Chinese hamster V79	similar		7-ERF	0.28	0.27*	0.06*	0.20*	[608]	decreased activity
											Phenacetin	0.125					
F205V	-49.9	0.0	7.7	end of helix E	undefined	-1.4	neutral										unknown
S212C[1]	19.5	111.9	111.7	loop btwn helix E&F	undefined	-1.9	neutral	Chinese hamster V79	similar		7-ERF	similar	1.55*	0.78	0.41	[608]	decreased activity
											Phenacetin	similar					
R281W	54.4	85.2	80.1	loop btwn helix G & H	undefined	-0.2	neutral										unknown
S298R[1]	84.4	75.6	79.8	B-sheet 2	undefined	1.0	neutral										unknown
G299A	28.5	102.7	89.0	B-sheet 2	undefined	-5.2	damaging	Chinese hamster V79	similar		7-ERF	similar	1.03	0.94	0.90	[608]	neutral
											Phenacetin	similar					
G299S[1]	28.9	102.7	92.3	B-sheet 2	undefined	-4.6	damaging										unknown
I314V[1]	-26.7	5.2	8.3	α-helix I	haem contact/SRS 4	-1.0	neutral										unknown
D348N[1]	3.0	57.5	39.7	α-helix J	undefined	0.8	neutral	*E.coli*	slightly decreased	0.3	MelQ		0.65	1.37	2.10	[317]	substrate specific
											MelQx		1.22	1.67	1.40		
											IQ		1.29	2.00	1.60		
											PhIP		0.87	2.64	3.00		
											GluP-1		3.25	2.60	0.90		
											Phenacetin		0.72	0.60	0.90		
R356Q	-29.6	42.7	46.7	loop btwn	undefined	-4.1	damagin										unknown

				helix J & K			g										
R377Q[1]	-29.6	0.0	3.4	α-helix K	ERR triad/core stabilising motif	-1.1	neutral	Chinese hamster V79	decreased (0.3)	ND	7-ERF		8.92	0.01	0.00	[607]	inactive
											Phenacetin	ND					
I386F[1]	23.2	14.8	27.7	loop btwn helix k & B-sheet 3[1]	haem contact/SRS5	0.1	neutral	*E.coli*	slightly decreased	decreased (0.5)	MelQ		0.17	0.27	1.50	[317]	substrate specific
											MelQx		1.70	3.33	2.00		
											IQ		1.76	0.50	0.30		
											PhIP		4.80	1.82	0.40		
											GluP-1		4.00	2.80	0.70		
											Phenacetin		12.80	1.73	0.10		
T395M[1]	46.8	43.7	53.0	B-sheet 4[1]	undefined	-1.0	neutral										unknown
N397H[1]	39.1	55.9	78.0	loop btwn B-sheet 4[1]& 4[2]	undefined	-1.0	neutral										unknown
C406Y[1]	85.1	7.1	22.6	B-sheet 3[2]	undefined	0.0	neutral	*E.coli*	slightly decreased	similar (0.9)	MelQ		0.54	1.50	2.60	[317]	substrate specific
											MelQx		0.74	1.41	1.90		
											IQ		1.05	1.43	1.40		
											PhIP		13.20	2.36	2.50		
											GluP-1		2.75	2.60	1.00		
											Phenacetin		2.48	0.80	0.30		
R431W[1]	54.4	9.0	9.9	3_{10} helix	ERR triad/meander	2.3	damaging	*E.coli*	ND	ND						[317]	inactive
F432S[1]	-100.9	0.0	3.5	3_{10} helix	meander	-5.3	damaging										unknown
D436N[1]	3.0	82.5	102.5	loop btwn 3_{10} helix and helix L	undefined	-0.8	neutral										unknown
T438I[1]	50.6	68.2	87.8	loop btwn 3_{10} helix and helix L	undefined	-0.8	neutral	Chinese hamster V79	similar		7-ERF	similar	0.92	1.00	1.10	[608]	neutral
											Phenacetin	similar					
R456H[1]	-20.2	3.2	25.7	loop btwn	Cys	-1.5	neutral	Chinese	decreased	ND	7-ERF		3.60	0.02	0.01	[607]	inactive

				3_{10} helix and helix L	pocket/haem contact/CPR contact (ion charge cluster residue)			hamster V79/ Sf9	(0.3)		Phenacetin	>0.04					
R457W	54.4	44.0	46.8	loop btwn 3_{10} helix and helix L	Cys-pocket/CPR contact	-0.4	neutral										unknown
Q478H	-9.4	19.5	45.2	end of helix L	undefined	0.3	neutral	*E.coli*	similar	similar	7-ERF				0.90	[609]	neutral
P485R[1]	60.7	98.7	107.7	loop btwn B-sheet 5^1& 5^2	undefined	-1.6	neutral										unknown
R510Q[1]	-29.6	69.5	86.2	B-sheet 5^2	undefined	-0.3	neutral										unknown
I514V	-26.7	n/a	n/a	truncated C-terminal	undefined	n/a	n/a										unknown

Substrate abbreviations

7-ERF	7-ethoxyresorufin
Phenacetin	*N*-(4-Ethoxyphenyl)acetamide
MeIQ	2-amino-2,4-dimethylimidazo[4,5-f]quinoline
MeIQx	2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline
IQ	2-amino-3-methylimidazo[4,5-f]quinoline
PhIP	2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine
GluP-1	2-amino-6-methyldipyrido[1,2-a :3`,2`-d]imidazole

1 SNPs found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.
2 RSA refers to the relative solvent accessibility as calculated by SDM.
3 *Relative holoprotein* levels refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.
4 *Relative activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.
ND refers to protein levels, activity or kinetic parameters that were not detectable and the asterisk (*) indicates data reported to be significantly different from the wild-type enzyme. (Note that some studies provide quantitative data without indicating whether differences between wild-type and mutant proteins are statistically significant)

Table B2: CYP2A6 SNP table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY			EXPERIMENTAL DATA FROM THE LITERATURE									
	Δ Volume (Å^3)	RSA[2] (%)		2° structure location	Functional region location	SDM Score (kcal. mol^{-1})	SD[3]	Effect	Expression system	Relative expression levels	Relative holoprotein levels [4]	Substrate	Relative activity[5]	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
		wt	mut															
G5R [1]	113.3	n/a	n/a	truncated N-terminal	undefined	n/a	n/a	n/a										unknown
M6L[1]	3.8	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a	n/a										unknown
S29N[1]	25.1	n/a	n/a	truncated N-terminal	halting signal	n/a	n/a	n/a										unknown
V110L[1]	26.7	2.7	5.8	α-helix B'	SRS1	0.3	0.2	neutral	*E.coli*	decreased	decreased thermal stability	Nicotine		1.61	0.98	0.59*	[610]	decreased activity
F118L[1]	-23.2	7.2	4.3	B-C loop	SRS1/gating residue	-0.4	0.2	neutral	*In vivo*				similar				[611]	unknown
G121R	113.3	84.0	75.4	start of helix C	SRS1	2.7	0.0	damaging										unknown
R128Q[1]	-29.6	22.0	15.8	α-helix C	haem contact/CPR contact	-0.9	0.7	neutral	insect	similar	ND	Coumarin		5.36	0.10	0.02	[612]	inactive
R128L[1]	-6.7	22.0	17.4	α-helix C	haem contact/CPR contact	1.2	1.1	neutral (6/10)										unknown
S131A[1]	-0.4	0.3	0.1	α-helix C	SRS1	2.9	0.0	damaging										unknown

E151K	30.2	61.6	72.4	α-helix D	undefined	-1.1	0.0	neutral											unknown
D158E[1]	27.3	77.9	63.3	α-helix D	undefined	1.5	0.1	neutral											unknown
L160H[1]	-13.5	0.2	0.3	α-helix D	undefined	-1.9	0.1	neutral	HepG2		ND	Coumarin	ND				[613]	inactive
									In vivo			Coumarin	ND				[310]	
L160I[1]	0.0	0.2	0.8	α-helix D	undefined	-0.4	0.0	neutral											unknown
D169E	27.3	55.3	71.8	loop btwn helix D&E	undefined	-1.0	0.1	neutral											unknown
K194E[1]	-30.2	95.7	88.4	loop btwn helix E&F	undefined	-0.4	0.0	neutral	*E.coli*	similar	similar	Coumarin		2.10*	1.80*	0.85	[614]	neutral	
R203S[1]	-84.4	62.2	55.2	α-helix F	SRS(2,3)	-1.2	0.0	neutral	*E.coli*	similar		Coumarin	same				[615]	neutral	
												Nicotine	same	1.26	0.93	0.73			
									E.coli	similar	similar	Coumarin		0.99	1.05	1.05	[614]		
R203C[1]	-64.9	62.2	50.8	α-helix F	SRS(2,3)	-7.2	0.0	damaging	*E.coli*	decreased (0.6)		Coumarin	decreased (0.3)				[615]	decreased activity	
												Nicotine	decreased (0.3)	1.33	0.25	0.19			
									In vivo			Nicotine	Decreased (0.6)						
S224P[1]	23.7	24.5	12.8	α-helix G'	SRS(2,3)	1.6	0.4	neutral	*E.coli*		decreased (0.62)	Tegafur		1.42	0.59	0.41	[614]	decreased activity	
												Coumarin		1.71	0.97	0.59			
Q239K	24.8	79.6	74.0	α-helix G	SRS(2,3)	-0.1	0.0	neutral											unknown
R257C	-64.9	72.4	92.6	End of α-helix G	undefined	-4.6	2.4	damaging											unknown

T258K	52.5	67.8	70.1	loop btwn helix G & H	undefined	0.7	0.2	neutral											unknown
R265Q	-29.6	44.9	80.6	loop btwn helix G & H	undefined	-0.6	0.9	neutral											unknown
E279Q	5.4	74.4	75.2	loop btwn helix H & I	undefined	-0.1	0.3	neutral											unknown
V292M	22.9	9.3	24.4	α-helix I	SRS4	0.7	0.0	neutral											unknown
T294S	-27.1	0.1	0.5	α-helix I	SRS4	-1.0	0.5	neutral											unknown
G301A	28.5	77.3	58.2	α-helix I	catalytic groove/haem contact/SRS4	3.5	1.2	damaging											unknown
R311C	-64.9	7.4	22.5	α-helix I	SRS4	9.8	0.7	damaging											unknown
Y351H[1]	-40.4	15.3	34.9	α-helix K	undefined	-1.4	0.7	neutral (7/10)											unknown
V365M[1]	22.9	10.7	5.1	loop btwn helix K & B-sheet 3[1]	SRS5	1.5	0.3	neutral	*E.coli*			Coumarin		1.77*	1.05	0.58*	[616]	substrate specific	
												Nicotine		0.98	0.40*	0.40*			
									In vivo			Nicotine	decreased						
									E.coli	similar		Coumarin	similar				[615]		
												Nicotine	decreased (0.3)	1.52	0.47*	0.31*			
									E.coli	decreased thermal stability		Nicotine		0.47*	0.68	0.14*	[610]		
E390K	30.2	5.1	12.0	B-sheet 3[2]	undefined	-2.3	1.1	damaging											unknown

								g										
Y392F[1]	-82.5	7.4	6.5	B-sheet 3^2	undefined	1.3	0.1	neutral	*E.coli*		similar	Coumarin		1.7*	0.83*	0.50*	[313]	substrate specific
												Nicotine		1.41*	1.24*	0.88		
												Tegafur		2.33*	0.84*	0.36*		
Q409R	29.6	81.7	96.3	loop btwn B-sheet 3^2 & helix K'	meander	0.0	0.0	neutral										unknown
N418S	-25.1	50.9	57.1	loop btwn B-sheet 3^2 & helix K'	undefined	1.2	0.1	neutral										unknown
N418D[1]	-3.0	50.9	58.3	loop btwn B-sheet 3^2 & helix K'	undefined	1.6	0.0	neutral										unknown
E419D[1]	-27.3	106.9	114.8	loop btwn B-sheet 3^2 & helix K'	undefined	0.2	0.0	neutral	*E.coli*	similar	similar	Coumarin				1.06	[609]	neutral
G435R	113.3	18.3	6.1	loop btwn 3_{10} helix & helix L	Cys-pocket/CPR contact	-1.3	0.6	neutral										unknown
N438Y[1]	79.5	38.1	45.2	loop btwn 3_{10} helix & helix L	Cys-pocket/CPR contact	2.1	0.8	damaging	*E.coli*	decreased		Nicotine	decreased	1.41	0.92	0.63*	[610]	decreased activity
I471T[1]	-50.6	1.7	13.0	btwn B-sheet 5^1 & 5^2	SRS6	-2.3	0.2	damaging	*E.coli*	decreased	Decreased (decreased thermal stability)	Coumarin				0.63	[314]	substrate specific
												Nicotine				ND		
									E.coli		similar	Coumarin		2.5*	0.21*	0.08*	[313]	
												Nicotine		0.57*	0.17*	0.28*		
												Tegafur		3.33*	0.27*	0.08*		
									In vivo			Coumarin	decrease				[617]	

													d					
												Nicotine	decreased				[617, 618]	
K476R[1]	4.8	32.5	42.1	btwn B-sheet 5[1] & 5[2]	SRS6	1.3	0.6	neutral	*E.coli*	similar	similar	Coumarin		1.54*	1.41	0.91		decreased activity?
									In vivo			Nicotine	decreased				[611]	
G479V[1]	79.9	2.0	0.9	btwn B-sheet 5[1] & 5[2]	SRS6	2.1	0.0	damaging	*S. cerevisiae expression*	very low	ND	Coumarin				ND	[619]	inactive
R485L[1]	54.0	44.8	34.7	btwn B-sheet 5[1] & 5[2]	SRS6	1.3	0.1	neutral	*In vivo*			Coumarin	similar				[617]	neutral
									In vivo			Nicotine	similar					

Substrate abbreviations

Nicotine	3-[(2S)-1-methylpyrrolidin-2-yl]pyridine
Coumarin	chromen-2-one
Tegafur	5-fluoro-1-(oxolan-2-yl)pyrimidine-2,4-dione

1 SNPs found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.
2 *RSA* refers to the relative solvent accessibility as calculated by SDM.
3 *SD* refers to the standard deviation of the SDM score across different crystal structures of the same isoform.
4 *Relative holoprotein* levels refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.
5 *Relative activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.
ND refers to protein levels, activity or kinetic parameters that were not detectable and the asterisk (*) indicates data reported to be significantly different from the wild-type enzyme. (Note some studies provide quantitative data without indicating whether differences between wild-type and mutant proteins are statistically significant)

Table B3: CYP2B6 SNP table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY		EXPERIMENTAL DATA FROM THE LITERATURE									
	Δ Volume (Å^3)	RSA[2] (%)		2° structure	Functional Region	SDM Score (kcal.mol^{-1})	Effect	Expression system	Relative expression levels	Relative holoprotein levels [3]	Substrate	Relative activity [4]	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
Q21L[1]	22.9	n/a	n/a	Truncated N-terminal	signal-anchor sequence	n/a											unknown
R22C[1]	-64.9	n/a	n/a	Truncated N-terminal	halting signal	n/a		COS-1	similar		7-EFC	similar	1.14	0.89	0.78	[620]	neutral
								COS-7			7-EFC		similar	similar	similar	[448]	
											Selegiline Dem		similar	similar	similar		
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	increased	similar	increased	similar	[621]	
T26S[1]	-27.1	n/a	n/a	Truncated N-terminal	undefined	n/a		COS-1	1.2	similar	Bupropion		0.73	0.74	1.02	[622]	neutral
								COS-7			7-EFC		similar	similar	similar	[448]	
											Selegiline Dem		similar	similar	similar		
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	similar	similar	similar	similar	[621]	
D28G[1]	-51.0	n/a	n/a	Truncate	undefined	n/a		COS-1	1.2	similar	Bupropio		0.73	0.74	1.02	[622	neutral

				d N-terminal							n					]	
								COS-7			7-EFC		similar	similar	similar		
											Selegiline Dem		similar	similar	similar	[448]	
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	similar	similar	similar	similar	[621]	
R29T[1]	-57.3	n/a	n/a	Truncated N-terminal	halting signal	n/a		COS-1	1.2	similar	Bupropion		0.73	0.74	1.02	[622]	neutral
								COS-7			7-EFC		similar	similar	similar		
											Selegiline Dem		similar	similar	similar	[448]	
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	similar	similar	similar	similar	[621]	
M46V[1]	-22.9	7.1	16.5	loop btwn N-term & helix A	SRS1'a	-1.8	neutral	COS-1	0.16	decreased (mostly P420)	Bupropion		ND	0.02	0.02	[623]	inactive
								COS-7			7-EFC		ND	ND	ND		
											Selegiline Dem		ND	ND	ND	[448]	
											Selegiline Dep		ND	ND	ND		
								COS-7			Artemether	Very low	ND	ND	ND		
G99E[1]	78.3	3.1	13.0	B-C loop	SRS1	-1.0	neutral	COS-1	0.075	decreased (mostly P420)	Bupropion		ND	ND	ND	[623]	inactive
								COS-7			7-EFC		ND	ND	ND	[448	

											Selegiline Dem		ND	ND	ND	]	
											Selegiline Dep		ND	ND	ND		
								COS-7			Artemether		ND	ND	ND	[621]	
K139E[1]	-30.2	56.2	42.1	loop btwn helix C & D	undefined	0.6	neutral	COS-1	ND	ND	Bupropion		ND	ND	ND	[623]	inactive
								COS-7			7-EFC		ND	ND	ND	[448]	
											Selegiline Dem		ND	ND	ND		
											Selegiline Dep		ND	ND	ND		
								COS-7			Artemether	ND	ND	ND	ND	[621]	
R140Q[1]	-29.6	65.8	66.1	loop btwn helix C & D	undefined	0.4	neutral	COS-1	0.825	small decrease	Bupropion		1.80	0.23	0.13	[623]	decreased activity
								COS-7			7-EFC		similar	decreased	similar	[448]	
											Selegiline Dem		similar	similar	decreased		
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	decreased	similar	decreased	similar	[621]	
P167A[1]	-24.1	1.3	2.8	start of helix E	undefined	0.3	neutral										unknown
T168I[1]	50.6	9.2	11.0	start of helix E	undefined	3.4	damaging	COS-1	0.03		Bupropion		0.67	0.16	0.03	[622]	decreased activity
								insect cells		decreased (0.5)	Bupropion				0.95		
Q172H	9.4	4.8	2.2	α-helix E	undefined	2.4	damagin	COS-7			7-EFC		increased	similar	similar	[448	neutral

[1]							g				Selegiline Dem		similar	similar	similar	]	
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	similar	similar	similar	similar	[621]	
I182M	-3.8	6.6	11.7	α-helix E	undefined	-0.3	neutral										unknown
V183I	26.7	1.6	2.7	α-helix F	undefined	-0.1	neutral										unknown
M198T[1]	-46.8	34.6	32.8	α-helix E	SRS(2, 3)	-2.7	damaging	COS-7			7-EFC		decreased	decreased	similar	[448]	substrate specific
											Selegiline Dem		ND	ND	ND		
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	decreased	similar	decreased	similar	[621]	
Y226H	-40.4	77.1	73.2	F-G loop	SRS(2, 3)	0.6	neutral										unknown
D257N	3.0	71.2	74.3	loop btwn helix G & H	undefined	0.1	neutral										unknown
S259R[1]	84.4	107.9	104.7	loop btwn helix G & H	undefined	0.0	neutral	COS-1	similar		7-EFC	similar	1.16	1.61	1.39	[620]	neutral
								COS-7			7-EFC		Small increase			[448]	
											Selegiline Dem		similar				
											Selegiline Dep		similar				
K262R[1]	4.8	66.1	59.5	loop btwn helix G	undefined	-0.4	neutral	COS-1	similar		7-EFC	increased	1.09	1.96	1.80**	[620]	substrate specific
								COS-7			7-EFC		similar	similar	similar	[448	

				& H							Selegiline Dem		similar	similar	similar	]	
											Selegiline Dep		similar	similar	similar		
								Yeast	similar		Bupropion	decreased activity					
								COS-7			Artemether	increased	similar	increased	similar	[621]	
N289K	54.5	40.3	40.6	start of helix I	SRS4	1.3	neutral										unknown
T306S[1]	-27.1	3.1	7.9	α-helix J	catalytic groove/haem contact/SRS 4	-1.7	neutral										unknown
I328T[1]	-50.6	0.3	0.0	α-helix J	undefined	-3.9	damaging	COS-1	ND	ND	Bupropion		ND	ND	ND	[622]	inactive
								Baculovirus infected insect cells		decreased (0.45)	Bupropion				0.50	[622]	
								COS-7			7-EFC		ND	ND	ND	[448]	
											Selegiline Dem		ND	ND	ND		
											Selegiline Dep		ND	ND	ND		
								COS-7			Artemether	ND	ND	ND	ND	[621]	
								HEK293	very low		Bupropion					[624]	
								Yeast	very low			similar	similar	similar			
R336C[1]	-64.9	44.9	84.5	loop	undefined	-3.0	damagin	COS-1	ND	ND	Bupropio		1.18	0.03	0.03	[622	decreased

				btwn helix J & J’			g				n					]	
								Baculovirus infected insect cells		decreased (0.45)	Bupropion				0.78		
I391N[1]	-52.6	0.0	1.2	loop btwn B-sheet 3^2& helix K’	undefined	-4.6	damaging	COS-1	0.18	decreased (mostly P420)	Bupropion		ND	ND	ND	[623]	inactive
								COS-7			7-EFC		ND	ND	ND	[448]	
											Selegiline Dem		ND	ND	ND		
											Selegiline Dep		ND	ND	ND		
								COS-7			Artemether	very low	ND	ND	ND	[621]	
H397R	20.2	25.2	43.3	loop btwn helix K’& 3_{10} helix	undefined	-1.3	neutral										unknown
T423N	-2.0	23.4	28.5	loop btwn 3_{10} helix & α-helix L	undefined	-0.4	neutral										unknown
P428T[1]	3.4	10.0	15.1	loop btwn 3_{10} helix & helix L	haem contact	-1.2	neutral	COS-1	13.2		Bupropion		ND	ND	ND	[622]	inactive
								Baculovirus infected insect cells		decreased (0.5)	Bupropion				0.50	[622]	
								COS-7			7-EFC		ND	ND	ND	[448]	
											Selegiline Dem		ND	ND	ND		

											Selegiline Dep		ND	ND	ND		
								COS-7			Artemether	ND	ND	ND	ND	[621]	
M459V[1]	-22.9	2.3	4.6	B-sheet 5[1]	undefined	-0.5	neutral	COS-7			7-EFC		similar	similar	similar	[448]	neutral
											Selegiline Dem		similar	similar	similar		
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	similar	similar	similar	similar	[621]	
G476D[1]	51.0	1.5	0.7	loop bwn B- sheet 5[1] & 5[2]	SRS6	0.5	neutral	COS-7			7-EFC		ND	ND	ND	[448]	inactive
											Selegiline Dem		ND	ND	ND		
											Selegiline Dep		ND	ND	ND		
								COS-7			Artemether	ND	ND	ND	ND	[621]	
Q485L[1]	22.9	56.4	43.6	loop bwn B-sheet 5[1] & 5[2]	undefined	-1.7	neutral	COS-7			7-EFC		similar	similar	similar	[448]	neutral
											Selegiline Dem		similar	similar	similar		
											Selegiline Dep		similar	similar	similar		
								COS-7			Artemether	similar	similar	similar	similar	[621]	
R487C[1]	-64.9	42.5	12.6	Start of B-sheet 5[2]	undefined	-3.6	damaging	COS-1	similar		7-EFC	small increase	1.04	1.68	1.62	[620]	neutral
								COS-7			7-EFC		similar	similar	similar	[448]	
											Selegiline Dem		similar	similar	similar		
											Selegiline		similar	similar	similar		

											Dep								
								COS-7			Artemether	similar	similar	similar	similar	[621]	

Substrate abbreviations

7-EFC	7-ethoxy-4-trifluoromethylcoumarin *O*-deethylaton
Selegiline Dem	selegiline *N*-demethylation
Selegiline Dep	selegiline *N*-depropagylation
Artemether	Artemether demethylation
Bupropion	Bupropion hydroxylation

1 *SNPs* found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.
2 *RSA* refers to the relative solvent accessibility as calculated by SDM.
3 *Relative holoprotein levels* refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.
4 *Relative Activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.
ND refers to protein levels, activity or kinetic parameters that were not detectable and the asterisk (*) indicates data reported to be significantly different from the wild-type enzyme. (Note that some studies provide quantitative data without indicating whether differences between wild-type and mutant proteins are statistically significant)

Table B4: CYP2C8 SNP table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY			EXPERIMENTAL DATA FROM THE LITERATURE									
	Δ Volume (Å^3)	RSA^2 (%)		2° structure location	Functional region location	SDM Score (kcal.mol^{-1})	SD^3	Effect	Expression system	Relative expression levels	Relative holoprotein levels^4	Substrate	Relative activity^5	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
		WT	Mut															
A12S	0.4	n/a	n/a	Truncated N-terminal	signal-anchor sequence	n/a	n/a	n/a										unknown
A82S	0.4	0.2	3.0	α-helix B	undefined	-3.2	0.1	damaging										unknown
R139K[1]	-4.8	39.4	77.8	loop btwn helix C & D	undefined	-1.4	0.4	neutral	Hep G2 cells	similar		Paclitaxel		0.9	0.7*	0.7*	[625]	neutral
									Yeast				similar				[626]	
R144C	-64.9	2.2	18.8	α-helix D	undefined	-11.5	0.0	damaging										unknown
G171S[1]	28.9	23.0	11.1	α-helix E	undefined	1.5	0.4	neutral	COS-1	similar		Paclitaxel	similar	1.1	1.2	1.1	[627]	neutral
V181I	26.7	0.3	3.8	α-helix E	undefined	-0.1	0.0	neutral										unknown
R186G[1]	-113.3	17.8	29.8	loop btwn helix E & F	undefined	-4.6	0.3	damaging	COS-1	decreased (0.4)		Paclitaxel	decreased (0.2)	ND	ND	ND	[627]	inactive
									Insect cells		decreased							
I223M[1]	-3.8	14.7	18.2	F-G loop	SRS(2,3)	0.2	0.1	neutral	Yeast	similar	similar	Paclitaxel		0.8	0.94	1.22	[628]	neutral
A238P[1]	24.1	68.7	36.3	α-helix G	SRS(2,3)	-2.2	0.0	damaging	Yeast	decreased	decreased	Paclitaxel		2.9*	1.3	0.46	[628]	decreased activity

I244V	-26.7	0.9	1.3	α-helix G	SRS(2,3)	-1.0	0.0	neutral										unknown
K247R[1]	4.8	16.2	20.7	α-helix G	undefined	0.5	0.6	neutral	COS-1	similar		Paclitaxel	similar	0.9	1.0	1.1	[627]	neutral
I264M[1]	-3.8	0.1	0.4	α-helix H	undefined	-0.4	0.0	neutral	Yeast			Paclitaxel	decreased				[626]	decreased activity
									E.coli		decreased (decreased thermal stability)	Paclitaxel		1.5	0.4	0.3*	[629]	
I269F[1]	23.2	20.1	40.9	α-helix H	undefined	0.2	0.1	neutral	*E.coli*			Paclitaxel		2.1	1.1	0.5*	[630]	decreased activity
									E.coli			Arachidonic Acid	0.9					
									Yeast				decreased				[626]	
L361F	23.2	16.2	12.5	Loop btwn helix K & B-sheet3[1]	haem contact/SRS 5	0.1	0.2	neutral										unknown
K383N[1]	-54.5	68.7	71.0	loop btwn B-sheet 4[2] & 3[2]	CPR contact	-1.1	0.8	neutral	COS-1	similar	same	Paclitaxel		0.9	1.1	1.2	[627]	neutral
L390S	-77.7	0.7	9.8	between B-sheet 3[2] & helix K'	undefined	-5.2	0.0	damaging										unknown
K399R[1]	4.8	90.8	96.8	loop between	undefined	-0.4	0.0	neutral	Hep G2 cells	decreased		Paclitaxel		1.4	1.6*	1.2	[625]	decreased activity

				helix K' & 3_{10} helix					Yeast			Paclitaxel	decreased				[626]	
P404A	-24.1	12.6	5.4	loop btwn helix K' & 3_{10} helix	meander	1.2	0.0	neutral	Hep G2 cells	decreased (0.46)		Paclitaxel		1.3	1.7*	1.4*	[625]	decreased activity
									E.coli	decreased (decreased thermal stability)		Paclitaxel		1.9	0.6	0.3*	[629]	
M426V	-22.9	4.7	9.8	loop btwn 3_{10} helix & helix L	undefined	-3.0	0.0	damaging										unknown

Substrate abbreviations

Paclitaxel	paclitaxel 6á-hydroxylation
Arachidonic Acid	(5Z,8Z,11Z,14Z)-icosa-5,8,11,14-tetraenoic acid (various metabolites)

Notes
1 SNPs found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.
2 *RSA* refers to the relative solvent accessibility as calculated by SDM.
3 *SD* refers to the standard deviation of the SDM score across different crystal structures of the same isoform.
4 *Relative holoprotein levels* refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.
5 *Relative activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.
The asterisk (*) indicates data reported to be significantly different from the wild-type enzyme. (Note that some studies provide quantitative data without indicating whether differences between wild-type and mutant proteins are statistically significant)

Table B5: CYP2C9 SNP table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY			EXPERIMENTAL DATA FROM THE LITERATURE									
	Δ Volume (Å^3)	RSA^2 (%)		2° structure location	Functional region location	SDM Score (kcal.mol^{-1})	SD^3	Effect	Expression system	Relative expression levels	Relative holoprotein levels^4	Substrate	Relative activity^5	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
		wt	mut															
S3Y[1]	77.7	n/a	n/a	truncated N-terminal	undefined	n/a	n/a	n/a										unknown
L17I[1]	0	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a	n/a										unknown
L19I[1]	0.0	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a	n/a	*E.coli*			Tolbutamide				similar	[556]	neutral
P30L[1]	54.0	n/a	n/a	truncated N-terminal	proline rich motif	n/a	n/a	n/a										unknown
N41D[1]	-3.0	2.2	6.3	loop btwn N-term & helix A	undefined	1.7	0.0	neutral										unknown
G70R[1]	113.3	34.7	44.2	loop btwn B-sheet	SRS1'b	-4.3	0.6	damaging										unknown

				1^1 & 1^2														
V76M[1]	22.9	0.1	1.8	B-sheet 1^2	undefined	-1.2	0.0	neutral										unknown
L90P[1]	-54.0	34.0	26.2	loop btwn helix B & C	CPR contact	-2.5	1.1	damaging (2/3)	COS-7	decrease		Lornoxicam	decrease	2.25	0.27	0.18	[631]	decrease activity
									COS-7			Tolbutamide	decrease	11.0	similar	0.01	[553]	
												Diclofenac		5.0	0.1	0.02		
									Insect	similar	decreased (0.12)	Diclofenac		3.89*	0.19*	0.05*	[545]	
												Lorsartan		2.46*	0.06*	0.03*		
												Glimepiride		7.17*	0.10*	0.01*		
R124Q	-29.6	14.1	7.7	α-helix C	haem contact/CPR contact/ SRS1	-0.4	0.7	neutral										unknown
R125H[1]	-20.2	78.6	72.5	α-helix C	CPR contact/SRS1	-0.6	0.0	neutral	*E.coli*			Tolbutamide	Decreased (reduced affinity for CPR)	5.05	0.36	0.07	[632]	decreased activity
R125L[1]	-6.7	78.6	73.5	α-helix C	CPR contact/SRS1	-0.2	0.0	neutral										unknown
T130R[1]	57.3	3.7	4.4	α-helix C	SRS1	-2.1	0.7	damaging	COS-1			Diclofenac	(reducred affinity for CPR)	similar	0.10	decreased	[544]	decreased activity
									Insect	similar	similar	Diclofenac		1.72*	0.42*	0.24*	[545]	
												Lorsartan		1.34	0.05*	0.04*		
												Glimepiride		0.89	0.08*	0.10*		

R132Q[1]	-29.6	63.2	73.2	loop btwn helix C & D	CPR contact (charge cluster)/SRS2	-0.2	0.2	neutral	Insect	similar	similar	Diclofenac	(reduced affinity for CPR)	0.53	0.01*	0.18*	[633]	decreased activity
												Lorsartan	(reduced affinity for CPR)	0.92	0.01*	0.01*	[545]	
												Glimepiride	(reduced affinity for CPR)	1.11	0.01*	0.01*		
R144C[1]	-64.9	0.8	16.8	α-helix D	undefined	-11.5	0.0	damaging	COS-7			Tolbutamide	decreased (0.3)	similar			[634]	decreased activity
												Phenytoin	decreased (0.5)					
									HepG2			(S)-warfarin			decreased	decreased	[635]	
									HepG2			Tolbutamide	similar					
												(S)-Warfarin	(reduced affinity for CPR)	similar	decrease	decrease	[636]	
												Diclofenac		similar	decrease	decrease		
												Lauric acid		similar	decrease	decrease		
									Insect			(S)-warfarin		1.0	0.7	0.7	[549]	
												Phenytoin		1.0	0.7	0.7		
									Yeast			Diclofenac		0.7	0.4	0.6	[550]	
												Losartan		0.8	0.5	0.6		
									Insect			Lornoxicam		1.1	1.2	1.1	[637]	
R144H	-20.2	0.8	8.0	α-helix	undefined	0.1	0.1	neutral										unknown

				D														
R150C	-64.9	68.3	66.5	α-helix D	undefined	-2.6	4.0	damaging (1/3)										unknown
R150H[1]	-20.2	68.3	57.1	α-helix D	undefined	-0.6	0.0	neutral	*E.coli*			Tolbutamide	(improved affinity for CPR)	0.5	0.5	2.0	[556]	increased activity
												Phenytoin	decrease				[555]	
R150L[1]	-6.7	68.3	64.6	α-helix D	undefined	-0.2	0.0	neutral				Diclofenac		similar	similar	similar	[544]	neutral
N204H[1]	39.1	13.6	32.4	α-helix F	SRS(2,3)	1.5	0.0	neutral										unknown
L208V	-26.7	51.6	47.9	α-helix F	SRS(2,3)/gating residue	-0.6	0.0	neutral										unknown
Q214L[1]	106.6	6.6	14.9	α-helix F' (F-G loop)	SRS(2,3)	3.2	0.0	damaging (2/3)	COS-1			Diclofenac		2.0	0.33	0.17	[544]	decreased activity
									Insect	similar	similar	Diclofenac		4.06*	1.11	0.27*	[545]	
												Lorsartan		1.81*	0.23*	0.13*		
												Glimepiride		1.39	0.59*	0.44*		
H251R[1]	20.2	18.3	16.6	α-helix G	undefined	-0.1	0.0	neutral	*E.coli*			Tolbutamide		similar	similar	similar	[556]	neutral
E272G[1]	-78.3	49.4	48.4	α-helix H	undefined	-4.0	0.3	damaging	*E.coli*			Tolbutamide		similar	similar	similar	[556]	neutral
P279T[1]	3.4	90.2	111.8	loop btwn helix H & I	undefined	1.1	1.3	neutral (2/3)	COS-1			Diclofenac		similar	similar	similar	[544]	neutral
T299A	116.1	3.9	7.8	centre	catalytic	1.9	0.4	neutral	*E.coli*			Tolbutami	decrease	3.73	0.31	0.08	[632	decreased

[1]				of α-helix I	groove/SRS4			(2/3)				de	d				]	activity
I327T[1]	-50.6	2.7	3.9	α-helix J	undefined	-3.9	1.3	damaging										unknown
R335Q[1]	-29.6	57.9	101.7	loop btwn helix J & J'	undefined	-0.3	0.0	neutral	Insect	similar	similar	Diclofenac		0.88	0.86	0.94	[633]	neutral
												Losartan		0.95	0.70*	0.75*	[545]	
												Glimepiride		0.89	0.64*	0.72		
R335W[1]	54.4	57.9	99.6	loop btwn helix J & J'	undefined	-2.2	0.0	damaging	*In vivo*			Losartan	similar				[638]	decreased activity
									Invivo			Phenytoin	decrease				[555]	
									E.coli		decrease	Tolbutamide		3.0	0.8	0.3	[556]	
P337L	54.0	1.4	0.2	loop btwn helix J & J'	undefined	0.9	0.0	neutral										unknown
E354K[1]	30.2	0	2.3	α-helix K	ERR triad / core stabilizing motif	-4.0	0.0	damaging	*E.coli*	ND	ND						[639]	inactive
									HEK293	ND	ND							
Y358C	-85.1	13.4	6.6	α-helix K	SRS5	-6.8	1.5	damaging										unknown
I359L[1]	0.0	1.7	2.2	end α-helix K	SRS5	0.2	0.0	neutral	Yeast			Tolbutamide		increase	large decrease	large decrease	[547]	decreased activity
												(S)-Warfarin		increase	large decrease	large decrease		
									Yeast	similar		(S)-Warfarin		3.7*	0.4*	0.1*	[548]	

												Diclofenac		3.2*	0.9	0.3*		
												Piroxicam		1.5	0.05*	0.02*		
												Phenytoin		2.0*	0.08*	0.04*		
												Tenoxicam		3.2*	0.13*	0.04*		
												Mefenamic acid		4.9*	0.3*	0.06*		
												Tolbutamide		11.5*	1.1	0.1*		
									Insect			(S)-Warfarin		5.0	0.2	0.04	[549]	
												Phenytoin		4.0	0.2	0.06		
									Yeast			Diclofenac		3.2	1.0	0.3	[550]	
												Losartan		1.4	0.12	0.1		
									Yeast			Diclofenac		8.3	1.4	0.2	[551]	
									Insect			(S)-Warfarin		5.0	0.14	0.04	[552]	
												Diclofenac		3.5	0.4	0.1		
												Lauric Acid		3.7	0.2	0.05		
									COS-7	decrease		Lornoxicam		1.3	0.34*	0.28*	[631]	
									Insect	similar	similar	Diclofenac		2.94*	1.11	0.37*	[545]	
												Losartan		0.88	0.20*	0.23*		
												Glimepiride		0.72	0.13*	0.20*		
I359T[1]	-50.6	1.7	4.7	end α-helix K	SRS-5	-3.9	0.0	damaging	Yeast			Diclofenac		1.9	0.4	0.2	[551]	Decreased activity
D360E	27.3	8.0	13.5	loop	SRS-5	-1.3	0.3	neutral	Insect			(S)-		12.0	1.0	0.08	[552	Decreased

[1]				btwn helix K & β-sheet 3[1]								Warfarin					]	activity
												Diclofenac		4.6	0.9	0.2		
												Lauric acid		2.8	0.3	0.1		
									In vivo			Losartan	decrease				[638]	
									In vivo			Phenytoin	decrease				[555]	
D360N[1]	3.0	8.0	12.3	loop btwn helix K & β-sheet 3[1]	SRS-5	1.4	0.1	neutral										unknown
P382S[1]	-23.7	52.4	70.8	loop btwn β-sheet 4[2] & 3[2]	SRS5	-0.9	1.0	neutral	*E.coli*			Tolbutamide	similar	1.36	1.16	0.85	[632]	neutral
D397A[1]	-22.5	30.3	24.7	loop btwn helix K' & 3_{10} helix	undefined	1.2	0.0	neutral	*E.coli*	ND	ND	Tolbutamide		ND	ND	ND	[632]	inactive
L413P[1]	-54.0	2.6	7.2	Loop bwtn 3_{10} helix & helix L	meander	-4.1	0.0	damaging										unknown
L447F	23.2	0.9	1.2	α-helix L	undefined	-0.8	0.0	neutral										unknown
L447T	-50.6	0.9	4.1	α-helix L	undefined	-3.5	3.3	damaging										unknown

Q454H[1]	9.4	8.4	7.06	α-helix L	undefined	1.4	0.3	neutral	*E.coli*			Tolbutamide	similar	1.14	1.21	1.06	[632]	neutral
N474S[1]		31.4	20.3	loop btwn β-sheet 5[1] & 5[2]	SRS6	-3.5	3.3	damaging (2/3)										unknown
A477T[1]	27.5	41.0	29.0	loop btwn β-sheet 5[1] & 5[2]	SRS-6	0.6	0.5	neutral	COS-1			Diclofenac		2.0	0.33	0.17	[544]	decreased activity
									Insect	similar	similar	Diclofenac		4.28*	0.84	0.19*	[545]	
												Losartan		0.69	0.01*	0.01*		
												Glimepiride		0.78	0.17*	0.25*		
P489S[1]	-23.7	46.4	73.3	end of β-sheet 5[2]	undefined	1.7	0.5	neutral			decrease	Tolbutamide		0.9	0.7	0.75	[556]	neutral
V490F[1]	49.9	17.1	37.26	C-terminal	undefined	-0.1	0.8	neutral										unknown

Substrate abbreviations

(S)-warfarin	(S)-warfarin 7-hydroxylation
Diclofenac	diclofenac 4'-hydroxylation
Lauric acid	lauric acid 11-hydroxylation
Phenytoin	Phenytoin 4'-hydroxylation
Losartan	Losartan oxidation
Lornoxicam	lornoxicam 5-hydroxylation
Piroxicam	Piroxicam 5'-hydroxylation
Tenoxicam	Tenoxicam 5'-hydroxylation
Mefenamic acid	Mefenamic acid 39-hydroxylation

Tolbutamide	Tolbutamide hydroxylation
Glimepiride	glimepiride hydroxylation

Notes

1 SNPs found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.

2 *RSA* refers to the relative solvent accessibility as calculated by SDM.

3 *SD* refers to the standard deviation of the SDM score across different crystal structures of the same isoform.

4 *Relative holoprotein levels* refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.

5 *Relative activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.

ND refers to protein levels, activity or kinetic parameters that were not detectable and the asterisk (*) indicates data reported to be significantly different from the wild-type enzyme. (Note that some studies provide quantitative data without indicating whether differences between wild-type and mutant proteins are statistically significant)

Table B6: CYP2C19 SNP table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY		EXPERIMENTAL DATA FROM THE LITERATURE									
	Δ Volume (Å^3)	RSA^2 (%)		2° structure location	Functional region location	SDM Score (kcal.mol^{-1})	Effect	Expression system	Relative expression levels	Relative holoprotein levels ^{3}	Substrate	Relative activity ^{4}	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
		wt	mut														
M1V	22.9	n/a	n/a	truncated N-terminal	undefined	n/a	n/a										unknown
L17P[1]	-54.0	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a	Yeast	similar	similar	S-Mephenytoin		0.99	1.17	1.2	[316]	neutral
											Omeprazole		0.79	1.01	1.28		
											CEC		1.00	1.41	1.48	[315]	
I19L[1]	0.0	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a	Yeast	similar	similar?	S-Mephenytoin		1.25	1.41*	1.15	[316]	neutral
											Omeprazole		1.25	1.03	0.81		
											CEC		1.31	1.10	0.82	[315]	
S51G[1]	-28.9	16.8	24.1	α-helix A	SRS1'a	-2.0	neutral	Yeast			S-Mephenytoin		3.00*	similar	Decrease (0.29-0.47)	[640]	decreased activity
								Yeast			Omeprazole		1.47*	1.80*	1.24	[641]	
								Yeast	similar	similar	S-Mephenytoin		3.00*	1.78*	0.60	[316]	
											Omeprazole		2.70*	1.20	0.44*		
											CEC		1.57	0.46	0.30*	[315	

																]	
M74T[1]	-46.8	15.6	2.9	β-sheet 1[2]	SRS1'b	-2.7	damaging	Yeast	similar	similar	S-Mephenytoin		0.71	0.70	0.95	[316]	neutral
											Omeprazole		0.67	0.51	0.75		
											CEC		1.07	1.24	1.16	[315]	
G91R[1]	113.3	23.2	54.2	loop btwn helix B & B'	CPR contact	2.2	damaging										unknown
E92D[1]	-27.3	70.8	95.5	loop btwn helix B & B'	SRS1/ CPR contact	0.0	neutral	Yeast	similar	similar	S-Mephenytoin		0.60*	0.71	1.26	[316]	substrate specific
											Omeprazole		1.61*	0.87	0.53		
											CEC		1.27	0.76	0.56	[315]	
V113I[1]	26.7	52.3	45.0	loop btwn helix B ' & C	SRS1/CPR contact	0.4	neutral										unknown
W120R[1]	-54.4	4.1	4.5	α-helix C	haem contact/SRS 1	-2.7	damaging	*E.coli*			S-Mephenytoin	decreased (0.09)				[642]	decreased activity
											Tolbutamide	decreased (0.13)					
								Yeast	similar	similar	S-Mephenytoin		N.D.	N.D.	N.D.	[316]	
											Omeprazole		N.D.	N.D.	N.D.		
											CEC		0.88	0.10*	0.11*	[315]	
E122A	-49.8	40.2	31.9	α-helix C	SRS1	1.4	neutral	Yeast	similar	similar	S-Mephenytoin		1.08	1.09	0.93	[316]	substrate specific
											Omeprazole		2.28*	0.61*	0.27*		

											CEC		1.10	0.29*	0.26*	[315]	
R132Q[1]	-29.6	65.6	70.8	loop btwn helix C & D	CPR contact (charge cluster)/SRS 1	0.4	neutral	*E.coli*			S-Mephenytoin		N.D.	N.D.	ND	[460]	inactive
											Tolbutimide		N.D.	N.D.	ND		
								Yeast	similar	similar	S-Mephenytoin		N.D.	N.D.	N.D.	[316]	
											Omeprazole		N.D.	N.D.	N.D.		
R144C	-64.9	2.7	23.6	α-helix D	undefined	-4.6	damaging										unknown
R144H[1]	-20.2	2.7	12.8	α-helix D	undefined	0.2	neutral	*E.coli*			S-Mephenytoin		1.25*	0.73	0.58	[461]	decrease activity
								Yeast	similar	similar	S-Mephenytoin		1.90*	0.77	0.39*	[316]	
											Omeprazole		1.92*	0.78	0.44*		
											CEC		1.01	0.39	0.37*	[315]	
R150H[1]	-20.2	61.7	79.4	α-helix D	undefined	-0.6	neutral	*E.coli*			S-Mephenytoin		similar	similar	similar	[461]	neutral
								Yeast	similar	similar	S-Mephenytoin		1.01	1.13	1.12	[316]	
											Omeprazole		0.91	1.05	1.13		
											CEC		1.38	1.20	0.87	[315]	
A161P[1]	24.1	36.7	34.4	loop btwn helix D & E	undefined	-10.5	damaging	Yeast	similar	similar	S-Mephenytoin		1.57*	0.19*	0.12*	[316]	decreased activity
											Omeprazole		1.23	0.20*	0.16*		
											CEC		0.94	0.34	0.39*	[315]	
F168L[1]	-23.2	63.0	97.1	α-helix E	undefined	1.0	neutral	Yeast	similar	similar	S-Mephenytoin		0.67	0.81	1.15	[316]	neutral

											Omeprazole		1.21	0.76	0.63		
											CEC		1.27	0.98	0.76	[315]	
A173V	51.4	0.0	0.0	α-helix E	undefined	-0.2	neutral										unknown
N176S	-25.1	0.0	1.1	α-helix E	undefined	0.4	neutral										unknown
R186P[1]	-60.7	22.7	12.0	loop btwn α-helix E & F	undefined	1.4	neutral										unknown
W212C	-119.3	31.9	83.4	α-helix F'	SRS(2,3)	-2.1	damaging	Yeast	similar?	similar?	S-Mephenytoin		2.57*	0.34*	0.13*	[316]	decreased activity
											Omeprazole		2.07*	0.26*	0.12*		
P227L[1]	54.0	48.7	51.8	F-G loop	SRS(2,3)	0.6	neutral	*E.coli*			S-Mephenytoin		increased	decreased	decrease	[461]	decreased activity
								E.coli		similar	S-Mephenytoin	decreased	ND	ND	ND	[643]	
											Omeprazole	decreased	2.94*	0.62*	0.18*		
								Yeast	similar	similar	S-Mephenytoin		2.10*	0.34*	0.16*	[316]	
											Omeprazole		1.95*	0.69	0.34*		
											CEC		1.30	0.27*	0.23*	[315]	
D256N[1]	3.0	64.1	57.2	loop btwn helix G & H	undefined	-0.3	neutral	*E.coli*	similar?	similar	S-Mephenytoin		1.95	0.50*	0.24*	[643]	decreased activity
											Omeprazole		0.58	0.38*	0.64*		
Q279P	-31.1	91.4	73.3	loop btwn	undefined	0.1	neutral										unknown

				helix H & I													
S280Y	138.8	15.5	42.1	loop btwn helix H & I	undefined	-1.4	neutral										unknown
T302R[1]	57.3	5.4	7.4	α- helix I	catalytic groove/haem contact/CPR contact/SRS 4	-1.7	neutral										unknown
R329C	-64.9	69.4	78.0	end of α-helix J	undefined	-7.2	damaging										unknown
R329H[1]	-20.2	69.4	89.3	α-helix J	undefined	-0.6	neutral	Yeast			S-Mephenytoin		similar	similar	similar	[640]	neutral
								Yeast			Omeprazole		1.14	0.75	0.67	[641]	
								Yeast	similar	similar	S-Mephenytoin		2.40*	1.29	0.53	[644]	
											Omeprazole		1.11	0.97	0.88		
											CEC		1.51	1.34	0.88	[315]	
I331V[1] (WT 1.B)	-26.7	2.2	0.0	loop btwn helix J & J'	undefined	-0.4	neutral	Yeast	similar	similar	S-Mephenytoin	similar				[316]	neutral
											Omeprazole	similar					
R335Q[1]	-29.6	48.5	88.6	loop btwn helix J & J'	undefined	-0.3	neutral										unknown
H344Y	40.4	70.	72.	α-helix	CPR contact	0.1	neutral										unknown

		3	7	J'	(ionic charge cluster residue)												
D360N[1]	3.0	9.4	10.2	loop btwn helix K & B-sheet 3[1]	SRS5	1.5	neutral	Yeast	similar	decreased (0.6)	S-Mephenytoin		1.57*	0.51*	0.33*	[316]	decreased activity
											Omeprazole		1.44*	0.52*	0.36*		
V374I[1]	26.7	8.6	3.7	B-sheet 4[1]	undefined	0.8	neutral										unknown
V394M[1]	22.9	0.0	0.4	α-helix K'	undefined	-0.1	neutral	*E.coli*	similar?	similar	S-Mephenytoin		1.09	1.05	0.95	[643]	neutral
											Omeprazole		0.98	0.91	0.91		
E405K[1]	30.2	83.4	97.6	loop btwn helix k' & L	meander	0.4	neutral	*E.coli*	similar?	similar	S-Mephenytoin		1.64	0.81	0.48	[643]	unknown
											Omeprazole		0.66*	0.67	0.98		
R410C[1]	-64.9	85.5	88.9	3 $_{10}$ helix	meander	-0.3	neutral	*E.coli*			S-Mephenytoin		similar	similar	similar	[461]	neutral
								Yeast	similar	similar	S-Mephenytoin		1.12	1.06	0.96	[316]	
											Omeprazole		0.90	0.98	1.04		
											CEC		1.04	1.51	1.45	[315]	
R433W[1]	54.4	2.2	16.7	loop btwn 3 $_{10}$ helix & helix L	Cys-pocket/haem contact/CPR contact (ionic charge cluster	0.4	neutral	*E.coli*			S-mephenytoin	ND				[645]	inactive
											tolbutamide	ND					
								Yeast	similar	ND	S-Mephenytoin		N.D.	N.D.	N.D.	[316]	
											Omeprazole		N.D.	N.D.	N.D.		

					residue)												
R442C[1]	-64.9	14.1	39.6	α-helix L	CPR contact (ionic charge cluster residue)	-10.0	damaging	Yeast	similar	decrease (0.6)	S-Mephenytoin		1.21	0.13*	0.10*	[316]	decreased activity
											Omeprazole		0.77	0.21*	0.27*		
											CEC		1.36	0.51	0.37*	[315]	
								In vivo			Mephobarbital	decreased				[646]	
F448L[1]	-36.7	0.0	0.2	α-helix L	undefined	-0.2	neutral										unknown
P464T	3.4	37.1	35.5	loop btwn β-sheet 5[1] & 5[2]	SRS6	-0.6	neutral										unknown

Substrate abbreviations

CEC	3-Cyano-7-ethoxycoumarin
S-Mephenytoin	*S*-mephenytoin 4'-hydroxylation
Omeprazole	Omeprazole 5'-hydroxylation
Tolbutimide	Tolbutamide hydroxylation
Mephobarbital	Mephobarbital 4′-hydroxylation

Notes

1 SNPs found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.

2 *RSA* refers to the relative solvent accessibility as calculated by SDM.

3 *Relative holoprotein levels* refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.

4 *Relative activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.

ND refers to protein levels, activity or kinetic parameters that were not detectable and the asterisk (*) indicates data reported to be significantly different from the wild-type enzyme. (Note that some studies provide quantitative data without indicating whether differences between wild-type and mutant proteins are statistically significant)

Table B7: CYP2D6 SNP table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY		EXPERIMENTAL DATA FROM THE LITERATURE									
	Δ Volume ($Å^3$)	RSA[2] (%)		2° structure location	Functional region location	SDM Score (kcal.mol^{-1})	Effect	Expression system	Relative expression levels	Relative holoprotein levels [3]	Substrate	Relative activity [4]	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
		wt	mut														
V7M[1]	22.9	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a										unknown
V11M[1]	22.9	n/a	n/a	truncated N-terminal	signal-anchor sequence	n/a	n/a										unknown
R25W[1]	54.4	n/a	n/a	truncated N-terminal	halting signal	n/a	n/a										unknown
R26H	-20.2	n/a	n/a	truncated N-terminal	halting signal	n/a	n/a										unknown
R26C[1]	-64.9	n/a	n/a	truncated N-terminal	halting signal	n/a	n/a										unknown
P34S[1]	-23.7	28.5	16.0	btwn N-term & α-helix A	proline rich motif	4.0	damaging										unknown
G42R[1]	113.3			missing region	proline rich motif	n/a	n/a										unknown
R62W[1]	54.4	22.4	55.8	α-helix A	undefined	1.6	neutral										unknown

A85V[1]	51.4	63.0	68.4	α-helix B	undefined	-1.6	neutral												unknown
A90V[1]	51.4	0.8	0.8	α-helix B	undefined	-0.2	neutral	COS-7	similar		Bufuralol	similar	0.91	1.62	1.84	[647]	neutral
											DXM	similar					
L91M[1]	-3.8	1.5	3.1	α-helix B	CPR contact	-0.8	neutral										unknown
H94R[1]	20.2	27.3	47.7	loop btwn α-helix B & B'	CPR contact	-1.2	neutral										unknown
V104A[1]	-51.4	3.2	5.6	B-C loop	SRS1	-2.3	damaging										unknown
V104M[1]	22.9	3.2	2.5	B-C loop	SRS1	-0.6	neutral										unknown
T107I[1]	50.6	27.9	10.4	α-helix B'	SRS1	4.2	damaging	COS-1	slightly increased		Bufuralol	similar				[648]	substrate specific
								Yeast	similar	similar	Bufuralol		similar	similar	similar		
											codeine		increased	similar	decreased		
T107Y[1]	77.5	27.9	12.0	α-helix B'	SRS1	2.9	damaging										unknown
I109V[1]	-26.7	7.6	3.1	α-helix B'	SRS1	-0.4	neutral										unknown
V119M[1]	22.9	10.1	5.4	B-C loop	SRS1/haem contact	1.0	neutral										unknown
F120I[1]	-23.2	42.0	24.1	B-C loop	SRS1/haem contact/gating residue	-0.4	neutral										unknown
A122S[1]	0.4	0.1	0.0	B-C loop	SRS1	-2.3	damaging										unknown
R123L[1]	-6.7	44.3	46.4	B-C	SRS1	0.1	neutral										unknown

				loop													
R129G	-113.3	38.1	45.7	α-helix C	SRS1/CPR contact/gating residue	-2.1	damaging										unknown
V136I[1]	26.7	58.6	38.9	α-helix C	CPR contact	0.6	neutral										unknown
Q151E[1]	-5.4	74.1	61.6	α-helix D	undefined	0.4	neutral										unknown
W152G	-167.7	9.5	42.3	α-helix D	undefined	-3.9	damaging										unknown
W152R	-54.4	9.5	21.3	α-helix D	undefined	-2.1	damaging										unknown
E155K[1]	30.2	49.6	64.4	α-helix D	undefined	-1.0	neutral										unknown
E156A[1]	-49.8	1.9	4.7	α-helix D	undefined	0.42	neutral	COS-7	decreased		Bufuralol	decreased	0.34*	0.20	0.63	[647]	decreased activity
											DXM	decreased					
N166D[1]	-3	89.5	86.0	end of α-helix D	undefined	0.0	neutral										unknown
H167Q	-9.4	46.9	23.2	loop btwn α-helix D& E	undefined	0.2	neutral										unknown
S168A	-0.4	96.0	94.9	loop btwn α-helix D& E	undefined	-0.3	neutral										unknown
G169E	78.3	13.1	47.2	loop btwn α-helix	undefined	-5.9	damaging										unknown

				D& E															
G169R[1]	113.3	13.1	57.6	loop btwn α-helix D& E	undefined	-3.2	damaging												unknown
R173C[1]	-64.9	52.1	50.7	loop btwn α-helix D& E	undefined	-4.4	damaging												unknown
R201H[1]	-20.2	34.7	61.6	start of α- helix F	undefined	-1.1	neutral												unknown
G212E	78.3	0.0	0.0	α-helix F	SRS(2,3)	0.6	neutral												unknown
L213P	-54	32.5	19.1	α-helix F	SRS(2,3)	-3.7	damaging												unknown
L231P*[1]	-60.7/-54	35.4	70.2	α-helix F	SRS(2,3)/gating residue	-7.8	damaging												unknown
A237S[1]	0.4	27.7	18.7	start of α- helix G	SRS(2,3)	-2.8	damaging												unknown
E278K[1]	30.2	32.0	19.7	α-helix H	undefined	-0.9	neutral												unknown
M279K	5.7	18.6	30.3	α-helix H	undefined	-0.3	neutral												unknown
N285S[1]	-25.1	65.3	44.6	loop btwn α-helix H & I	undefined	0.7	neutral												unknown
L295V	-26.7	0.9	5.3	α-helix I	SRS1	-1.4	neutral												unknown
R296C[1]	-64.9	27.7	9.5	α-helix I	SRS1	-8.8	damagin	Yeast		slightly	Bufuralol		similar	similar	similar	[648	neutral		

							g			decreased							]	
I297L[1]	0	4.5	9.7	α-helix I	SRS1	0.2	neutral	Insect			DXM		1.13	6.55*	5.76*	[649	substrate specific
											Codeine		2.87*	1.57	0.57	]	
A300G	-28.5	5.5	4.6	α-helix I	SRS1	-4.2	damaging										unknown
D301N	3	20.9	23.7	α-helix I	SRS1	0.6	neutral										unknown
S311L	77.7	5.5	2.2	α-helix I	SRS1	2.7	damaging										unknown
H324P[1]	-40.5	35.8	23.1	loop btwn helix H & I	undefined	-4.6	damaging	*In vivo*			Sparteine	ND				[650]	inactive
								Insect	similar	ND	Bufuralol	ND				[651]	
R329L	-6.7	74.1	74.6	α-helix J	undefined	-0.2	neutral										unknown
R330P	-60.7	63.4	25.5	α-helix J	undefined	-1.8	neutral										unknown
E334A[1]	-49.8	19.6	28.2	α-helix J	undefined	-0.5	neutral										unknown
V338M[1]	22.9	24.9	36.2	End of α-helix J	undefined	1.0	neutral										unknown
R343G[1]	-113.3	46.5	93.4	loop btwn helix I & K	undefined	-2.4	damaging										unknown
Y355C	-85.1	10.8	37.2	α-helix K	undefined	-8.8	damaging										unknown
R365H	-20.2	0.0	3.3	α-helix K	ERR triad/core stabilising motif	0.2	neutral										unknown

I369T[1]	-50.6	12.8	5.6	loop btwn helix K & β-sheet 3[1]	Haem contact/SRS5	-3.2	damaging	Insect	similar	slightly decreased	DXM		6.55*	4.08*	0.62	[649]	substrate specific
											Codeine		2.05	3.00*	1.46		
V370I	26.7	29.7	27.5	loop btwn helix K & β-sheet 3[1]	Haem contact/SRS5	0.4	neutral										unknown
G373S	28.9	16.6	7.1	loop btwn helix K & β-sheet 3[1]	Haem contact/SRS5	-3.3	damaging										unknown
R380H	-20.2	66.0	60.5	loop btwn B-sheet 3[1] & 4[1]	CPR contact (ionic charge cluster tesidue)	0.0	neutral										unknown
E383K[1]	30.2	62.0	76.5	B-sheet 4[1]	undefined	-1.2	neutral										unknown
K404Q[1]	-24.8	18.8	36.0	loop btwn α-helix K'& 3_{10} helix	undefined	-0.2	neutral										unknown
E410K[1]	30.2	90.1	81.9	loop btwn α-helix K' & 3_{10} helix	undefined	0.4	neutral	COS-7	slightly decreased		Bufuralol	similar	1.00	1.34	1.34	[647]	substrate specific
											DXM	similar					
								Insect	slightly decreased	slightly decreased	DXM		2.42*	0.86	0.35*	[649]	
											Codeine	similar	0.84	1.20	1.43		
E418K[1]	30.2	56.8	70.9	3_{10} helix	meander	-1.1	neutral										unknown

E418Q[1]	5.4	56.8	46.1	3_{10} helix	meander	-0.6	neutral												unknown
L421P[1]	-54	11.0	19.0	loop btwn 3_{10} helix & helix L	meander	-4.4	damaging												unknown
P430L	54	31.6	45.9	loop btwn 3_{10} helix & helix L	undefined	-0.9	neutral												unknown
R440H[1]	-20.2	52.2	54.2	loop btwn 3_{10} helix & helix L	Cys-pocket/CPR contact	1.5	neutral												unknown
R441C[1]	-64.9	1.4	1.8	loop btwn 3_{10} helix & helix L	Cys-pocket/haem contact/CPR contact (charge cluster residue)	-10.4	damaging	COS-1		ND	Propafenone	ND				[652]	inactive		
R441H[1]	-20.2	1.4	3.3	loop btwn 3_{10} helix & helix L	cys-pocket/Haem contact/CPR contact	-1.0	neutral												unknown
G445E[1]	78.3	61.3	17.3	start of α-helix L	Cys-pocket/haem contact/CPR contact	3.9	damaging												unknown
A449D[1]	22.5	19.3	12.5	α-helix L	haem contact	-0.5	neutral												unknown
P469A[1]	-24.1	6.4	10.5	loop btwn B-sheet 5[1]	undefined	-0.1	neutral												unknown

				& 5[2]														
T470A[1]	-27.5	111.0	108.6	loop btwn B-sheet 5[1] & 5[2]	undefined	-0.3	neutral											unknown
H478S[1]	-64.2	88.2	95.2	loop btwn B-sheet 5[1] & 5[2]	SRS6	0.2	neutral											unknown
H478Y[1]	40.4	88.2	99.2	loop btwn B-sheet 5[1] & 5[2]	SRS6	1.5	neutral											unknown
G479A[1]	28.5	2.6	4.2	loop btwn B-sheet 5[1] & 5[2]	SRS6	1.3	neutral											unknown
G479R	113.3	2.6	6.0	loop btwn B-sheet 5[1] & 5[2]	SRS6	-0.2	neutral											unknown
F481V[1]	-49.9	21.1	35.8	loop btwn B-sheet 5[1] & 5[2]	SRS6	0.5	neutral											unknown
A482S[1]	0.4	70.6	64.8	loop btwn B-sheet 5[1] & 5[2]	SRS6	-0.8	neutral											unknown
S486T[1]	27.1	51.4	48.1	B-sheet 5[2]	SRS6	0.7	neutral	Yeast	similar	similar	Bufuralol		similar	similar	similar	[648]		neutral

								COS-7	decreased (<0.5)		Bufuralol	similar	1.26	1.38	1.07	[647]	
											DXM	similar					

Substrate abbreviations

Sparteine	Sparteine 2-oxidase/Sparteine 5-oxidase
Bufuralol	Bufuralol 1'-hydroxylation
DXM	Dextromethorphan O-demethylation
Codeine	Codeine O-demethylation
Propafenone	Propafenone hydroxylation

Notes

1 SNPs found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.
2 *RSA* refers to the relative solvent accessibility as calculated by SDM.
3 *Relative holoprotein levels* refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.
4 *Relative activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.
ND refers to protein levels, activity or kinetic parameters that were not detectable and the asterisk (*) indicates data reported to be significantly different from the wild-type enzyme. .
(Note that some studies provide quantitative data without indicating whether differences between wild-type and mutant proteins are statistically significant)

Table B8: CYP2E1 SNP table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY			EXPERIMENTAL DATA FROM THE LITERATURE									
	Δ Volume (Å^3)	RSA[2] (%)		2° structure location	Functional region location	SDM Score (kcal.mol^{-1})	SD[3]	Effect	Expression system	Relative expression levels	Relative holoprotein levels [4]	Substrate	Relative activity [5]	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
		wt	mut															
V72L	26.7	7.5	2.7	B-sheet 1[1]	SRS1'b	-0.4	0.3	neutral										unknown
R76H[1]	-20.2	32.8	23.1	B-sheet 1[2]	SRS1'b	0.8	1.7	neutral	COS-1	decreased (0.37)		Chloroxazone	0.36				[653]	decreased activity
									COS-1	decrease (0.29)		Chloroxazone		similar	increased?	increased?	[654]	
												Nitrophenol		similar	increased?	increased?		
R126Q	-29.6	13.0	13.0	α-helix C	haem contact/ CPR contact/SRS 1	-0.1	0.6	neutral										unknown
R126W	54.4	13.0	12.4	α-helix C	haem contact/ CPR contact/SRS 1	2.0	0.2	damaging										unknown
G173S	28.9	4.9	6.7	α-helix E	undefined	0.6	0.5	neutral										unknown
A175T	27.5	0.2	0.4	α-helix E	undefined	-2.8	0.3	damaging										unknown
V179I[1]	26.7	1.4	0.7	α-helix	undefined	-0.1	0.0	neutral	lymphobl	similar		Chloroxazo	similar				[655	neutral

				E					a-stoid			ne					]	
									COS-1			Chloroxazone		similar	similar	similar	[654]	
												Nitrophenol		similar	similar	similar		
N219D	-3.0	29.5	16.9	α-helix F'	SRS(2,3)	0.4	1.2	neutral										unknown
G288S	-54.8	17.2	12.5	α-helix I	SRS4	1.5	0.6	neutral										unknown
I341M	-3.8	42.0	16.0	3_{10} helix J'	undefined	0.2	0.1	neutral										unknown
S366C	19.5	24.8	16.5	btwn α-helix K & B-sheet 3^1	SRS5	1.9	0.0	neutral										unknown
T373A	-27.5	50.8	56.0	loop btwn B-sheet 3^1 & 4^1	CPR contact	0.2	0.0	neutral										unknown
V389I[1]	26.7	0.2	0.9	B-sheet 3^2	undefined	-0.3	0.0	neutral	COS-1	similar		Chloroxazone	similar				[653]	neutral
									COS-1			Chloroxazone		similar	similar	similar	[654]	
												Nitrophenol		similar	similar	similar		
D394G	-51.0	22.0	32.6	α-helix K'	haem contact	-2.2	0.9	damaging										unknown
V396L	26.7	2.4	0.0	α-helix K'	undefined	0.2	0.0	neutral										unknown
N418S	-25.1	102.2	71.6	loop btwn 3_{10} helix &	undefined	0.4	0.1	neutral										unknown

				L															
H457L[1]	13.5	17.1	19.3	end α-helix L	undefined	1.6	0.8	neutral											unknown
R484H	-20.2	86.8	73.4	loop btwn B-sheet 5^1 & 5^2	SRS6	-1.3	0.3	neutral											unknown

Substrate abbreviations

Chlorzoxazone	Chlorzoxazone 6-hydroxylation
Nitrophenol	Nitrophenol 2-hydroxylation

Notes

1 SNPs found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.

2 **RSA** refers to the relative solvent accessibility as calculated by SDM.

3 *SD* refers to the standard deviation of the SDM score across different crystal structures of the same isoform.

4 *Relative holoprotein levels* refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.

5 *Relative activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.

Table B9: CYP3A4 SNP table

SNP	STRUCTURAL INFORMATION					PROTEIN STABILITY			EXPERIMENTAL DATA FROM THE LITERATURE										
	Δ Volume (Å^3)	RSA[2] (%)		2° structure location	Functional region location	SDM Score (kcal.mol^{-1})	SD[3]	Effect	Expression system	Relative expression levels	Relative holoprotein levels [4]	Substrate	Relative activity[5]	Km_{mut}/Km_{wt}	$Vmax_{mut}/Vmax_{wt}$	Vmax/Km (mut/wt)	Ref	Classification
		wt	mut															
L15P[1]	-54.0	n/a		truncated N-terminal	signal-anchor sequence	n/a												unknown
H30D	-42.1	13.5	28.5	btwn N-term & α-helix A	halting signal	-3.1	2.5	damaging (5/7)										unknown
G56D[1]	51.0	13.4	3.5	btwn N-term & α-helix A	SRS1'a	-2.3	1.6	damaging (4/7)	*E.coli*	similar		Testosterone	similar				[540]	decreased activity
									E.coli	similar		Testosterone		2.10	0.77	0.37	[541]	
												Midazolam		1.59	0.60	0.38		
												Nifedipine		0.91	0.41	0.45		
Y68C	-85.1	22.6	22.8	α-helix A	undefined	-5.5	1.4	damaging										unknown
K96E	-30.2	60.7	62.9	B-C loop	CPR contact	-0.3	0.4	neutral										unknown
I118V[1]	-26.7	6.8	6.9	B-C loop	haem contact/SRS1	-1.1	0.4	neutral	*In vivo*			Cortisol	decrease ?				[656]	unknown
R130Q[1]	-29.6	18.6	19.8	α-helix D	haem contact/CPR contact/SRS1	-1.4	0.2	neutral	*E.coli*	decrease	ND						[540]	inactive
R162Q[1]	-26.9	60.6	72.6	α-helix D	undefined	-0.5	0.3	neutral										unknown

R162W	54.4	60.4	70.7	α-helix D	undefined	0.9	0.3	neutral											unknown
V170I[1]	26.7	3.4	2.9	loop btwn helix D & E	undefined	-0.3	0.0	neutral	*E.coli*	similar		Testosterone	similar				[540]	neutral
D174H[1]	3.0	78.7	79.9	α-helix E	undefined	0.2	0.2	neutal	*E.coli*	similar		Testosterone	similar				[540]	neutral
F176V[1]	-49.9	0.7	4.6	α-helix E	undefined	-1.4	0.0	neutral										unknown
T185S[1]	-27.1	1.3	12.6	α-helix E	undefined	-1.4	0.5	neutral	HepG2	decreased		Testosterone		0.73	0.38	0.52	[542]	decreased activity
									E.coli	similar		Testosterone		3.75	0.33	0.09	[541]	
												Midazolam		3.61	0.33	0.09		
												Nifedipine		5.94	0.40	0.07		
T187S	-27.1	3.4	2.6	end of α-helix E	undefined	-1.6	0.3	neutral										unknown
F189S[1]	-100.9	1.7	17.9	loop btwn helix E & F	undefined	-5.2	1.1	damaging	*E.coli*	decreased		Testosterone	decrease (0.27*)				[539]	decreased activity
												Chlorpyrifos	Decrease (<0.1*)					
I193V	-26.7	15.4	17.6	loop btwn helix E & F	undefined	-0.5	0.4	neutral										unknown
P218R[1]	60.7	85.8	98.0	α-helix F'	SRS(2,3)	1.3	0.0	neutral	*In vivo*			Cortisol	decrease ?				[656]	unknown
S222P[1]	23.7	52.	41.	α-helix	SRS(2,3)	-0.9	0.7	neutral	Insect			Nifedipine		4.68*	0.64*	0.14*	[538	decreased

		1	3	F'								Testosterone		1.46	0.84	0.58	]	activity
									E.coli	similar		Testosterone		1.32	0.40	0.30	[541]	
												Midazolam		1.90	0.33	0.17		
												Nifedipine		3.43	0.36	0.10		
I223R[1]	6.7	0.0	0.1	α-helix F'	SRS(2,3)/gating residue	-2.9	0.8	damaging										unknown
S252A	-0.4	16.0	16.4	α-helix G	SRS(2,3)	2.7	0.4	damaging										unknown
L293P[1]	-54.0	54.1	62.8	start of helix I	start of SRS4	-0.5	0.1	neutral	HepG2	similar		Testosterone		0.91	1.50	1.64	[542]	increased activity?
									E.coli			Testosterone	increase (1.76*)				[539]	
												Chlorpyrifos	increase (>2.0*)					
												Testosterone		0.85	0.75	0.88	[541]	
												Midazolam		0.95	0.80	0.84		
												Nifedipine		0.84	0.67	0.80		
A305S	0.4	68.8	64.2	α-helix I	catalytic groove/haem contact/SRS4	-2.2	0.0	damaging										unknown
M318I	3.8	0.0	0.1	α-helix I	undefined	-0.5	0.0	neutral										unknown
Y319C[1]	-85.1	9.7	8.9	α-helix I	undefined	-7.0	1.9	damaging										unknown
T349N	-2.0	20.1	26.1	α-helix J'	CPR contact	-0.7	1.0	neutral										unknown
V359E	-6.1	0.4	2.0	α-helix	undefined	-2.7	0.0	damagin										unknown

				K				g											
T363M[1]	46.8	0.4	0.5	α-helix K	core-stabilising motif	2.1	0.2	damaging	*E.coli*	decrease	decrease						[540]	decreased activity
									HepG2	decrease		Testosterone		0.38	0.22	0.59	[542]	
L373F[1]	23.2	13.5	12.2	start of β-sheet 3[1]	haem contact/SRS5	0.0	0.0	neutral	*E.coli*	same		Testosterone				altered metabolic profile	[540]	decreased activity
												Midazolam		4.0				
P411L	54.0	12.0	12.9	btwn α-helix K' & 3_{10} helix	meander	2.3	0.3	damaging (5/7)										unknown
P416L[1]	54.0	5.2	4.2	start of 3_{10} helix	meander	3.9	0.2	damaging	*E.coli*	decrease	ND						[540]	inactive
I431T	-50.6	0.3	1.2	btwn 3_{10} helix & helix L	undefined	-2.9	0.7	damaging									[539]	unknown
M445T[1]	-46.8	39.6	28.4	start of α-helix L	CPR contact	-2.6	0.3	damaging	*E.coli*	similar		Testosterone	similar				[540]	neutral
									E.coli			Testosterone	similar (0.84)				[539]	
												Chlorpyrifos	similar				[539]	
R446K	-4.8	53.1	57.4	α-helix L	CPR contact	0.0	0.1	neutral										unknown
F463C	-81.4	0.1	5.2	start of β-sheet 5[1]	undefined	-8.1	0.0	damaging										unknown

P467S[1]	-23.7	30.3	32.5	loop btwn β-sheet 5[1] & 5[2]	undefined	-1.4	1.1	neutral	*E.coli*			Testosterone	Similar (0.84)				[539]	neutral
												Chlorpyrifos	similar					
S478C	19.5	23.5	33.5	loop btwn β-sheet 5[1] & 5[2]	SRS6	-8.7	2.3	damaging										unknown

Substrate abbreviations

Testosterone	Testosterone – 6β hydroxylation
Cortisol	Cortisol 6β- hydroxylation
Nifedipine	Nifedipine oxidation
Midazolam	Midazolam 1' hydroxylation
Chlorpyrifos	Chlorpyrifos-oxon/trichloropyridinol

Notes

1 SNPs found in the Human Cytochrome P450 (CYP) Allele Nomenclature Database, http://www.cypalleles.ki.se/.
2 *RSA* refers to the relative solvent accessibility as calculated by SDM.
3 *SD* refers to the standard deviation of the SDM score across different crystal structures of the same isoform.
4 *Relative holoprotein levels* refers to the P450:P420 ratio (usually determined by a carbon monoxide difference spectra) relative to wild-type enzyme.
5 *Relative activity* refers to the activity relative to wild-type enzyme where full kinetics is not available and turnover has only been measured for a limited number of substrate concentrations.
ND refers to protein levels, activity or kinetic parameters that were not detectable and the asterisk (*) indicates data reported to be significantly different from the wild-type enzyme. .
(Note that some studies provide quantitative data without indicating whether differences between wild-type and mutant proteins are statistically significant)

Appendix C. **Expression construct sequences**

The DNA sequences for N-terminal truncated CYP3A4, CYP2C9 and CYP3A4 cloned into the pBJW102.2 parent expression vector are shown below. The start and stop codons are underlined and the coding region is shown in bold with sequences coding for the CYP3A4/CYP2C9/CPR in red, BCCP-tag in blue, His-tag in green and the G-S linkers in yellow. Position of SNPs for variants expressed in this study are boxed in black with the nucleotide substitution and corresponding amino acid substitution shown in the right margin. The translated protein sequences, with the position of variants indicated, are also provided

C-terminal tagged CYP3A4 expression construct

```
   1 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA
  61 ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGGCA
 121 CTTACGGCCT ATGGAACCCA TTCACATGGA CTTTTTAAGA AGCTTGGAAT TCCAGGGCCC
 181 ACACCTCTGC CTTTTTTGGG AAATATTTTG TCCTACCATA AGGGCTTTTG TATGTTTGAC  G>A  G56D
 241 ATGGAATGTC ATAAAAAGTA TGGAAAAGTG TGGGGCTTTT ATGATGGTCA ACAGCCTGTG
 301 CTGGCTATCA CAGATCCTGA CATGATCAAA ACAGTGCTAG TGAAAGAATG TTATTCTGTC  A>G  K96E
 361 TTCACAAACC GGAGGCCTTT TGGTCCAGTG GGATTTATGA AAAGTGCCAT CTCTATAGCT
 421 GAGGATGAAG AATGGAAGAG ATTACGATCA TTGCTGTCTC CAACCTTCAC CAGTGGAAAA
 481 CTCAAGGAGA TGGTCCCTAT CATTGCCCAG TATGGAGATG TGTTGGTGAG AAATCTGAGG
 541 CGGGAAGCAG AGACAGGCAA GCCTGTCACC TTGAAAGACA TCTTTGGGGC CTACAGCATG
 601 GATGTGATCA CTAGCACATC ATTTGGAGTG AACATCGACT CTCTCAACAA TCCACAAGAC
 661 CCCTTTGTGG AAAACACCAA GAAGCTTTTA AGATTTGATT TTTTGGATCC ATTCTTTCTC  C>G  P218R
 721 TCAATAACAG TCTTTCCATT CCTCATCCCA ATTCTTGAAG TATTAAATAT CTGTGTGTTT  T>C  S222P
 781 CCAAGGGAAG TTACAAATTT TTTAAGAAAA TCTATAAAAA GGATGAAAGA AAGTCGCCTC  T>G  I223R
 841 GAAGATACAC AAAAGCACCG AGTGGATTTC CTTCAGCTGA TGATTGACTC TCAGAATTCA
 901 AAAGAAACTG AGTCCCACAA AGCTCTGTCC GATCTGGAGC TCGTGGCCCA ATCAATTATC  T>C  L293P
 961 TTTATTTTTG CTGGCTATGA AACCACGAGC AGTGTTCTCT CCTTCATTAT GTATGAACTG
1021 GCCACTCACC CTGATGTCCA GCAGAAACTG CAGGAGGAAA TTGATGCAGT TTTACCCAAT
1081 AAGGCACCAC CCACCTATGA TACTGTGCTA CAGATGGAGT ATCTTGACAT GGTGGTGAAT
1141 GAAACGCTCA GATTATTCCC AATTGCTATG AGACTTGAGA GGGTCTGCAA AAAAGATGTT  C>T  L373F
1201 GAGATCAATG GGATGTTCAT TCCCAAAGGG GTGGTGGTGA TGATTCCAAG CTATGCTCTT
1261 CACCGTGACC CAAAGTACTG GACAGAGCCT GAGAAGTTCC TCCCTGAAAG ATTCAGCAAG
1321 AAGAACAAGG ACAACATAGA TCCTTACATA TACACACCCT TTGGAAGTGG ACCCAGAAAC
1381 TGCATTGGTA TGAGGTTTGC TCTCATGAAC ATGAAACTTG CTCTAATCAG AGTCCTTCAG  T>C  M445T
1441 AACTTCTCCT TCAAACCTTG TAAAGAAACA CAGATCCCCC TGAAATTAAG CTTAGGAGGA
1501 CTTCTTCAAC CAGAAAAACC CGTTGTTCTA AAGGTTGAGT CAAGGGATGG CACCGTAAGT
1561 GGAGCCGGGG GTGGCAGCGG TTCTGGCGCA GCAGCGGAAA TCAGTGGTCA CATCGTACGT
1621 TCCCCGATGG TTGGTACTTT CTACCGCACC CCAAGCCCGG ACGCAAAAGC GTTCATCGAA
1681 GTGGGTCAGA AAGTCAACGT GGGCGATACC CTGTGCATCG TTGAAGCCAT GAAAATGATG
1741 AACCAGATCG AAGCGGACAA ATCCGGTACC GTGAAAGCAA TTCTGGTCGA AAGTGGACAA
1801 CCGGTAGAAT TTGACGAGCC GCTGGTCGTC ATCGAGGGTG GCAGCGGTTC TGGCCACCAT
1861 CACCATCACC ATAAGCTTAA TTAGCTGAGC TTGGACTCCT GTTGATAGAT CCAGTAATGA
1921 CCTCAGAACT CCATCTGGAT TTGTTCAGAA CGCTCGGTTG CCGCCGGGCG TTTTTTATTG
1981 GTGAGAATCC AAGCTAGCTT GGCGAGATTT TCAGGAGCTA AGGAAGCTAA AATGGAGAAA
2041 AAAATCACTG GATATACCAC CGTTGATATA TCCCAATGGC ATCGTAAAGA ACATTTTGAG
2101 GCATTTCAGT CAGTTGCTCA ATGTACCTAT AACCAGACCG TTCAGCTGGA TATTACGGCC
2161 TTTTTAAAGA CCGTAAAGAA AAATAAGCAC AAGTTTTATC CGGCCTTTAT TCACATTCTT
2221 GCCCGCCTGA TGAATGCTCA TCCGGAATTT CGTATGGCAA TGAAAGACGG TGAGCTGGTG
2281 ATATGGGATA GTGTTCACCC TTGTTACACC GTTTTCCATG AGCAAACTGA AACGTTTTCA
2341 TCGCTCTGGA GTGAATACCA CGACGATTTC CGGCAGTTTC TACACATATA TTCGCAAGAT
2401 GTGGCGTGTT ACGGTGAAAA CCTGGCCTAT TTCCCTAAAG GGTTTATTGA GAATATGTTT
```

```
2461 TTCGTCTCAG CCAATCCCTG GGTGAGTTTC ACCAGTTTTG ATTTAAACGT GGCCAATATG
2521 GACAACTTCT TCGCCCCCGT TTTCACCATG GGCAAATATT ATACGCAAGG CGACAAGGTG
2581 CTGATGCCGC TGGCGATTCA GGTTCATCAT GCCGTTTGTG ATGGCTTCCA TGTCGGCAGA
2641 ATGCTTAATG AATTACAACA GTACTGCGAT GAGTGGCAGG GCGGGGCGTA ATTTTTTTAA
2701 GGCAGTTATT GGTGCCCTTA AACGCCTGGG GTAATGACTC TCTAGCTTGA GGCATCAAAT
2761 AAAACGAAAG GCTCAGTCGA AAGACTGGGC CTTTCGTTTT ATCTGTTGTT TGTCGGTGAA
2821 CGCTCTCCTG AGTAGGACAA ATCCGCCCTC TAGATTACGT GCAGTCGATG ATAAGCTGTC
2881 AAACATGAGA ATTGTGCCTA ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG
2941 CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG
3001 GGGAGAGGCG GTTTGCGTAT TGGGCGCCAG GGTGGTTTTT CTTTTCACCA GTGAGACGGG
3061 CAACAGCTGA TTGCCCTTCA CCGCCTGGCC CTGAGAGAGT TGCAGCAAGC GGTCCACGCT
3121 GGTTTGCCCC AGCAGGCGAA AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA
3181 GCTGTCTTCG GTATCGTCGT ATCCCACTAC CGAGATATCC GCACCAACGC GCAGCCCGGA
3241 CTCGGTAATG GCGCGCATTG CGCCCAGCGC CATCTGATCG TTGGCAACCA GCATCGCAGT
3301 GGGAACGATG CCCTCATTCA GCATTTGCAT GGTTTGTTGA AAACCGGACA TGGCACTCCA
3361 GTCGCCTTCC CGTTCCGCTA TCGGCTGAAT TTGATTGCGA GTGAGATATT TATGCCAGCC
3421 AGCCAGACGC AGACGCGCCG AGACAGAACT TAATGGGCCC GCTAACAGCG CGATTTGCTG
3481 GTGACCCAAT GCGACCAGAT GCTCCACGCC CAGTCGCGTA CCGTCTTCAT GGGAGAAAAT
3541 AATACTGTTG ATGGGTGTCT GGTCAGAGAC ATCAAGAAAT AACGCCGGAA CATTAGTGCA
3601 GGCAGCTTCC ACAGCAATGG CATCCTGGTC ATCCAGCGGA TAGTTAATGA TCAGCCCACT
3661 GACGCGTTGC GCGAGAAGAT TGTGCACCGC CGCTTTACAG GCTTCGACGC CGCTTCGTTC
3721 TACCATCGAC ACCACCACGC TGGCACCCAG TTGATCGGCG CGAGATTTAA TCGCCGCGAC
3781 AATTTGCGAC GGCGCGTGCA GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG
3841 TTTGCCCGCC AGTTGTTGTG CCACGCGGTT GGGAATGTAA TTCAGCTCCG CCATCGCCGC
3901 TTCCACTTTT TCCCGCGTTT TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGGAAAC
3961 GGTCTGATAA GAGACACCGG CATACTCTGC GACATCGTAT AACGTTACTG GTTTCACATT
4021 CACCACCCTG AATTGACTCT CTTCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTGCA
4081 CCATTCGATG GTGTCGGAAT TTCGGGCAGC GTTGGGTCCT GGCCACGGGT GCGCATGATC
4141 TAGAGCTGCC TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
4201 GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG
4261 TCAGCGGGTG TTGGCGGGTG TCGGGGCGCA GCCATGACCC AGTCACGTAG CGATAGCGGA
4321 GTGTATACTG GCTTAACTAT GCGGCATCAG AGCAGATTGT ACTGAGAGTG CACCATATGC
4381 GGTGTGAAAT ACCGCACAGA TGCGTAAGGA GAAAATACCG CATCAGGCGC TCTTCCGCTT
4441 CCTCGCTCAC TGACTCGCTG CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT
4501 CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG
4561 CAAAAGGCCA GCAAAAGGCC AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA
4621 GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC
4681 CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG
4741 TTCCGACCCT GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC
4801 TTTCTCATAG CTCACGCTGT AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG
4861 GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC
4921 TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA
4981 TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG
5041 GCTACACTAG AAGGACAGTA TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA
5101 AAAGAGTTGG TAGCTCTTGA TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG
5161 TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT
5221 CTACGGGGTC TGACGCTCAG TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT
5281 TATCAAAAAG GATCTTCACC TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT
5341 AAAGTATATA TGAGTAAACT TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA
5401 TCTCAGCGAT CTGTCTATTT CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA
5461 CTACGATACG GGAGGGCTTA CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC
5521 GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA
5581 GTGGTCCTGC AACTTTATCC GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG
5641 TAAGTAGTTC GCCAGTTAAT AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG
5701 TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG
5761 TTACATGATC CCCCATGTTG TGCAAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG
5821 TCAGAAGTAA GTTGGCCGCA GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC
5881 TTACTGTCAT GCCATCCGTA AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT
5941 TCTGAGAATA GTGTATGCGG CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA
6001 CCGCGCCACA TAGCAGAACT TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA
```

```
6061 AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA
6121 ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC
6181 AAAATGCCGC AAAAAAGGGA ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC
6241 TTTTTCAATA TTATTGAAGC ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG
6301 AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC
6361 CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA
6421 GGCCCTTTCG TCTTCAC
```

C-terminal tagged CYP2C9 expression construct

```
   1 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA
  61 ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGGCA
 121 CTTACGGCCC TCCCTCCTGG CCCCACTCCT CTCCCAGTGA TTGGAAATAT CCTACAGATA
 181 GGTATTAAGG ACATCAGCAA ATCCTTAACC AATCTCTCAA AGGTCTATGG CCCTGTGTTC
 241 ACTCTGTATT TTGGCCTGAA ACCCATAGTG GTGCTGCATG GATATGAAGC AGTGAAGGAA G>A  V76M
 301 GCCCTGATTG ATCTTGGAGA GGAGTTTTCT GGAAGAGGCA TTTTCCCACT GGCTGAAAGA
 361 GCTAACAGAG GATTTGGAAT TGTTTTCAGC AATGGAAAGA AATGGAAGGA GATCCGGCGT
 421 TTCTCCCTCA TGACGCTGCG GAATTTTGGG ATGGGGAAGA GGAGCATTGA GGACCGTGTT G>A  R132Q
 481 CAAGAGGAAG CCCGCTGCCT TGTGGAGGAG TTGAGAAAAA CCAAGGCCTC ACCCTGTGAT G>A  R150H
 541 CCCACTTTCA TCCTGGGCTG TGCTCCCTGC AATGTGATCT GCTCCATTAT TTTCCATAAA
 601 CGTTTTGATT ATAAAGATCA GCAATTTCTT AACTTAATGG AAAAGTTGAA TGAAAACATC
 661 AAGATTTTGA GCAGCCCCTG GATCCAGATC TGCAATAATT TTTCTCCTAT CATTGATTAC A>T  Q214L
 721 TTCCCGGGAA CTCACAACAA ATTACTTAAA AACGTTGCTT TTATGAAAAG TTATATTTTG
 781 GAAAAAGTAA AAGAACACCA AGAATCAATG GACATGAACA ACCCTCAGGA CTTTATTGAT
 841 TGCTTCCTGA TGAAAATGGA GAAGGAAAAG CACAACCAAC CATCTGAATT TACTATTGAA C>A P297T
 901 AGCTTGGAAA ACACTGCAGT TGACTTGTTT GGAGCTGGGA CAGAGACGAC AAGCACAACC
 961 CTGAGATATG CTCTCCTTCT CCTGCTGAAG CACCCAGAGG TCACAGCTAA AGTCCAGGAA
1021 GAGATTGAAC GTGTGATTGG CAGAAACCGG AGCCCCTGCA TGCAAGACAG GAGCCACATG
1081 CCCTACACAG ATGCTGTGGT GCACGAGGTC CAGAGATACA TTGACCTTCT CCCCACCAGC A>C  I359L
1141 CTGCCCCATG CAGTGACCTG TGACATTAAA TTCAGAAACT ATCTCATTCC CAAGGGCACA T>C  I359T
1201 ACCATATTAA TTTCCCTGAC TTCTGTGCTA CATGACAACA AAGAATTTCC CAACCCAGAG
1261 ATGTTTGACC CTCATCACTT TCTGGATGAA GGTGGCAATT TTAAGAAAAG TAAATACTTC
1321 ATGCCTTTCT CAGCAGGAAA ACGGATTTGT GTGGGAGAAG CCCTGGCCGG CATGGAGCTG
1381 TTTTTATTCC TGACCTCCAT TTTACAGAAC TTTAACCTGA AATCTCTGGT TGACCCAAAG
1441 AACCTTGACA CCACTCCAGT TGTCAATGGA TTTGCCTCTG TGCCGCCCTT CTACCAGCTG
1501 TGCTTCATTC CTGTCGGGGG TGGCAGCGGT TCTGGCGCAG CAGCGGAAAT CAGTGGTCAC
1561 ATCGTACGTT CCCCGATGGT TGGTACTTTC TACCGCACCC CAAGCCCGGA CGCAAAAGCG
1621 TTCATCGAAG TGGGTCAGAA AGTCAACGTG GGCGATACCC TGTGCATCGT TGAAGCCATG
1681 AAAATGATGA ACCAGATCGA AGCGGACAAA TCCGGTACCG TGAAAGCAAT TCTGGTCGAA
1741 AGTGGACAAC CGGTAGAATT TGACGAGCCG CTGGTCGTCA TCGAGGGTGG CAGCGGTTCT
1801 GGCCACCATC ACCATCACCA TAAGCTTAAT TAGCTGAGCT TGGACTCCTG TTGATAGATC
1861 CAGTAATGAC CTCAGAACTC CATCTGGATT TGTTCAGAAC GCTCGGTTGC CGCCGGGCGT
1921 TTTTTATTGG TGAGAATCCA AGCTAGCTTG GCGAGATTTT CAGGAGCTAA GGAAGCTAAA
1981 ATGGAGAAAA AAATCACTGG ATATACCACC GTTGATATAT CCCAATGGCA TCGTAAAGAA
2041 CATTTTGAGG CATTTCAGTC AGTTGCTCAA TGTACCTATA ACCAGACCGT TCAGCTGGAT
2101 ATTACGGCCT TTTTAAAGAC CGTAAAGAAA AATAAGCACA AGTTTTATCC GGCCTTTATT
2161 CACATTCTTG CCCGCCTGAT GAATGCTCAT CCGGAATTTC GTATGGCAAT GAAAGACGGT
2221 GAGCTGGTGA TATGGGATAG TGTTCACCCT TGTTACACCG TTTTCCATGA GCAAACTGAA
2281 ACGTTTTCAT CGCTCTGGAG TGAATACCAC GACGATTTCC GGCAGTTTCT ACACATATAT
2341 TCGCAAGATG TGGCGTGTTA CGGTGAAAAC CTGGCCTATT TCCCTAAAGG GTTTATTGAG
2401 AATATGTTTT TCGTCTCAGC CAATCCCTGG GTGAGTTTCA CCAGTTTTGA TTTAAACGTG
2461 GCCAATATGG ACAACTTCTT CGCCCCCGTT TTCACCATGG GCAAATATTA TACGCAAGGC
2521 GACAAGGTGC TGATGCCGCT GGCGATTCAG GTTCATCATG CCGTTTGTGA TGGCTTCCAT
2581 GTCGGCAGAA TGCTTAATGA ATTACAACAG TACTGCGATG AGTGGCAGGG CGGGGCGTAA
2641 TTTTTTTAAG GCAGTTATTG GTGCCCTTAA ACGCCTGGGG TAATGACTCT CTAGCTTGAG
2701 GCATCAAATA AAACGAAAGG CTCAGTCGAA AGACTGGGCC TTTCGTTTTA TCTGTTGTTT
2761 GTCGGTGAAC GCTCTCCTGA GTAGGACAAA TCCGCCCTCT AGATTACGTG CAGTCGATGA
2821 TAAGCTGTCA AACATGAGAA TTGTGCCTAA TGAGTGAGCT AACTTACATT AATTGCGTTG
2881 CGCTCACTGC CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC AGCTGCATTA ATGAATCGGC
2941 CAACGCGCGG GGAGAGGCGG TTTGCGTATT GGGCGCCAGG GTGGTTTTTC TTTTCACCAG
```

3001 TGAGACGGGC AACAGCTGAT TGCCCTTCAC CGCCTGGCCC TGAGAGAGTT GCAGCAAGCG
3061 GTCCACGCTG GTTTGCCCCA GCAGGCGAAA ATCCTGTTTG ATGGTGGTTA ACGGCGGGAT
3121 ATAACATGAG CTGTCTTCGG TATCGTCGTA TCCCACTACC GAGATATCCG CACCAACGCG
3181 CAGCCCGGAC TCGGTAATGG CGCGCATTGC GCCCAGCGCC ATCTGATCGT TGGCAACCAG
3241 CATCGCAGTG GGAACGATGC CCTCATTCAG CATTTGCATG GTTTGTTGAA AACCGGACAT
3301 GGCACTCCAG TCGCCTTCCC GTTCCGCTAT CGGCTGAATT TGATTGCGAG TGAGATATTT
3361 ATGCCAGCCA GCCAGACGCA GACGCGCCGA GACAGAACTT AATGGGCCCG CTAACAGCGC
3421 GATTTGCTGG TGACCCAATG CGACCAGATG CTCCACGCCC AGTCGCGTAC CGTCTTCATG
3481 GGAGAAAATA ATACTGTTGA TGGGTGTCTG GTCAGAGACA TCAAGAAATA ACGCCGGAAC
3541 ATTAGTGCAG GCAGCTTCCA CAGCAATGGC ATCCTGGTCA TCCAGCGGAT AGTTAATGAT
3601 CAGCCCACTG ACGCGTTGCG CGAGAAGATT GTGCACCGCC GCTTTACAGG CTTCGACGCC
3661 GCTTCGTTCT ACCATCGACA CCACCACGCT GGCACCCAGT TGATCGGCGC GAGATTTAAT
3721 CGCCGCGACA ATTTGCGACG GCGCGTGCAG GGCCAGACTG GAGGTGGCAA CGCCAATCAG
3781 CAACGACTGT TTGCCCGCCA GTTGTTGTGC CACGCGGTTG GGAATGTAAT TCAGCTCCGC
3841 CATCGCCGCT TCCACTTTTT CCCGCGTTTT CGCAGAAACG TGGCTGGCCT GGTTCACCAC
3901 GCGGGAAACG GTCTGATAAG AGACACCGGC ATACTCTGCG ACATCGTATA ACGTTACTGG
3961 TTTCACATTC ACCACCCTGA ATTGACTCTC TTCCGGGCGC TATCATGCCA TACCGCGAAA
4021 GGTTTTGCAC CATTCGATGG TGTCGGAATT TCGGGCAGCG TTGGGTCCTG GCCACGGGTG
4081 CGCATGATCT AGAGCTGCCT CGCGCGTTTC GGTGATGACG GTGAAAACCT CTGACACATG
4141 CAGCTCCCGG AGACGGTCAC AGCTTGTCTG TAAGCGGATG CCGGGAGCAG ACAAGCCCGT
4201 CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT CGGGGCGCAG CCATGACCCA GTCACGTAGC
4261 GATAGCGGAG TGTATACTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
4321 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCT
4381 CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT
4441 CAGCTCACTC AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA
4501 ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT
4561 TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT
4621 GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC
4681 GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA
4741 GCGTGGCGCT TTCTCATAGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT
4801 CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA
4861 ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG
4921 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC
4981 CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA
5041 CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG
5101 GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT
5161 TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG
5221 TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTTA
5281 AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG
5341 AGGCACCTAT CTCAGCGATC TGTCTATTTC GTTCATCCAT AGTTGCCTGA CTCCCCGTCG
5401 TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC
5461 GAGACCCACG CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG
5521 AGCGCAGAAG TGGTCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT TGTTGCCGGG
5581 AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG
5641 GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT
5701 CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC
5761 CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC
5821 ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT GAGTACTCAA
5881 CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC
5941 GGGATAATAC CGCGCCACAT AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT
6001 CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC
6061 GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA
6121 CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA
6181 TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT
6241 ACATATTTGA ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA
6301 AAGTGCCACC TGACGTCTAA GAAACCATTA TTATCATGAC ATTAACCTAT AAAAATAGGC
6361 GTATCACGAG GCCCTTTCGT CTTCAC

C-terminal tagged CPR expression construct

```
   1 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA
  61 ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGGCA
 121 CTTACGGCCT TCAGAAAGAA AAAAGAAGAA GTCCCCGAGT TCACCAAAAT TCAGACATTG
 181 ACCTCCTCTG TCAGAGAGAG CAGCTTTGTG GAAAAGATGA AGAAAACGGG GAGGAACATC
 241 ATCGTGTTCT ACGGCTCCCA GACGGGGACT GCAGAGGAGT TTGCCAACCG CCTGTCCAAG
 301 GACGCCCACC GCTACGGGAT GCGAGGCATG TCAGCGGACC CTGAGGAGTA TGACCTGGCC
 361 GACCTGAGCA GCCTGCCAGA GATCGACAAC GCCCTGGTGG TTTTCTGCAT GGCCACCTAC
 421 GGTGAGGGAG ACCCCACCGA CAATGCCCAG GACTTCTACG ACTGGCTGCA GGAGACAGAC
 481 GTGGATCTCT CTGGGGTCAA GTTCGCGGTG TTTGGTCTTG GGAACAAGAC CTACGAGCAC
 541 TTCAATGCCA TGGGCAAGTA CGTGGACAAG CGGCTGGAGC AGCTCGGCGC CCAGCGCATC
 601 TTTGAGCTGG GGTTGGGCGA CGACGATGGG AACTTGGAGG AGGACTTCAT CACCTGGCGA
 661 GAGCAGTTCT GGCCGGCCGT GTGTGAACAC TTTGGGGTGG AAGCCACTGG CGAGGAGTCC
 721 AGCATTCGCC AGTACGAGCT TGTGGTCCAC ACCGACATAG ATGCGGCCAA GGTGTACATG
 781 GGGGAGATGG GCCGGCTGAA GAGCTACGAG AACCAGAAGC CCCCCTTTGA TGCCAAGAAT
 841 CCGTTCCTGG CTGCAGTCAC CACCAACCGG AAGCTGAACC AGGGAACCGA GCGCCACCTC
 901 ATGCACCTGG AATTGGACAT CTCGGACTCC AAAATCAGGT ATGAATCTGG GGACCACGTG
 961 GCTGTGTACC CAGCCAACGA CTCTGCTCTC GTCAACCAGC TGGGCAAAAT CCTGGGTGCC
1021 GACCTGGACG TCGTCATGTC CCTGAACAAC CTGGATGAGG AGTCCAACAA GAAGCACCCA
1081 TTCCCGTGCC CTACGTCCTA CCGCACGGCC CTCACCTACT ACCTGGACAT CACCAACCCG
1141 CCGCGTACCA ACGTGCTGTA CGAGCTGGCG CAGTACGCCT CGGAGCCCTC GGAGCAGGAG
1201 CTGCTGCGCA AGATGGCCTC CTCCTCCGGC GAGGGCAAGG AGCTGTACCT GAGCTGGGTG
1261 GTGGAGGCCC GGAGGCACAT CCTGGCCATC CTGCAGGACT GCCCGTCCCT GCGGCCCCCC
1321 ATCGACCACC TGTGTGAGCT GCTGCCGCGC CTGCAGGCCC GCTACTACTC CATCGCCTCA
1381 TCCTCCAAGG TCCACCCCAA CTCTGTGCAC ATCTGTGCGG TGGTTGTGGA GTACGAGACC
1441 AAGGCCGGCC GCATCAACAA GGGCGTGGCC ACCAACTGGC TGCGGGCCAA GGAGCCTGCC
1501 GGGGAGAACG GCGGCCGTGC GCTGGTGCCC ATGTTCGTGC GCAAGTCCCA GTTCCGCCTG
1561 CCCTTCAAGG CCACCACGCC TGTCATCATG GTGGGCCCCG GCACCGGGGT GGCACCCTTC
1621 ATAGGCTTCA TCCAGGAGCG GGCCTGGCTG CGACAGCAGG GCAAGGAGGT GGGGGAGACG
1681 CTGCTGTACT ACGGCTGCCG CCGCTCGGAT GAGGACTACC TGTACCGGGA GGAGCTGGCG
1741 CAGTTCCACA GGGACGGTGC GCTCACCCAG CTCAACGTGG CCTTCTCCCG GGAGCAGTCC
1801 CACAAGGTCT ACGTCCAGCA CCTGCTAAAG CAAGACCGAG AGCACCTGTG GAAGTTGATC
1861 GAAGGCGGTG CCCACATCTA CGTCTGTGGG GATGCACGGA ACATGGCCAG GGATGTGCAG
1921 AACACCTTCT ACGACATCGT GGCTGAGCTC GGGGCCATGG AGCACGCGCA GGCGGTGGAC
1981 TACATCAAGA AACTGATGAC CAAGGGCCGC TACTCCCTGG ACGTGTGGAG CGGGGGTGGC
2041 AGCGGTTCTG GCGCAGCAGC GGAAATCAGT GGTCACATCG TACGTTCCCC GATGGTTGGT
2101 ACTTTCTACC GCACCCCAAG CCCGGACGCA AAAGCGTTCA TCGAAGTGGG TCAGAAAGTC
2161 AACGTGGGCG ATACCCTGTG CATCGTTGAA GCCATGAAAA TGATGAACCA GATCGAAGCG
2221 GACAAATCCG GTACCGTGAA AGCAATTCTG GTCGAAAGTG GACAACCGGT AGAATTTGAC
2281 GAGCCGCTGG TCGTCATCGA GGGTGGCAGC GGTTCTGGCC ACCATCACCA TCACCATAAG
2341 CTTAATTAGC TGAGCTTGGA CTCCTGTTGA TAGATCCAGT AATGACCTCA GAACTCCATC
2401 TGGATTTGTT CAGAACGCTC GGTTGCCGCC GGGCGTTTTT TATTGGTGAG AATCCAAGCT
2461 AGCTTGGCGA GATTTTCAGG AGCTAAGGAA GCTAAAATGG AGAAAAAAAT CACTGGATAT
2521 ACCACCGTTG ATATATCCCA ATGGCATCGT AAAGAACATT TTGAGGCATT TCAGTCAGTT
2581 GCTCAATGTA CCTATAACCA GACCGTTCAG CTGGATATTA CGGCCTTTTT AAAGACCGTA
2641 AAGAAAAATA AGCACAAGTT TTATCCGGCC TTTATTCACA TTCTTGCCCG CCTGATGAAT
2701 GCTCATCCGG AATTTCGTAT GGCAATGAAA GACGGTGAGC TGGTGATATG GGATAGTGTT
2761 CACCCTTGTT ACACCGTTTT CCATGAGCAA ACTGAAACGT TTTCATCGCT CTGGAGTGAA
2821 TACCACGACG ATTTCCGGCA GTTTCTACAC ATATATTCGC AAGATGTGGC GTGTTACGGT
2881 GAAAACCTGG CCTATTTCCC TAAAGGGTTT ATTGAGAATA TGTTTTTCGT CTCAGCCAAT
2941 CCCTGGGTGA GTTTCACCAG TTTTGATTTA AACGTGGCCA ATATGGACAA CTTCTTCGCC
3001 CCCGTTTTCA CCATGGGCAA ATATTATACG CAAGGCGACA AGGTGCTGAT GCCGCTGGCG
3061 ATTCAGGTTC ATCATGCCGT TTGTGATGGC TTCCATGTCG GCAGAATGCT TAATGAATTA
3121 CAACAGTACT GCGATGAGTG GCAGGGCGGG GCGTAATTTT TTTAAGGCAG TTATTGGTGC
3181 CCTTAAACGC CTGGGGTAAT GACTCTCTAG CTTGAGGCAT CAAATAAAAC GAAAGGCTCA
3241 GTCGAAAGAC TGGGCCTTTC GTTTTATCTG TTGTTTGTCG GTGAACGCTC TCCTGAGTAG
3301 GACAAATCCG CCCTCTAGAT TACGTGCAGT CGATGATAAG CTGTCAAACA TGAGAATTGT
3361 GCCTAATGAG TGAGCTAACT TACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG
3421 GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG
```

3481 CGTATTGGGC GCCAGGGTGG TTTTTCTTTT CACCAGTGAG ACGGGCAACA GCTGATTGCC
3541 CTTCACCGCC TGGCCCTGAG AGAGTTGCAG CAAGCGGTCC ACGCTGGTTT GCCCCAGCAG
3601 GCGAAAATCC TGTTTGATGG TGGTTAACGG CGGGATATAA CATGAGCTGT CTTCGGTATC
3661 GTCGTATCCC ACTACCGAGA TATCCGCACC AACGCGCAGC CCGGACTCGG TAATGGCGCG
3721 CATTGCGCCC AGCGCCATCT GATCGTTGGC AACCAGCATC GCAGTGGGAA CGATGCCCTC
3781 ATTCAGCATT TGCATGGTTT GTTGAAAACC GGACATGGCA CTCCAGTCGC CTTCCCGTTC
3841 CGCTATCGGC TGAATTTGAT TGCGAGTGAG ATATTTATGC CAGCCAGCCA GACGCAGACG
3901 CGCCGAGACA GAACTTAATG GGCCCGCTAA CAGCGCGATT TGCTGGTGAC CCAATGCGAC
3961 CAGATGCTCC ACGCCCAGTC GCGTACCGTC TTCATGGGAG AAAATAATAC TGTTGATGGG
4021 TGTCTGGTCA GAGACATCAA GAAATAACGC CGGAACATTA GTGCAGGCAG CTTCCACAGC
4081 AATGGCATCC TGGTCATCCA GCGGATAGTT AATGATCAGC CCACTGACGC GTTGCGCGAG
4141 AAGATTGTGC ACCGCCGCTT TACAGGCTTC GACGCCGCTT CGTTCTACCA TCGACACCAC
4201 CACGCTGGCA CCCAGTTGAT CGGCGCGAGA TTTAATCGCC GCGACAATTT GCGACGGCGC
4261 GTGCAGGGCC AGACTGGAGG TGGCAACGCC AATCAGCAAC GACTGTTTGC CCGCCAGTTG
4321 TTGTGCCACG CGGTTGGGAA TGTAATTCAG CTCCGCCATC GCCGCTTCCA CTTTTTCCCG
4381 CGTTTTCGCA GAAACGTGGC TGGCCTGGTT CACCACGCGG GAAACGGTCT GATAAGAGAC
4441 ACCGGCATAC TCTGCGACAT CGTATAACGT TACTGGTTTC ACATTCACCA CCCTGAATTG
4501 ACTCTCTTCC GGGCGCTATC ATGCCATACC GCGAAAGGTT TTGCACCATT CGATGGTGTC
4561 GGAATTTCGG GCAGCGTTGG GTCCTGGCCA CGGGTGCGCA TGATCTAGAG CTGCCTCGCG
4621 CGTTTCGGTG ATGACGGTGA AAACCTCTGA CACATGCAGC TCCCGGAGAC GGTCACAGCT
4681 TGTCTGTAAG CGGATGCCGG GAGCAGACAA GCCCGTCAGG GCGCGTCAGC GGGTGTTGGC
4741 GGGTGTCGGG GCGCAGCCAT GACCCAGTCA CGTAGCGATA GCGGAGTGTA TACTGGCTTA
4801 ACTATGCGGC ATCAGAGCAG ATTGTACTGA GAGTGCACCA TATGCGGTGT GAAATACCGC
4861 ACAGATGCGT AAGGAGAAAA TACCGCATCA GGCGCTCTTC CGCTTCCTCG CTCACTGACT
4921 CGCTGCGCTC GGTCGTTCGG CTGCGGCGAG CGGTATCAGC TCACTCAAAG GCGGTAATAC
4981 GGTTATCCAC AGAATCAGGG GATAACGCAG GAAAGAACAT GTGAGCAAAA GGCCAGCAAA
5041 AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT CCATAGGCTC CGCCCCCCTG
5101 ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA
5161 GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC TCCTGTTCCG ACCCTGCCGC
5221 TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT GGCGCTTTCT CATAGCTCAC
5281 GCTGTAGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT GTGCACGAAC
5341 CCCCCGTTCA GCCCGACCGC TGCGCCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG
5401 TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC AGAGCGAGGT
5461 ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCCTAA CTACGGCTAC ACTAGAAGGA
5521 CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA GTTGGTAGCT
5581 CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT TTTTGTTTGC AAGCAGCAGA
5641 TTACGCGCAG AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGGTCTGACG
5701 CTCAGTGGAA CGAAAACTCA CGTTAAGGGA TTTTGGTCAT GAGATTATCA AAAAGGATCT
5761 TCACCTAGAT CCTTTTAAAT TAAAAATGAA GTTTTAAATC AATCTAAAGT ATATATGAGT
5821 AAACTTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC ACCTATCTCA GCGATCTGTC
5881 TATTTCGTTC ATCCATAGTT GCCTGACTCC CCGTCGTGTA GATAACTACG ATACGGGAGG
5941 GCTTACCATC TGGCCCCAGT GCTGCAATGA TACCGCGAGA CCCACGCTCA CCGGCTCCAG
6001 ATTTATCAGC AATAAACCAG CCAGCCGGAA GGGCCGAGCG CAGAAGTGGT CCTGCAACTT
6061 TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC TAGAGTAAGT AGTTCGCCAG
6121 TTAATAGTTT GCGCAACGTT GTTGCCATTG CTACAGGCAT CGTGGTGTCA CGCTCGTCGT
6181 TTGGTATGGC TTCATTCAGC TCCGGTTCCC AACGATCAAG GCGAGTTACA TGATCCCCCA
6241 TGTTGTGCAA AAAAGCGGTT AGCTCCTTCG GTCCTCCGAT CGTTGTCAGA AGTAAGTTGG
6301 CCGCAGTGTT ATCACTCATG GTTATGGCAG CACTGCATAA TTCTCTTACT GTCATGCCAT
6361 CCGTAAGATG CTTTTCTGTG ACTGGTGAGT ACTCAACCAA GTCATTCTGA GAATAGTGTA
6421 TGCGGCGACC GAGTTGCTCT TGCCCGGCGT CAATACGGGA TAATACCGCG CCACATAGCA
6481 GAACTTTAAA AGTGCTCATC ATTGGAAAAC GTTCTTCGGG GCGAAAACTC TCAAGGATCT
6541 TACCGCTGTT GAGATCCAGT TCGATGTAAC CCACTCGTGC ACCCAACTGA TCTTCAGCAT
6601 CTTTTACTTT CACCAGCGTT TCTGGGTGAG CAAAAACAGG AAGGCAAAAT GCCGCAAAAA
6661 AGGGAATAAG GGCGACACGG AAATGTTGAA TACTCATACT CTTCCTTTTT CAATATTATT
6721 GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT ATTTAGAAAA
6781 ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT GCCACCTGAC GTCTAAGAAA
6841 CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCGTAT CACGAGGCCC TTTCGTCTTC
6901 AC

CYP3A4 recombinant protein sequence

```
  1 MALTAYGTHS HGLFKKLGIP GPTPLPFLGN ILSYHKGFCM FDMECHKKYG KVWGFYDGQQ  G56D
 61 PVLAITDPDM IKTVLVKECY SVFTNRRPFG PVGFMKSAIS IAEDEEWKRL RSLLSPTFTS  K96E
121 GKLKEMVPII AQYGDVLVRN LRREAETGKP VTLKDIFGAY SMDVITSTSF GVNIDSLNNP
181 QDPFVENTKK LLRFDFLDPF FLSITVFPFL IPILEVLNIC VFPREVTNFL RKSIKRMKES P218R; S222P; I223R
241 RLEDTQKHRV DFLQLMIDSQ NSKETESHKA LSDLELVAQS IIFIFAGYET TSSVLSFIMY L293P
301 ELATHPDVQQ KLQEEIDAVL PNKAPPTYDT VLQMEYLDMV VNETLRLFPI AMRLERVCKK L373F
361 DVEINGMFIP KGVVVMIPSY ALHRDPKYWT EPEKFLPERF SKKNKDNIDP YIYTPFGSGP
421 RNCIGMRFAL MNMKLALIRV LQNFSFKPCK ETQIPLKLSL GGLLQPEKPV VLKVESRDGT M445T
481 VSGAGGGSGS GAAAEISGHI VRSPMVGTFY RTPSPDAKAF IEVGQKVNVG DTLCIVEAMK
541 MMNQIEADKS GTVKAILVES GQPVEFDEPL VVIEGGSGSG HHHHHHKLN*
```

CYP2C9 recombinant protein sequence

```
  1 MALTALPPGP TPLPVIGNIL QIGIKDISKS LTNLSKVYGP VFTLYFGLKP IVVLHGYEAV V76M
 61 KEALIDLGEE FSGRGIFPLA ERANRGFGIV FSNGKKWKEI RRFSLMTLRN FGMGKRSIED R132Q
121 RVQEEARCLV EELRKTKASP CDPTFILGCA PCNVICSIIF HKRFDYKDQQ FLNLMEKLNE R150H
181 NIKILSSPWI QICNNFSPII DYFPGTHNKL LKNVAFMKSY ILEKVKEHQE SMDMNNPQDF Q214L
241 IDCFLMKMEK EKHNQESEFT IESLENTAVD LFGAGTETTS TTLRYALLLL LKHPEVTAKV P279T
301 QEEIERVIGR NRSPCMQDRS HMPYTDAVVH EVQRYTDLLP TSLPHAVTCD IKFRNYLIPK I359L/T
361 GTTILISLTS VLHDNKEFPN PEMFDPHHFL DEGGNFKKSK YFMPFSAGKR ICVGEALAGM
421 ELFLFLTSIL QNFNLKSLVD PKNLDTTPVV NGFASVPPFY QLCFIPVGGG SGSGAAAEIS
481 GHIVRSPMVG TFYRTPSPDA KAFIEVGQKV NVGDTLCIVE AMKMMNQIEA DKSGTVKAIL
541 VESGQPVEFD EPLVVIEGGS GSGHHHHHHK LN*
```

CPR recombinant protein sequence

```
  1 MALTAFRKKK EEVPEFTKIQ TLTSSVRESS FVEKMKKTGR NIIVFYGSQT GTAEEFANRL
 61 SKDAHRYGMR GMSADPEEYD LADLSSLPEI DNALVVFCMA TYGEGDPTDN AQDFYDWLQE
121 TDVDLSGVKF AVFGLGNKTY EHFNAMGKYV DKRLEQLGAQ RIFELGLGDD DGNLEEDFIT
181 WREQFWPAVC EHFGVEATGE ESSIRQYELV VHTDIDAAKV YMGEMGRLKS YENQKPPFDA
241 KNPFLAAVTT NRKLNQGTER HLMHLELDIS DSKIRYESGD HVAVYPANDS ALVNQLGKIL
301 GADLDVVMSL NNLDEESNKK HPFPCPTSYR TALTYYLDIT NPPRTNVLYE LAQYASEPSE
361 QELLRKMASS SGEGKELYLS WVVEARRHIL AILQDCPSLR PPIDHLCELL PRLQARYYSI
421 ASSSKVHPNS VHICAVVVEY ETKAGRINKG VATNWLRAKE PAGENGGRAL VPMFVRKSQF
481 RLPFKATTPV IMVGPGTGVA PFIGFIQERA WLRQQGKEVG ETLLYYGCRR SDEDYLYREE
541 LAQFHRDGAL TQLNVAFSRE QSHKVYVQHL LKQDREHLWK LIEGGAHIYV CGDARNMARD
601 VQNTFYDIVA ELGAMEHAQA VDYIKKLMTK GRYSLDVWSG GGSGSGAAAE ISGHIVRSPM
661 VGTFYRTPSP DAKAFIEVGQ KVNVGDTLCI VEAMKMMNQI EADKSGTVKA ILVESGQPVE
721 FDEPLVVIEG GSGSGHHHHH HKLN*
```

References

[1] Nelson, D.R.; Nebert, D.W. Cytochrome P450 (CYP) gene superfamily. *eLS,* **2011**.

[2] Omura, T. Contribution of cytochrome P450 to the diversification of eukaryotic organisms. *Biotechnol. Appl. Biochem.,* **2013***, 60*, 4-8.

[3] Omura, T.; Sato, R. A new cytochrome in liver microsomes. *J. Biol. Chem.,* **1962***, 237*, 1375-1376.

[4] Omura, T.; Sato, R. The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *J. Biol. Chem.,* **1964***, 239*, 2370-2378.

[5] Omura, T.; Sato, R. The carbon monoxide-binding pigment of liver microsomes II. Solubilization, purification, and properties. *J. Biol. Chem.,* **1964***, 239*, 2379-85.

[6] Kumar, S. Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. *Expert opinion on drug metabolism & toxicology,* **2010***, 6*, 115-131.

[7] Ingelman-sundberg, M.; Sim, S.C.; Gomez, A.; Rodriguez-antona, C. Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. *Cancer,* **2007***, 116*, 496-526.

[8] Danielson, P.B. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metab.,* **2002***, 3*, 561-97.

[9] Chothia, C. Protein families in the metazoan genome. *Development,* **1994***, 1994*, 27-33.

[10] Gonzalez, F.J.; Nebert, D.W. Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. *Trends in Genetics,* **1990***, 6*, 182-186.

[11] Nelson, D.R. The cytochrome p450 homepage. http://drnelson.uthsc.edu/CytochromeP450.html (Accessed June 6, **2014**).

[12] Nelson, D.R. The cytochrome p450 homepage. *Human genomics,* **2009***, 4*, 59.

[13] Guengerich, F.P. Mechanisms of cytochrome P450 substrate oxidation: MiniReview. *J. Biochem. Mol. Toxicol.,* **2007***, 21*, 163-168.

[14] Munro, A.W.; Girvan, H.M.; Mason, A.E.; Dunford, A.J.; McLean, K.J. What makes a P450 tick? *Trends Biochem. Sci.,* **2013***, 38*, 140-150.

[15] Meunier, B.; de Visser, S.P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. *Chem. Rev.,* **2004***, 104*, 3947-80.

[16] Makris, T.M.; Koenig, K.v.; Schlichting, I.; Sligar, S.G. The status of high-valent metal oxo complexes in the P450 cytochromes. *J. Inorg. Biochem.,* **2006***, 100*, 507-518.

[17] Rittle, J.; Green, M.T. Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science,* **2010***, 330*, 933-937.

[18] Krest, C.M.; Onderko, E.L.; Yosca, T.H.; Calixto, J.C.; Karp, R.F.; Livada, J.; Rittle, J.; Green, M.T. Reactive intermediates in cytochrome p450 catalysis. *J. Biol. Chem.,* **2013***, 288*, 17074-17081.

[19] Denisov, I.G.; Makris, T.M.; Sligar, S.G.; Schlichting, I. Structure and chemistry of cytochrome P450. *Chem. Rev.,* **2005***, 105*, 2253-2278.

[20] Klingenberg, M. Pigments of rat liver microsomes. *Arch. Biochem. Biophys.,* **1958***, 75*, 376-386.

[21] Sun, Y.; Zeng, W.; Benabbas, A.; Ye, X.; Denisov, I.; Sligar, S.G.; Du, J.; Dawson, J.H.; Champion, P.M. Investigations of heme ligation and ligand switching in cytochromes P450 and P420. *Biochemistry (N. Y.),* **2013***, 52*, 5941-5951.

[22] Sligar, S.G. Coupling of spin, substrate, and redox equilibriums in cytochrome P450. *Biochemistry (N. Y.),* **1976***, 15*, 5399-5406.

[23] Lange, R.; Bonfrils, C.; Debey, P. The Low-Spin⇌ High-Spin Transition of Camphor-Bound Cytochrome P-450. *Eur. J. Biochem.,* **1977***, 79*, 623-628.
[24] Lange, R.; Balny, C. UV-visible derivative spectroscopy under high pressure. *BBA-Protein Structure and Molecular Enzymology,* **2002***, 1595*, 80-93.
[25] Goeptar, A.R.; Scheerens, H.; Vermeulen, N.P. Oxygen and xenobiotic reductase activities of cytochrome P450. *CRC Crit. Rev. Toxicol.,* **1995***, 25*, 25-65.
[26] White, R.E.; McCarthy, M.B.; Egeberg, K.D.; Sligar, S.G. Regioselectivity in the cytochromes P-450: control by protein constraints and by chemical reactivities. *Arch. Biochem. Biophys.,* **1984***, 228*, 493-502.
[27] Raucy, J.L.; Kraner, J.C.; Lasker, J.M. Bioactivation of halogenated hydrocarbons by cytochrome P4502E1. *CRC Crit. Rev. Toxicol.,* **1993***, 23*, 1-20.
[28] Vaz, A.D.; McGinnity, D.F.; Coon, M.J. Epoxidation of olefins by cytochrome P450: evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant. *Proc. Natl. Acad. Sci. U. S. A.,* **1998***, 95*, 3555-3560.
[29] Guengerich, F.P.; Munro, A.W. Unusual cytochrome p450 enzymes and reactions. *J. Biol. Chem.,* **2013***, 288*, 17065-17073.
[30] Poulos, T.L.; Finzel, B.C.; Gunsalus, I.C. The 2.6-Å crystal structure of Pseudomonas putida cytochrome P-450. *J. Biol. Chem.,* **1985***, 260*, 16122-16130.
[31] Research Collaboratory for Structural Bioinformatics. Protein Data Bank: A Resource for Studying Biological Macromolecules. (http://www.pdb.org/).
[32] Hasemann, C.A.; Kurumbail, R.G.; Boddupalli, S.S.; Peterson, J.A.; Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure,* **1995***, 3*, 41-62.
[33] Otyepka, M.; Berka, K.; Anzenbacher, P. Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450? *Curr. Drug Metab.,* **2012***, 13*, 130-142.
[34] Graham, S.E.; Peterson, J.a. How similar are P450s and what can their differences teach us? *Arch. Biochem. Biophys.,* **1999***, 369*, 24-9.
[35] Raag, R.; Martinis, S.A.; Sligar, S.G.; Poulos, T.L. Crystal structure of the cytochrome P-450CAM active site mutant Thr252Ala. *Biochemistry,* **1991***, 30*, 11420-11429.
[36] Gerber, N.C.; Sligar, S.G. Catalytic mechanism of cytochrome P-450: evidence for a distal charge relay. *J. Am. Chem. Soc.,* **1992***, 114*, 8742-8743.
[37] Poulos, T.L.; Finzel, B.C.; Howard, A.J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.,* **1987***, 195*, 687-700.
[38] Neve, E.P.A.; Ingelman-Sundberg, M. Intracellular transport and localization of microsomal cytochrome P450. *Anal. Bioanal. Chem.,* **2008***, 392*, 1075-84.
[39] Chen, C.D.; Kemper, B. Different structural requirements at specific proline residue positions in the conserved proline-rich region of cytochrome P450 2C2. *J. Biol. Chem.,* **1996***, 271*, 28607-11.
[40] Chen, C.D.; Doray, B.; Kemper, B. A conserved proline-rich sequence between the N-terminal signal-anchor and catalytic domains is required for assembly of functional cytochrome P450 2C2. *Arch. Biochem. Biophys.,* **1998***, 350*, 233-8.
[41] Kemper, B. Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. *Toxicol. Appl. Pharmacol.,* **2004***, 199*, 305-15.
[42] Otyepka, M.; Skopalík, J.; Anzenbacherová, E.; Anzenbacher, P. What common structural features and variations of mammalian P450s are known to date? *Biochim. Biophys. Acta,* **2007***, 1770*, 376-389.

[43] Cosme, J.; Johnson, E.F. Engineering microsomal cytochrome P450 2C5 to be a soluble, monomeric enzyme. Mutations that alter aggregation, phospholipid dependence of catalysis, and membrane binding. *J. Biol. Chem.,* **2000***, 275*, 2545-2553.
[44] Williams, P.A.; Cosme, J.; Sridhar, V.; Johnson, E.F.; McRee, D.E. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell,* **2000***, 5*, 121-131.
[45] Shank-Retzlaff, M.L.; Raner, G.M.; Coon, M.J.; Sligar, S.G. Membrane topology of cytochrome P450 2B4 in Langmuir–Blodgett monolayers. *Arch. Biochem. Biophys.,* **1998***, 359*, 82-88.
[46] Black, S.D.; Martin, S.T.; Smith, C.A. Membrane topology of liver microsomal cytochrome P450 2B4 determined via monoclonal antibodies directed to the halt-transfer signal. *Biochemistry (N. Y.),* **1994***, 33*, 6945-6951.
[47] De Lemos-Chiarandini, C.; Frey, A.B.; Sabatini, D.D.; Kreibich, G. Determination of the membrane topology of the phenobarbital-inducible rat liver cytochrome P-450 isoenzyme PB-4 using site-specific antibodies. *J. Cell Biol.,* **1987***, 104*, 209-219.
[48] Ohta, Y.; Kawato, S.; Tagashira, H.; Takemori, S.; Kominami, S. Dynamic structures of adrenocortical cytochrome P-450 in proteoliposomes and microsomes: protein rotation study. *Biochemistry (N. Y.),* **1992***, 31*, 12680-12687.
[49] Bayburt, T.H.; Sligar, S.G. Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks. *Proc. Natl. Acad. Sci. U. S. A.,* **2002***, 99*, 6725-30.
[50] Berka, K.; Hendrychová, T.; Anzenbacher, P.; Otyepka, M. Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site. *J. Phys. Chem. A,* **2011***, 115*, 11248-11255.
[51] Cojocaru, V.; Balali-Mood, K.; Sansom, M.S.; Wade, R.C. Structure and dynamics of the membrane-bound cytochrome P450 2C9. *PLoS Comput. Biol.,* **2011***, 7*, e1002152.
[52] Yamamoto, K.; Gildenberg, M.; Ahuja, S.; Im, S.; Pearcy, P.; Waskell, L.; Ramamoorthy, A. Probing the transmembrane structure and topology of microsomal cytochrome-p450 by solid-state NMR on temperature-resistant bicelles. *Sci. Rep.,* **2013***, 3*.
[53] Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.,* **1992***, 267*, 83-90.
[54] Ekroos, M.; Sjogren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. U. S. A.,* **2006***, 103*, 13682-13687.
[55] Schoch, G.A.; Yano, J.K.; Sansen, S.; Dansette, P.M.; Stout, C.D.; Johnson, E.F. Determinants of cytochrome P450 2C8 substrate binding: structures of complexes with montelukast, troglitazone, felodipine, and 9-cis-retinoic acid. *J. Biol. Chem.,* **2008***, 283*, 17227-17237.
[56] Shah, M.B.; Wilderman, P.R.; Pascual, J.; Zhang, Q.; Stout, C.D.; Halpert, J.R. Conformational adaptation of human cytochrome P450 2B6 and rabbit cytochrome P450 2B4 revealed upon binding multiple amlodipine molecules. *Biochemistry (N. Y.),* **2012***, 51*, 7225-7238.
[57] Sevrioukova, I.F.; Poulos, T.L. Dissecting Cytochrome P450 3A4–Ligand Interactions Using Ritonavir Analogues. *Biochemistry (N. Y.),* **2013***, 52*, 4474-4481.
[58] 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*, 683-686.
[59] Williams, P.A.; Cosme, J.; Ward, A.; Angove, H.C.; Vinković, D.M.; Jhoti, H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature,* **2003***, 424*, 464-468.

[60] Davydov, D.R.; Halpert, J.R. Allosteric P450 mechanisms: multiple binding sites, multiple conformers or both? *Expert Opin. Drug Metab. Toxicol.,* **2008***, 4*, 1523-1535.
[61] Korzekwa, K.; Krishnamachary, N.; Shou, M.; Ogai, A.; Parise, R.; Rettie, A.; Gonzalez, F.; Tracy, T. Evaluation of atypical cytochrome P450 kinetics with two-substrate models: evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites. *Biochemistry (N. Y.),* **1998***, 37*, 4137-4147.
[62] Atkins, W.M.; Wang, R.W.; Lu, A.Y. Allosteric behavior in cytochrome P450-dependent in vitro drug-drug interactions: A prospective based on conformational dynamics. *Chem. Res. Toxicol.,* **2001***, 14*, 338-347.
[63] Atkins, W.M. Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annu. Rev. Pharmacol. Toxicol.,* **2005***, 45*, 291-310.
[64] Gay, S.C.; Roberts, A.G.; Maekawa, K.; Talakad, J.C.; Hong, W.; Zhang, Q.; Stout, C.D.; Halpert, J.R. Structures of cytochrome P450 2B4 complexed with the antiplatelet drugs ticlopidine and clopidogrel. *Biochemistry (N. Y.),* **2010***, 49*, 8709-8720.
[65] Lee, Y.T.; Glazer, E.C.; Wilson, R.F.; Stout, D.D.; Goodin, D.B. Three clusters of conformational states in p450cam reveal a multistep pathway for closing of the substrate access channel. *Biochemistry,* **2011***, 50*, 693-703.
[66] Cojocaru, V.; Winn, P.J.; Wade, R.C. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta,* **2007***, 1770*, 390-401.
[67] Bakan, A.; Meireles, L.M.; Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics,* **2011***, 27*, 1575-7.
[68] Lampe, J.N.; Brandman, R.; Sivaramakrishnan, S.; de Montellano, P.R. Two-dimensional NMR and all-atom molecular dynamics of cytochrome P450 CYP119 reveal hidden conformational substates. *J. Biol. Chem.,* **2010***, 285*, 9594-9603.
[69] Lampe, J.N.; Floor, S.N.; Gross, J.D.; Nishida, C.R.; Jiang, Y.; Trnka, M.J.; Ortiz de Montellano, Paul R Ligand-induced conformational heterogeneity of cytochrome P450 CYP119 identified by 2D NMR spectroscopy with the unnatural amino acid 13C-p-methoxyphenylalanine. *J. Am. Chem. Soc.,* **2008***, 130*, 16168-16169.
[70] Ross Wilderman, P.; R Halpert, J. Plasticity of CYP2B Enzymes: Structural and Solution Biophysical Methods. *Curr. Drug Metab.,* **2012***, 13*, 167-176.
[71] Pochapsky, T.C.; Kazanis, S.; Dang, M. Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxid. Redox Signaling,* **2010***, 13*, 1273-1296.
[72] Wilderman, P.R.; Shah, M.B.; Liu, T.; Li, S.; Hsu, S.; Roberts, A.G.; Goodlett, D.R.; Zhang, Q.; Woods, V.L.; Stout, C.D.; Halpert, J.R. Plasticity of Cytochrome P450 2B4 as Investigated by Hydrogen-Deuterium Exchange Mass Spectrometry and X-ray Crystallography. *J. Biol. Chem.,* **2010***, 285*, 38602-38611.
[73] Eyles, S.J.; Kaltashov, I.A. Methods to study protein dynamics and folding by mass spectrometry. *Methods,* **2004***, 34*, 88-99.
[74] Hui Bon Hoa, G.; McLean, M.A.; Sligar, S.G. High pressure, a tool for exploring heme protein active sites. *Biochim. Biophys. Acta,* **2002***, 1595*, 297-308.
[75] Anzenbacherová, E.; Bec, N.; Anzenbacher, P.; Hudeček, J.; Souček, P.; Jung, C.; Munro, A.W.; Lange, R. Flexibility and stability of the structure of cytochromes P450 3A4 and BM-3. *Eur. J. Biochem.,* **2000***, 267*, 2916-2920.
[76] Hendrychová, T.; Anzenbacherová, E.; Hudeček, J.; Skopalík, J.; Lange, R.; Hildebrandt, P.; Otyepka, M.; Anzenbacher, P. Flexibility of human cytochrome P450 enzymes: Molecular dynamics and spectroscopy reveal important function-related variations. *Biochim. Biophys. Acta,* **2011***, 1814*, 58-68.
[77] Davydov, D.R.; Baas, B.J.; Sligar, S.G.; Halpert, J.R. Allosteric mechanisms in cytochrome P450 3A4 studied by high-pressure spectroscopy: pivotal role of substrate-induced changes

in the accessibility and degree of hydration of the heme pocket. *Biochemistry,* **2007***, 46*, 7852-64.
[78] Mak, P.J.; Denisov, I.G.; Grinkova, Y.V.; Sligar, S.G.; Kincaid, J.R. Defining CYP3A4 Structural Responses to Substrate Binding. Raman Spectroscopic Studies of a Nanodisc-Incorporated Mammalian Cytochrome P450. *J. Am. Chem. Soc.,* **2011** *, 133,* 1357-1366.
[79] Hudeček, J.; Hodek, P.; Anzenbacherová, E.; Anzenbacher, P. Structural analysis of cytochromes P450 shows differences in flexibility of heme 2-and 4-vinyls. *Biochim. Biophys. Acta,* **2007***, 1770*, 413-419.
[80] Skopalík, 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*, 8165-8173.
[81] Lüdemann, S. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J. Mol. Biol.,* **2000***, 303*, 813-830.
[82] Lüdemann, S.K.; Lounnas, V.; Wade, R.C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.,* **2000***, 303*, 797-811.
[83] Oda, A.; Yamaotsu, N.; Hirono, S. New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J. Comput. Chem.,* **2005***, 26*, 818-826.
[84] Autenrieth, F.; Tajkhorshid, E.; Baudry, J.; Luthey-Schulten, Z. Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Comput. Chem.,* **2004***, 25*, 1613-1622.
[85] Favia, A.D.; Cavalli, A.; Masetti, M.; Carotti, A.; Recanatini, M. Three-dimensional model of the human aromatase enzyme and density functional parameterization of the iron-containing protoporphyrin IX for a molecular dynamics study of heme-cysteinato cytochromes. *Proteins,* **2006***, 62*, 1074-1087.
[86] Harris, D.L.; Loew, G.H. A Role for Thr 252 in Cytochrome P450cam Oxygen Activation. *J. Am. Chem. Soc.,* **1994***, 116*, 11671-11674.
[87] Hendrychova, T.; Berka, K.; Navratilova, V.; Anzenbacher, P.; Otyepka, M. Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. *Curr. Drug Metab.,* **2012***, 13*, 177.
[88] Sevrioukova, I.F.; Poulos, T.L. Understanding the mechanism of cytochrome P450 3A4: recent advances and remaining problems. *Dalton Trans.,* **2013***, 42*, 3116-3126.
[89] Zharkova, M.S.; Sobolev, B.N.; Veselovsky, A.V.; Archakov, A.I. Prediction of amino acid residues participated in substrate recognition by cytochrome P450 subfamilies with broad substrate specificity. *J. Mol. Recogn.,* **2013***, 26*, 86-91.
[90] Sevrioukova, I.F.; Li, H.; Zhang, H.; Peterson, J.A.; Poulos, T.L. Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc. Natl. Acad. Sci. U. S. A.,* **1999***, 96*, 1863-1868.
[91] Wang, M.; Roberts, D.L.; Paschke, R.; Shea, T.M.; Masters, B.S.; Kim, J.J. Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc. Natl. Acad. Sci. U. S. A.,* **1997***, 94*, 8411-6.
[92] Miwa, G. T., West, S. B., Huang, M. T., & Lu, A. Y. Studies on the association of cytochrome P-450 and NADPH-cytochrome c reductase during catalysis in a reconstituted hydroxylating system. *J. Biol. Chem.,* **1979***, 254*, 5695-5700.
[93] Bernhardt, R.; Kraft, R.; Otto, A.; Ruckpaul, K. Electrostatic interactions between cytochrome P-450 LM2 and NADPH-cytochrome P-450 reductase. *Biomed. Biochim. Acta,* **1988***, 47*, 581-92.

[94] Zhao, Q.; Modi, S.; Smith, G.; Paine, M.; McDonagh, P.D.; Wolf, C.R.; Tew, D.; Lian, L.Y.; Roberts, G.C.; Driessen, H.P. Crystal structure of the FMN-binding domain of human cytochrome P450 reductase at 1.93 A resolution. *Protein Sci.,* **1999***, 8*, 298-306.
[95] Shen, A.L.; Kasper, C.B. Role of acidic residues in the interaction of NADPH-cytochrome P450 oxidoreductase with cytochrome P450 and cytochrome c. *J. Biol. Chem.,* **1995***, 270*, 27475-80.
[96] Davydov, D.R.; Kariakin, A.A.; Petushkova, N.A.; Peterson, J.A. Association of cytochromes P450 with their reductases: opposite sign of the electrostatic interactions in P450BM-3 as compared with the microsomal 2B4 system. *Biochemistry,* **2000***, 39*, 6489-97.
[97] Bridges, A.; Gruenke, L.; Chang, Y.T.; Vakser, I.A.; Loew, G.; Waskell, L. Identification of the binding site on cytochrome P450 2B4 for cytochrome b5 and cytochrome P450 reductase. *J. Biol. Chem.,* **1998***, 273*, 17036-49.
[98] Zawaira, A.; Gallotta, M.; Beeton-Kempen, N.; Coulson, L.; Marais, P.; Kuttel, M.; Blackburn, J. Exhaustive computational search of ionic-charge clusters that mediate interactions between mammalian cytochrome P450 (CYP) and P450-oxidoreductase (POR) proteins. *Comput. Biol. Chem.,* **2010***, 34*, 42-52.
[99] Sevrioukova, I.F.; Hazzard, J.T.; Tollin, G.; Poulos, T.L. The FMN to heme electron transfer in cytochrome P450BM-3. Effect of chemical modification of cysteines engineered at the FMN-heme domain interaction site. *J. Biol. Chem.,* **1999***, 274*, 36097-36106.
[100] Voznesensky, A.I.; Schenkman, J.B. The cytochrome P450 2B4-NADPH cytochrome P450 reductase electron transfer complex is not formed by charge-pairing. *J. Biol. Chem.,* **1992***, 267*, 14669-76.
[101] Johnson, E.F.; Muller-Eberhard, U. Resolution of two forms of cytochrome P-450 from liver microsomes of rabbit treated with 2,3,7,8-tetrachlorodibenzo-p-dioxin. *J. Biol. Chem.,* **1977***, 252*, 2839-2845.
[102] Estabrook, R.W.; Shet, M.S.; Fisher, C.W.; Jenkins, C.M.; Waterman, M.R. The interaction of NADPH-P450 reductase with P450: an electrochemical study of the role of the flavin mononucleotide-binding domain. *Arch. Biochem. Biophys.,* **1996***, 333*, 308-315.
[103] Muller-Enoch, D.; Churchill, P.; Fleischer, S.; Guengerich, F.P. Interaction of liver microsomal cytochrome P-450 and NADPH-cytochrome P-450 reductase in the presence and absence of lipid. *J. Biol. Chem.,* **1984***, 259*, 8174-8182.
[104] Ingelman-Sundberg, M. Phospholipids and detergents as effectors in the liver microsomal hydroxylase system. *Biochim. Biophys. Acta,* **1977***, 488*, 225-234.
[105] Ingelman-Sundberg, M.; Haaparanta, T.; Rydstroem, J. Membrane charge as effector of cytochrome P-450LM2 catalyzed reactions in reconstituted liposomes. *Biochemistry (N. Y),* **1981***, 20*, 4100-4106.
[106] Ingelman-Sunberg, M.; Blanck, J.; Smettan, G.; Ruckpaul, K. Reduction of Cytochrome P-450 LM2 by NADPH in Reconstituted Phospholipid Vesicles is Dependent on Membrane Charge. *Eur. J. Biochem.,* **1983***, 134*, 157-162.
[107] Causey, K.M.; Eyer, C.S.; Backes, W.L. Dual role of phospholipid in the reconstitution of cytochrome P-450 LM2-dependent activities. *Mol. Pharmacol.,* **1990***, 38*, 134-142.
[108] Lu, A.Y.H.; Junk, K.; Coon, M.J. Resolution of the cytochrome P-450-containing omega-hydroxylation system of liver microsomes into three components. *The Journal of Biological Chemistry,* **1969***, 244*, 3714-3721.
[109] Miwa, T.; Lu, A.Y. The association of cytochrome P-450 and NADPH-cytochrome P-450 reductase in phospholipid membranes. *Arch. Biochem. Biophys.,* **1984***, 234*, 161-166.
[110] Estabrook, R.W.; Franklin, M.R.; Cohen, B.; Shigamatzu, A.; Hildebrant, A.G. Bicohemical and genetic factors influencing drug metabolism. Influence of hepatic

microsomal mixed function oxidation reactions on cellular metabolic control. *Metabolism,* **1971***, 20*, 187-199.
[111] Backes, W.L.; Kelley, R.W. Organization of multiple cytochrome P450s with NADPH-cytochrome P450 reductase in membranes. *Pharmacol. Ther.,* **2003***, 98*, 221-233.
[112] Reed, J.R.; Backes, W.L. Formation of P450· P450 complexes and their effect on P450 function. *Pharmacol. Ther.,* **2012***, 133*, 299-310.
[113] Peterson, J.A.; Ebel, R.E.; O'Keeffe, D.H.; Matsubara, T.; Estabrook, R.W. Temperature dependence of cytochrome P-450 reduction. A model for NADPH-cytochrome P-450 reductase:cytochrome P-450 interaction. *J. Biol. Chem.,* **1976***, 251*, 4010-4016.
[114] Kawato, S.; Gut, J.; Cherry, R.J.; Winterhalter, K.H.; Richter, C. Rotation of cytochrome P-450. I. Investigations of protein-protein interactions of cytochrome P-450 in phospholipid vesicles and liver microsomes. *J. Biol. Chem.,* **1982***, 257*, 7023-7029.
[115] Gut, J.; Richter, C.; Cherry, R.J.; Winterhalter, K.H.; Kawato, S. Rotation of cytochrome P-450. II. Specific interactions of cytochrome P-450 with NADPH-cytochrome P-450 reductase in phospholipid vesicles. *J. Biol. Chem.,* **1982***, 257*, 7030-7036.
[116] Kawato, S.; Ashikawa, I.; Iwase, T.; Hara, E. Drug-induction decreases the mobility of cytochrome P-450 in rat liver microsomes: protein rotation study. *J. Biochem.,* **1991***, 109*, 587-593.
[117] Gut, J.; Richter, C.; Cherry, R.J.; Winterhalter, K.H.; Kawato, S. Rotation of cytochrome P-450. Complex formation of cytochrome P-450 with NADPH-cytochrome P-450 reductase in liposomes demonstrated by combining protein rotation with antibody-induced cross-linking. *J. Biol. Chem.,* **1983***, 258*, 8588-8594.
[118] Dean, W.L.; Gray, R.D. Relationship between state of aggregation and catalytic activity for cytochrome P-450LM2 and NADPH-cytochrome P-450 reductase. *J. Biol. Chem.,* **1982***, 257*, 14679-14685.
[119] Myasoedova, K.; Tsuprun, V. Cytochrome P-450: hexameric structure of the purified LM4 form. *FEBS Lett.,* **1993***, 325*, 251-254.
[120] Myasoedova, K.N.; Berndt, P. Immobilized cytochrome P-450 LM2: Dissociation and reassociation of oligomers. *FEBS Lett.,* **1990***, 270*, 177-180.
[121] Tsuprun, V.; Myasoedova, K.; Berndt, P.; Sograf, O.; Orlova, E.; Chernyak, V.Y.; Archakov, A.; Skulachev, V. Quaternary structure of the liver microsomal cytochrome P-450. *FEBS Lett.,* **1986***, 205*, 35-40.
[122] Wagner, S.L.; Dean, W.L.; Gray, R.D. Zwitterionic detergent mediated interaction of purified cytochrome P-450LM4 from 5, 6-benzoflavone-treated rabbits with NADPH-cytochrome P-450 reductase. *Biochemistry (N. Y.),* **1987***, 26*, 2343-2348.
[123] Sevrukova, I.; Kanaeva, I.; Koen, Y.; Samenkova, N.; Bachmanova, G.; Archakov, A. Catalytic activity of cytochrome P4501A2 in reconstituted system with Emulgen 913. *Arch. Biochem. Biophys.,* **1994***, 311*, 133-143.
[124] Davydov, D.R. Conformational heterogeneity of cytochrome P450 3A4 revealed by high pressure spectroscopy. *Biochem. Biophys. Res. Commun.,* **2003***, 312*, 121-130.
[125] Fernando, H.; Davydov, D.; Chin, C.; Halpert, J. Role of subunit interactions in P450 oligomers in the loss of homotropic cooperativity in the cytochrome P450 3A4 mutant L211F/D214E/F304W. *Arch. Biochem. Biophys.,* **2007***, 460*, 129-140.
[126] Davydov, D.R.; Fernando, H.; Baas, B.J.; Sligar, S.G.; Halpert, J.R. Kinetics of dithionite-dependent reduction of cytochrome P450 3A4: heterogeneity of the enzyme caused by its oligomerization. *Biochemistry,* **2005***, 44*, 13902-13913.
[127] Koley, A.P.; Buters, J.T.; Robinson, R.C.; Markowitz, A.; Friedman, F.K. CO binding kinetics of human cytochrome P450 3A4. Specific interaction of substrates with kinetically distinguishable conformers. *J. Biol. Chem.,* **1995***, 270*, 5014-5018.

[128] Koley, A.P.; Robinson, R.C.; Friedman, F.K. Cytochrome P450 conformation and substrate interactions as probed by CO binding kinetics. *Biochimie,* **1996***, 78*, 706-713.
[129] West, S.B.; Lu, A.Y. Reconstituted liver microsomal enzyme system that hydroxylates drugs, other foreign compounds and endogenous substrates: V. Competition between cytochromes P-450 and P-448 for reductase in 3, 4-benzpyrene hydroxylation. *Arch. Biochem. Biophys.,* **1972***, 153*, 298-303.
[130] Li, D.N.; Pritchard, M.P.; Hanlon, S.P.; Burchell, B.; Wolf, C.R.; Friedberg, T. Competition between cytochrome P-450 isozymes for NADPH-cytochrome P-450 oxidoreductase affects drug metabolism. *J. Pharmacol. Exp. Ther.,* **1999***, 289*, 661-667.
[131] Tan, Y.; Patten, C.J.; Smith, T.; Yang, C.S. Competitive interactions between cytochromes P450 2A6 and 2E1 for NADPH-cytochrome P450 oxidoreductase in the microsomal membranes produced by a baculovirus expression system. *Arch. Biochem. Biophys.,* **1997***, 342*, 82-91.
[132] Backes, W.L.; Batie, C.J.; Cawley, G.F. Interactions among P450 enzymes when combined in reconstituted systems: formation of a 2B4-1A2 complex with a high affinity for NADPH-cytochrome P450 reductase. *Biochemistry (N. Y.),* **1998***, 37*, 12852-12859.
[133] Cawley, G.F.; Zhang, S.; Kelley, R.W.; Backes, W.L. Evidence supporting the interaction of CYP2B4 and CYP1A2 in microsomal preparations. *Drug Metab. Dispos.,* **2001***, 29*, 1529-1534.
[134] Yamazaki, H.; Gillam, E.; Dong, M.; Johnson, W.; Guengerich, F.; Shimada, T. Reconstitution of Recombinant Cytochrome P450 2C10 (2C9) and Comparison with Cytochrome P450 3A4 and Other Forms: Effects of Cytochrome P450-P450 and Cytochrome P450-b 5 Interactions. *Arch. Biochem. Biophys.,* **1997***, 342*, 329-337.
[135] Hazai, E.; Kupfer, D. Interactions between CYP2C9 and CYP2C19 in reconstituted binary systems influence their catalytic activity: possible rationale for the inability of CYP2C19 to catalyze methoxychlor demethylation in human liver microsomes. *Drug Metab. Dispos.,* **2005***, 33*, 157-164.
[136] Subramanian, M.; Low, M.; Locuson, C.W.; Tracy, T.S. CYP2D6-CYP2C9 protein-protein interactions and isoform-selective effects on substrate binding and catalysis. *Drug Metab. Dispos.,* **2009***, 37*, 1682-1689.
[137] Reed, J.R.; Eyer, M.; Backes, W.L. Functional interactions between cytochromes P450 1A2 and 2B4 require both enzymes to reside in the same phospholipid vesicle: evidence for physical complex formation. *J. Biol. Chem.,* **2010***, 285*, 8942-8952.
[138] Kelley, R.W.; Cheng, D.; Backes, W.L. Heteromeric complex formation between CYP2E1 and CYP1A2: evidence for the involvement of electrostatic interactions. *Biochemistry (N. Y.),* **2006***, 45*, 15807-15816.
[139] Subramanian, M.; Zhang, H.; Tracy, T.S. CYP2C9-CYP3A4 protein-protein interactions in a reconstituted expressed enzyme system. *FASEB J.,* **2008***, 22*, 919.7.
[140] Denison, M.S.; Nagy, S.R. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals*. *Annu. Rev. Pharmacol. Toxicol.,* **2003***, 43*, 309-334.
[141] Mizuno, N.; Niwa, T.; Yotsumoto, Y.; Sugiyama, Y. Impact of drug transporter studies on drug discovery and development. *Pharmacol. Rev.,* **2003***, 55*, 425-461.
[142] Tanigawara, Y. Role of P-glycoprotein in drug disposition. *Ther. Drug Monit.,* **2000***, 22*, 137-140.
[143] Christians, U. Transport proteins and intestinal metabolism: P-glycoprotein and cytochrome P4503A. *Ther. Drug Monit.,* **2004***, 26*, 104-106.
[144] Gibson, G.G.; Skett, P. *Introduction to drug metabolism*, 3rd ed.; Nelson Thornes: Cheltenham, UK, **2001**.

[145] Váradi, A.; Sarkadi, B. Multidrug resistance-associated proteins: Export pumps for conjugates with glutathione, glucuronate or sulfate. *Biofactors,* **2003***, 17*, 103-114.
[146] Jancova, P.; Anzenbacher, P.; Anzenbacherova, E. Phase II drug metabolizing enzymes. *Biomedical Papers,* **2010***, 154*, 103-116.
[147] 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*, 103-141.
[148] Pavek, P.; Dvorak, Z. Xenobiotic-induced transcriptional regulation of xenobiotic metabolizing enzymes of the cytochrome P450 superfamily in human extrahepatic tissues. *Curr. Drug Metab.,* **2008***, 9*, 129-143.
[149] Ingelman-Sundberg, M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol. Sci.,* **2004***, 25*, 193-200.
[150] Johansson, I.; Ingelman-Sundberg, M. Genetic polymorphism and toxicology--with emphasis on cytochrome p450. *Toxicol. Sci.,* **2011***, 120*, 1-13.
[151] Ding, X.; Kaminsky, L.S. Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. *Annu. Rev. Pharmacol. Toxicol.,* **2003***, 43*, 149-173.
[152] Bieche, I.; Narjoz, C.; Asselah, T.; Vacher, S.; Marcellin, P.; Lidereau, R.; Beaune, P.; de Waziers, I. Reverse transcriptase-PCR quantification of mRNA levels from cytochrome (CYP)1, CYP2 and CYP3 families in 22 different human tissues. *Pharmacogenet. Genomics,* **2007***, 17*, 731-742.
[153] Shimada, T.; Yun, C.H.; Yamazaki, H.; Gautier, J.C.; Beaune, P.H.; Guengerich, F.P. Characterization of human lung microsomal cytochrome P-450 1A1 and its role in the oxidation of chemical carcinogens. *Mol. Pharmacol.,* **1992***, 41*, 856-864.
[154] Nishimura, M.; Yaguti, H.; Yoshitsugu, H.; Naito, S.; Satoh, T. Tissue distribution of mRNA expression of human cytochrome P450 isoforms assessed by high-sensitivity real-time reverse transcription PCR. *Yakugaku Zasshi,* **2003***, 123*, 369-375.
[155] Stiborová, M.; Martínek, V.; Rýdlová, H.; Koblas, T.; Hodek, P. Expression of cytochrome P450 1A1 and its contribution to oxidation of a potential human carcinogen 1-phenylazo-2-naphthol (Sudan I) in human livers. *Cancer Lett.,* **2005***, 220*, 145-154.
[156] Chang, T.K.; Chen, J.; Pillay, V.; Ho, J.Y.; Bandiera, S.M. Real-time polymerase chain reaction analysis of CYP1B1 gene expression in human liver. *Toxicol. Sci.,* **2003***, 71*, 11-19.
[157] Nebert, D.W.; Dalton, T.P.; Okey, A.B.; Gonzalez, F.J. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J. Biol. Chem.,* **2004***, 279*, 23847-23850.
[158] Nebert, D.W.; Dalton, T.P. The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. *Nat. Rev. Cancer,* **2006***, 6*, 947-960.
[159] Zhou, S.; Wang, B.; Yang, L.; Liu, J. Structure, function, regulation and polymorphism and the clinical significance of human cytochrome P450 1A2. *Drug Metab. Rev.,* **2010***, 42*, 268-354.
[160] Li, J.K.; He, F.; Bi, H.C.; Zuo, Z.; Liu, B.D.; Luo, H.B.; Huang, M. Inhibition of human cytochrome P-450 CYP1A2 by flavonoids: a quantitative structure-activity relationship study. *Yao Xue Xue Bao,* **2008***, 43*, 1198-1204.
[161] Hiemke, C.; Härtter, S. Pharmacokinetics of selective serotonin reuptake inhibitors. *Pharmacol. Ther.,* **2000***, 85*, 11-28.
[162] Shimada, T.; Yamazaki, H.; Foroozesh, M.; Hopkins, N.E.; Alworth, W.L.; Guengerich, F.P. Selectivity of polycyclic inhibitors for human cytochrome P450s 1A1, 1A2, and 1B1. *Chem. Res. Toxicol.,* **1998***, 11*, 1048-1056.

[163] Sansen, S.; Yano, J.K.; Reynald, R.L.; Schoch, G.A.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *J. Biol. Chem.,* **2007***, 282*, 14348-14355.
[164] Wang, A.; Savas, U.; Stout, C.D.; Johnson, E.F. Structural characterization of the complex between alpha-naphthoflavone and human cytochrome P450 1B1. *J. Biol. Chem.,* **2011***, 286*, 5736-5743.
[165] Walsh, A.A.; Szklarz, G.D.; Scott, E.E. Human cytochrome P450 1A1 structure and utility in understanding drug and xenobiotic metabolism. *J. Biol. Chem.,* **2013***, 288*, 12932-12943.
[166] Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F.P. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.,* **1994***, 270*, 414-423.
[167] Haberl, M.; Anwald, B.; Klein, K.; Weil, R.; Fuss, C.; Gepdiremen, A.; Zanger, U.M.; Meyer, U.A.; Wojnowski, L. Three haplotypes associated with CYP2A6 phenotypes in Caucasians. *Pharmacogenet. Genomics,* **2005***, 15*, 609-624.
[168] Di, Y.M.; Chow, V.D.; Yang, L.; Zhou, S. Structure, function, regulation and polymorphism of human cytochrome P450 2A6. *Curr. Drug Metab.,* **2009***, 10*, 754-780.
[169] Yano, J.K.; Hsu, M.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat. Struct. Mol. Biol.,* **2005***, 12*, 822-823.
[170] Abu-Bakar, A.; Arthur, D.M.; Wikman, A.S.; Rahnasto, M.; Juvonen, R.O.; Vepsalainen, J.; Raunio, H.; Ng, J.C.; Lang, M.A. Metabolism of bilirubin by human cytochrome P450 2A6. *Toxicol. Appl. Pharmacol.,* **2012***, 261*, 50-58.
[171] Komatsu, T.; Yamazaki, H.; Shimada, N.; Nakajima, M.; Yokoi, T. Roles of cytochromes P450 1A2, 2A6, and 2C8 in 5-fluorouracil formation from tegafur, an anticancer prodrug, in human liver microsomes. *Drug Metab. Dispos.,* **2000***, 28*, 1457-1463.
[172] Murai, K.; Yamazaki, H.; Nakagawa, K.; Kawai, R.; Kamataki, T. Deactivation of anti-cancer drug letrozole to a carbinol metabolite by polymorphic cytochrome P450 2A6 in human liver microsomes. *Xenobiotica,* **2009***, 39*, 795-802.
[173] Sueyoshi, T.; Kawamoto, T.; Zelko, I.; Honkakoski, P.; Negishi, M. The repressed nuclear receptor CAR responds to phenobarbital in activating the human CYP2B6 gene. *J. Biol. Chem.,* **1999***, 274*, 6043-6046.
[174] Wang, H.; Faucette, S.; Sueyoshi, T.; Moore, R.; Ferguson, S.; Negishi, M.; LeCluyse, E.L. A novel distal enhancer module regulated by pregnane X receptor/constitutive androstane receptor is essential for the maximal induction of CYP2B6 gene expression. *J. Biol. Chem.,* **2003***, 278*, 14146-14152.
[175] Wilderman, P.R.; Shah, M.B.; Jang, H.; Stout, C.D.; Halpert, J.R. Structural and thermodynamic basis of (+)-α-pinene binding to human cytochrome P450 2B6. *J. Am. Chem. Soc.,* **2013***, 135*, 10433-10440.
[176] Shah, M.B.; Pascual, J.; Zhang, Q.; Stout, C.D.; Halpert, J.R. Structures of Cytochrome P450 2B6 Bound to 4-Benzylpyridine and 4-(4-Nitrobenzyl)pyridine: Insight into Inhibitor Binding and Rearrangement of Active Site Side Chains. *Mol. Pharmacol.,* **2011**.
[177] Gay, S.C.; Shah, M.B.; Talakad, J.C.; Maekawa, K.; Roberts, A.G.; Wilderman, P.R.; Sun, L.; Yang, J.Y.; Huelga, S.C.; Hong, W.; Zhang, Q.; Stout, C.D.; Halpert, J.R. Crystal structure of a cytochrome P450 2B6 genetic variant in complex with the inhibitor 4-(4-chlorophenyl) imidazole at 2.0-Å resolution. *Mol. Pharmacol.,* **2010***, 77*, 529-538.
[178] Zhao, Y.; Halpert, J.R. Structure–function analysis of cytochromes P450 2B. *Biochim. Biophys. Acta,* **2007***, 1770*, 402-412.

[179] Hodgson, E.; Rose, R.L. The importance of cytochrome P450 2B6 in the human metabolism of environmental chemicals. *Pharmacol. Ther.,* **2007***, 113*, 420-428.
[180] Zanger, U.M.; Klein, K. Pharmacogenetics of cytochrome P450 2B6 (CYP2B6): advances on polymorphisms, mechanisms, and clinical relevance. *Frontiers in genetics,* **2013***, 4*.
[181] Kreth, K.; Kovar, K.; Schwab, M.; Zanger, U.M. Identification of the human cytochromes P450 involved in the oxidative metabolism of "Ecstasy"-related designer drugs. *Biochem. Pharmacol.,* **2000***, 59*, 1563-1571.
[182] Edwards, R.J.; Adams, D.A.; Watts, P.S.; Davies, D.S.; Boobis, A.R. Development of a comprehensive panel of antibodies against the major xenobiotic metabolising forms of cytochrome P450 in humans. *Biochem. Pharmacol.,* **1998***, 56*, 377-387.
[183] Delozier, T.C.; Kissling, G.E.; Coulter, S.J.; Dai, D.; Foley, J.F.; Bradbury, J.A.; Murphy, E.; Steenbergen, C.; Zeldin, D.C.; Goldstein, J.A. Detection of human CYP2C8, CYP2C9, and CYP2J2 in cardiovascular tissues. *Drug Metab. Dispos.,* **2007***, 35*, 682-688.
[184] Thelen, K.; Dressman, J.B. Cytochrome P450-mediated metabolism in the human gut wall. *J. Pharm. Pharmacol.,* **2009***, 61*, 541-558.
[185] Paine, M.F.; Hart, H.L.; Ludington, S.S.; Haining, R.L.; Rettie, A.E.; Zeldin, D.C. The human intestinal cytochrome P450 "pie". *Drug Metab. Dispos.,* **2006***, 34*, 880-886.
[186] Chen, Y.; Goldstein, J.A. The transcriptional regulation of the human CYP2C genes. *Curr. Drug Metab.,* **2009***, 10*, 567-578.
[187] Reynald, R.; Sansen, S.; Stout, C.; Johnson, E. Structural characterization of human cytochrome P450 2C19: active site differences between P450s 2C8, 2C9, and 2C19. *The Journal of biological chemistry,* **2012***, 287*, 44581-44591.
[188] Dauchy, S.; Dutheil, F.; Weaver, R.J.; Chassoux, F.; Daumas-Duport, C.; Couraud, P.; Scherrmann, J.; De Waziers, I.; Declèves, X. ABC transporters, cytochromes P450 and their main transcription factors: expression at the human blood–brain barrier. *J. Neurochem.,* **2008***, 107*, 1518-1528.
[189] Ingelman-Sundberg, M.; Evans, W.E. Unravelling the functional genomics of the human CYP2D6 gene locus. *Pharmacogenet. Genomics,* **2001***, 11*, 553-554.
[190] Bock, K.W.; Schrenk, D.; Forster, A.; Griese, E.; Mörike, K.; Brockmeier, D.; Eichelbaum, M. The influence of environmental and genetic factors on CYP2D6, CYP1A2 and UDP-glucuronosyltransferases in man using sparteine, caffeine, and paracetamol as probes. *Pharmacogenet. Genomics,* **1994***, 4*, 209-218.
[191] Cairns, W.; Smith, C.A.; McLaren, A.W.; Wolf, C.R. Characterization of the human cytochrome P4502D6 promoter. A potential role for antagonistic interactions between members of the nuclear receptor family. *J. Biol. Chem.,* **1996***, 271*, 25269-25276.
[192] Caraco, Y.; Sheller, J.; Wood, A.J. Pharmacogenetic determinants of codeine induction by rifampin: the impact on codeine's respiratory, psychomotor and miotic effects. *J. Pharmacol. Exp. Ther.,* **1997***, 281*, 330-336.
[193] Hara, H.; Adachi, T. Contribution of hepatocyte nuclear factor-4 to down-regulation of CYP2D6 gene expression by nitric oxide. *Mol. Pharmacol.,* **2002***, 61*, 194-200.
[194] Wang, A.; Savas, U.; Hsu, M.; Stout, C.D.; Johnson, E.F. Crystal structure of human cytochrome P450 2D6 with prinomastat bound. *J. Biol. Chem.,* **2012***, 287*, 10834-10843.
[195] Rowland, P.; Blaney, F.E.; Smyth, M.G.; Jones, J.J.; Leydon, V.R.; Oxbrow, A.K.; Lewis, C.J.; Tennant, M.G.; Modi, S.; Eggleston, D.S.; Chenery, R.J.; Bridges, A.M. Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.,* **2006***, 281*, 7614-7622.
[196] Masuda, K.; Tamagake, K.; Okuda, Y.; Torigoe, F.; Tsuzuki, D.; Isobe, T.; Hichiya, H.; Hanioka, N.; Yamamoto, S.; Narimatsu, S. Change in enantioselectivity in bufuralol 1 ″-hydroxylation by the substitution of phenylalanine-120 by alanine in cytochrome P450 2D6. *Chirality,* **2005***, 17*, 37-43.

[197] Yu, A.; Idle, J.R.; Byrd, L.G.; Krausz, K.W.; Küpfer, A.; Gonzalez, F.J. Regeneration of serotonin from 5-methoxytryptamine by polymorphic human CYP2D6. *Pharmacogenet. Genomics,* **2003***, 13*, 173-181.
[198] Yu, A.; Idle, J.R.; Herraiz, T.; Küpfer, A.; Gonzalez, F.J. Screening for endogenous substrates reveals that CYP2D6 is a 5-methoxyindolethylamine O-demethylase. *Pharmacogenet. Genomics,* **2003***, 13*, 307-319.
[199] Frank, D.; Jaehde, U.; Fuhr, U. Evaluation of probe drugs and pharmacokinetic metrics for CYP2D6 phenotyping. *Eur. J. Clin. Pharmacol.,* **2007***, 63*, 321-333.
[200] Peter, R.; Boecker, R.; Beaune, P.H.; Iwasaki, M.; Guengerich, F.P.; Yang, C.S. Hydroxylation of chlorzoxazone as a specific probe for human liver cytochrome P-450IIE1. *Chem. Res. Toxicol.,* **1990***, 3*, 566-573.
[201] Nedelcheva, V.; Gut, I.; Souček, P.; Tichavská, B.; Týnkova, L.; Mráz, J.; Guengerich, F.P.; Ingelman-Sundberg, M. Metabolism of benzene in human liver microsomes: individual variations in relation to CYP2E1 expression. *Arch. Toxicol.,* **1999***, 73*, 33-40.
[202] Powell, H.; Kitteringham, N.; Pirmohamed, M.; Smith, D.; Park, B. Expression of cytochrome P4502E1 in human liver: assessment by mRNA, genotype and phenotype. *Pharmacogenet. Genomics,* **1998***, 8*, 411-421.
[203] Tompkins, L.M.; Wallace, A.D. Mechanisms of cytochrome P450 induction. *J. Biochem. Mol. Toxicol.,* **2007***, 21*, 176-181.
[204] Mohri, T.; Nakajima, M.; Fukami, T.; Takamiya, M.; Aoki, Y.; Yokoi, T. Human CYP2E1 is regulated by miR-378. *Biochem. Pharmacol.,* **2010***, 79*, 1045-1052.
[205] Porubsky, P.R.; Battaile, K.P.; Scott, E.E. Human cytochrome P450 2E1 structures with fatty acid analogs reveal a previously unobserved binding mode. *J. Biol. Chem.,* **2010***, 285*, 22282-90.
[206] Lu, Y.; Cederbaum, A.I. CYP2E1 and oxidative liver injury by alcohol. *Free Radic. Biol. Med.,* **2008***, 44*, 723-738.
[207] Bolt, H.M.; Roos, P.H.; Thier, R. The cytochrome P-450 isoenzyme CYP2E1 in the biological processing of industrial chemicals: consequences for occupational and environmental medicine. *Int. Arch. Occup. Environ. Health,* **2003***, 76*, 174-185.
[208] Caro, A.A.; Cederbaum, A.I. Oxidative stress, toxicology, and pharmacology of cyp2e1. *Annu. Rev. Pharmacol. Toxicol.,* **2004***, 44*, 27-42.
[209] Daly, A.K. Significance of the minor cytochrome P450 3A isoforms. *Clin. Pharmacokinet.,* **2006***, 45*, 13-31.
[210] Guengerich, F.P. Mechanism-based inactivation of human liver microsomal cytochrome P-450 IIIA4 by gestodene. *Chem. Res. Toxicol.,* **1990***, 3*, 363-371.
[211] Guengerich, F. Reactions and significance of cytochrome P-450 enzymes. *J. Biol. Chem.,* **1991***, 266*.
[212] Wrighton, S.A.; Brian, W.R.; Sari, M.A.; Iwasaki, M.; Guengerich, F.P.; Raucy, J.L.; Molowa, D.T.; Vandenbranden, M. Studies on the expression and metabolic capabilities of human liver cytochrome P450IIIA5 (HLp3). *Mol. Pharmacol.,* **1990***, 38*, 207-213.
[213] Yamazaki, H.; Inui, Y.; Wrighton, S.A.; Guengerich, F.P.; Shimada, T. Procarcinogen activation by cytochrome P450 3A4 and 3A5 expressed in Escherichia coli and by human liver microsomes. *Carcinogenesis,* **1995***, 16*, 2167-2170.
[214] Kitada, M.; Kamataki, T.; Itahashi, K.; Rikihisa, T.; Kanakubo, Y. P-450 HFLa, a form of cytochrome P-450 purified from human fetal livers, is the 16 alpha-hydroxylase of dehydroepiandrosterone 3-sulfate. *J. Biol. Chem.,* **1987***, 262*, 13534-13537.
[215] Wrighton, S.A.; Vandenbranden, M. Isolation and characterization of human fetal liver cytochrome P450HLp2: a third member of the P450III gene family. *Arch. Biochem. Biophys.,* **1989***, 268*, 144-151.

[216] Timsit, Y.E.; Negishi, M. CAR and PXR: the xenobiotic-sensing receptors. *Steroids,* **2007***, 72*, 231-246.
[217] Pascussi, J.; Gerbal-Chaloin, S.; Duret, C.; Daujat-Chavanieu, M.; Vilarem, M.; Maurel, P. The tangle of nuclear receptors that controls xenobiotic metabolism and transport: crosstalk and consequences. *Annu. Rev. Pharmacol. Toxicol.,* **2008***, 48*, 1-32.
[218] Gnerre, C.; Blättler, S.; Kaufmann, M.R.; Looser, R.; Meyer, U.A. Regulation of CYP3A4 by the bile acid receptor FXR: evidence for functional binding sites in the CYP3A4 gene. *Pharmacogenet. Genomics,* **2004***, 14*, 635-645.
[219] Duniec-Dmuchowski, Z.; Ellis, E.; Strom, S.C.; Kocarek, T.A. Regulation of CYP3A4 and CYP2B6 expression by liver X receptor agonists. *Biochem. Pharmacol.,* **2007***, 74*, 1535-1540.
[220] Matsubara, T.; Yoshinari, K.; Aoyama, K.; Sugawara, M.; Sekiya, Y.; Nagata, K.; Yamazoe, Y. Role of vitamin D receptor in the lithocholic acid-mediated CYP3A induction in vitro and in vivo. *Drug Metab. Dispos.,* **2008***, 36*, 2058-2063.
[221] Klein, K.; Thomas, M.; Winter, S.; Nussler, A.K.; Niemi, M.; Schwab, M.; Zanger, U.M. PPARA: a novel genetic determinant of CYP3A4 in vitro and in vivo. *Clin. Pharmacol. Ther.,* **2012***, 91*, 1044-1052.
[222] Rakhshandehroo, M.; Hooiveld, G.; Müller, M.; Kersten, S. Comparative analysis of gene regulation by the transcription factor PPARα between mouse and human. *PloS one,* **2009***, 4*, e6796.
[223] Yano, J.K.; Wester, M.R.; Schoch, G.A.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution. *J. Biol. Chem.,* **2004***, 279*, 38091-38094.
[224] Sevrioukova, I.F.; Poulos, T.L. Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. *Proc. Natl. Acad. Sci. U.S.A.,* **2010***, 107*, 18422-18427.
[225] Sevrioukova, I.F.; Poulos, T.L. Pyridine-Substituted Desoxyritonavir Is a More Potent Inhibitor of Cytochrome P450 3A4 than Ritonavir. *J. Med. Chem.,* **2013***, 56*, 3733-3741.
[226] Sevrioukova, I.F.; Poulos, T.L. Interaction of human cytochrome P4503A4 with ritonavir analogs. *Arch. Biochem. Biophys.,* **2012***, 520*, 108-116.
[227] Sevrioukova, I.F.; Poulos, T.L. Structural and mechanistic insights into the interaction of cytochrome P4503A4 with bromoergocryptine, a type I ligand. *J. Biol. Chem.,* **2012***, 287*, 3510-3517.
[228] Niwa, T.; Murayama, N.; Yamazaki, H. Heterotropic cooperativity in oxidation mediated by cytochrome P450. *Curr. Drug Metab.,* **2008***, 9*, 453-462.
[229] Foti, R.S.; Rock, D.A.; Wienkers, L.C.; Wahlstrom, J.L. Selection of alternative CYP3A4 probe substrates for clinical drug interaction studies using in vitro data and in vivo simulation. *Drug Metab. Dispos.,* **2010***, 38*, 981-987.
[230] Roberts, A.G.; Yang, J.; Halpert, J.R.; Nelson, S.D.; Thummel, K.T.; Atkins, W.M. The structural basis for homotropic and heterotropic cooperativity of midazolam metabolism by human cytochrome P450 3A4. *Biochemistry (N. Y.),* **2011***, 50*, 10804-10818.
[231] Huang, S.; Temple, R.; Throckmorton, D.; Lesko, L. Drug interaction studies: study design, data analysis, and implications for dosing and labeling. *Clin. Pharmacol. Ther.,* **2007***, 81*, 298-304.
[232] Monahan, B.P.; Ferguson, C.L.; Killeavy, E.S.; Lloyd, B.K.; Troy, J.; Cantilena, L.R. Torsades de pointes occurring in association with terfenadine use. *JAMA,* **1990***, 264*, 2788-2790.
[233] Honig, P.K.; Wortham, D.C.; Zamani, K.; Conner, D.P.; Mullin, J.C.; Cantilena, L.R. Terfenadine-ketoconazole interaction: pharmacokinetic and electrocardiographic consequences. *JAMA,* **1993***, 269*, 1513-1518.

[234] Honig, P.K.; Woosley, R.L.; Zamani, K.; Conner, D.P.; Cantilena, L.R. Changes in the pharmacokinetics and electrocardiographic pharmacodynamics of terfenadine with concomitant administration of erythromycin. *Clin. Pharmacol. Ther.,* **1992**, *52*, 231-238.
[235] Thompson, D.; Oster, G. Use of terfenadine and contraindicated drugs. *JAMA,* **1996**, *275*, 1339-1341.
[236] Markham, A.; Wagstaff, A.J. Fexofenadine. *Drugs,* **1998**, *55*, 269-274.
[237] Dooley, K.E.; Kim, P.S.; Williams, S.D.; Hafner, R. TB and HIV therapeutics: pharmacology research priorities. *AIDS research and treatment,* **2012**, *2012*.
[238] Niemi, M.; Backman, J.T.; Fromm, M.F.; Neuvonen, P.J.; Kivistö, K.T. Pharmacokinetic interactions with rifampicin. *Clin. Pharmacokinet.,* **2003**, *42*, 819-850.
[239] Swaminathan, S.; Padmapriyadarsini, C.; Venkatesan, P.; Narendran, G.; Ramesh Kumar, S.; Iliayas, S.; Menon, P.A.; Selvaraju, S.; Pooranagangadevi, N.P.; Bhavani, P.K.; Ponnuraja, C.; Dilip, M.; Ramachandran, R. Efficacy and safety of once-daily nevirapine- or efavirenz-based antiretroviral therapy in HIV-associated tuberculosis: a randomized clinical trial. *Clin. Infect. Dis.,* **2011**, *53*, 716-724.
[240] Boulle, A.; Van Cutsem, G.; Cohen, K.; Hilderbrand, K.; Mathee, S.; Abrahams, M.; Goemaere, E.; Coetzee, D.; Maartens, G. Outcomes of nevirapine-and efavirenz-based antiretroviral therapy when coadministered with rifampicin-based antitubercular therapy. *JAMA,* **2008**, *300*, 530-539.
[241] Manosuthi, W.; Sungkanuparph, S.; Tantanathip, P.; Lueangniyomkul, A.; Mankatitham, W.; Prasithsirskul, W.; Burapatarawong, S.; Thongyen, S.; Likanonsakul, S.; Thawornwa, U.; Prommool, V.; Ruxrungtham, K.; N2R Study Team A randomized trial comparing plasma drug concentrations and efficacies between 2 nonnucleoside reverse-transcriptase inhibitor-based regimens in HIV-infected patients receiving rifampicin: the N2R Study. *Clin. Infect. Dis.,* **2009**, *48*, 1752-1759.
[242] Nowack, R. Review Article: Cytochrome P450 enzyme, and transport protein mediated herb–drug interactions in renal transplant patients: Grapefruit juice, St John's Wort–and beyond!(Review Article). *Nephrology,* **2008**, *13*, 337-347.
[243] Mills, E.; Cooper, C.; Seely, D.; Kanfer, I. African herbal medicines in the treatment of HIV: Hypoxis and Sutherlandia. An overview of evidence and pharmacology. *Nutr J,* **2005**, *4*, 19.
[244] Huang, S.; Zhao, H.; Lee, J.; Reynolds, K.; Zhang, L.; Temple, R.; Lesko, L. Therapeutic protein–drug interactions and implications for drug development. *Clin. Pharmacol. Ther.,* **2010**, *87*, 497-503.
[245] Huang, S.; Temple, R. Is this the drug or dose for you?: impact and consideration of ethnic factors in global drug development, regulatory review, and clinical practice. *Clin. Pharmacol. Ther.,* **2008**, *84*, 287-295.
[246] Urquhart, B.L.; Tirona, R.G.; Kim, R.B. Nuclear Receptors and the Regulation of Drug-Metabolizing Enzymes and Drug Transporters: Implications for Interindividual Variability in Response to Drugs. *J. Clin. Pharmacol.,* **2007**, *47*, 566-578.
[247] Wood, A.J.; Evans, W.E.; McLeod, H.L. Pharmacogenomics—drug disposition, drug targets, and side effects. *N. Engl. J. Med.,* **2003**, *348*, 538-549.
[248] Nakajima, M.; Fukami, T.; Yamanaka, H.; Higashi, E.; Sakai, H.; Yoshida, R.; Kwon, J.; McLeod, H.L.; Yokoi, T. Comprehensive evaluation of variability in nicotine metabolism and CYP2A6 polymorphic alleles in four ethnic populations. *Clin. Pharmacol. Ther.,* **2006**, *80*, 282-297.
[249] Lamba, J.K.; Lin, Y.S.; Schuetz, E.G.; Thummel, K.E. Genetic contribution to variable human CYP3A-mediated metabolism. *Adv. Drug Deliv. Rev.,* **2012**, *64*, 256-269.

[250] Fowler, S.M.; Riley, R.J.; Pritchard, M.P.; Sutcliffe, M.J.; Friedberg, T.; Wolf, C.R. Amino acid 305 determines catalytic center accessibility in CYP3A4. *Biochemistry (N. Y.),* **2000***, 39*, 4406-4414.
[251] 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*, 92-101.
[252] Xue, L.; Wang, H.F.; Wang, Q.; Szklarz, G.D.; Domanski, T.L.; Halpert, J.R.; Correia, M.A. Influence of P450 3A4 SRS-2 residues on cooperativity and/or regioselectivity of aflatoxin B1 oxidation. *Chem. Res. Toxicol.,* **2001***, 14*, 483-491.
[253] Nakajima, M.; Yokoi, T.; Mizutani, M.; Kinoshita, M.; Funayama, M.; Kamataki, T. Genetic polymorphism in the 5'-flanking region of human CYP1A2 gene: effect on the CYP1A2 inducibility in humans. *J. Biochem.,* **1999***, 125*, 803-808.
[254] Sachse, C.; Brockmöller, J.; Bauer, S.; Roots, I. Functional significance of a C→ A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br. J. Clin. Pharmacol.,* **1999***, 47*, 445-449.
[255] Ghotbi, R.; Christensen, M.; Roh, H.; Ingelman-Sundberg, M.; Aklillu, E.; Bertilsson, L. Comparisons of CYP1A2 genetic polymorphisms, enzyme activity and the genotype-phenotype relationship in Swedes and Koreans. *Eur. J. Clin. Pharmacol.,* **2007***, 63*, 537-546.
[256] Djordjevic, N.; Ghotbi, R.; Jankovic, S.; Aklillu, E. Induction of CYP1A2 by heavy coffee consumption is associated with the CYP1A2− 163C> A polymorphism. *Eur. J. Clin. Pharmacol.,* **2010***, 66*, 697-703.
[257] Laika, B.; Leucht, S.; Heres, S.; Schneider, H.; Steimer, W. Pharmacogenetics and olanzapine treatment: CYP1A2*1F and serotonergic polymorphisms influence therapeutic outcome. *Pharmacogenomics J.,* **2010***, 10*, 20-29.
[258] Moonen, H.; Engels, L.; Kleinjans, J.; Kok, T.d. The CYP1A2-164A→ C polymorphism (CYP1A2* 1F) is associated with the risk for colorectal adenomas in humans. *Cancer Lett.,* **2005***, 229*, 25-31.
[259] Cornelis, M.C.; El-Sohemy, A.; Campos, H. Genetic polymorphism of CYP1A2 increases the risk of myocardial infarction. *J. Med. Genet.,* **2004***, 41*, 758-762.
[260] Wang, H.; Zhang, Z.; Han, S.; Lu, Y.; Feng, F.; Yuan, J. CYP1A2 rs762551 polymorphism contributes to cancer susceptibility: a meta-analysis from 19 case-control studies. *BMC Cancer,* **2012***, 12*, 528-2407-12-528.
[261] Malaiyandi, V.; Sellers, E.M.; Tyndale, R.F. Implications of CYP2A6 Genetic Variation for Smoking Behaviors and Nicotine Dependence. *Clin. Pharmacol. Ther.,* **2005***, 77*, 145-158.
[262] Mwenifumbo, J.C.; Tyndale, R.F. Genetic variability in CYP2A6 and the pharmacokinetics of nicotine. *Pharmacogenomics,* **2007***, 8*, 1385-1402.
[263] Kirchheiner, J.; Klein, C.; Meineke, I.; Sasse, J.; Zanger, U.M.; Mürdter, T.E.; Roots, I.; Brockmöller, J. Bupropion and 4-OH-bupropion pharmacokinetics in relation to genetic polymorphisms in CYP2B6. *Pharmacogenet Genomics,* **2003***, 13*, 619-626.
[264] Johnstone, E.; Benowitz, N.; Cargill, A.; Jacob, R.; Hinks, L.; Day, I.; Murphy, M.; Walton, R. Determinants of the rate of nicotine metabolism and effects on smoking behavior*. *Clin. Pharmacol. Ther.,* **2006***, 80*, 319-330.
[265] Rotger, M.; Tegude, H.; Colombo, S.; Cavassini, M.; Furrer, H.; Decosterd, L.; Blievernicht, J.; Saussele, T.; Günthard, H.; Schwab, M. Predictive value of known and novel alleles of CYP2B6 for efavirenz plasma concentrations in HIV-infected individuals. *Clin. Pharmacol. Ther.,* **2007***, 81*, 557-566.

[266] Lang, T.; Klein, K.; Fischer, J.; Nüssler, A.K.; Neuhaus, P.; Hofmann, U.; Eichelbaum, M.; Schwab, M.; Zanger, U.M. Extensive genetic polymorphism in the human CYP2B6 gene with impact on expression and function in human liver. *Pharmacogenet. Genomics,* **2001***, 11*, 399-415.
[267] King, J.; Aberg, J.A. Clinical impact of patient population differences and genomic variation in efavirenz therapy. *AIDS,* **2008***, 22*, 1709-1717.
[268] Telenti, A.; Zanger, U. Pharmacogenetics of anti-HIV drugs. *Annu. Rev. Pharmacol. Toxicol.,* **2008***, 48*, 227-256.
[269] Rakhmanina, N.Y.; van den Anker, John N Efavirenz in the therapy of HIV infection. *Expert opinion on drug metabolism & toxicology,* **2010***, 6*, 95-103.
[270] Crettol, S.; Déglon, J.; Besson, J.; Croquette-Krokkar, M.; Gothuey, I.; Hämmig, R.; Monnat, M.; Hüttemann, H.; Baumann, P.; Eap, C.B. Methadone enantiomer plasma levels, CYP2B6, CYP2C19, and CYP2C9 genotypes, and response to treatment. *Clin. Pharmacol. Ther.,* **2005***, 78*, 593-604.
[271] Eap, C.B.; Crettol, S.; Rougier, J.; Schläpfer, J.; Grilo, L.S.; Déglon, J.; Besson, J.; Croquette-Krokar, M.; Carrupt, P.; Abriel, H. Stereoselective block of hERG channel by (S)-methadone and QT interval prolongation in CYP2B6 slow metabolizers. *Clin. Pharmacol. Ther.,* **2007***, 81*, 719-728.
[272] Bunten, H.; Liang, W.; Pounder, D.; Seneviratne, C.; Osselton, M.D. CYP2B6 and OPRM1 gene variations predict methadone-related deaths. *Addict. Biol.,* **2011***, 16*, 142-144.
[273] Holstein, A.; Plaschke, A.; Ptak, M.; Egberts, E.; El-Din, J.; Brockmöller, J.; Kirchheiner, J. Association between CYP2C9 slow metabolizer genotypes and severe hypoglycaemia on medication with sulphonylurea hypoglycaemic agents. *Br. J. Clin. Pharmacol.,* **2005***, 60*, 103-106.
[274] Xu, H.; Murray, M.; McLachlan, A.J. Influence of genetic polymorphisms on the pharmacokinetics and pharmacodynamics of sulfonylurea drugs. *Curr. Drug Metab.,* **2009***, 10*, 643-658.
[275] Jonas, D.E.; McLeod, H.L. Genetic and clinical factors relating to warfarin dosing. *Trends Pharmacol. Sci.,* **2009***, 30*, 375-386.
[276] Teichert, M.; Eijgelsheim, M.; Rivadeneira, F.; Uitterlinden, A.G.; van Schaik, R.H.; Hofman, A.; De Smet, P.A.; van Gelder, T.; Visser, L.E.; Stricker, B.H. A genome-wide association study of acenocoumarol maintenance dosage. *Hum. Mol. Genet.,* **2009***, 18*, 3758-3768.
[277] Teichert, M.; Eijgelsheim, M.; Uitterlinden, A.G.; Buhre, P.N.; Hofman, A.; De Smet, P.A.; Visser, L.E.; Stricker, B.H. Dependency of phenprocoumon dosage on polymorphisms in the VKORC1, CYP2C9, and CYP4F2 genes. *Pharmacogenet. Genomics,* **2011***, 21*, 26-34.
[278] Manolopoulos, V.G.; Ragia, G.; Tavridou, A. Pharmacogenomics of oral antidiabetic medications: current data and pharmacoepigenomic perspective. *Pharmacogenomics,* **2011***, 12*, 1161-1191.
[279] Surendiran, A.; Pradhan, S.; Agrawal, A.; Subrahmanyam, D.; Rajan, S.; Anichavezhi, D.; Adithan, C. Influence of CYP2C9 gene polymorphisms on response to glibenclamide in type 2 diabetes mellitus patients. *Eur. J. Clin. Pharmacol.,* **2011***, 67*, 797-801.
[280] He, S.; Zhou, Z.; Li, X.; Zhou, S. Clinical drugs undergoing polymorphic metabolism by human cytochrome P450 2C9 and the implication in drug development. *Curr. Med. Chem.,* **2011***, 18*, 667-713.
[281] Furuta, T.; Shirai, N.; Kodaira, M.; Sugimoto, M.; Nogaki, A.; Kuriyama, S.; Iwaizumi, M.; Yamade, M.; Terakawa, I.; Ohashi, K. Pharmacogenomics-based tailored versus

standard therapeutic regimen for eradication of H. pylori. *Clin. Pharmacol. Ther.,* **2007***, 81*, 521-528.
[282] Furuta, T.; Shirai, N.; Sugimoto, M.; Ohashi, K.; Ishizaki, T. Pharmacogenomics of proton pump inhibitors. *Pharmacogenomics,* **2004***, 5*, 181-202.
[283] Klotz, U. Clinical impact of CYP2C19 polymorphism on the action of proton pump inhibitors: a review of a special problem. *Int. J. Clin. Pharmacol. Ther.,* **2006***, 44*, 297-302.
[284] Klotz, U.; Schwab, M.; Treiber, G. CYP2C19 polymorphism and proton pump inhibitors. *Pharmacol. Toxicol.,* **2004***, 95*, 2-8.
[285] Nielsen, K.K.; Brøsen, K.; Hansen, M.J.; Gram, L.F. Single-dose kinetics of clomipramine: relationship to the sparteine and S-mephenytoin oxidation polymorphisms. *Clin. Pharmacol. Ther.,* **1994***, 55*, 518-527.
[286] Steimer, W.; Zopf, K.; von Amelunxen, S.; Pfeiffer, H.; Bachofer, J.; Popp, J.; Messner, B.; Kissling, W.; Leucht, S. Amitriptyline or not, that is the question: pharmacogenetic testing of CYP2D6 and CYP2C19 identifies patients with low or high risk for side effects in amitriptyline therapy. *Clin. Chem.,* **2005***, 51*, 376-385.
[287] Yu, K.; Yim, D.; Cho, J.; Park, S.S.; Park, J.Y.; Lee, K.; Jang, I.; Yi, S.; Bae, K.; Shin, S. Effect of omeprazole on the pharmacokinetics of moclobemide according to the genetic polymorphism of CYP2C19. *Clin. Pharmacol. Ther.,* **2001***, 69*, 266-273.
[288] Bertilsson, L.; Henthorn, T.K.; Sanz, E.; Tybring, G.; Säwe, J.; Villén, T. Importance of genetic factors in the regulation of diazepam metabolism: relationship to S-mephenytoin, but not debrisoquin, hydroxylation phenotype. *Clin. Pharmacol. Ther.,* **1989***, 45*, 348-355.
[289] Kosaki, K.; Tamura, K.; Sato, R.; Samejima, H.; Tanigawara, Y.; Takahashi, T. A major influence of CYP2C19 genotype on the steady-state concentration of N-desmethylclobazam. *Brain Dev.,* **2004***, 26*, 530-534.
[290] Mikus, G.; Scholz, I.M.; Weiss, J. Pharmacogenomics of the triazole antifungal agent voriconazole. *Pharmacogenomics,* **2011***, 12*, 861-872.
[291] Kerb, R.; Fux, R.; Mörike, K.; Kremsner, P.G.; Gil, J.P.; Gleiter, C.H.; Schwab, M. Pharmacogenetics of antimalarial drugs: effect on metabolism and transport. *Lancet Infect. Dis.,* **2009***, 9*, 760-774.
[292] Collet, J.; Hulot, J.; Pena, A.; Villard, E.; Esteve, J.; Silvain, J.; Payot, L.; Brugier, D.; Cayla, G.; Beygui, F. Cytochrome P450 2C19 polymorphism in young patients treated with clopidogrel after myocardial infarction: a cohort study. *The Lancet,* **2009***, 373*, 309-317.
[293] Mega, J.L.; Close, S.L.; Wiviott, S.D.; Shen, L.; Hockett, R.D.; Brandt, J.T.; Walker, J.R.; Antman, E.M.; Macias, W.; Braunwald, E. Cytochrome p-450 polymorphisms and response to clopidogrel. *N. Engl. J. Med.,* **2009***, 360*, 354-362.
[294] Simon, T.; Verstuyft, C.; Mary-Krause, M.; Quteineh, L.; Drouet, E.; Méneveau, N.; Steg, P.G.; Ferrières, J.; Danchin, N.; Becquemont, L. Genetic determinants of response to clopidogrel and cardiovascular events. *N. Engl. J. Med.,* **2009***, 360*, 363-375.
[295] Sofi, F.; Giusti, B.; Marcucci, R.; Gori, A.M.; Abbate, R.; Gensini, G.F. Cytochrome P450 2C19*2 polymorphism and cardiovascular recurrences in patients taking clopidogrel: a meta-analysis. *Pharmacogenomics J.,* **2011***, 11*, 199-206.
[296] Sim, S.C.; Risinger, C.; Dahl, M.; Aklillu, E.; Christensen, M.; Bertilsson, L.; Ingelman-Sundberg, M. A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants*. *Clin. Pharmacol. Ther.,* **2006***, 79*, 103-113.
[297] Sibbing, D.; Koch, W.; Gebhard, D.; Schuster, T.; Braun, S.; Stegherr, J.; Morath, T.; Schomig, A.; von Beckerath, N.; Kastrati, A. Cytochrome 2C19*17 allelic variant, platelet aggregation, bleeding events, and stent thrombosis in clopidogrel-treated patients with coronary stent placement. *Circulation,* **2010***, 121*, 512-518.

[298] Darbar, D.; Roden, D.M. Pharmacogenetics of antiarrhythmic therapy. *Expert opinion on pharmacotherapy,* **2006***, 7*, 1583-1590.
[299] Stingl, J.; Brockmöller, J.; Viviani, R. Genetic variability of drug-metabolizing enzymes: the dual impact on psychiatric therapy and regulation of brain function. *Mol. Psychiatry,* **2012***, 18*, 273-287.
[300] Leppert, W. CYP2D6 in the metabolism of opioids for mild to moderate pain. *Pharmacology,* **2011***, 87*, 274-285.
[301] Stamer, U.M.; Zhang, L.; Stüber, F. Personalized therapy in pain management: where do we stand? *Pharmacogenomics,* **2010***, 11*, 843-864.
[302] Brauch, H.; Murdter, T.E.; Eichelbaum, M.; Schwab, M. Pharmacogenomics of tamoxifen therapy. *Clin. Chem.,* **2009***, 55*, 1770-1782.
[303] Gasche, Y.; Daali, Y.; Fathi, M.; Chiappe, A.; Cottini, S.; Dayer, P.; Desmeules, J. Codeine intoxication associated with ultrarapid CYP2D6 metabolism. *N. Engl. J. Med.,* **2004***, 351*, 2827-2831.
[304] Koren, G.; Cairns, J.; Chitayat, D.; Gaedigk, A.; Leeder, S.J. Pharmacogenetics of morphine poisoning in a breastfed neonate of a codeine-prescribed mother. *Lancet,* **2006***, 368*, 704.
[305] Madadi, P.; Avard, D.; Koren, G. Pharmacogenetics of opioids for the treatment of acute maternal pain during pregnancy and lactation. *Curr. Drug Metab.,* **2012***, 13*, 721-727.
[306] Wang, D.; Guo, Y.; Wrighton, S.A.; Cooke, G.E.; Sadee, W. Intronic polymorphism in CYP3A4 affects hepatic expression and response to statin drugs. *Pharmacogenomics J.,* **2011***, 11*, 274-286.
[307] Elens, L.; van Schaik, R.H.; Panin, N.; De Meyer, M.; Wallemacq, P.; Lison, D.; Mourad, M.; Haufroid, V. Effect of a new functional CYP3A4 polymorphism on calcineurin inhibitors' dose requirements and trough blood levels in stable renal transplant patients. *Pharmacogenomics,* **2011***, 12*, 1383-1396.
[308] Elens, L.; Becker, M.L.; Haufroid, V.; Hofman, A.; Visser, L.E.; Uitterlinden, A.G.; Stricker, B.C.; van Schaik, R.H. Novel CYP3A4 intron 6 single nucleotide polymorphism is associated with simvastatin-mediated cholesterol reduction in the Rotterdam Study. *Pharmacogenet. Genomics,* **2011***, 21*, 861-866.
[309] Human Cytochrome P450 Allele Nomenclature Committee home page. (www.cypalleles.ki.se).
[310] Hadidi, H.; Zahlsen, K.; Idle, J.R.; Cholerton, S. A single amino acid substitution (Leu160His) in cytochrome P450 CYP2A6 causes switching from 7-hydroxylation to 3-hydroxylation of coumarin. *Food Chem. Toxicol.,* **1997***, 35*, 903-907.
[311] Richardson, T.H.; Johnson, E.F. Alterations of the regiospecificity of progesterone metabolism by the mutagenesis of two key amino acid residues in rabbit cytochrome P450 2C3v. *J. Biol. Chem.,* **1994***, 269*, 23937-23943.
[312] Paine, M.J.; McLaughlin, L.A.; Flanagan, J.U.; Kemp, C.A.; Sutcliffe, M.J.; Roberts, G.C.; Wolf, C.R. Residues glutamate 216 and aspartate 301 are key determinants of substrate specificity and product regioselectivity in cytochrome P450 2D6. *J. Biol. Chem.,* **2003***, 278*, 4021-4027.
[313] Fukami, T.; Nakajima, M.; Higashi, E.; Yamanaka, H.; Sakai, H.; McLeod, H.L.; Yokoi, T. Characterization of novel CYP2A6 polymorphic alleles (CYP2A6*18 and CYP2A6*19) that affect enzymatic activity. *Drug Metab. Dispos.,* **2005***, 33*, 1202-1210.
[314] Ariyoshi, N.; Sawamura, Y.; Kamataki, T. A novel single nucleotide polymorphism altering stability and activity of CYP2a6. *Biochem. Biophys. Res. Commun.,* **2001***, 281*, 810-814.

[315] Wang, H.; Kim, R.a.; Sun, D.; Gao, Y.; Wang, H.; Zhu, J.; Chen, C. Evaluation of the effects of 18 non-synonymous single-nucleotide polymorphisms of CYP450 2C19 on in vitro drug inhibition potential by a fluorescence-based high-throughput assay. *Xenobiotica,* **2011***, 41*, 826-35.
[316] Wang, H.; An, N.; Wang, H.; Gao, Y.; Liu, D.; Bian, T.; Zhu, J.; Chen, C. Evaluation of the Effects of 20 nonsynonymous single nucleotide polymorphisms of CYP2C19 on S-mephenytoin 4′-hydroxylation and omeprazole 5′-hydroxylation. *Drug Metab. Dispos.,* **2011***, 39*, 830-837.
[317] Zhou, H.; Josephy, P.D.; Kim, D.; Guengerich, F.P. Functional characterization of four allelic variants of human cytochrome P450 1A2. *Arch. Biochem. Biophys.,* **2004***, 422*, 23-30.
[318] Paul, S.M.; Mytelka, D.S.; Dunwiddie, C.T.; Persinger, C.C.; Munos, B.H.; Lindborg, S.R.; Schacht, A.L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.,* **2010***, 9*, 203-214.
[319] Lindgardt, Z.; Reeves, M.; Wallenstein, J. Waking the giant: business model innovation in the drug industry. *IN VIVO-NEW SERIES-,* **2008***, 26*, 54.
[320] Caskey, C.T. The drug development crisis: efficiency and safety. *Annu. Rev. Med.,* **2007***, 58*, 1-16.
[321] Beeton-Kempen, N.; Shoko, A.; Blackburn, J. Personalizing protein-drug interactions. *Pure Appl. Chem.,* **2008***, 80*, 1811-1820.
[322] Hou, T.; Wang, J.; Zhang, W.; Wang, W.; Xu, X. Recent advances in computational prediction of drug absorption and permeability in drug discovery. *Curr. Med. Chem.,* **2006***, 13*, 2653-2667.
[323] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.,* **2012***, 64*, 4-17.
[324] Baldrick, P. Safety evaluation to support First-In-Man investigations I: Kinetic and safety pharmacology studies. *Regul. Toxicol. Pharmacol.,* **2008***, 51*, 230-236.
[325] Gad, S.C. *Clinical trials handbook*, illustrated ed.; John Wiley & Sons: New Jersey, **2009**.
[326] Amstutz, U.; Carleton, B. Pharmacogenetic testing: time for clinical practice guidelines. *Clin. Pharmacol. Ther.,* **2011***, 89*, 924.
[327] Relling, M.; Guchelaar, H.; Roden, D.; Klein, T. Pharmacogenetics: call to action. *Clin. Pharmacol. Ther.,* **2011***, 90*, 507-507.
[328] Relling, M.V.; Klein, T.E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin. Pharmacol. Ther.,* **2011***, 89*, 464-467.
[329] Swen, J.; Nijenhuis, M.; De Boer, A.; Grandia, L.; Maitland-Van Der Zee, A.; Mulder, H.; Rongen, G.; van Schaik, R.; Schalekamp, T.; Touw, D. Pharmacogenetics: from bench to byte—an update of guidelines. *Clin. Pharmacol. Ther.,* **2011***, 89*, 662-673.
[330] PharmGKB. The Pharmacogenomics Knowledgebase. (http://www.pharmgkb.org/) (Accessed July 3, **2014**).
[331] PharmGKB. Well-Known Pharmacogenomic Associations. (http://www.pharmgkb.org/search/knownPairs.action) (Accessed July 3, **2014**).
[332] FDA. Table of Pharmacogenomic Biomarkers in Drug Labeling. (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm) (Accessed July 3, **2014**).
[333] PharmGKB. Genetic Tests. (http://www.pharmgkb.org/views/viewGeneticTests.action) (Accessed July 3, **2014**).
[334] Johnson, J.; Gong, L.; Whirl-Carrillo, M.; Gage, B.; Scott, S.; Stein, C.; Anderson, J.; Kimmel, S.; Lee, M.; Pirmohamed, M. Clinical Pharmacogenetics Implementation

Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clin. Pharmacol. Ther.,* **2011***, 90*, 625-629.
[335] Palareti, G.; Leali, N.; Coccheri, S.; Poggi, M.; Manotti, C.; D'Angelo, A.; Pengo, V.; Erba, N.; Moia, M.; Ciavarella, N. Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study (ISCOAT). *The Lancet,* **1996***, 348*, 423-428.
[336] Hafner Jr, J.W.; Belknap, S.M.; Squillante, M.D.; Bucheit, K.A. Adverse drug events in emergency department patients. *Ann. Emerg. Med.,* **2002***, 39*, 258-267.
[337] Fihn, S.D.; McDonell, M.; Martin, D.; Henikoff, J.; Vermes, D.; Kent, D.; White, R.H. Risk factors for complications of chronic anticoagulation: a multicenter study. *Ann. Intern. Med.,* **1993***, 118*, 511-520.
[338] McWilliam, A.; Lutter, R.W.; Nardinelli, C. *Health care savings from personalizing medicine using genetic testing: the case of warfarin*; AEI-Brookings Joint Center for Regulatory Studies: Washington, DC, **2006**.
[339] Bash, P.; Singh, U.; Langridge, R.; Kollman, P. Free energy calculations by computer simulation. *Science,* **1987***, 236*, 564-568.
[340] Park, H.; Lee, S. Prediction of the mutation-induced change in thermodynamic stabilities of membrane proteins from free energy simulations. *Biophys. Chem.,* **2005***, 114*, 191-7.
[341] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T.E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.,* **2000***, 33*, 889-97.
[342] Bordner, A.J.; Abagyan, R.A. Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. *Proteins,* **2004***, 57*, 400-13.
[343] Guerois, R.; Nielsen, J.E.; Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.,* **2002***, 320*, 369-87.
[344] Topham, C.; Srinivasan, N.; Blundell, T. Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng.,* **1997***, 10*, 7-21.
[345] Gilis, D.; Rooman, M. Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. *J. Mol. Biol.,* **1997***, 272*, 276-90.
[346] Saraboji, K.; Gromiha, M.M.; Ponnuswamy, M.N. Average assignment method for predicting the stability of protein mutants. *Biopolymers,* **2006***, 82*, 80-92.
[347] Frenz, C.M. Neural network-based prediction of mutation-induced protein stability changes in Staphylococcal nuclease at 20 residue positions. *Proteins,* **2005***, 59*, 147-51.
[348] Capriotti, E.; Fariselli, P.; Casadio, R. A neural-network-based method for predicting protein stability changes upon single point mutations. *Bioinformatics,* **2004***, 20 Suppl 1*, i63-8.
[349] Cheng, J.; Randall, A.; Baldi, P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins,* **2006***, 62*, 1125-32.
[350] Capriotti, E.; Fariselli, P.; Calabrese, R.; Casadio, R. Predicting protein stability changes from sequences using support vector machines. *Bioinformatics,* **2005***, 21 Suppl 2*, ii54-8.
[351] Wang, Z.; Moult, J. SNPs, protein structure, and disease. *Hum. Mutat.,* **2001***, 17*, 263-70.
[352] Ng, P.C.; Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.,* **2003***, 31*, 3812-3814.
[353] Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols,* **2009***, 4*, 1073-1081.

[354] Bromberg, Y.; Rost, B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.,* **2007***, 35*, 3823-3835.
[355] Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: an online force field. *Nucleic Acids Res.,* **2005***, 33*, W382-W388.
[356] Worth, C.L.; Burke, D.; Blundell, T.L. Estimating the Effects Of Single Nucleotide Polymorphisms On Protein Structure : How Good are we at Identifying Likely Disease Associated Mutations ? *Proceedings of Molecular Interactions—Bringing Chemistry to Life,* **2007**, 11-26.
[357] Worth, C.L.; Schreyer, A.; Forman, J.R.; Cheng, T.M.K.; Lee, S.; Gong, S.; Burke, D.F. A structural bioinformatics approach to the analysis of nonsynonymous single nucleotide polymorphisms (nsSNPs) and their relation to disease. *JBCB,* **2007***, 5*, 1297-1318.
[358] Worth, C.L.; Preissner, R.; Blundell, T.L. SDM--a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.,* **2011***, 39*, W215-22.
[359] Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nature methods,* **2010***, 7*, 248-249.
[360] Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.,* **2005***, 33*, W306-W310.
[361] Wang, L.; Li, Y.; Zhou, S. A Bioinformatics Approach for the Phenotype Prediction of Nonsynonymous Single Nucleotide Polymorphisms in Human Cytochromes P450. *Drug Metab. Dispos.,* **2009***, 37*, 977-991.
[362] Brandon, E.F.; Raap, C.D.; Meijerman, I.; Beijnen, J.H.; Schellens, J.H. An update on in vitro test methods in human hepatic drug biotransformation research: pros and cons. *Toxicol. Appl. Pharmacol.,* **2003***, 189*, 233-246.
[363] Donato, M.T.; Lahoz, A.; Castell, J.V.; Gomez-Lechon, M.J. Cell lines: a tool for in vitro drug metabolism studies. *Curr. Drug Metab.,* **2008***, 9*, 1-11.
[364] Cohen, L.H.; Remley, M.J.; Raunig, D.; Vaz, A.D.N. In vitro drug interactions of cytochrome p450: an evaluation of fluorogenic to conventional substrates. *Drug Metab. Dispos.,* **2003***, 31*, 1005-15.
[365] Di, L.; Kerns, E.H.; Li, S.Q.; Carter, G.T. Comparison of cytochrome P450 inhibition assays for drug discovery using human liver microsomes with LC-MS, rhCYP450 isozymes with fluorescence, and double cocktail with LC-MS. *Int. J. Pharm.,* **2007***, 335*, 1-11.
[366] Lahoz, A.; Donato, M.T.; Castell, J.V.; Gómez-Lechón, M.J. Strategies to in vitro assessment of major human CYP enzyme activities by using liquid chromatography tandem mass spectrometry. *Curr. Drug Metab.,* **2008***, 9*, 12-9.
[367] Brian Houston, J.; Carlile, D.J. Prediction of hepatic clearance from microsomes, hepatocytes, and liver slices. *Drug Metab. Rev.,* **1997***, 29*, 891-922.
[368] Crewe, H.; Barter, Z.; Rowland Yeo, K.; Rostami-Hodjegan, A. Are there differences in the catalytic activity per unit enzyme of recombinantly expressed and human liver microsomal cytochrome P450 2C9? A systematic investigation into inter-system extrapolation factors. *Biopharm. Drug Dispos.,* **2011***, 32*, 303-318.
[369] Wienkers, L.C.; Heath, T.G. Predicting in vivo drug interactions from in vitro drug discovery data. *Nat. Rev. Drug Discov.,* **2005***, 4*, 825-833.
[370] George, J.; Goodwin, B.; Liddle, C.; Tapner, M.; Farrell, G.C. Time-dependent expression of cytochrome P450 genes in primary cultures of well-differentiated human hepatocytes. *J. Lab. Clin. Med.,* **1997***, 129*, 638-648.
[371] Zinchenko, Y.S.; Schrum, L.W.; Clemens, M.; Coger, R.N. Hepatocyte and kupffer cells co-cultured on micropatterned surfaces to optimize hepatocyte function. *Tissue Eng.,* **2006***, 12*, 751-761.

[372] Bhatia, S.; Balis, U.; Yarmush, M.; Toner, M. Microfabrication of Hepatocyte/Fibroblast Co-cultures: Role of Homotypic Cell Interactions. *Biotechnol. Prog.,* **1998***, 14*, 378-387.
[373] Otsuka, H.; Sasaki, K.; Okimura, S.; Nagamura, M.; Watanabe, R.; Kawabe, M. Contribution of Fibroblasts Cultured on 3D Silica Nonwoven Fabrics to Cocultured Hepatocytes Function. *Chem. Lett.,* **2014***, 43*, 343-345.
[374] Kim, B.; Park, I.; Hoshiba, T.; Jiang, H.; Choi, Y.; Akaike, T.; Cho, C. Design of artificial extracellular matrices for tissue engineering. *Prog. Polym. Sci.,* **2011***, 36*, 238-268.
[375] Khojasteh, S.C.; Prabhu, S.; Kenny, J.R.; Halladay, J.S.; Lu, A.Y. Chemical inhibitors of cytochrome P450 isoforms in human liver microsomes: a re-evaluation of P450 isoform selectivity. *Eur. J. Drug Metab. Pharmacokinet.,* **2011***, 36*, 1-16.
[376] Ghaemmaghami, A.M.; Hancock, M.J.; Harrington, H.; Kaji, H.; Khademhosseini, A. Biomimetic tissues on a chip for drug discovery. *Drug Discov. Today,* **2012***, 17*, 173-181.
[377] Schütte, J.; Hagmeyer, B.; Holzner, F.; Kubon, M.; Werner, S.; Freudigmann, C.; Benz, K.; Böttger, J.; Gebhardt, R.; Becker, H. "Artificial micro organs"—a microfluidic device for dielectrophoretic assembly of liver sinusoids. *Biomed. Microdevices,* **2011***, 13*, 493-501.
[378] Elliott, N.T.; Yuan, F. A review of three-dimensional in vitro tissue models for drug discovery and transport studies. *J. Pharm. Sci.,* **2011***, 100*, 59-74.
[379] van Midwoud, P.M.; Verpoorte, E.; Groothuis, G.M. Microfluidic devices for in vitro studies on liver drug metabolism and toxicity. *Integrative Biology,* **2011***, 3*, 509-521.
[380] Pelkonen, O.; Raunio, H. Metabolic activation of toxins: tissue-specific expression and metabolism in target organs. *Environ. Health Perspect.,* **1997***, 105 Suppl 4*, 767-774.
[381] Sukumaran, S.M.; Potsaid, B.; Lee, M.; Clark, D.S.; Dordick, J.S. Development of a fluorescence-based, ultra high-throughput screening platform for nanoliter-scale cytochrome p450 microarrays. *J. Biomol. Screen.,* **2009***, 14*, 668-78.
[382] Arrabito, G.; Galati, C.; Castellano, S.; Pignataro, B. Luminometric sub-nanoliter droplet-to-droplet array (LUMDA) and its application to drug screening by phase I metabolism enzymes. *Lab on a Chip,* **2013***, 13*, 68-72.
[383] Kronbach, T.; Johnson, E.F. An inhibitory monoclonal antibody binds in close proximity to a determinant for substrate binding in cytochrome P450IIC5. *J. Biol. Chem.,* **1991***, 266*, 6215-6220.
[384] Wang, H.; Dick, R.; Yin, H.; Licad-Coles, E.; Kroetz, D.L.; Szklarz, G.; Harlow, G.; Halpert, J.R.; Correia, M.A. Structure-function relationships of human liver cytochromes P450 3A: aflatoxin B1 metabolism as a probe. *Biochemistry (N. Y.),* **1998***, 37*, 12536-12545.
[385] Ibeanu, G.C.; Ghanayem, B.I.; Linko, P.; Li, L.; Pederson, L.G.; Goldstein, J.A. Identification of residues 99, 220, and 221 of human cytochrome P450 2C19 as key determinants of omeprazole activity. *J. Biol. Chem.,* **1996***, 271*, 12496-12501.
[386] Fraser, D.J.; He, Y.Q.; Harlow, G.R.; Halpert, J.R. Use of chimeric enzymes and site-directed mutagenesis for identification of three key residues responsible for differences in steroid hydroxylation between canine cytochromes P-450 3A12 and 3A26. *Mol. Pharmacol.,* **1999***, 55*, 241-247.
[387] He, Y.; Luo, Z.; Klekotka, P.A.; Burnett, V.L.; Halpert, J.R. Structural determinants of cytochrome P450 2B1 specificity: evidence for five substrate recognition sites. *Biochemistry (N. Y.),* **1994***, 33*, 4419-4424.
[388] 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*, 5396-5402.
[389] He, Y.Q.; He, Y.A.; Halpert, J.R. Escherichia coli expression of site-directed mutants of cytochrome P450 2B1 from six substrate recognition sites: substrate specificity and inhibitor selectivity studies. *Chem. Res. Toxicol.,* **1995***, 8*, 574-579.

[390] Roussel, F.; Khan, K.K.; Halpert, J.R. The importance of SRS-1 residues in catalytic specificity of human cytochrome P450 3A4. *Arch. Biochem. Biophys.,* **2000***, 374*, 269-278.
[391] 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 (N. Y.),* **1997***, 36*, 8831-8839.
[392] Conley, A.; Mapes, S.; Corbin, C.J.; Greger, D.; Graham, S. Structural determinants of aromatase cytochrome P450 inhibition in substrate recognition site-1. *Molecular Endocrinology,* **2002***, 16*, 1456-1468.
[393] He, Y.Q.; Harlow, G.R.; Szklarz, G.D.; Halpert, J.R. Structural determinants of progesterone hydroxylation by cytochrome P450 2B5: the role of nonsubstrate recognition site residues. *Arch. Biochem. Biophys.,* **1998***, 350*, 333-339.
[394] Biagini, C.P.; Philpot, R.M.; Célier, C.M. Nonsubstrate recognition site residues are involved in testosterone hydroxylation by cytochrome P450 CYP 2C11. *Arch. Biochem. Biophys.,* **1999***, 361*, 309-314.
[395] Wallace, A.; Laskowski, R.; Thornton, J. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.,* **1995***, 8*, 127-134.
[396] Kjellander, B.; Masimirembwa, C.M.; Zamora, I. Exploration of enzyme-ligand interactions in CYP2D6 & 3A4 homology models and crystal structures using a novel computational approach. *J. Chem. Inf. Model.,* **2007***, 47*, 1234-1247.
[397] Schreyer, A.; Blundell, T. CREDO: a protein-ligand interaction database for drug discovery. *Chem. Biol Drug Des.,* **2009***, 73*, 157-67.
[398] Schreyer, A.M.; Blundell, T.L. CREDO: a structural interactomics database for drug discovery. *Database,* **2013***, 2013*, bat049.
[399] Gallina, A.M.; Bisignano, P.; Bergamino, M.; Bordo, D. PLI: a web-based tool for the comparison of protein-ligand interactions observed on PDB structures. *Bioinformatics,* **2013***, 29*, 395-397.
[400] Zhou, D.; Afzelius, L.; Grimm, S.W.; Andersson, T.B.; Zauhar, R.J.; Zamora, I. Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. *Drug Metab. Dispos.,* **2006***, 34*, 976-983.
[401] Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.,* **2005***, 48*, 6970-9.
[402] Zawaira, A.; Ching, L.Y.; Coulson, L.; Blackburn, J.; Wei, Y.C. An Expanded, Unified Substrate Recognition Site Map for Mammalian Cytochrome P450s: Analysis of Molecular Interactions Between 15 Mammalian CYP450 Isoforms and 868 Substrates. *Curr. Drug Metab.,* **2011***, 12*, 684-700.
[403] Yun, C.H.; Hammons, G.J.; Jones, G.; Martin, M.V.; Hopkins, N.E.; Alworth, W.L.; Guengerich, F.P. Modification of cytochrome P450 1A2 enzymes by the mechanism-based inactivator 2-ethynylnaphthalene and the photoaffinity label 4-azidobiphenyl. *Biochemistry (N. Y.),* **1992***, 31*, 10556-10563.
[404] Aoyama, T.; Korzekwa, K.; Nagata, K.; Adesnik, M.; Reiss, A.; Lapenson, D.P.; Gillette, J.; Gelboin, H.V.; Waxman, D.J.; Gonzalez, F.J. Sequence requirements for cytochrome P-450IIB1 catalytic activity. Alteration of the stereospecificity and regioselectivity of steroid hydroxylation by a simultaneous change of two hydrophobic amino acid residues to phenylalanine. *J. Biol. Chem.,* **1989***, 264*, 21327-21333.
[405] Von Weymarn, L.B.; Sridar, C.; Hollenberg, P.F. Identification of amino acid residues involved in the inactivation of cytochrome P450 2B1 by two acetylenic compounds: the role of three residues in nonsubstrate recognition Sites. *J. Pharmacol. Exp. Ther.,* **2004***, 311*, 71-79.

[406] DeLano, W.L. The PyMOL Molecular Graphics System. Version 1.4, Schrödinger, LLC.
[407] Ioannides, C. *Cytochromes P450: Role in the metabolism and toxicity of drugs and other xenobiotics*; Royal Society of Chemistry: Cambridge, UK, **2008**.
[408] Coon, M.J. Cytochrome P450: nature's most versatile biological catalyst. *Annu. Rev. Pharmacol. Toxicol.,* **2005***, 45*, 1-25.
[409] Zhang, H.; Myshkin, E.; Waskell, L. Role of cytochrome b5 in catalysis by cytochrome P450 2B4. *Biochem. Biophys. Res. Commun.,* **2005***, 338*, 499-506.
[410] Department of Medicine Indiana University. P450 Drug Interaction Table. http://medicine.iupui.edu/clinpharm/ddis/table.asp (Accessed July 1, **2010**).
[411] Hayes, E. Cytochrome P-450 and Drug Interactions. (http://www.edhayes.com/CYP450-3.html) (Accessed April 14, **2010**).
[412] Genelex Corporation. P450 Drug-Interactions Substrates Table. http://www.healthanddna.com/Druglist.pdf (Accessed June 1, **2010**).
[413] Pharmacy Society of Wisconsin. Clinically Significant Cytochrome P450 Drug Interactions. (http://www.pswi.org/professional/pharmaco/Cytochrome.pdf) (Accessed June 1, **2010**).
[414] Structural Genomics Consortium. Human cytochrome P450, family 2, R1. (http://www.sgc.utoronto.ca/SGC-WebPages/StructureDescription/2OJD.php) (Accessed June 1, **2010**).
[415] Just the Berries. Clinically Significant Cytochrome P-450 (CYP450) Drug Interactions. (http://www.theberries.ca/archives/cyp450.html) (Accessed June 1, **2010**).
[416] UniProt Consortium. Q6VVX0 (CP2R1_HUMAN). (www.uniprot.org/uniprot/Q6VVX0) (Accessed June 1, **2010**).
[417] Thomson Reuters. MicroMedex 2.0. (http://www.micromedex.com) (Accessed June 1, **2010**).
[418] Accerlrys. Pipeline Pilot. (http://accelrys.com/products/pipeline-pilot/) (Accessed June 1, **2010**).
[419] National Center for Biotechnology Information. PubChem. (http://pubchem.ncbi.nlm.nih.gov) (Accessed June 1, **2010**).
[420] Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.,* **2011***, 32*, 1466-1474.
[421] Umetrics. SIMCA-P+. (http://www.umetrics.com/simca) (Accessed June 1, **2010**).
[422] Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.,* **1994***, 22*, 4673-4680.
[423] Sali, A.; Blundell, T.L. Definition of general topological equivalence in protein structures: A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *J. Mol. Biol.,* **1990***, 212*, 403-428.
[424] Trott, O.; Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.,* **2010***, 31*, 455-461.
[425] The Scripps Research Institute. AutoDock. http://autodock.scripps.edu/ (Accessed June 1, **2010**).
[426] National University of Signapore. PaDEL-ADV. (http://padel.nus.edu.sg/software/padeladv/) (Accessed June 1, **2010**).
[427] Wade, R.C.; Winn, P.J.; Schlichting, I.; Sudarko. A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.,* **2004***, 98*, 1175-1182.

[428] Park, S.; Shimizu, H.; Adachi, S.; Nakagawa, A.; Tanaka, I.; Nakahara, K.; Shoun, H.; Obayashi, E.; Nakamura, H.; Shiro, Y. Crystal structure of nitric oxide reductase from denitrifying fungus Fusarium oxysporum. *Nat. Struct. Mol. Biol.,* **1997***, 4*, 827-832.
[429] Ravichandran, K.G.; Boddupalli, S.S.; Hasermann, C.A.; Peterson, J.A.; Deisenhofer, J. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science,* **1993***, 261*, 731-736.
[430] Gora, A.; Brezovsky, J.; Damborsky, J. Gates of enzymes. *Chem. Rev.,* **2013***, 113*, 5871-5923.
[431] Lüdemann, S.K.; Carugo, O.; Wade, R.C. Substrate Access to Cytochrome P450cam: A Comparison of a Thermal Motion Pathway Analysis with Molecular Dynamics Simulation Data. *J. Mol. Model.,* **1997***, 3*, 369-374.
[432] Winn, P.J.; Lüdemann, S.K.; Gauges, R.; Lounnas, V.; Wade, R.C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc. Natl. Acad. Sci. U. S. A.,* **2002***, 99*, 5361-5366.
[433] Schleinkofer, K.; Winn, P.J.; Lüdemann, S.K.; Wade, R.C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep.,* **2005***, 6*, 584-589.
[434] Fishelovitch, D.; Shaik, S.; Wolfson, H.J.; Nussinov, R. Theoretical Characterization of Substrate Access/Exit Channels in the Human Cytochrome P450 3A4 Enzyme: Involvement of Phenylalanine Residues in the Gating Mechanism. *J. Phys. Chem. B.,* **2009***, 113*, 13018-13025.
[435] Denisov, I.G.; Shih, A.Y.; Sligar, S.G. Structural differences between soluble and membrane bound cytochrome P450s. *J. Inorg. Biochem.,* **2012***, 108*, 150-158.
[436] Oostenbrink, C.; de Ruiter, A.; Hritz, J.; Vermeulen, N. Malleability and versatility of cytochrome P450 active sites studied by molecular simulations. *Curr. Drug Metab.,* **2012***, 13*, 190.
[437] Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics,* **2006***, 7*.
[438] Petrek, M.; Kosinova, P.; Koca, J.; Otyepka, M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure,* **2007***, 15*, 1357-1363.
[439] Yaffe, E.; Fishelovitch, D.; Wolfson, H.J.; Halperin, D.; Nussinov, R. MolAxis: efficient and accurate identification of channels in macromolecules. *Proteins,* **2008***, 73*, 72-86.
[440] Berka, K.; Hanák, O.; Sehnal, D.; Banáš, P.; Navrátilová, V.; Jaiswal, D.; Ionescu, C.; Vařeková, R.S.; Koča, J.; Otyepka, M. MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.,* **2012***, 40*, W222-W227.
[441] Haines, D.C.; Tomchick, D.R.; Machius, M.; Peterson, J.A. Pivotal role of water in the mechanism of P450BM-3. *Biochemistry,* **2001***, 40*, 13456-13465.
[442] Fishelovitch, D.; Shaik, S.; Wolfson, H.J.; Nussinov, R. How does the reductase help to regulate the catalytic cycle of cytochrome P450 3A4 using the conserved water channel? *J. Phys. Chem. B.,* **2010***, 114*, 5964-70.
[443] Zawaira, A.; Coulson, L.; Gallotta, M.; Karimanzira, O.; Blackburn, J. On the deduction and analysis of singlet and two-state gating-models from the static structures of mammalian CYP450. *J. Struct. Biol.,* **2011***, 173*, 282-293.
[444] Tina, K.G.; Bhadra, R.; Srinivasan, N. PIC: Protein Interactions Calculator. *Nucleic Acids Res.,* **2007***, 35*, W473-W476.
[445] Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.6. http://evolution.genetics.washington.edu/phylip/phylipweb.html (Accessed Jan 1, **2014**).

[446] Mayrose, I.; Graur, D.; Ben-Tal, N.; Pupko, T. Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol. Biol. Evol.,* **2004***, 21*, 1781-1791.
[447] Arendse, L.; L Blundell, T.; Blackburn, J. Combining In Silico Protein Stability Calculations with Structure-Function Relationships to Explore the Effect of Polymorphic Variation on Cytochrome P450 Drug Metabolism. *Curr. Drug Metab.,* **2013***, 14*, 745-763.
[448] Watanabe, T.; Sakuyama, K.; Sasaki, T.; Ishii, Y.; Ishikawa, M.; Hirasawa, N.; Hiratsuka, M. Functional characterization of 26 CYP2B6 allelic variants (CYP2B6. 2-CYP2B6. 28, except CYP2B6. 22). *Pharmacogenet. Genomics,* **2010***, 20*, 459-462.
[449] Mizuguchi, K.; Deane, C.M.; Blundell, T.L.; Johnson, M.S.; Overington, J.P. JOY: protein sequence-structure representation and analysis. *Bioinformatics,* **1998***, 14*, 617-623.
[450] Couñago, R.; Wilson, C.J.; Peña, M.I.; Wittung-Stafshede, P.; Shamoo, Y. An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability–activity trade-offs. *Protein Eng. Des. Sel.,* **2008***, 21*, 19-27.
[451] Mukaiyama, A.; Haruki, M.; Ota, M.; Koga, Y.; Takano, K.; Kanaya, S. A hyperthermophilic protein acquires function at the cost of stability. *Biochemistry,* **2006***, 45*, 12673-12679.
[452] Beadle, B.M.; Shoichet, B.K. Structural bases of stability–function tradeoffs in enzymes. *J. Mol. Biol.,* **2002***, 321*, 285-296.
[453] Yutani, K.; Ogasahara, K.; Tsujita, T.; Sugino, Y. Dependence of conformational stability on hydrophobicity of the amino acid residue in a series of variant proteins substituted at a unique position of tryptophan synthase alpha subunit. *Proc. Natl. Acad. Sci.,* **1987***, 84*, 4441-4444.
[454] Yun, C.; Yim, S.; Kim, D.; Ahn, T. Functional expression of human cytochrome P450 enzymes in Escherichia coli. *Curr. Drug Metab.,* **2006***, 7*, 411-29.
[455] Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.,* **2003***, 31*, 3812-3814.
[456] Sano, E.; Li, W.; Yuki, H.; Liu, X.; Furihata, T.; Kobayashi, K.; Chiba, K.; Neya, S.; Hoshino, T. Mechanism of the decrease in catalytic activity of human cytochrome P450 2C9 polymorphic variants investigated by computational analysis. *J.Comput.Chem.,* **2010***, 31*, 2746-2758.
[457] Oda, A.; Yamaotsu, N.; Hirono, S. Studies of Binding Modes of (S)-Mephenytoin to Wild Types and Mutants of Cytochrome P450 2C19 and 2C9 Using Homology Modeling and Computational Docking. *Pharm. Res.,* **2004***, 21*, 2270-2278.
[458] Zhou, Y.-.; Zheng, Q.-.; Li, Z.-.; Zhang, Y.; Sun, M.; Sun, C.-.; Si, D.; Cai, L.; Guo, Y.; Zhou, H. On the human CYP2C9* 13 variant activity reduction: a molecular dynamics simulation and docking study. *Biochimie,* **2006***, 88*, 1457-1465.
[459] Burke, D.F.; Worth, C.L.; Priego, E.; Cheng, T.; Smink, L.J.; Todd, J.A.; Blundell, T.L. Genome bioinformatic analysis of nonsynonymous SNPs. *BMC Bioinformatics,* **2007***, 8*, 301.
[460] Ibeanu, G.; Goldstein, J.; Meyer, U.; Benhamou, S.; Bouchardy, C.; Dayer, P.; Ghanayem, B.; Blaisdell, J. Identification of new human CYP2C19 alleles (CYP2C19*6 and CYP2C19*2B) in a Caucasian poor metabolizer of mephenytoin. *J. Pharmacol. Exp. Ther.,* **1998***, 286*, 1490-1495.
[461] Blaisdell, J.; Mohrenweiser, H.; Jackson, J.; Ferguson, S.; Coulter, S.; Chanas, B.; Xi, T.; Ghanayem, B.; Goldstein, J.A. Identification and functional characterization of new potentially defective alleles of human CYP2C19. *Pharmacogenetics,* **2002***, 12*, 703-711.
[462] Hiratsuka, M. In vitro assessment of the allelic variants of cytochrome P450. *Drug Metab. Pharmacokinet.,* **2012***, 27*, 68-84.

[463] NCBI. dbSNP Home Page. (http://www.ncbi.nlm.nih.gov/SNP/index.html) (Accessed June 1, **2011**).
[464] Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.,* **2007***, 47*, 195-207.
[465] Smith, R.E.; Lovell, S.C.; Burke, D.F.; Montalvao, R.W.; Blundell, T.L. Andante: reducing side-chain rotamer search space during comparative modeling using environment-specific substitution probabilities. *Bioinformatics,* **2007***, 23*, 1099-1105.
[466] Barnes, H.J.; Arlotto, M.P.; Waterman, M.R. Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in Escherichia coli. *Proceedings of the National Academy of Sciences,* **1991***, 88*, 5597-5601.
[467] Aguiar, M.; Masse, R.; Gibbs, B.F. Regulation of cytochrome P450 by posttranslational modification. *Drug Metab. Rev.,* **2005***, 37*, 379-404.
[468] Wu, Z.; Qiao, J.; Zhang, Z.; Guengerich, F.P.; Liu, Y.; Pei, X. Enhanced bacterial expression of several mammalian cytochrome P450s by codon optimization and chaperone coexpression. *Biotechnol. Lett.,* **2009***, 31*, 1589-1593.
[469] Larson, J.R.; Coon, M.J.; Porter, T.D. Purification and properties of a shortened form of cytochrome P-450 2E1: deletion of the NH2-terminal membrane-insertion signal peptide does not alter the catalytic activities. *Proc. Natl. Acad. Sci. U. S. A.,* **1991***, 88*, 9141-5.
[470] Dong, M.; Yamazaki, H.; Guo, Z.; Guengerich, F.P. Recombinant human cytochrome P450 1A2 and an N-terminal-truncated form: construction, purification, aggregation properties, and interactions with flavodoxin, ferredoxin, and NADPH-cytochrome P450 reductase. *Arch. Biochem. Biophys.,* **1996***, 327*, 11-19.
[471] Sandhu, P.; Baba, T.; Guengerich, F.P. Expression of Modified Cytochrome P450 2C10 (2C9) in *Escherichia coli*, Purification, and Reconstitution of Catalytic Activity. *Arch. Biochem. Biophys.,* **1993***, 306*, 443-450.
[472] Ueng, Y.; Shimada, T.; Yamazaki, H.; Guengerich, F.P. Oxidation of aflatoxin B1 by bacterial recombinant human cytochrome P450 enzymes. *Chem. Res. Toxicol.,* **1995***, 8*, 218-225.
[473] Beeton-Kempen, N. *P450 Biochips: Development of a Protein Microarray Platform for Investigating Cytochrome P450 Clinical Drug Metabolism*. PhD Thesis. University of Cape Town: Cape Town, March **2010**.
[474] Roberts, E.L.; Shu, N.; Howard, M.J.; Broadhurst, R.W.; Chapman-Smith, A.; Wallace, J.C.; Morris, T.; Cronan, J.E.; Perham, R.N. Solution structures of apo and holo biotinyl domains from acetyl coenzyme A carboxylase of Escherichia coli determined by triple-resonance nuclear magnetic resonance spectroscopy. *Biochemistry (N. Y.),* **1999***, 38*, 5045-5053.
[475] Samaddar, M.; Blackburn, J.; Hart, D.; Dyson, M. Protein tag comprising a biotinylation domain and method for increasing solubility and determining folding state. 10/502,581[US 2005/0221308 A1]. US, Google Patents. **2003**.
[476] Fall, R.R. Analysis of microbial biotin proteins. *Methods Enzymol.,* **1979***, 62*, 390-398.
[477] Guengerich, F.P.; Martin, M.V.; Sohl, C.D.; Cheng, Q. Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase. *Nature protocols,* **2009***, 4*, 1245-51.
[478] Gillam, E.M.; Baba, T.; Kim, B.; Ohmori, S.; Guengerich, F.P. Expression of Modified Human Cytochrome P450 3A4 in *Escherichia coli* and Purification and Reconstitution of the Enzyme. *Arch. Biochem. Biophys.,* **1993***, 305*, 123-131.
[479] Idicula-Thomas, S.; Balaji, P.V. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. *Protein science,* **2005***, 14*, 582-592.

[480] Hunt, I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expr. Purif.,* **2005***, 40*, 1-22.
[481] Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.,* **1976***, 72*, 248-254.
[482] Pace, C.N.; Scholtz, J.M. Measuring the conformational stability of a protein. *Protein structure: A practical approach,* **1997***, 2*, 299-321.
[483] Shortle, D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.,* **1996***, 10*, 27-34.
[484] Maves, S.A.; Sligar, S.G. Understanding thermostability in cytochrome P450 by combinatorial mutagenesis. *Protein Science,* **2001***, 10*, 161-168.
[485] Sabat, J.; Stuehr, D.J.; Yeh, S.; Rousseau, D.L. Characterization of the proximal ligand in the P420 form of inducible nitric oxide synthase. *J. Am. Chem. Soc.,* **2009***, 131*, 12186-12192.
[486] Perera, R.; Sono, M.; Sigman, J.A.; Pfister, T.D.; Lu, Y.; Dawson, J.H. Neutral thiol as a proximal ligand to ferrous heme iron: implications for heme proteins that lose cysteine thiolate ligation on reduction. *Proc. Natl. Acad. Sci. U. S. A.,* **2003***, 100*, 3641-3646.
[487] Martinis, S.A.; Blanke, S.R.; Hager, L.P.; Sligar, S.G.; Hui Bon Hoa, G.; Rux, J.J.; Dawson, J.H. Probing the heme iron coordination structure of pressure-induced cytochrome P420cam. *Biochemistry (N. Y.),* **1996***, 35*, 14530-14536.
[488] Wells, A.V.; Li, P.; Champion, P.M.; Martinis, S.A.; Sligar, S.G. Resonance Raman investigations of Escherichia coli-expressed Pseudomonas putida cytochrome P450 and P420. *Biochemistry (N. Y.),* **1992***, 31*, 4384-4393.
[489] Murugan, R.; Mazumdar, S. Role of substrate on the conformational stability of the heme active site of cytochrome P450cam: effect of temperature and low concentrations of denaturants. *JBIC Journal of Biological Inorganic Chemistry,* **2004***, 9*, 477-488.
[490] Egawa, T.; Hishiki, T.; Ichikawa, Y.; Kanamori, Y.; Shimada, H.; Takahashi, S.; Kitagawa, T.; Ishimura, Y. Refolding processes of cytochrome P450cam from ferric and ferrous acid forms to the native conformation. Formations of folding intermediates with non-native heme coordination state. *J. Biol. Chem.,* **2004***, 279*, 32008-32017.
[491] Todorovic, S.; Jung, C.; Hildebrandt, P.; Murgida, D.H. Conformational transitions and redox potential shifts of cytochrome P450 induced by immobilization. *JBIC Journal of Biological Inorganic Chemistry,* **2006***, 11*, 119-127.
[492] Sun, L.; Wang, Z.; Jiang, H.; Tan, X.; Huang, Z. Novel Conformational Transitions of Human Cytochrome P450 2C8 during Thermal and Acid-induced Unfolding. *Chinese Journal of Chemistry,* **2010***, 28*, 1491-1502.
[493] Kumar, S.; Sun, L.; Liu, H.; Muralidhara, B.K.; Halpert, J.R. Engineering mammalian cytochrome P450 2B1 by directed evolution for enhanced catalytic tolerance to temperature and dimethyl sulfoxide. *Protein Eng. Des. Sel.,* **2006***, 19*, 547-554.
[494] Anzenbacher, P.; Hudeček, J.; Stružinský, R. Study of thermal stability of cytochrome P450 by differential scanning calorimetry. *FEBS Lett.,* **1982***, 149*, 208-210.
[495] Koley, A.P.; Buters, J.T.; Robinson, R.C.; Markowitz, A.; Friedman, F.K. Differential mechanisms of cytochrome P450 inhibition and activation by alpha-naphthoflavone. *J. Biol. Chem.,* **1997***, 272*, 3149-3152.
[496] Koley, A.P.; Robinson, R.C.; Markowitz, A.; Friedman, F.K. Drug-drug interactions: effect of quinidine on nifedipine binding to human cytochrome P450 3A4. *Biochem. Pharmacol.,* **1997***, 53*, 455-460.
[497] Davydov, D.R.; Knyushko, T.V.; Hui Bon Hoa, G. High pressure induced inactivation of ferrous cytochrome P-450 LM2 (11B4) CO complex: Evidence for the presence of two conformers in the oligomer. *Biochem. Biophys. Res. Commun.,* **1992***, 188*, 216-221.

[498] Davydov, D.R. High-pressure-induced transitions in microsomal cytochrome P450 2B4 in solution: Evidence for conformational inhomogeneity in the oligomers. *Arch. Biochem. Biophys.,* **1995***, 320*, 330-344.
[499] Fisher, M.T.; Scarlata, S.F.; Sligar, S.G. High-pressure investigations of cytochrome P-450 spin and substrate binding equilibria. *Arch. Biochem. Biophys.,* **1985***, 240*, 456-463.
[500] Bancel, F.; Bec, N.; Ebel, C.; Lange, R. A Central Role for Water in the Control of the Spin State of Cytochrome P-450scc. *Eur. J. Biochem.,* **1997***, 250*, 276-285.
[501] Di Primo, C.; Deprez, E.; Hoa, G.; Douzou, P. Antagonistic effects of hydrostatic pressure and osmotic pressure on cytochrome P-450cam spin transition. *Biophys. J.,* **1995***, 68*, 2056-2061.
[502] Davydov, D.R.; Hui Bon Hoa, G.; Peterson, J.A. Dynamics of protein-bound water in the heme domain of P450BM3 studied by high-pressure spectroscopy: Comparison with P450cam and P450 2B4. *Biochemistry (N. Y.),* **1999***, 38*, 751-761.
[503] Di Primo, C.; Hui Bon Hoa, G.; Douzou, P.; Sligar, S.G. Heme-pocket-hydration change during the inactivation of cytochrome P-450camphor by hydrostatic pressure. *Eur. J. Biochem.,* **1992***, 209*, 583-588.
[504] Hui Bon Hoa, G.; Di Primo, C.; Dondaine, I.; Sligar, S.G.; Gunsalus, I.; Douzou, P. Conformational changes of cytochromes P-450cam and P-450lin induced by high pressure. *Biochemistry (N. Y.),* **1989***, 28*, 651-656.
[505] Viner, R.; Novikov, K.N.; Ritov, V.B.; Kagan, V.E.; Alterman, M.A. Effect of different solubilizing agents on the aggregation state and catalytic activity of two purified rabbit cytochrome P450 isozymes, CYP1A2 (LM4) and CYP2B4 (LM2). *Biochem. Biophys. Res. Commun.,* **1995***, 217*, 886-891.
[506] Scott, E.E.; He, Y.A.; Wester, M.R.; White, M.A.; Chin, C.C.; Halpert, J.R.; Johnson, E.F.; Stout, C.D. An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. *Proc. Natl. Acad. Sci. U. S. A.,* **2003***, 100*, 13196-13201.
[507] Hu, G.; Johnson, E.F.; Kemper, B. CYP2C8 exists as a dimer in natural membranes. *Drug Metab. Dispos.,* **2010***, 38*, 1976-83.
[508] Schoch, G.A.; Yano, J.K.; Wester, M.R.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *J. Biol. Chem.,* **2004***, 279*, 9497-9503.
[509] Hu, G.; Johnson, E.F.; Kemper, B. CYP2C8 exists as a dimer in natural membranes. *Drug Metab. Dispos.,* **2010***, 38*, 1976-1983.
[510] Houston, J.B.; Kenworthy, K.E. In vitro-in vivo scaling of CYP kinetic data not consistent with the classical Michaelis-Menten model. *Drug Metab. Disposition,* **2000***, 28*, 246-254.
[511] Hutzler, J.M.; Tracy, T.S. Atypical kinetic profiles in drug metabolism reactions. *Drug Metab. Disposition,* **2002***, 30*, 355-362.
[512] Schrag, M.L.; Wienkers, L.C. Covalent alteration of the CYP3A4 active site: evidence for multiple substrate binding domains. *Arch. Biochem. Biophys.,* **2001***, 391*, 49-55.
[513] Lin, Y.; Lu, P.; Tang, C.; Mei, Q.; Sandig, G.; Rodrigues, A.D.; Rushmore, T.H.; Shou, M. Substrate Inhibition Kinetics for Cytochrome P450-Catalyzed Reactions. *Drug Metab. Dispos.,* **2001***, 29*, 368-374.
[514] Harlow, G.R.; Halpert, J.R. Analysis of human cytochrome P450 3A4 cooperativity: construction and characterization of a site-directed mutant that displays hyperbolic steroid hydroxylation kinetics. *Proc. Natl. Acad. Sci.,* **1998***, 95*, 6636-6641.
[515] Dabrowski, M.J.; Schrag, M.L.; Wienkers, L.C.; Atkins, W.M. Pyrene⊙ pyrene complexes at the active site of cytochrome P450 3A4: Evidence for a multiple substrate binding site. *J. Am. Chem. Soc.,* **2002***, 124*, 11866-11867.

[516] Hosea, N.A.; Miller, G.P.; Guengerich, F.P. Elucidation of distinct ligand binding sites for cytochrome P450 3A4. *Biochemistry,* **2000***, 39*, 5929-5939.
[517] Fernando, H.; Halpert, J.R.; Davydov, D.R. Resolution of multiple substrate binding sites in cytochrome P450 3A4: the stoichiometry of the enzyme-substrate complexes probed by FRET and Job's titration. *Biochemistry (N. Y.),* **2006***, 45*, 4199-4209.
[518] Schrag, M.L.; Wienkers, L.C. Triazolam substrate inhibition: evidence of competition for heme-bound reactive oxygen within the CYP3A4 active site. *Drug Metab. Disposition,* **2001***, 29*, 70-75.
[519] Galetin, A.; Clarke, S.E.; Houston, J.B. Multisite kinetic analysis of interactions between prototypical CYP3A4 subgroup substrates: midazolam, testosterone, and nifedipine. *Drug Metab. Disposition,* **2003***, 31*, 1108-1116.
[520] Galetin, A.; Clarke, S.E.; Houston, J.B. Quinidine and haloperidol as modifiers of CYP3A4 activity: multisite kinetic model approach. *Drug Metab. Disposition,* **2002***, 30*, 1512-1522.
[521] Houston, J.B.; Galetin, A. Modelling atypical CYP3A4 kinetics: principles and pragmatism. *Arch. Biochem. Biophys.,* **2005***, 433*, 351-60.
[522] Pernecky, S.J.; Larson, J.R.; Philpot, R.M.; Coon, M.J. Expression of truncated forms of liver microsomal P450 cytochromes 2B4 and 2E1 in Escherichia coli: influence of NH2-terminal region on localization in cytosol and membranes. *Proc. Natl. Acad. Sci. U. S. A.,* **1993***, 90*, 2651-2655.
[523] Pernecky, S.J.; Olken, N.M.; Bestervelt, L.L.; Coon, M.J. Subcellular localization, aggregation state, and catalytic activity of microsomal P450 cytochromes modified in the NH2-terminal region and expressed in Escherichia coli. *Arch. Biochem. Biophys.,* **1995***, 318*, 446-456.
[524] Scott, E.E.; Spatzenegger, M.; Halpert, J.R. A truncation of 2B subfamily cytochromes P450 yields increased expression levels, increased solubility, and decreased aggregation while retaining function. *Arch. Biochem. Biophys.,* **2001***, 395*, 57-68.
[525] von Wachenfeldt, C.; Richardson, T.H.; Cosme, J.; Johnson, E.F. Microsomal P450 2C3 is expressed as a soluble dimer in Escherichia coli following modification of its N-terminus. *Arch. Biochem. Biophys.,* **1997***, 339*, 107-114.
[526] Ekins, S.; Ring, B.; Binkley, S.; Hall, S.; Wrighton, S. Autoactivation and activation of the cytochrome P450s. *Int. J. Clin. Pharmacol. Ther.,* **1998***, 36*, 642-651.
[527] Tracy, T.S.; Hutzler, J.M.; Haining, R.L.; Rettie, A.E.; Hummel, M.A.; Dickmann, L.J. Polymorphic variants (CYP2C9* 3 and CYP2C9* 5) and the F114L active site mutation of CYP2C9: effect on atypical kinetic metabolism profiles. *Drug Metab. Disposition,* **2002***, 30*, 385-390.
[528] Tracy, T.S.; Hummel, M.A. Modeling kinetic data from in vitro drug metabolism enzyme experiments. *Drug Metab. Rev.,* **2004***, 36*, 231-242.
[529] Copeland, R.A., Ed. *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, 2nd ed.; Wiley-VCH, Inc.: New York, **2000**.
[530] Gekko, K.; Timasheff, S.N. Thermodynamic and kinetic examination of protein stabilization by glycerol. *Biochemistry (N. Y.),* **1981***, 20*, 4677-4686.
[531] Gekko, K.; Timasheff, S.N. Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. *Biochemistry (N. Y.),* **1981***, 20*, 4667-4676.
[532] Arakawa, T.; Timasheff, S.N. Stabilization of protein structure by sugars. *Biochemistry (N. Y.),* **1982***, 21*, 6536-6544.
[533] Headlam, M.J.; Tuckey, R.C. The effect of glycerol on cytochrome P450scc (CYP11A1) spin state, activity, and hydration. *Arch. Biochem. Biophys.,* **2002***, 407*, 95-102.

[534] Voznesensky, A.I.; Schenkman, J.B. Inhibition of cytochrome-P450 reductase by polyols has an electrostatic nature. *Eur. J. Biochem.,* **1992***, 210*, 741-746.
[535] Marks, B.; Thompson, D.; Volak, L.; Zlokarnik, G and Trubetskoy, O. Characterization of in vitro HTS assays with Vivid® Fluorogenic substrates and recombinant human CYP1A2, 2D6, 2C9, 2C19 and 3A4. Product literature poster. http://tools.lifetechnologies.com/downloads/L0863.pdf (Accessed October 20, **2013**).
[536] Makings, L.R.; Zlokarnik, G. Optical molecular sensors for cytochrome P450 activity. Paris Patent European Patent 1 140 888 B1, June 14, **2003**.
[537] Suman, G.; Jamil, K.; Suseela, K.; Vamsy, M.C. Novel mutations of CYP3A4 in fine needle aspiration cytology samples of breast cancer patients and its clinical correlations. *Cancer Biomarkers,* **2009***, 5*, 33-40.
[538] Sata, F.; Sapone, A.; Elizondo, G.; Stocker, P.; Miller, V.P.; Zheng, W.; Raunio, H.; Crespi, C.L.; Gonzalez, F.J. CYP3A4 allelic variants with amino acid substitutions in exons 7 and 12: Evidence for an allelic variant with altered catalytic activity. *Clin. Pharmacol. Ther.,* **2000***, 67*, 48-56.
[539] Dai, D.; Tang, J.; Rose, R.; Hodgson, E.; Bienstock, R.J.; Mohrenweiser, H.W.; Goldstein, J.A. Identification of Variants of CYP3A4 and Characterization of Their Abilities to Metabolize Testosterone and Chlorpyrifos. *J. Pharmacol. Exp. Ther.,* **2001***, 299*, 825-831.
[540] Eiselt, R.; Domanski, T.; Zibat, A.; Mueller, R.; Presecan-Siedel, E.; Hustert, E.; Zanger, U.; Brockmoller, J.; Klenk, H.; Meyer, U.; Khan, K.; He, Y.; Halpert, J.; Wojnowski, L. Identification and functional characterization of eight CYP3A4 protein variants. *Pharmacogenetics,* **2001***, 11*, 447-458.
[541] Miyazaki, M.; Nakamura, K.; Fujita, Y.; Guengerich, F.P.; Horiuchi, R.; Yamamoto, K. Defective activity of recombinant cytochromes P450 3A4.2 and 3A4.16 in oxidation of midazolam, nifedipine, and testosterone. *Drug Metab. Dispos.,* **2008***, 36*, 2287-91.
[542] Murayama, N.; Nakamura, T.; Saeki, M.; Soyama, A.; Saito, Y.; Sai, K.; Ishida, S.; Nakajima, O.; Itoda, M.; Ohno, Y.; Ozawa, S.; Sawada, J. CYP3A4 gene polymorphisms influence testosterone 6beta-hydroxylation. *Drug Metab. Pharmacokinet.,* **2002***, 17*, 150-156.
[543] Veenstra, D.L.; Blough, D.K.; Higashi, M.K.; Farin, F.M.; Srinouanprachan, S.; Rieder, M.J.; Rettie, A.E. CYP2C9 haplotype structure in European American warfarin patients and association with clinical outcomes. *Clin. Pharmacol. Ther.,* **2005***, 77*, 353-364.
[544] Maekawa, K.; Fukushima-Uesaka, H.; Tohkin, M.; Hasegawa, R.; Kajio, H.; Kuzuya, N.; Yasuda, K.; Kawamoto, M.; Kamatani, N.; Suzuki, K.; Yanagawa, T.; Saito, Y.; Sawada, J. Four novel defective alleles and comprehensive haplotype analysis of CYP2C9 in Japanese. *Pharmacogenet. Genomics,* **2006***, 16*, 497-514.
[545] Maekawa, K.; Harakawa, N.; Sugiyama, E.; Tohkin, M.; Kim, S.R.; Kaniwa, N.; Katori, N.; Hasegawa, R.; Yasuda, K.; Kamide, K. Substrate-dependent functional alterations of seven CYP2C9 variants found in Japanese subjects. *Drug Metab. Dispos.,* **2009***, 37*, 1895-1903.
[546] Aithal, G.; Day, C.; Kesteven, P.; Daly, A. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet,* **1999***, 353*, 717-719.
[547] Sullivan-Klose, T.; Ghanayem, B.; Bell, D.; Zhang, Z.; Kaminsky, L.; Shenfield, G.; Miners, J.; Birkett, D.; Goldstein, J. The role of the CYP2C9-Leu359 allelic variant in the tolbutamide polymorphism. *Pharmacogenetics,* **1996***, 6*, 341-349.
[548] Takanashi, K.; Tainaka, H.; Kobayashi, K.; Yasumori, T.; Hosakawa, M.; Chiba, K. CYP2C9 Ile359 and Leu359 variants: enzyme kinetic study with seven substrates. *Pharmacogenetics,* **2000***, 10*, 95-104.

[549] Rettie, A.E.; Haining, R.L.; Bajpai, M.; Levy, R.H. A common genetic basis for idiosyncratic toxicity of warfarin and phenytoin. *Epilepsy Res.,* **1999***, 35*, 253-255.
[550] Yasar, Ü.; Tybring, G.; Hidestrand, M.; Oscarson, M.; Ingelman-Sundberg, M.; Dahl, M.L.; Eliasson, E. Role of CYP2C9 polymorphism in losartan oxidation. *Drug Metab. Dispos.,* **2001***, 29*, 1051-1056.
[551] Ieiri, I.; Tainaka, H.; Morita, T.; Hadama, A.; Mamiya, K.; Hayashibara, M.; Ninomiya, H.; Ohmori, S.; Kitada, M.; Tashiro, N. Catalytic activity of three variants (Ile, Leu, and Thr) at amino acid residue 359 in human CYP2C9 gene and simultaneous detection using single-strand conformation polymorphism analysis. *Ther. Drug Monit.,* **2000***, 22*, 237-244.
[552] Dickmann, L.; Rettie, A.; Kneller, M.; Kim, R.; Wood, A.; Stein, C.; Wilkinson, G.; Schwarz, U. Identification and functional characterization of a new CYP2C9 variant (CYP2C9*5) expressed among African Americans. *Mol. Pharmacol.,* **2001***, 60*, 382-387.
[553] Guo, Y.; Wang, Y.; Si, D.; Fawcett, P.J.; Zhong, D.; Zhou, H. Catalytic activities of human cytochrome P450 2C9*1, 2C9*3 and 2C9*13. *Xenobiotica,* **2005***, 35*, 853-861.
[554] Liu, Y.; Jeong, H.; Takahashi, H.; Drozda, K.; Patel, S.R.; Shapiro, N.L.; Nutescu, E.A.; Cavallari, L.H. Decreased warfarin clearance associated with the CYP2C9 R150H (* 8) polymorphism. *Clin. Pharmacol. Ther.,* **2012***, 91*, 660-665.
[555] Allabi, A.C.; Gala, J.L.; Horsmans, Y. CYP2C9, CYP2C19, ABCB1 (MDR1) genetic polymorphisms and phenytoin metabolism in a Black Beninese population. *Pharmacogenet. Genomics,* **2005***, 15*, 779-786.
[556] Blaisdell, J.; Jorge-Nebert, L.F.; Coulter, S.; Ferguson, S.S.; Lee, S.J.; Chanas, B.; Xi, T.; Mohrenweiser, H.; Ghanayem, B.; Goldstein, J.A. Discovery of new potentially defective alleles of human CYP2C9. *Pharmacogenetics,* **2004***, 14*, 527-537.
[557] Michaelis, L.; Menten, M.L. Die kinetik der invertinwirkung. *Biochem. Z.,* **1913***, 49*, 352.
[558] Díaz-Mochón, J.J.; Tourniaire, G.; Bradley, M. Microarray platforms for enzymatic and cell-based assays. *Chem. Soc. Rev.,* **2007***, 36*, 449-457.
[559] Chen, C.; Zhu, H. Protein microarrays. *BioTechniques,* **2006***, 40*, 423.
[560] Poetz, O.; Schwenk, J.M.; Kramer, S.; Stoll, D.; Templin, M.F.; Joos, T.O. Protein microarrays: catching the proteome. *Mech. Ageing Dev.,* **2005***, 126*, 161-170.
[561] Paweletz, C.P.; Charboneau, L.; Bichsel, V.E.; Simone, N.L.; Chen, T.; Gillespie, J.W.; Emmert-Buck, M.R.; Roth, M.J.; Petricoin, E.; Liotta, L.A. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene,* **2001***, 20*, 1981-1989.
[562] Ciaccio, M.F.; Wagner, J.P.; Chuu, C.; Lauffenburger, D.A.; Jones, R.B. Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nature methods,* **2010***, 7*, 148-155.
[563] MacBeath, G.; Schreiber, S.L. Printing proteins as microarrays for high-throughput function determination. *Science,* **2000***, 289*, 1760-1763.
[564] Zhu, H.; Snyder, M. Protein chip technology. *Curr. Opin. Chem. Biol.,* **2003***, 7*, 55-63.
[565] Sutandy, F.; Qian, J.; Chen, C.; Zhu, H. Overview of Protein Microarrays. *Current Protocols in Protein Science,* **2013**, 27.1. 1-27.1. 16.
[566] Peluso, P.; Wilson, D.S.; Do, D.; Tran, H.; Venkatasubbaiah, M.; Quincy, D.; Heidecker, B.; Poindexter, K.; Tolani, N.; Phelan, M. Optimizing antibody immobilization strategies for the construction of protein microarrays. *Anal. Biochem.,* **2003***, 312*, 113-124.
[567] Cha, T.; Guo, A.; Zhu, X. Enzymatic activity on a chip: the critical role of protein orientation. *Proteomics,* **2005***, 5*, 416-419.
[568] Nath, N.; Hurst, R.; Hook, B.; Meisenheimer, P.; Zhao, K.Q.; Nassif, N.; Bulleit, R.F.; Storts, D.R. Improving protein array performance: focus on washing and storage conditions. *J. Proteome Res.,* **2008***, 7*, 4475-4482.

[569] Jonkheijm, P.; Weinrich, D.; Schröder, H.; Niemeyer, C.M.; Waldmann, H. Chemical strategies for generating protein biochips. *Angew. Chem. Int. Ed.,* **2008***, 47*, 9618-9647.
[570] Espina, V.; Woodhouse, E.C.; Wulfkuhle, J.; Asmussen, H.D.; Petricoin III, E.F.; Liotta, L.A. Protein microarray detection strategies: focus on direct detection technologies. *J. Immunol. Methods,* **2004***, 290*, 121-133.
[571] Chen, G.Y.; Uttamchandani, M.; Zhu, Q.; Wang, G.; Yao, S.Q. Developing a Strategy for Activity-Based Detection of Enzymes in a Protein Microarray. *ChemBioChem,* **2003***, 4*, 336-339.
[572] Funeriu, D.P.; Eppinger, J.; Denizot, L.; Miyake, M.; Miyake, J. Enzyme family–specific and activity-based screening of chemical libraries using enzyme microarrays. *Nat. Biotechnol.,* **2005***, 23*, 622-627.
[573] Srinivasan, R.; Huang, X.; Ng, S.L.; Yao, S.Q. Activity-Based Fingerprinting of Proteases. *ChemBioChem,* **2006***, 7*, 32-36.
[574] Eppinger, J.; Funeriu, D.P.; Miyake, M.; Denizot, L.; Miyake, J. Enzyme Microarrays: On-Chip Determination of Inhibition Constants Based on Affinity-Label Detection of Enzymatic Activity. *Angewandte Chemie,* **2004***, 116*, 3894-3898.
[575] Mugherli, L., Burchak, O. N., Balakireva, L. A., Thomas, A., Chatelain, F., & Balakirev, M. Y. In Situ Assembly and Screening of Enzyme Inhibitors with Surface-Tension Microarrays. *Angewandte Chemie,* **2009***, 121*, 7775-7780.
[576] Sakakihara, S.; Araki, S.; Iino, R.; Noji, H. A single-molecule enzymatic assay in a directly accessible femtoliter droplet array. *Lab on a Chip,* **2010***, 10*, 3355-3362.
[577] Pompano, R.R.; Liu, W.; Du, W.; Ismagilov, R.F. Microfluidics using spatially defined arrays of droplets in one, two, and three dimensions. *Annu. Rev. Anal. Chem.,* **2011***, 4*, 59-81.
[578] Ehrl, B.N.; Liebherr, R.B.; Gorris, H.H. Single molecule kinetics of horseradish peroxidase exposed in large arrays of femtoliter-sized fused silica chambers. *Analyst,* **2013***, 138*, 4260-4265.
[579] Bistolas, N.; Wollenberger, U.; Jung, C.; Scheller, F.W. Cytochrome P450 biosensors—a review. *Biosensors and Bioelectronics,* **2005***, 20*, 2408-2423.
[580] Sadeghi, S.J.; Fantuzzi, A.; Gilardi, G. Breakthrough in P450 bioelectrochemistry and future perspectives. *Biochim. Biophys. Acta,* **2011***, 1814*, 237-248.
[581] Trubetskoy, O.V.; Gibson, J.R.; Marks, B.D. Highly miniaturized formats for in vitro drug metabolism assays using vivid fluorescent substrates and recombinant human cytochrome P450 enzymes. *J. Biomol. Screen.,* **2005***, 10*, 56-66.
[582] Lee, M.; Park, C.B.; Dordick, J.S.; Clark, D.S. Metabolizing enzyme toxicology assay chip (MetaChip) for high-throughput microscale toxicity analyses. *Proc. Natl. Acad. Sci. U. S. A.,* **2005***, 102*, 983-7.
[583] Schott. Nexterion® Slide H Protein application. (http://www.schott.com/nexterion/) (Accessed Aug 2, **2013**).
[584] Gonzalez, M.; Bagatolli, L.A.; Echabe, I.; Arrondo, J.L.; Argarana, C.E.; Cantor, C.R.; Fidelio, G.D. Interaction of biotin with streptavidin. Thermostability and conformational changes upon binding. *J. Biol. Chem.,* **1997***, 272*, 11288-11294.
[585] Boutell, J.M.; Hart, D.J.; Godber, B.L.; Kozlowski, R.Z.; Blackburn, J.M. Functional protein microarrays for parallel characterisation of p53 mutants. *Proteomics,* **2004***, 4*, 1950-1958.
[586] Beeton-Kempen, N.; Duarte, J.; Shoko, A.; Serufuri, J.; John, T.; Cebon, J.; Blackburn, J. Development of a novel, quantitative protein microarray platform for the multiplexed serological analysis of autoantibodies to cancer-testis antigens. *Int. J. Cancer,* **2014**.

[587] Greenblatt, D.J.; Venkatakrishnan, K.; Harmatz, J.S.; Parent, S.J.; von Moltke, L.L. Sources of variability in ketoconazole inhibition of human cytochrome P450 3A in vitro. *Xenobiotica,* **2010***, 40*, 713-20.
[588] Greenblatt, D.J.; Zhao, Y.; Venkatakrishnan, K.; Duan, S.X.; Harmatz, J.S.; Parent, S.J.; Court, M.H.; von Moltke, L.L. Mechanism of cytochrome P450-3A inhibition by ketoconazole. *J. Pharm. Pharmacol.,* **2011***, 63*, 214-221.
[589] Ray, A.; Johnson, R.; Souyri, A. Dynamic behavior of single glycerol droplets in humid air streams. *Langmuir,* **1989***, 5*, 133-140.
[590] Soap and Detergent Association Glycerine: An Overview. *Glycerine & Oleochemical Division,* **1990***, New York*.
[591] Yim, S.; Yun, C.; Ahn, T.; Jung, H.; Pan, J. A continuous spectrophotometric assay for NADPH-cytochrome P450 reductase activity using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide. *J. Biochem. Mol. Biol.,* **2005***, 38*, 366-9.
[592] Back, J.F.; Oakenfull, D.; Smith, M.B. Increased thermal stability of proteins in the presence of sugars and polyols. *Biochemistry (N. Y.),* **1979***, 18*, 5191-5196.
[593] Street, T.O.; Bolen, D.W.; Rose, G.D. A molecular mechanism for osmolyte-induced protein stability. *Proc. Natl. Acad. Sci. U. S. A.,* **2006***, 103*, 13997-14002.
[594] Nair, O. *Functional effects of cytochrome P450 variants on drug metabolism and adverse drug reactions: developing and extending high throughput P450 protein technology platforms*. MSc Thesis. University of Cape Town: Cape Town, February **2014**.
[595] Dutton, D.R.; Reed, G.A.; Parkinson, A. Redox cycling of resorufin catalyzed by rat liver microsomal NADPH-cytochrome P450 reductase. *Arch. Biochem. Biophys.,* **1989***, 268*, 605-616.
[596] Dutton, D.R.; Parkinson, A. Reduction of 7-alkoxyresorufins by NADPH-cytochrome P450 reductase and its differential effects on their O-dealkylation by rat liver microsomal cytochrome P450. *Arch. Biochem. Biophys.,* **1989***, 268*, 617-629.
[597] Hollfelder, F.; Kirby, A.J.; Tawfik, D.S. Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature,* **1996***, 383*, 60-63.
[598] Hollfelder, F.; Kirby, A.J.; Tawfik, D.S.; Kikuchi, K.; Hilvert, D. Characterization of proton-transfer catalysis by serum albumins. *J. Am. Chem. Soc.,* **2000***, 122*, 1022-1029.
[599] Jackson, R.M.; Gabb, H.A.; Sternberg, M.J. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.,* **1998***, 276*, 265-285.
[600] Lu, H.; Lu, L.; Skolnick, J. Development of unified statistical potentials describing protein-protein interactions. *Biophys. J.,* **2003***, 84*, 1895-1901.
[601] Tovchigrechko, A.; Vakser, I.A. Development and testing of an automated approach to protein docking. *Proteins,* **2005***, 60*, 296-301.
[602] Sano, T.; Cantor, C.R. Intersubunit contacts made by tryptophan 120 with biotin are essential for both strong biotin binding and biotin-induced tighter subunit association of streptavidin. *Proc. Natl. Acad. Sci. U. S. A.,* **1995***, 92*, 3180-3184.
[603] Jones, M.L.; Kurzban, G.P. Noncooperativity of biotin binding to tetrameric streptavidin. *Biochemistry (N. Y.),* **1995***, 34*, 11750-11756.
[604] Loosli, A.; Rusbandi, U.E.; Gradinaru, J.; Bernauer, K.; Schlaepfer, C.W.; Meyer, M.; Mazurek, S.; Novic, M.; Ward, T.R. (Strept) avidin as host for biotinylated coordination complexes: stability, chiral discrimination, and cooperativity. *Inorg. Chem.,* **2006***, 45*, 660-668.
[605] Czerwinski, M.; Krop-Watorek, A.; Wasniowska, K.; Smolarek, D.; Spitalnik, S.L. Construction of an agglutination tool: recombinant Fab fragments biotinylated in vitro. *New biotechnology,* **2009***, 26*, 215-221.

[606] Neish, C.S.; Martin, I.L.; Henderson, R.M.; Edwardson, J.M. Direct visualization of ligand-protein interactions using atomic force microscopy. *Br. J. Pharmacol.,* **2002***, 135*, 1943-1950.
[607] Saito, Y.; Hanioka, N.; Maekawa, K.; Isobe, T.; Tsuneto, Y.; Narimatsu, S.; Sawada, J. Functional analysis of three CYP1A2 variants found in a Japanese population. *Pharmacology,* **2005***, 33*, 1905-1910.
[608] Murayama, N.; Soyama, A.; Saito, Y.; Nakajima, Y.; Komamura, K.; Ueno, K.; Kamakura, S.; Kitakaze, M.; Kimura, H.; Goto, Y.; Saitoh, O.; Katoh, M.; Ohnuma, T.; Kawai, M.; Sugai, K.; Ohtsuki, T.; Suzuki, C.; Minami, N.; Ozawa, S.; Sawada, J. Six Novel Nonsynonymous CYP1A2 Gene Polymorphisms : Catalytic Activities of the Naturally Occurring Variant Enzymes. *Pharmacology,* **2004***, 308*, 300-306.
[609] Iwasaki, M.; Yoshimura, Y.; Asahi, S.; Saito, K.; Sakai, S.; Morita, S.; Takenaka, O.; Inoda, T.; Kashiyama, E.; Aoyama, A.; Nakabayashi, T.; Omori, S.; Kuwabara, T.; Izumi, T.; Nakamura, K.; Takanaka, K.; Nakayama, Y.; Takeuchi, M.; Nakamura, H.; Kametani, S.; Terauchi, Y.; Hashizume, T.; Nagayama, S.; Kume, T.; Achira, M.; Kawai, H.; Kawashiro, T.; Nakamura, A.; Nakai, Y.; Kagayama, A.; Shiraga, T.; Niwa, T.; Yoshimura, T.; Morita, J.; Ohsawa, F.; Tani, M.; Osawa, N.; Ida, K.; Noguchi, K. Functional characterization of single nucleotide polymorphisms with amino acid substitution in CYP1A2, CYP2A6, and CYP2B6 found in the Japanese population. *Drug Metab. Pharmacokinet.,* **2004***, 19*, 444-452.
[610] Al Koudsi, N.; Ahluwalia, J.S.; Lin, S.K.; Sellers, E.M.; Tyndale, R.F. A novel CYP2A6 allele (CYP2A6*35) resulting in an amino-acid substitution (Asn438Tyr) is associated with lower CYP2A6 activity in vivo. *Pharmacogenomics J,* **2009***, 9*, 274-282.
[611] Mwenifumbo, J.C.; Al Koudsi, N.; Ho, M.K.; Zhou, Q.; Hoffmann, E.B.; Sellers, E.M.; Tyndale, R.F. Novel and established CYP2A6 alleles impair in vivo nicotine metabolism in a population of Black African descent. *Hum. Mutat.,* **2008***, 29*, 679-688.
[612] Kitagawa, K.; Kunugita, N.; Kitagawa, M.; Kawamoto, T. CYP2A6*6, a novel polymorphism in cytochrome p450 2A6, has a single amino acid substitution (R128Q) that inactivates enzymatic activity. *J. Biol. Chem.,* **2001***, 276*, 17830-17835.
[613] Yamano, S.; Tatsuno, J.; Gonzalez, F. The CYP2A3 gene product catalyzes coumarin 7-hydroxylation in human liver microsomes. *Biochemistry,* **1990***, 29*, 1322-1329.
[614] Tiong, K.H.; Yiap, B.C.; Tan, E.L.; Ismail, R.; Ong, C.E. Functional Characterization of Cytochrome P450 2A6 Allelic Variants CYP2A6*15, CYP2A6*16, CYP2A6*21, and CYP2A6*22. *Drug Metab. Dispos.,* **2010***, 38*, 745-751.
[615] Ho, M.K.K.; Mwenifumbo, J.C.; Zhao, B.; Gillam, E.M.; Tyndale, R.F. A novel CYP2A6 allele, CYP2A6*23, impairs enzyme function in vitro and in vivo and decreases smoking in a population of Black-African descent. *Pharmacogenet. Genomics,* **2008***, 18*, 67-75.
[616] Fukami, T.; Nakajima, M.; Yoshida, R.; Tsuchiya, Y.; Fujiki, Y.; Katoh, M.; McLeod, H.L.; Yokoi, T. A novel polymorphism of human CYP2A6 gene CYP2A6*17 has an amino acid substitution (V365M) that decreases enzymatic activity in vitro and in vivo. *Clin. Pharmacol. Ther.,* **2004***, 76*, 519-527.
[617] Xu, C.; Rao, Y.S.; Xu, B.; Hoffmann, E.; Jones, J.; Sellers, E.M.; Tyndale, R.F. An in Vivo Pilot Study Characterizing the New CYP2A6*7, *8, and *10 Alleles. *Biochem. Biophys. Res. Commun.,* **2002***, 290*, 318-324.
[618] Yoshida, R.; Nakajima, M.; Watanabe, Y.; Kwon, J.; Yokoi, T. Genetic polymorphisms in human CYP2A6 gene causing impaired nicotine metabolism. *Br. J. Clin. Pharmacol.,* **2002***, 54*, 511-517.
[619] Oscarson, M.; McLellan, R.; Gullstén, H.; Agúndez, J.; Benítez, J.; Rautio, A.; Raunio, H.; Pelkonen, O.; Ingelman-Sundberg, M. Identification and characterisation of novel

polymorphisms in the CYP2A locus: implications for nicotine metabolism. *FEBS Lett.,* **1999***, 460*, 321-327.

[620] Jinno, H.; Tanaka-Kagawa, T.; Ohno, A.; Makino, Y.; Matsushima, E.; Hanioka, N.; Ando, M. Functional Characterization of Cytochrome P450 2B6 Allelic Variants. *Drug Metab. Dispos.,* **2003***, 31*, 398-403.

[621] Honda, M.; Muroi, Y.; Tamaki, Y.; Saigusa, D.; Suzuki, N.; Tomioka, Y.; Matsubara, Y.; Oda, A.; Hirasawa, N.; Hiratsuka, M. Functional characterization of CYP2B6 allelic variants in demethylation of antimalarial artemether. *Drug Metab. Dispos.,* **2011***, 39*, 1860-1865.

[622] Klein, K.; Lang, T.; Saussele, T.; Barbosa-Sicard, E.; Schunck, W.H.; Eichelbaum, M.; Schwab, M.; Zanger, U.M. Genetic variability of CYP2B6 in populations of African and Asian origin: allele frequencies, novel functional variants, and possible implications for anti-HIV therapy with efavirenz. *Pharmacogenet. Genomics,* **2005***, 15*, 861-873.

[623] Lang, T.; Klein, K.; Richter, T.; Zibat, A.; Kerb, R.; Eichelbaum, M.; Schwab, M.; Zanger, U.M. Multiple Novel Nonsynonymous CYP2B6 Gene Polymorphisms in Caucasians : Demonstration of Phenotypic Null Alleles. *Pharmacology,* **2004***, 311*, 34-43.

[624] Wang, J.; Sönnerborg, A.; Rane, A.; Josephson, F.; Lundgren, S.; Ståhle, L.; Ingelman-Sundberg, M. Identification of a novel specific CYP2B6 allele in Africans causing impaired metabolism of the HIV drug efavirenz. *Pharmacogenet. Genomics,* **2006***, 16*, 191.

[625] Soyama, A.; Saito, Y.; Hanioka, N.; Murayama, N.; Nakajima, O.; Katori, N.; Ishida, S.; Sai, K.; Ozawa, S.; Sawada, J. Non-synonymous Single Nucleotide Alterations Found in the CYP2C8 Gene Result in Reduced in Vitro Paclitaxel Metabolism. *Biol. Pharm. Bull.,* **2001***, 24*, 1427-1430.

[626] Gao, Y.; Liu, D.; Wang, H.; Zhu, J.; Chen, C. Functional characterization of five CYP2C8 variants and prediction of CYP2C8 genotype-dependent effects on in vitro and in vivo drug-drug interactions. *Xenobiotica,* **2010***, 40*, 467-475.

[627] Hichiya, H.; Tanaka-Kagawa, T.; Soyama, A.; Jinno, H.; Koyano, S.; Katori, N.; Matsushima, E.; Uchiyama, S.; Tokunaga, H.; Kimura, H.; Minami, N.; Katoh, M.; Sugai, K.; Goto, Y.; Tamura, T.; Yamamoto, N.; Ohe, Y.; Kunitoh, H.; Nokihara, H.; Yoshida, T.; Minami, H.; Saijo, N.; Ando, M.; Ozawa, S.; Saito, Y.; Sawada, J. Functional characterization of five novel CYP2C8 variants, G171S, R186X, R186G, K247R, and K383N, found in a Japanese population. *Drug Metab. Dispos.,* **2005***, 33*, 630-636.

[628] Hanioka, N.; Matsumoto, K.; Saito, Y.; Narimatsu, S. Functional characterization of CYP2C8. 13 and CYP2C8. 14: catalytic activities toward paclitaxel. *Basic Clin. Pharmacol. Toxicol.,* **2010***, 107*, 565-569.

[629] Jiang, H.; Zhong, F.; Sun, L.; Feng, W.; Huang, Z.X.; Tan, X. Structural and functional insights into polymorphic enzymes of cytochrome P450 2C8. *Amino Acids,* **2011***, 40*, 1195-1204.

[630] Dai, D.; Zeldin, D.; Blaisdell, J.; Chanas, B.; Coulter, S.; Ghanayem, B.; Goldstein, J. Polymorphisms in human CYP2C8 decrease metabolism of the anticancer drug paclitaxel and arachidonic acid. *Pharmacogenetics,* **2001***, 11*, 597-607.

[631] Guo, Y.; Zhang, Y.; Wang, Y.; Chen, X.; Si, D.; Zhong, D.; Fawcett, J.P.; Zhou, H. Role of CYP2C9 and its variants (CYP2C9* 3 and CYP2C9* 13) in the metabolism of lornoxicam in humans. *Drug Metab. Dispos.,* **2005***, 33*, 749-753.

[632] DeLozier, T.C.; Lee, S.C.; Coulter, S.J.; Goh, B.C.C.; Goldstein, J.A. Functional characterization of novel allelic variants of CYP2C9 recently discovered in southeast Asians. *J. Pharmacol. Exp. Ther.,* **2005***, 315*, 1085-1090.

[633] Yin, T.; Maekawa, K.; Kamide, K.; Saito, Y.; Hanada, H.; Miyashita, K.; Kokubo, Y.; Akaiwa, Y.; Otsubo, R.; Nagatsuka, K.; Otsuki, T.; Horio, T.; Takiuchi, S.; Kawano, Y.;

Minematsu, K.; Naritomi, H.; Tomoike, H.; Sawada, J.; Miyata, T. Genetic variations of CYP2C9 in 724 Japanese individuals and their impact on the antihypertensive effects of losartan. *Hypertension research,* **2008***, 31*, 1549-1557.

[634] Veronese, F.M. Peptide and protein PEGylation: a review of problems and solutions. *Biomaterials,* **2001***, 22*, 405-17.

[635] Rettie, A.E.; Wienkers, L.C.; Gonzalez, F.J.; Trager, W.F.; Korzekwa, K.R. Impaired (S)-warfarin metabolism catalysed by the R144C allelic variant of CYP2C9. *Pharmacogenet. Genomics,* **1994***, 4*, 39-42.

[636] Crespi, C.L.; Miller, V.P. The R144C change in the CYP2C9* 2 allele alters interaction of the cytochrome P450 with NADPH: cytochrome P450 oxidoreductase. *Pharmacogenetics,* **1997***, 7*, 203.

[637] Iida, I.; Miyata, A.; Arai, M.; Hirota, M.; Akimoto, M.; Higuchi, S.; Kobayashi, K.; Chiba, K. Catalytic roles of CYP2C9 and its variants (CYP2C9* 2 and CYP2C9* 3) in lornoxicam 5′-hydroxylation. *Drug Metab. Dispos.,* **2004***, 32*, 7-9.

[638] Allabi, A.C.; Gala, J.L.; Horsmans, Y.; Babaoglu, M.O.; Bozkurt, A.; Heusterspreute, M.; Yasar, U. Functional impact of CYP2C95, CYP2C96, CYP2C98, and CYP2C911 in vivo among black Africans. *Clin. Pharmacol. Ther.,* **2004***, 76*, 113-118.

[639] Herman, D.; Dolzan, V.; Ingelman-Sundberg, M. Characterization of the novel defective CYP2C9* 24 allele. *Drug Metab. Dispos.,* **2007***, 35*, 831-834.

[640] Hanioka, N.; Hanioka, N.; Tsuneto, Y.; Saito, Y.; Sumada, T.; Maekawa, K.; Saito, K.; Sawada, J.; Narimatsu, S. Functional characterization of two novel CYP2C19 variants (CYP2C19* 18 and CYP2C19* 19) found in a Japanese population. *Xenobiotica,* **2007***, 37*, 342-355.

[641] Hanioka, N.; Tsuneto, Y.; Saito, Y.; Maekawa, K.; Sawada, J.; Narimatsu, S. Influence of CYP2C19* 18 and CYP2C19* 19 Alleles on Omeprazole 5-Hydroxylation: In vitro Functional Analysis of Recombinant Enzymes Expressed in Saccharomyces cerevisiae. *Basic Clin. Pharmacol. Toxicol.,* **2008***, 102*, 388-393.

[642] Ibeanu, G.C.; Blaisdell, J.; Ferguson, R.J.; Ghanayem, B.I.; Brøsen, K.; Benhamou, S.; Bouchardy, C.; Wilkinson, G.R.; Dayer, P.; Goldstein, J.A. A Novel Transversion in the Intron 5 Donor Splice Junction ofCYP2C19 and a Sequence Polymorphism in Exon 3 Contribute to the Poor Metabolizer Phenotype for the Anticonvulsant DrugS-Mephenytoin. *J. Pharmacol. Exp. Ther.,* **1999***, 290*, 635-640.

[643] Lee, S.; Kim, W.; Kim, H.; Shon, J.; Lee, S.S. Identification of New CYP2C19 Variants Exhibiting Decreased Enzyme Activity in the Metabolism of S-Mephenytoin and Omeprazole. *Pharmacology,* **2009***, 37*, 2262-2269.

[644] Zhang, Z.; Wang, L.; Gao, Y.; Zhang, J.; Zhenirovskyy, M.; Alexov, E. Predicting folding free energy changes upon single point mutations. *Bioinformatics,* **2012***, 28*, 664-71.

[645] Ibeanu, G.C.; Blaisdell, J.; Ghanayem, B.I.; Beyeler, C.; Benhamou, S.; Bouchardy, C.; Wilkinson, G.R.; Dayer, P.; Daly, A.K.; Goldstein, J.A. An additional defective allele, CYP2C19* 5, contributes to the S-mephenytoin poor metabolizer phenotype in Caucasians. *Pharmacogenetics,* **1998***, 8*, 129.

[646] Morita J, Kobayashi K, Wanibuchi A, Kimura M, Irie S, Ishizaki T, Chiba K. A novel single nucleotide polymorphism (SNP) of the CYP2C19 gene in a Japanese subject with lowered capacity of mephobarbital 4′-hydroxylation. *Drug Metab. Pharmacokinet.,* **2004***, 19*, 236-238.

[647] Sakuyama, K.; Sasaki, T.; Ujiie, S.; Obata, K.; Mizugaki, M.; Ishikawa, M.; Hiratsuka, M. Functional characterization of 17 CYP2D6 allelic variants (CYP2D6.2, 10, 14A-B, 18, 27, 36, 39, 47-51, 53-55, and 57). *Drug Metab. Dispos.,* **2008***, 36*, 2460-2467.

[648] Oscarson, M.; Hidestrand, M.; Johansson, I.; Ingelman-Sundberg, M. A combination of mutations in the CYP2D6*17 (CYP2D6Z) allele causes alterations in enzyme function. *Mol. Pharmacol.,* **1997***, 52*, 1034-40.
[649] Zhang, W.Y.; Tu, Y.B.; Haining, R.L.; Yu, A.M. Expression and functional analysis of CYP2D6. 24, CYP2D6. 26, CYP2D6. 27, and CYP2D7 isozymes. *Drug Metab. Disposition,* **2009***, 37*, 1-4.
[650] Evert, B.; Griese, E.U.; Eichelbaum, M. A missense mutation in exon 6 of the CYP2D6 gene leading to a histidine 324 to proline exchange is associated with the poor metabolizer phenotype of sparteine. *Naunyn Schmiedebergs Arch. Pharmacol.,* **1994***, 350*, 434-439.
[651] Evert, B.; Eichelbaum, M.; Haubruck, H.; Zanger, U.M. Functional properties of CYP2D6 1 (wild-type) and CYP2D6 7 (His324Pro) expressed by recombinant baculovirus in insect cells. *Naunyn Schmiedebergs Arch. Pharmacol.,* **1997***, 355*, 309-318.
[652] Klein, K.; Tatzel, S.; Raimundo, S.; Saussele, T.; Hustert, E.; Pleiss, J.; Eichelbaum, M.; Zanger, U.M. A natural variant of the heme-binding signature (R441C) resulting in complete loss of function of CYP2D6. *Drug Metab. Disposition,* **2007***, 35*, 1247-1250.
[653] Hu, Y.; Oscarson, M.; Johansson, I.; Yue, Q.Y.; Dahl, M.L.; Tabone, M.; Arincò, S.; Albano, E.; Ingelman-Sundberg, M. Genetic polymorphism of human CYP2E1: characterization of two variant alleles. *Mol. Pharmacol.,* **1997***, 51*, 370-376.
[654] Hanioka, N.; Tanaka-Kagawa, T.; Miyata, Y.; Matsushima, E.; Makino, Y.; Ohno, A.; Yoda, R.; Jinno, H.; Ando, M. Functional characterization of three human cytochrome p450 2E1 variants with amino acid substitutions. *Xenobiotica,* **2003***, 33*, 575-586.
[655] Fairbrother, K.S.; Grove, J.; de Waziers, I.; Steimel, D.T.; Day, C.P.; Crespi, C.L.; Daly, A.K. Detection and characterization of novel polymorphisms in the CYP2E1 gene. *Pharmacogenetics,* **1998***, 8*, 543.
[656] Hsieh, K.; Lin, Y.; Cheng, C.; Lai, M.; Lin, M.; Siest, J.; Huang, J. Novel mutations of CYP3A4 in Chinese. *Drug Metab. Dispos.,* **2001***, 29*, 268-273.

An Expanded, Unified Substrate Recognition Site Map for Mammalian Cytochrome P450s: Analysis of Molecular Interactions Between 15 Mammalian CYP450 Isoforms and 868 Substrates

Alexander Zawaira[&†], Lim Yen Ching[‡], Lauren Coulson[*], Jonathan Blackburn[*†] and Yap Chun Wei[‡†]

[&]*Gene Expression & Biophysics Group, Synthetic Biology - ERA. Building 20, CSIR Biosciences. Meiring Naude Road. Brummeria, Pretoria 0001. South Africa, *Division of Medical Biochemistry, Institute of Infectious Disease & Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa, [‡]Department of Pharmacy, Faculty of Science, National University of Singapore, Blk S7-02-06, 18 Science Drive 4, Singapore 117543*

Abstract: The original map of mammalian cytochrome P450 (CYP450) residues involved in substrate recognition was prepared for the CYP2 family by Gotoh in 1992 by manual alignment of mammalian CYP450 residues with substrate recognition site (SRS) residues manually delimited from a bacterial cytochrome P450-substrate complex. Using modern structural bioinformatics tools, we have identified CYP450-ligand interactions in mammalian complexes to create a "X-ray structures" SRS map. In a parallel approach, we have built a "docking" SRS map by successful docking of 868 known substrates of 10 mammalian CYP450 isoforms and analysis of contacts made in docking solutions. We subsequently combined these maps to create a unified description of SRSs. The new map largely agrees with the original map by Gotoh with the six original SRS regions appearing in similar locations along the CYP450 sequence as in Gotoh's map. However, important differences also occur: Two new SRS regions appear before SRS1 and we have assigned them as SRS1'a and SRS1'b; SRS1 is much bigger in our map than in Gotoh's (49 aligned positions versus 28); & SRS2 and SRS3 are co-joined in our map to give a single large SRS region (60 aligned positions) we have designated as SRS(2,3), in contrast to the 9 and 10 aligned positions individually covered by SRS2 and SRS3 respectively in Gotoh's original map. These differences result in the SRS zone covering 33 % of the mammalian CYP450 sequence in our map as opposed to 16 % in Gotoh's map.

Keywords: Cytochrome P450, contact residues, docking complex, Gotoh's map, site of metabolism, substrate recognition site.

INTRODUCTION

The Cytochromes P450 (CYP450, EC 1.14.14.1) are a superfamily of heme-thiolate monooxygenases that catalyze the oxidation of a wide variety of hydrophobic endogenous and xenobiotic substrates [1] and CYP450 homologs have been identified from all lineages of life including eukaryotes and bacteria [2] and viruses [3]. CYP450 genes have been organised into families and subfamilies on the basis of sequence identity – members of the same family share more than 40 % amino acid identity whereas members of the same sub-family share more than 55 % amino acid identity [4]. The nomenclature system for CYP450 genes and proteins is as follows: the abbreviation 'CYP', followed by an Arabic numeral indicating the gene family, a capital Roman letter indicating the sub-family and another Arabic numeral for the individual gene [4].

Enzymes in CYP450 family 1 through 3 are the major xenobiotic metabolizing enzymes in mammals and are characterized by broad substrate specificities [5, 6] – for example, a search of the PubChem database shows that CYP3A4 has ~ 400 known xenobiotic substrates, CYP2D6 has ~ 210, CYP2C9 has ~ 160, CYP1A2 has ~ 65, CYP2A6 has ~ 30, CYP2B4 has ~ 10 and CYP2C5 has ~ 5. This evolutionary strategy has allowed few catalysts to handle/process relatively large numbers of exogenous molecules encountered by mammals. Notably, whilst the substrate specificities of these xenobiotic metabolizing CYP450 enzymes partially overlap, they are also distinguishable from each other [5, 7, 8].

[†]Address correspondence to these authors at the Gene Expression & Biophysics Group, Synthetic Biology – ERA. Building 20, CSIR Biosciences. Meiring Naude Road. Brummeria, Pretoria 0001. South Africa; Tel: +2712 841 33 46; Fax: 00 27 12 841 4790; E-mails: azawaira@gmail.com; azawaira@yahoo.com; Institute of Infectious Disease & Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa; Tel: +27 21 406 6071; Fax: 00 27 21 650 4833; E-mail: jonathan.blackburn@uct.ac.za; Department of Pharmacy, Faculty of Science, National University of Singapore, Blk S7-02-06, 18 Science Drive 4, Singapore 117543. Tel: +65 65165971; Fax: 00 65 6779 1554; E-mail: phayapc@nus.edu.sg

An important goal in CYP450 research has been the elucidation of the molecular mechanisms underlying the broad specificities of the mammalian xenobiotic metabolizing CYP450 systems and an important step towards this is the identification of parts of the CYP450 protein that are involved in the recognition/binding of substrates [5, 7]. A comprehensive delimitation of mammalian substrate recognition site (SRS) amino acid residues was carried out by Gotoh [5], who inferred a map of SRSs in mammalian CYP450 proteins by performing group-to-group alignment (the modern day equivalent of profile-profile alignment) of a group of mammalian CYP450 sequences to a group of bacterial CYP450 sequences that included *Pseudomonas putida* CYP450 101A (also known as P450cam); CYP450 101A was the only CYP whose substrate-binding residues were known at the time – these had been determined by X-ray crystallography of a substrate-bound form [2, 5].

Gotoh's work stimulated experimental studies that sought to investigate the existence of the SRSs he had predicted in mammalian CYP450. Generally, the strategy employed in these studies has involved identification of amino acid differences that lie within Gotoh's SRSs and that are principally responsible for characteristic functional differences between closely-related enzymes [7]. Such amino acids have been identified on the basis of either functional inter-conversions achieved by inter-changing specific residues between closely-related enzymes or by studying the functional properties of chimeras of closely-related enzymes [7-9]. Examples of these studies include work from Halpert's laboratory which provided support for the existence in CYP2B1 of five out the six SRSs predicted by Gotoh – the only exception being SRS3 [10, 11]. Halpert's laboratory also provided evidence for the existence in CYP3A4 of SRS2 [12], SRS1 [13] & SRS5 [14] and provided agreement between substrate recognition studies with CYP450 family 2 and family 3 enzymes [9], whilst Conley *et al.* provided evidence for the existence of SRS1 in porcine aromatase cytochrome P450 (P450arom) [15]. Together, these works have validated Gotoh's prediction of the six SRSs and have established that these regions are involved in substrate-binding and contribute to the

1389-2002/11 $58.00+.00

differences in substrate specificities observed between isoforms [14, 15].

However, other experimental studies have identified amino acid differences which lie outside Gotoh's SRSs but which are also apparently responsible for characteristic functional differences between closely-related CYP450 enzymes [7, 8, 16, 17]. Whilst almost all of the residues identified by analysis of closely-related CYP450 isoforms as being critical for substrate specificity are located in or near SRSs [7], the identification of residues outside Gotoh's SRSs that influence substrate specificity suggests that there is scope to re-evaluate and extend Gotoh's original SRS map. For example, von Weymarn *et al*. showed that substitution of the amino acid triplet Ser407, Asn417 and Ala419 (all lying between SRS5 and SRS6) protects CYP2B1 against mechanism-based inactivation by acetylenic compounds [16]; Biagini *et al*. showed that the non-SRS residues Val4 (occurs before SRS1) and Phe187 (occurs closer to SRS2, between SRS1 and SRS2) have a role in testosterone hydroxylation by CYP2C11 [17]; He *et al.* showed that the non-SRS residues His120 (occurs just C-terminal to SRS1) and Pro221 (occurs between SRS2 and SRS3) have a role in progesterone hydroxylation by CYP2B5 [7]; & Ibeanu *et al.* showed that the non-SRS residues Pro220 and Thr221 (both lying in the F-G loop between SRS2 and SRS3) are important determinants of the omeprazole hydroxylase activity of CYP2C19 [8].

Gotoh's SRS map undoubtedly provided a good framework for researchers to contextualise their experimental results. We note however that the strategy of studying characteristic functional differences between closely-related CYP450 enzymes had been used prior to Gotoh's work to delimit specificity determining residue positions in CYP450 sequences. For example, Aoyama *et al*. showed that CYP2B1 positions Leu58 and Ile114 modulate the stereo- and regio-specificity of steroid hydroxylation by the enzyme [18]; of these, Ile114 lies within Gotoh's SRS1 but Leu58 occurs before SRS1 in Gotoh's map. Linberg & Negishi showed that positions 117, 209 and 365 in mouse P450coh (now called CYP2A5) modulate substrate specificity [19, 20]; of these, residue 117 lies in Gotoh's SRS1, residue 209 lies in Gotoh's SRS2 and residue 365 lies in Gotoh's SRS5. Kedzie *et al*. showed that CYP2B1 position 478 modulates androstenedione 16-beta hydroxylase activity [21]; this residue lies in Gotoh's SRS6.

Mechanism-based inhibitors were also used to delimit substrate recognition sites prior to Gotoh's work. For example, using mechanism based inhibitors Yun *et al.* discovered that CYP1A2 positions 175-184 (peptide FQELMAAVGR) and 67-78 (peptide L(S)QQYGDVLQIR) include substrate recognition/binding regions [22]; of these, residues 175-184 comprise a non-SRS region lying between Gotoh's SRS1 and SRS2, whilst residues 67-78 comprise a non-SRS region occurring before Gotoh's SRS1. More recently, methods using mechanism-based inhibitors have continued to yield insights into substrate recognition in CYP450, an example being the work of Lin *et al.* which demonstrated the role of the CYP2B1 residue Thr205 in androgen 16-beta hydroxylation activity [23]; this residue lies within Gotoh's SRS2.

A sizeable number of mammalian structures have been deposited into the RCSB Protein Data Bank (www.pdb.org) since Gotoh published his map – in 2011 there are around 40 substrate/inhibitor-bound mammalian CYP450 structures and around 10 substrate/inhibitor-free structures. Examples of substrate-bound structures include: CYP2C5 in complex with the drug Diclofenac [24]; CYP2C9 in complex with the anti-coagulant Warfarin [25]; CYP2E1 in complex with fatty-acid substrate analogs [26]; CYP2R1 in complex with vitamin D3 [27]; CYP2A6 in complex with Methoxsalen and with Coumarin [28]; CYP1A2 in complex with alpha-naphthoflavone [29]; CYP3A4 in complex with Metyrapone and Progesterone [25]; human brain CYP46A1 in complex with cholesterol substrate [30]; human placental aromatase (CYP19A1) in complex with Androstenedione [31]; human Prostacyclin synthase (CYP8A1) in complex with the inhibitor Minoxidil [32]; and rat CYP24A1 docked with 1α,25-dihydroxyvitamin D_3 [33].

Furthermore, since publication of Gotoh's SRS map, numerous structural bioinformatics tools which enable the precise delimitation of interactions between protein and ligand in protein-ligand complexes have been developed. These tools include the program Ligplot - which delimits hydrophobic and hydrogen-bond interactions between protein amino acid side chains and the bound ligand [34] - and Zamora's energetics functions/molecular interaction fields which calculate ligand-protein interaction energies [35]. Powerful docking tools such as FlexX, Gold, Glide *etc* have also been developed which can be used for example to generate *in silico* CYP450-substrate complexes that can then be evaluated on the basis of considerations of the likely site-of-metabolism [36]. The site of metabolism may be experimentally determined or can today be predicted using powerful programs such as Metasite [37-39]. Furthermore, molecular dynamics tools and systems have also been developed specifically for CYP450s that can be applied to elucidate the roles of different parts of CYP450 molecules (especially the F-G loop) in substrate recognition [40-42].

Gotoh's SRS map continues to be a highly relevant tool in CYP450 research today. For example, it can be used to guide prioritisation of mutants when studying the effect of clinically-relevant CYP450 polymorphic variation on drug metabolism or on adverse drug reactions using reconstituted recombinant protein systems. In addition, when the importance of the different positions within SRS regions as determinants of CYP450 substrate specificity is scored, say, by using the frequency with which they are used in CYP450-ligand complexes, such a map can be used to inform protein engineering studies of CYP450 substrate specificity. However, the development of such a quantitative position-scored SRS map has to be preceded, among other things, by the definition of as complete an SRS map as possible. In the present work, we have therefore applied a range of new computational tools to the wealth of new structural data with the goal of creating a more complete mammalian CYP450 substrate recognition site map which can, amongst others, explain aspects of the substrate specificity data that did not otherwise accord with Gotoh's original SRSs.

MATERIALS AND METHODS

Structures and Sequences

Structures of ligand-bound mammalian CYP450 proteins were retrieved from the RCSB Protein Data Bank [43]. The structure PDB codes are as follows. CYP1A2 (*Homo sapiens*): 2HI4. CYP2A6 (*Homo sapiens*): 3EBS, 1Z11, 2FDU, 2FDV, 2FDW and 2FDY. CYP2A13 (*Homo sapiens*): 2P85. CYP2B4 (*Oryctolagus cuniculus*): 2Q6N, 2BDM, 1SUO, 3KW4 and 3ME6. CYP2B6 (*Homo sapiens*): 3IBD. CYP2C5 (*Oryctolagus cuniculus*): 1N6B and 1NR6. CYP2C8 (*Homo sapiens*): 2NNJ, 2NNI, 2NNH and 2VN0. CYP2C9 (*Homo sapiens*): 1OG5 and 1R9O. CYP2E1 (*Homo sapiens*): 3GPH, 3KOH, 3LC4, 3E4E and 3E6I. CYP2R1 (*Homo sapiens*): 3C6G, 3DL9 and 3CZH. CYP3A4 (*Homo sapiens*): 1W0F, 1W0G, 2J0D, 2V0M and 3NXU. CYP46A1 (*Homo sapiens*): 2Q9F, 3MDM, 3MDR, 3MDT and 3MDV. CYP19A1 (*Homo sapiens*): 3EQM. CYP8A1 (*Homo sapiens*): 3B6H. CYP11A1 (*Homo sapiens*): 3NA0. We note that 3EBS is a structure of the I208S/I300F/G301A/S369G mutant of CYP2A6 bound to the drug phenacetin. We also note that 1NR6 is a structure of the N202H/R206E/I207L/S209G/S210T mutant of CYP2C5 bound to the drug diclofenac.

Sequences were retrieved from the Universal Protein Resource Knowledgebase (UniProtKB/Swiss-Prot) database via the European Bioinformatics Institute website [44]. The accession numbers were according to the UniProtKB database unless otherwise stated. The sequence accession numbers used herein are as follows. CYP1A2: P05177 (*Homo sapiens*), CYP2A6: P11509 (*Homo sapiens*),

CYP2A13: Q16696 (*Homo sapiens*), CYP2B4: P00178 (*Oryctolagus cuniculus*), CYP2B6: P20813 (*Homo sapiens*), CYP2C5: P00179 (*Oryctolagus cuniculus*), CYP2C8: P10632 (*Homo sapiens*), CYP2C9: P11712 (*Homo sapiens*), CYP2E1: P05181 (*Homo sapiens*), CYP2R1: Q6VVX0 (*Homo sapiens*), CYP3A4: P08684 (*Homo sapiens*), CYP46A1: Q9Y6A2 (*Homo sapiens*), CYP19A1: P11511 (*Homo sapiens*), CYP8A1: Q16647 (*Homo sapiens*) and CYP11A1: P05108 (*Homo sapiens*).

This study includes complexes of substrate- and inhibitor-bound CYP450s. Substrate complexes are included only if the substrate presents a known site of metabolism, or one predicted by MetaSite (see below), within 5 Å of the heme iron. Similarly, inhibitor complexes are included only if the inhibitor presents a lone pair of electrons (most commonly *via* an inhibitor nitrogen atom) within 5 Å of the heme iron. Using these criteria, the following exclusions were made: (CYP3A4, 2V0M, KLN 1501), (CYP2C8, 2NNH, REA 502), (CYP2C8, 2NNH, PLM 503), (CYP2C8, 2NNI, PLM 502), (CYP2C8, 2NNJ, PLM 502), (CYP8A1, 3B6H, BOG 701), (CYP2B4, 2BDM, TMI 502) and (CYP2B4, 2BDM, TMI 503).

Delimitation of Protein-ligand Interactions

The program LIGPLOT [34] was used in default mode to delimit interactions between the ligand and the protein in X-ray crystal structure complexes used in this study and in docking complexes generated in this study. LIGPLOT delimits and reports amino acid residues that make hydrophobic interactions and hydrogen-bond interactions with the bound ligand. In order to confirm that the delimitations made using LIGPLOT were consistent with those made using other related methods, we also used Zamora's molecular interaction fields (MIFs) [35] to delimit amino acid residues that interact with ligand in a limited number of X-ray protein-ligand complexes; in this quality control exercise, only amino acids whose MIF-calculated interaction energies with the ligand were $\leq$ -1 kcal/mol were regarded as "contact residues". Through this exercise, we quickly established that Zamora's MIF were delimiting exactly the same set of residues as LIGPLOT in the considered structures, confirming that LIGPLOT's delimitations can be trusted. Given that Zamora's MIFs appeared to provide no additional delimitations over LIGPLOT, for simplicity we thereafter only used LIGPLOT in all subsequent work.

Sequence Alignments

CYP450 protein sequences were aligned using the program ClustalW (default parameters) [45-47]. The quality of the multiple sequence alignments was assessed using various cues that included: (i) successful alignment of the heme iron-binding cysteine residue, (ii) successful alignment of secondary structure elements such as the I, F and G helices and (iii) similarity with Gotoh's alignment of CYP2 family proteins and bacterial CYPs [5].

The CYP Substrate Space: Molecular Descriptors

A list of known substrates for CYP1A2, CYP2A6, CYP2B4, CYP2C5, CYP2C8, CYP2C9, CYP2D6, CYP2R1, CYP3A4 and CYP46A1 was compiled from primary literature, Micromedex, RCSB Protein Data Bank and other online databases and resources. Online resources that were used include the twelve sources listed as references [43, 48-58]**.** Chemical structures for these substrates were obtained from PubChem database and were processed using the "Standardize Molecule", "Strip Salts", and "Add Hydrogens" modules of Pipeline Pilot Student Edition v6.1.5. [59]. PaDEL-Descriptor v2.4 [60] was then used to compute the PubChem fingerprints [61] for the substrates. A principal component analysis (PCA) plot of the substrates using the PubChem fingerprints was performed using SIMCA-P+ v12.0.1 [62] to show the distribution of the substrates in the chemical space.

Prediction of Sites of Metabolism Using MetaSite

MetaSite is a program designed to compute and rank the likely structure of metabolites formed by CYP-mediated primary metabolism [37]. To do this, MetaSite calculates two sets of descriptors, one for the CYP enzyme and one for the potential substrate, respectively representing the chemical fingerprint of the enzyme and the substrate. The set of descriptors used to characterize the CYP enzyme is based on GRID flexible molecular interaction fields (GRID-MIFs). Using the 3D structure of a given compound, MetaSite then automatically calculates charges and reactivities, pharmacophoric recognition, descriptor handling, and similarity. Validation of the method has shown that the primary site of metabolism is found in the top three MetaSite predictions in more than 85% of cases.

In our work, SMILES strings for CYP450 ligands were prepared from PDB files of the entities using Open Babel [63]. PDB files of ligands were extracted from PDB files of CYP450-ligand complexes. Sites of metabolism were predicted by MetaSite using default parameters in both CYP450 isoform-specific predictions and in organ-specific (e.g. liver) predictions.

Generation of a SRS Map Based on X-ray Structures: the "X-ray Structures" SRS Map

CYP450 protein-ligand contacts were delimited using the program LIGPLOT [34] as described above. This identified, for a given CYP450-ligand complex, the CYP450 amino acid residues making hydrophobic and/or H-bond interactions with the ligand. The protein sequences of the CYP450 isoforms whose complexes with ligands had been analysed using LIGPLOT were retrieved from the UniProtKB database and aligned using ClustalW as described above. The alignment included sequences of other mammalian CYP450 isoforms of interest whose structures in complex with ligands have not yet been determined. The CYP450 contact residues identified using LIGPLOT were annotated on the multiple sequence alignment to generate a contact map (simply an alignment of sequences where contact residues and frequency of contact in structures of the same CYP450 isoform is indicated) as follows: (a) The structures for a given CYP450 isoform were grouped and analysed using LIGPLOT to identify contact residues; (b) continuous sequence sub-strings of about 15 amino acid residues were defined that include the residues delimited by LIGPLOT as interaction residues. This step facilitates the identification of contact residues on the sequence of the appropriate isoform in the multiple sequence alignment; (c) the contact residues for each isoform in the multiple sequence alignment were identified using the sub-strings and shaded using the following colour scheme: (i) yellow if the use of that residue position in ligand contacts has been observed only once in the LIGPLOT contact maps for the CYP450 isoform under consideration, (ii) blue if observed twice, (iii) turquoise if observed thrice, (iv) green if observed four times, (v) red if observed five times and (vi) dark red if observed six times. N.B. The highest number of structures collected for the individual CYP450 isoforms in this study was six. The resultant contact map is shown in Fig. **S1** of the Supplementary Materials. An implicit assumption here is that whenever a residue position in a given CYP450 isoform is delimited as an interaction/contact residue using LIGPLOT, all residues that are evolutionarily equivalent to that residue position in other CYP450 isoforms are also contact residues. Evolutionarily equivalent residue positions are in the same column of a multiple sequence alignment and these are residue positions that descend from the same sequence position in an ancestral sequence [64]. Hence, the delimitation of a contact residue in a multiple sequence alignment automatically delimits a "contact column" of residues – namely the column in which that contact residue lies in the multiple sequence alignment.

We generated a SRS map from the contact map as follows: (a) SRSs were assembled as blocks of "contact columns" which are within three amino acid residues of each other. The use of three amino acids is arbitrary and is inspired by Gotoh's use of the same value in delimiting the precise boundaries for his SRSs [5]; (b) the precise N-terminal and C-terminal boundaries of the SRSs are delimited by calculating column positions that are three amino acid residues away (counting to the left for the N-terminal boundary and to the right for the C-terminal boundary, the third amino acid position is included in the SRS) from the outer-most "contact columns" in the SRS assembled so far. The farthest column position among the several options that can sometimes be calculated for a SRS is taken as the boundary of the SRS.

Generation of CYP450-ligand Complexes by Docking

Structures of CYP1A2 (2HI4), CYP2A6 (1Z10), CYP2B4 (2Q6N), CYP2C5 (1NR6), CYP2C8 (2VN0), CYP2C9 (1R9O), CYP2D6 (2F9Q), CYP2R1 (3CZH), CYP3A4 (2V0M) and CYP46A1 (2Q9F) were retrieved from the RCSB Protein Data Bank. These structures were prepared for docking using AutoDock Vina [65] by (a) removing multiple copies of the CYP450 isozyme, (b) removing water molecules, (c) removing buffer molecules, (d) removing bound substrate, (e) addition of hydrogen atoms, and (f) converting pdb file to pdbqt format using AutoDockTools [66].

The docking space in each CYP450 isozyme was defined using a grid box whose size was determined by requiring that the grid box accommodates the largest substrate from a comprehensive list of substrates of the CYP450 isoform.

In general, in the absence of precedence in the literature, it is essential before docking of the ligands begins to determine if the docking software, here AutoDock Vina, is suitable for docking ligands in the active site of the enzyme system in question, here CYP450. It is also essential to identify the general set of parameters that permit accurate docking for a given docking system – i.e. it is essential to calibrate the docking system. In that light, a re-docking study was performed in which substrates whose structures bound to CYP450 isoforms are known were re-docked into the structure of the appropriate CYP450 isozyme using AutoDock Vina. RMSD values - representing the structural difference between docking complexes and X-ray structure complexes - were calculated for the substrate in each substrate-CYP450 isoform combination. Note that the CYP450 protein structure in each re-docking study was identical to that in the relevant X-ray complex since it was regarded as a rigid object during docking.

All RMSD values for the structural difference between the substrate in docking- and X-ray complexes in the re-docking study were less than 2 Å. Wang *et al.* [67] have previously reported a comparative evaluation of 11 scoring functions for molecular docking and used a 2 Å value to compare the similarity of outputs from the different docking software. It is therefore reasonable to conclude that substrates in the re-docking complexes generated here had similar orientations to substrates in the relevant X-ray structures, indicating that AutoDock Vina run under default parameters can provide a suitable platform for the docking of ligands into CYP450 proteins. We therefore proceeded to generate CYP450-ligand complexes for delimiting the docking SRS map by applying AutoDock Vina (default conditions) onto the remaining substrates for each of the CYP450 isoforms. PaDEL-ADV v1.7 [68] was used to automate the docking of substrates into CYP450 active sites using the AutoDock Vina software.

We generated a maximum of 9 docking modes for each substrate (note that some substrates had less than 9 docking modes because the software could not find additional unique binding modes). Due to the large number of docking modes generated (approximately 7000), visual inspection of the docking modes of the substrates was generally not done. Docking modes for a few substrates were however visualized randomly for quality control purposes. In addition, the docking modes of some substrates with high binding energy (near to -1 kcal/mol) were also visualized to check for errors in their chemical representation. In carrying out such visualization, we were only concerned about the correct chemical and structural representation of the substrates and did not in general examine the docking modes to explain why the compound was a substrate; this was only done for glibenclamide and vardenafil as the docking complexes of these two suggested that SRS2 and SRS3 are co-joined.

Generation of a SRS Map Based on Docking Complexes: the "Docking" SRS Map

An in-house version of PaDEL-ADV (v1.7) [68] was used to automate both the AutoDock Vina [65] docking process and the LIGPLOT [34] identification of protein-ligand contacts in complexes generated using AutoDock Vina. In this regard, the internal workings of PaDEL-ADV are briefly as follows. PaDEL-ADV starts by reading a directory containing substrate files. For each substrate, the structure file was converted into a PDB file using The Chemistry Development Kit [69]. The PDB file was subsequently converted to pdbqt format using the "prepare_ligand4.py" python script provided by AutoDockTools [66]. Following that, AutoDock Vina was used to dock the substrate into the active site of the appropriate CYP450 protein. Individual binding modes were then extracted from the output pdbqt file using the "vina_split" program of AutoDock Vina. These pdbqt files were then converted to PDB files using the "pdbqt_to_pdb.py" python script [66]. LIGPLOT was then applied to the PDB files of the individual binding modes to identify protein-ligand contacts for all the poses of each substrate. Information about these protein-ligand contacts for all the substrates of a CYP450 protein was then consolidated into a protein-ligand contact matrix where each row represents a pose of a substrate and each column represents a protein residue that is contacted by substrate. This protein-ligand contact matrix was used directly to delimit the SRS docking map. The protein-ligand contact matrix used here was processed to remove those poses with a computed ligand binding free energy greater than -1 kcal/mol and those poses in which the atoms of the substrate are more than 5 Å away from the heme iron.

The remaining poses were then used to generate a contact map and then a SRS map using procedures such as those used for the X-ray structures SRS map. The docking complexes contacts map is shown in Fig. **S2** of the Supplementary Materials.

The 5 Å distance cut-off was chosen because it is a commonly accepted upper limit for the distance between the site of metabolism and the heme iron at a specific committed stage in the CYP450 catalytic cycle. The -1 kcal/mol cut-off used was strictly speaking an arbitrary value and was chosen to include as many of the known substrates for a given enzyme in the analysis as possible without introducing too many incorrect poses [Note that since all scoring functions aim to reproduce the Gibbs free energy of binding, any cut-off must be < 0 kcal/mol in order for the binding to be favorable]. Thus for example, in CYP2C9 there are 6 known substrates, the lowest energy of which was -1.1 kcal/mol; if we had set a lower cut-off (e.g. -2 kcal/mol), we would exclude these known substrates from analysis, whereas if we had used a higher cut-off (e.g. -0.1 kcal/mol), we would have risked including poses that may be incorrect.

Assessment of SRS Map Completeness: Calculation of the Coverage of the CYP450 Substrate Space Achieved by Ligands in X-ray Structures

Our objective here was to determine the coverage of the CYP450 substrate space achieved by the ligands that we used to delineate the X-ray structures SRS map. Calculation of this coverage gives us a ligand-based measure of the likely extent of completeness of the X-ray structures SRS map. For example, a map

derived from contacts made by a higher percentage of ligands in the CYP450 substrate space is naturally expected to yield a more complete description of SRSs than a map derived from a smaller percentage of ligands. The ligand-based measure of SRS map completeness can then be compared with other measures of completeness such as the projection of SRSs onto the CYP450 fold (*vide infra*). Our strategy for achieving a calculation of this coverage was to divide the CYP450 substrate space calculated above into clusters of similar molecules and to calculate the percentage of these clusters that contain at least one molecule from the set used to delimit the X-ray structures SRS map. This percentage value was taken as a quantitative measure of the coverage of the CYP450 substrate space achieved by the ligands used to delineate the X-ray structures SRS map. We used *k*-means clustering to divide-up the CYP450 substrate space into clusters.

K-means clustering [70, 71] is an unsupervised, heuristic method of cluster analysis which aims to partition *n* observations into *k* clusters, in which each observation belongs to the cluster with the nearest mean, the aim being to find the centres of natural clusters in the data using an iterative refinement approach. Using this method, we clustered all the molecules in the CYP450 substrate space into structurally-related chemical families. We performed the *k*-means clustering as follows: Centroids for *k* clusters were defined using *k* randomly selected ligands; Each ligand was then assigned to the cluster with a centroid nearest to it. The positions of the centroids were then recalculated and the ligands re-assigned to the clusters. The recalculation of the centroids and reassignment of the ligands continued until the centroid positions became stable. In this study, the software RapidMiner [72] was used to perform the *k*-means clustering.

An important challenge when performing *k*-means clustering is the choice of the optimum value of *k* – the number of clusters. In this study, we devised a simple strategy to guide the selection of *k* for the *k* -means clustering. In this strategy, we defined four chemical types/families into which some molecules in the CYP450 substrate space could be easily assigned manually on the basis of structural similarity, the idea being to use these four chemical families as "markers" in the evolution of the clusters. The four chemical families chosen were: dihydropyridines; selective serotonin reuptake inhibitors; beta-blockers; & benzodiazepines. The satisfactory assignment of molecules from these four families into clusters of closely-related members was taken to be indicative of equally satisfactory evolution of clusters in all parts of the CYP450 substrate space. There are 48 molecules that belong to the families described above and the distribution of these 48 chemical entities across clusters was followed during the course of the clustering procedure. Hence in our strategy for selection of the optimum *k*-value (i.e. termination of the clustering algorithm), the clustering of the 48 molecules from the four chemical families described above was followed in the clusters that were generated by RapidMiner as the value of *k* was increased during the course of *k*-means clustering; manually assessment was used to establish the extent of similarity within clusters containing at least one molecule from the 48 marker molecules. Our strategy therefore achieved the definition of algorithm/search termination criteria by simply assessing some of the *k*-means clusters' verisimilitude to a pre-defined and manually-ascribed notion of chemical similarity that had been applied to a small part of the CYP450 substrate chemical space. Achievement of satisfactory clustering of the 48 marker molecules (a small part of the CYP450 substrate space, 48/868 ~ 6 %) was taken to be indicative of equally satisfactory clustering in all other parts of the CYP450 substrate space. We note that a major source of the disparities between PubChem fingerprints (used here to calculate the co-ordinates of the CYP450 substrates in chemical space) and our visual assignment of the 48 marker molecules on the basis of structural similarity is that PubChem fingerprints place greater direct emphasis on electronic structure similarity and less emphasis on structural similarity while the opposite is true in the case of our visual assignments.

Assessment of SRS Map Completeness: Projection of the SRSs Delimited in the Final Unified Map Onto the CYP450 Fold

The CYP2C9 structure 1OG2 (chain A) was chosen as a representative mammalian CYP450 fold because the structure does not have gaps in between its N-term end (sequence position Pro30) and its C-term end (Val490). This absence of gaps in the CYP2C9 structure allows visual assessment of the completeness of the unified SRS map by identification of parts of the CYP450 fold that are proximal to and unscreened from the catalytic centre and which either have or have not been assigned as SRS zones. For this purpose, CYP2C9 SRS boundaries were extracted from the final unified SRS map and applied onto the CYP2C9 structure 1OG2.

RESULTS

Generation of a SRS Map Based on X-ray Structures: the "X-ray Structures" SRS Map

Figure **1** shows a comparison between the X-ray structures map and Gotoh's map. Contact residues in CYP450-ligand complexes have been projected onto an alignment of CYP2C9, CYP2B4 and CYP3A4 and the X-ray structures map delimited as described in Materials and Methods.

The CYP450 Substrate Space: Molecular Descriptors Assessment of SRS Map Completeness

Figure **2** shows the distribution of known CYP450 substrates/ ligands in chemical descriptor space (see Materials and Methods for further details). The plot allows comparative visualization of the chemical spread of the ligands in CYP450-ligand complexes.

Calculation of the Coverage of the CYP450 Substrate Space Achieved by Ligands in X-ray Structures

We have found that the automatically-assigned clusters that contain the 48 "marker" chemical entities (belonging to the four chemical families defined above) bear the greatest agreement with manually-assigned chemical similarity at the *k* -value of 86. Hence the percentage spread of the CYP450 substrate space achieved by the ligands that we used to delineate the X-ray structures SRS map was calculated at $k = 86$. To determine the coverage of the ligands that were used to delineate the X-ray structures SRS map, each ligand was assigned to the cluster with a centroid nearest to it. The percentage coverage of these ligands was then computed by dividing the total number of clusters that are occupied by these ligands (22) by the total number of clusters (86), revealing that the coverage of 'total known' CYP450 substrate space by the ligands that were used to delineate the X-ray structures SRS map is 26%. Considered in isolation, this result suggests that the coverage of the CYP450 substrate space achieved by the ligands from X-ray structures alone is low. In turn, this suggests that the completeness of the SRS map delimited from the X-ray structures may also be low and therefore that X-ray structures of more diverse CYP450-ligand complexes need to be observed before high SRS map completeness could be attained.

Generation of a SRS Map Based on Docking: the "Docking" SRS Map

The overall statistics of the substrates and CYP450 isoforms used in the docking experiments are provided in Table **1** below. The relevant sections of the docking SRS map are shown in Fig. **S3** in the Supplementary Materials.

Are the Newly-delimited SRS Substantive?

We have performed structural analysis to assess the likelihood that the newly delimited SRS zones constitute substantive SRSs -

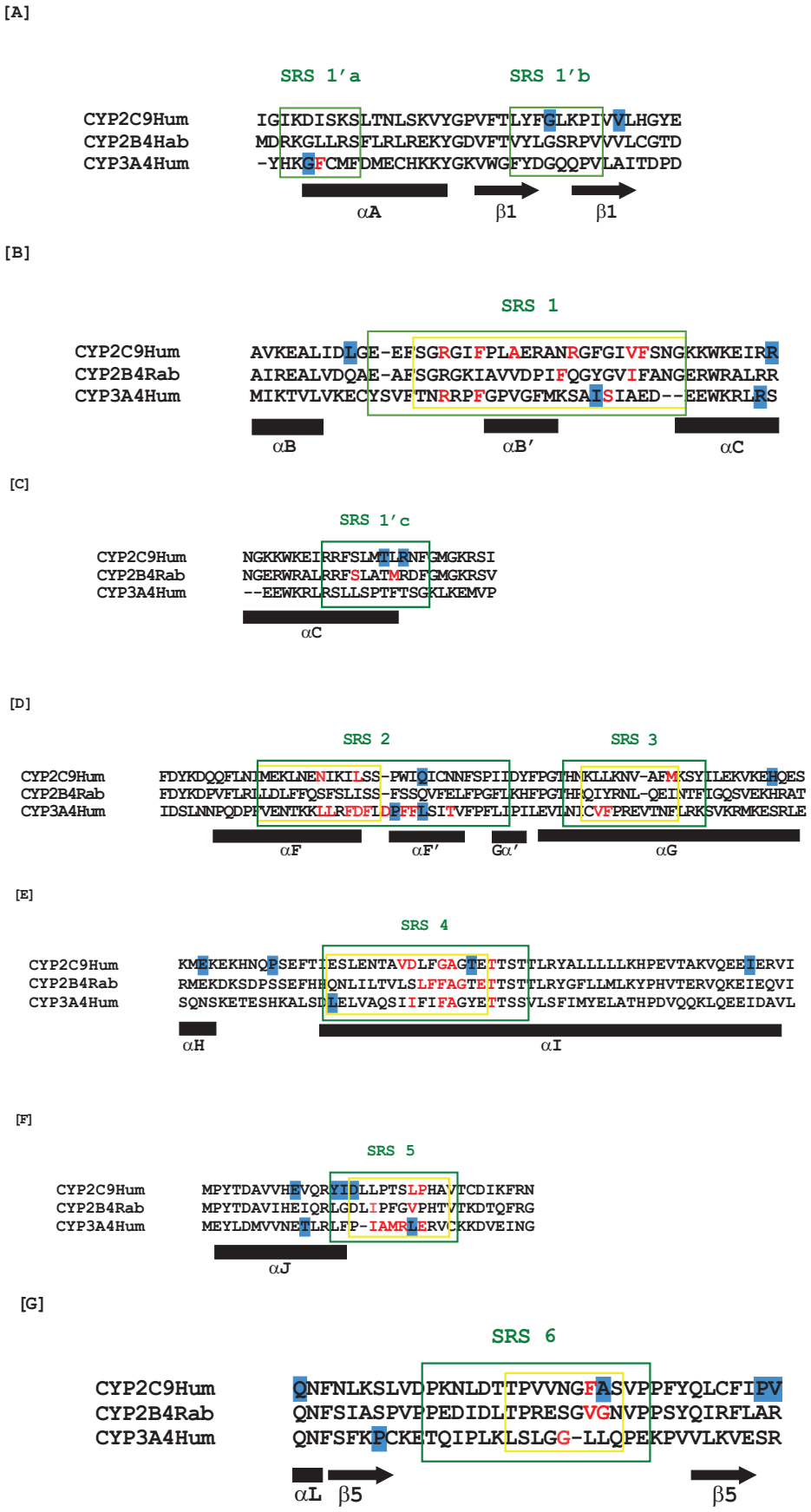


Fig. (1). Comparison of the X-ray structures SRS map delimited here and Gotoh's SRS map Contact residues in CYP450-ligand complexes have been projected onto an alignment of CYP2C9, CYP2B4 and CYP3A4 and the X-ray structures map delimited as described in Materials and Methods.

The newly defined X-ray structures SRS map is illustrated here using sections from an alignment of CYP2C9, CYP2B4 and CYP3A4. The complete set of sequences and contacts, from which the SRS map was delimited, is shown in Fig.S1 of Supplementary Materials. SRS regions delimited in this work are boxed in green while those delimited by Gotoh are boxed in yellow. Contact residues identified in crystal structures are shown in red characters. Secondary structure assignment is shown below the sequence (black rectangular boxes represent α- helices and black arrows represent β-sheets). Panel [**A**] shows the newly-identified SRS zones designated as SRS1'a and SRS1'b. Panel [**B**] shows SRS1. Panel [**C**] shows another newly-identified SRS zone designated as SRS1'c. Panel [**D**] shows close proximity between SRS2 and SRS3 and Panel [**E**, **F**] and [**G**] show SRS4, SRS5 and SRS6, respectively. The positions of CYP2C9 and CYP3A4 polymorphic variants are highlighted in blue.

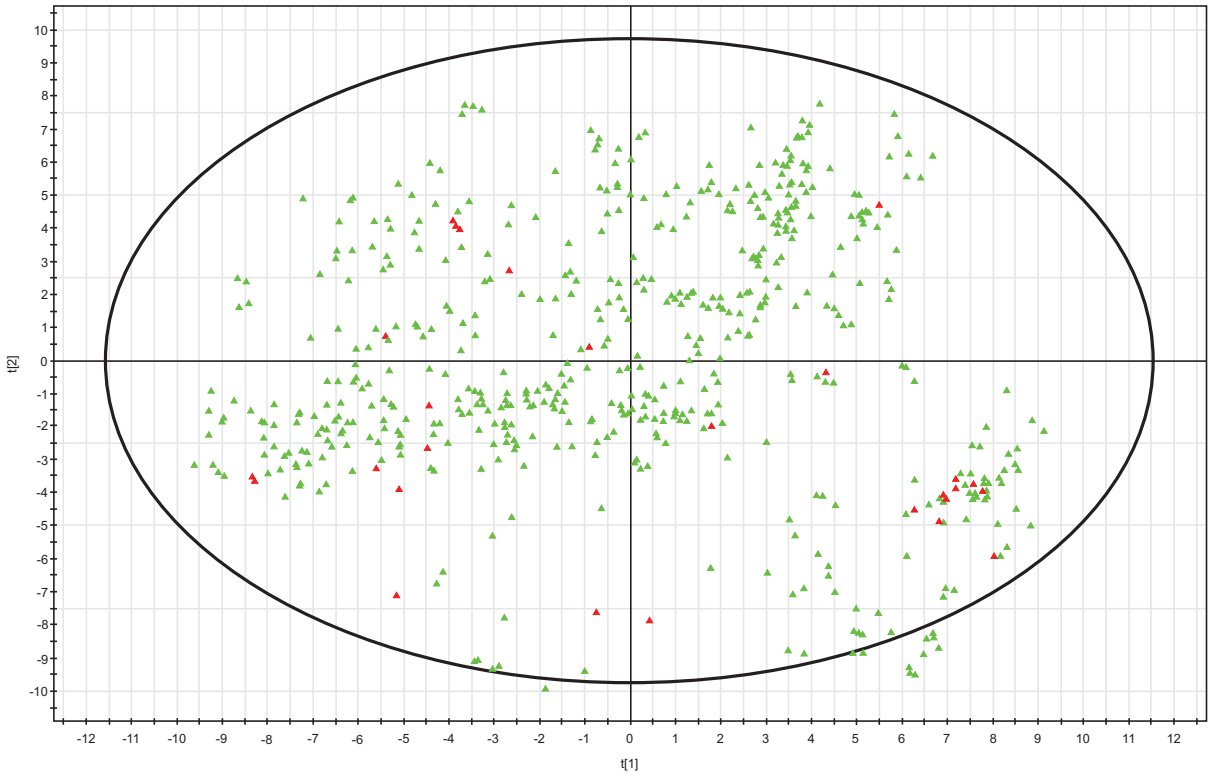


Fig. (2). Comparative analysis of the chemical diversity of ligands from CYP450-ligand complexes used in this study
This plot shows the distribution of human CYP450 substrates in chemical space using PubChem fingerprints as chemical descriptors. Structures of chemical entities were sourced, chemical descriptors calculated and principal component analysis performed as described in Materials and Methods. The known substrates for CYP1A2, CYP2A6, CYP2B4, CYP2C5, CYP2C8, CYP2C9, CYP2D6, CYP2R1, CYP3A4 and CYP46A1 are represented by triangles: Ligands used in this study to delineate the X-ray structures SRS map (i.e. ligands bound to X-ray structures) are represented by red triangles and all other known ligands are represented by green triangles, illustrating the coverage of CYP450 substrate space achieved by the ligands in X-ray structures.

Table 1. Statistics of Substrates and CYP450 Isoforms Used to Generate the Docking SRS Map

CYP450	No. of Known Substrates	No. of Substrates Used to Generate Contact Map[a]	% Substrates Used to Generate Contact Map
1A2	65	61	93.8
46A1	11	11	100.0
2A6	33	26	78.8
2B4	12	12	100.0
2C5	5	5	100.0
2C8	28	27	96.4
2C9	156	152	97.4
2D6	212	186	87.7
2R1	2	2	100.0
3A4	401	386	96.3

[a]Substrates that, (1) are more than 5Å from hem, (2) have greater than -1kcal/mol binding affinity with CYP450 protein, (3) are unable to be docked using AutoDock Vina, were therefore not used to generate the contact map.

i.e. to assess if they are stand-alone SRSs that are distinct from each other (where relevant) as well as from the known ones.

We carried out a qualitative assessment of the likelihood of each newly-delimited SRS zone constituting a stand-alone SRS by searching for structural elements that are removed and/or screened from the active site and are interposed between the newly-delimited SRS zone and other SRS zones (newly-delimited and/or previously-defined). Such intervening regions achieve the separation of SRS zones by non-SRS zones and their presence supports the conclusion that the newly-delimited SRS zone is substantive and stand-alone.

Structural Analysis of Newly-delimited SRS Regions: SRS1'c

Figure **3** shows the relationship between SRS1 and the newly-delimited SRS1'c.

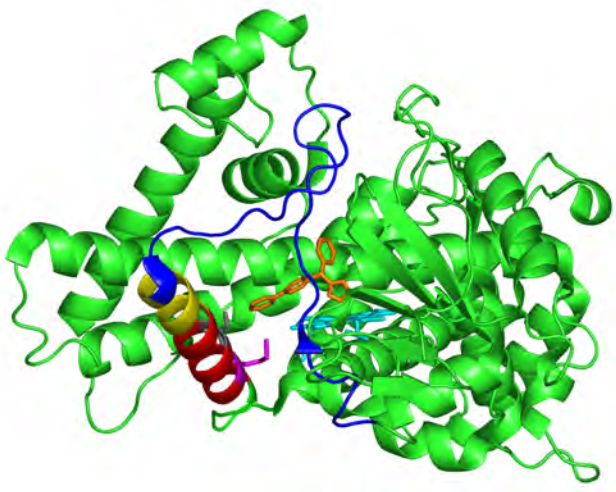


Fig. (3). The relationship between SRS1 and SRS1'c on the CYP450 fold

CYP2B4 protein is shown in green, the heme group is shown in turquoise and bifonazole is shown in orange (PDB structure 2BDM). SRS1 residues are shown in blue (CYP2B4 residues 92-118 inclusive) and X-ray structures SRS map boundaries delimited herein have been used for SRS1. SRS1'c is shown in red (residues 125-135 inclusive). SRS1'c residues Ser128 and Met132 are shown respectively in grey and magenta and sticks have also been rendered for these two residues. The delimitation of SRS1'c in the X-ray structures SRS map is based upon the observed contacts between the bifonazole ligand and CYP2B4 Ser128 and CYP2B4 Met132 in the structure 2BDM. The residues that lie between SRS1 and SRS1'c are shown in yellow (residues 119-124 inclusive). The structural model was rendered in pymol (www.pymol.org).

SRS1 and SRS1'c are quite proximal and there are no structural elements between SRS1 and SRS1'c that would obviously obstruct interaction between the residues that intervene between SRS1 and SRS1'c and the substrate. The intervening residues are Glu119-Leu124 inclusive along the CYP2B4 sequence. It is conceivable that the intervening residues (shown in yellow in Fig. **4**) may in fact be SRS residues, thereby giving a much bigger SRS1 than previously delimited. We therefore analysed the results of docking twelve CYP2B4 substrates into the CYP2B4 structure 2BDM (the structure rendered in Fig. **3**) to explore the possible involvement of SRS1-SRS1'c intervening residues (rendered in yellow) in substrate binding. The following are the PubChem codes of the CYP2B4 substrates used in this study: 2341 = Benzphetamine, 4116 = Methoxyflurane, 6009 = Aminopyrine, 6013 = Testosterone, 6115 = Aniline, 7095 = Biphenyl, 7485 = p-nitroanisole, 35703 = 7-ethoxycoumarin, 107683 = 7-pentoxyresorufin, 114982 = 7-benzyloxyresorufin, 130779 = 7-ethoxy-4-trifluoromethylcoumarin and 5359271 = Ethylmorphine.

Contacts between CYP2B4 substrates and SRS1-SRS1'c intervening residues are summarized in Table **2** below.

The results in Table **2** show involvement of SRS1-SRS1'c intervening residues in substrate contact. This suggests that SRS1 extends from Ala92 to Phe135 along the CYP2B4 sequence (the blue, yellow and red sections in Fig. **3**). In the final SRS map delimited here, SRS1'c is therefore not given as a stand-alone SRS region but is fused to SRS1.

Structural Analysis of Newly-delimited SRS Regions: SRS1'a and SRS1'b

Figure **4** shows the relationship between SRS1 and the newly-delimited SRS regions - SRS1'a and SRS1'b.

SRS1'a and SRS1'b are separated from SRS1 and from each other by structural elements that include residues that are far removed from the CYP450 active site and whose interaction with bound ligand is screened by other residues (see Fig. **4**). Hence SRS1'a and SRS1'b are truly distinct from SRS1 and from each other as SRS regions (i.e. they do not form a single continuous SRS zone). Furthermore, SRS1, SRS1'a and SRS1'b appear as distinct SRS contacts zones in the large-scale docking study upon which the delimitation of the docking SRS map is based (see Figs. **S2** and **S3**). By way of example, the large scale docking study includes the CYP46A1 structure 2Q9F and the CYP3A4 structure 2V0M; The X-ray structure hydrophobic contact between CYP3A4 Phe57 and ketoconazole that results in the delimitation of SRS1'a is also observed in docking complexes of the CYP3A4 structure 2V0M and gives rise to a standalone SRS1'a in the docking SRS map (see Figs. **S2** and **S3**); Furthermore, the X-ray structure contacts between CYP46A1 Phe80 (hydrophobic) and CYP46A1 His81 (hydrogen-bond) and Cholesterol-3-sulphate which result in the delimitation of SRS1'b are also observed in docking complexes of the CYP46A1 structure 2Q9F and give rise to a stand-alone SRS1'b (see Figs. **S2** and **S3**). Hence the view arising from structural analysis that SRS1'a, SRS1'b and SRS1 are distinct SRS zones is directly supported by their emergence as separate zones in the docking SRS map.

Structural Analysis of Newly-delimited SRS Regions: the Overlap of SRS2 and SRS3

The overlap between Gotoh's SRS2 and SRS3 in the docking SRS map (Fig. **S3**, Panel D) as opposed to their close proximity in the X-ray structures SRS map (Fig. **S3**, Panel D and Fig. **1**, panel D) is due to the contacts involving Tyr225 and His230 identified in CYP2C9 docking complexes. Tyr225 makes contact with ligand 110634 (Vardenafil) and His230 makes contact with ligand 3488 (Glibenclamide) and 110634 (Vardenafil). Tyr225 accounts for 2 % of the total number of contacts delimited in CYP2C9 docking complexes in this study while His230 accounts for 4.6 %.

We evaluated the docking complexes that give ligand contacts with residues Tyr225 and His230 by analysing the agreement between the experimentally validated and/or MetaSite-predicted sites of metabolism of the ligand and the sites of metabolism suggested by the pose of the docked ligand in the CYP2C9 active site. The sites of metabolism (experimental and predicted) for Glibenclamide and Vardenafil are shown in Fig. **S4** in the Supplementary Materials.

Glibenclamide (also known as Glyburide) is a second generation sulfonylurea used for treatment of type II and gestational diabetes mellitus [73-76]. It targets ATP-dependent ion channels [77, 78] and is mainly metabolised by CYP2C9, CYP2C19 and CYP3A4 in the human liver [74]. Experimentally-verified sites of metabolism in the Glibenclamide cyclo-hexyl moiety (indicated by arrows (5), (6a) and (7a) in Fig. **S4**, Panel A) have been identified in various publications [76]. Experimentally-verified sites of metabolism in the Glibenclamide ethyl bridge moiety (indicated by arrows (1) and (4) in Fig. **S4**, Panel A) have also been identified

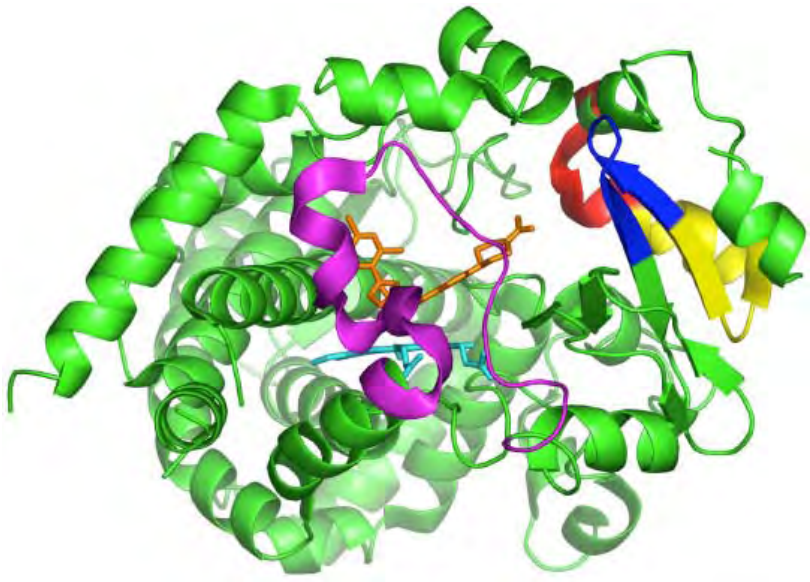


Fig. (4). The relationship between SRS1, SRS1'a and SRS1'b on the CYP450 fold.

CYP3A4 protein is shown in green, the heme group is shown in turquoise and ketoconazole is shown in orange (PDB structure 2V0M). SRS1 residues are shown in magenta (CYP3A4 residues 99-125 inclusive) and X-ray structures SRS map boundaries delimited herein have been used for SRS1. SRS1'a residues are shown in red (residues 54-60 inclusive) and SRS1'b residues are shown in blue (residues 74-81 inclusive). The residues that lie between SRS1'a residues and SRS1'b residues are shown in yellow (residues 61-73 inclusive). The structural model was rendered in pymol (www.pymol.org).

[76]. The active-site orientation of Glibenclamide in the highest-scoring (and only) docking complex/solution from which the contact between Glibenclamide and CYP2C9 His230 was delimited is shown in Fig. **S5** in the Supplementary Materials. It can be seen that the Glibenclamide molecule presents the cyclo-hexyl moiety towards the heme iron in the catalytic centre. This docking solution is in accord with the known metabolites of Glibenclamide.

Vardenafil is a potent and selective inhibitor of cyclic GMP-specific phosphodiesterase type 5 (PDE5) found in smooth muscle cells that line blood vessels – especially in the corpus cavernosum of the penis. Vardenafil is used for the treatment of erectile dysfunction [79-81]. Vardenafil is predominantly metabolised by hepatic CYP3A4 and to a lesser extent by CYP3A5 and CYP2C isoforms [79]. Although the main Vardenafil metabolite is N-desethylvardenafil (Vardenafil with the ethyl group removed from the piperazine moiety by hydrolysis) [80, 81], there is evidence for the existence of hydroxylated metabolites [81].

The three docking solutions for the docking of Vardenafil into the CYP2C9 active site have generally the same orientation in which a predicted site of metabolism (site (5) in Fig. **S4** Panels C & D) comes close to the catalytic heme iron. Vardenafil makes contact with the residue Tyr225 in each of these three complexes; the orientation of Vardenafil in the lowest-energy (highest ranking) docking complex is shown in Fig. **S6** in the Supplementary Materials. These three docking solutions are in accord with the predicted metabolites of Vardenafil.

The Unified SRS Map: Set Union of the X-ray Structures and Docking SRS Maps

The SRS zones of the X-ray structures SRS map and the docking SRS map are added together in set-theoretic fashion to give the unified SRS map shown in Fig. **5** below. The unified map also incorporates results from structural analyses of SRS zones delimited in the X-ray structures map such as the amalgamation of SRS1'c and SRS1 to give a large SRS1 zone.

The SRS boundaries shown pictorially in Fig. **5** above are given along the CYP450 sequence for CYP2C9 isoform (SRS boundaries are also given in Table **3**). A version of the unified SRS map that includes more CYP450 sequences is shown in Fig. **S11** in the Supplementary Materials.

Assessment of SRS Map Completeness: Projection of the SRSs Delimited in the Final Unified Map Onto the CYP450 Fold

The SRSs in the final unified map shown in Fig. **5** above were projected onto the CYP2C9 structure 1OG2.pdb as described in the Materials and Methods section above. The projections are shown in Fig. **6** below.

Table 2. Summary of Contacts Between CYP2B4 Substrates and SRS1-SRS1'c Intervening Residues

CYP2B4 Residue	List of Substrates Making Contact with Residue	Percentage of Substrates Making Contact with Residue
Trp121	2341, 6009, 6013, 7485, 107683, 114982 and 5359271	7/12 = 58.3 %
Leu124	2341, 6009, 35703, 107683, 114982, 130779 and 5359271	7/12 = 58.3 %
Arg125	2341, 6009, 107683, 114982 and 130779	5/12 = 41.7 %

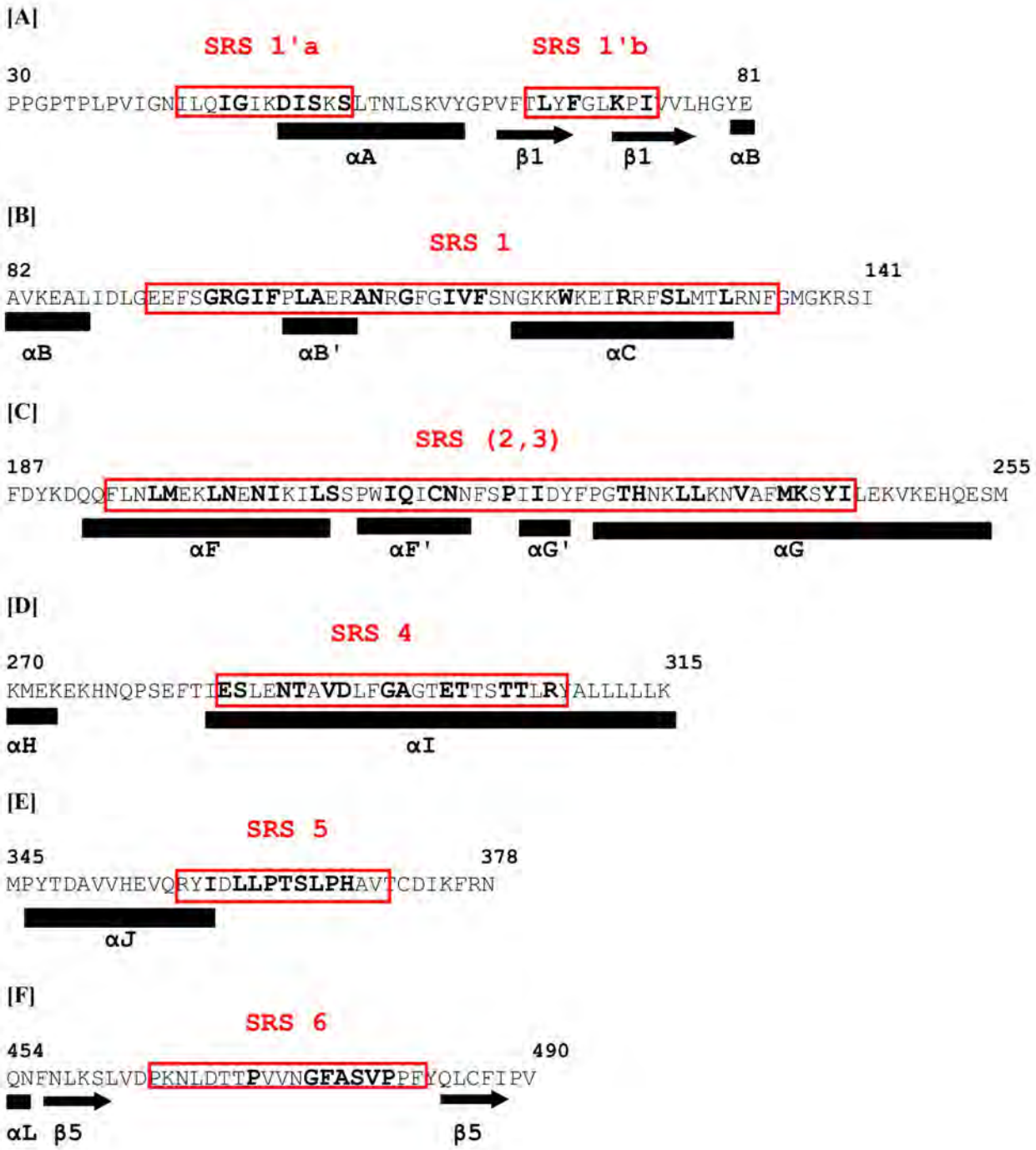


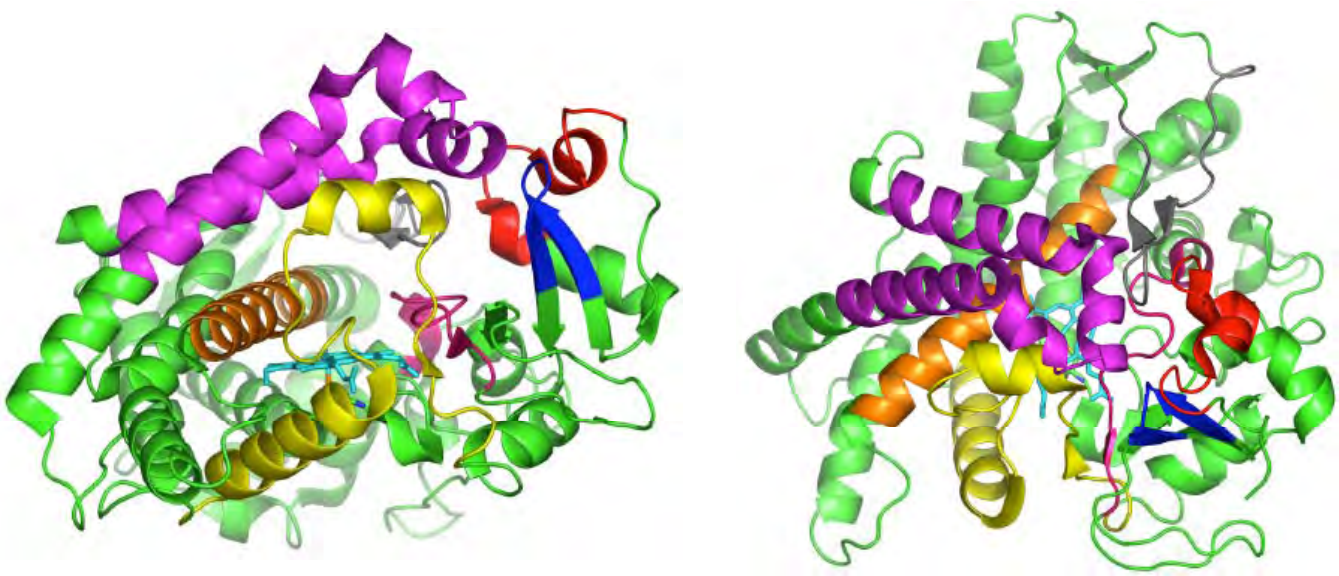
Fig. (5). The unified SRS map: set union of the X-ray structures and docking SRS maps

The final SRS map is built by taking the N term-most limit and the C term-most limit for each SRS in the comparison between the X-ray structures SRS map and the docking SRS map. Regions of the CYP2C9 sequence that fall in the final SRS zones 1'a to 6 are shown in panels [A] – [F]. SRS zones are boxed in red and secondary structures are shown below the sequence (black rectangular boxes represent α- helices and black arrows represent β-sheets). Based on visual inspection of the CYP2C9 1OG2 crystal structure, SRS residues with side chains facing towards the active site are shown in bold face while those pointing away are shown in plain font. Panel [A] shows SRS1'a and SRS1'b; panel [B] shows the extended SRS1 which results from combining SRS1 in Fig. **1** and SRS1'c; panel [C] shows the co-joining of SRS2 and SRS3 regions given as SRS(2,3); and panels [D], [E] and [F] show SRS4, SRS5 and SRS6, respectively. Numbering of sequence positions is according to the CYP2C9 sequence.

Table 3. SRS Boundaries in CYP2B4, CYP2C9 and CYP3A4

CYP450 Isoform	SRS Boundary Positions (Boundary Residues are Part of SRSs)
CYP2B4	(SRS1'a: L43-S54), (SRS1'b: T67-V75), (SRS1: E93-F135), (SRS(2,3): F195-I245), (SRS4: Q286-R308), (SRS5: R358-V371) and (SRS6: P465-S483). That is 171/491 = 35 % of CYP2B4 residues are SRS residues.
CYP2C9	(SRS1'a: I37-S53), (SRS1'b: T66-I74), (SRS1: E92-F134), (SRS(2,3): F194-I244), (SRS4: E285-R307), (SRS5: R357-V370) and (SRS6: P464-F482). That is 176/490 = 36 % of CYP2C9 residues are SRS residues.
CYP3A4	(SRS1'a: I50-F60), (SRS1'b: G73-V81), (SRS1: Y99-G140), (SRS(2,3): Q200-S252), (SRS4: L293-S315), (SRS5: R365-C377) and (SRS6: T471-P488). That is 169/503 = 34 % of CYP3A4 residues are SRS residues.

The percentage of SRS residue positions in the alignment of Fig.**S11** (Supplementary Materials) is 197/589 = **33 %**. Note that aligned sequences have, by definition, the same length (here 589 aligned positions).



A **B**

Fig. (6). Projection of SRS zones in the final unified map onto the CYP450 fold
SRS boundaries along the CYP2C9 sequence (given in Table 3 above) were applied onto the CYP2C9 structure 1OG2. The colour codes for the SRSs are as follows: SRS1'a is in red, SRS1'b is in blue, SRS1 is in yellow, SRS(2,3) is in magenta, SRS4 is in orange, SRS5 is in hot-pink and SRS6 is in grey. The heme group is shown in turquoise. Parts of the CYP450 structure that have not been assigned to SRSs are in green. Panel (A) shows a "side-view" (relative to the heme group) and Panel (B) shows a "top view" (relative to the heme group). The structural model was rendered in pymol (www.pymol.org).

Projection of SRSs delimited in the final unified map onto the mammalian CYP450 fold (as typified by the CYP2C9 structure 1OG2) reveals that the majority of regions on the mammalian CYP450 fold that are proximal to the active site and that are not screened from the active site by other CYP450 structural elements are assigned as SRS zones. This suggests that the SRS map delimited herein is highly complete. Projection of SRSs onto the CYP450 fold provides a more direct and therefore more reliable measure for assessing the completeness of the SRS map than the ligand-based measure. Thus, conclusions drawn from observing the distribution of SRSs on the CYP450 fold override and out-weigh those inferred from the ligand-based chemical diversity measure. We therefore conclude that the new SRS map delimited herein is highly complete.

Review of the Explanatory Power of the New Map: Filling the Gaps in Gotoh's Map

We have assessed the explanatory power of the new unified map delimited herein by reviewing studies in the literature where residues, falling outside the SRS regions in Gotoh's map, have been shown through experiment to play a role in the determination of the substrate specificity of CYP450 isoforms. We are interested to see if the role of these 'non-Gotoh-SRS' residues in the determination of CYP450 substrate specificity can be explained by the reassignment of these residues to SRS zones in our map. The 11 cases we identified in the literature are reported in Table **4**, some of which have already been discussed in the Introduction (*vide supra*). We found that in five out of eleven of these cases, residues previously identified as non-Gotoh-SRS residues, now lie within SRS zones in our new map.

Table 4. Review of the Explanatory Power of the New Map

Study Number: Reference	Isoform: Residue(s)	Description of Findings in Study	Position in Gotoh's Map	Position in New Map
[a]1: [8]	CYP2C19: Pro220	Important determinant of omeprazole hydroxylase activity.	Occurs between Gotoh's SRS1 and SRS2 (F-G loop)	SRS(2,3)
[a]2: [8]	CYP2C19: Thr221	Important determinant of omeprazole hydroxylase activity.	Occurs between Gotoh's SRS1 and SRS2 (F-G loop)	SRS(2,3)
[a]3: [7]	CYP2B5: Pro221	Important determinant of progesterone hydroxylation activity.	Occurs between Gotoh's SRS2 and SRS3 (F-G loop)	SRS(2,3)
[a]4: [7]	CYP2B5: His120	Important determinant of progesterone hydroxylation activity.	C-term of Gotoh's SRS1	SRS1
[a]5: [22]	CYP1A2: residues 67-78 (L(S)QQYGDVLQIR)	Using mechanism-based inhibitors, this study identifies a stretch of rat CYP1A2 residues that includes residues involved in substrate recognition region.	N-term of Gotoh's SRS1	Occurs between SRS1'a and SRS1'b. Residues 67-78 (shown in bold) fall within SRS1'b
6: [22]	CYP1A2: residues 175-184 (FQEL-MAAVGR)	Identified as a substrate recognition region using mechanism-based inhibitors.	Occurs between Gotoh's SRS1 and SRS2	Non-SRS residues occurring mid-way between SRS1 and SRS(2,3)
7: [9]	CYP3A12: Thr187	Using chimeric enzymes and site-directed mutagenesis, this study identifies Thr187 as one of three residue positions responsible for differences in steroid hydroxylation between canine CYP3A12 and CYP3A26.	Occurs between Gotoh's SRS1 and SRS2	Non-SRS residue occurring closer to SRS(2,3) in the space between SRS1 and SRS(2,3)
8: [17]	CYP2C11: Phe187	In this study, site-directed mutagenesis on allelic variants in the Gun and Wistar rats identified Phe187 as an important residue for steroid hydroxylation	Occurs between Gotoh's SRS1 and SRS2	Non-SRS residue between SRS1 and SRS(2,3)
9: [17]	CYP2C11: Val4	In this study, site-directed mutagenesis on allelic variants in the Gun and Wistar rats identified Val4 as an important residue for steroid hydroxylation.	Occurs in the hydrophobic transmembrane domain	Non-SRS residue in hydrophobic transmembrane domain
10: [16]	CYP2B1: Ser407 Asn417 Ala419	Chimeric constructs and site directed mutagenesis studies show that substitution of these three residues in CYP2B1 protects the enzyme against inactivation by two acetylenic mechanism-based inhibitors.	Occur between SRS5 and SRS6	Non-SRS residues occurring between SRS5 and SRS6
11: [18]	CYP2B1: Leu58	CYP2B1 variant L58F, I114F affects the metabolism of testosterone, eliminating the production of 16 beta-hydroxylated metabolites but having no effect on the production of 6 alpha-hydroxylated and 17 keto metabolites.	C-term of Gotoh's SRS1	Non-SRS residue between SRS1'a and SRS1'b (4th residue after of SRS1'a)

[a] Case numbers shown in bold type denote CYP450 non-SRS residue positions (Gotoh's map) that have been experimentally shown to influence substrate specificity and whose role in determining substrate specificity we have successfully explained in this study by their re-assignment to SRS zones in our unified map.

The most well-known family of non-SRS regions that influence substrate specificity lie between Gotoh's SRS2 and SRS3, as illustrated in cases numbered 1, 2 and 3 in Table **4**. It is clear that our map resolves these cases since it combines SRS2 & SRS3 as well as the intervening sequence into a single SRS region.

He *et al.* (case no 4, Table **4**) previously identified the non-Gotoh-SRS residue CYP2B5 His120 as an important determinant of CYP2B5 progesterone hydroxylation activity [7]. This role of this residue in substrate specificity can be explained by the assignment of CYP2B5 His120 to the enlarged SRS1 in our new map.

A fascinating case that highlights the explanatory power of our new map is provided by the work of Yun *et al.* [22] (case no. 5, Table **4**) who used mechanism-based inhibitors to suggest that rat CYP1A2 positions 67-78 [L(S)QQYGDVLQIR] were involved in substrate recognition. Our work reveals that the three last residues of this peptide ["QIR"] lie within CYP1A2 SRS1'b in our new map, consistent with the experimental results of Yun *et al*.

In the remaining 6 cases our new map does not offer an immediate explanation for the role of these non-SRS residues in substrate specificity, as discussed briefly below.

Rat CYP1A2 positions 175-184 (FQELMAAVGR) identified by Yun *et al.* [22] (case no. 6, Table **4**) lie almost mid-way between the enlarged SRS1 and SRS(2,3) in our new map. The homologous residues in CYP2C9 (residues 156-163; "LRKTKASP") are surface residues that are distal to the active-site and that are extensively screened from the active-site by other parts of the CYP2C9 structure (Fig. **S7**). It is therefore difficult to speculate on a mechanism by which these residues could gain access to the active site and it seems possible instead that they have been mis-identified as substrate recognition regions in the previous work.

Fraser *et al.* identified canine CYP3A12 Thr187 as an important determinant in CYP3A12-catalysed steroid hydroxylation [9] (Case no. 7, Table **4**). Examination of the homologous residue positions in structures of CYP3A4 (Thr187) and CYP2C9 (Asn176) reveals that this residue is screened from the active site by the I-helix and that it either lies just on the plane of the heme co-factor in some CYP structures (e.g. in the CYP3A4 structure 1TQN) or quite beneath the plane of the heme co-factor in others (e.g. in the CYP2C9 structure 1R9O; Fig. **S8**). It is difficult to assess the likelihood of substrate access/contact of this position solely from static structures.

Biagini *et al.* (case no. 8, Table **4**) showed that the non-SRS residue CYP2C11 Phe187 is an important determinant of CYP2C11 testosterone hydroxylation [17]. This residue lies 6 amino acids upstream of SRS(2,3) in our new map and lies in the same structural neighbourhood as CYP3A12 Thr187 (case no. 7, Table **4**), but we note that these positions in CYP2C11 and CYP3A12 are infact not homologues of each other. Examination of the position of the homologous residue to CYP3A12 Thr187 in the CYP2C9 structure (Phe187) shows that it is screened from the active site by the I-helix and that it lies quite above the plane of the heme co-factor and in close proximity to SRS(2,3) (Fig. **S9**). It is therefore attractive to speculate that Biagini's data indicates a truer limit of SRS(2,3) than has been possible to delimit with the methods used in this study.

Biagini *et al.* also identified rat CYP2C11 Val4 as an important determinant in CYP2C11-catalysed steroid hydroxylation [17] (case no. 9, Table **4**). Early sequence positions such as CYP2C11 Val4 lie in the hydrophobic CYP450 membrane anchor polypeptide and do not form part of the CYP450 catalytic domain. As such, this residue position probably influences enzyme performance by mechanisms outside the scope of those we are studying here.

The role of CYP2B1 residues Ser407, Asn417 and Ala419 (case no. 10, Table **4**) in providing protection against mechanism-based inactivation by acetylenic inhibitors [62] cannot be explained by our new SRS map. However, we note that Asn417 and Ala419 lie in the range of residues CYP2B1 Phe413-Ile435, a segment that immediately precedes the strictly conserved Cys436 which forms an obligate thiolate with the heme, suggesting possible alternative explanations for the experimental observations based on chemical modification of Cys436 by the acetylenic inhibitors.

Finally, we note that the involvement of rat CYP2B1 Leu58 in modulating CYP2B1 testosterone hydroxylation activity [66] (Case number 11, Table **4**) cannot be explained by way of re-assignment to our unified SRS zones since it lies 4 residue after the C-terminal boundary of SRS1'a in CYP2B1.

DISCUSSION

Taking advantage of the wealth of new CYP450 structural data and new computational tools that have become available over the last 10 years, we have created an expanded, unified Substrate Recognition Site map for mammalian Cytochrome P450s. To do this, we collected 43 structures of mammalian CYP450-ligand complexes and used the program Ligplot [34] to identify contact residues in a suitable subset of those structures. From this, we created an 'X-ray structures' SRS map by projecting the identified contact residues onto an alignment of sequences of mammalian CYP450s (Fig. **1**). In a parallel approach we generated plausible CYP450-ligand complexes by docking known CYP450 substrates into structures of CYP450 isoforms using the program AutoDock Vina [65]. In analysing the docking modes, we used arbitrary cut-offs to exclude those where the computed binding free energy was too low (> -1 kcal/mol) or where no atom of the ligand was within 5 Å from the heme iron. In considering these cut-offs, we note that, given the very high energy of the iron (IV) oxo species generated during the catalytic cycle, once a compound has bound and is in close-enough proximity to the iron (IV) oxo species, the chemical question becomes which atom will be oxidized - which is probably controlled as much by precise available local conformations as it is by bond energies in the substrate. We did not attempt to further evaluate in this work why each compound was a substrate, but from this *in silico* study, we created a 'docking' SRS map by projecting contact residues onto an alignment of sequences of mammalian CYP450s (Fig. **S3**). The X-ray structures SRS map and the docking SRS map largely agreed with each other and importantly include the main features in Gotoh's original SRS map, but equally importantly they also exhibit new features that were absent in Gotoh's map. Features in common with Gotoh's map included the appearance of Gotoh's six original SRS regions in the same locations as they appear in Gotoh's map.

We subsequently combined our docking- and X-ray structures SRS maps to obtain the new, unified SRS map described in this work (Fig. **5**). Important differences between the new map delimited herein and the old map by Gotoh include the following: (i) two new regions appear before SRS1 in our map – we have designated the new regions as SRS1'a and SRS1'b; (ii) SRS1 is much wider in our map than in Gotoh's – it covers 49 aligned-sequence positions in our map versus 28 aligned-sequence positions in Gotoh's map; (iii) SRS2 and SRS3 are co-joined in our map to give a single large SRS region which we have designated as SRS(2,3) to maintain consistency with Gotoh's numbering. SRS2 covers 9 aligned-sequence positions in Gotoh's map while SRS3 covers 10 positions whereas our newly-designated SRS(2,3) covers 60 aligned-sequence positions; (iv) SRS4 in our map has a C-terminal extension of 3 aligned-sequence positions relative to SRS4 in Gotoh's map; (v) SRS5 in our map has a N-terminal extension of 2 aligned-sequence positions and a C-terminal extension of 1 aligned-sequence position relative to SRS5 in Gotoh's map; & (vi) SRS6 in our map has a N-terminal extension of 7 aligned-sequence positions and a C-terminal extension of 2 aligned-sequence positions relative to SRS6 in Gotoh's map.

Our work shows that in mammalian CYP450, Gotoh's SRS2 and SRS3 are in fact, one continuous SRS region [SRS (2,3)]. This is consistent with the observation that the conformation of the F-G loop - which lies between Gotoh's SRS2 and SRS3 - is highly variable in X-ray structures of mammalian CYP450 isoforms and is able to adopt conformations that are more proximal to (and unscreened from) the active site than is possible in many bacterial CYP450s (including, notably, structures of CYP101A – the template in Gotoh's delimitations). For example, in the CYP2C9 structure 1OG2 the F-G loop is almost completely folded into two small helices that are connected by a short loop that allows a bent connec-

tion between the two small helices. In that same structure the F-helix and the G-helix are completely folded in helical conformation. Furthermore, the F-helix, F-G loop and G-helix in 1OG2 closely associate with each other and with the other secondary structure elements of the CYP450 protein to form a "lid" above the active site cleft. This configuration brings the large F-G loop close enough to the active site to make it likely that it participates in substrate recognition. Similarly, in the CYP3A4 structures 1TQN, 1W0F and 2V0M, the F-helix is partially unwound and part of the F-G loop is folded into a helix separated from the G-helix by a "loop part" of the F-G loop. Again, the F-helix, F-G loop and G-helix form close associations with each other and with the other secondary structure elements of the CYP450 protein and form a "lid" above the active site cleft. This configuration of unscreened proximity to the active site suggests that the mammalian F-G loop could be involved in substrate recognition, as suggested by our unified SRS map.

An important difference between mammalian and bacterial F-G loops is that F-G loops are generally much smaller in bacterial CYP450 than in mammalian CYP450: They typically comprise ~ 21 amino acids in mammalian CYP450 (see CYP2C9 structure 1R9O, Ser209-Thr229 inclusive for example) but typically comprise only ~ 8 amino acids in bacterial CYP450 (see CYP101A structure 3CPP, Thr185-Thr192 inclusive for example), as has been noted previously [82]. We also observe, however, that in a subset of bacterial CYP450 structures - importantly, including the one used by Gotoh to delimit his mammalian SRSs (structure 3CPP) - the F-G loop is folded away from the rest of the protein and forms a hoop that lies in a plane almost perpendicular to the plane defined by the F-helix and the first half of the G-helix, as shown in Fig. **S10**. As a result of this configuration, Gotoh's approach of identifying residues within 10 Å of any atom of the bound camphor substrate in structure 3CPP as SRS residues had the effect of excluding F-G residues Asp188-Thr192, giving rise to the separation between SRS2 and SRS3 calculated in bacterial structures that was subsequently transferred (erroneously) to mammalian CYP450s in Gotoh's map. We note that the inclusion of Asp188-Thr192 as SRS residues would have required Gotoh to set the maximum distance threshold for interaction at 18 Å, well beyond the commonly-accepted thresholds for interaction (interaction distance cut-offs are typically less than 10 Å [34, 83, 84]).

An interesting prediction of our unified map is that 33 % of the CYP450 sequence is involved in substrate contact (*c.f.* Gotoh's figure of 16 %), suggesting that the overall spread of SRS regions had previously been substantially underestimated. This observation naturally leads to the questions "how complete is the description of SRS regions in the new map" and "how can we assess the degree of completeness"? Projection of the SRS regions delimited in our map onto the CYP450 fold (Fig. **6**) reveals that the majority of locations on the CYP450 fold that are unobstructed from and are proximal to the catalytic heme iron have been included as SRS regions in our new unified map. This suggests that the new map given here represents a high degree of completeness in the description of CYP450 SRS regions. An important caveat here though is that it is well appreciated that docking algorithms are not particularly good at accurately predicting the energetics of different possible ligand binding modes, plus even programs such as MetaSite make prediction of multiple likely oxidation sites on a given substrate but do not necessarily get these in the right order compared to experimental observations. Taken together, it is possible that our docking studies will therefore tend to implicate a wider range of residues as part of SRSs than are actually involved through side chain-substrate interactions. In this regard, we note that many residues in our unified SRSs are facing outwards from the CYP450 active site (see Fig. **5** for examples); whilst this does not automatically preclude their involvement in substrate binding as main chain atoms can also participate, it may mean that, in the context of substrate recognition, such residues are more permissive to mutation. Further experimentation will be required to explore this.

In calculating that the coverage achieved by the X-ray structure ligands is only 26 % of CYP450 substrate space, it is interesting to note that the X-ray- and docking SRS maps determined here are in close agreement, with both identifying for example the new SRSs 1'a, 1'b and 1'c, the notable exceptions being that the co-joining of SRS1 & SRS1'c and of SRSs 2 & 3 were only identified in the latter. Thus, despite the apparently substantially incomplete representation of the chemical types that constitute the CYP450 substrate space, the limited set of X-ray structures currently available seemingly already give a reasonably complete picture of how mammalian CYP450s recognise their substrates. In other words, it appears that most parts of the complete SRS zone on the CYP450 fold are sampled by a relatively small number of ligands, which in turn suggests that it may not be necessary to observe a large number of CYP450-ligand complexes (for example, > 50 % of the CYP450 substrate universe) in order to achieve the definition of a highly complete SRS map. It is worth emphasising of course that the docking studies performed here have played a critical role, both in confirming the usage of and extending the span of SRSs identified in the X-ray map, as well as in verifying the existence and usage of, for example, SRS(2,3) in a range of CYP450s for which there are currently no appropriate ligand-bound X-ray structures.

We have re-examined the 11 cases we identified in the literature of non-Gotoh-SRSs that are apparently involved in determining substrate recognition (Table **4**) and found that our new unified SRS map provides direct and ready explanation for 5 of these 11 cases. Of the remaining 6 cases, one involves surface residues that are distal to and extensively screened from the active site, another lies completely outside the catalytic domain in the N-terminal hydrophobic membrane anchor peptide, two more are screened from the active site by the I-helix, and a further one is in proximity to the cysteine that forms the obligate heme thiolate. It seems likely therefore that in at least 5 of the 6 cases that remain unexplained by our unified SRS-map, the origin of their effect infact lies in longer-range, 'second sphere' cooperative effects rather than in direct interaction with substrate molecules. Thus, our new, unified SRS map has neatly accounted for each of the literature examples of non-Gotoh-SRS regions that could plausibly directly contact substrates, thereby immediately demonstrating the utility of our expanded, unified SRS map.

Further development of our new SRS maps by scoring the individual positions in SRS zones according to how frequently their use has been observed in X-ray crystal structures and/or in docking complexes will provide a useful resource in studies that seek to modulate CYP450 specificity by mutagenesis; such "scored position" maps - of which Fig. **S2** is a preliminary example - would for example inform future structure-activity studies on CYP450s by providing ranked lists of specificity-determining residues to be targeted by mutagenesis.

In considering other future possible uses of our expanded, unified SRS map, it is interesting to reflect on the fact that the human CYP450 enzymes that are involved in primary drug metabolism are known to be highly polymorphic in the population, with the distribution of polymorphisms varying between ethnic groups [85, 86]. The impact of the vast majority of such polymorphic variations in human CYP450 genes on the metabolism of new and existing drugs - as well as on adverse drug-drug reactions - remains poorly understood today, but clear individual examples exist in the literature. For example, CYP2C9 has over 35 non-synonymous single nucleotide polymorphisms (SNPs) [www.cypalleles.ki.se] distributed throughout the protein sequence (Fig. **1**) and the effect of individual CYP2C9 variant alleles on warfarin dose requirement has been documented, with patients with R144C and I359L mutations having

considerably lower warfarin clearance rates [87-89]. Interestingly, site-directed mutagenesis studies have shown that single amino acid changes in CYP450s can affect not only rates of reaction, but also substrate specificity, enantioselectivity and regioselectivity of oxidation [90-92]. We therefore anticipate that our new, unified SRS map may help in gaining a better understanding of structure-function relationships in CYP450s, as an important step towards predicting the effects of SNPs on drug metabolism and on adverse drug reactions; indeed, we note that that numerous SNPs in just CYP2C9, CYP2B4 and CYP3A4 lie within our new, expanded SRSs (Fig. **5**), suggesting a potential direct influence of such SNPs on substrate recognition.

ABBREVIATIONS

SRS = Substrate recognition site

CYP450 = Cytochrome P450

ACKNOWLEDGEMENTS

A.Z. wishes to thank the University of Cape Town Clinical Infectious Diseases Research Initiative (CIDRI) for a research fellowship. Y.C.W. wishes to acknowledge the MOE grant AcRF Tier 1 (R-148-000-136-112). L.C wishes to thank the National Research Foundation and the Harry Crossley Foundation for bursaries. J.B. thanks the South African Research Chair Initiative (SARChI) for a research chair.

CONFLICT OF INTEREST

No conflict of interest declared.

AUTHOR CONTRIBUTIONS

A.Z., J.B. & Y.C.W designed research; A.Z., L.Y.C, L.C. & Y.C.W performed research; A.Z., L.Y.C., L.C., J.B. & Y.C.W wrote the paper.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

REFERENCES

[1] Bridges, A.; Gruenke, L.; Chang, Y.T.; Vakser, I.A.; Loew, G.; Waskell, L. Identification of the binding site on cytochrome P450 2B4 for cytochrome b5 and cytochrome P450 reductase. *J. Biol. Chem.*, **1998**, *273*, 17036-49.

[2] Zawaira, A.; Matimba, A.; Masimirembwa, C. Prediction of sites under adaptive evolution in cytochrome P450 sequences and their relationship to substrate recognition sites. *Pharmacogenet. Genomics*, **2008**, *18*, 467-76.

[3] Lamb, D.C.; Lei, L.; Warrilow, A.G.; Lepesheva, G.I.; Mullins, J.G.; Waterman, M.R.; Kelly, S.L. The first virally encoded cytochrome p450. *J. Virol.*, **2009**, *83*, 8266-9.

[4] Nelson, D.R.; Koymans, L.; Kamataki, T.; Stegeman, J.J.; Feyereisen, R.; Waxman, D.J.; Waterman, M.R.; Gotoh, O.; Coon, M.J.; Estabrook, R.W.; Gunsalus, I.C.; Nebert, D.W. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, **1996**, *6*, 1-42.

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

[6] Urlacher, V.B.; Schmid, R.D. Protein engineering of the cytochrome P450 monooxygenase from Bacillus megaterium. *Methods Enzymol.*, **2004**, *388*, 208-24.

[7] He, Y.Q.; Harlow, G.R.; Szklarz, G.D.; Halpert, J.R. Structural determinants of progesterone hydroxylation by cytochrome P450 2B5: the role of nonsubstrate recognition site residues. *Arch. Biochem. Biophys.*, **1998**, *350*, 333-9.

[8] Ibeanu, G.C.; Ghanayem, B.I.; Linko, P.; Li, L.; Pederson, L.G.; Goldstein, J.A. Identification of residues 99, 220, and 221 of human cytochrome P450 2C19 as key determinants of omeprazole activity. *J. Biol. Chem.*, **1996**, *271*, 12496-501.

[9] Fraser, D.J.; He, Y.Q.; Harlow, G.R.; Halpert, J.R. Use of chimeric enzymes and site-directed mutagenesis for identification of three key residues responsible for differences in steroid hydroxylation between canine cytochromes P-450 3A12 and 3A26. *Mol. Pharmacol.*, **1999**, *55*, 241-7.

[10] He, Y.; Luo, Z.; Klekotka, P.A.; Burnett, V.L.; Halpert, J.R. Structural determinants of cytochrome P450 2B1 specificity: evidence for five substrate recognition sites. *Biochemistry*, **1994**, *33*, 4419-24.

[11] He, Y.Q.; He, Y.A.; Halpert, J.R. Escherichia coli expression of site-directed mutants of cytochrome P450 2B1 from six substrate recognition sites: substrate specificity and inhibitor selectivity studies. *Chem. Res. Toxicol.*, **1995**, *8*, 574-9.

[12] 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*, 5396-402.

[13] Roussel, F.; Khan, K.K.; Halpert, J.R. The importance of SRS-1 residues in catalytic specificity of human cytochrome P450 3A4. *Arch. Biochem. Biophys.*, **2000**, *374*, 269-78.

[14] 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*, 8831-9.

[15] Conley, A.; Mapes, S.; Corbin, C.J.; Greger, D.; Graham, S. Structural determinants of aromatase cytochrome p450 inhibition in substrate recognition site-1. *Mol. Endocrinol.*, **2002**, *16*, 1456-68.

[16] Von Weymarn, L.B.; Sridar, C.; Hollenberg, P.F. Identification of amino acid residues involved in the inactivation of cytochrome P450 2B1 by two acetylenic compounds: the role of three residues in nonsubstrate recognition Sites. *J. Pharmacol. Exp. Ther.*, **2004**, *311*, 71-9.

[17] Biagini, C.P.; Philpot, R.M.; Celier, C.M. Nonsubstrate recognition site residues are involved in testosterone hydroxylation by cytochrome P450 CYP 2C11. *Arch. Biochem. Biophys.*, **1999**, *361*, 309-14.

[18] Aoyama, T.; Korzekwa, K.; Nagata, K.; Adesnik, M.; Reiss, A.; Lapenson, D.P.; Gillette, J.; Gelboin, H.V.; Waxman, D.J.; Gonzalez, F.J. Sequence requirements for cytochrome P-450IIB1 catalytic activity. Alteration of the stereospecificity and regioselectivity of steroid hydroxylation by a simultaneous change of two hydrophobic amino acid residues to phenylalanine. *J. Biol. Chem.*, **1989**, *264*, 21327-33.

[19] Lindberg, R.L.; Negishi, M. Alteration of mouse cytochrome P450coh substrate specificity by mutation of a single amino-acid residue. *Nature*, **1989**, *339*, 632-4.

[20] Juvonen, R.O.; Iwasaki, M.; Negishi, M. Structural function of residue-209 in coumarin 7-hydroxylase (P450coh). Enzyme-kinetic studies and site-directed mutagenesis. *J. Biol. Chem.*, **1991**, *266*, 16431-5.

[21] Kedzie, K.M.; Balfour, C.A.; Escobar, G.Y.; Grimm, S.W.; He, Y.A.; Pepperl, D.J.; Regan, J.W.; Stevens, J.C.; Halpert, J.R. Molecular basis for a functionally unique cytochrome P450IIB1 variant. *J. Biol. Chem.*, **1991**, *266*, 22515-21.

[22] Yun, C.H.; Hammons, G.J.; Jones, G.; Martin, M.V.; Hopkins, N.E.; Alworth, W.L.; Guengerich, F.P. Modification of cytochrome P450 1A2 enzymes by the mechanism-based inactivator 2-ethynylnaphthalene and the photoaffinity label 4-azidobiphenyl. *Biochemistry*, **1992**, *31*, 10556-63.

[23] Lin, H.L.; Zhang, H.; Waskell, L.; Hollenberg, P.F. Threonine-205 in the F helix of p450 2B1 contributes to androgen 16 beta-hydroxylation activity and mechanism-based inactivation. *J. Pharmacol. Exp. Ther.*, **2003**, *306*, 744-51.

[24] Wester, M.R.; Johnson, E.F.; Marques-Soares, C.; Dijols, S.; Dansette, P.M.; Mansuy, D.; Stout, C.D. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 A resolution: evidence for an induced fit model of substrate binding. *Biochemistry*, **2003**, *42*, 9335-45.

[25] Williams, P.A.; Cosme, J.; Ward, A.; Angove, H.C.; Matak Vinkovic, D.; Jhoti, H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature*, **2003**, *424*, 464-8.

[26] Porubsky, P.R.; Battaile, K.P.; Scott, E.E. Human cytochrome P450 2E1 structures with fatty acid analogs reveal a previously unobserved binding mode. *J. Biol. Chem.*, *285*, 22282-90.

[27] Strushkevich, N.; Usanov, S.A.; Plotnikov, A.N.; Jones, G.; Park, H.W. Structural analysis of CYP2R1 in complex with vitamin D3. *J. Mol. Biol.*, **2008**, *380*, 95-106.

[28] Yano, J.K.; Hsu, M.H.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat. Struct. Mol. Biol.*, **2005**, *12*, 822-3.

[29] Sansen, S.; Yano, J.K.; Reynald, R.L.; Schoch, G.A.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *J. Biol. Chem.*, **2007**, *282*, 14348-55.

[30] Mast, N.; White, M.A.; Bjorkhem, I.; Johnson, E.F.; Stout, C.D.; Pikuleva, I.A. Crystal structures of substrate-bound and substrate-free cytochrome P450 46A1, the principal cholesterol hydroxylase in the brain. *Proc. Natl. Acad. Sci. U S A*, **2008**, *105*, 9546-51.

[31] Ghosh, D.; Griswold, J.; Erman, M.; Pangborn, W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature*, **2009**, *457*, 219-23.

[32] Li, Y.C.; Chiang, C.W.; Yeh, H.C.; Hsu, P.Y.; Whitby, F.G.; Wang, L.H.; Chan, N.L. Structures of prostacyclin synthase and its complexes with substrate analog and inhibitor reveal a ligand-specific heme conformation change. *J. Biol. Chem.*, **2008**, *283*, 2917-26.

[33] Annalora, A.J.; Goodin, D.B.; Hong, W.X.; Zhang, Q.; Johnson, E.F.; Stout, C.D. Crystal structure of CYP24A1, a mitochondrial cytochrome P450 involved in vitamin D metabolism. *J. Mol. Biol.*, **2010**, *396*, 441-51.

[34] Wallace, A.C.; Laskowski, R.A.; Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, **1995**, *8*, 127-34.

[35] Kjellander, B.; Masimirembwa, C.M.; Zamora, I. Exploration of enzyme-ligand interactions in CYP2D6 & 3A4 homology models and crystal structures using a novel computational approach. *J. Chem. Inf. Model.*, **2007**, *47*, 1234-47.

[36] Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins*, **2004**, *57*, 225-42.

[37] Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.*, **2005**, *48*, 6970-9.

[38] Zhou, D.; Afzelius, L.; Grimm, S.W.; Andersson, T.B.; Zauhar, R.J.; Zamora, I. Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. *Drug Metab. Dispos.*, **2006**, *34*, 976-83.

[39] Ahlstrom, M.M.; Ridderstrom, M.; Zamora, I.; Luthman, K. CYP2C9 structure-metabolism relationships: optimizing the metabolic stability of COX-2 inhibitors. *J. Med. Chem.*, **2007**, *50*, 4444-52.

[40] Schleinkofer, K.; Sudarko; Winn, P.J.; Ludemann, S.K.; Wade, R.C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep.*, **2005**, *6*, 584-9.

[41] Wade, R.C.; Motiejunas, D.; Schleinkofer, K.; Sudarko; Winn, P.J.; Banerjee, A.; Kariakin, A.; Jung, C. Multiple molecular recognition mechanisms. Cytochrome P450--a case study. *Biochim. Biophys. Acta*, **2005**, *1754*, 239-44.

[42] Fishelovitch, D.; Shaik, S.; Wolfson, H.J.; Nussinov, R. Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism. *J. Phys. Chem. B*, **2009**, *113*, 13018-25.

[43] Research Collaboratory for Structural Bioinformatics, Protein Data Bank: A Resource for Studying Biological Macromolecules. http://www.pdb.org (Accessed 1 June, **2011**).

[44] European Bioinformatics Institute: Data Resoures and Tools. http://www.ebi.ac.uk (Accessed 1 June, **2011**).

[45] Jeanmougin, F.; Thompson, J.D.; Gouy, M.; Higgins, D.G.; Gibson, T.J. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.*, **1998**, *23*, 403-5.

[46] Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **1994**, *22*, 4673-80.

[47] Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **1997**, *25*, 4876-82.

[48] Ioannides, C. *Cytochromes P450: Role in the Metabolism and Toxicity of Drugs and other Xenobiotics.* RSC Publishing: Cambridge, **2008**.

[49] Coon, M.J. CYTOCHROME P450: Nature's Most Versatile Biological Catalyst. *Ann. Rev. Pharmacol. Toxicol.*, **2005**, *45*, 1-25.

[50] Haoming, Z.; Myshkin, E.; Waskell, L. Role of cytochrome b5 in catalysis by cytochrome P450 2B4. *Biochem. Biophys. Res. Commun.*, **2005**, *338*, 499-506.

[51] Department of Medicine, Indiana University: P450 Drug Interaction Table. http://medicine.iupui.edu/clinpharm/ddis/table.asp (Accessed 1 June, **2011**).

[52] Edmund Hayes: Cytochrome P-450 and Drug Interactions. http://www.edhayes.com/CYP450-3.html (Accessed 14 October, **2010**).

[53] Genelex Corporation: P450 Drug-Interactions Substrates Table. http://www.healthanddna.com/Druglist.pdf (Accessed 1 June, **2011**).

[54] Pharmacy Society of Wisconsin: Clinically Significant Cytochrome P450 Drug Interactions. http://www.pswi.org/professional/pharmaco/Cytochrome.pdf (Accessed 1 June, **2011**).

[55] Structural Genomics Consortium: Human cytochrome P450, family 2, R1. http://www.sgc.utoronto.ca/SGC-WebPages/Structure Description/2OJD.php (Accessed 1 June, **2011**).

[56] Just the Berries: Clinically Significant Cytochrome P-450 (CYP 450) Drug Interactions. http://www.theberries.ca/archives/cyp450.html (Accessed 1 June, **2011**).

[57] UniProt Consortium: Q6VVX0 (CP2R1_HUMAN). http://www.uniprot.org/uniprot/Q6VVX0 (Accessed 1 June, **2011**).

[58] Thomson Reuters: MicroMedex 2.0. http://www.micromedex.com (Accessed 1 June, **2011**).

[59] Accerlrys: Pipeline Pilot. http://accelrys.com/products/pipeline-pilot/ (Accessed 1 June, **2011**).

[60] Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.*, **2011**, *32*, 1466-74.

[61] National Center for Biotechnology Information: PubChem. http://pubchem.ncbi.nlm.nih.gov (Accessed 1 June, **2011**).

[62] Umetrics: SIMCA-P+. http://www.umetrics.com/simca (Accessed 1 June, **2011**).

[63] Open Babel: The Open Source Chemistry Toolbox. http://openbabel.org/wiki/Main_Page (Accessed 1 June, **2011**).

[64] Gruber, J.; Zawaira, A.; Saunders, R.; Barrett, C.P.; Noble, M.E. Computational analyses of the surface properties of protein-protein interfaces. *Acta Crystallogr. D Biol. Crystallogr.*, **2007**, *63*, 50-7.

[65] Trott, O.; Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **2010**, *31*, 455-61.

[66] The Scripps Research Institute: AutoDock. http://autodock.scripps.edu (Accessed 1 June, **2011**).

[67] Wang, R.; Lu, Y.; Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *J. Med. Chem.*, **2003**, *46*, 2287-2303.

[68] National University of Signapore: PaDEL-ADV. http://padel.nus.edu.sg/software/padeladv/ (Accessed 1 June, **2011**).

[69] Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E.L. Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. *Curr. Pharm. Des.*, **2006**, *12*, 2111-20.

[70] MacQueen, J.B. *Some methods for classification and analysis of multivariate observations. In Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press: Berkeley, CA, **1967**.

[71] Hartigan, J.A.; Wong, M.A. A K-means clustering algorithm. *Applied Statistics*, **1979**, *28*, 100-108.

[72] Rapid-I GmbH: RapidMiner. http://rapid-i.com/content/view/181/-190 (Accessed 1 June, **2011**).

[73] Ravindran, S.; Zharikova, O.L.; Hill, R.A.; Nanovskaya, T.N.; Hankins, G.D.; Ahmed, M.S. Identification of glyburide metabolites formed by hepatic and placental microsomes of humans and baboons. *Biochem. Pharmacol.*, **2006**, *72*, 1730-7.

[74] Zharikova, O.L.; Ravindran, S.; Nanovskaya, T.N.; Hill, R.A.; Hankins, G.D.; Ahmed, M.S. Kinetics of glyburide metabolism by hepatic and placental microsomes of human and baboon. *Biochem. Pharmacol.*, **2007**, *73*, 2012-9.

[75] Zhang, H.; Henion, J.; Yang, Y.; Spooner, N. Application of atmospheric pressure ionization time-of-flight mass spectrometry coupled with liquid chromatography for the characterization of *in vitro* drug metabolites. *Anal. Chem.*, **2000**, *72*, 3342-8.

[76] Tiller, P.R.; Land, A.P.; Jardine, I.; Murphy, D.M.; Sozio, R.; Ayrton, A.; Schaefer, W.H. Application of liquid chromatography-mass spectrometry(n) analyses to the characterization of novel glyburide metabolites formed *in vitro*. *J. Chromatogr. A*, **1998**, *794*, 15-25.

[77] Boyd, A.E., 3rd; Aguilar-Bryan, L.; Nelson, D.A. Molecular mechanisms of action of glyburide on the beta cell. *Am. J. Med.*, **1990**, *89*, 3S-10S; discussion 51S-53S.

[78] Englert, H.C.; Heitsch, H.; Gerlach, U.; Knieps, S. Blockers of the ATP-sensitive potassium channel SUR2A/Kir6.2: a new approach to prevent sudden cardiac death. *Curr. Med. Chem. Cardiovasc. Hematol. Agents*, **2003**, *1*, 253-71.

[79] Montorsi, F.; Salonia, A.; Briganti, A.; Barbieri, L.; Zanni, G.; Suardi, N.; Cestari, A.; Montorsi, P.; Rigatti, P. Vardenafil for the treatment of erectile dysfunction: a critical review of the literature based on personal clinical experience. *Eur. Urol.*, **2005**, *47*, 612-21.

[80] Ku, H.Y.; Ahn, H.J.; Seo, K.A.; Kim, H.; Oh, M.; Bae, S.K.; Shin, J.G.; Shon, J.H.; Liu, K.H. The contributions of cytochromes P450 3A4 and 3A5 to the metabolism of the phosphodiesterase type 5 inhibitors sildenafil, udenafil, and vardenafil. *Drug Metab. Dispos.*, **2008**, *36*, 986-90.

[81] Strano-Rossi, S.; Anzillotti, L.; de la Torre, X.; Botre, F. A gas chromatography/mass spectrometry method for the determination of sildenafil, vardenafil and tadalafil and their metabolites in human urine. *Rapid Commun. Mass Spectrom.*, **2010**, *24*, 1697-706.

[82] Li, H.; Poulos, T.L. Conformational dynamics in cytochrome P450-substrate interactions. *Biochimie*, **1996**, *78*, 695-9.

[83] Tina, K.G.; Bhadra, R.; Srinivasan, N. PIC: Protein Interactions Calculator. *Nucleic Acids Res.*, **2007**, *35*, W473-6.

[84] Babu, M.M. NCI: A server to identify non-canonical interactions in protein structures. *Nucleic Acids Res.*, **2003**, *31*, 3345-8.

[85] Ingelman-Sundberg, M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol. Sci.*, **2004**, *25*, 193-200.

[86] Ingelman-Sundberg, M.; Sim, S.C.; Gomez, A.; Rodriguez-Antona, C. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacol. Ther.*, **2007**, *116*, 496-526.

[87] Rettie, A.; Wienkers, L.; Gonzalez, F.; Trager, W.; Korzekwa, K. Impaired (S)-warfarin metabolism catalysed by the R144C allelic variant of CYP2C9. *Pharmacogenetics*, **1994**, *4*, 39-42.

[88] Haining, R.; Hunter, A.; Veronese, M.; Trager, W.; Rettie, A. Allelic variants of human cytochrome P450 2C9: baculovirus-mediated expression, purification, structural characterization, substrate stereoselectivity, and prochiral selectivity of the wild-type and I359L mutant forms. *Arch. Biochem. Biophys.*, **1996**, *333*, 447-458.

[89] Aithal, G.; Day, C.; Kesteven, P.; Daly, A. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet*, **1999**, *353*, 717-719.

[90] Jung, F.; Griffin, K.J.; Song, W.; Richardson, T.H.; Yang, M.; Johnson, E.F. Identification of amino acid substitutions that confer a high affinity for sulfaphenazole binding and a high catalytic efficiency for warfarin metabolism to P450 2C19. *Biochemistry*, **1998**, *37*, 16270-16279.

[91] Miyazaki, M.; Nakamura, K.; Fujita, Y.; Guengerich, F.P.; Horiuchi, R.; Yamamoto, K. Defective activity of recombinant cytochromes P450 3A4.2 and 3A4.16 in oxidation of midazolam, nifedipine, and testosterone *Drug metab. Dispos.*, **2008**, *36*, 2287-2291.

[92] Sata, F.; Sapone, A.; Elizondo, G.; Stocker, P.; Miller, V.P.; Zheng, W.; Raunio, H.; Crespi, C.L.; Gonzalez, F.J. CYP3A4 allelic variants with amino acid substitutions in exons 7 and 12: evidence for an allelic variant with altered catalytic activity. *Clin. Pharmacol. Ther.*, **2000**, 48-56.

Received: March 20, 2011 **Revised: May 14, 2011** **Accepted: May 20, 2011**

Journal of Structural Biology 173 (2011) 282–293

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

On the deduction and analysis of singlet and two-state gating-models from the static structures of mammalian CYP450

Alexander Zawaira [a,*], Lauren Coulson [a], Marco Gallotta [b], Owen Karimanzira [a], Jonathan Blackburn [a,*]

[a] Division of Medical Biochemistry, Institute for Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa
[b] Department of Computer Science, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa

ARTICLE INFO

Article history:
Received 9 July 2010
Received in revised form 22 September 2010
Accepted 29 September 2010
Available online 12 October 2010

Keywords:
Cytochrome P450
Tunnels
Gating residues
Molecular dynamics
Random expulsion molecular dynamics

ABSTRACT

Differential tunnel-opening patterns were established in static structures of mammalian CYP450 isoforms and subsequently applied to identify tunnel-intersecting residues. The identified tunnel-intersecting residues permitted the subsequent construction of gating models via the identification of intra-protein interactions. We define 28 two-state gating models and 37 singlet gating-residue models. Our results reveal the preponderance of aromatic gating residues in CYP3A4 and CYP2A6, whereas we find a preponderance of polar/charged residues in CYP2C5. In CYP2C8 there is balanced presence of polar/charged and hydrophobic aliphatic residues in gating models, whilst in CYP2C9 there is balanced presence of all residue-types. These patterns suggest fast evolutionary dynamics for gating residues and we find that the average rate of evolution of gating residues in CYP2C5, CYP2C8, CYP2C9 and CYP2A6 is significantly faster than the average rate of evolution of the entire sequence.

Our study identifies 67% of calculable gating models identified in the literature by molecular dynamics approaches and 92% of residues appearing in literature models appear in our models. However, only 6% of the models identified in this work had been previously-described in the literature. This suggests that our study has defined the most comprehensive list yet of tunnel-gating models in mammalian CYP450 and in doing so have created a benchmark for molecular dynamics approaches to the ligand-tunnelling problem in CYP450.

1. Introduction

The Cytochromes P450 (CYP450, EC 1.14.14.1) are a superfamily of hemoprotein monooxygenases that catalyse the oxidation of hydrophobic endogenous and xenobiotic substrates (Bridges et al., 1998). CYP450[1] homologs have been sequenced from all lineages of life – including eukaryotes and bacteria (Werck-Reichhart and Feyereisen, 2000; Zawaira et al., 2008). Furthermore, a viral CYP450 from the giant virus *Acanthamoeba polyphaga* has recently been characterized (Lamb et al., 2009).

The crystal structures of CYP450 determined so far show the same overall fold in which the active site is buried in the core of the protein (Hasemann et al., 1995; Sansen et al., 2007; Schoch et al., 2004; Wester et al., 2004; Yano et al., 2004). In some of these crystal structures, such as that of 6-deoxyerythronolide-bound CYP450eryF (CYP107A1, PDB code 1OXA; (Cupp-Vickery and Poulos, 1995)), the active site is completely cut-off from the bulk solvent. That is, the protein surface is not accessible from the active site and there are absolutely no access routes that would permit molecules of at least the size of water molecules, represented as spherical probes of radius 1.4 Å in calculations, to access the active site (Wade et al., 2004). Such structures have been denoted as "closed" structures (Wade et al., 2004) and are consistent with the CYP450 catalytic cycle requirement of complete isolation of the active site from the bulk solvent in order to prevent uncoupling reactions and to ensure reaction specificity.

In other crystal structures tunnels wide enough to accommodate only molecules at most the size of water molecules, connect the buried active site to the protein surface (Wade et al., 2004). In still other crystal structures, large tunnels from the active site to the protein surface that are sterically accessible for typical size substrate molecules have been delimited. Such crystal structures have been denoted as "wide-open" (Wade et al., 2004) and include a structure of substrate-free CYP55A1 (PDB code 1ROM; (Park et al., 1997a,b)) and a structure of substrate-free CYP450BM-3 (PDB code 2HPD; (Boddupalli et al., 1992; Fulco, 1991)). In these "wide-open" crystal structures, sterically accessible access routes have been delimited in distinct parts of the CYP450 structure/fold (Wade et al., 2004).

* Corresponding authors. Address: Room N3.03, Wernher & Beit Building North, Department of Medical Biochemistry, Institute for Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory 7925, South Africa (A. Zawaira).
E-mail address: Alexander.Zawaira@uct.ac.za (A. Zawaira).

[1] *Abbreviations used:* CYP450, cytochrome P450; REMD, random expulsion molecular dynamics; MD, molecular dynamics.

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.09.026

This complex picture of the presence/absence of tunnels linking the CYP450 active site to the bulk solvent has driven the emergence of an entire field of CYP450 research in which it is sought to elucidate how reactants (substrate, protons, oxygen) enter and how products (metabolites, water) exit the active site of CYP450 (Schleinkofer et al., 2005; Wade et al., 2004); a number of putative active-site access channels have been proposed, based on use of various theoretical approaches, as exemplified below.

Standard molecular dynamics (MD) approaches and extensions thereof, such as random expulsion molecular dynamics (REMD) and steered molecular dynamics (SMD) have been used in long-standing efforts to solve the CYP450 active-site access puzzle (Ludemann et al., 1997, 2000a,b; Schleinkofer et al., 2005; Wade et al., 2004, 2005; Winn et al., 2002). Pathways delimited from the MD approaches are supported by results from other independent theoretical approaches such as thermal motion pathway (TMP) analysis (Ludemann et al., 1997) and adiabatic mapping (Ludemann et al., 2000b).

Channels in several CYP450 isoforms accessible to spherical probes the size of water molecules have been found using MD approaches and have been named as pathway (pw) pw1, pw2 and pw3 (Ludemann et al., 1997, 2000a,b; Wade et al., 2004, 2005; Winn et al., 2002). These pathways are delimited by the secondary structure elements lining them at their exit points from the CYP450 structure. The pathway pw2 has been further sub-divided into the sub-types pw2a, pw2b, pw2c, pw2d and pw2e on the basis of REMD ligand egress trajectories (Schleinkofer et al., 2005). A ligand following pw1 from the active site exits close to the heme plane, passes under the I helix and egresses between the C/C′ and H or L helices – close to the G–H loop and β2 strand (Wade et al., 2004). The subclasses of pw2 share proximity with the B–C loop/B′ helix (Schleinkofer et al., 2005). Pw2a exits the CYP450 structure between the F–G loop, B–C loop/B′ helix/BB′ loop and the β1 strand (Wade et al., 2004). Pw2b exits between the B/B′ loop and the β1 and β3 strands whilst pw2c exits between the G and I helices and the B′ helix/B–C loop. Pw2d exits between the N-terminus and helices a and A whilst pw2e exits through the B–C loop. Pw3 exits between the F and G helices or at the EF loop (Wade et al., 2004). In addition, a number of CYP450 crystal structures have a small channel from the active site to the protein surface that has been referred to as the "solvent channel", abbreviated as channel S, which exits between the F, E and I helices and the β4 strand (Haines et al., 2001; Wade et al., 2004). Channel S is thought to be important in conveying and controlling proton access to the CYP450 active site (Wade et al., 2004). The relative positions of these tunnels are illustrated in Fig. S2. A water channel, named channel W, has also been identified which exits the CYP450 structure at the base of the B–C loop near the C-terminus of the B-helix (Cojocaru et al., 2007).

Cojocaru and co-workers applied the computational geometry tool, CAVER (Petrek et al., 2006), to conduct the largest survey to-date of the distribution and pattern of tunnels in static CYP450 structures (Cojocaru et al., 2007). In that study two new pathway families (pw4 and pw5) and two further pw2 subfamilies (pw2f and pw2ac) were identified (Cojocaru et al., 2007). This work showed that the computational geometry algorithm, CAVER, applied on static structures of CYP450s recapitulates largely the same spectrum of access routes as that from molecular dynamics simulations. Hence it has been possible to assign tunnels calculated using CAVER to the tunnel systems calculated using REMD approaches. It is known that CAVER can give a number of possible tunnels emanating from the same user-defined starting point but terminating at different CAVER-calculated exit points on the protein surface. This occurs owing to the fact that several choices for tunnel-exit points may exist in the calculations of the CAVER algorithm. The CAVER algorithm divides three-dimensional space into a grid and calculations are based on grid-points. An envelope (convex hull) is calculated around the protein molecule and grid points that fall on this envelope are assigned as possible exit-points for tunnels. Hence, when several such points fall in the same space as a REMD tunnel, several "variants" of that REMD tunnel will appear in CAVER calculations (Cojocaru et al., 2007; Petrek et al., 2006). In principle, only one "variant" need be considered and the greatest priority is given to the widest tunnel that bears the greatest verisimilitude to the associated REMD tunnel. In cases where it is not straightforward to assign the CAVER-calculated tunnel that bears the greatest verisimilitude to REMD tunnels all the candidates may be assigned as "variants" of that REMD tunnel.

Gating residues have been identified for some of the pathways in CYP3A4 and CYP2C5 using MD approaches. Fishelovitch and co-workers identified the following CYP3A4 gating residue pairs in SMD simulations of both Temazepam (TMZ) and Testosterone-6OH (TST-OH): pw2a, F57-F215; pw2b, F108-F220; pw2c, F108-F241; pw3, F213-F241 and S, R212-L482. However, SMD simulations in the same study identified different pw2e-gating residues for TMZ and TST-OH: F108-I120 and R105-S119, respectively (Fishelovitch et al., 2009). In CYP2C5 REMD simulations that used progesterone and 21-hydroxy-progesterone as substrate, Schleinkofer and co-workers identified K241-V106 as a pw2c gating residue pair (Schleinkofer et al., 2005). Interestingly, the SMD simulations on CYP3A4 suggest that certain residues play a role in gating more than one tunnel [e.g. F108 (pw2b, pw2c & pw2e) and F241 (pw2c and pw3)] and interact with different residues in doing so (e.g. F108-F220 vs F108-F241 vs F108-I120). It is also interesting that some of the SMD-predicted gating residues apparently vary in a substrate-dependent manner.

Taken together, these observations suggest that the picture of tunnel-gating residues in cytochrome P450 enzymes is far from complete and that further analysis using different approaches is warranted. We have therefore built a strategy for the identification of tunnel-gating residues based on the assumption that amino-acid side chain motions are the main mechanism by which CYP450 molecules open or close ligand access tunnels. Usefully, a constant characteristic of the SMD-predicted gating residues is that the gating interactions that close the tunnel of interest are present in the start configuration of the MD simulation and always appear to persist throughout the course of the MD simulation, their eventual breakage accompanying the escape of the ligand from the protein (Fishelovitch et al., 2009; Schleinkofer et al., 2005). Hence it is plausible that additional tunnel-gating residues which can be identified from static structures by other approaches may also be expected to make interactions that show the same trends.

Our strategy begins with the use of the computational geometry tool, CAVER, to perform a focused and detailed analysis of the distribution of tunnels in static CYP3A4, CYP2C5, CYP2C8, CYP2C9 and CYP2A6 structures with the view of building "tunnel-open"/"tunnel-closed" CYP450 structure pairs for as many of the known (putative) pathways as possible. These "tunnel-open"/"tunnel-closed" pairs are subsequently used to identify amino-acids whose side chains obstruct a given REMD tunnel of interest by searching for side chains in the "tunnel-closed" structure that intersect with the REMD tunnel of interest calculated in the "tunnel-open" structure. The "tunnel-open" and "tunnel-closed" structures are aligned to facilitate delimitation of the amino-acids whose side chains intersect the tunnel of interest. Gating interactions are subsequently delimited by tracking-down intra-protein interactions made by the tunnel-intersecting residues. These interactions form part of the two-state models described in this work when delimitations are performed using structures of the same CYP450 isoform. In our strict regime for delimitation of two-state gating models, we stipulate that tunnel-closure interactions occur exclusively in the "tunnel-closed" structure and tunnel-opening interactions occur exclusively in the "tunnel-open" structure. When

"tunnel-open"/"tunnel-closed" pairs are comprised of structures of different CYP450 isoforms, only the tunnel-intersecting residue may be delimited but not interactions that occur exclusively in either the "tunnel-open" or "tunnel-closed" structure. The gating models we derive in this case are termed singlet gating models to denote that, unlike in the primary description of a two-state gating model where two interacting amino acids and the tunnel they obstruct are specified, here only one amino acid and the tunnel it intersects are specified.

Since our goal is to define as many "tunnel-open"/"tunnel-closed" pairs as possible that are based on CAVER-calculated tunnels which bear the greatest verisimilitude to REMD tunnels, it is necessary to optimise the trade-off between the cut-off gorge widths for the CAVER-calculated tunnels that are to be included in the study versus the verisimilitude of the included tunnels to REMD tunnels. We note that in the optimisation process, for any given gorge width cut-off, the assignment of CAVER-calculated tunnels to REMD tunnels (i.e. tunnel verisimilitude – done visually in this work) would be more subjective than the assignment of the number of "tunnel-open"/"tunnel-closed" pairs. We also note as a reminder that CAVER should not be expected, within the frameworks of the CYP450 ligand-tunnelling problem, to calculate tunnels of large gorge widths – typically taken as $\geqslant$2.8 Å i.e. large enough for at least water molecules to pass through. Hence verisimilitude to REMD tunnels is an important criterion and guide within the context of the CYP450 ligand-tunnelling problem. In addition to the assumption described above that amino-acid side chains play a paramount role in the gating of tunnels, we also assume that the CAVER-calculated tunnels that we subsequently assign to REMD tunnels are sufficiently related to REMD tunnels enough to allow the accurate extraction of further information about REMD tunnels – such as gating models.

Two-state gating models (where they can be built) offer the most parsimonious rationalisation for differential tunnel-opening patterns in static structures. We have explored the relationship between gating models identified in this work and gating models identified in MD approaches described in the literature. Furthermore, the large survey of gating residues presented herein has permitted analysis of the side chain chemical types, rate of evolution and divergence and conservation of the tunnel-intersecting gating residues and the gating interaction networks they make in CYP450.

2. Materials and methods

2.1. Structures and sequences

Structures were retrieved from the RCSB Protein Data Bank (www.pdb.org). The structure PDB codes are as follows. CYP3A4 (*Homo sapiens*): 1TQN, 1W0F, 2J0D and 2V0M. CYP2C5 (*Oryctolagus cuniculus*): 1DT6, 1N6B and 1NR6. CYP2C8 (*Homo sapiens*): 1PQ2, 2NNH, 2NNI, 2NNJ and 2VN0. CYP2C9 (*Homo sapiens*): 1OG2, 1OG5 and 1R90. CYP2A6 (*Homo sapiens*): 1Z10, 1Z11, 2FDU, 2FDV, 2FDY, 2FDW and 2PG5. Sequences were retrieved from the Universal Protein Resource Knowledgebase (UniProtKB/Swiss-Prot) database via the European Bioinformatics Institute website (http://www.ebi.ac.uk). The accession numbers were according to the UniProtKB database unless otherwise stated. The sequence accession numbers used herein are as follows. CYP2C5: P00179 (*Oryctolagus cuniculus*), CYP2C8: P10632 (*Homo sapiens*), CYP2C9: P11712 (*Homo sapiens*), CYP2A6: P11509 (*Homo sapiens*) and CYP3A4: P08684 (*Homo sapiens*).

2.2. Calculation of tunnels using CAVER

Tunnels were searched in each of the CYP450 structures using the program CAVER (Petrek et al., 2006). CAVER is a computational geometry tool that uses the Dijkstra algorithm to calculate the shortest low-cost accessible path from a user-defined starting point located within the interior of the protein to the bulk solvent. Several accessible routes may be returned and in such situations CAVER returns a ranked/prioritised list of possible accessible routes. CAVER gives greatest priority to the access route with the widest gorge width irrespective of the length of the route. Furthermore, when two access channels have the same gorge width, the shorter channel has greater priority (personal communication, Petr Benes, CAVER support) (Petrek et al., 2006). We used the heme Iron (heme Fe) in CYP450 structures as the starting-point in tunnel searches. The heme Fe co-ordinates were read directly from PDB files. Water molecules were ignored in the searches.

For a given CYP450 structure, the determination of the variation of the number of "tunnel-open"/"tunnel-closed" pairs with the gorge width cut-off began with issuing a request to CAVER for a large number of tunnels – typically 60. The number of "tunnel-open"/"tunnel-closed" pairs was recorded as the gorge width of accepted tunnels (accepted as being significant and therefore included in the study) was increased from 0.7 Å to 3.3 Å (inclusive) in steps of 0.1 Å. Qualitative notes about the overall similarity of accepted tunnels to REMD tunnels were also made to assist in the assignment of the verisimilitude of the CAVER-calculated tunnels to REMD tunnels at given gorge width cut-offs. At cut-offs $\leqslant$ 1 Å, the majority of accepted tunnels were not akin to the previously-defined REMD tunnels. The maximum number of "tunnel-open"/"tunnel-closed" pairs, calculated at each point as the gorge-width cut-off is incremented in steps of 0.1 Å from 0.7 to 3.3 Å (inclusive) in each CYP450 structure, was found to be 4. There was no differential tunnel-opening in all CYP450 structures outside the gorge width cut-off range [0.7–3.3 Å]. We also found that a gorge width cut-off of 1.4 Å allowed the best balance of factors (verisimilitude versus number of differential pairs) in the majority of the CYP450 structures that were analysed and was subsequently applied in all tunnel calculations in this study.

2.3. Structure alignments

Structure visualisations and alignments were done using the program Pymol (www.pymol.org). Structure alignments of CYP450 catalytic domains were evaluated by applying a panel of visual criteria such as the simultaneous achievement of stacking of heme co-factors in parallel planes with less than 3 Å distance separation between corresponding atoms and complete superimposition of I-helices. General good alignment of F and G helices was also considered in accepting or rejecting alignments of structures. The Root Mean Square Deviation (RMSD) of the alignments was also considered. Protein sequences were aligned using the program ClustalW (default parameters) (Jeanmougin et al., 1998; Thompson et al., 1994, 1997).

2.4. Delimitation of two-state gating models

Two-state gating models here are delimited from different structures of the same CYP450 isoform. The search for two-state gating models begins with the identification of tunnel-open/tunnel-closed pairs of structures for a given tunnel and a given CYP450 isoform. These pairs are identified from profiles calculated using CAVER (Petrek et al., 2006) as described above. Tunnel-open/tunnel-closed pairs are aligned using Pymol (www.pymol.org) and only the tunnel of interest is displayed. Amino-acid side-chains from the tunnel-closed structure that intersect the tunnel in the tunnel-open structure are manually identified by searching along the tunnel from the tunnel's start-point at the heme Iron (Fe) to the tunnel's end-point at the protein surface. This manual search usually identifies only one residue whose side-chain intersects

with the tunnel of interest; residues that directly interact with the tunnel-intersecting residue are subsequently identified using the Protein Interaction Calculator (PIC) server (default parameters) (Tina et al., 2007). This delimitation of residues interacting with the tunnel-intersecting residue is done independently for the "tunnel-open" and the "tunnel-closed" structures.

The interactions delimited in the tunnel-closed structure using PIC have to satisfy the requirement that they are absent in the tunnel-open structure in order to constitute "tunnel-closure" interactions in a two-state gating model. Similarly, interactions delimited in the tunnel-open structure using PIC have to satisfy the requirement that they are absent in the tunnel-closed structure in order to constitute "tunnel-opening" interactions in a two-state gating model. Hence interactions common to both "tunnel-open" and "tunnel-closed" structures are omitted from our strict two-state gating models. We note that such two-state gating models offer a detailed maximum parsimony mechanistic explanation for the differential channel-opening patterns in structures of a given CYP450 isoform.

The description of a given two-state gating model includes a 3-tuple, (Aa1, Aa2, pw), where Aa1 is the manually-delimited amino acid whose side-chain intersects the REMD tunnel pw and Aa2 is an amino acid delimited to be involved in a non-covalent interaction (e.g. hydrophobic, ionic, main chain–main chain H-bond etc.) with the manually-delimited amino acid Aa1. These 3-tuples are reported directly in Table 2 and Table S2.

2.5. Delimitation of singlet gating models from structures of different CYP450 isoforms

In close analogy to the delimitation of two-state gating-models, the delimitation of singlet gating models in this approach begins with the alignment of structures of *different* CYP450 isoforms with differential tunnel-opening patterns for a given tunnel of interest. Tunnel-open/tunnel-closed pairs of different CYP450 isoforms are aligned using Pymol (www.pymol.org) and only the tunnel of interest is displayed. Amino-acid side-chains from the tunnel-closed structure that intersect the tunnel in the tunnel-open structure are manually identified by searching along the tunnel from the tunnel's start-point at the heme Iron (Fe) to the tunnel's end-point at the protein surface. This manual search usually identifies only one residue whose side-chain intersects with the tunnel of interest.

The description of a gating-residue model includes a 2-tuple (Aa1, pw) where Aa1 is the manually-delimited amino acid whose side-chain intersects the REMD tunnel pw. We do not directly report 3-tuple models here as there is no simple way of delimiting interactions that occur exclusively in tunnel-open or tunnel-closed structures. However, it is possible to assemble possible 3-tuple gating-residue models from these results by appending to the 2-tuple models (Aa1, pw), residues delimited to be in direct interaction with Aa1. These residues are described in the comments column of Table 3 and Table S3.

2.6. Calculation of intra-protein interactions

For a given tunnel-intersecting residue identified manually as described above, the intra-protein interaction partners in main chain-main chain, main chain-side chain, side chain-side chain, hydrophobic aliphatic, ionic and hydrophobic aromatic interactions are identified using Protein Interaction Calculator Server (PIC) (Tina et al., 2007). Default parameters are used unless otherwise stated.

Table 1
CAVER-calculated tunnels in CYP3A4 structures.

CYP450: PDB code	CAVER tunnel number	REMD equivalent and tunnel gorge width							
		pw2a	pw2b	pw2c	pw2e	pw3	S	W	Width/Å
CYP3A4: 1TQN	1		×						2.3
	2				×				1.9
	3[a]						×		1.4
	4[b]						×		1.4
CYP3A4: 1W0F	1		×						2.2
	2				×				1.9
CYP3A4: 2J0D	1[c1]	×							3.3
	2[c2]								3.2
	3		×						2.4
	4[d]	×							2.2
	5[e]								2.1
	6[f]						×		1.7
	7				×				1.7
CYP3A4: 2V0M	1		×						2.4
	2	×							1.6
	3				×				1.5
	4					×			1.5
	5	×							1.5

Tunnels calculated in a structure of a given CYP450 isoform are numbered sequentially in Arabic numerals. The cross sign (×) is placed under the name of the REMD pathway to which the numbered CAVER-calculated tunnel is assigned. The gorge width of the numbered tunnel is also reported in the same row. For example, four significant tunnels are calculated in the CYP3A4 structure 1TQN and the first tunnel (No. 1) is assigned to the REMD pathway pw2b. This tunnel has a gorge width of 2.330 Å. The REMD pathways pw1 and pw2d were not observed.

[a] This is a variant of the solvent channel (S) that is positioned forward more towards the F–G loop.

[b] This is a variant of the solvent channel (S) that is positioned backwards – farther removed from the F–G loop.

[c1] This tunnel is probably an artefact of the gap created by the missing residues (214–218 inclusive) in 2J0D. It lies in close proximity to the space of the canonical pw2a and can therefore be assigned as pw2a. However, others have assigned it as pw2f (Cojocaru et al., 2007).

[c2] This tunnel is probably an artefact of the gap created by the missing residues (214 – 218 inclusive) in 2J0D. This tunnel lies midway between the space of the canonical tunnel S and pw2a. Unlike the canonical S it does not bend over the I-helix nor does it encroach towards the β4 strand. The assignment of this tunnel is therefore not trivial and is not attempted here.

[d] This is probably the true pw2a tunnel in 2J0D as it is closest to the β1 strand and is not adjacent to the missing part of the F–G loop.

[e] This tunnel is probably an artefact of the gap created by the missing residues (214 – 218 inclusive) in 2J0D. This tunnel lies midway between the space of the canonical tunnel S and pw2a. Unlike the canonical S it does not bend over the I-helix nor does it encroach towards the β4 strand. The assignment of this tunnel is therefore not trivial and is not attempted here.

[f] This variant of the solvent channel (S) in 2J0D is more proximal to the F helix than the variants in 1TQN.

Table 2
Two-state gating models in CYP3A4.

Model number	Structures: open/closed	Tunnel-closing interaction	Gated tunnel
1	1TQN/2V0M (CYP3A4)	**Arg 212**-Glu 308	S
Comment: see comments in model number 2 below			
2	1TQN/2V0M (CYP3A4)	**Arg 212**-Asp 214	S
Comment: the Arg 212-Glu 308 ionic interaction, the Arg 212 O-Asp 214 OD2 main chain-side chain interaction and the Arg 212 O-Asp 214N main chain–main chain interaction that close tunnel S in 2V0M are absent in 1TQN as determined by Protein Interaction Calculator. Instead, the interactions are replaced by: (a) Arg 212NH1-Phe 304 O (carbonyl oxygen) interaction. This main chain (Phe 304 O)-side chain (Arg 212 NH1) interaction stretches the Arg 212 side-chain downwards parallel to channel S. This interaction is accompanied by greater unwinding of the F helix in 1TQN. (b) Side interaction between the N nitrogen of Arg 212 and the OE2 oxygen of Glu 308. This main chain (Arg 212 N)-side chain (Glu 308 OE2) interaction co-operates with the interaction in (a) to position the Arg 212 side-chain parallel to channel S			
3	1TQN/2V0M (CYP3A4)	Leu 482-Ile 369	S
Comments: see comments in model number 4 below			
4	1TQN/2V0M (CYP3A4)	**Leu 482**-Phe 304	S
Comments: the Leu 482-Ile 369 and Leu 482-Phe 304 hydrophobic aliphatic interactions that closes tunnel S in 2V0M are absent in 1TQN as determined by Protein Interaction Calculator. Instead, Leu 482 makes hydrophobic contact with Met 371 in 1TQN. The Leu 482 O-Gln 484N main chain–main chain interaction is present in both 1TQN and 2V0M and is not listed as an interaction specific for either tunnel closure or opening			
5	2V0M/1W0F (CYP3A4)	**Phe 213**-Phe 108	pw3
Comment: see comments in model number 9 below			
6	2V0M/1W0F (CYP3A4)	**Phe 213**-Phe 215	pw3
Comments: see comments in model number 9 below			
7	2V0M/1W0F (CYP3A4)	**Phe 213**-Phe 241	pw3
Comments: see comments in model number 9 below			
8	2V0M/1W0F (CYP3A4)	**Phe 213**-Phe 304	pw3
Comments: see comments in model number 9 below			
9	2V0M/1W0F (CYP3A4)	**Phe 213**-Val 240	pw3
Comments: the Phe 213-Phe 108, Phe 213-Phe 215, Phe 213-Phe 241 and Phe 213-Phe 304 hydrophobic aromatic interactions and the Phe 213-Val 240 hydrophobic interaction that close tunnel pw3 in 1W0F are absent in 2V0M as determined by Protein Interaction Calculator. Instead, they are replaced in 2V0M by the Phe 213-Leu 211 hydrophobic interaction and the Phe 213N-Leu 211 O main chain–main chain interaction. We note that the Phe 213-Phe 220 interaction is present in both 2V0M and 1W0F			
10	2V0M/1TQN (CYP3A4)	**Phe 215**-Phe 108	pw2a
Comments: see comments in model number 12 below			
11	2V0M/1TQN (CYP3A4)	**Phe 215**-Phe 213	pw2a
Comments: see comments in model number 12 below			
12	2V0M/1TQN (CYP3A4)	**Phe 215**-Phe 220	pw2a
Comments: the Phe 215-Phe 108, Phe 215-Phe 213 and Phe 215-Phe 220 hydrophobic aromatic interactions that close tunnel pw2a in 1TQN are absent in 2V0M as determined by Protein Interaction Calculator. Instead, Phe 215 makes hydrophobic interactions with Leu 216 and with Leu 479 in 2V0M			

The two-state gating models we delimited in CYP3A4 structures are numbered in Arabic numerals. The tunnel-open/tunnel-closed pairs of CYP3A4 structures were constructed using the profiles in Table 1 above. Each of the numbered models corresponds to a manually-delimited residue whose side-chain intersects the tunnel of interest (listed in the fourth column from the right) and the intra-protein interactions it makes. The manually-delimited tunnel-intersecting residue in each model is shown underlined and in boldface.

2.7. Comparison of gating models delimited in this work with those delimited in the literature by molecular dynamics approaches

Gating models delimited by MD (for a given tunnel) were collated from the literature and tabulated. The tabulated literature gating models were analysed to establish the extent of agreement and overlap with gating models identified in this work using various simple metrics/measures. Gating model information is given in the literature in the form of 3-tuples (Aa1, Aa2, pw). Several metrics were devised to compare 3-tuples identified in this work with those reported in the literature. The first metric, Q1, expresses the percentage of all literature gating 3-tuples (identified by molecular dynamics approaches) that have also been identified in this work.

In the calculation of the second metric, Q2, the set of literature 3-tuples is trimmed to include only those 3-tuples that include pathways for which it has been possible to delimit gating residues for a given CYP450 isoform using our approach described herein. Q2 is calculated in the same manner as Q1 over this trimmed/limited set of literature 3-tuples. We also calculate P1, the percentage of 3-tuples models identified in this work that have been previously described in the literature. For purposes of calculation of P1, the 2-tuple models delimited in this work are converted to 3-tuples by assuming that only one 3-tuple model arises from a given 2-tuple. This "pessimistic" model effectively gives the smallest possible number of 3-tuple gating-residue models since all the interactions involving the tunnel-intersecting residue Aa1 in the 2-tuple models may be realised as exclusively tunnel-closure interactions (these interactions are listed in the comments section of Table 3 and Table S3).

2.8. Analysis of gating models

Tunnel-intersecting residues in the gating models delimited as described above were classified according to side chain chemical type as hydrophobic aromatic (Phe, Trp, Tyr), Polar/Charged (Arg, Lys, Glu, Asp, Thr, Gln, Asn, Ser, His, Cys) or hydrophobic aliphatic (Leu, Ile, Val, Met, Pro). Glycine (no side-chain) and Alanine (amphipathic) were not considered as residues that fall in any of these three classes and therefore would not have contributed to counts of residues had they been delimited as tunnel-intersecting residues. The abovementioned assignment of amino-acid residues as hydrophobic aromatic, polar/charged or hydrophobic aliphatic was along the lines discussed by Valdar for conservation analysis (Valdar, 2002). The distribution of side chain chemical-types occurring as tunnel-intersecting residues in each of CYP2C5, CYP2C8, CYP2C9 and CYP2A6 was determined by calculating the percentage of tunnel-intersecting residues that are hydrophobic aromatic, polar/charged or hydrophobic aliphatic.

Table 3
Singlet gating models in CYP3A4.

Model number	Structures: open/closed	Gating residue	Gated tunnel	Maximum number of 3-tuple models
1	1DT6 (CYP2C5)/1TQN (CYP3A4)	CYP3A4 Phe113	pw2c	(8)
Comments: Phe 113 makes hydrophobic contacts with the following residues in 1TQN: Met 114, Leu 293, Val 296, Ala 297 and Ile 300 Phe 113 also makes the following main chain–main chain interactions in 1TQN: Phe 113N-Val 111 O, Phe 113 O-Lys 115N and Phe 113 O-Ser 116N Finally Phe 113 makes a Phe 113 O-Ser 116 OG main chain-side chain interaction.				
2	1DT6 (CYP2C5)/1TQN (CYP3A4)	CYP3A4 Phe 241	pw2c	(11)
Comments: Phe 241 makes hydrophobic aromatic interactions with the following residues in 1TQN: Phe 108, Phe 213 and Phe 304 Phe 241 makes the following hydrophobic interactions in 1TQN: Val 111, Met 114, Ile 120, Pro 242, Ile 300 and Ile 301. Phe 241 also makes Phe 241N-Cys 239 O and Phe 241 O-Arg 243N main chain–main chain interactions in 1TQN Furthermore, Phe 241 makes Phe 241N-Cys 239 SG and Phe 241 O-Cys 239 SG main chain-side chain interactions in 1TQN				
3	2FDV (CYP2A6)/2V0M(CYP3A4)	CYP3A4 Trp 126	W	(10)
Comments: Trp 126 makes aromatic hydrophobic contact with Tyr 99. Trp 126 makes hydrophobic contact with Ile 118, Ala 121 and Pro 439. Trp 126N makes main chain–main chain interactions with Glu 122 O, Asp 123 O and Glu 124 O. Trp 126 O makes main chain–main chain interactions with Arg 128N, Leu 129N and Arg 130N				
4	2FDV (CYP2A6)/2V0M(CYP3A4)	Lys 127	W	(4)
Comments: Lys 127N makes main chain–main chain interactions with Asp 123 O and Glu 124 O Lys 127 O makes main chain–main chain interactions Arg 130N and Ser 131N Lys 127 O makes main chain–side chain interactions with Ser 131 OG				

The singlet gating models we delimited in CYP3A4 structures are numbered in Arabic numerals. The tunnel-open/tunnel-closed pairs of CYP structures were constructed using the profiles in Table 1 above. Each of the numbered models corresponds to a manually-delimited residue whose side-chain intersects the tunnel of interest (listed in the fourth column) and the intra-protein interactions it makes. Unlike in the two-state models defined in Table 2 and Table S2, interactions in the "tunnel-open" mode cannot be defined here. Furthermore, unlike in the two-state models, the interactions of tunnel-intersecting residues occurring exclusively in the tunnel-open structure or tunnel-closed structure cannot be labelled here. The manually-delimited tunnel-intersecting residue in each model is shown underlined and in boldface. The singlet gating models are given here as the 2-tuples (Aa1, pw) where Aa1 is the manually-delimited tunnel-intersecting residue that intersects tunnel pw. The number in parentheses ( ) in the column "Maximum number of 3-tuple models" gives the maximum number of 3-tuple gating-reside models (Aa1, Aa2, pw) that can be generated by pairing the residue Aa2 delimited to interact with the tunnel-intersecting residue Aa1. Residues of type Aa2 are listed in the "comments" section of Table 3 for a given residue Aa1.

We also analysed the rate of evolution of the tunnel-intersecting residues in CYP2C5, CYP2C8, CYP2C9 and CYP2A6. The sequences of CYP2C5, CYP2C8, CYP2C9 and CYP2A6 were aligned using the software ClustaW (default parameters) (Jeanmougin et al., 1998; Thompson et al., 1994, 1997) and the alignment was used to calculate a tree using the program PROML in the PHYLIP suite of programs (version 3.69, default parameters, http://evolution.genetics.washington.edu/phylip.html). The PROML-calculated tree and the ClustalW-calculated alignment were used as input in the program Rate4Site (Pupko et al., 2002) (version 2.01 for windows, default parameters) to calculate the relative rates of evolution of all the sequence positions in the alignment of CYP2C5, CYP2C8, CYP2C9 and CYP2A6. The average rate of evolution of the tunnel-intersecting residues in each of the isoforms CYP2C5, CYP2C8, CYP2C9 and CYP2A6 was calculated as the arithmetic average of the rates of evolution of the corresponding positions in the ClustalW-calculated alignment. The statistical significance of the difference between the average rate of evolution of the entire sequence and the average rate of evolution of the tunnel-intersecting residue was calculated for each of CYP2C5, CYP2C8, CYP2C9 and CYP2A6 using the Student's *t*-test at the 5% level of statistical significance. Here the null hypothesis that the average rate of evolution of the entire sequence is the same as the average rate of evolution of the tunnel-intersecting residues was tested against the alternative hypothesis that the average rate of evolution of the entire sequence is less than the average rate of evolution of the tunnel-intersecting residues in each of CYP2C5, CYP2C8, CYP2C9 and CYP2A6. Student's *t*-test calculations were performed using functions in Microsoft Excel spreadsheets.

Finally, we annotated an alignment of PDB-extracted sequences of CYP2C5, CYP2C8, CYP2C9 and CYP2A6 to show the relationship between conserved positions in the CYP2 family, CYP secondary structure elements, Gotoh's substrate recognition sites (Gotoh, 1992), the residues delimited in the two-state gating models of this work and the tunnel-intersecting residues in the gating models delimited herein.

3. Results

3.1. Calculation of tunnels using CAVER

The distribution of tunnels calculated in CYP3A4 structures using CAVER is shown in Table 1. The distribution of tunnels in CYP2C5, CYP2C8, CYP2C9 and CYP2A6 structures is shown in Table S1.

3.2. Delimitation of two-state gating models

A list of the two-state gating models that we delimited in CYP3A4 is given in Table 2. Two-state models delimited in CYP2C5, CYP2C8 and CYP2C9 are given in Table S2. Furthermore, an example of the manual delimitation of a tunnel-intersecting residue is shown in Fig. 1 below for the case of the CYP3A4 tunnel-open/tunnel closed pair 1TQN/2V0M (see Table 1 above). This example considers the opening and closure of tunnel S. Analysis of this pair of structures, after calculation of tunnels in 1TQN and subsequent alignment to 2V0M, identifies Arginine 212 as the 2V0M residue whose side-chain intersects 1TQN tunnel S.

3.3. A note on the gating of pw2b

The tunnel pw2b is open in all the four structures of CYP3A4 considered in this study (the structures 1 W0E and 1 W0G give the same tunnel pattern as 1 W0F). As such it is not possible to define tunnel-open/tunnel-close pairs of structures for this tunnel. However, we have identified a conserved network of interactions present in all four CYP3A4 structures that restricts the space available to the pw2b tunnel as it extends from the active-site to the protein surface. We speculate that this space restriction then allows the swinging motions of side chains of gating residues (candidates are residues such as Phe 220 and Phe 219) to efficiently shut-down or open pw2b. In the absence of this network of space-restricting residues, several combinations of gating residues or large protein

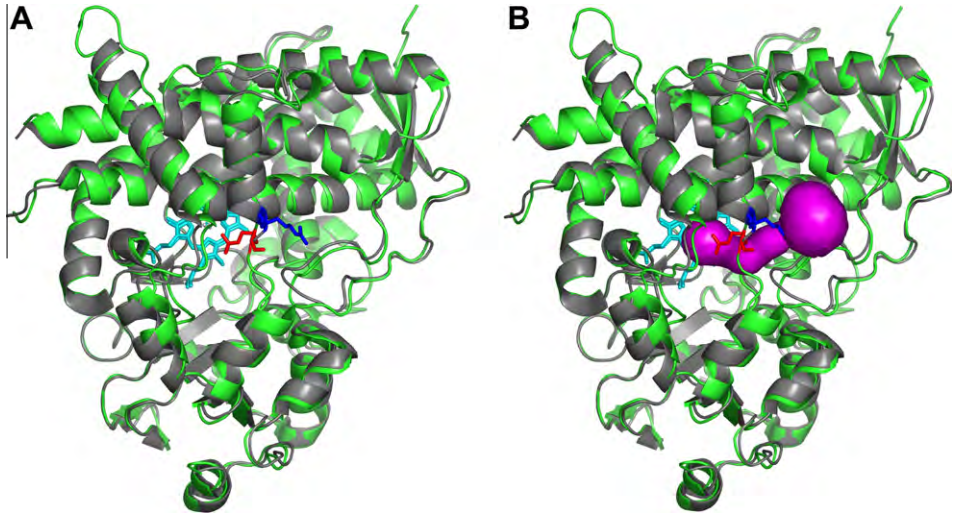


Fig. 1. The manual delimitation of the tunnel-intersecting residue in a two-state gating model for tunnel S in CYP3A4. Panel (A) shows the alignment of the CYP3A4 structures 1TQN and 2V0M without any CAVER-calculated tunnels shown. Panel (B) shows the same alignment of 1TQN and 2V0M in the same orientation but with the CAVER-calculated tunnel S shown in purple. In both panels (A) and (B), 1TQN is shown in green, 2V0M in grey, the heme group in turquoise, Arg 212 in 2V0M is shown in blue and Arg 212 in 1TQN is shown in red. Panel (B) shows that Arg 212 in 2V0M intersects the CAVER-calculated tunnel S in 1TQN. The two structures were aligned using Pymol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

movements would be required to shut-down all the spaces on the protein surface the tunnel pw2b can exit-from. This network of interactions comprises the following pairs of ionic interactions: Arg 106-Asp 76, Arg 106-Glu 374, Arg 372-Asp 76 and Arg 372-Glu 374.

3.4. Delimitation of singlet gating models from structures of different CYP450 isoforms

A list of the singlet gating models that we delimited is given in Table 3 and in Table S3. Furthermore, an example of the manual delimitation of the gating residue whose side-chain intersects a tunnel is shown in Fig. 2 below for the case of 1DT6.pdb (CYP2C5) and 1TQN.pdb (CYP3A4) which comprise a pw2c tunnel-open/tunnel-closed pair of structures. The pathway pw2c is open in 1DT6 and closed in 1TQN. Analysis of this pair of structures, after calculation of tunnels in 1DT6 and subsequent alignment to 1TQN, identifies Phe 113 and Phe 241 as 1TQN.pdb residues whose side chains intersect tunnel pw2c in 1DT6.pdb.

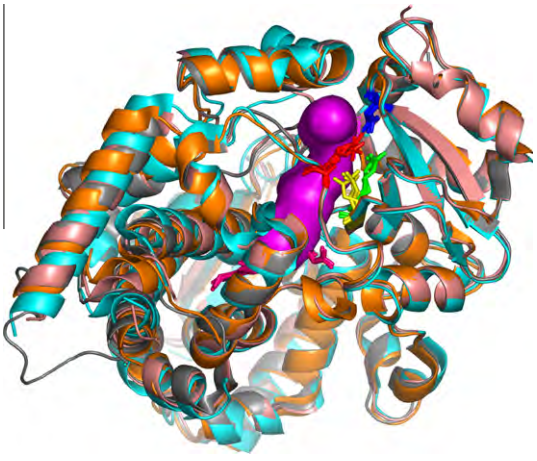


Fig. 2. Conserved network of interactions that restricts the space available for tunnel pw2b exit on the CYP450 protein surface. Tunnel pw2b calculated in the CYP3A4 structure 1TQN using CAVER is shown in purple. The amino-acid residues that constitute the conserved networks of interacting residues that are thought to restrict the space available for tunnel pw2b are: Arg 106 (in red), Asp 76 (in blue), Arg 372 (in green) and Glu 374 (in yellow). Arg 106 makes ionic contact with Asp 76 and with Glu 374. Arg 372 makes ionic contact with Asp 76 and with Glu 374. The aligned CYP3A4 structures are as follows: 1TQN (in grey), 1W0F (in salmon), 2J0D (in orange) and 2V0M (in cyan). The heme group is rendered in hot-pink. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.5. Comparison of gating models delimited in this work with those delimited in the literature by molecular dynamics approaches

Gating models delimited by molecular dynamics approaches in the literature are shown in Table 4 below. The percentage of all literature gating model 3-tuples that have also been identified in this work (defined as Q1 in Section 2) is calculated by comparing Tables 2 and 3, S2 and S3 with Table 4 as: 4/9 = 44%. As noted above, gating models for pw2b and pw2e cannot be deduced for the CYP3A4 isoform using our approach because pw2b and pw2e are open in all the CYP3A4 structures. After excluding pw2b and pw2e gating models from the set of literature gating models in Table 4, six gating models remain (model 3, 5 and 6 are omitted in Table 4). Therefore a more objective measure of the percentage of literature gating model 3-tuples (limited set) that have also been identified in this work (defined as Q2 in Section 2) is re-calculated as: 4/6 = 67%. We also note that 11/12 = 92% of residues that appear in the limited set of literature molecular dynamics gating models (i.e. the set that excludes gating residues for CYP3A4 tunnel pw2b and pw2e) have also appeared in the models delimited in this work – matched here to the same tunnel as in the literature (Fig. 3).

A total of 65 gating models have been identified in this work (28 two-state gating models plus 37 singlet gating models) across CYP3A4, CYP2C5, CYP2C8, CYP2C9 and CYP2A6; 4 of these are in common with those identified by molecular dynamics in the literature (across CYP3A4 (Fishelovitch et al., 2009) and CYP2C5 (Schleinkofer et al., 2005)), hence the percentage of gating models identified in this work which are also described in the MD literature (defined as P1 in Section 2) is calculated as 4/65 = 6%. We have identified a total of 27 gating models for CYP3A4 and CYP2C5 (16 two-state models plus 11 singlet gating models). Hence P1 is calculated across the same spread of CYP isoforms as 4/27 = 15%.

Table 4
Gating models identified from molecular dynamics studies in the literature.

Model number	Gating residues	Tunnel	Comments	Reference
	CYP2C5			
1	K241-V106	pw2c	Breakage of the Val 106 O-K 241 NZ main chain-side chain H-bond was found to be a pre-requisite for pw2c channel opening in REMD and MD simulations	Schleinkofer et al. (2005)
	CYP3A4			
2	F57-**F215**	**pw2a**	This gating model was identified from SMD simulations using temazepam and testosterone-6OH as products	Fishelovitch et al. (2009)
	CYP3A4			
3	F108-F220	pw2b	This gating model was identified from SMD simulations using temazepam and testosterone-6OH as products	Fishelovitch et al. (2009)
	CYP3A4			
4	F108-F241	pw2c	This gating model was identified from SMD simulations using temazepam and testosterone-6OH as products	Fishelovitch et al. (2009)
	CYP3A4			
5	F108-I120	pw2e	This gating model was identified from SMD simulations using temazepam as product	Fishelovitch et al. (2009)
	CYP3A4			
6	R105-S119	pw2e	This gating model was identified from SMD simulations using testosterone-6OH as product	Fishelovitch et al. (2009)
	CYP3A4			
7	F213-F241	pw3	This gating model was identified from SMD simulations using temazepam as product	Fishelovitch et al. (2009)
	CYP3A4			
8	F213-V240	pw3	This gating model was identified from SMD simulations using testosterone-6OH as product	Fishelovitch et al. (2009)
	CYP3A4			
9	R212-L482	**S**	This gating model was identified from SMD simulations using temazepam and testosterone-6OH as products	Fishelovitch et al. (2009)

Amino-acid residues and tunnels that constitute a literature molecular dynamics gating model (specified as a 3-tuple of two amino acids and the REMD tunnel they gate) which has also been identified here are underlined.
Gating model residues from literature molecular dynamics gating models that have also appeared in the gating models delimited here, matched here to the same tunnel as in the literature but not to the same amino acid interaction partner, are shown in bold type.

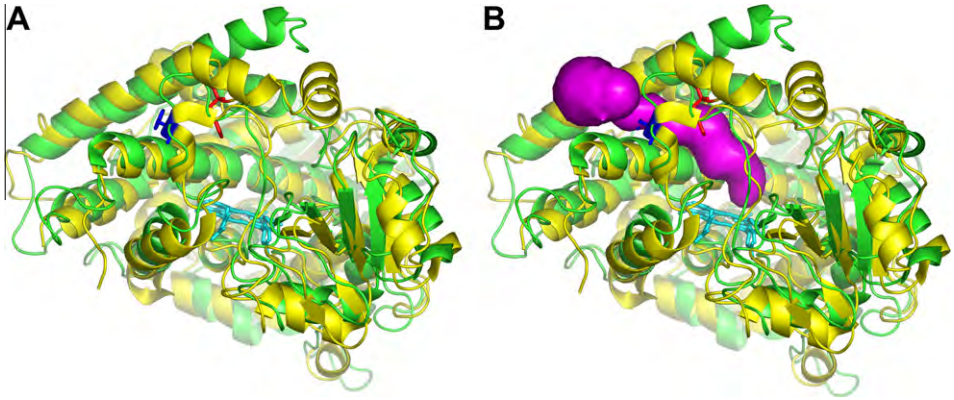


Fig. 3. The manual delimitation of tunnel-intersecting residues in a singlet gating model for tunnel pw2c in CYP3A4. Panel (A) shows the alignment of the CYP3A4 structure 1TQN and the CYP2C5 structure 1DT6 without any CAVER-calculated tunnels shown. Panel (B) shows the same alignment of 1TQN and 1DT6 in the same orientation but with the CAVER-calculated tunnel pw2c shown in purple. The tunnel pw2c is open in the structure 1DT6. In both panels (A) and (B), 1TQN is shown in yellow, 1DT6 in green, the heme group in turquoise, CYP3A4 Phe 113 in 1TQN is shown in blue and CYP3A4 Phe 241 in 1TQN is shown in red. Panel (B) shows that Phe 113 and Phe 241 in 1TQN intersect the CAVER-calculated tunnel pw2c in 1DT6. The two structures were aligned using Pymol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
The distribution of amino-acid side chain-types occurring as tunnel-intersecting residues in gating models delimited herein.

CYP isoform	Hydrophobic aromatic (%)	Polar/charged (%)	Hydrophobic aliphatic (%)
CYP3A4	**11/16 = 68.75**	3/16 = 18.75	2/16 = 12.5
CYP2C5	2/11 = 18.18	**6/11 = 54.55**	3/11 = 27.27
CYP2C8	2/14 = 14.29	6/14 = 42.86	6/14 = 42.86
CYP2C9	5/13 = 38.46	4/13 = 30.77	4/13 = 30.77
CYP2A6	**6/11 = 54.55**	2/11 = 18.18	3/11 = 27.27

For each CYP isoform, the side chain chemical-type with the highest percentage of tunnel-intersecting residues binned into it is underlined (and shown in boldface if it accounts for 50% or more of the tunnel-intersecting residues).

3.6. Analysis of gating models

Tunnel-intersecting residues in the gating models listed in Tables 2 and 3, S2 and S3 were classified according to side chain chemical types as described in Section 2. The distribution of side chain chemical-types occurring as tunnel-intersecting residues is reported in Table 5 below.

The levels of sequence similarity within the CYP2C sub-family and the different usage of amino-acid side chain chemical types in the CYP2 family and within the CYP2C sub-family highlighted in Table 5 suggest that the rate of evolution of the tunnel-gating residue positions is faster than average rate of evolution of the entire sequence for a given CYP450 isoform. We proceeded to

Table 6
The statistical significance of the average rate of evolution of tunnel-intersecting residues in CYP2 isoforms in this study.

CYP450 isoform	Average relative rate of evolution of tunnel-intersecting residues	Student *t*-test *p*-Value
CYP2C8	1.466283	0.0193
CYP2C9	1.613267	0.0032
CYP2C5	1.388589	0.0428
CYP2A6	1.891163	0.0005

The null hypothesis that the average rate of evolution of the entire sequence (1.0055) is the same as the average rate of evolution of the tunnel-intersecting residues was tested against the alternative hypothesis that the average rate of evolution of the tunnel-intersecting residues is greater than the average rate of evolution of the entire sequence (1.0055) in each of CYP2C5, CYP2C8, CYP2C9 and CYP2A6. Student's *t*-test calculations were performed using functions in Microsoft Excel spreadsheets. The null hypothesis is rejected in favour of the alternative hypothesis at the 5% level for each of CYP2C5, CYP2C8, CYP2C9 and CYP2A6.

investigate the rate of evolution of sequence positions in CYP2 sequences.

The average relative rate of evolution of all sequence positions in the alignment of CYP2C5, CYP2C8, CYP2C9 and CYP2A6 was calculated as 1.0055. The average relative rate of evolution of the tunnel-intersecting residues in CYP2C5, CYP2C8, CYP2C9 and CYP2A6 and the associated *p*-values calculated from Student *t*-test are reported in Table 6.

Fig. S1 shows the annotated alignment of PDB file-derived sequences of CYP2C5, CYP2C8, CYP2C9 and CYP2A6 showing the relationship between conserved positions in the CYP2 family, CYP secondary structure elements, Gotoh's substrate recognition sites (Gotoh, 1992), the residues delimited in the two-state gating models of this work and the tunnel-intersecting residues in the singlet gating models delimited herein. The CYP2C5 sequence was extracted from the PDB file 1N6B, CYP2C8 from 2NNH, CYP2C9 from 1R9O and CYP2A6 from 2FDV.

4. Discussion

We have used the computational geometry tool, CAVER (Petrek et al., 2006), to calculate the distribution of tunnels in CYP2C5, CYP2C8, CYP2C9, CYP2A6 and CYP3A4 structures. In this work, we have accepted as significant, only those tunnels with a gorge width of 1.4 Å. This cut-off was determined from a thresholding exercise designed to strike an optimal balance between the number of "tunnel-open"/"tunnel-closed" pairs generated for subsequent analysis and the verisimilitude the CAVER-calculated tunnels bear to REMD tunnels. It has largely been possible to assign the CAVER-calculated tunnels to REMD pathways. The assignments to pw2b, pw2c, pw2e, pw3 and W have been straightforward for all the significant CAVER-calculated tunnels in this work. On the other hand, there are cases where it has not been clear if assignment of CAVER-calculated tunnels to pw2a would be more appropriate than assignment to tunnel S and vice versa. An extreme example of this is tunnel number 2 in the CYP3A4 structure 2J0D (see Table 1), which falls midway in the space between tunnel S and tunnel pw2a; we have therefore not assigned it to any REMD tunnel in this work. It is also noted that a number of CAVER-calculated tunnels may be assigned to the same REMD pathway as explained in the Introduction.

The distribution of tunnels in Tables 1 and S1 was subsequently used to select tunnel-open/tunnel-closed pairs of structures for calculation of two-state gating models and singlet gating models. We have deduced two-state gating models (reported in Table 2 and Table S2) from tunnel-open/tunnel-closed structures of the same CYP450 isoform and singlet gating models (reported in Table 3 and Table S3) from tunnel-open/tunnel-closed structures of different CYP450 isoforms. It is not always possible to successfully deduce gating models from tunnel-open/tunnel-closed pairs identified from Table 1 and Table S1. The most common reason for failure has been substantial intersections of main chain atoms of the tunnel-closed structure with the tunnel of interest calculated in the tunnel-open structure after alignment of the tunnel-open and tunnel-closed structures. This occurrence causes a breakdown in our approach since we assume for the purposes of this work that tunnel closure is solely due to obstructions by side-chains (i.e. we use side chain motions to explain the tunnel patterns in CYP450 structures). In this manner, pw2e gating residues in CYP2A6 cannot be deduced from our analysis of the pw2e tunnel-open CYP2C9 structure 1R9O and the pw2e tunnel-closed CYP2A6 structure 2FDV because residues 105 to 107 inclusive in 2FDV clash with the CAVER-calculated pw2e tunnel in 1R9O after alignment in Pymol. Similarly, the delimitation of pw2b-gating residues from the tunnel-open/tunnel-closed pair of structures 2NNH/2FDV (CYP2C8/CYP2A6) is not possible because CYP2A6 residues Glu 103 to Ala 105 inclusive clash with the CAVER-calculated pw2b tunnel in 2NNH after alignment of 2NNH and 2FDV in Pymol.

It is obvious that singlet gating models and two-state gating models cannot be defined for a given REMD tunnel if that tunnel is open in all structures of a given CYP450 isoform. For that reason, our approach cannot delimit singlet gating models and two-state gating models for pw2b and pw2e in CYP3A4 as these tunnels are open in all the CYP3A4 structures. This has been taken into account in the definition of Q2, a metric that expresses the percentage of literature molecular dynamics 3-tuple gating models that have also been identified on this work. In the calculation of Q2, the three models in the literature (models 3, 5 and 6 in Table 4) have been omitted to leave a total of 6 MD models in the literature set. This limited set is expected to achieve a more unbiased assessment of the overlap between models delimited herein and those described in the literature.

The value of Q2 is calculated over the limited set of literature MD 3-tuple gating-residue models as 67%. This highlights the high degree of coverage of literature gating models by the set of models identified in our work. The singlet gating models identified in this work are reported as 2-tuple models (see Table 3 and Table S3) whilst on the other hand MD gating models in the literature are reported as 3-tuples. In order to calculate P1, the percentage of singlet gating models identified in this work that have also been identified in the literature, it is necessary to define a model for the extraction of 3-tuples from the 2-tuple models listed in Table 3 and Table S3. Such a model would then permit the comparison of like objects in the calculation of P1. We have chosen a 'pessimistic' model for the extraction of 3-tuples that assumes that only one 3-tuple model can be generated from each of the 2-tuple models listed in Table 3 and Table S3. That is, of the all the amino-acid residues making interactions with the tunnel-intersecting residue as listed in the 'comments' section of Table 3 and Table S3, only one residue makes interactions that occur exclusively in the tunnel-closed structure. These 3-tuple models would be generated by pairing a given tunnel-intersecting residue with a residue delimited to be involved in an interaction of sorts with the tunnel-intersecting residue. Alternatively, it is possible to base calculations on an "optimistic" model that generates 3-tuples from all the residues delimited to interact with a given tunnel-intersecting residue. That is, here it is assumed that all of the interactions involving the tunnel-intersecting residue listed in the "comments" section of Table 3 and Table S3 occur exclusively in the tunnel-closed structure. Using this pessimistic model, we describe 61 new potential gating residue pairs for cytochrome P450 enzymes, as well as independently re-finding the majority (4 out of 6) of such gating residue pairs previously described in the literature. One trivial explanation for the large number of new gating models found here is that we have

extended the analysis over a larger number of CYP450 enzymes than previously analysed. However, even if we just consider CYP3A4, we found 16 gating models in this work, of which only 4 had previously been described in the literature. The statistic P1 is calculated as 4/65 = 4% across the entire set of CYP450 isoforms analysed here and as 4/27 = 15% across the same spread of CYP450 isoforms as in the literature (CYP3A4 & CYP2C5). This suggests that the body of new gating models described for the first time here will be of significant value in the further study of ligand-tunnelling problems in CYP450s; it also suggests that the combination of tools such as CAVER and Protein Interaction Calculator (PIC) might provide a simple, rapid and comprehensive alternative to complex MD simulations as a means to explore tunnels and tunnel gating in other classes of protein.

It is interesting to note that for those literature CYP450 gating residue pairs which we did not re-find directly in this work, in each case we did find the same pair of individual residues, but predict that they instead interact directly with different partners. Indeed, in general we find multiple potential gating interactions for each individual gating residue, suggesting either that previous MD simulations have been remarkably adept at choosing between different potential interactors, or that the picture is more complex than first thought. Further evidence of such potential complexity arises from the observation that our work has identified clusters of aromatic gating residues which might provide a mechanism for coordinated opening/closing of different tunnels; we have recently initiated correlated motion analyses to investigate this possibility further, results of which will be presented elsewhere. The high value of the statistic Q2 (67%) and the fact that 11/12 = 92% of residues that appear in the limited set of literature molecular dynamics gating models (i.e. the set that excludes gating residues for CYP3A4 tunnel pw2b and pw2e) have also appeared in the models delimited in this work, underscores the utility of the two major assumptions we have made in the development of a frame-work for the identification of gating models: that amino-acid side chains play a paramount role in the gating of tunnels and that the CAVER-calculated tunnels that we subsequently assign to REMD tunnels are sufficiently related to REMD tunnels and relevant enough to allow the accurate extraction of gating models.

In addition, through this work we have identified a network of ionic interactions present in all the CYP3A4 structures that restricts the space of the CYP3A4 surface available for exit of the pw2b tunnel. The pairs of ionic interactions that make-up the interaction network are as follows: Arg 106-Asp 76, Arg 106-Glu374, Arg 372-Asp 76 and Arg 372-Glu374. We speculate that this space-restricting network of interactions may operate together with proximal bulky residues such as Phe 220 to create an effective pw2b gating system. Fishelovitch and co-workers have found that the ionic interactions Asp 76-Arg 106 and Asp 76-Arg 372 are highly stable and remain unperturbed during product exit in SMD simulations (Fishelovitch et al., 2009). This is in harmony with the notion that the ionic interactions are a component of a wider pw2b-gating mechanism in which they specifically serve to restrict the exit space available to pw2b on the CYP3A4 surface. We have not managed to delimit pw2b-gating residues in this work as pw2b is open in all CYP3A4 structures. However, Phe 220 is proximal to CAVER-calculated pw2b tunnels in all the four CYP3A4 structures and side chain rotamers of Phe 220 that intersect pw2b can readily be selected from rotamer databases (data not shown). It is therefore plausible that the swinging motions of the Phe 220 side chain form the basis for gating of the space-restricted pw2b tunnel. We also note that Fishelovitch and co-workers identified Phe 220 as a pw2b gating-residue (3-tuple model is (Phe 220, Phe 108, pw2b)) in SMD simulations (Fishelovitch et al., 2009).

A survey of the gating models delimited in this work reveals that the most conserved gating model is the gating of tunnel W by a tunnel-intersecting Tryptophan residue whose position on the CYP450 fold is also conserved; this residue is Trp 126 in CYP3A4, Trp 120 in CYP2C5, Trp 120 in CYP2C8 and Trp 120 CYP2C9. Interestingly, the gating of tunnel W by Trp 126/Trp 120 is the only gating model that is conserved between the CYP2 and CYP3 families. It is also the only gating model in the CYP2C sub-family where the tunnel-intersecting residue is absolutely conserved both in residue position and in residue identity. There are tunnel-intersecting residue positions and residue types that show a reasonable degree of conservation in the CYP2C sub-family, notably those involved in the gating of pw3: in CYP2C5 these residues are Leu 207, Ile 229 and Leu 233; in CYP2C8 these residues are Ile 207, Leu 208, Thr 229 and Val 233; and in CYP2C9 these residues are Ile 207, Leu 208, Thr 229 and Tyr 225. It can be seen that the sequence positions 207 and 229 are involved in gating in all the CYP2C isoforms considered here; not surprisingly, there is greater conservation of pw3-intersecting residues between CYP2C8 and CYP2C9 – Ile 207, Leu 208 and Thr 229 intersect pw3 in both – in accord with the close sequence resemblance of CYP2C8 and CYP2C9 (78% identity). The gating models of tunnel pw2c are another class of models that show interesting features including a fairly high degree of conservation in position and residue identity of the tunnel-intersecting residue: In CYP2C5 the pw2c-intersecting residue Lys 241 is located on the G-helix, interacting with Val 106, a residue located on the BC-loop/B′ helix; and in CYP2C8 the pw2c-intersecting residue is Arg 241, interacting with the I-helix (N-term) residue Glu 285. Hence in CYP2C5 and CYP2C8, the position 241 is a conserved tunnel-intersecting position where a basic amino-acid side-chain intersects pw2c. However, an interesting charge switch occurs in CYP2C9 whereby the tunnel-intersecting residue is Arg 108 – a residue located on the BC-loop/B′-helix. Furthermore, CYP2C9 Arg 108 interacts with, among others (see Table S2), Asp 293 – a residue located in the N-terminal region of the I-helix.

The analysis of amino acid chemical types (summarized in Table 5) of the entire set of tunnel-intersecting residues (listed in Tables 2 and 3, S2 and S3) reveals a number of interesting trends. Table 5 reveals a strong presence of hydrophobic aromatic tunnel-intersecting residues in CYP3A4. Here hydrophobic aromatic side chains account for 68.75% of the tunnel-intersecting residues in CYP3A4 gating models. In contrast to the situation in CYP3A4, in CYP2C5 there is a more predominant presence (54.55%, Table 5) of polar/charged tunnel-intersecting residues in the gating models, whilst in CYP2C8 there is equal presence of polar/charged (42.86%, Table 5) and hydrophobic aliphatic (42.86%, Table 5) tunnel-intersecting residues in the gating models. The combined percentage presence of polar/charged and hydrophobic aliphatic residues (85.72%) is much greater than the percentage presence of hydrophobic aromatic residues in the CYP2C8 gating models (14.29%). Unlike in CYP3A4, CYP2C5 and CYP2C8, in CYP2C9 gating models there is a more equitable presence of hydrophobic aromatic (38.46%), polar/charged (30.77%) and hydrophobic aliphatic (30.77%) tunnel-intersecting residues with a slight bias for hydrophobic aromatic residues. Finally, in CYP2A6, like in CYP3A4, there is a predominant presence (54.55%) of hydrophobic aromatic tunnel-intersecting residues. It is interesting to note that CYP2A6 is more akin to CYP3A4 than it is to other CYP2 isoforms studied here (CYP2C5, CYP2C8 and CYP2C9) in its use of hydrophobic aromatic tunnel-intersecting residues. It is clear from Table 5 that there is very different usage of amino-acid side chain chemical types in the CYP2 family and within the CYP2C sub-family. Given the levels of sequence similarity within the CYP2C sub-family and given the different usage of side chain chemical types of tunnel-intersecting residues, we expected that the rate of evolution of the tunnel-intersecting residues would be greater than the rate of evolution of the entire sequence in each of the CYP2 isoforms considered in this

work. Indeed the average rate of evolution of the tunnel-intersecting residues was found to be significantly faster than the average rate of evolution of the entire sequence for each of CYP2C5, CYP2C8, CYP2C9 and CYP2A6 (see Table 6). In keeping with the relatively fast evolutionary dynamics of tunnel-intersecting residues, the tunnel-intersecting residue positions are not among the most conserved positions in sequence alignments (see Fig. S1).

We previously built a model for the interaction between mammalian CYP450 and P450-oxidoreductase (POR) proteins (Zawaira et al., 2010). The following CYP2B4 residues are special in the model because they have been experimentally-delimited by others as interfacial residues: R85, M137, E143, K384, K421, R422, K433 and R443. Analysis of the corresponding residues in CYP2A6; R88, V140, E146, K387, K424, K425, K436 and R446, respectively, CYP2C5; K84, M136, E142, K380, K417, K418, K429 and R439, respectively, CYP2C8; K84, M136, E142, K383, K420, K421, K432 and R442, respectively, and CYP2C9; K84, M136, E142, K383, K420, K421, K432 and G442, reveals a very high degree of conservation of these residues – unlike the much lower conservation of tunnel-gating residues. Current models for the evolution of protein function assert that the rate of evolution increases from the catalytic site (centre of chemical function) to the protein core (structural core), to surface regulation regions such as protein–protein recognition sites and to ordinary surface sites (Valencia, 2005). On the basis of the conservation patterns described above, we cast tunnel-gating residues into this hierarchical evolution model by placing them between surface regulation regions and ordinary surface sites. When cast in this manner, this gives rise to the view that the evolution of viable gating interactions occurs more spontaneously within the CYP450 protein than the co-evolution of interactions between CYP450 and POR that support functional CYP450-POR recognition. In turn, the co-evolution of interactions that support functional CYP450-POR recognition is assumed to be more spontaneous than the evolution of the structural core of CYP450. Indeed, the average relative rate of evolution of the above-mentioned eight interfacial positions in CYP2A6, CYP2C5, CYP2C8 and CYP2C9 is much lower than that of the tunnel-gating positions (0.6488625 versus the values in Table 6). One could speculate that the higher rate of evolution of gating residues is permitted because viable gating residue networks may arise at several points along the tunnel hence, unlike surface regulation sites such as protein–protein binding sites, many permutations can be attempted faster until a viable one is attained. However, we note that residues that line these tunnels are probably constrained in their evolution by selection against combinations that may cause local structure collapse or any other irreversible alteration of tunnels.

Finally, the amino acid residue interaction patterns delimited in this work suggests the presence of clusters of residues in CYP450 isoforms which we hypothesise to form regulatory networks through which the enzyme achieves co-ordinated tunnel gating for the multiple tunnels in CYP450. We are currently investigating these clusters using correlated motions analysis and preliminary data strongly supports the hypothesis that the clusters comprise a regulatory network. This work will be reported elsewhere.

Author contributions

A.Z. and J.B. designed research; A.Z., L.C., M.G. and O.K., performed research; A.Z., L.C., M.G. and O.K. analysed data; A.Z., L.C., M.G., O.K., and J.B. wrote the paper.

Conflict of interest

No conflict of interest declared.

Acknowledgments

L.C and O.K thank the National Research Foundation (NRF) of South Africa for financial support through bursaries. A.Z. wishes to thank the University of Cape Town Clinical Infectious Diseases Research Initiative (CIDRI) for a research fellowship. J.B. wishes to thank the South African Research Chair Initiative (SARChI) for a research chair.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2010.09.026.

References

Boddupalli, S.S., Hasemann, C.A., Ravichandran, K.G., Lu, J.Y., Goldsmith, E.J., Deisenhofer, J., Peterson, J.A., 1992. Crystallization and preliminary x-ray diffraction analysis of P450terp and the hemoprotein domain of P450BM-3, enzymes belonging to two distinct classes of the cytochrome P450 superfamily. Proc. Natl Acad. Sci. USA 89, 5567–5571.

Bridges, A., Gruenke, L., Chang, Y.T., Vakser, I.A., Loew, G., Waskell, L., 1998. Identification of the binding site on cytochrome P450 2B4 for cytochrome b5 and cytochrome P450 reductase. J. Biol. Chem. 273, 17036–17049.

Cojocaru, V., Winn, P.J., Wade, R.C., 2007. The ins and outs of cytochrome P450s. Biochim. Biophys. Acta 1770, 390–401.

Cupp-Vickery, J.R., Poulos, T.L., 1995. Structure of cytochrome P450eryF involved in erythromycin biosynthesis. Nat. Struct. Biol. 2, 144–153.

Fishelovitch, D., Shaik, S., Wolfson, H.J., Nussinov, R., 2009. Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism. J. Phys. Chem. B 113, 13018–13025.

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

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

Haines, D.C., Tomchick, D.R., Machius, M., Peterson, J.A., 2001. Pivotal role of water in the mechanism of P450BM-3. Biochemistry 40, 13456–13465.

Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A., Deisenhofer, J., 1995. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure 3, 41–62.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J., 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. 23, 403–405.

Lamb, D.C., Lei, L., Warrilow, A.G., Lepesheva, G.I., Mullins, J.G., Waterman, M.R., Kelly, S.L., 2009. The first virally encoded cytochrome p450. J. Virol. 83, 8266–8269.

Ludemann, S.K., Carugo, O., Wade, R.C., 1997. Substrate access to cytochrome P450cam: a comparison of thermal motion pathway analysis with molecular dynamics simulation data. J. Mol. Model. 3, 369–374.

Ludemann, S.K., Lounnas, V., Wade, R.C., 2000a. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. J. Mol. Biol. 303, 797–811.

Ludemann, S.K., Lounnas, V., Wade, R.C., 2000b. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. J. Mol. Biol. 303, 813–830.

Park, S.Y., Shimizu, H., Adachi, S., Shiro, Y., Iizuka, T., Nakagawa, A., Tanaka, I., Shoun, H., Hori, H., 1997a. Crystallization, preliminary diffraction and electron paramagnetic resonance studies of a single crystal of cytochrome P450nor. FEBS Lett. 412, 346–350.

Park, S.Y., Shimizu, H., Adachi, S., Nakagawa, A., Tanaka, I., Nakahara, K., Shoun, H., Obayashi, E., Nakamura, H., Iizuka, T., Shiro, Y., 1997b. Crystal structure of nitric oxide reductase from denitrifying fungus *Fusarium oxysporum*. Nat. Struct. Biol. 4, 827–832.

Petrek, M., Otyepka, M., Banas, P., Kosinova, P., Koca, J., Damborsky, J., 2006. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinf. 7, 316.

Pupko, T., Bell, R.E., Mayrose, I., Glaser, F., Ben-Tal, N., 2002. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics 18 (Suppl. 1), S71–S77.

Sansen, S., Yano, J.K., Reynald, R.L., Schoch, G.A., Griffin, K.J., Stout, C.D., Johnson, E.F., 2007. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. J. Biol. Chem. 282, 14348–14355.

Schleinkofer, K., Sudarko, P.J., Winn, P.J., Ludemann, S.K., Wade, R.C., 2005. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? EMBO Rep. 6, 584–589.

Schoch, G.A., Yano, J.K., Wester, M.R., Griffin, K.J., Stout, C.D., Johnson, E.F., 2004. Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. J. Biol. Chem. 279, 9497–9503.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Tina, K.G., Bhadra, R., Srinivasan, N., 2007. PIC: Protein Interactions Calculator. Nucleic Acids Res. 35, W473–W476.

Valdar, W.S., 2002. Scoring residue conservation. Proteins 48, 227–241.

Valencia, A., 2005. Automatic annotation of protein function. Curr. Opin. Struct. Biol. 15, 267–274.

Wade, R.C., Winn, P.J., Schlichting, I., Sudarko, P.J., 2004. A survey of active site access channels in cytochromes P450. J. Inorg. Biochem. 98, 1175–1182.

Wade, R.C., Motiejunas, D., Schleinkofer, K., Sudarko, P.J., Winn, P.J., Banerjee, A., Kariakin, A., Jung, C., 2005. Multiple molecular recognition mechanisms. Cytochrome P450 – a case study. Biochim. Biophys. Acta 1754, 239–244.

Werck-Reichhart, D., Feyereisen, R., 2000. Cytochromes P450: a success story. Genome Biol 1: REVIEWS3003.

Wester, M.R., Yano, J.K., Schoch, G.A., Yang, C., Griffin, K.J., Stout, C.D., Johnson, E.F., 2004. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. J. Biol. Chem. 279, 35630–35637.

Winn, P.J., Ludemann, S.K., Gauges, R., Lounnas, V., Wade, R.C., 2002. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. Proc. Natl Acad. Sci. USA 99, 5361–5366.

Yano, J.K., Wester, M.R., Schoch, G.A., Griffin, K.J., Stout, C.D., Johnson, E.F., 2004. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. J. Biol. Chem. 279, 38091–38094.

Zawaira, A., Matimba, A., Masimirembwa, C., 2008. Prediction of sites under adaptive evolution in cytochrome P450 sequences and their relationship to substrate recognition sites. Pharmacogenet. Genomics 18, 467–476.

Zawaira, A., Gallotta, M., Beeton-Kempen, N., Coulson, L., Marais, P., Kuttel, M., Blackburn, J., 2010. Exhaustive computational search of ionic-charge clusters that mediate interactions between mammalian cytochrome P450 (CYP) and P450-oxidoreductase (POR) proteins. Comput. Biol. Chem. 34, 42–52.

Send Orders for Reprints to reprints@benthamscience.net

Current Drug Metabolism, **2013,** *14,* **745-763**

Combining *In Silico* Protein Stability Calculations with Structure-Function Relationships to Explore the Effect of Polymorphic Variation on Cytochrome P450 Drug Metabolism

Lauren Arendse*[,‡], Tom L. Blundell[‡] and Jonathan Blackburn*[,†]

**Institute for Infectious Disease & Molecular Medicine, Division of Medical Biochemistry, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa; [‡]Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, United Kingdom*

Abstract: We carried out an *in silico* structural analysis of 348 non-synonymous single nucleotide polymorphisms, found across nine of the major human drug metabolising cytochrome P450 isoforms, to determine the effects of mutations on enzyme structure and function. Previous functional studies in our group have delineated regions of the cytochrome P450 structure important for substrate recognition, substrate and product access and egress from the active site and interaction with the cytochrome P450 reductase. Here we combine the information from those studies with new *in silico* calculations on the effect of mutations on protein stability and we compare our results to experimental data in order to establish the likely causes of altered drug metabolism observed for cytochrome P450 variants in functional assays to date, in the process creating a cytochrome P450 polymorphic variant map.

Using the computational tool Site Directed Mutator we predicted destabilising mutations that result in altered enzyme function *in vitro* with a specificity of 83%. We found that 75% of all cytochrome P450 mutations that show altered activity *in vitro* are either predicted to be destabilising to protein structure or are found within regions predicted to be important for catalytic activity. Furthermore, we found that 70% of the mutations that showed similar activity to the wild-type enzyme in *in vitro* studies lie outside of functional regions important for catalytic activity and are predicted to have no effect on protein stability. Our resultant cytochrome P450 polymorphic variant map should therefore find utility in predicting the likely functional effect of uncharacterised variants on drug metabolism.

Keywords: Cytochrome P450, drug metabolism, protein stability, polymorphic variation, single nucleotide polymorphisms (SNPs).

INTRODUCTION

Cytochrome P450s (CYP450s), a large functionally diverse superfamily of haem-thiolate enzymes found across all lineages of life [1], are responsible for the metabolism of a wide range of structurally diverse endogenous and exogenous compounds. CYP450s are most well known for their important role in phase I drug metabolism, oxidising 70-80% of pharmaceutical drugs [2, 3], with substrate oxidation being achieved by the insertion of an active oxygen molecule, derived from molecular oxygen, into the substrate.

Camphor-metabolising CYP101 (P450cam) from *Pseudomonas putida* was the first CYP450 to have its crystal structure solved [4] and extensive studies on this enzyme, combined with other subsequently solved bacterial CYP450 structures, provided the initial framework for our understanding of structure-function relationships in CYP450 enzymes. CYP450s have a globular three-dimensional fold typically made up of six β-strands and thirteen α-helices surrounding a large buried hydrophobic active site [5]. Today there are over 500 CYP450 structures available in the Protein Data Bank (PDB), including over 100 mammalian structures. As predicted by Hasemann and co-workers [6], despite very low sequence identities, the CYP450 fold is highly conserved in eukaryote structures. The haem-binding region is the most structurally conserved, apart from local differences in the I-helix. A number of key conserved structural features have been described including the conserved ERR triad [6], the I-helix catalytic groove [5, 7, 8] and the hydrophobic Cys-pocket or the B-bulge surrounding the catalytic cysteine residue that coordinates to the haem ion. By comparison the substrate binding region is a more variable and flexible region. Whilst all prokaryotic CYP450s are soluble and located in the cytosol, most eukaryotic CYP450s are bound to endoplasmic reticulum membrane by an N-terminal anchor and are known as microsomal CYP450s; hydrophobic regions in the F and G helices are also thought to facilitate membrane binding. Similarities and differences found across different CYP450 structures have been discussed in detail in a number of recent reviews [9-11].

Most CYP450 crystal structures are in a "closed" conformation with the active site completely cut-off from the bulk solvent but some structures show "open" or "wide-open" conformations [12, 13], suggesting that CYP450s undergo considerable conformational changes to allow substrate/products and water molecules in and out of the active site. A number of approaches, including thermal motion pathway analysis [14], adiabatic mapping [15], molecular dynamics [14, 16-22] and computational geometry tools [23-28] have been used in an attempt to understand CYP450 flexibility, differential tunnel opening and gating patterns observed in these enzymes. These independent methods are in general agreement and several pathways found in distinct parts of the CYP450 fold have been identified as putative substrate access and solvent tunnels; however the gating patterns observed for these tunnels are not necessarily conserved between isoforms, perhaps suggesting fast evolutionary dynamics [25].

Substrate recognition sites (SRS) are of particular interest in the CYP450 field because drug metabolizing mammalian CYP450s are highly promiscuous enzymes that are able to metabolise a wide range of diverse substrates, yet are still able to maintain a degree of specificity. An updated SRS map based on CYP450-ligand interactions in mammalian crystal complexes and docking studies has

[†]Address correspondence to this author at the Institute for Infectious Disease & Molecular Medicine, Division of Medical Biochemistry, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa; Tel: +27 21 406 6071; E-mail: jonathan.blackburn@uct.ac.za

1875-5453/13 $58.00+.00

recently been described [29] that identifies a number of new SRS regions not predicted in Gotoh's original SRS map [30].

During the catalytic cycle, mammalian microsomal CYP450s source electrons from their redox partner NADPH dependent CYP450 reductase (CPR), which forms a transient complex with CYP450 during electron transfer [31]. CPR is made up of four domains: FMN-binding, connecting, FAD-binding and NADPH-binding. A number of studies have contributed to elucidating the CYP450-CPR interface and have identified clusters of basic residues on the proximal surface of CYP450s that interact with exposed acidic clusters on the CPR FNM domain surface [32-37].

The human genome encodes 57 different CYP450 enzymes but the majority of human drug metabolism can be attributed to a small number of highly polymorphic isoforms, namely CYP1A2 (5%), CYP2A6 (2%), CYP2B6 (2-4%), CYP2C8 (1%), CYP2C9 (10%), CYP2C19 (5%), CYP2D6 (20-30%), CYP2E1(2-4%) and CYP3A4 (40-45%) [38], where the number in brackets indicates the percentage of drug metabolism carried out by each isoform. The distribution of this polymorphic variation differs between ethnic groups and has been shown to affect drug response [39]. Single nucleotide polymorphisms (SNPs) occur both in the coding and non-coding regions of these genes, resulting in effects on CYP450 expression as well as function [38, 40]. Phenotypic effects of non-synonymous polymorphisms have been shown to be substrate specific in a number of cases [41-44] making prediction of the effect of polymorphisms on drug metabolism difficult. Importantly though, structural stability, haem binding, substrate access and recognition, as well as interactions with CPR, are all aspects of CYP450 function that can be directly affected by changes in the amino acid sequence.

In vitro approaches have been used to measure the effect of non-synonymous polymorphic variation on CYP450 structure and function and a variety of expression systems such as bacteria, yeast, insect cells and mammalian cells are now commonly used to express stable CYP450 enzymes with similar catalytic properties to human liver microsomes. Recombinant expression systems for individual CYP450 enzymes have enabled site-directed mutagenesis studies to investigate the effect of SNPs on protein function. CYP450-holoprotein levels are routinely obtained from carbon monoxide difference spectra [45] allowing the quantification of total haem incorporation and differentiating between correctly folded holoprotein (P450) and misfolded holoprotein (P420). *In vitro* screening techniques have been developed to monitor CYP450 substrate turnover, including radiometric, fluorogenic and chromatography/mass spectrometry based activity assays [46-48]; and miniaturised high throughput platforms have begun to make CYP450 activity assays more time and cost effective [49-51]. A high throughput screening platform has been used recently to assay 18 CYP2C19 variants against 10 different drugs [43]; however few such studies have been reported to date and as a result, the effect of non-synonymous mutations on drug metabolism remains poorly understood experimentally for all except a limited set of mutations and substrates.

The sheer number of CYP450 polymorphisms, combined with the many diverse potential drugs, highlights the need for improved *in silico* prediction tools and experimental approaches for determining the effects of single amino acid substitutions in CYP450 enzymes on drug metabolism. Here we analyse the impact of polymorphic variation on enzyme stability and function in mammalian CYP450 drug metabolising enzymes: We use representative *in silico* modeling tools to predict the qualitative effect of mutations on protein stability; We then combine the resulting data with previous *in silico* analyses of conserved structural features, SRS regions, CPR contacts and tunnel gating residues to create a unified CYP450 SNP map; and finally we use this CYP450 SNP map to correlate the location of individual mutations with their experimentally observed impact on CYP450 function.

METHODS

Crystal Structures

Structures of human CYP450 proteins and rat CYP2B4 were retrieved from the RCSB Protein Data Bank [52]. The PDB codes for the structures used in this study are as follows. **CYP1A2**: 2HI4. **CYP2A6**: 1Z10, 1Z11, 2FDU, 2FDV, 2FDW, 2FDY, 2PG5, 2PG6, 2PG7, 3EBS. **CYP2B6**: 3IBD. **CYP2C8**: 1PQ2, 2NNH, 2NNI, 2NNJ, 2VN0. **CYP2C9:** 1OG2, 1OG5, 1R9O. **CYP2C19:** 4GQS. **CYP2D6:** 2F9Q. **CYP2E1:** 3E6I, 3E4E, 3GPH, 3KOH, 3LC4. **CYP3A4**: 1TQN, 1W0E, 1W0F, 1W0G, 2J0D, 2V0M, 3NXU. **CYP2B4** (rat):2BDM. Only chain A was used for structures that have more than one CYP450 chain in the asymmetric unit.

Non-synonymous Single Nucleotide Polymorphisms

Non-synonymous single nucleotide polymorphisms (SNPs) were sourced from the Human Cytochrome P450 Allele Nomenclature Committee home page [53] and NCBI dbSNP [54] for isoforms CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and CYP3A4.

Comparative Modeling of Polymorphisms Using ANDANTE

Polymorphisms were modelled using ANDANTE, a program that makes use of environmental substitution tables (ESSTs) to reduce the side chain rotamer search space during comparative modeling of single amino acid mutations [55]. A Python script was written to run ANDANTE in batch mode. Mutant structures were generated for each SNP using all the available PDB crystal structures for the respective isoform. ANDANTE was run using the default parameters and the "- local" option which attempts to minimise all rotamer clashes in the local area that result from the mutation. Note that ANDANTE does not take co-factors or ligands into account when modeling side chain positions. All mutant structures for each SNP were visualised and compared to the original template crystal structures using PyMOL [56].

Predicting the Effect of Polymorphisms on Protein Stability Using SDM

Site directed mutator (SDM) [57], a statistical potential energy function first developed by Topham *et al* [58], was used as a representative *in silico* tool to predict the effect of each SNP on protein stability. SDM uses environment-specific substitution frequencies within homologous protein families to calculate a stability score. The environment specific substitution tables (ESSTs) provide information on the amino acids local structural environment and the probability of it being substituted by another amino acid. The algorithm uses these ESSTs to calculate a stability score, which predicts the difference in free energy between a wild-type and mutant protein, by analogy to a reversible folding-unfolding thermodynamic cycle.

SDM input: SDM requires a wild-type and mutant structure (which has only one single amino acid substitution) in PDB format as input. The mutant structures built using ANDANTE, together with their template structures, were used as input for the mutant and wild-type structures respectively.

SDM output: SDM output gives details of the local structural environment of the wild-type and mutant residue which includes secondary structure prediction and percentage solvent accessibility. SDM also outputs a pseudo $\Delta\Delta G$ score or stability score and a prediction of disease association. Negative scores suggest that the mutation is destabilising and positive scores indicate a stabilising effect. A cut-off of 2 $kcal.mol^{-1}$ had previously been used to indicate that a mutation is damaging to the protein structure. Both stabilising and destabilising mutations that reach this cut-off were therefore categorised as damaging mutations whilst mutation with scores between -2 and 2 $kcal.mol^{-1}$ are classified as neutral. It must be noted that like ANDANTE, SDM does not consider co-factors or ligands when calculating solvent accessibility or stability scores.

SDM was run on all ANDANTE derived mutant structures. If ANDANTE generated more than one structure for the same SNP, an average SDM score and standard deviation were calculated. In subsequent analysis, we used the output from SDM in a qualitative manner, describing mutations as damaging or neutral and made no further specific use of the absolute numerical SDM score.

Generating an Annotated Structural Alignment

A multiple CYP450 structural alignment was generated including one representative structure for each isoform using the program Baton (D.Burke, unpublished data and [59]). The PDB codes for representative structures were as follows CYP1A2: 2HI4, CYP2A6: 2FDV, CYP2B6: 3IBD, CYP2C8: 1PQ2, CYP2C9**:** 1R9O, CYP2C19: 4GQS, CYP2E1: 3E4E, CYP2D6: 2F9Q, CYP3A4: 1TQN.

The program JOY [60] was used to annotate the protein sequence alignment with 3D structural features. Here is a brief description of the structural features represented by JOY.

Solvent accessibility is measured using the program PSA which implements the algorithm developed by Lee and Richards (1971). A relative solvent accessibility (RSA) is calculated by comparing an extended conformation of a residue to the residue conformation in question. The default cut-off setting of 7% RSA was used. This classifies residues with less than 7% RSA as inaccessible and residues of >7% as accessible. In the Joy format, buried residues are displayed in upper case and accessible residues in lower case.

Secondary structure assignment is calculated using the program SSTRUC (D.K.Smith, unpublished). In the JOY format, residues forming α, 3_{10} and π helices, and β strands are displayed in different colours. The default consensus secondary structure definition was used which means a consensus secondary structure is assigned, underneath the alignment, if more than 70% of the residues at that position are in a particular conformational state. α-helices A – L and β-sheets 1-5 are labelled below the alignment according to CYP450 nomenclature of Poulos and co-workers [5].

Main chain dihedral angles are also calculated by SSTRUC. Residues with a positive main chain φ angle are shown in italics.

Hydrogen bonding within a structure is calculated using the program HBOND (J.Overington, unpublished), which calculates the associated energy and angles for interactions between heavy atom donor and acceptor pairs. JOY uses a lenient distant cut-off of 3.5Å for hydrogen bonds. The hydrogen bond data is split into three classes: 1) side chain to main chain amide; 2) side chain to main chain carbonyl; 3) side chain to other side chain. Residues whose side chain is bonded to a main chain amide are denoted in bold face. Residues that are underlined indicate that their side chain is hydrogen bonded to a main chain carbonyl. Residues whose side chain is hydrogen bonded to another side chain are not displayed in the JOY format used here.

Defining CYP450 Functional Regions

The structural alignment was further annotated with known functional regions. If residues did not fall within any of these regions then their functional region was classified as undefined. Functional regions can overlap so one residue may fall within more than one functional region. The definitions of functional regions are as follows.

Conserved Motifs

The hydrophobic *signal-anchor sequence,* usually preceded by an acidic amino acid residue that targets the protein to the endoplasmic reticulum, is the 20-25 amino acid stretch near the N-terminus [61]. The *halting-signal* refers to the basic residues found between the signal-anchor sequence and the proline-rich motif, which ensure the correct orientation of the protein in the membrane [61]. The first two functional regions are truncated in the crystal structures and therefore do not appear in the structural alignment. The *proline rich motif* is defined as the N-terminal sequence motif PPGPXPXPXXGN, which has been shown to be important for protein folding rather than catalytic activity [62-64]. The conserved consensus motif (G/A)GX(D/E)T, located in the middle of the I-helix, has been labelled the *I-helix catalytic groove*. This region plays an important role in proton transport [7, 8] and is responsible for the formation of the oxygen binding pocket [5]. The *K-helix core stabilising motif* is the invariant EXXR motif within the K-helix that interacts with the meander region [6]. The *meander* is found between the K'-helix and the L-helix and is highly conserved in structure despite the lack of any secondary structure formation. The absolutely conserved Asp and Arg residues in the core stabilising motif interact with an Arg/His residue in the meander region, forming a set of salt bridges known as the *ERR triad*. The *Cys-pocket* refers to the sequence motif FXXGXXXCXG that forms a β-bulge around the absolutely conserved catalytic Cys residue.

Haem Contacts

Haem-residue interactions were identified using the protein-ligand interaction database CREDO [65]. CREDO assigns each contact residues to an interaction type based on the geometry and distance thresholds using criteria adapted from Marcou and Rognan [66]. Residues interacting with the haem group were identified in all the available PDB structures for each isoform, combined and mapped onto the representative structure in the sequence alignment. This was done to account for any differences in the haem binding regions for structures of the same isoform due to conformational changes.

SRS Regions

SRS regions were assigned based on the recently described mammalian CYP450 expanded and unified substrate recognition map [29]. This map was created by identifying interacting residues in mammalian crystal structure complexes as well as interacting residues observed in docking studies.

Cytochrome P450 Reductase (CPR) Contacts

CPR contacts were assigned based on the recent model for interactions between CYP2B4 and FMN domain of rat CYP450 reductase [37]. This model was built using the bacterial redox complex structure (CYP102: 1BVY, chain A) as a template and was used to identify the key residues involved in ionic interactions at the interface. All CYP2B4–P450-oxidoreductase interfacial residues calculated from the 1BVY.pdb-derived structural model at 6Å distance cut-off were mapped onto the human CYP450 representative structures *via* structural alignment. Residues forming part of ionic charge clusters where further categorised as *ionic charge cluster residues* to distinguish them from the other CPR contact residues at the interface.

Gating Residues

Tunnel-gating residues were assigned based on residues previously identified as tunnel-gating residues in the literature [19, 25]. Our group recently identified residues involved in putative gating models responsible for the differential tunnel-opening patterns seen in static structures of mammalian CYP450 [25]. (**CYP2A6**: F107, F111, F118, M205, L241, Q242, Q245, M293; **CYP2C8**: N99, S114, W120, R125, F205, I207, L208, N209, T229, V233, R241, E285, E300, V362, P367, M388, I476; **CYP2C9**: I99, A103, R108, F110, G111, F114, W120, I207, L208, Y225, T229, D293, F476; **CYP3A4**: F108, F113, W126, K127, R212, F213, D214, F215, F220, V240, F241, F304, E308, I369, L482.) Previous studies had used steered molecular dynamics simulations to identify residues serving as gate keepers to substrate/product tunnels in CYP3A4 enzymes [19]. (CYP3A4: F57, R105, F108, S119, I120, R212, F213, F215, F220, V240, F241, L482.)

Furthermore, solvent explicit molecular dynamics have previously been used to identify an aqueduct gating mechanism that is regulated by binding of the CYP450 reductase [67]. (**CYP3A4**: R375) As R375 is highly conserved in CYP450 is has been sug-

gested that this could be a general aqueduct gating mechanism. All residues at this position in the alignment have been defined as *aqueduct gating residues*. (**CYP1A2:** H388**, CYP2A6**: R372, **CYP2B6**: H369, **CYP2C8:** H368, **CYP2C9:** H368**, CYP2C19:** H368**, CYP2D6:** H376**, CYP2E1:** H370**, CYP3A4:** R375)

Experimental Data Collection

Available experimental data on protein expression, holoprotein levels, enzyme activity and kinetics (K_m, V_{max} and V_{max}/K_m) was gathered from the literature and tabulated for each SNP. (Table S1-S9 in the Supplementary Materials). Due to the wide range of data types and the great diversity in methodology all data was expressed as a fraction (variant data/ wild-type data) measured in the same experiment. If the difference between wild-type and variant data was reported to be statistically significant, it is indicated in Tables S1-S9 with an asterisks (some studies provide quantitative data without statistical evidence). If no quantitative data was available then data was expressed qualitatively as similar to wild-type "similar", significantly lower than wild-type "decreased", significantly higher than wild-type "increased" or not detectable "ND". All *in vivo* data was recorded qualitatively. If more than one non-synonymous SNP coexist in one allele and have therefore been tested as a double or triple mutant (as is common for CYP2D6), data was included only if it is clear which SNP is dominant in the resulting phenotype.

Classifying SNPs Based on Experimental Data

After careful analysis of all available data, SNPs were classified into six groups based primarily on the *in vitro* experimental evidence. *In vivo* data was however, considered if it was the only data available. "Inactive" refers to mutations that resulted in protein/holoprotein levels that are not detectable or activity that is less than 2% of wild-type activity for all expression systems and substrates tested. "Decreased activity" refers to SNPs that resulted in significantly decreased expression levels, haem incorporation or altered kinetic parameters, leading to impaired enzyme function. "Increased activity" refers to mutants that show significant increases in activity. "Substrate specific" refers to mutants that show more than 50% difference in mutant/wild-type activity ratios for different substrates. "Neutral" refers to mutants that show similar holoprotein levels and activity to that of the wild-type enzyme. "Unknown" refers to SNPs that have not yet been tested *in vitro* or for which the experimental data is inadequate or contradictory.

RESULTS

The effect of 348 individual CYP450 mutations on protein stability was calculated using the tool SDM [57]. Pseudo ΔΔG scores - together with predicted effect on stability – and structural information, including the relative solvent accessibility (RSA) of the wild-type and mutant residue, the secondary structure location and the functional region within which the mutation falls, is recorded for each of these CYP450 variants in Tables S1-S9 (Supplementary Materials). A visual representation of CYP450 secondary structure elements, functional regions and positions of SNPs is shown for CYP3A4 in (Fig. **1**).

Experimental Data

Experimental activity data was available for 38% of the SNPs analysed in this study (Fig. **2**); these data are collated in Tables S1-S9 (Supplementary Materials). 17% of these SNPs are categorised as inactive, 34% of SNPs lead to decreased activity and 11% displayed substrate specific effects on activity. It must be noted though that only a small percentage of SNPs have been tested on more than one substrate in the same *in vitro* experiment, so it is possible that some SNPs that currently fall within the neutral or decreased activity category, based on available data, would show substrate specific effects if tested on a broader range of different substrates. CYP3A4 L293P and CYP2C9 R150H were the only SNPs reported to give increased activity *in vitro*. The remaining 37% of SNPs tested showed no significant effect on enzyme expression or active. Table **1** lists the SNPs that could be classified based on experimental data, complete with their secondary structure, functional region location, their predicted effect on protein stability, and their classification based on experimental data.

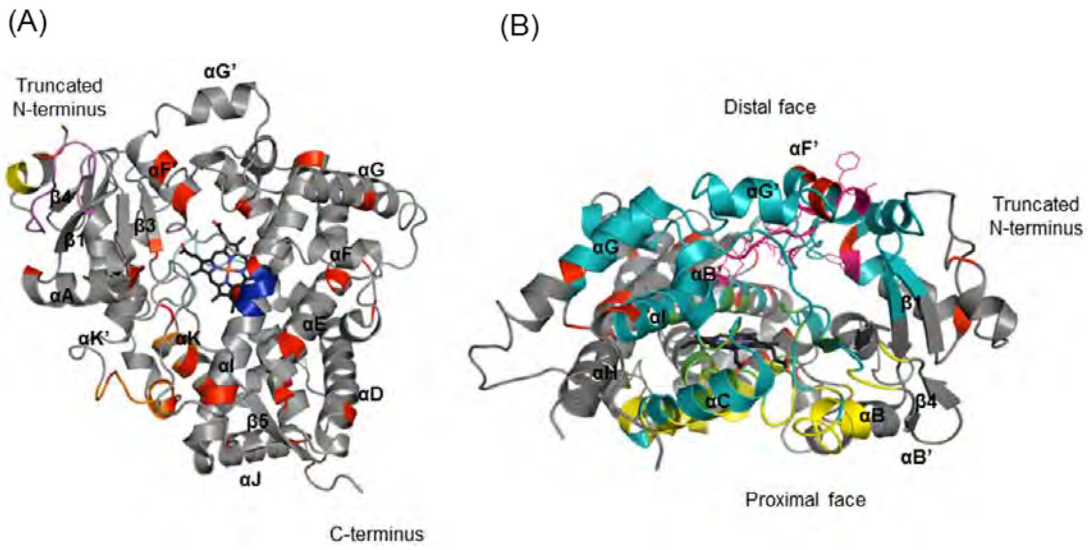


Fig. (1). CYP450 Structural overview

CYP3A4, PDB structure 1TQN, is shown in grey with the haem group in black. Secondary structure elements, α-helices and β-strands are labelled using the nomenclature of Poulos and co-workers [5]. Positions of polymorphic variants found in CYP3A4 are indicated in red. (**A**) Distal face view showing the position of the halting signal (yellow), the proline-rich motif (magnenta), the I-helix catalytic groove (blue), the core stabilising motif, the meander stabilised by the ERR triad (orange) and the cys-pocket (cyan). (**B**) Side view showing the position of the SRS regions (cyan), the gating residues (magnenta), the haem contacts (green) and the CPR interface (yellow) on the proximal face of the protein.

Table 1. Structural Information and Protein Stability Predictions for SNPs that could be Classified According to their Effect on Enzyme Activity Based on Experimental Data Collected from the Literature.

SNP	STRUCTURAL INFORMATION		PROTEIN STABILITY		Classification based on experimental data
	2° structure location	Functional region location	SDM Score kcal.mol^{-1}	Prediction	
CYP1A2					
P42R	loop btwn N-term & helix A	proline rich motif	-0.3	neutral	inactive
T83M	B-sheet1^{2}	SRS1'b	0.1	neutral	neutral
E168Q	α-helix D	undefined	0.0	neutral	decreased activity
F186L	loop btwn Helix D & E	undefined	-1.0	neutral	decreased activity
S212C	loop btwn helix E & F	undefined	-1.9	neutral	decreased activity
G299A	B-sheet 2	undefined	-5.2	damaging	neutral
D348N	α-helix J	undefined	0.8	neutral	substrate specific
R377Q	α-helix K	ERR triad/core stabilising motif	-1.1	neutral	inactive
I386F	loop btwn helix K & B-sheet 3^{1}	haem contact/SRS5	0.1	neutral	substrate specific
C406Y	B-sheet 3^{2}	undefined	0.0	neutral	substrate specific
R431W	3_{10} helix	ERR triad/meander	2.3	damaging	inactive
T438I	loop btwn 3_{10} helix & helix L	undefined	-0.8	neutral	neutral
R456H	loop btwn 3_{10} helix & helix L	Cys pocket/haem contact/CPR contact (charge cluster residue)	-1.5	neutral	inactive
Q478H	end of helix L	undefined	0.3	neutral	neutral
CYP2A6					
V110L	α-helix B'	SRS1	0.3	neutral	decreased activity
R128Q	α-helix C	haem contact/CPR contact	-0.9	neutral	inactive
L160H	α-helix D	undefined	-1.9	neutral	inactive
K194E	loop btwn helix E & F	undefined	-0.4	neutral	neutral
R203S	α-helix F	SRS(2,3)	-1.2	neutral	neutral
R203C	α-helix F	SRS(2,3)	-7.2	damaging	decreased activity
S224P	α-helix G'	SRS(2,3)	1.6	neutral	decreased activity
V365M	loop btwn helix K & B-sheet 3^{1}	SRS5	1.5	neutral	substrate specific
Y392F	B-sheet 3^{2}	undefined	1.3	neutral	substrate specific
E419D	loop btwn B-sheet 3^{2} & helix K'	undefined	0.2	neutral	neutral
N438Y	loop btwn 3_{10} helix & helix L	Cys-pocket/CPR contact	2.1	damaging	decreased activity
I471T	btwn B-sheet 5^{1} & 5^{2}	SRS6	-2.3	damaging	substrate specific
K476R	btwn B-sheet 5^{1} & 5^{2}	SRS6	1.3	neutral	decreased activity
G479V	btwn B-sheet 5^{1} & 5^{2}	SRS6	2.1	damaging	inactive
R485L	btwn B-sheet 5^{1} & 5^{2}	SRS6	1.3	neutral	neutral

(Table 1) Contd….

SNP	STRUCTURAL INFORMATION		PROTEIN STABILITY		Classification based on experimental data
	2° structure location	Functional region location	SDM Score kcal.mol^{-1}	Prediction	
CYP2B6					
R22C	Truncated N-terminal	halting signal	n/a	n/a	neutral
T26S	Truncated N-terminal	undefined	n/a	n/a	neutral
D28G	Truncated N-terminal	undefined	n/a	n/a	neutral
R29T	Truncated N-terminal	halting signal	n/a	n/a	neutral
M46V	loop btwn N-term & helix A	SRS1'a	-1.8	neutral	inactive
G99E	B-C loop	SRS1	-1.0	neutral	inactive
K139E	loop btwn helix C & D	undefined	0.6	neutral	inactive
R140Q	loop btwn helix C & D	undefined	0.4	neutral	decreased activity
T168I	start of helix E	undefined	3.4	damaging	decreased activity
Q172H	α-helix E	undefined	2.4	damaging	neutral
M198T	α-helix E	SRS(2, 3)	-2.7	damaging	substrate specific
S259R	loop btwn helix G & H	undefined	0.0	neutral	neutral
K262R	loop btwn helix G & H	undefined	-0.4	neutral	substrate specific
I328T	α-helix J	undefined	-3.9	damaging	inactive
R336C	loop btwn helix J & J'	undefined	-3.0	damaging	decreased activity
I391N	loop btwn B-sheet 3^2& helix K'	undefined	-4.6	damaging	inactive
P428T	loop btwn 3_{10} helix & helix L	haem contact	-1.2	neutral	inactive
M459V	B-sheet 5^1	undefined	-0.5	neutral	neutral
G476D	loop bwn B-sheet 5^1 & 5^2	SRS6	0.5	neutral	inactive
Q485L	loop bwn B-sheet 5^1 & 5^2	undefined	-1.7	neutral	neutral
R487C	Start of B-sheet 5^2	undefined	-3.6	damaging	neutral
CYP2C8					
R139K	loop btwn helix C & D	undefined	-1.4	neutral	neutral
G171S	α-helix E	undefined	1.5	neutral	neutral
R186G	loop btwn helix E & F	undefined	-4.6	damaging	inactive
I223M	F-G loop	SRS(2,3)	0.2	neutral	neutral
A238P	α-helix G	SRS(2,3)	-2.2	damaging	decreased activity
K247R	α-helix G	undefined	0.5	neutral	neutral
I264M	α-helix H	undefined	-0.4	neutral	decreased activity
I269F	α-helix H	undefined	0.2	neutral	decreased activity
K383N	loop btwn B-sheet 4^2 & 3^2	CPR contact	-1.1	neutral	neutral
K399R	loop btwn helix K' & 3_{10} helix	undefined	-0.4	neutral	decreased activity
P404A	loop btwn helix K' & 3_{10} helix	meander	1.2	neutral	decreased activity

(Table 1) Contd….

SNP	STRUCTURAL INFORMATION		PROTEIN STABILITY		Classification based on experimental data
	2° structure location	Functional region location	SDM Score kcal.mol^{-1}	Prediction	
CYP2C9					
L19I	truncated N-terminal	signal-anchor sequence	n/a	n/a	neutral
L90P	loop btwn helix B & C	CPR contact	-2.5	damaging	decreased activity
R125H	α-helix C	CPR contact/SRS1	-0.6	neutral	decreased activity
T130R	α-helix C	SRS1	-2.1	damaging	decreased activity
R132Q	loop btwn helix C & D	CPR contact (charge cluster residue)/SRS2	-0.2	neutral	decreased activity
R144C	α-helix D	undefined	-11.5	damaging	decreased activity
R150H	α-helix D	undefined	-0.6	neutral	increased activity
R150L	α-helix D	undefined	-0.2	neutral	neutral
Q214L	α-helix F' (F-G loop)	SRS(2,3)	3.2	damaging	decreased activity
H251R	α-helix G	undefined	-0.1	neutral	neutral
E272G	α-helix H	undefined	-4.0	damaging	neutral
P279T	loop btwn helix H & I	undefined	1.1	neutral	neutral
T299A	centre of α-helix I	catalytic groove/SRS4	1.9	neutral	decreased activity
R335Q	loop btwn helix J & J'	undefined	-0.3	neutral	neutral
R335W	loop btwn helix J & J'	undefined	-2.2	damaging	decreased activity
E354K	α-helix K	ERR triad / core stabilizing motif	-4.0	damaging	inactive
I359L	end α-helix K	SRS5	0.2	neutral	decreased activity
I359T	end α-helix K	SRS-5	-3.9	damaging	decreased activity
D360E	loop btwn helix K & B-sheet 3^1	SRS-5	-1.3	neutral	decreased activity
P382S	loop btwn B-sheet 4^2 & 3^2	SRS5	-0.9	neutral	neutral
D397A	loop btwn helix K' & 3_{10} helix	undefined	1.2	neutral	inactive
Q454H	α-helix L	undefined	1.4	neutral	neutral
A477T	loop btwn B-sheet 5^1 & 5^2	SRS-6	0.6	neutral	decreased activity
P489S	end of B-sheet 5^2	undefined	1.7	neutral	neutral
CYP2C19					
L17P	truncated N-terminal	signal-anchor sequence	n/a	n/a	neutral
I19L	truncated N-terminal	signal-anchor sequence	n/a	n/a	neutral
S51G	α-helix A	SRS1'a	-2.0	neutral	decreased activity
M74T	B-sheet 1^2	SRS1'b	-2.7	damaging	neutral
E92D	loop btwn helix B & B'	CPR contact/SRS1	0.0	neutral	substrate specific
W120R	α-helix C	haem contact/SRS1	-2.7	damaging	decreased activity
E122A	α-helix C	SRS1	1.4	neutral	substrate specific
R132Q	loop btwn helix C & D	CPR contact (charge cluster residue)/SRS1	0.4	neutral	inactive

(Table 1) Contd….

SNP	STRUCTURAL INFORMATION		PROTEIN STABILITY		Classification based on experimental data
	2° structure location	Functional region location	SDM Score kcal.mol^{-1}	Prediction	
R144H	α-helix D	undefined	0.2	neutral	decreased activity
R150H	α-helix D	undefined	-0.6	neutral	neutral
A161P	loop btwn helix D & E	undefined	-10.5	damaging	decreased activity
F168L	α-helix E	undefined	1.0	neutral	neutral
W212C	α-helix F'	SRS(2,3)	-2.1	damaging	decreased activity
P227L	F-G loop	SRS(2,3)	0.6	neutral	decreased activity
D256N	loop btwn helix G & H	undefined	-0.3	neutral	decreased activity
R329H	α-helix J	undefined	-0.6	neutral	neutral
I331V[1] (WT 1.B)	loop btwn helix J & J'	undefined	-0.4	neutral	neutral
D360N	loop btwn helix K & B-sheet 3[1]	SRS5	1.5	neutral	decreased activity
V394M	α-helix K'	undefined	-0.1	neutral	neutral
R410C	3 $_{10}$ helix	meander	-0.3	neutral	neutral
R433W	loop btwn 3 $_{10}$ helix & helix L	Cys-pocket/haem contact/CPR contact (charge cluster residue)	0.4	neutral	inactive
R442C	α-helix L	CPR contact (charge cluster residue)	-10.0	damaging	decreased activity
CYP2D6					
A90V	α-helix B	undefined	-0.2	neutral	neutral
T107I	α-helix B'	SRS1	4.2	damaging	substrate specific
E156A	α-helix D	undefined	0.42	neutral	decreased activity
R296C	α-helix I	SRS1	-8.8	damaging	neutral
I297L	α-helix I	SRS1	0.2	neutral	substrate specific
H324P	loop btwn helix H & I	undefined	-4.6	damaging	inactive
I369T	loop btwn helix K & B-sheet 3[1]	Haem contact/SRS5	-3.2	damaging	substrate specific
E410K	loop btwn α-helix K' & 3_{10} helix	undefined	0.4	neutral	substrate specific
R441C	loop btwn 3_{10} helix & helix L	Cys-pocket/haem contact/CPR contact (charge cluster residue)	-10.4	damaging	inactive
S486T	B-sheet 5[2]	SRS6	0.7	neutral	neutral
CYP2E1					
R76H	B-sheet 1[2]	SRS1'b	0.8	neutral	decreased activity
V179I	α-helix E	undefined	-0.1	neutral	neutral
V389I	B-sheet 3[2]	undefined	-0.3	neutral	neutral

(Table 1) Contd….

SNP	STRUCTURAL INFORMATION		PROTEIN STABILITY		Classification based on experimental data
	2° structure location	Functional region location	SDM Score kcal.mol^{-1}	Prediction	
CYP3A4					
G56D	btwn N-term & α-helix A	SRS1'a	-2.3	damaging	decreased activity
R130Q	α-helix D	haem contact/CPR contact/SRS1	-1.4	neutral	inactive
V170I	loop btwn helix D & E	undefined	-0.3	neutral	neutral
D174H	α-helix E	undefined	0.2	neutral	neutral
T185S	α-helix E	undefined	-1.4	neutral	decreased activity
F189S	loop btwn helix E & F	undefined	-5.2	damaging	decreased activity
S222P	α-helix F'	SRS(2,3)	-0.9	neutral	decreased activity
L293P	loop btwn helix H & I	undefined	-0.5	neutral	increased activity
T363M	α-helix K	core-stabilising motif	2.1	damaging	decreased activity
L373F	start of B-sheet 3^1	haem contact/SRS5	0.0	neutral	decreased activity
P416L	start of 3_{10} helix	meander	3.9	damaging	inactive
M445T	start of α-helix L	CPR contact	-2.6	damaging	neutral
P467S	loop btwn B-sheet 5^1 & 5^2	undefined	-1.4	neutral	neutral

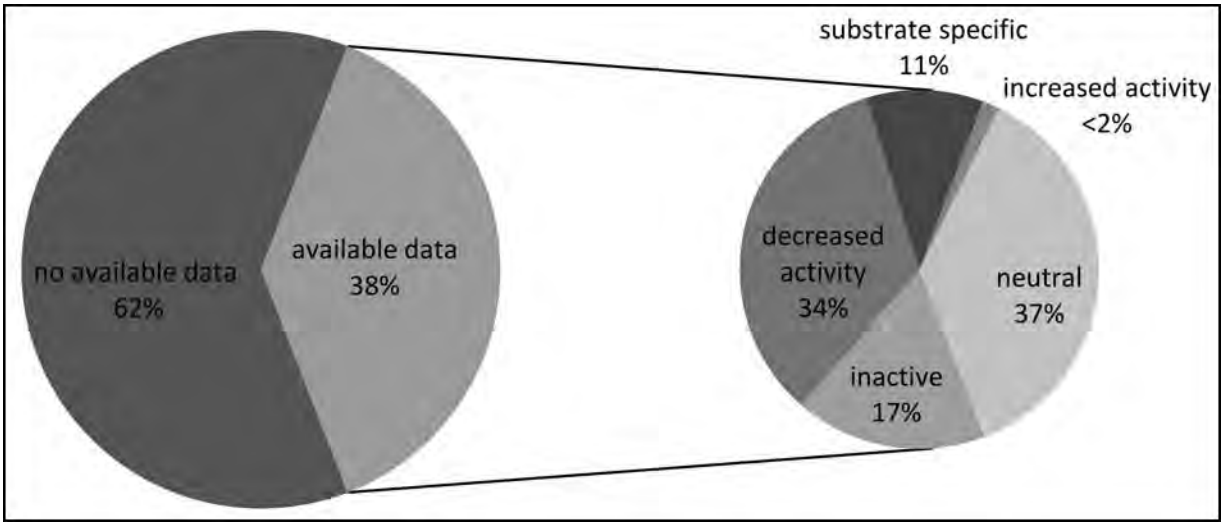


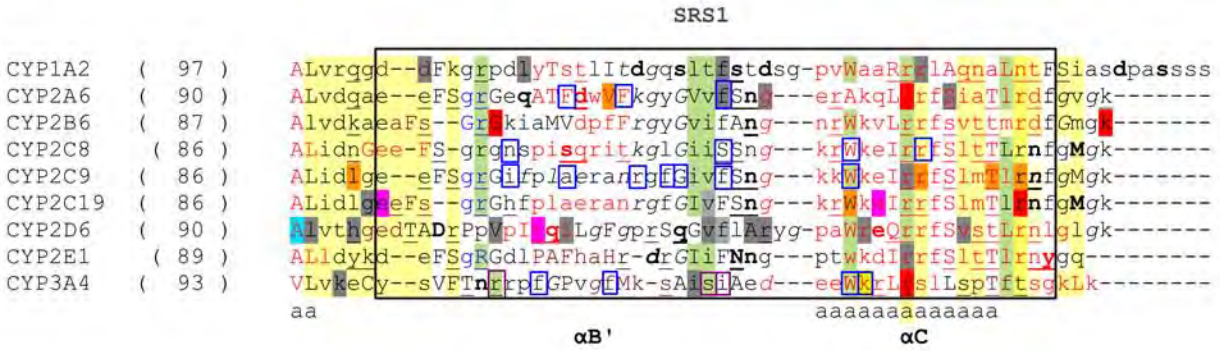
Fig. (2). SNPs categorised according to available experimental data.

The number of SNPs per isoform, classified into 6 categories on the basis of available experimental data, is shown in Table **2**. To date, CYP2B6 and CYP2C19 polymorphic variants have been the most thoroughly characterised, the largest contribution for CYP2B6 variant data coming from a study by Watanabe *et al*, who investigated the effect of 26 variant CYP2B6 alleles, expressed in COS-7 cells, on 7-ethoxy-4-trifluoromethylcoumarin (7-EFC) and selegiline metabolism [68]. There is *in vitro* data for 24 out of the 39 CYP2C19 SNPs and 20 of these were tested by Wang *et al* who reported expression levels, P420:P450 holoprotein levels and kinetic data for 3 different substrates and 10 inhibitors using high through-put activity assays [43, 69]. The functional data available for individual CYP2D6 SNPs is the most incomplete and as a result, we were able to classify only 12% of the SNPs found in CYP2D6. CYP2D6 shows the highest levels of polymorphic variation and although there is a large amount of clinical data available, as well as data from several *in vitro* studies, most CYP2D6 alleles contain multiple SNPs in the coding and/or non-coding regions of the gene, which makes it difficult to determine the contribution of individual mutations to the observed phenotype.

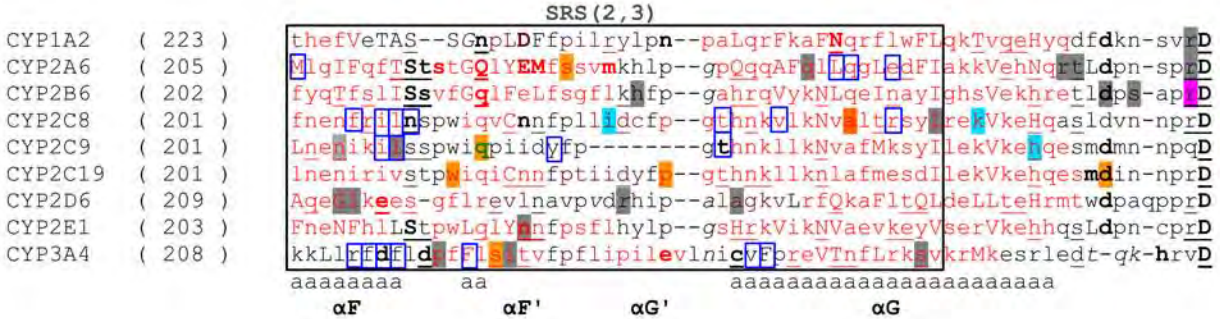
The CYP450 SNP Map

A multiple structural alignment, including a representative structure for each of the nine isoforms, is shown in (Fig. **3**). This alignment has been annotated with structural features using the program JOY [60]. Secondary structure assignment and regions important for the enzyme's structural integrity and function are indicated as described in Methods. The positions of SNPs are


SRS1'a
SRS1'b
CYP1A2 (34)
CYP2A6 (30)
CYP2B6 (28)
CYP2C8 (28)
CYP2C9 (26)
CYP2C19 (28)
CYP2D6 (34)
CYP2E1 (31)
CYP3A4 (28)
αA
β1¹
β1²
αB
proline rich motif


SRS1
CYP1A2 (97)
CYP2A6 (90)
CYP2B6 (87)
CYP2C8 (86)
CYP2C9 (86)
CYP2C19 (86)
CYP2D6 (90)
CYP2E1 (89)
CYP3A4 (93)
αB'
αC


CYP1A2 (159)
CYP2A6 (143)
CYP2B6 (140)
CYP2C8 (139)
CYP2C9 (139)
CYP2C19 (139)
CYP2D6 (147)
CYP2E1 (142)
CYP3A4 (144)
αD
αE
αF


SRS(2,3)
CYP1A2 (223)
CYP2A6 (205)
CYP2B6 (202)
CYP2C8 (201)
CYP2C9 (201)
CYP2C19 (201)
CYP2D6 (209)
CYP2E1 (203)
CYP3A4 (208)
αF
αF'
αG'
αG

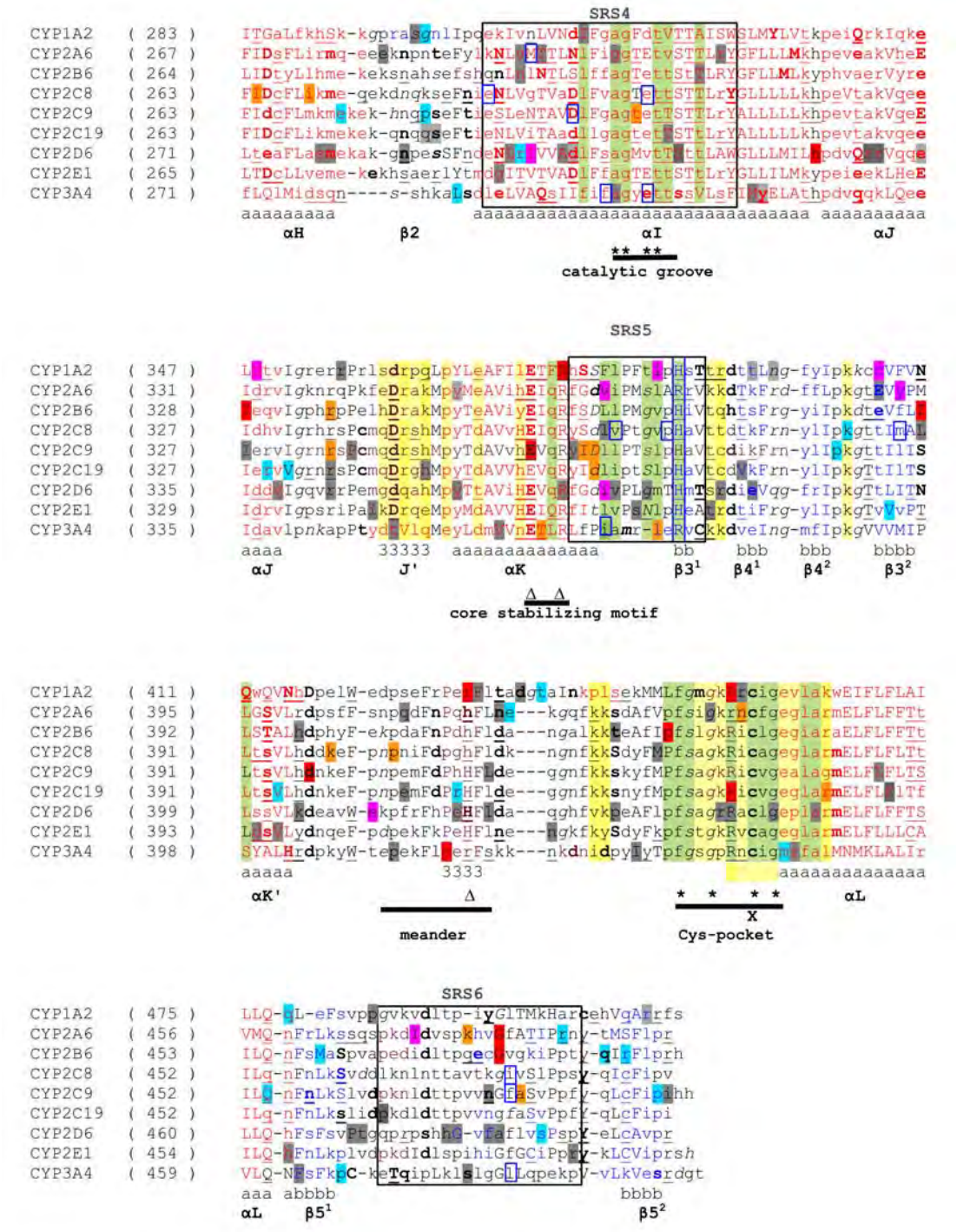


Fig. (3). The CYP450 SNP map consists of a multiple structural alignment, including a representative structure for each of the nine drug metabolising CYP450 isoforms.

The alignment is annotated with the following structural features according to the JOY format: Solvent inaccessible/buried residues (upper case), Solvent accessible (lower case), Positive Φ (italics), Hydrogen bond to main chain amide (bold), Hydrogen bond to main chain carbonyl (underlined), α-helix (red), β-sheet (blue), 3^{10}-helix (dark red).

Secondary structures are named according to the nomenclature of Poulos *et al* [5]. The positions of conserved motifs are also indicated below the alignment.

Functional regions as described in materials and methods are indicated as follows: SRS regions (boxed in grey), haem contacts (shaded in green), CPR contacts (shaded in yellow- light yellow for residues within 6Å of interface and bright yellow for charge cluster residues), gating residues (boxed in blue) and aqueduct gating residues (boxed in dashed-blue).

The positions of known SNPs are highlighted according to their classifications based on experimental data available in the literature. Inactive (red), decreased activity (orange), substrate specific (magenta), increased activity (bright green), neutral (blue) and unknown (grey). If more than one SNP can occur at the same residue position then the position is highlighted according to the SNP that shows the most drastic effect on enzyme function.

Table 2. Summary of Experimental Data for Drug Metabolising CYP450 Isoforms

Experimental Data	CYP1A2	CYP2A6	CYP2B6	CYP2C8	CYP2C9	CYP2C19	CYP2D6	CYP2E1	CYP3A4	Total
inactive	4	3	7	1	2	2	2	0	2	23
decreased activity	3	5	3	5	12	9	1	1	6	45
substrate specific	3	3	2	0	0	2	4	0	0	14
increased activity	0	0	0	0	1	0	0	0	1	2
neutral	4	4	9	5	9	9	3	2	4	49
unknown	24	26	10	8	21	17	71	15	23	215
Total SNPs	38	41	31	19	45	39	81	18	36	348

highlighted according to their assignment based on available experimental data. The alignment does not give information on the mutant residue type but this information can be found in the Tables S1-S9 in the Supplementary Materials. If more than one SNP occurs at the same position, the wild-type residue is highlighted according to the mutant residue type that has the greatest known effect on enzyme function.

Approximately half the SNPs analysed fell within ordered secondary structure regions (helices A-L or β-sheets 1-5), while the other half occurred in loop/disordered regions. Ordered helices and β-strands make up about 50% of the CYP450 structure, indicating that the SNPs are evenly spread across the protein structure (Table S10 in the Supplementary Materials).

Effect of SNPs on Protein Stability

SDM predicted 28% of the SNPs analysed to be damaging mutations that are likely to result in a functionally impaired enzyme (Fig. **4**). The majority of these mutations are predicted as damaging due to destabilising effects (-ΔΔG scores) however 6% of the SNPs are predicted as damaging due to a stabilising effect (+ΔΔG scores); the literature rationalisation on positive ΔΔG scores as damaging is based on the premise that mutations that are highly stabilising may reduce conformational flexibility that is important for function [70-73]. 64% of the SNPs were predicted to be neutral by SDM. SDM could not be used to predict the effect of the remaining 8% of the SNPs on stability as these mutations lie within the N-terminal domain, which is truncated or disordered in crystal structures. Here we have considered such mutations as being neutral since they are unlikely to affect the stability of the catalytic domain and since *in vitro* studies have shown that the N-terminal membrane anchor region is not important for enzyme activity [74]. A detailed summary of the SDM data obtained for SNPs found in each isoform can be found in Table S11 in the Supplementary Material.

Correlation Between Stability Predictions and Experimental Data

The correlation between SDM results and experimental data is shown in Figure **5**, from which it can be seen that of the 38 CYP450 SNPs for which there is experimental data and which were predicted to be damaging by SDM, 31 showed altered activity in the experimental data, giving a positive predictor value (PPV) of 82% for SDM here. Conversely, of the 84 SNPs in our dataset that showed altered activity experimentally, SDM predicted that 31 would be damaging, meaning that used on its own, SDM has a sensitivity (i.e. the percentage of correctly classified true positives) of 37% in analysis of CYP450 SNPs. Finally, of the 42 SNPs classified as neutral based on experimental data and which could be analysed by SDM, 35 were predicted to have no significant effect on protein stability, meaning that used on its own, SDM has a specificity (i.e. the percentage of correctly classified true negatives) of 83% in analysis of CYP450 SNPs.

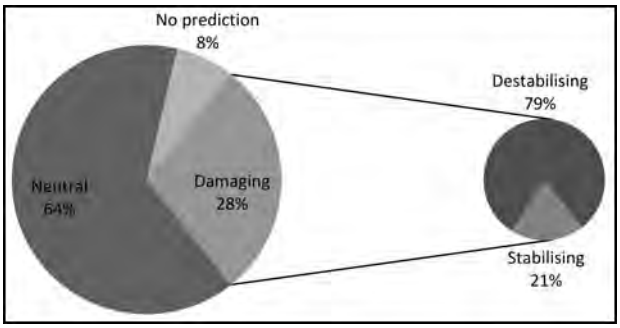


Fig. (4). Proportion of SNPs predicted to be damaging to protein structure

Effect of SNPs on Function

According to our analysis, 54% of the CYP450 sequence space forms part of one or more functional regions (Fig. **6**): The seven conserved motifs cover ~13% of the CYP450 sequence space; just over 5% of CYP450 residues form haem contacts; ~8% form the CPR interface; & SRS's are the largest functional regions, covering one third of the sequence space. Gating residues have been identified only in four of the nine isoforms so they currently only make up an average of 0.5% of the CYP450 sequence.

Figure **6** also shows the percentage of SNPs falling within each functional region. Based on the percentage sequence space covered by each functional region, SNPs are evenly spread throughout all functional regions, including the conserved motifs, with 52% of SNPs falling within one or more defined regions and the other 48% falling within undefined functional regions. A breakdown of the number of SNPs falling within each functional region for each isoform is given in Table S12 in the Supplementary Material.

Correlation Between Functional Region Analysis and Experimental Data

The correlation between the functional region location of SNPs and experimental data is shown in Figure **7**, from which it can be seen that of the 64 CYP450 SNPs for which there is experimental data and which fall in functional regions important for catalytic activity (and which are therefore potentially damaging), 53 showed altered activity in the experimental data, giving a positive predictor value of 83% for this functional region analysis. Conversely, of the 84 SNPs in our dataset that showed altered activity experimentally, 53 fall in functional regions important for catalytic activity (and are therefore potentially damaging), giving a sensitivity of 63%. Finally, of the 49 SNPs classified as neutral based on experimental data, 38 were found to fall within undefined regions or regions that are not important for catalytic activity (and are therefore less likely to be damaging), giving a specificity of 78%.

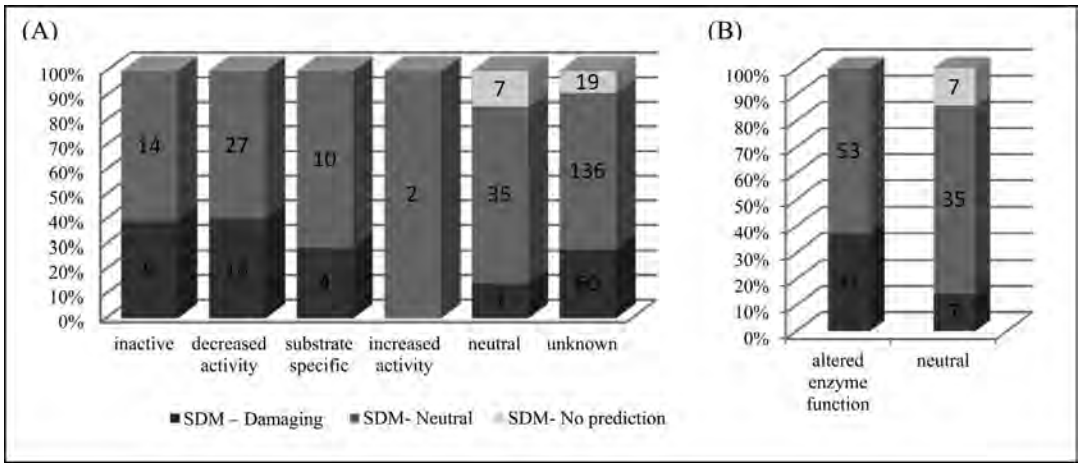


Fig. (5). Correlation between SDM stability predictions and experimental data.

The y-axis shows the percentages of SNPs predicted to be damaging to protein structure (dark grey columns), those predicted to be neutral (medium grey columns) and those that fell within truncated regions of the structure that could not be analysed by SDM (light grey columns) for (**A**) each experimentally classified group and (**B**) for a combination of all groups showing altered enzyme function vs. those classified as neutral, as shown along the x-axis. The numbers of SNPs accounting for these percentages are shown within the columns.

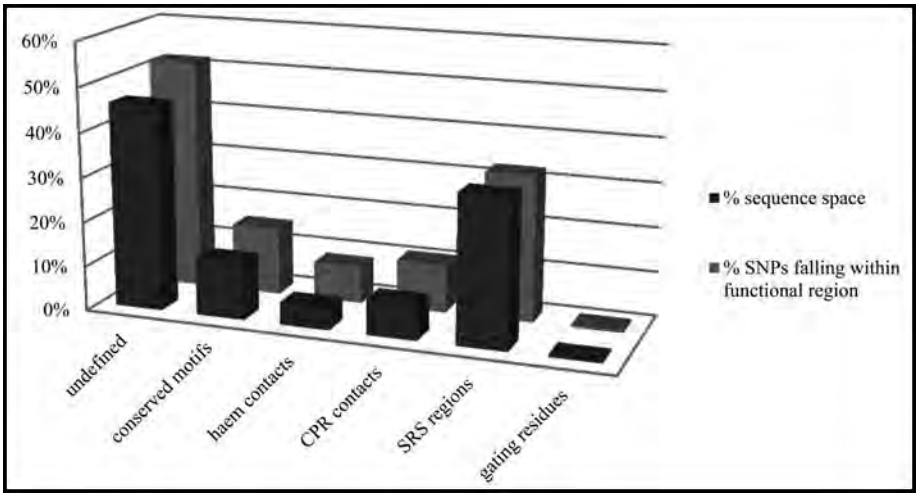


Fig. (6). Percentage of SNPs falling within each functional region and the percentage sequence space covered by each functional region, illustrating that SNPs are spread evenly across functional regions.

Table **3** shows the number of SNPs occurring in each functional region as classified according to experimental data. Where a SNP occurs in more than one region, it was assigned to the functional region that corresponds best with the observed experimental data. For example if a mutation occurs at the position of a haem contact and it falls within a SRS region but the experimental data shows that no holoprotein could be detected for this variant then this SNP is assigned to the haem contact region instead of the SRS region. If it is unclear which functional region best fits the data then the SNP is assigned to both regions thus explaining the differences in the absolute numbers in Figure **7** and Table **3**. The numbers in brackets indicates the number of these SNPs predicted to be damaging by SDM.

Five SNPs within the N-terminal signal anchor sequence and halting signal have been functionally characterised *in vitro* and all of them were classified as neutral according to experimental data. N-terminal regions play a role in membrane binding but mutations in these regions are not expected to have an effect on protein folding or catalytic activity. The majority of the other neutral SNPs (67%) occurred within undefined regions, while 16% occurred within SRS regions.

Of the 13 mutations falling within the other conserved motifs - namely the proline rich motif, the I-helix catalytic groove, the Cys-pocket, the core stabilising motif and the meander region – 12 resulted in an inactive protein or a protein with decreased activity. Three of these inactive SNPs fell within the ERR triad.

Approximately 82% of SNPs falling within SRS regions had some effect on protein function with the majority causing decreased activity. Experimental data showed that 31 SNPs falling within undefined functional regions had an effect on enzyme activity, of which 10 were predicted to be damaging by SDM.

The correlation between the experimental data and the combined SDM stability- and functional analysis predictions is shown in Figure **8**, from which it can be seen that of the 78 CYP450 SNPs for which there is experimental data and which fall in functional regions important for catalytic activity and/or are predicted to be damaging by SDM, 63 showed altered activity in the experimental data, giving a positive predictor value of 81% for this combined analysis. Conversely, of the 84 SNPs in our dataset that showed altered activity experimentally, 63 fall in functional regions important for catalytic activity and/or are predicted to be damaging by SDM, giving a sensitivity of 75% for this combined analysis. Fi-

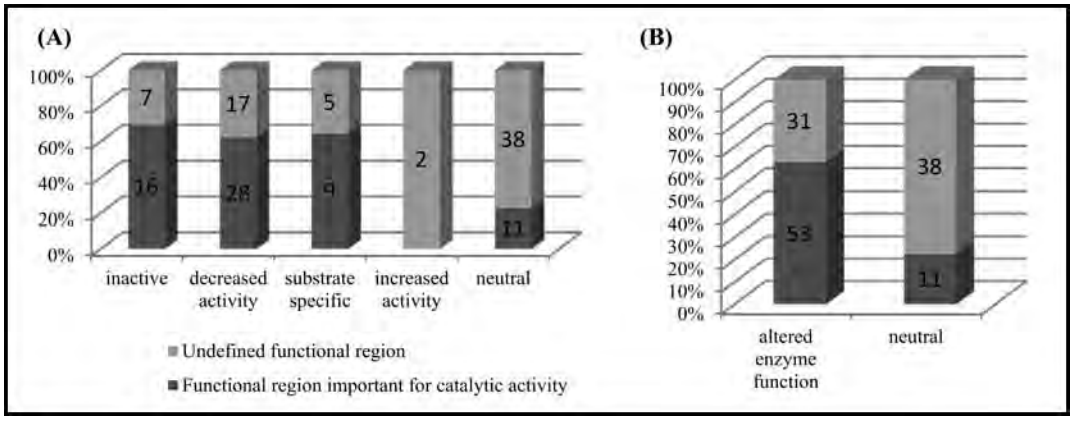


Fig. (7). Correlation between functional analysis and experimental data.

The y-axis shows the percentages of SNPs falling within functional regions important for catalytic activity (dark grey) vs. those falling within undefined functional regions or regions that do not affect catalytic activity (light grey) for (**A**) each experimentally classified group and for (**B**) a combination of all groups showing altered enzyme function vs. those classified as neutral, as shown along the x-axis. The numbers of SNPs accounting for these percentages are shown within the columns. Note that functional regions important for catalytic activity include all defined regions excluding the N-terminal signal- anchor motif and halting signal.

Table 3. Correlation between Experimental Data and Functional Region Localisation of SNPs

Functional Region	Inactive	Decreased activity	Substrate specific	Increased activity	Neutral
signal anchor sequence	0 (0)	0 (0)	0 (0)	0 (0)	3 (n/a)
halting signal	0 (0)	0 (0)	0 (0)	0 (0)	2 (n/a)
proline rich motif	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)
I-helix catalytic groove	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)
cys-pocket	3 (1)	1 (1)	0 (0)	0 (0)	0 (0)
core stabilizing motif	2 (1)	1 (1)	0 (0)	0 (0)	0 (0)
meander	2 (2)	1 (0)	0 (0)	0 (0)	1 (0)
ERR-Triad	3 (2)	0 (0)	0 (0)	0 (0)	0 (0)
haem contact	6 (1)	2 (1)	0 (0)	0 (0)	0 (0)
CPR contact	1 (0)	5 (3)	0 (0)	0 (0)	2 (1)
SRS	5 (1)	23 (8)	9 (4)	0 (0)	8 (2)
gating residue	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
undefined	7 (4)	17 (6)	5 (0)	2 (0)	33 (4)

nally, of the 49 SNPs classified as neutral based on experimental data, 34 were found to fall within undefined regions or regions that are not important for catalytic activity and are predicted to have no effect on stability, giving a specificity of 69% for this combined analysis.

DISCUSSION

Previous functional studies in our group have delineated regions of the cytochrome P450 structure important for substrate recognition [29], substrate and product access and egress from the active site [25] and interaction with the cytochrome P450 reductase [37]. Here we have combined the information from those studies with new *in silico* calculations on the effect of polymorphic variations in the major human drug metabolising CYP450 enzymes on protein stability and have compared our results to experimental data in order to establish the likely causes of altered drug metabolism observed for cytochrome P450 variants in functional assays to date. In the process, we have created a cytochrome P450 polymorphic variant map (Fig. **3**), which gives an overview of the location of known CYP450 polymorphic variation in relation to important structural and functional regions, whilst detailed SNP tables (Tables S1-S9) provide information on the structural environments of each variant residue, the predicted effect of the mutation on protein stability and a summary of the available *in vitro* data for each SNP.

A number of *in silico* tools are available that aim to distinguish between disease associated mutations and benign mutations based on the residues position relative to functional sites, protein stability calculations or a combination of both. Using the tools PolyPhen (http://genetics.bwh.harvard.edu/pph) and SIFT (http://blocks.fhcrc.org/sift/SIFT.html), Wang *et al* previously predicted that 39-43% of the human CYP450 SNPs they analysed would have an impact on protein function [75] and, correlation with known experimental data suggested a PPV of 70%. Other CYP450 studies have used a combination of docking and molecular dynamics to assess changes to

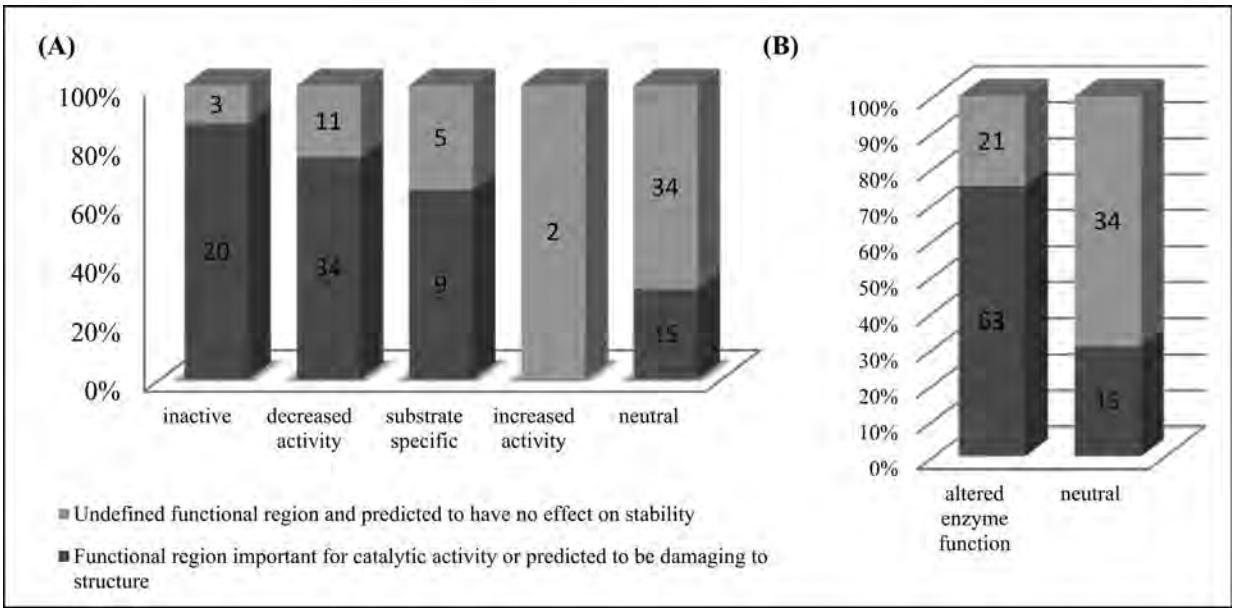


Fig. (8). Correlation between experimental data and the combined SDM stability- and functional analysis predictions.
The y-axis shows the percentage of SNPs found in functional regions important for catalytic activity or predicted by SDM to be damaging to the structure (dark grey) vs. those found in undefined functional region and predicted to have no effect on stability (light grey) (**A**) for each experimentally classified group and (**B**) for combination of all groups showing altered enzyme function vs. those classified as neutral. The numbers of SNPs accounting for these percentages are shown within the columns. Note that functional regions important for catalytic activity include all defined regions excluding the N-terminal signal- anchor motif and halting signal.

the binding pocket and ligand interactions for small sets of SNPs [76-78]. However, whilst tools such as PolyPhen take the positions of some functional sites into account, quantitatively predicting the effect of mutations on ligand and co-factor binding is still a huge challenge for computational biology as accurate calculations are very computationally intensive, limiting their use for large interactions surfaces.

In the present study we aimed to break down the analysis of possible functional effects of SNPs into different components, enabling us to distinguish between mutations affecting stability and those lying within functional regions of the enzyme; in turn this enabling us to more precisely predict the molecular mechanisms underlying changes in enzyme activity caused by different CYP450 SNPs. There are a number of *in silico* tools which aim to predict the effect of single point mutations on protein stability, based on methods such as physical effective energy functions [79-81], empirical potential energy functions [82, 83], statistical potential energy functions [58, 84, 85] and machine learning [86-89]. Of these, physical potential energy functions are likely to be the most accurate but are generally computationally expensive and time consuming whilst machine learning methods and empirical potential energy functions use experimental data to fit their function and generally suffer from over fitting the function to the training data set. Here, we therefore chose to use Site Directed Mutator (SDM) as a representative tool to predict the effect of mutations on the stability of the apoprotein, enabling us to distinguish between mutations affecting the inherent stability of the protein fold as appose to those at functional sites important for haem or substrate binding. SDM - which is a statistical potential energy function - provides a quick high-throughput method for calculating the effect of single point mutations on protein stability and performs comparably or better to a number of other published methods [90]. SDM is not based on prior knowledge of the mutant's thermodynamic measurements and has been found to demonstrate better sensitivity in predicting stabilizing mutations than other methods [57]. Importantly for our purposes, SDM focuses solely on predicting effects of mutations on protein stability ($\Delta\Delta G$), ignoring effects on functional regions such a ligand contacts. Interestingly, we found that when SDM was used in isolation, we observed a PPV of 82%, apparently immediately exceeding the performance of PolyPhen or SIFT for CYP450 enzymes.

By way of illustration of our approach, CYP2C19*5, corresponding to the single amino acid substitution R433W, has been identified as a defective allele that contributes to the s-mephenytoin poor metaboliser phenotype in Caucasian and Oriental populations [91]. *In vitro* functional assays have shown that expression levels in yeast are similar to the wild-type but that P450 holoprotein is not detected [43]. Figure **9** shows a model of the CYP2C19 polymorphic variant R433W aligned to the wild-type structure. Arg^{433} is a buried residue with a relative solvent accessibility of 5.3%. When this residue is mutated to a tryptophan the side chain volume increases by 54.4 $Å^3$ while the relative solvent accessibility changes from 2% to 17%. The $\Delta\Delta G$ SDM score for this mutation is 0.4 $kcal.mol^{-1}$, which indicates that this mutation is neutral and does not have a significant effect on the stability of the apoprotein. However, Arg^{433} is part of the Cys-pocket or β-bulge, which is a highly conserved region important for haem binding. Based on the protein-ligand interaction data for CYP2C19, Arg^{433} is a haem contact residue that interacts through its 5-guanidino group with the haem propionate via a salt bridge. Mutating this residue to a tryptophan disrupts this interaction and the increase in volume and the position of the side chain may plausibly result in a steric clash between tryptophan and the haem group, providing a possible explanation for the lack of holoprotein observed *in vitro*.

Reliable *in silico* prediction of the effect of SNP variants on CYP450 protein function remains a challenging endeavour today. In the present study, using a representative, computationally-light *in silico* tool in isolation to predict the effects of individual SNPs on the stability of the CYP450 apoproteins, we were able to demonstrate an encouraging specificity of 83% and a positive predictor value of 82% for our predictions when compared to available experimental data, but a sensitivity of only 37% for CYP450 SNPs (Fig. **5**, Table **4**), suggesting that altered CYP450 activity is only partially explained by changes in protein stability. This is in agree-

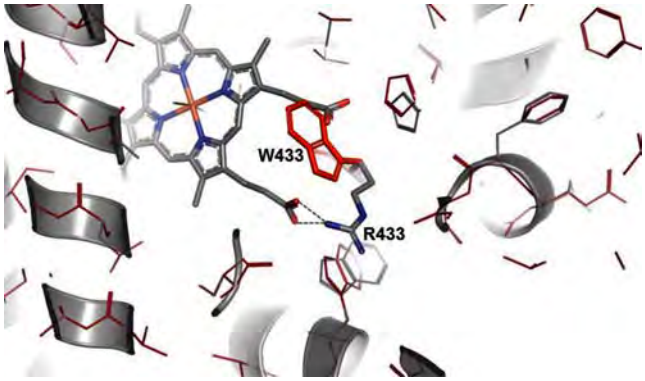


Fig. (9). CYP2C19 R433W model aligned to the wild-type structure.
The CYP2C19 R433W model (red) generated by ANDANTE is aligned to the wild type CYP2C19 pdb structure 4GQS (grey). Wild-type residue Arg 433 is shown interacting with the haem group.

ment with studies by Blundell *et al* who have shown that although changes in protein stability can be relatively accurately predicted *in silico*, these predictions alone are not sufficient to enable accurate large-scale classification of benign and disruptive single amino acid substitutions [57, 90, 92]. However, by combining the SDM predictions on effect of each SNP on protein stability with an analysis of whether or not the relevant SNP fell within a known functional region important for catalysis, we were able to significantly improve the sensitivity of our predictions to 75%, whilst still retaining very good positive predictor and specificity values (Table **4**). In more detail, our combined 'SDM plus functional region analysis' approach enabled us to account for 87% of the SNPs resulting in very low or inactive enzyme activity; 76% of the SNPs resulting in decreased activity and 64% of the SNPs resulting in substrate specific affects *in vitro* (Fig. **8**).

Table 4. Summary performance of SDM and functional region analysis on CYP450 SNPs

	Sensitivity (%)	Specificity (%)	PPV (%)
SDM alone	37	83	82
Functional region analysis alone	63	78	83
SDM plus functional region analysis	75	69	81

Although we have annotated our alignment in Figure **3** with the known CYP450 functional regions, CYP450 function is by no means completely understood. Information about residues important for conformational change, oligomerisation and interactions with effector molecules is still limited. Investigating further mutations that show altered *in vitro* activity but do not affect protein stability and lie outside known functional regions may shed light on hinge regions and peripheral or allosteric binding sites important for conformational change and substrate access in CYP450s. In this context, it is interesting that our analyses reveal that 5 out of the 14 SNPs that show altered substrate specific *in vitro* activity, together with both SNPs that showed increased *in vitro* activity, lie within currently undefined functional regions of CYP450 structure (Table **3**). For example, CYP2C19 variant D256N shows similar expression and holoprotein levels to the wild type enzyme but shows substrate specific changes in K_m and reduced V_{max} [93]; Asp^{256} lies in the loop region between helices G and H (Fig. **10**). This G/H loop has shown high conformation variability in MD studies on mammalian CYP450 structures [18, 94] and we therefore speculate that Asp^{256} may form part of a hinge region important for conformational changes within the enzyme.

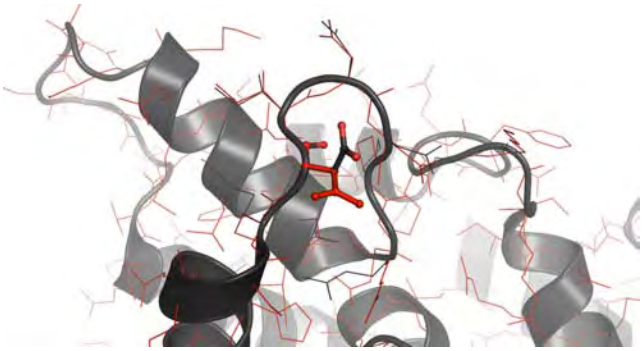


Fig. (10). CYP2C19 D256N model aligned to the wild-type structure.
The CYP2C19 D256N model (red) generated by ANDANTE is aligned to the wild type CYP2C19 pdb structure 4GQS (grey). Asp^{256} and Asn^{256} are shown as sticks.

A significant limitation in our analysis on damaging SNPs is the quality, availability and variability of experimental data found in the literature. In particular, heterogonous expression system, total protein quantification, methods for generating kinetic data, as well as kinetic models for data analysis, vary between studies and can all profoundly influence results, as recently reviewed by Hiratsuka [95]. For example, most CYP450 substrate turnover is still analysed today using Michaelis-Menten kinetics, yet it is well documented that some CYP450 isoforms such as CYP3A4 show atypical kinetics due to multiple binding sites, substrate inhibition, allosteric regulation or multiple active forms of the enzyme. Depending on the model to which the experimental data was fitted, substantially different kinetic parameters might be reported for any given system, complicating quantitative comparisons between different studies on mammalian CYP450 SNPs.

The sum of the defined functional regions drawn together in our CYP450 SNP Map make up over half of the protein, whilst the substrate recognition regions cover ~33% of the sequence space, reflecting the broad substrate specificities of mammalian CYP450s. It is important to recognise therefore that not all mutations in these functional regions will affect enzyme activity for all possible substrates. It is also important to note that certain residues within SRS regions have side chains facing away from the active site (see for example Fig. **5** in reference [29]); some of these residues make contact with the substrate *via* main chain interactions rather than side chain interactions and mutation of such residues might directly affect ligand interactions (and hence enzyme function) as a result either of altered positioning of main chain atoms relative to the bound ligand within the active site, or alternatively by directly affecting protein stability. In other instances, such residues with side chains facing away from the active site do not obviously make main chain contacts to the ligand; here mutational effects on enzyme function are more likely to be manifested through either direct effects on protein stability or through co-operative conformational changes that indirectly perturb the active site. By way of example of this latter type, CYP2A6 variant R203S, lying within the F helix and SRS(2,3), shows similar activity to the wild type enzyme in *in vitro* studies; the side chain of Arg^{203} faces away from the active site (Fig. **11**) and SDM predicts that mutation of Arginine 203 to Serine will not be damaging, consistent with experimental data. By contrast, variant R203C does show reduced activity *in vitro;* this mutation is predicted to be highly destabilising by SDM and is likely to affect substrate binding by disrupting the F helix. It could

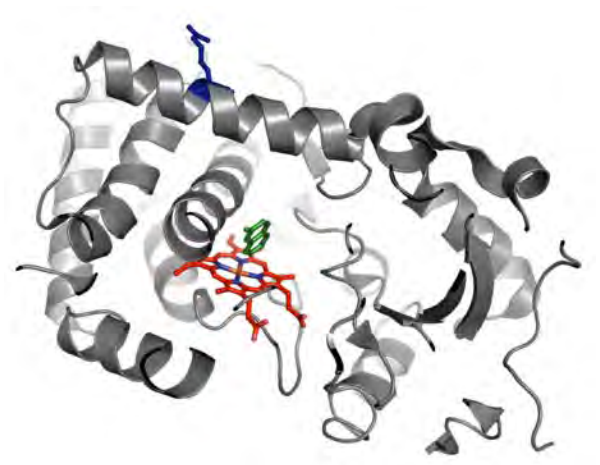


Fig. (11). CYP2A6 model showing the position of R203.
CYP2A6, PDB structure 1Z10 chain A, is shown in grey, the haem group in red and a coumarin substrate molecule bound within the active site shown in green. R203 is shown in blue.

be argued that such residues that make neither side chain nor main chain contacts to the ligand in known CYP450 structures should be excluded from the annotated functional regions in our CYP450 SNP Map. However, we have chosen to leave these residues annotated as part of the substrate recognition site regions here for consistency with the existing literature on CYP450s [29, 30]; we also note that as further CYP450 structures are determined, the precise positioning of equivalent residues relative to bound ligands may change, again arguing for their retention as part of the SRS regions at this point.

Inclusion of SDM data in our CYP450 SNP Map has proved useful in extending predictions of the likely consequence of mutation on CYP450 function, including those variants that fall within known functional regions. Importantly, most SNPs analysed in this study have only been tested to date for one or a small subset of substrates so it is plausible that more SNPs lying within SRS regions will show altered metabolism or substrate specific effects when tested on a larger panel of drugs. Our study therefore provides a useful rationalisation of CYP450 structure-function relationships by combining both *in silico* and experimental data into a unified CYP450 SNP Map that has surprisingly good predictive capability (judged by the sensitivity and specificity data presented here) and which should therefore find utility in predicting the likely effects of uncharacterised CYP450 variants on drug metabolism.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

L.A. thanks the National Research Foundation of South Africa (NRF), the Harry Crossley Foundation and the Commonwealth Scholarship Commission for PhD bursaries. J.B. thanks the NRF for a research chair and for funding for the research. T.L.B. thanks the BBSRC and Cambridge University for funding.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

REFERENCES

[1] Danielson, P. B. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metab.*, **2002**, *3*, 561-97.

[2] Bertz, R. J.; Granneman, G. R. Use of *in vitro* and *in vivo* data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin. Pharmacokinet.*, **1997**, *32*, 210-58.

[3] Evans, W. E. Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics. *Science*, **1999**, *286*, 487-491.

[4] Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C. The 2.6-Å crystal structure of Pseudomonas putida cytochrome P-450. *J. Biol. Chem.*, **1985**, *260*, 16122-16130.

[5] Poulos, T. L.; Finzel, B. C.; Howard, A. J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.*, **1987**, *195*, 687-700.

[6] Hasemann, C. A.; Kurumbail, R. G.; Boddupalli, S. S.; Peterson, J. A.; Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure*, **1995**, *3*, 41-62.

[7] Raag, R.; Martinis, S. A.; Sligar, S. G.; Poulos, T. L. Crystal structure of the cytochrome P-450CAM active site mutant Thr252Ala. *Biochemistry*, **1991**, *30*, 11420-11429.

[8] Gerber, N. C.; Sligar, S. G. Catalytic mechanism of cytochrome P-450: evidence for a distal charge relay. *J. Am. Chem. Soc.*, **1992**, *114*, 8742-8743.

[9] Johnson, E. F.; Stout, C. D. Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun.*, **2005**, *338*, 331-336.

[10] Otyepka, M.; Skopalík, J.; Anzenbacherová, E.; Anzenbacher, P. What common structural features and variations of mammalian P450s are known to date? *Biochim. Biophys. Acta*, **2007**, *1770*, 376-389.

[11] Pochapsky, T. C.; Kazanis, S.; Dang, M. Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxid. Redox Signaling*, **2010**, *13*, 1273-1296.

[12] Wade, R. C.; Winn, P. J.; Schlichting, I.; Sudarko. A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.*, **2004**, *98*, 1175-1182.

[13] Ross Wilderman, P.; R Halpert, J. Plasticity of CYP2B Enzymes: Structural and Solution Biophysical Methods. *Curr. Drug Metab.*, **2012**, *13*, 167-176.

[14] Lüdemann, S. K.; Carugo, O.; Wade, R. C. Substrate Access to Cytochrome P450cam: A Comparison of a Thermal Motion Pathway Analysis with Molecular Dynamics Simulation Data. *J. Mol. Model.*, **1997**, *3*, 369-374.

[15] Lüdemann, S. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J. Mol. Biol.*, **2000**, *303*, 813-830.

[16] Winn, P. J.; Lüdemann, S. K.; Gauges, R.; Lounnas, V.; Wade, R. C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc. Natl. Acad. Sci. U. S. A.*, **2002**, *99*, 5361-5366.

[17] Schleinkofer, K.; Winn, P. J.; Lüdemann, S. K.; Wade, R. C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep.*, **2005**, *6*, 584-589.

[18] Skopalík, 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*, 8165-8173.

[19] Fishelovitch, D.; Shaik, S.; Wolfson, H. J.; Nussinov, R. Theoretical Characterization of Substrate Access/Exit Channels in the Human Cytochrome P450 3A4 Enzyme: Involvement of Phenylalanine Residues in the Gating Mechanism. *J. Phys. Chem. B.*, **2009**, *113*, 13018-13025.

[20] Denisov, I. G.; Shih, A. Y.; Sligar, S. G. Structural differences between soluble and membrane bound cytochrome P450s. *J. Inorg. Biochem.*, **2012**, *108*, 150-158.

[21] Hendrychova, T.; Berka, K.; Navratilova, V.; Anzenbacher, P.; Otyepka, M. Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. *Curr. Drug Metab.*, **2012**, *13*, 177.

[22] Oostenbrink, C.; de Ruiter, A.; Hritz, J.; Vermeulen, N. Malleability and versatility of cytochrome P450 active sites studied by molecular simulations. *Curr. Drug Metab.*, **2012**, *13*, 190.

[23] Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, **2006**, *7*.

[24] Cojocaru, V.; Winn, P. J.; Wade, R. C. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta*, **2007**, *1770*, 390-401.

[25] Zawaira, A.; Coulson, L.; Gallotta, M.; Karimanzira, O.; Blackburn, J. On the deduction and analysis of singlet and two-state gating-models from the static structures of mammalian CYP450. *J. Struct. Biol.*, **2011**, *173*, 282-293.

[26] Yaffe, E.; Fishelovitch, D.; Wolfson, H. J.; Halperin, D.; Nussinov, R. MolAxis: efficient and accurate identification of channels in macromolecules. *Proteins*, **2008**, *73*, 72-86.

[27] Petrek, M.; Kosinova, P.; Koca, J.; Otyepka, M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure*, **2007**, *15*, 1357-1363.

[28] Berka, K.; Hanák, O.; Sehnal, D.; Banáš, P.; Navrátilová, V.; Jaiswal, D.; Ionescu, C.; Vařeková, R. S.; Koča, J.; Otyepka, M. MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.*, **2012**, *40*, W222-W227.

[29] Zawaira, A.; Ching, L. Y.; Coulson, L.; Blackburn, J.; Wei, Y. C. An Expanded, Unified Substrate Recognition Site Map for Mammalian Cytochrome P450s: Analysis of Molecular Interactions Between 15 Mammalian CYP450 Isoforms and 868 Substrates. *Curr. Drug Metab.*, **2011**, *12*, 684-700.

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

[31] Wang, M.; Roberts, D. L.; Paschke, R.; Shea, T. M.; Masters, B. S.; Kim, J. J. Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. *Proc. Natl. Acad. Sci. U. S. A.*, **1997**, *94*, 8411-6.

[32] Bernhardt, R.; Kraft, R.; Otto, A.; Ruckpaul, K. Electrostatic interactions between cytochrome P-450 LM2 and NADPH-cytochrome P-450 reductase. *Biomed. Biochim. Acta*, **1988**, *47*, 581-92.

[33] Zhao, Q.; Modi, S.; Smith, G.; Paine, M.; McDonagh, P. D.; Wolf, C. R.; Tew, D.; Lian, L. Y.; Roberts, G. C.; Driessen, H. P. Crystal structure of the FMN-binding domain of human cytochrome P450 reductase at 1.93 A resolution. *Protein Sci.*, **1999**, *8*, 298-306.

[34] Shen, A. L.; Kasper, C. B. Role of acidic residues in the interaction of NADPH-cytochrome P450 oxidoreductase with cytochrome P450 and cytochrome c. *J. Biol. Chem.*, **1995**, *270*, 27475-80.

[35] Davydov, D. R.; Kariakin, a. a.; Petushkova, N. a.; Peterson, J. a. Association of cytochromes P450 with their reductases: opposite sign of the electrostatic interactions in P450BM-3 as compared with the microsomal 2B4 system. *Biochemistry*, **2000**, *39*, 6489-97.

[36] Bridges, a.; Gruenke, L.; Chang, Y. T.; Vakser, I. a.; Loew, G.; Waskell, L. Identification of the binding site on cytochrome P450 2B4 for cytochrome b5 and cytochrome P450 reductase. *J. Biol. Chem.*, **1998**, *273*, 17036-49.

[37] Zawaira, A.; Gallotta, M.; Beeton-Kempen, N.; Coulson, L.; Marais, P.; Kuttel, M.; Blackburn, J. Exhaustive computational search of ionic-charge clusters that mediate interactions between mammalian cytochrome P450 (CYP) and P450-oxidoreductase (POR) proteins. *Comput. Biol. Chem.*, **2010**, *34*, 42-52.

[38] Ingelman-Sundberg, M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol. Sci.*, **2004**, *25*, 193-200.

[39] Daly, A. K. Pharmacogenetics of the cytochromes P450. *Curr. Top. Med. Chem.*, **2004**, *4*, 1733-44.

[40] Ingelman-sundberg, M.; Sim, S. C.; Gomez, A.; Rodriguez-antona, C. Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. *Cancer*, **2007**, *116*, 496-526.

[41] Fukami, T.; Nakajima, M.; Higashi, E.; Yamanaka, H.; Sakai, H.; McLeod, H. L.; Yokoi, T. Characterization of novel CYP2A6 polymorphic alleles (CYP2A6*18 and CYP2A6*19) that affect enzymatic activity. *Drug Metab. Dispos.*, **2005**, *33*, 1202-1210.

[42] Ariyoshi, N.; Sawamura, Y.; Kamataki, T. A novel single nucleotide polymorphism altering stability and activity of CYP2a6. *Biochem. Biophys. Res. Commun.*, **2001**, *281*, 810-814.

[43] Wang, H.; Kim, R. a.; Sun, D.; Gao, Y.; Wang, H.; Zhu, J.; Chen, C. Evaluation of the effects of 18 non-synonymous single-nucleotide polymorphisms of CYP450 2C19 on *in vitro* drug inhibition potential by a fluorescence-based high-throughput assay. *Xenobiotica*, **2011**, *41*, 826-35.

[44] Zhou, H.; Josephy, P. D.; Kim, D.; Guengerich, F. P. Functional characterization of four allelic variants of human cytochrome P450 1A2. *Arch. Biochem. Biophys.*, **2004**, *422*, 23-30.

[45] Omura, T.; Sato, R. The carbon monoxide-binding pigment of liver microsomes II. Solubilization, purification, and properties. *J. Biol. Chem.*, **1964**, *239*, 2379-85.

[46] Cohen, L. H.; Remley, M. J.; Raunig, D.; Vaz, A. D. N. *In vitro* drug interactions of cytochrome p450: an evaluation of fluorogenic to conventional substrates. *Drug Metab. Dispos.*, **2003**, *31*, 1005-15.

[47] Di, L.; Kerns, E. H.; Li, S. Q.; Carter, G. T. Comparison of cytochrome P450 inhibition assays for drug discovery using human liver microsomes with LC-MS, rhCYP450 isozymes with fluorescence, and double cocktail with LC-MS. *Int. J. Pharm.*, **2007**, *335*, 1-11.

[48] Lahoz, A.; Donato, M. T.; Castell, J. V.; Gómez-Lechón, M. J. Strategies to *in vitro* assessment of major human CYP enzyme activities by using liquid chromatography tandem mass spectrometry. *Curr. Drug Metab.*, **2008**, *9*, 12-9.

[49] Crespi, C. L.; Miller, V. P.; Penman, B. W. Microtiter plate assays for inhibition of human, drug-metabolizing cytochromes P450. *Anal. Biochem.*, **1997**, *248*, 188-90.

[50] Trubetskoy, O. V.; Gibson, J. R.; Marks, B. D. Highly miniaturized formats for *in vitro* drug metabolism assays using vivid fluorescent substrates and recombinant human cytochrome P450 enzymes. *J. Biomol. Screen.*, **2005**, *10*, 56-66.

[51] Sukumaran, S. M.; Potsaid, B.; Lee, M.; Clark, D. S.; Dordick, J. S. Development of a fluorescence-based, ultra high-throughput screening platform for nanoliter-scale cytochrome p450 microarrays. *J. Biomol. Screen.*, **2009**, *14*, 668-78.

[52] Research Collaboratory for Structural Bioinformatics, Protein Data Bank: A Resource for Studying Biological Macromolecules. *http://www.pdb.org/* (accessed 1 June, 2011).

[53] Human Cytochrome P450 Allele Nomenclature Committee Home Page. *www.cypalleles.ki.se* (accessed 10 Oct, 2012).

[54] NCBI dbSNP Home Page. *http://www.ncbi.nlm.nih.gov/SNP/index.html* (accessed 1 June, 2011).

[55] Smith, R. E.; Lovell, S. C.; Burke, D. F.; Montalvao, R. W.; Blundell, T. L. Andante: reducing side-chain rotamer search space during comparative modeling using environment-specific substitution probabilities. *Bioinformatics*, **2007**, *23*, 1099-1105.

[56] DeLano, W. L. The PyMOL Molecular Graphics System. Version 1.4, Schrödinger, LLC.

[57] Worth, C. L.; Burke, D.; Blundell, T. L. Estimating the Effects Of Single Nucleotide Polymorphisms On Protein Structure: How Good are we at Identifying Likely Disease Associated Mutations ? *Proceedings of Molecular Interactions—Bringing Chemistry to Life*, **2007**, 11-26.

[58] Topham, C.; Srinivasan, N.; Blundell, T. Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng.*, **1997**, *10*, 7-21.

[59] Sali, A.; Blundell, T. L. Definition of general topological equivalence in protein structures: A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *J. Mol. Biol.*, **1990**, *212*, 403-428.

[60] Mizuguchi, K.; Deane, C. M.; Blundell, T. L.; Johnson, M. S.; Overington, J. P. JOY: protein sequence-structure representation and analysis. *Bioinformatics*, **1998**, *14*, 617-623.

[61] Neve, E. P. A.; Ingelman-Sundberg, M. Intracellular transport and localization of microsomal cytochrome P450. *Anal. Bioanal. Chem.*, **2008**, *392*, 1075-84.

[62] Chen, C. D.; Kemper, B. Different structural requirements at specific proline residue positions in the conserved proline-rich region of cytochrome P450 2C2. *J. Biol. Chem.*, **1996**, *271*, 28607-11.

[63] Chen, C. D.; Doray, B.; Kemper, B. A conserved proline-rich sequence between the N-terminal signal-anchor and catalytic domains is required for assembly of functional cytochrome P450 2C2. *Arch. Biochem. Biophys.*, **1998**, *350*, 233-8.

[64] Kemper, B. Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. *Toxicol. Appl. Pharmacol.*, **2004**, *199*, 305-15.

[65] Schreyer, A.; Blundell, T. CREDO: a protein-ligand interaction database for drug discovery. *Chem. Biol Drug Des.*, **2009**, *73*, 157-67.

[66] Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.*, **2007**, *47*, 195-207.

[67] Fishelovitch, D.; Shaik, S.; Wolfson, H. J.; Nussinov, R. How does the reductase help to regulate the catalytic cycle of cytochrome P450 3A4 using the conserved water channel? *J. Phys. Chem. B.*, **2010**, *114*, 5964-70.

[68] Watanabe, T.; Sakuyama, K.; Sasaki, T.; Ishii, Y.; Ishikawa, M.; Hirasawa, N.; Hiratsuka, M. Functional characterization of 26 CYP2B6 allelic variants (CYP2B6. 2–CYP2B6. 28, except CYP2B6. 22). *Pharmacogenet. Genomics*, **2010**, *20*, 459-462.

[69] Wang, H.; An, N.; Wang, H.; Gao, Y.; Liu, D.; Bian, T.; Zhu, J.; Chen, C. Evaluation of the Effects of 20 nonsynonymous single nucleotide polymorphisms of CYP2C19 on S-mephenytoin 4′-hydroxylation and omeprazole 5′-hydroxylation. *Drug Metab. Dispos.*, **2011**, *39*, 830-837.

[70] Couñago, R.; Wilson, C. J.; Peña, M. I.; Wittung-Stafshede, P.; Shamoo, Y. An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability–activity trade-offs. *Protein Eng. Des. Sel.*, **2008**, *21*, 19-27.

[71] Mukaiyama, A.; Haruki, M.; Ota, M.; Koga, Y.; Takano, K.; Kanaya, S. A hyperthermophilic protein acquires function at the cost of stability. *Biochemistry*, **2006**, *45*, 12673-12679.

[72] Beadle, B. M.; Shoichet, B. K. Structural bases of stability–function tradeoffs in enzymes. *J. Mol. Biol.*, **2002**, *321*, 285-296.

[73] Yutani, K.; Ogasahara, K.; Tsujita, T.; Sugino, Y. Dependence of conformational stability on hydrophobicity of the amino acid residue in a series of variant proteins substituted at a unique position of tryptophan synthase alpha subunit. *Proc. Natl. Acad. Sci.*, **1987**, *84*, 4441-4444.

[74] Yun, C.; Yim, S.; Kim, D.; Ahn, T. Functional expression of human cytochrome P450 enzymes in Escherichia coli. *Curr. Drug Metab.*, **2006**, *7*, 411-29.

[75] Wang, L.; Li, Y.; Zhou, S. A Bioinformatics Approach for the Phenotype Prediction of Nonsynonymous Single Nucleotide Polymorphisms in Human Cytochromes P450. *Drug Metab. Dispos.*, **2009**, *37*, 977-991.

[76] Sano, E.; Li, W.; Yuki, H.; Liu, X.; Furihata, T.; Kobayashi, K.; Chiba, K.; Neya, S.; Hoshino, T. Mechanism of the decrease in catalytic activity of human cytochrome P450 2C9 polymorphic variants investigated by computational analysis. *J.Comput.Chem.*, **2010**, *31*, 2746-2758.

[77] Oda, A.; Yamaotsu, N.; Hirono, S. Studies of Binding Modes of (S)-Mephenytoin to Wild Types and Mutants of Cytochrome P450 2C19 and 2C9 Using Homology Modeling and Computational Docking. *Pharm. Res.*, **2004**, *21*, 2270-2278.

[78] Zhou, Y. -.; Zheng, Q. -.; Li, Z. -.; Zhang, Y.; Sun, M.; Sun, C. -.; Si, D.; Cai, L.; Guo, Y.; Zhou, H. On the human CYP2C9* 13 variant activity reduction: a molecular dynamics simulation and docking study. *Biochimie*, **2006**, *88*, 1457-1465.

[79] Bash, P.; Singh, U.; Langridge, R.; Kollman, P. Free energy calculations by computer simulation. *Science*, **1987**, *236*, 564-568.

[80] Park, H.; Lee, S. Prediction of the mutation-induced change in thermodynamic stabilities of membrane proteins from free energy simulations. *Biophys. Chem.*, **2005**, *114*, 191-7.

[81] Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **2000**, *33*, 889-97.

[82] Bordner, A. J.; Abagyan, R. A. Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. *Proteins*, **2004**, *57*, 400-13.

[83] Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, **2002**, *320*, 369-87.

[84] Gilis, D.; Rooman, M. Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. *J. Mol. Biol.*, **1997**, *272*, 276-90.

[85] Saraboji, K.; Gromiha, M. M.; Ponnuswamy, M. N. Average assignment method for predicting the stability of protein mutants. *Biopolymers*, **2006**, *82*, 80-92.

[86] Frenz, C. M. Neural network-based prediction of mutation-induced protein stability changes in Staphylococcal nuclease at 20 residue positions. *Proteins*, **2005**, *59*, 147-51.

[87] Capriotti, E.; Fariselli, P.; Casadio, R. A neural-network-based method for predicting protein stability changes upon single point mutations. *Bioinformatics*, **2004**, *20 Suppl 1*, i63-8.

[88] Cheng, J.; Randall, A.; Baldi, P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, **2006**, *62*, 1125-32.

[89] Capriotti, E.; Fariselli, P.; Calabrese, R.; Casadio, R. Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*, **2005**, *21 Suppl 2*, ii54-8.

[90] Worth, C. L.; Schreyer, A.; Forman, J. R.; Cheng, T. M. K.; Lee, S.; Gong, S.; Burke, D. F. A structural bioinformatics approach to the analysis of nonsynonymous single nucleotide polymorphisms (nsSNPs) and their relation to disease. *JBCB*, **2007**, *5*, 1297-1318.

[91] Ibeanu, G.; Goldstein, J.; Meyer, U.; Benhamou, S.; Bouchardy, C.; Dayer, P.; Ghanayem, B.; Blaisdell, J. Identification of new human CYP2C19 alleles (CYP2C19*6 and CYP2C19*2B) in a Caucasian poor metabolizer of mephenytoin. *J. Pharmacol. Exp. Ther.*, **1998**, *286*, 1490-1495.

[92] Burke, D. F.; Worth, C. L.; Priego, E.; Cheng, T.; Smink, L. J.; Todd, J. A.; Blundell, T. L. Genome bioinformatic analysis of nonsynonymous SNPs. *BMC Bioinformatics*, **2007**, *8*, 301.

[93] Blaisdell, J.; Mohrenweiser, H.; Jackson, J.; Ferguson, S.; Coulter, S.; Chanas, B.; Xi, T.; Ghanayem, B.; Goldstein, J. A. Identification and functional characterization of new potentially defective alleles of human CYP2C19. *Pharmacogenetics*, **2002**, *12*, 703-711.

[94] Hendrychová, T.; Anzenbacherová, E.; Hudeček, J.; Skopalík, J.; Lange, R.; Hildebrandt, P.; Otyepka, M.; Anzenbacher, P. Flexibility of human cytochrome P450 enzymes: Molecular dynamics and spectroscopy reveal important function-related variations. *Biochim. Biophys. Acta*, **2011**, *1814*, 58-68.

[95] Hiratsuka, M. *In vitro* assessment of the allelic variants of cytochrome P450. *Drug Metab. Pharmacokinet.*, **2012**, *27*, 68-84.

Received: November 21, 2012 **Revised: June 21, 2013** **Accepted: June 27, 2013**