COMPUTATIONAL MODELING AND STRUCTURE-ACTIVITY RELATIONSHIPS OF CYTOCHROME P450 1A1: AN APPROACH TO CYP1A1 GENE DIRECTED ENZYME PRODRUG THERAPY

Thesis submitted to the School of Medicine, Faculty of Health Science, Flinders University, Adelaide, Australia, in fulfilling the requirements for the degree of Doctor of Philosophy

August 2010

BENJAMIN C LEWIS

Supervisors: Professor John O Miners, Professor Peter I Mackenzie

TABLE OF CONTENTS

COMPUTATIONAL MODELING AND STRUCTURE-ACTIVITY RELATIONSHIPS OF CYTOCHROME P450 1A1: AN APPROACH TO CYP1A1 GENE DIRECTED ENZYME PRODRUG THERAPY

LIST OF FIGURES

LIST OF TABLES

DECLARATION

"I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text."

Benjamin C Lewis

Dated:

ACKNOWLEDGEMENTS

I would like to thank those people who have endured or are suffering with cancer for inspiring me throughout this study.

I sincerely thank my supervisors, Professor John O Miners and Professor Peter I Mackenzie for their guidance, phenomenal expertise, and patience. Their perpetual enthusiasm for research motivates me to be a better scientist; I shall never forget you.

I dedicate this thesis to the two people who know me best, my wonderful wife Emily and my dear daughter May; you complete me.

PEER-REVIEWED WORK RESULTING FROM THIS THESIS

Refereed journal articles

1. **Lewis BC**, Mackenzie PI, Miners JO. Comparative homology modeling of human cytochrome P4501A1 (CYP1A1) and confirmation of residues involved in 7-ethoxyresorufin *O*-deethylation by site-directed mutagenesis and enzyme kinetic analysis. *Archives of Biochemistry and Biophysics*. 2007 Dec 1; 468(1): 58-69.

Conference proceedings

1. Homology modeling of Cytochrome P4501A1 (CYP1A1). **B.C. Lewis**, P.I. Mackenzie, and J.O. Miners. Joint meeting of the Australasian Society of Experimental and Clinical Pharmacologists and Toxicologists and the Australasian Pharmaceutical Sciences Association; Melbourne, Australia (2005).

2. Homology modeling of Cytochrome P4501A1 (CYP1A1). **B.C. Lewis**, P.I. Mackenzie, and J.O. Miners. International Conference on Pharmacogenetics; Changsha, China (2006).

3. Identification of the cytochrome P4501A1 (CYP1A1) residues involved in 7-ethoxyresorufin *O*-deethylation. **B.C. Lewis**, P.I. Mackenzie, and J.O. Miners. Joint meeting of the Australasian Society of Clinical and Experimental Pharmacologists and Toxicologists, and the South East Asian, Western Pacific Regional Meeting of Pharmacologists; Adelaide, Australia (2007).

4. Structure-function relationships of cytochrome P4501A1 (CYP1A1). **B.C. Lewis**, P.I. Mackenzie, and J.O. Miners. Adelaide Pharmacology Group meeting; National Wine Centre of Australia, Adelaide, Australia (2008).

5. Homology modeling guided mutagenesis of cytochrome P4501A1 (CYP1A1) to identify residues involved in 7-ethoxyresorufin *O*-deethylation. **B.C. Lewis**, P.I. Mackenzie, and J.O. Miners. Joint meeting of the Australian Society for Biochemistry and Molecular Biology, and the Australia and New Zealand Society for Cell and Developmental Biology; (ComBio), Canberra, Australia (2008).

Invited lectures and presentations

1. Adelaide Pharmacology Group meeting, National Wine Centre of Australia, Adelaide, Australia (2008).

2. Combio2008. The Annual Conferences of the Australian Society for Biochemistry and Molecular Biology, the Australian Society of Plant Scientists and the Australia and New Zealand Society for Cell and Developmental Biology; Canberra, Australia (21st-25th September, 2008).

Scientific awards

1. **2007 ASCEPT Oral Prize**. The ASCEPT Oral Prize is awarded annually for the best oral communication by a Higher Degree student member of the Australasian Society of Clinical and Experimental Pharmacologists and Toxicologists.

2. **2007 Percy Prize finalist.** In memory of Dr Neville Percy, the Percy prize is awarded to a higher degree student for a poster presentation.

ABSTRACT

Given the paucity of data relating to the structure of human CYP1A1, an enzyme of considerable toxicological significance, the initial aim of this thesis was to generate a chemically and structurally valid CYP1A1 homology model. CYP1A1 homology models based on the CYP2C5 X-ray crystal structure and a composite of the CYP2C5, CYP2C8, and CYP2C9 X-ray crystal structures were compared to a model generated using the crystal coordinates of CYP1A2. The model using the CYP1A2 coordinates gave near ideal stereochemical quality and was favored energetically. Automated *in silico* docking studies identified active-site residues potentially involved in the orientation and binding of the prototypic CYP1A1 substrate, 7-ethoxyresorufin. The most energetically favorable pose placed the carbon atom adjacent to the ether oxygen of 7-ethoxyresorufin at 4.4Å from the heme iron, at an angle of 106.4º to the plane of the heme. The CYP1A1 mutants S122A, F123A, F224A, A317Y, T321G, and I386G were generated to explore the roles of these residues in 7-ethoxyresorufin binding and turnover, and generally confirmed the importance of aromatic interactions over hydrogen bonding in orientating 7-ethoxyresorufin in a catalytically favorable orientation.

The validated 3-dimensional structure of CYP1A1 was subsequently employed to elucidate structure-activity relationships with the anticancer prodrug, dacarbazine (DTIC). *In silico* docking experiments demonstrated that DTIC orientates in close proximity to S122, F123, D313, A317, I386, Y259, and L496. Docking located the site of metabolism of DTIC at 5.6Å from the heme iron at an angle of 105.3º to the plane of the heme. Binding of DTIC in the active-site was stabilized by H-bonding between Y259 and the N2 position of the imidazole ring. Structural modification of

the CYP1A1 enzyme to increase its catalytic efficiency (V_{max}/K_m) for *N*-demethylation and the subsequent activation of DTIC was facilitated by the CYP1A1 homology model. Twenty-nine CYP1A1 mutants were generated and expressed in *E. coli*. DTIC *N*-demethylation by the CYP1A1 E161K, E256K, and I458V mutants exhibited Michaelis-Menten kinetics, with decreases in K_m that doubled the catalytic efficiency relative to wild-type ($P < 0.05$). The kinetics of DTIC *N*-demethylation by the CYP1A1*2C and CYP1A1*4 polymorphic variants was additionally characterized. There was an approximate 30% reduction in catalytic efficiency of the CYP1A1*2C and CYP1A1*4 variants relative to wild-type. Thus, patients with malignancies who carry either polymorphism may not respond as well to DTIC treatment compared to those expressing the wild-type enzyme.

As a chemotherapeutic agent, DTIC has relatively poor clinical activity in human malignancies and exhibits numerous serious adverse effects, which presumably arise from bioactivation in the liver and other tissues resulting in systemic exposure to the cytotoxic metabolite. Gene directed enzyme prodrug therapy (GDEPT) provides a means to enhance the efficacy and reduce the systemic toxicities associated with conventional chemotherapy. Thus, COS-7 and SK-MEL-28 cells were transfected with cDNA encoding an open reading frame (ORF) comprising the CYP1A1 (wild-type or mutant) coding sequence (CDS), a picornaviral 2A cleavage peptide, followed by the cytochrome P450 oxidoreductase (OxR) CDS. Cell line sensitization assays using the wild-type, E161K, E256K, and I458V CYP1A1-2A-OxR constructs were attempted, but the generation of stable cell lines was not successful. However, data obtained from transiently expressed CYP1A1-2A-OxR motifs in COS-7 and SK-MEL-28 cells revealed the synthesis of both holo-CYP1A1 and OxR proteins.

This study elucidated important structural characteristics of human CYP1A1 and how manipulation of protein tertiary structure can enhance enzyme function. Combination of kinetic analyses with *in silico* docking data has allowed interpretation of the structure-activity relationships of CYP1A1 and DTIC. Moreover, the successful generation of CYP1A1 enzymes with catalytically enhanced DTIC activation highlights their potential use as a strategy for P450-based GDEPT in the treatment of metastatic malignant melanoma.

ABBREVIATIONS AND SYMBOLS

±	plus or minus	**EROD**	ethoxyresorufin O-deethylase
17α	P45017A hydroxylase leader sequence	**FAD**	flavin adenine dinucleotide
4-MU	4-methylumbeliferone	**FF**	atomic force field
Å	angstroms (1.0 x 10^{-10} m)	**FMN**	flavin mononucleotide
AFM	atomic force microscopy	**FxR**	ferrodoxin reductase
AhR	aromatic hydrocarbon receptor	**g**	gram
AIC	Amino-imidazole-carboxamide	*g*	gravitational force
Amp^R	ampicillin resistance	**GDEPT**	gene-directed enzyme prodrug therapy
bp	basepairs	**h**	hour(s)
°C	degrees Celsius	**HLM**	human liver microsomes
cDNA	complementary deoxyribonucleic acid	**HMM**	hidden Markov model
CDS	coding sequence	**HMMTIC**	hydroxy-dimethyl-triazenyl-imidazole-carboxamide
Chlor^R	chloramphenicol resistance	**HPLC**	high performance liquid chromatography
cm	centimetre	**HRP**	horseradish peroxidase
CYP	cytochrome P450	**HTE**	high transformation efficiency
CYP1A1	cytochrome P4501A1	**IU**	international units
Cα	alpha-carbon atom	**kb**	kilobases
Da	Daltons	**kcal**	kilocalorie
DTIC	dacarbazine	**kDa**	kilodaltons
DTT	dithiothreitol	K_m	Michaelis constant
E. coli	*Escherichia coli*	K_{si}	reversible inhibition
Eres	7-ethoxyresorufin		

	constant	**NMR**	nuclear magnetic resonance
L	litre		
LB	Lauria-Bertani broth or agar	**OD**	optical density
		OmpA	outer membrane protein-A
M	Moles per litre		
mA	milliamps	**OxR**	cytochrome P450 oxidoreductase
mg	milligram	**P450**	cytochrome P450
min	minute(s)		
mL	millilitre	**PAH**	polycyclic aromatic hydrocarbon
μL	microlitre	**PDB**	RCSB protein databank
mM	millimolar	**pmol**	picomole
mm	millimetre	**PMSF**	phenylmethylsulfonyl fluoride
μM	micromolar		
μm	micrometer	r^2	coefficient of determination
MOPS	3-morpholino propane sulfonic acid	**Res**	resorufin
mRNA	messenger ribonucleic acid	**RMSD**	root mean square deviation
		rpm	revolutions per minute
MTIC	methyl-triazenyl-imidazole-carboxamide	**R-value**	statistical measure of fit
n	Hill coefficient	**[S]**	substrate concentration
$NADP^+$	nicotinamide dinucleotide phosphate	S_{50}	substrate concentration at half V_{max} in the Hill equation
NADPH	nicotinamide dinucleotide phosphate, reduced form	**SCR**	sequence conserved region
		SD	standard deviation
nd	not detected		
ng	nanogram	**SDS PAGE**	sodium dodecyl sulfate polyacrylamide gel electrophoresis
nM	nanomolar		
nmole	nanomole	**sec**	second(s)

SNP	single nucleotide polymorphism
SRS	substrate recognition site
TBS	tris-buffered saline
TBST	TBS + tween-20
TEA	triethylamine
UV	Ultraviolet
V	volts
v	enzymatic rate of product formation
V	volume (L)
v/v	volume per volume
vdW	van der Waals force
Vis	visible
V_{max}	maximal velocity of enzyme reaction rate
W	watts
w/v	weight per volume
w/w	weight per weight
ϕ	phi torsional angle
ψ	psi torsional angle

1.1 Overview

The concept of overexpression of individual cytochrome P450 enzymes in tumor cells is now becoming well recognized as a means to develop novel targets for anticancer therapy. The outcome in terms of cytotoxicity or lack of response to an anticancer drug is dependant on both the relative amount and the activity of the individual P450(s) within the tumor. Clearly, enhanced expression of individual P450s in tumor cells increases the potential for activation of anticancer prodrugs by the tumor cells directly.

For successful anticancer treatment using a prodrug, it is important to establish whether activation occurs at the tumor site, in the liver (or other 'non-target' tissues), or a combination of both. Extensive activation of a chemotherapeutic agent by an organ remote to the tumor site can lead to undesirable systemic exposure to the cytotoxic drug metabolites that kill host cells as opposed to the tumor cell target. In an attempt to combat this problem, several "next generation" selective cytotoxic agents entering clinical-trial have been designed around P450s, e.g. aminophenylbenzothiazoles and aliphatic amine N-oxides (Patterson & Murray 2002). However, these leads still appear to be inferior since most of the drugs are subject to more complex metabolic activation pathways than many conventional drugs.

Since the activation of many prodrugs used in clinical practice is dependant on the intrinsic ability of the liver to metabolize the drug, activation is sensitive to changes in hepatic enzyme activity. Despite this, few studies examine the metabolic activation of anticancer agents in patients carrying variant P450 alleles. Interindividual variation in P450-mediated metabolism, resulting from genetic polymorphism, effectively renders each patient unique with respect to prodrug activation, as well as drug dosage and kinetics. An area that potentially improves tumor/drug selectivity and one which minimises problems associated with interindividual variation is gene directed enzyme prodrug therapy (GDEPT).

Chemotherapy has long been the sole treatment for metastatic melanoma. In Australia and the United States the prodrug dacarbazine (DTIC) is currently the only approved chemotherapeutic agent targeted specifically to the treatment of metastatic melanoma. However, treatment with DTIC is generally associated with relatively poor outcomes and the co-administration of other therapies offers minimal clinical advantages. DTIC is primarily activated by the cytochromes P450 CYP1A1 and CYP1A2. To date, no studies have characterized the kinetics of DTIC metabolism in patients carrying wild-type or polymorphic CYP1A alleles, or their cellular response after bioactivation. However, numerous studies have investigated the substrate selectivity of the CYP1A enzymes (Lewis et al. 1999; Liu et al. 2003; Liu et al. 2004; Taly et al. 2007). Therefore, an opportunity exists to examine the structure-activity relationships between CYP1A1 and DTIC, however, the chemical structure of CYP1A1 has not been elucidated. Thus, a valid 3-dimensional structure of CYP1A1 would not only aid our understanding of the mechanism(s) involved in DTIC activation, but also aid our understanding of interactions with numerous other CYP1A1 drug substrates.

The purpose of this thesis was to identify important chemical and structural characteristics of human CYP1A1 and how manipulation of the enzyme's tertiary structure may be used to enhance the activation of DTIC. Molecular modeling of CYP1A1 elucidated a valid homology model that was utilized in the successful generation of a range of catalytically altered enzymes. By increasing enzyme-substrate affinity these enzymes doubled the catalytic efficiency of DTIC activation. The CYP1A1 mutants described are potential leads for use in P450-based GDEPT for the treatment of metastatic malignant melanoma.

Despite ongoing attempts by the pharmaceutical industry to discover and develop novel anti-cancer therapies, there have been no new approved therapies for metastatic melanoma use since 1998 (Proleukin, recombinant form of interleukin-2; FDA Center for Drug Evaluation and Research; last updated: July 2010; http://www.centerwatch.com/drug-information/fda-approvals/). The subsequent chapters of this thesis highlight that specialized and focused *in vitro* studies may potentially result in tailoring the activity of a preexisting drug to one with enhanced capabilities that may ultimately be employed in the clinic by means of gene therapy.

1.2 Rationale for the study

The research described in this thesis proceeded from the following broad concepts:

Little chemical or structural information is known regarding human CYP1A1, a key enzyme involved in the activation of the widely used chemotherapeutic agent DTIC. DTIC is currently the only approved non-biologic chemotherapy drug for metastatic melanoma. While it is the best available treatment and the standard against which new melanoma drugs are evaluated, it is unfortunately not particularly effective with

an average response of only 19% with no significant improvement in overall patient survival. Furthermore, DTIC is associated with severe adverse reactions, including liver damage. An improved understanding of the enzyme-substrate interactions and the kinetics of DTIC activation may provide a more targeted approach to cancer treatment.

1.3 Cytochromes P450

1.3.1 Background

P450s are represented by 781 gene families and constitute a superfamily of cysteinato-heme enzymes. These enzymes are present in all forms of life, with fungi being the largest contributor (310 families), followed by bacteria (205 families, including 10 Archaeal), animals (110 families), plants (95 families), and protists (61 families). Excluding variants and pseudogenes, there are approximately 8128 uniquely named P450 amino acid sequences. Of these, 2565 are found in animals (last updated: July 2010; http://drnelson.uthsc.edu/P450.statsfile.html).

Individual P450 enzymes are discrete gene products with molecular masses of approximately 57,000 Daltons (Da) (~500 amino acid residues). All P450s are related but differ from one-another in amino acid sequence and the mechanism by which each gene is regulated. Coding sequence (CDS) variations give rise to differences in secondary- and tertiary-structure (and ultimately quaternary structure), and subsequently determine the chemistry of the protein active-site. It is the chemistry of the protein active-site that establishes P450 substrate and inhibitor selectivities. Different members of the P450 superfamily have distinct, but often

overlapping substrate selectivities, with some enzymes acting on the same substrates, but possessing different stereo-selectivity or different kinetic mechanisms.

P450s are classified according to the electron-transfer protein each enzyme requires for catalysis. Electrons can be supplied to the P450 enzyme from either NADH/FAD (class I), NADPH/FAD-FMN (class II), no electron donor (class III), or NADPH directly (class IV) (Meunier, de Visser & Shaik 2004; Werck-Reichhart & Feyereisen 2000). Mammalian class I P450s are located in the inner mitochondrial membrane and are involved in steroid synthesis. The mitochondrial mammalian P450s, along with the soluble bacterial P450s, utilize the class I electron transport chain which involves the reduction of a flavin adenine dinucleotide (FAD)-containing reductase by NADH. Two electrons are then transferred from FAD to ferrodoxin reductase (FxR; Fe_2-S_2). The FxR protein subsequently acts as an electron shuttle by transferring each electron to the P450 enzyme for catalysis. Class II P450s are involved in the metabolism of drugs and other xenobiotics and are localized in the endoplasmic reticulum (ER) membrane (Peterson & Prough 1986). The electron transport chain of class II P450s, which are found in most eukaryotic cell types, is composed of a complex oxidoreductase (OxR) flavoprotein which has both FAD and flavin mononucleotide (FMN) domains (Meunier, de Visser & Shaik 2004; Peterson & Prough 1986). Potentiometric studies have shown that FAD serves as the initial electron acceptor from NADPH, while FMN serves to reduce the P450 holoenzyme (Figure 1.1) (Fleming et al. 2003). The electron flow to the heme iron is presumed to be regulated by a substrate dependant increase in the redox potential of the heme (Johnson et al. 2005b). An additional COOH-terminally anchored electron donor, cytochrome b_5, which also utilizes electrons from NADPH, has been found to enhance the activity of some class II P450 enzymes (Werck-Reichhart & Feyereisen

2000). Class III enzymes are self-sufficient and require no electron donor. They are involved in the synthesis of endogenous mediators, such as prostaglandins, in mammals. P450s that receive electrons directly from NADPH belong to class IV. Unique to fungi, these soluble P450s have the ability to reduce nitric oxide (NO) to nitrous oxide (N_2O) (Werck-Reichhart & Feyereisen 2000).

This thesis focuses on the mammalian microsomal class II P450 enzymes. Further classification of the class II P450 enzymes is achieved on the basis of their catalytic activity. Activities may be classified in three distinct groups: (i) monooxygenase activity, where an oxygen atom is incorporated into a substrate; (ii) oxidase activity, resulting in the formation of superoxide anion radicals or hydrogenperoxide (uncoupling of the catalytic cycle); and (iii) the production of free radical intermediates from substrate/reductase activity (under anaerobic conditions).
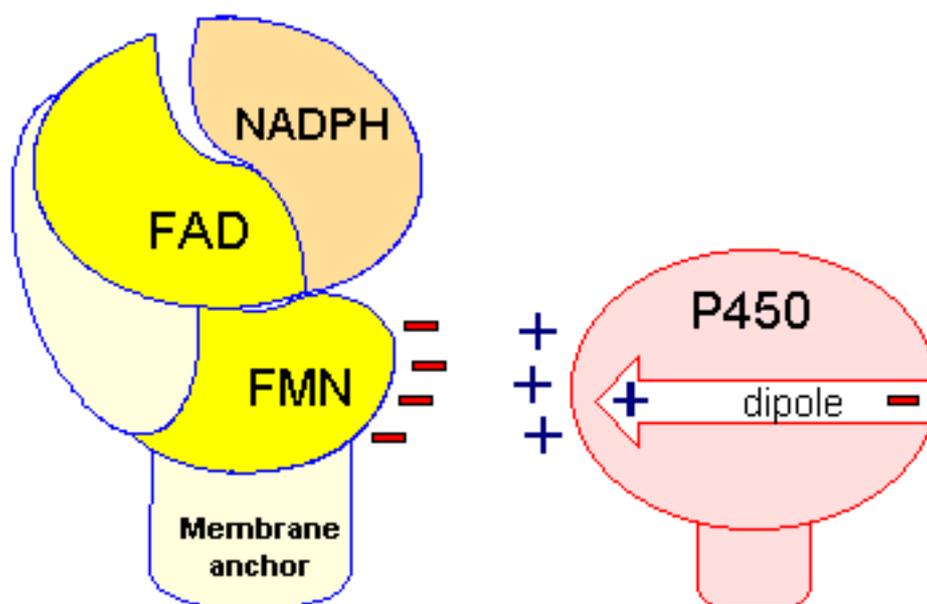
Figure 1.1 The P450 oxidoreductase, flavin adenine dinucleotide and flavin mononucleotide complex. Electrons from the membrane bound NADPH-OxR complex migrate from NADPH to FAD, to FMN, and then finally to the P450 heme Fe^{3+}.

The primary reaction catalyzed by class II P450s is the hydroxylation of a carbon atom, according to the catalytic cycle represented in Figure 1.2. For hydroxylation to proceed, the activation of molecular oxygen (O_2) is required. This is achieved by the heme prosthetic group situated within the active-site of the P450 holoenzyme, which utilizes electrons donated from the NADPH-OxR complex (Fleming et al. 2003). The heme group is comprised of an Fe^{3+} protoporphyrin-IX-complex covalently linked to an atom of sulfur provided by a highly conserved cysteine residue within the carboxy-terminal domain of the P450 enzyme (Figure 1.3). In the non-catalytic state, the sixth coordination site of the Fe^{3+} atom is stabilized by a single water molecule (de Graaf, Vermeulen & Feenstra 2005). In the presence of a substrate, the reductive activation of oxygen occurs from two one-electron donations, and is targeted toward the sixth coordination position. The first electron reduces the substrate-bound ferric heme (Fe^{3+}--R), facilitating the rapid binding of dioxygen and the formation of a ferrous-dioxygen intermediate (Fe^{2+}--O_2--R). It is at this point that other lone pair electron donors, particularly carbon monoxide (CO), can compete with O_2 for the iron electron pair acceptor site (Omura & Sato 1964b). The second electron donation forms a ferric hydroperoxy complex that is subsequently protonated (Fe^{3+}--H_2O_2--R). The O--O bond is then cleaved leaving one atom of oxygen incorporated into the substrate (R) with the remaining atom of oxygen reduced to form water (eq 1.1) (Fleming et al. 2003; Guengerich 2001; Guengerich et al. 2000; Meunier, de Visser & Shaik 2004).

$$\text{R-H} + O_2 + 2\acute{e} + 2H^+ \xrightarrow{\text{cytochrome P450}} \text{ROH} + H_2O$$

$$(1.1)$$
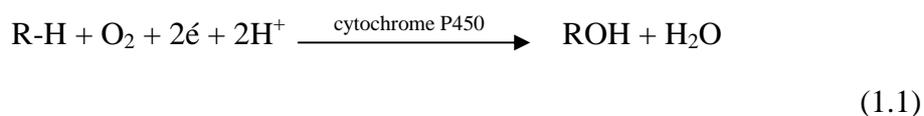
where R is the C-H bond of any P450 substrate and é is a single electron.
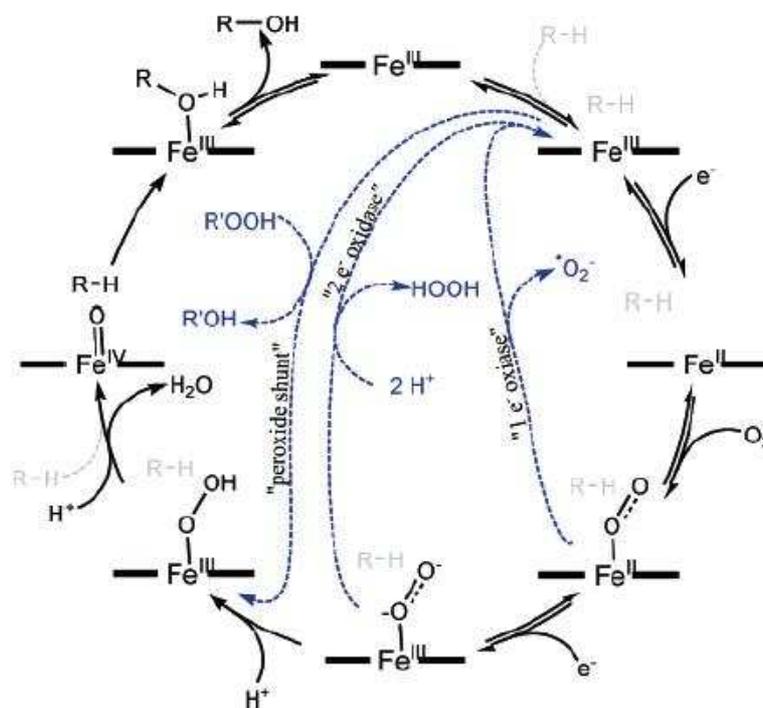
Figure 1.2 Schematic representation of the catalytic cycle of class II P450 enzymes. The main reaction is the monooxygenation pathway used in hydroxylation (black). Three alternative pathways are shown (blue) corresponding to the one- and two-electron oxidase uncoupling reactions and the peroxide shunt. The one-electron oxidase produces a superoxide radical, which can subsequently generate a radical derivative of the substrate. The peroxide shunt is the pathway used in generating an activated iron-oxygen species. Taken from de Graaf, Vermeulen & Feenstra (2005).
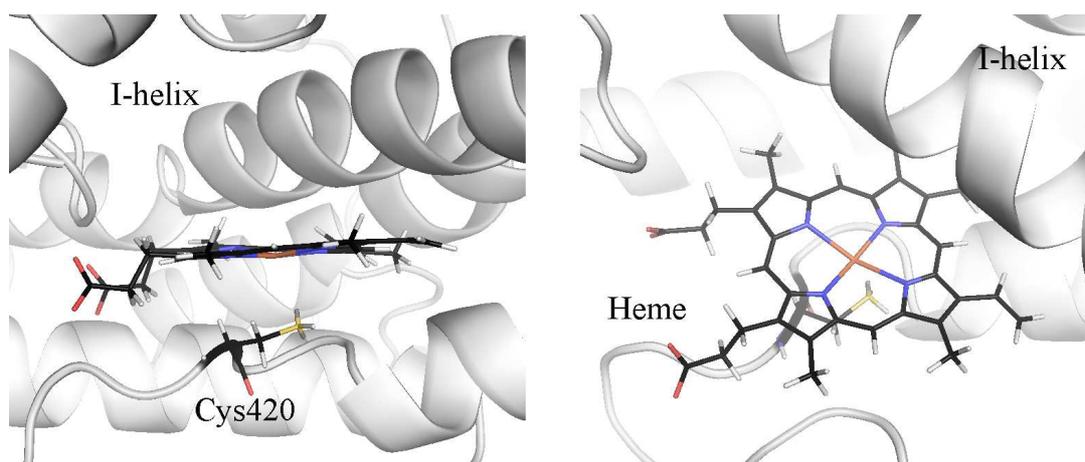
Figure 1.3 The heme prosthetic group of the cytochromes P450. Orthogonal views of the Fe^{3+} protoporphyrin-IX-complex covalently linked to the highly conserved cysteine residue (labeled Cys420). Sulfur atoms are colored yellow. For simplification, the 6th co-ordination site of the octahedral Fe^{3+} atom is not included.

1.3.2 Orientation and membrane topology of cytochromes P450

The membrane topology of P450s has been extensively studied by numerous chemical and computational approaches (Szczesna-Skorupa et al. 1995). Class II P450s reside on the cytosolic face of the ER and do not undergo recycling through the pre-Golgi compartment (Szczesna-Skorupa et al. 1998). Initial data suggested that P450s of the ER contained four to eight transmembrane helical domains, with the active-site positioned within the membrane itself. Subsequent findings identified microsomal P450s as being bound to the ER by a single 'anchor' site located at the NH_2-terminus. Moreover, the active-site was shown to be part of a large cytoplasmic domain that potentially contains one or two additional membrane contact sites. The presence of an NH_2-terminal signal peptide is strong evidence for targeting microsomal P450s to the ER membrane. The apparent signal peptide is not proteolytically cleaved, but is actually retained and utilized as the membrane anchor (Black 1992; Szczesna-Skorupa et al. 1998). Once associated with the ER membrane, the mobility of the membrane bound P450 has been likened to that of Golgi and plasma-membrane proteins ($2\text{-}6\mathrm{x}10^{-10}\mathrm{cm}^2/\mathrm{s}$) (Werck-Reichhart & Feyereisen 2000). However, P450s are thought to be hindered from traversing the membrane surface. Szczesna-Skorupa *et al.* (1998) hypothesized that the mechanism of P450 retention in the ER membrane may be related to the formation of protein-protein aggregates. Since P450s are known to closely associate with their redox partners, this appears highly probable.

The orientation of microsomal P450s in the ER membrane is clearly dependant on the location of the insertion signal, NH_2-terminal anchor, and the halt-transfer signal that serves to arrest polypeptide translocation through the membrane. In addition, the heme-containing domain may utilize amphipathic helices in binding to the cytosolic

side of the membrane. Black (1992) hypothesized that the plane of the active-site heme is likely positioned at an angle somewhat greater than parallel to the membrane surface. Likewise, research conducted by Bayburt & Sligar (2002) and Edwards et al. (1991) concluded that P450s are most likely orientated with the heme positioned perpendicular to the membrane. However, the exact nature of the peripheral membrane interactions of P450s remain incompletely characterized (Edwards et al. 1991; Szczesna-Skorupa et al. 1998).

Due to the challenges associated with crystallizing membrane-bound proteins, mammalian P450 X-ray crystal structures have become available only in the last decade. Crystallization of mammalian CYP1A2 (Sansen et al. 2007), CYP2A6 (Yano et al. 2005), CYP2C8 (Schoch et al. 2004), CYP2C9 (Wester et al. 2004; Williams et al. 2003), CYP3A4 (Yano et al. 2005), CYP2B4 (Scott et al. 2004b), CYP2C5 (Wester et al. 2003a; Wester et al. 2003b; Williams et al. 2000b), and CYP2D6 (Rowland et al. 2006) was accomplished by truncating the NH_2-terminus, thus generating a soluble form of the enzyme. Additionally, atomic force microscopy (AMF) has been employed to directly investigate the topography of membrane-bound proteins. To determine the position of the native membrane-bound enzyme relative to the membrane itself, rabbit CYP2B4 was incorporated into nanometer-scale phospholipids bilayer disks and visualized by AFM (Bayburt & Sligar 2002; Black 1992). Protein insertion occurred to a depth of 3.5nm. The orientation of the heme was estimated from results of anisotropy decay measurements after flash photolysis. Angles of 41º, 55º, or 71º between the heme and membrane were observed for purified CYP2B4 from phenobarbital-induced rabbits. For CYP2B4 from 3-methylcholanthrene induced animals, angles of 48º or 62º were observed, therefore suggesting orientation of the heme lays more perpendicular to the

membrane than parallel to it. Bayburt and Sligar (2002) additionally investigated the orientation of the redox transfer complexes. This was accomplished by overlaying the known structure of bacterial CYP102 with rabbit CYP2B4. They suggested that the overall fold of the CYP102-FMN domain resembles that of the mammalian reductase FMN domain. If the CYP102-FMN domain is positioned on the membrane surface analogous to that proposed for mammalian P450 reductase, then the heme domain is orientated perpendicular to the membrane and the hydrophobic tip of CYP2B4 is inserted into the membrane surface (Figure 1.4).
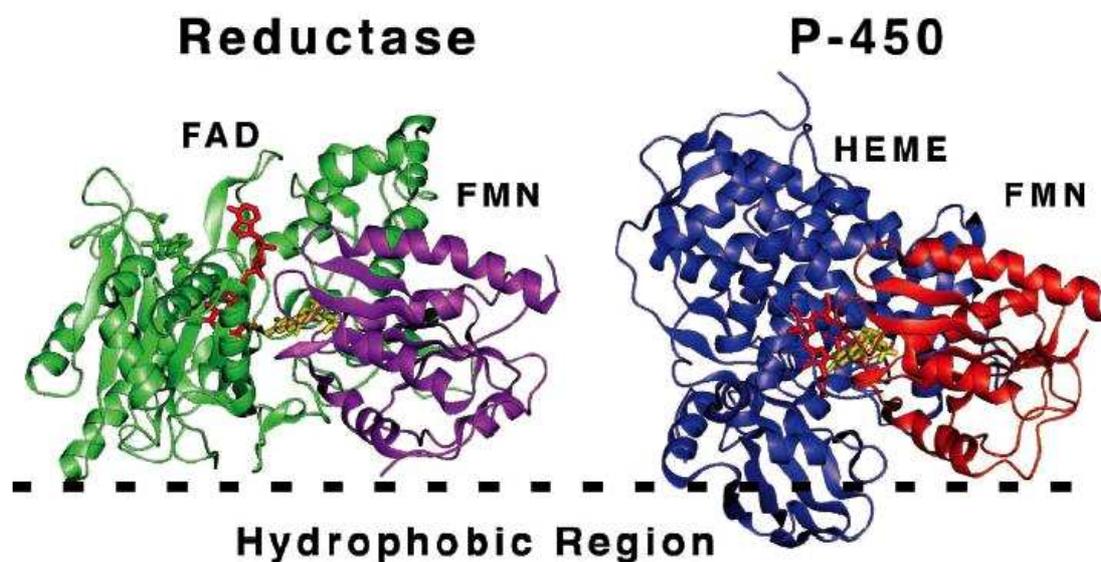
Figure 1.4 Orientation of mammalian OxR and P450 relative to the membrane surface. Cytochrome P450 (blue) is orientated with the heme (red) perpendicular to the ER membrane. Note the position (yellow) where FMN (purple) accepts electrons from FAD (green); in the P450 this position (FMN red) is in close proximity to the heme cofactor.

1.3.3 Human cytochromes P450

Human P450s are integral in regulating the steady-state levels of endogenous hormones and the numerous physiological substrates required for cell proliferation and differentiation (Guengerich 2001; Nebert & Nelson 1991). In addition, human P450s are the major catalysts involved in the biotransformation of pharmaceuticals and are therefore of considerable interest in terms of drug therapy. Notably, the class II P450 enzymes typically insert an atom of atmospheric molecular oxygen (O_2) into a substrate which results in the hydroxylation, epoxidation, and dealkylation of substrates to name a few.

Due to the large number of human P450s, individual enzymes require naming on the basis of evolutionary relatedness to their primary amino acid sequence (Figure 1.5). Members in a given gene family generally have more than 40% identity with other members of the same family, while members of the same human P450 subfamily have greater than 55% amino acid sequence identity and appear to lie on the same chromosome cluster (Nelson et al. 1996).

P450s can be categorized as either xenobiotic-metabolizing enzymes comprising CYP1, CYP2, and CYP3 with some involvement from CYP4, or as P450s involved in the synthesis of endogenous substrates such as steroids, fatty acids, and prostaglandins, namely CYP11, CYP17, CYP19, and CYP21 families (Daly 2003; McFadyen, Melvin & Murray 2004; Petushkova et al. 2006; Rendic 2002). The principal human P450s in CYP1, CYP2, and CYP3 families contribute primarily to the conversion of exogenous lipophilic molecules to water-soluble forms for excretion from the body (eq 1.2).
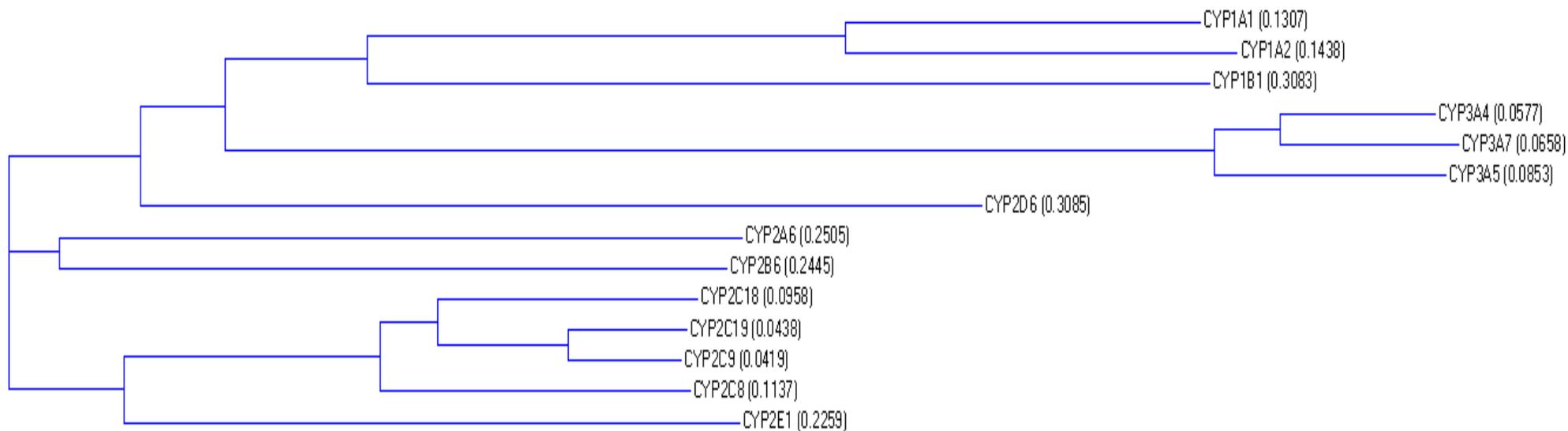
Figure 1.5 Evolutionary tree showing the relatedness of the main human P450s involved in the clearance of hydrophobic compounds.
The Tree was built using the Neighbor Joining method (NJ) of Saitou and Nei (Saitou & Nei 1987). The NJ method utilizes a matrix of distances between all pairs of sequence to be analyzed. These distances are related to the degree of divergence between each gene. The Tree is calculated after the sequences are aligned. Distance values are in parenthesis.

$$R\text{-}H + O_2 + 2\acute{e} + NADPH + H^+ \xrightarrow[\text{\textit{(oxidation)}}]{\text{cytochrome P450}} R\text{-}OH + H_2O + NADP^+$$

$$R\text{-}OH \xrightarrow[\text{\textit{(conjugation)}}]{\text{transferases}} R\text{-}OX \dashrightarrow \text{Renal excretion}$$

(1.2)
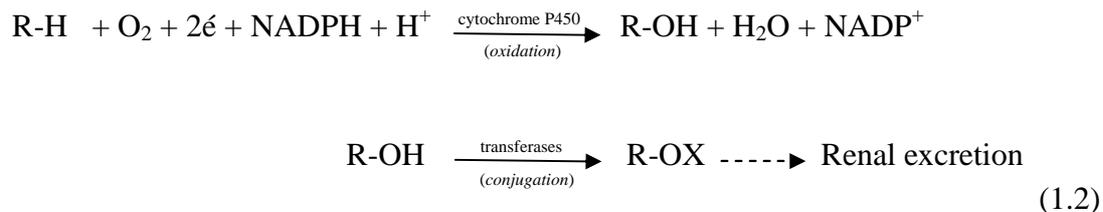
where R is any P450 substrate and X is any conjugation molecule (e.g. glucuronic acid).

The CYP1 family is comprised of enzymes involved in the metabolism of polycyclic aromatic hydrocarbons, heterocyclic compounds, methylxanthines, and aromatic amines. CYP1 is comprised of two subfamilies; CYP1A and CYP1B. The CYP2 family includes enzymes that metabolize numerous organic compounds, namely drugs, non-drug xenobiotics, and steroids, and consists of eight subfamilies comprising mammalian genes CYP2A through CYP2G and CYP2J. The CYP3 family similarly metabolizes drugs and other xenobiotics as well as being responsible for the 6β-hydroxylation of testosterone (Levy et al. 2000). Family CYP3 currently consists of one subfamily, CYP3A, which includes 27 genes. Examples of drug substrates of mammalian P450 enzymes are given in Table 1.1.

Human P450s are expressed in varying amounts throughout many tissues, with the highest levels found in the liver. The relative abundance of P450s in human liver can be seen in Figure 1.6a (Yeo, Rostami-Hodjegan & Tucker 2004). Some 57 different P450 genes (plus 58 pseudogenes) are present in the human genome, with eight accounting for more than 90% of drug oxidations (1A2, 2A6, 2C8, 2C9, 2C19, 2D6, 3A4, 3A5) (Figure 1.6b) (Code et al. 1997).

Table 1.1 Representative substrates of the principal P450 enzymes from families CYP1, CYP2, and CYP3.

P450 Enzyme	Drug Substrate
CYP1A1	phenacetin, dacarbazine, pregnenolone
CYP1A2	caffeine, clozapine, phenacetin, tacrine, theophylline
CYP2A6	nicotine
CYP2C8	cerivastatin, paclitaxel, repaglinide
CYP2C9	ibuprofen, phenytoin, tolbutamide, *S*-warfarin
CYP2C19	omeprazole, proguanil, *S*-mephenytoin
CYP2D6	codeine, debrisoquine, fluoxetine, metoprolol, perhexiline, tamoxifen
CYP2E1	ethanol, enflurane, halothane
CYP3A4/5	Ca-channel antagonists, cyclosporine, HIV-protease inhibitors, midazolam, nifedipine, simvastatin, tacrolimus
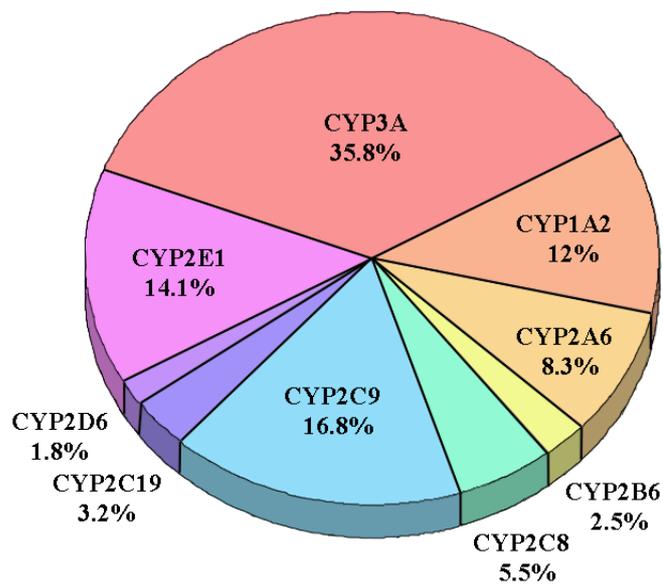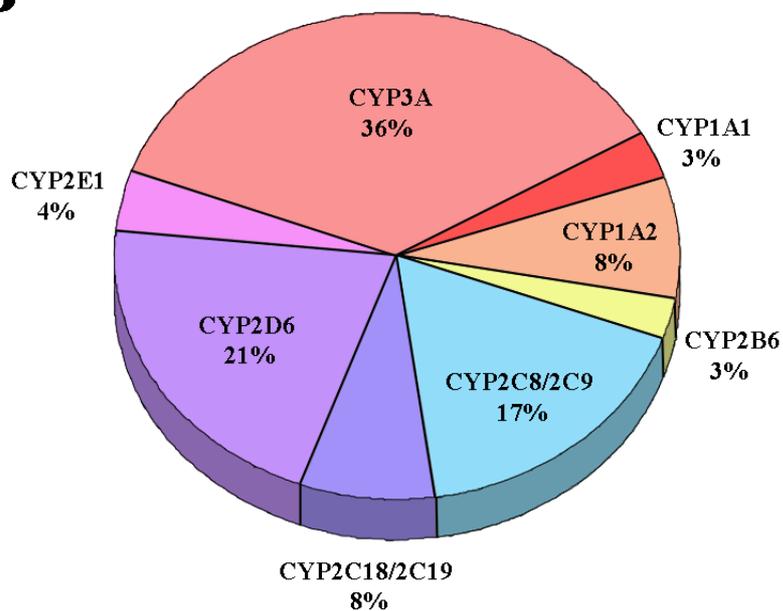
Figure 1.6 The relative abundance of human P450 enzymes in the liver (a); and the percentage of pharmaceuticals metabolized by respective P450 enzymes (b).

1.3.4 Human cytochrome P4501A subfamily

Human CYP1A1 and CYP1A2 share 72% amino acid sequence identity, but display differing substrate and inhibitor profiles, although some substrates can be metabolized by both enzymes. For example, 7-ethoxyresorufin is *O*-deethylated by both enzymes, albeit more effectively by CYP1A1 than CYP1A2, whereas CYP1A2 exhibits a preference for 7-methoxyresorufin (Burke et al. 1994; Hanioka et al. 2000). The basis of this overlapping substrate selectivity was investigated by Liu et al. (2004) whereby five reciprocal active-site mutations between CYP1A1 and CYP1A2 (viz. residues 122/124, 221/223, 225/227, 312, and 382) were used to identify key residues involved in alkoxyresorufin metabolism (Figure 1.7). Residue substitution led to loss of 7-methoxy- and 7-ethoxyresorufin *O*-deethylation activity compared to wild-type, except for the CYP1A1 S122T mutation which increased both activities. In addition, it was shown that mutations at position 382 in both CYP1A1 and CYP1A2 shifted the substrate selectivity from one enzyme to the other, confirming the importance of this residue. Interestingly, the CYP1A1 G225V mutant gave rise to a significant loss of heme and high levels of apoenzyme. Notably, mutagenesis of the CYP1A1 active-site residues involved in the stabilization and orientations of DTIC have not been reported.

CYP1A1-WT

| S122 | N221 G225 | L312 | V382 |

CYP1A2-WT

| T124 | T223 V227 | N312 | L382 |
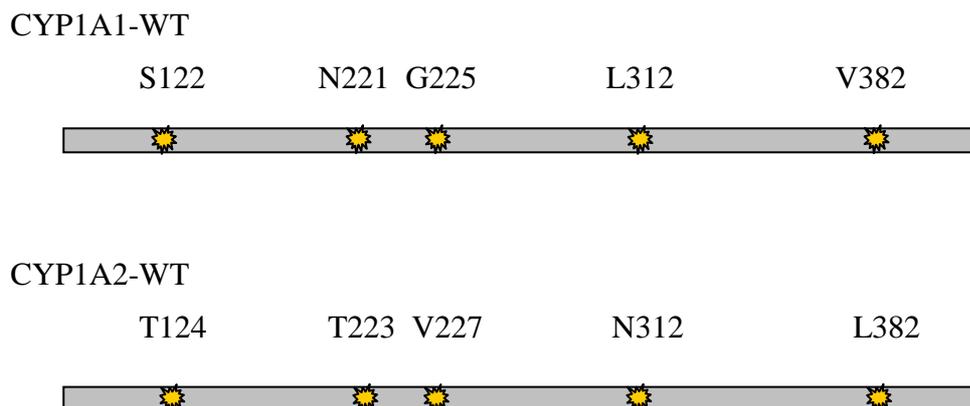
Figure 1.7 Location of the reciprocal mutation sites in CYP1A1 and CYP1A2 employed in the study of Liu *et al.* (2004).

1.3.5 Human cytochrome P4501A1

1.3.5.1 Background

CYP1A1 is the focus of this thesis. The *CYP1A1* (P1-450) gene, located at 15q22-q24, comprises seven exons and six introns and spans 5810 base pairs (Jaiswal, Gonzalez & Nebert 1985; Kawajiri et al. 1986). Human CYP1A1 (EC:1.14.14.1) is mainly present in the skin, lungs, placenta, and lymphocytes, and plays an important role in the metabolic activation of chemical carcinogens. Basal CYP1A1 protein expression in all tissues is thought to be low (Nebert et al. 2004), but varying levels of CYP1A1 mRNA have been detected following induction by polycyclic aromatic hydrocarbons (PAH's). Table A1.1 (Appendix 1) lists most substrates, inhibitors, and exogenous inducers of CYP1A1. CYP1A1 is capable of oxidizing benzo[a]pyrene and other PAH's to carcinogenic species (Chua et al. 2000; Chun, Shimada & Guengerich 1996; Liu et al. 2003; Sparfel et al. 2004). This enzyme is strongly induced by cigarette smoke and potentially associated with lung cancer (Han, Pentecost & Spivack 2003; Patterson & Murray 2002). Like many other P450s,

CYP1A1 is polymorphic and genetic variation is thought to play a role in determining cancer susceptibility (Balasubramanian et al. 2004; Han, Pentecost & Spivack 2003; Miyoshi et al. 2002).

1.3.5.2 Interindividual variability in human CYP1A1 expression

Nine CDS polymorphic alleles of the *CYP1A1* gene have been identified (Table 1.2). The frequencies of the four most common variant alleles, CYP1A1*2A, CYP1A1*2C, CYP1A1*3, and CYP1A1*4, vary between different populations: 2.7-5.1% in Caucasians; 10-15% in Japanese; 13.6-25.5% in Africans; and 2.7-22% in African-Americans (Garte et al. 2001). CYP1A1*2A and CYP1A1*2B are in linkage disequilibrium. Some polymorphisms in the mRNA and protein coding sequences of CYP1A1 have been hypothesized to be responsible for interindividual differences in the susceptibility to chemically induced disease, but have not been extensively studied in terms of both functional and clinical consequences. Of the polymorphisms reported in the coding region of CYP1A1, only a few have measurable functional consequences (Han, Pentecost & Spivack 2003). Han, Pentecost and Spivack (2003) examined 1.5kb of the promoter region of *CYP1A1*. Thirteen single nucleotide polymorphisms (SNP's) were identified, with the majority occurring as multi-SNP combinations in individual patients. The majority of CYP1A1 haplotypes were shown to have no functional effects compared to the wild-type promoter sequences. In contrast, two constructs of composite polymorphisms (C2923T-G2875A-T3777G and C2923T-T3777G-G4335A) appeared to result in a significant increase in basal promoter activity (1.38- and 1.50-fold, respectively).

Table 1.2 *CYP1A1* polymorphisms.

Data were obtained from http://www.imm.ki.se/CYPalleles/.

Allele	Protein	Nucleotide Changes	Trivial Name	Effect
CYP1A1*1A	CYP1A1.1	None	Wild-type	-
CYP1A1*1B	CYP1A1.1	C-3219T	-	-
CYP1A1*1C	CYP1A1.1	G-3229A	-	-
[a]**CYP1A1*2A**	CYP1A1.1	T3801C (MspI)	m1	-
[a]CYP1A1*2B	CYP1A1.2	A2455G; T3801C (MspI)	-	I462V
CYP1A1*2C	CYP1A1.2	A2455G	m2	I462V
CYP1A1*3	CYP1A1.1	T3205C	m3	-
CYP1A1*4	CYP1A1.4	C2453A	m4	T461N
CYP1A1*5	CYP1A1.5	C2461A	-	R464S
CYP1A1*6	CYP1A1.6	G1636T	-	M331I
CYP1A1*7		2346_2347 T insertion	-	Frame shift
CYP1A1*8	CYP1A1.8	T2414A	-	I448N
CYP1A1*9	CYP1A1.9	C2461T	-	R464C
CYP1A1*10	CYP1A1.10	C2500T	-	R477W
CYP1A1*11	CYP1A1.11	C2546G	-	P492R

[a]Denotes linkage disequilibrium.

1.3.5.3 Regulation of human CYP1A1

CYP1A1 (along with CYP1A2 and CYP1B1) is regulated by the aromatic hydrocarbon receptor (AhR). The AhR is a ligand activated transcription factor that mediates a toxic response toward specific chemical pollutants, including PAH's and polychlorinated dioxins, most notably benzo[a]pyrene (BaP), 3-methylcholanthrene (3-MC), 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), and β-naphthoflavone (βNF) (Denison & Nagy 2003; Nebert et al. 2004; Shimada & Fujii-Kuriyama 2004). The AhR is a member of the basic helix loop, helix (bHLH/PAS) protein superfamily (Galijatovic et al. 2004), whose members play an important part in facilitating intercellular signaling via sensory pathways (Roblin, Okey & Harper 2004). The AhR is a critical mediator of a cell signaling system, whereby its activation gives rise to the altered regulation of numerous genes. AhR is a soluble cytosolic protein which forms a complex with the chaperone proteins hsp90 and hsp23 and an immunophillin-like protein (involved in its immunosuppressive behavior) (Figure 1.8). Upon ligand activation, the AhR translocates to the nucleus, dissociates from the hsp proteins, and forms a heterodimer with another bHLH/PAS protein, the AhR nuclear translocator (ARNT). The ligand-AhR-ARNT complex then interacts with a 5′-GCGCT-3′ DNA core binding motif, commonly termed the xenobiotic (or drug) response element (XRE; DRE), present in multiple copies upstream of the *CYP1A1* gene promoter (Le Ferrec et al. 2002). The DNA bound AhR-ARNT dimer then recruits cofactors, allowing the complex to regulate expression of *CYP1A1* (Levine-Fridman, Chen & Elferink 2004; Roblin, Okey & Harper 2004). Interestingly, the human *CYP1A1* gene also contains a negative regulatory domain about -800 bases from the transcriptional start site (Galijatovic et al. 2004). This is not evident in other species.

Research conducted by Harper *et al* (2004) shows that CYP1A2 and CYP1B1 are constitutively expressed in the liver of AhR-null mice. In addition, CYP1A2 and CYP1B1 are induced in AhR-null mice by phenobarbitone, indicating that the expression of CYP1A2 and CYP1B1 can be regulated by factors other than the AhR (e.g. the constitutive androstane receptor, CAR). In contrast, CYP1A1 expression appears to be highly dependent on the AhR. Transfection of the full-length human AhR cDNA into BP8 rat hepatoma cells (AhR deficient) was sufficient to induce substantial CYP1A1 mRNA levels without the presence of an exogenous AhR ligand. This was achieved using the human elongation factor 1α-subunit promoter (hEF-1α). Additionally, the AhR antagonist, 3,4-dimethoxyflavone, is known to inhibit CYP1A1 expression in a concentration dependent manner (Roblin, Okey & Harper 2004).

The AhR transcription factor is additionally involved in cell cycle regulation (Levine-Fridman, Chen & Elferink 2004). This suggests that in the absence of an exogenous ligand, the AhR functions to promote cell growth. Studies with 5L rat hepatoma cells (AhR-positive) demonstrated that TCDD induces a G_1 phase cell cycle arrest not detected in BP8 cells. Other findings suggest that some endogenous AhR agonists may also be substrates for the CYP1A1 enzyme (Chang & Puga 1998). CYP1A1-mediated depletion of the endogenous AhR agonist therefore creates a negative feedback mechanism which suppresses prolonged AhR activity under normal physiological conditions. Evidence presented by Levine-Fridman, Chen and Elferink (2004) further suggests that the duration of AhR activity can dramatically impact on the cell cycle response to growth factors and other extra-cellular signals (consistent with the hypothesis that the AhR functions as a regulator of cell cycle progression through G_1 phase) and that growth factor-stimulated CYP1A1 induction

during the G1-to-S phase transition inactivates the AhR by rapidly depleting the endogenous ligand. In addition, failure to inhibit AhR activity resulted in cell cycle arrest.

Differences in the susceptibility of individuals to the adverse actions of PAH's may be due to differing CYP1A1 expression and genetic variations in *CYP1A1*. The same can be said for interindividual differences in the levels of expression of the AhR, and the occurrence of AhR polymorphism. In addition to the multiplicity of genotypic transcriptional malfunctions, further exogenous compounds can account for the disruption of post-transcriptional inhibition. For example, fluasterone (16α-fluoro-5-androsten-17-one), a synthetic analogue of the chemo-preventative hormone dehydroepiandrosterone (DHEA), has been shown to inhibit the rate of CYP1A1 promoter-controlled transcription. Fluasterone moderately inhibited TCDD-induced transcription of CYP1A1 by hindering the XRE binding motif, as measured by ethoxyresorufin *O*-deethylase (EROD) activity. In the presence of the RNA polymerase II inhibitor actinomycin D, fluasterone caused an increase in the degradation of CYP1A1 mRNA (Ciolino, MacDonald & Yeh 2002). In contrast, CYP1B1 mRNA remained unaffected, further indicating that the expression of the CYP1A1 gene is regulated by unique factors.
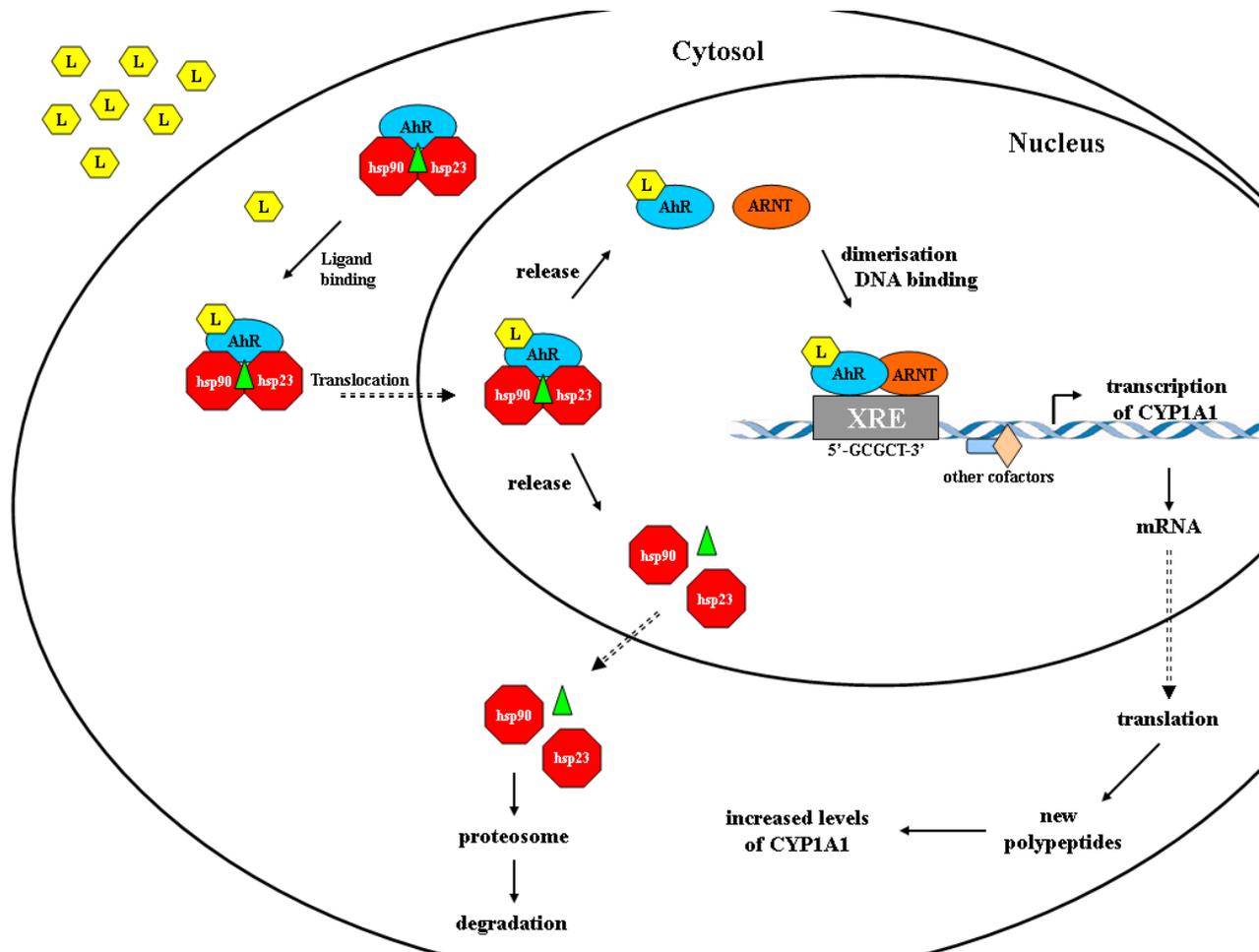
Figure 1.8 The molecular mechanism of activation of gene expression by the Aromatic Hydrocarbon Receptor (AhR).

L; ligand, hsp90; heat shock protein 90, hsp23; heat shock protein 23, ARNT; AhR nuclear translocator, XRE; xenobiotic response element.

1.3.5.4 Inhibition of human CYP1A1

Many chemicals including α-naphthoflavone, 7-hydroxyflavone, synthetic chemo-preventative organoselenium compounds such as 1,2-, 1,3-, and 1,4-phenylenebis(methylene)selenocyanate, vinylic and acetylenic PAH's such as 1-(1-propynyl)pyrene, and 4-(1-propynyl)biphenyl have been identified as inhibitors of CYP1A1 (Appendix 1, Table A1.1). Recently, Chun, Kim and Guengerich (1999) reported that the natural products resveratrol and rhapontigenin are also inhibitors of human CYP1A1. For example, the IC_{50} value for rhapontigenin (3,3′,5-trihydroxy-4′methyl-thoxystilbene) inhibition of CYP1A1 was 0.4 μM. As a mechanism-based inhibitor, rhapontigenin showed a 400-fold selectivity for CYP1A1 over CYP1A2 and a 23-fold selectivity over CYP1B1. In addition, rhapontigenin did not show any significant inhibition of EROD activity in human liver microsomes which contain negligible amounts of CYP1A1 (Chun et al. 2001). With respect to resveratrol (*trans*-3,4′,5-trihydrostilbene), there was a 50-fold selectivity in its inhibition of CYP1A1 over CYP1A2, with an IC_{50} value of 23μM. Prior to this study, resveratrol was perceived to be a potential cancer chemo-preventative agent due to its ability to inhibit oxidation/reduction enzymes and induce conjugation enzymes. However, the mechanism-based inactivation of CYP1A1 with rhapontigenin suggests that this compound would be a more potent and selective chemo-preventative agent.

1.3.6 Heterologous expression of human P450 enzymes

1.3.6.1 Background

The heterologous expression of recombinant proteins provides a means of obtaining higher yields of a target protein compared to those levels found endogenously in cells. Mammalian expression systems are routinely used for a variety of recombinant proteins, however the over-expression of P450s in mammalian cell lines has been less successful (McManus et al. 1990). Greater success has been achieved using yeast (Ikushiro et al. 2004), insect (Ong et al. 1998), and bacterial expression systems, with bacteria being the simplest and most cost effective (Boye et al. 2004; Gonzalez & Korzekwa 1995). Bacteria also has the advantage of accommodating foreign genetic material (Fujita & Kamataki 2002) and producing very high yields of protein in a relatively short amount of time due to competent cell replication (Baneyx & Mujacic 2004). However, the production of functional proteins in *E. coli*, especially from eukaryotic origin, can be problematic due to differences in codon usage and the formation of inclusion bodies (Jana & Deb 2005). Table 1.3 summarizes some of the molecular tools used to increase expression yields of functional recombinant proteins.

Table 1.3 Molecular tools and techniques used for optimizing heterologous protein expression in *Escherichia coli.* Modified from Jana & Deb (2005).

Component	Remarks
Host strain	Choice of host strain can impact expression via the host's genotype. Sufficient cellular machinery is required for correct protein translation and folding.
Plasmid copy number	Gene dosage, as manipulated through plasmid copy number, can greatly affect protein expression.
Antibiotic selection	Choice of antibiotic resistance on the expression plasmid can influence heterologous expression.
Promoter	Strong or weak, inducible or constitutive; promoter and regulation are major influences on protein expression, which is also affected by relative orientation and strength of promoters on the plasmid.
Transcription termination	Effectiveness and spacing of transcription terminators affect expression.
mRNA stability	The stability of the mRNA affects yield. Secondary structure at the 5' end of the message often plays a critical role.
Translation signal	The ribosome binding-site affects the amount of ribosome loading and clearance, and hence, expression. Secondary structure at the 5' end can affect the accessibility of the ribosome binding-site.
Codon usage (*E. coli*)	Numerous differences occur between human and bacterial codon usage. Utilizing the optimal *E. coli* codons often improves yield.
Signal peptide	An amino signal peptide, used for targeting a protein to a specific region of the bacterial cell, can greatly enhance the expression yield through protein folding and stability.
Temperature	Temperature has a profound affect on protein folding and stability. Improves yields of soluble proteins.
Growth conditions	Growth conditions, oxygen levels, growth rate, carbon source and fermentor configuration all affect yield.
Media and supplements	Supplementation with rare elements can aid the correct folding of recombinant proteins and stabilization during transport across membranes.

1.3.6.2 Expression of modified P450s in Escherichia coli

Unlike most eukaryotes and many bacteria, enterobacteria such as *E. coli* lack P450s, which makes them a useful organism for recombinant expression. Successful synthesis of P450s in *E. coli* requires the nucleotide CDS of the desired protein to be cloned downstream of an efficient promoter and ribosome binding-site in a high copy number plasmid. Transcriptional activation of the promoter gives rise to mRNA molecules that await ribosome-mRNA complexation for efficient translation. As described by Barnes (1996), the nascent polypeptide then associates with the *E. coli* membrane, binds a molecule of heme, and folds into its final tertiary structure.

The bacterial plasmid of choice for many laboratories expressing P450s is pCW ori(+) (Figure 1.9), a derivative of the pHSe5 plasmid backbone. pCW ori(+) contains a pBR322 origin of replication, the *lac* I^q gene that encodes a repressor protein which binds to the *lac* operator and blocks transcription down-stream, the β-lactamase gene for conferring ampicillin resistance, and a bacterophage origin of replication. Transcription and translation of pCW ori(+) is controlled by a *lac*UV5 promoter and two copies of the *tac* promoter cassette with translation initiation coming from the phage T4 lysozyme gene. A trpA transcription termination cassette is located down-stream of the multiple cloning site (MCS). Protein synthesis is induced with isopropyl β-D-thiogalactopyranoside (IPTG), a molecular mimic of allolactose, a lactose metabolite that triggers transcription of the *lac* operon. The modified CheW chemotaxis gene of *E. coli* and the β-galactosidase gene aid insert selection.
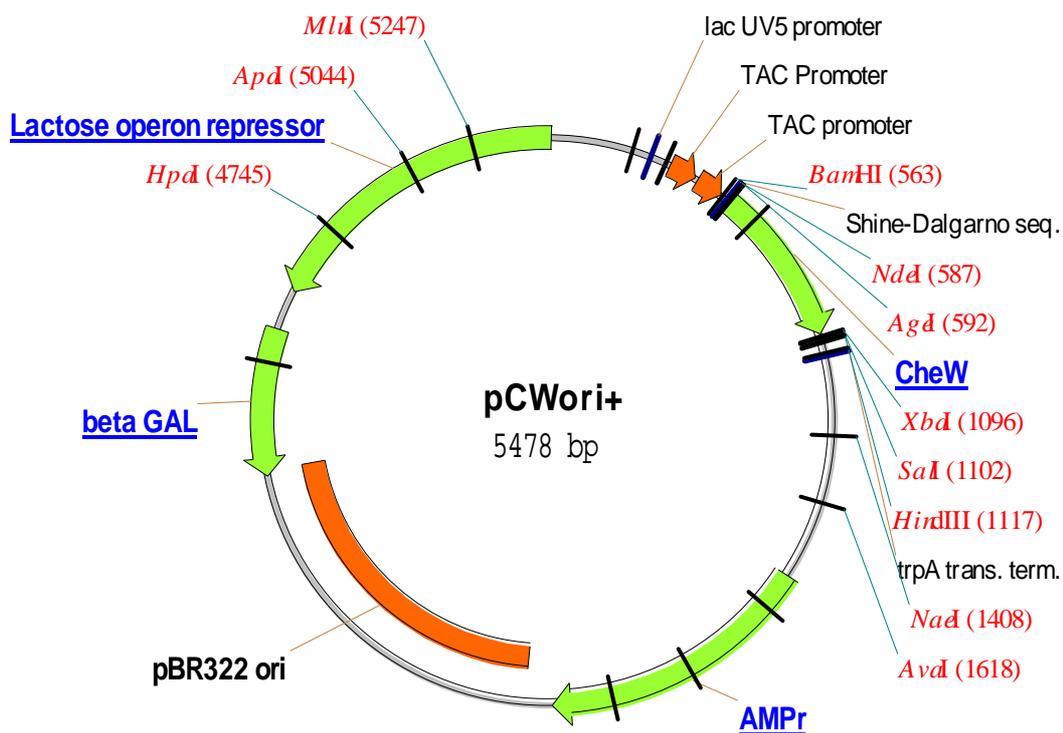
Figure 1.9 Plasmid map of the pCW ori(+) bacterial expression plasmid. Elements are drawn to scale.

Due to the fact that all class II P450s bind to the ER membrane, truncation of the 5' terminus of the CDS to remove the hydrophobic transmembrane domain typically increases the yield of many P450s in bacteria. This has been observed by several investigators, primarily utilizing CYP2C family proteins (Boye et al. 2004; Guo et al. 1994; Iwata et al. 1998; Sandhu, Baba & Guengerich 1993). Additionally, the NH_2-terminus can be modified to target expression of a P450 to a desired *E. coli* organelle (Luirink et al. 2005), which in turn decreases the likelihood of secondary-structure formation of message around the ribosome assembly site (and the ATG). Table 1.4 lists several signal sequences used for targeting message to different locations within the gram-negative bacterial cell (Figure 1.10). In addition, the leader sequence

obtained from the bovine microsomal 17α-hydroxylase cytochrome P450, P45017A is required. In an attempt to optimize bacterial expression, Barnes, Arlotto and Waterman (1991) showed that modification to the first seven residues of P45017A greatly enhanced holoenzyme formation. Specifically, the second codon was changed from TGG (Trp) to GCT (Ala), the preferred second codon for expression of the bacterial *lac*Z gene. In addition, amending the DNA CDS of residues 4, 5, 6 and 7 of P45017A to utilize the bacterial codon usage reduced the Gibbs free energy (ΔG) of mRNA self-hybridization. Use of the P45017A-signal peptide (17α-leader sequence) has a profound effect on the yield of functional P45017A in *E. coli* and has subsequently been used by numerous groups to successfully express large amounts of mammalian P450s in *E. coli* (Boye et al. 2004; Fisher et al. 1992; Gillam et al. 1997; Lewis, Mackenzie & Miners 2007; Pritchard et al. 1997; Richardson et al. 1995; Shet et al. 1994).

The heterologous expression of P450s in *E. coli* is usually conducted in K-12 strains where heme is impermeable to the gram-negative phenotype. This makes holoenzyme P450 expression somewhat limited in addition to the fact that heme is potentially toxic to the cells (Harnastai, Gilep & Usanov 2006). The ability of most bacteria to produce the heme precursor 5-aminolevulinic acid (δ-ALA) via ALA synthase (C-4 path) is not shared by *E. coli*, which synthesize δ-ALA from glutamate (C-5 path) (Avissar & Beale 1989). Since cellular heme content plays a role in the regulation of δ-ALA production, endogenous heme biosynthesis is tightly controlled with almost no free heme present under normal culture conditions (Woodard & Dailey 1995). Moreover, heme has a relatively low affinity for the nascent P450 polypeptide. To combat this problem bacterial cultures are commonly supplemented with exogenous δ-ALA.

Table 1.4 Common signal peptides used in the expression of eukaryotic proteins in bacteria. Basic amino acids, blue; hydrophobic amino acids, yellow. Red stars represent cleavage sites that mark the end of each signal peptide. Refer to Figure 1.10 for a schematic representation of gram-negative bacteria.

Inner Membrane Proteins	Signal Peptide
Phage fd, major coat protein	M**KK**SLVL**K**ASVAVATLVPMLSFA*****AE
Phage fd, minor coat protein	M**KK**LLFAIPLVVPFYSHS*****AE

Periplasmic Proteins	Signal Peptide
Alkaline phosphatase	M**K**QSTIALALLPLLFTPVTKA*****RT
Leucine-specific binding protein	M**K**ANA**K**TIIAGMIALAISHTAMA*****DD
β-lactamase of pBR322	MSIQ**HFR**VALIPFFAAFCLPVFA*****HP

Outer Membrane Proteins	Signal Peptide
Lipoprotein	M**K**AT**K**LVLGAVILGSTLLAG*****CS
LamB	L**RK**LPLAVAVAAGVMSAQAMA*****VD
OmpA	MMITM**KK**TAIAIAVALAGFATVAQA*****AP

In the bacterial heterologous system used for P450s, it is not uncommon for overexpressed proteins to form insoluble aggregates called 'inclusion bodies'. Inclusion bodies usually result from proteins that fail to fold correctly and tend to accumulate in the cytoplasm or periplasm (Figure 1.10) depending on the signal peptide used (Table 1.4). The recombinant protein usually accounts for 80-95% of the inclusion body and is contaminated with outer membrane proteins, ribosomal components, and a small amount of phospholipids and nucleic acids (Baneyx &

Mujacic 2004). In the case of P450s, heme incorporation is critical for obtaining the

correct folding conformation.
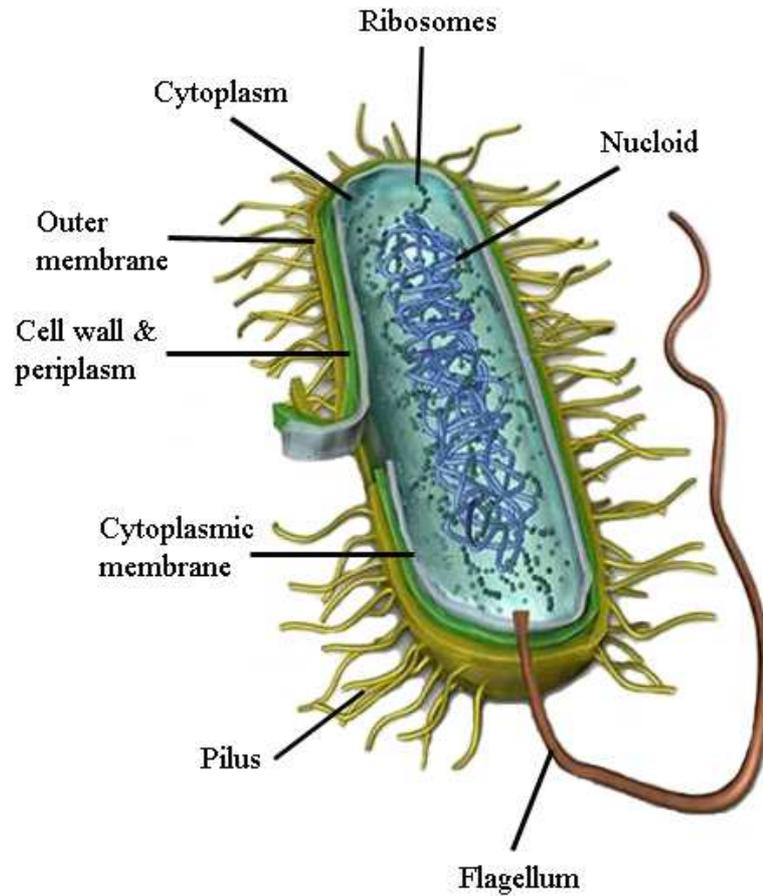
Figure 1.10 Schematic representation of the structural organisation of gram-negative bacteria. Amended from Pugsley (1993).

Once successful expression of P450 has been achieved, membranous extracts usually undergo spectral analysis to determine the functional holoenzyme concentration. This is achieved by reducing the protein preparation containing the oxidized P450 with sodium dithionite ($Na_2S_2O_4$) followed by exposure to $CO_{(g)}$ to uncouple the P450 catalytic cycle (Figure 1.2). The initial oxidized P450 present in the native protein preparation gives a Soret band at 414nm (extinction coefficient; 124mM^{-1}cm^{-1}). Reduction of the heme Fe^{3+} to Fe^{2+} shifts the Soret band to 427nm (149mM^{-1}cm^{-1}). Treatment with $CO_{(g)}$ also displaces any $O_{2(g)}$ bound to the Fe^{2+} allowing the absorption spectra to be determined for both apo- and holo-enzyme forms (Figure 1.11). In the reduced apo- form bound with $CO_{(g)}$, a shift in the Soret band to 420nm is observed along with a marked increase in absorbance intensity (213mM^{-1}cm^{-1}). The reduced P450 with bound $CO_{(g)}$ gives rise to an absorbance peak at 450nm (91mM^{-1}cm^{-1}) characteristic of the holo- form of the enzyme (Miyake & Takayama 1975; Omura & Sato 1964a; Omura & Sato 1964b).

To generate an active system *in vitro*, P450s require the donation of electrons, utilized for reducing the heme Fe^{3+}, from the redox partner (section 1.3.1). Bacteria, however, do not contain a suitable endogenous redox partner, hence, the heterologous expression of P450s additionally requires expression of OxR. Dual expression may be achieved by a number of methods: (i) a CYP-OxR fusion construct (Deeni et al. 2001; Fisher et al. 1992; Fisher, Shet & Estabrook 1996; Shet et al. 1994; Shet et al. 1993), (ii) a bicistronic plasmid (Boye et al. 2004; Gillam et al. 1997; Parikh, Gillam & Guengerich 1997), or (iii) the simultaneous co-transformation of different P450 and OxR plasmids (Dong & Porter 1996; Fujita & Kamataki 2002; Iwata et al. 1998; Lewis, Mackenzie & Miners 2007; Pritchard et al. 1998). It should be noted that mixing of bacterial membrane preparations generated

from separate P450 and OxR expressions can reconstitute enzymatic activity (Boye et al. 2004). However, the generation of an active mixed-membrane system is not successful with those P450 enzymes that contain a hydrophobic signal peptide, but is feasible for those P450s expressed with a truncated NH_2-terminus (e.g. CYP2C9 Δ3-20 truncation) (unpublished observation). It appears that the more hydrophilic nature of the truncated P450 allows greater flexibility in the bacterial membrane and can therefore align and couple more freely with the supplemented OxR.
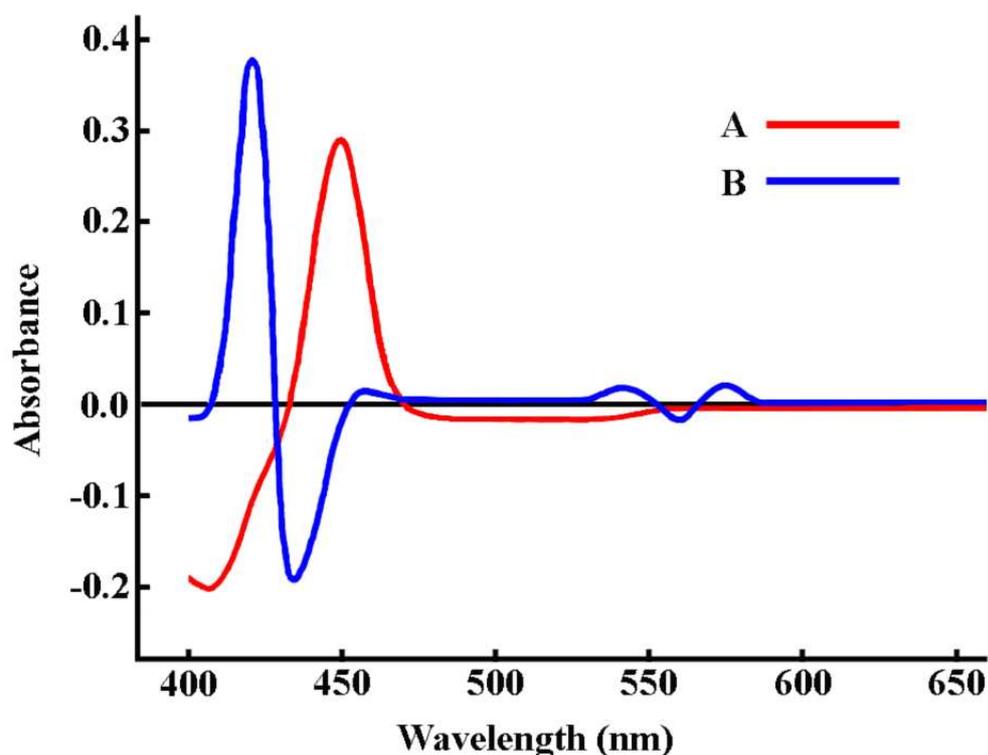
Figure 1.11 P450 CO difference spectra. Spectrum A (red) shows the formation of the ferrous Fe^{2+}-CO complex of a typical holo-P450 after reduction with sodium dithionate and bubbling with $CO_{(g)}$ (450 peak). Spectrum B (blue) shows the inactive reduced apo-enzyme peak at 420nm with a deep trough at 435nm, two small peaks at 540nm and 575nm, and a trough at 560nm. The P420 species has a substantially larger Soret molar absorption coefficient than the holo- form.

Like P450, the measurement of OxR concentration is equally important and is commonly attained by the method outlined by Yasukochi & Masters (1976). NADPH-dependent reduction of cytochrome *c* is monitored by the increase of absorbance at 550nm. The rate of reduction of cytochrome *c* (nmol/mg protein/min) is used as a measure of OxR activity in expressed membrane fractions. The OxR:P450 ratio can be critical for the derivation of kinetic parameters for a drug substrate, whereby an excess of OxR over P450 can enhance catalytic activity (Venkatakrishnan et al. 2000; Yamazaki et al. 1999). Additionally, many P450-dependant reactions have been shown to be stimulated by cytochrome b_5. This possibly occurs via direct electron transfer to the P450 from cytochrome b_5 (holo-b_5) or through a conformational effect on the P450 in the absence of electron transfer (apo-b_5) (Yamazaki et al. 1997; Yamazaki et al. 2001).

1.3.6.3 Recombinant expression of human CYP1A1

Guo et al. (1994) were the first to express the full-length cDNA encoding human CYP1A1, along with several modified constructs, including one carrying the 17α-leader sequence, in *E. coli*. Little expression was observed with the native sequence and several modified constructs, but successful expression (20-25nmol P450 L^{-1} culture) was achieved with a construct in which the Ala codon GCT was placed in the second position and the 5′-terminal codons were maximized for AT content and minimized for potential secondary structure formation of the message transcript. α-Naphthoflavone was found to protect against denaturation by detergents during the solubilization of membranes and was therefore added to all buffers used for purification. The recombinant CYP1A1 catalyzed 7-ethoxyresorufin *O*-deethylation and benzo[a]pyrene 3-hydroxylation with K_m (V_{max}) values of 0.58μM (8.3 nmol

min^{-1}/nmol P450) and 15μM (2.5 nmol min^{-1}/nmol P450), respectively. To further optimize expression, Chun, Shimada and Guengerich (1996) utilized the pCW ori(+) plasmid to code for a fusion protein consisting of the complete CYP1A1 coding sequence (with second amino acid change, Ala) connected by a Ser-Thr linker to rat OxR (beginning at amino acid 57). Expression was typically 150nmol P450 L^{-1} culture, similar to the levels observed with CYP1B1 (Shimada et al. 1998). The favorable difference in CYP1A1 expression compared to that of Guo *et al* (1994) was thought to be due to the addition of the heme precursor δ-ALA to cultures.

The purified fusion protein was shown to catalyze benzo[a]pyrene 3-hydroxylation, 7-ethoxyresorufin *O*-deethylation, and zoxazolamine 6-hydroxylation according to Michaelis-Menten kinetics. However, the catalytic activity of the fusion protein was markedly lower than that of a similar fusion protein constructed with rat CYP1A1. This was thought to reflect the inherently lower activity of human CYP1A1 compared to the rat enzyme, reflecting structural differences arising from the differing coding sequences in each species (Figure 1.12). Notably, no increase in turnover was observed when purified OxR or cytochrome b_5 was added. Furthermore, the authors found that the catalytic activity of the enzyme was considerably higher within purified fractions in contrast to enzyme still localized in membranes.

```
                    1                                                       50
human CYP1A1    M....LFPIS MSATEFLLAS VIFCLVFWVI RASRPQVPKG KNPPGPWGW
  rat CYP1A1    MPSVYGFPAF TSATELLLAV TTFCLGFWVV RVTRTWVPKG KSPPGPWGL
                    51                                                     100
human CYP1A1    PLIGHMLTLG KNPHLALSRM SQQYGDVLQI RIGSTPVVVL GLDTIRQAL
  rat CYP1A1    PFIGHVLTLG KNPHLSLTKL SQQYGDVLQI RIGSTPVVVL GLNTIKQAL
                    101                                                    150
human CYP1A1    VRQGDDFKGR PDLYTFTLIS NGQSMSFSPD SGPVWAARRR AQNGLKSFS
  rat CYP1A1    VKQGDDFKGR PDLYSFTLIA NGQSMTFNPD SGPLWAARRR AQNALKSFS
                    151                                                    200
human CYP1A1    IASDPASSTS CYLEEHVSKE AEVLISTLQE LMAGPGHFNP RYVVVSVTN
  rat CYP1A1    IASDPTLASS CYLEEHVSKE AEYLISKFQK LMAEVGHFDP KYLVVSVAN
                    201                                                    250
human CYP1A1    VICAICFGRR YDHNHQELLS LVNLNNNFGE VVGSGNPADF PILRYLPNP
  rat CYP1A1    VICAICFGRR YDHDDQELLS IVNLSNEFGE VTGSGYPADF PILRYLPNS
                    251                                                    300
human CYP1A1    SLNAFKDLNE KFYSFMQKMV KEHYKTFEKG HIRDITDSLI HCQEKQLDE
  rat CYP1A1    SLDAFKDLNK KFYSFMKKLI KEHYRTFEKG HIRDITDSLI HCQDRRLDE
                    301                                                    350
human CYP1A1    NANVQLSDEK IINIVLDLFG AGFDTVTTAI SWSLMYLVMN RVQRKIQEE
  rat CYP1A1    NANVQLSDDK VITIVFDLFG AGFDTITTAI SWSLMYLVTN RIQRKIQEE
                    351                                                    400
human CYP1A1    LDTVIGRSRR PRLSDRSHLP YMEAFILETF RHSSFVPFTI HSTTRDTSL
  rat CYP1A1    LDTVIGRDRQ PRLSDRPQLP YLEAFILETF RHSSFVPFTI HSTIRDTSL
                    401                                                    450
human CYP1A1    KGFYIPKGRC VFVNQWQINH DQKLWVNPSE FLPERFLTPD AIDKVLSEK
  rat CYP1A1    NGFYIPKGHC VFVNQWQVNH DQELWGDPNE FRPERFLTSS TLDKHLSEK
                    451                                                    500
human CYP1A1    VIIFGMGKRK CIGETIARWE VFLFLAILLQ RVEFSVPLGV VDMTPIYGL
  rat CYP1A1    VILFGLGKRK CIGETIGRLE VFLFLAILLQ QMEFNVSPGE VDMTPAYGL
                    501                          524
human CYP1A1    TMKHACCEHF QMQLRS.... ....
  rat CYP1A1    TLKHARCEHF QVQMRSSGPQ HLQA
```

Figure 1.12 Sequence alignment of human CYP1A1 and rat CYP1A1. Alignment similarity is 85.1% (red and green) while identity is 77.7% (red only).

1.4 Cytochromes P450 and neoplastic disease

1.4.1 Metabolic conversion of anti-cancer drugs

P450s are responsible for the metabolism of most prescribed cancer chemotherapeutic agents (Guengerich et al. 2000) by generating inactive or metabolically activated products, or both. While P450 enzymes are predominantly located in the liver, there is substantial evidence that individual P450s are also expressed in other tissues such as the brain, lung, kidney, and the gastrointestinal tract (Guengerich 2000; Murray et al. 1992; Patterson & Murray 2002; Rooney et al. 2004). In addition, there is also evidence showing P450 expression in resected human tumors. Of particular interest is expression in a variety of solid tumors including breast, colon, lung, esophagus, ovarian, bladder, prostate, stomach, and soft tissue sarcomas (Lord, Bongiovanni & Bralley 2002; McFadyen, Melvin & Murray 2004; Miyoshi et al. 2002; Rooney et al. 2004). Indeed CYP1A1, CYP1B1, and CYP3A4 are known to be over-expressed in tumors (Murray et al. 1992; Patterson & Murray 2002; Rooney et al. 2004; Schwartz, Chen & Waxman 2002) and may influence the fate of certain cytotoxic drugs within the tumor.

1.4.2 Cytochrome P450-based prodrug cancer therapy

One promising area for the improvement of tumor/drug selectivity, and one which escapes the problems associated with inter-individual variation, is Gene-Directed Enzyme Prodrug Therapy (GDEPT) (Aghi, Hochberg & Breakefield 2000; Cass et al. 2001; Daly 2003; Denny 2002; Hughes et al. 2002; Johnson 2003; Xu & McLeod 2001). GDEPT is a two step approach. In the first step a drug-activating enzyme is targeted and expressed in the tumor tissue. The gene(s) encoding the enzyme(s)

should be of either non-human origin or human protein that is deficient or expressed in low concentrations in normal tissue (Denny 2002). The protein must achieve sufficient expression in the tumors and have high catalytic activity (Daly 2003). In the second step a non-toxic prodrug, which is a substrate of the exogenous enzyme that is now expressed in the tumor, is administered systemically (Xu & McLeod 2001). The net gain is that the prodrug can be activated within the tumor, producing high local concentrations of the active moiety.

Although GDEPT is susceptible to the same technical challenges as other areas of medicine utilizing gene therapy, it has several advantages over other cancer gene therapies due to its ability to not only reduce systemic exposure, but also transiently replace a polymorphic gene that may confer a 'poor (or extensive) metabolizer' phenotype. In addition, GDEPT is known to produce a 'bystander effect', which extends the cytotoxic response beyond those cells transfected with the prodrug-activating P450. A limitation of GDEPT, however, is that only a small proportion of tumor cells become activated using current enzyme prodrug combinations. In an effort to overcome this problem, the approach of many researchers is to generate new activation competent drugs, i.e. prodrugs that are efficiently activated and effective at killing neighboring cells via a bystander effect (Xu & McLeod 2001). Since recombinant DNA technology can be utilized in GDEPT, an alternative solution is to design specific P450 enzymes with enhanced prodrug activating abilities. Thus, mutant forms of human enzymes can be generated by site-directed mutagenesis and, in doing so, avoid the expected immune response seen with the use of non-human proteins [e.g. rat CYP2B1 (Kan, Kingsman & Naylor 2002)].

Much remains unknown about the use of P450 enzymes in GDEPT, with the majority of questions relating to the design of novel substrates. An increased

understanding of the metabolism of antineoplastic agents within the body is also essential in order to develop more highly specialized and directed therapies. As such, enhanced tumor mediated metabolism by P450s requires further investigation to utilize the specific activation of prodrugs by enzymes, whose expression, or kinetic activity is limited to, or can be enhanced in, cancer cells. GDEPT utilizing P450's may offer the possibility of a truly targeted chemotherapy so long as overproduction of the toxic metabolite is constrained to the tumor environment, thereby minimizing systemic toxicity.

1.5 Structure and function of the mammalian cytochromes P450

1.5.1 Background

Structure-activity analyses in P450s have been essential in understanding the mechanisms responsible for substrate selectivity, regioselectivity, stereoselectivity, and substrate orientation (Lewis 2003; Lewis, Modi & Dickins 2002; Lewis, Lake & Dickins 2004; Miles et al. 2000; Ridderstrom et al. 2001). Using various experimental and analytical approaches, links between structure and function have elucidated numerous relationships, from the substrate recognition sites (SRS) proposed by Gotoh (1992) to the interactions affecting flexibility of the heme prosthetic group by Hudecek et al. (2007). Technical advances in structural and computational techniques, such as nuclear magnetic resonance (NMR), X-ray diffraction, and computational modeling, have substantially increased the P450 knowledge base at the molecular level. Of further importance was the development of site-directed mutagenesis, which allows the targeted substitution of active-site

amino acids and hence a better understanding of the chemistry of substrate access, catalysis, product egress, and enzyme inhibition.

1.5.2 Crystal structure determination

Due to the large molecular size of P450s, application of NMR for structure determination has been somewhat limited. In contrast, the use of X-ray diffraction has proved to be a more powerful method for elucidating the structures of P450s, including those with bound substrate. The generation of mammalian P450 structures has been challenging largely due to the difficulty in obtaining large quantities of protein in a homogeneous state for growing diffraction quality crystals (Zhao & Halpert 2007). The first P450 crystal structure solved was that of CYP101 (P450$_{cam}$) from *Pseudomonos putida* in 1985 (Poulos et al. 1985). The structure of CYP101 was bound with the bicyclic terpene, camphor, which was located in a buried pocket adjacent to the oxygen binding site of the heme. The CYP101 structure was refined to an R-value (statistical measure of fit) of 0.23 at a resolution of 2.6Å. Availability of the CYP101 crystal structure was particularly useful for resolving the topology and secondary structure in the region surrounding the heme. Furthermore, elucidation of the CYP101 crystal structure was an important milestone allowing researchers to make comparisons with the eukaryotic P450s. However, extrapolation of the CYP101 structural data to eukaryotic P450s proved difficult, primarily due to the membrane binding of mammalian P450s and the limited sequence identity among P450 enzymes (generally 10-30%) (Lewis 2002a). In addition, as a class I P450 the residues responsible for the alignment of CYP101 and its redox partner (FAD) are

markedly different to that of the mammalian class II proteins (FAD-FMN; section 1.3.1).

Hasemann et al. (1995) proposed that regions of functional significance among P450s should be conserved at the secondary structure level despite differences in the primary sequence. To this end, they published structures of the class I P450 enzyme CYP108 (P450$_{terp}$; *Pseudomonas sp.*) and the first structure of a soluble class II P450 enzyme, CYP505 (P450$_{BM-3}$; *Bacillus magaterium*). The comparison between all three P450 crystal structures (CYP101, CYP108, and CYP505) revealed a tertiary arrangement with a highly conserved core of secondary-structure elements. The substrate recognition regions exhibited the greatest differences, with large shifts in the B', F and G helices (Figure 1.13).

The first mammalian P450 to be crystallized was rabbit CYP2C5 (Williams et al. 2000a). CYP2C5 was modified to increase its solubility and aid crystallization. The crystallized form of CYP2C5 differed from the native enzyme by deletion of the transmembrane domain (residues 3-21), in addition to incorporating D2A, Q22K, N23T, G25S, and R26K mutations that were introduced to improve heterologous expression. A further five substitutions were made: N202H, R206E, I207L, S209G, and S210T, to reduce membrane association, phospholipid dependence for catalysis, and protein aggregation (Cosme & Johnson 2000). Cosme and Johnson's (2000) results identified that the residues of the F and G helix, and the F-G loop were responsible for the membrane interaction of CYP2C5 and the aggregation state of truncated CYP2C5 in solution. Structurally, the CYP2C5 crystal revealed that attachment to the ER membrane occurs through a broad hydrophobic surface that neighbors the N-terminal transmembrane anchor, and that orientation of the electrostatic dipole in the membrane bound enzyme appears optimal for maximizing

OxR alignment and electron transfer (section 1.3.1). In addition, the entrance to the putative substrate access channel between the F-G loop and the N-terminal β-sheet system was also shown to be located in the membrane attachment surface (Figure 1.14a).

The CYP2C5 crystal structure shows the enzyme in a closed conformation. It is also clear that significant structural differences occur between the mammalian CYP2C5 compared to the soluble microbial CYP101, CYP108, and CYP505 structures (Figures 1.13 and 1.14), particularly in the active-site. In a subsequent study, Wester et al. (2003a) co-crystallized CYP2C5 with the sulfaphenazole analogue, 4-methyl-*N*-methyl-*N*-(2-phenyl-2*H*-pyrazol-3-yl)benzenesulfonamide (DMZ), at 2.3Å resolution (Figure 1.14b).

Like rabbit CYP2C5, the crystallization of rabbit CYP2B4 was achieved in both the presence and absence of substrate. The initial CYP2B4 structure was obtained by (Scott et al. 2004a) at 1.6Å resolution (R-value 0.217), the most accurate mammalian structure at that time. The crystallized protein differed from the native enzyme by truncation of the N-terminus (3-21) and mutation of residues E2A, G22K, H23K, P24T, K25S, A26S, H27K, and R29K to facilitate protein expression. The structure of CYP2B4 revealed an open cleft, formed by the B', C, F, and G helices, which extends from the protein surface toward the heme Fe. This structure is representative of an open enzyme conformation (Figure 1.15a).
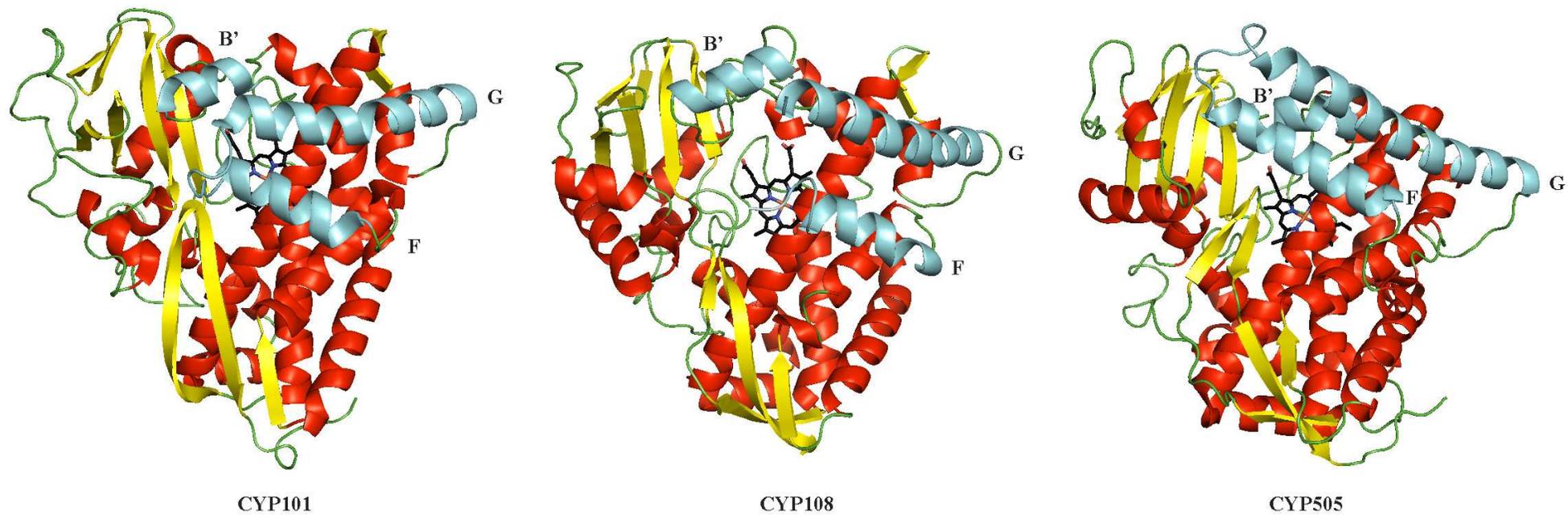
Figure 1.13 Crystal structures of CYP101 (P450$_{cam}$), CYP108 (P450$_{terp}$), and CYP505 (P450$_{BM-3}$). The largest differences occur in the B', F, and G helicies (cyan).
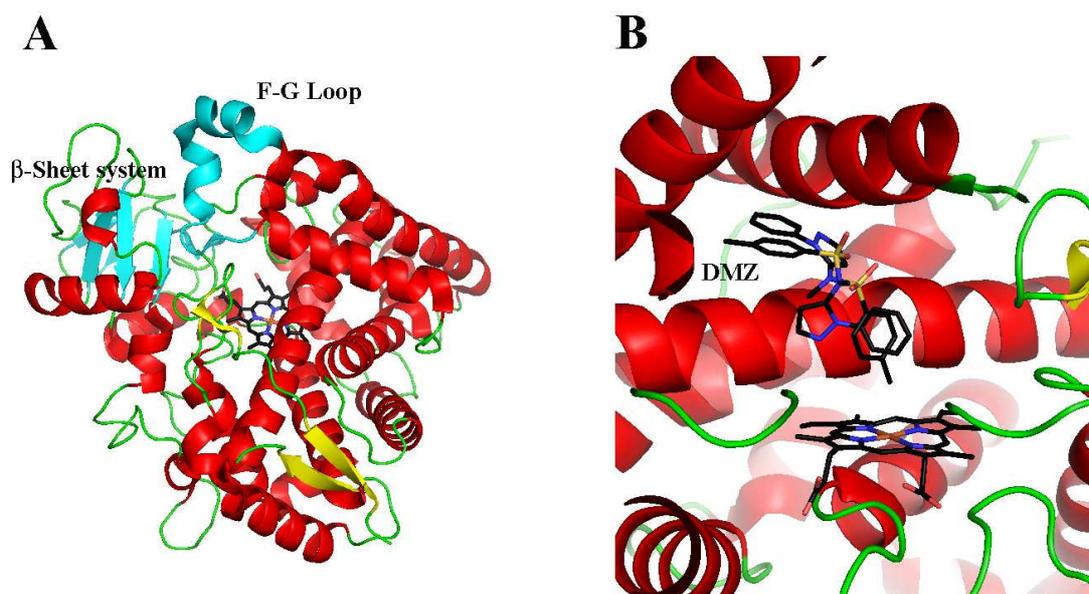
Figure 1.14 The tertiary structure of rabbit CYP2C5 crystalized with and without substrate. (a) The overall fold of rabbit CYP2C5 (pdb; 1NR6) highlighting the F-G loop and the N-terminal β-sheet system shown in cyan. The F-G loop was omitted from the final crystal structure and is subsequently not represented as a loop using the model viewer, PyMOL (seen as a helix). (b) The recrystallized structure (pdb; 1N6B) bound with 4-methyl-N-methyl-N-(2-phenyl-2H-pyrazol-3-yl)benzenesulfonamide (DMZ) in two different catalytic orientations.

Scott et al. (2004a) found that, compared to the CYP2C5 structure, the open cleft in CYP2B4 is generated by the repositioning of highly conserved secondary structures. Interestingly, CYP2B4 crystallized as a dimer, with the F' and G' helices of one molecule filling the open cleft of a second molecule, therefore trapping the enzyme in the open conformation. The structure of CYP2B4 co-crystallized with the specific inhibitor, 4-(4-chloro-phenyl)imidazole (CPI) (1.9Å resolution; R-value 0.215), adopted a closed conformation similar to that seen for mammalian CYP2C5 (Figure 1.15b). The differences between the open and closed structures of CYP2B4 were primarily due to the relocation of the B', C, and F through G helices. Collectively, this was termed the 'lid' domain (Scott et al. 2004b).
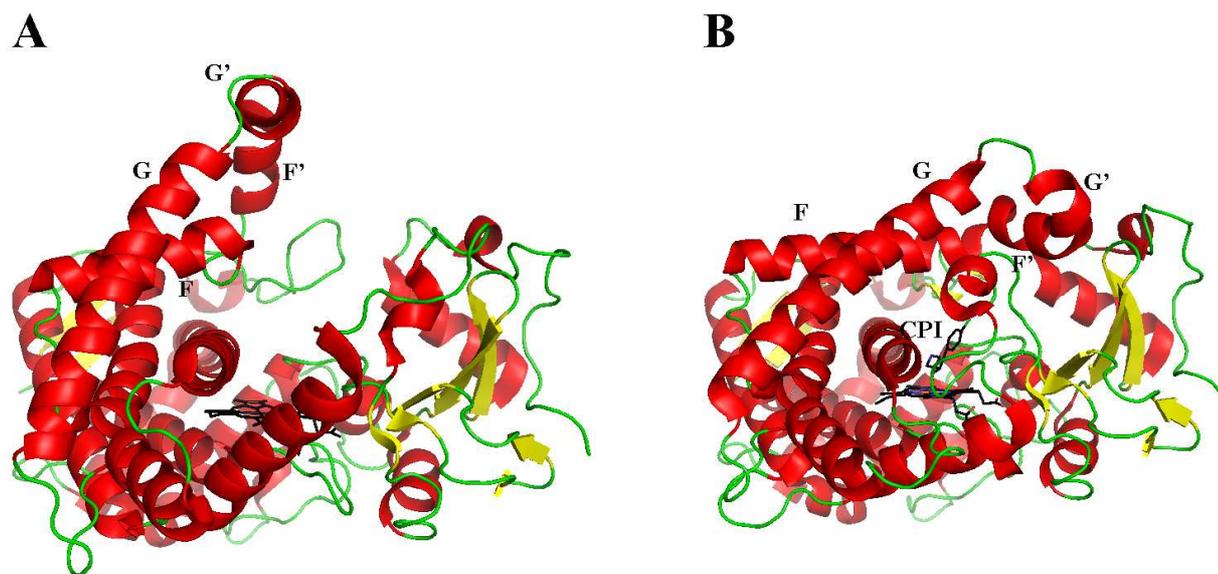
Figure 1.15 The open and closed conformations of rabbit CYP2B4. (a) The open conformation of CYP2B4 (pdb; 1PO5). (b) CYP2B4 (pdb; 1SUO) co-crystallized with 4-(4-chloro-phenyl)imidazole (CPI) highlighting the orgasnization of the F-G loop 'lid'.

Elucidation of the structures of the human drug metabolizing P450s CYP2A6, CYP2C9, CYP2C8, CYP3A4, and CYP2D6 followed in reasonably quick succession. All CDS required truncation of the membrane bound N-terminal domain in addition to mutations necessary for protein solubilization. Three structures of CYP2C9 have been solved; the first structure was unliganded (pdb; 1OG2, 2.6Å) (Figure 1.16a) and the second was co-crystallized with the anti-coagulant drug warfarin (pdb; 1OG5, 2.55Å) (Figure 1.16b) (Williams et al. 2003). The third CYP2C9 structure solved by Wester et al. (2004) showed significant differences compared to the 1OG5 crystal (RMSD 0.74Å). The crystal structure of Wester *et al.* (2003) (pdb; 1R9O, 2Å) (Figures 1.16c and 1.16d) was complexed with flurbiprofen and contained fewer modifications to the CYP2C9 CDS. The 1OG5 structure reported by Williams *et al.* (2003) was essentially CYP2C5-like, with extensive modifications in residues 30-53, 97-121, 196-233, and 467-478. Specifically, the

K206E, I215V, C216Y, S220P, P221A, I223L, and I234L substitutions all corresponded to residues found in CYP2C5. In addition, the location of the bound warfarin in the 1OG5 structure was positioned at the distal end of the active-site cavity with the site of hydroxylation in an unproductive orientation some 10Å from the heme Fe (Figure 1.16b). The refined 1R9O CYP2C9 structure displays the native conformation of the F to G helix region and exhibits an extra turn at the N-terminus of the A helix. In addition, the distinct conformation of the B to C helix region allows R108 to hydrogen bond with D293 and N289 on the I-helix. Furthermore, the 1R9O structure described by Wester *et al.* (2003) highlights the role of R108 and its importance in stabilizing and orienting the binding of the acidic flurbiprofen in the CYP2C9 active-site for productive catalysis (Figure 1.16d).

Like CYP2B4, human CYP2C8 was crystallized as a symmetric dimer formed by the interaction of the F, F', G', and G helices (Schoch et al. 2004). Interestingly, two palmitic acid molecules were bound at the dimer interface, suggesting a peripheral binding site that may contribute to drug interactions during CYP2C8 catalyzed biotransformation (Figure 1.17). The 2.7Å (R-value 0.248) structure of CYP2C8 (pdb; 1PQ2) additionally identified an active-site volume almost twice that of rabbit CYP2C5, which is consistent with the large size of its preferred substrate paclitaxel (853.9 gmol^{-1}) (Kerdpin et al. 2004). The increased volume of the CYP2C8 active-site reveals a shift in the F' helix which establishes a cavity above β-sheet 1. However, flexibility in the F-G helix region is likely to be reduced in CYP2C8 dimers (Schoch et al. 2004).
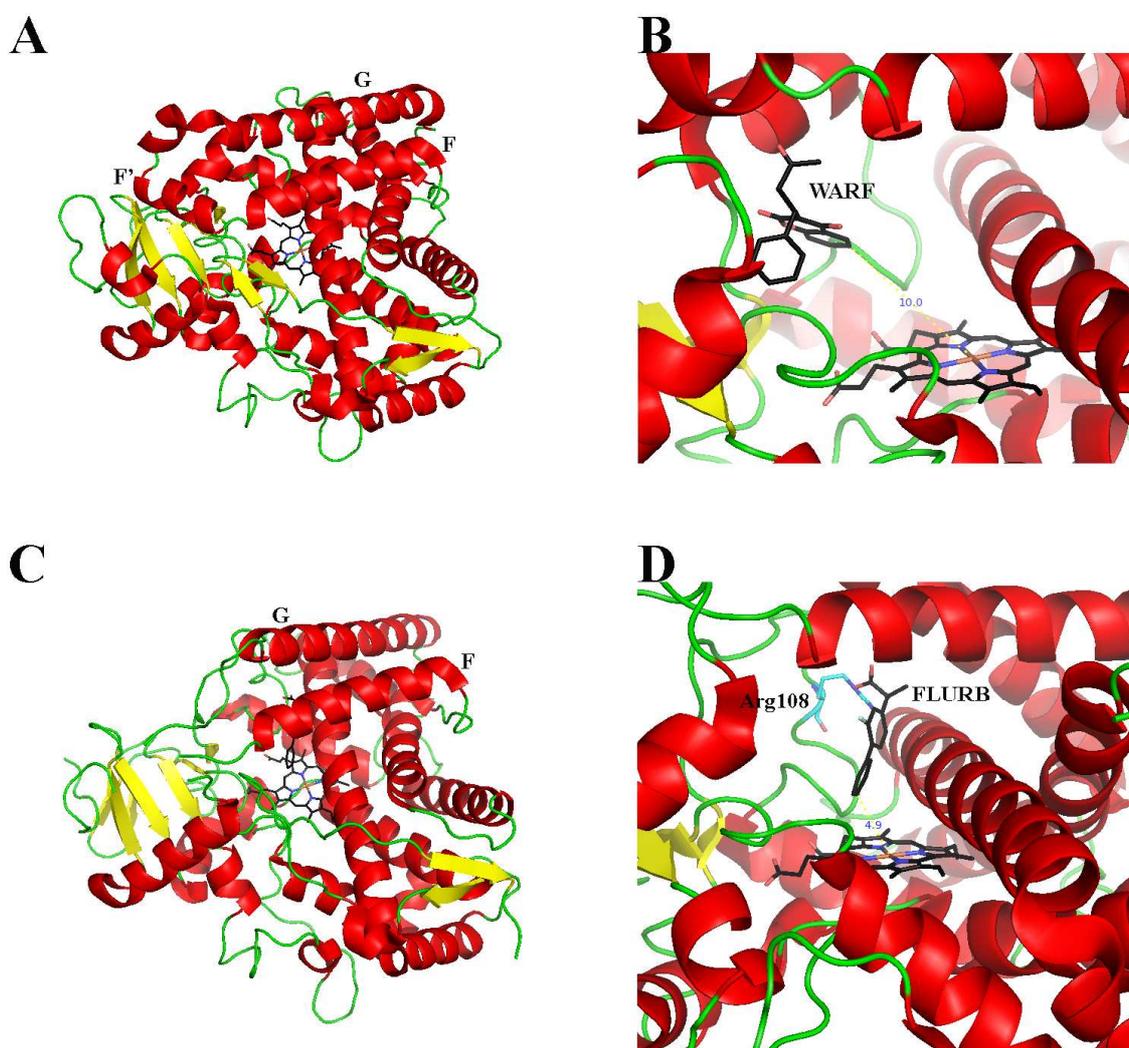
Figure 1.16 Bound and unbound crystal structures of human CYP2C9. (a) The crystal strucuture of CYP2C9 in the absence of ligand (pdb; 1OG2, 2.6Å) and (b) cocrystallized with warfarin (WARF) located 10Å from the site of catalysis (pdb; 1OG5, 2.55Å). (c) The refined crystal structure of CYP2C9 generated by Wester *et al.* (2003) (pdb; 1R9O, 2Å), (d) 1R9O complexed with flurbiprofen (FLURB), the site of catalysis positioned 4.9Å from the heme Fe.
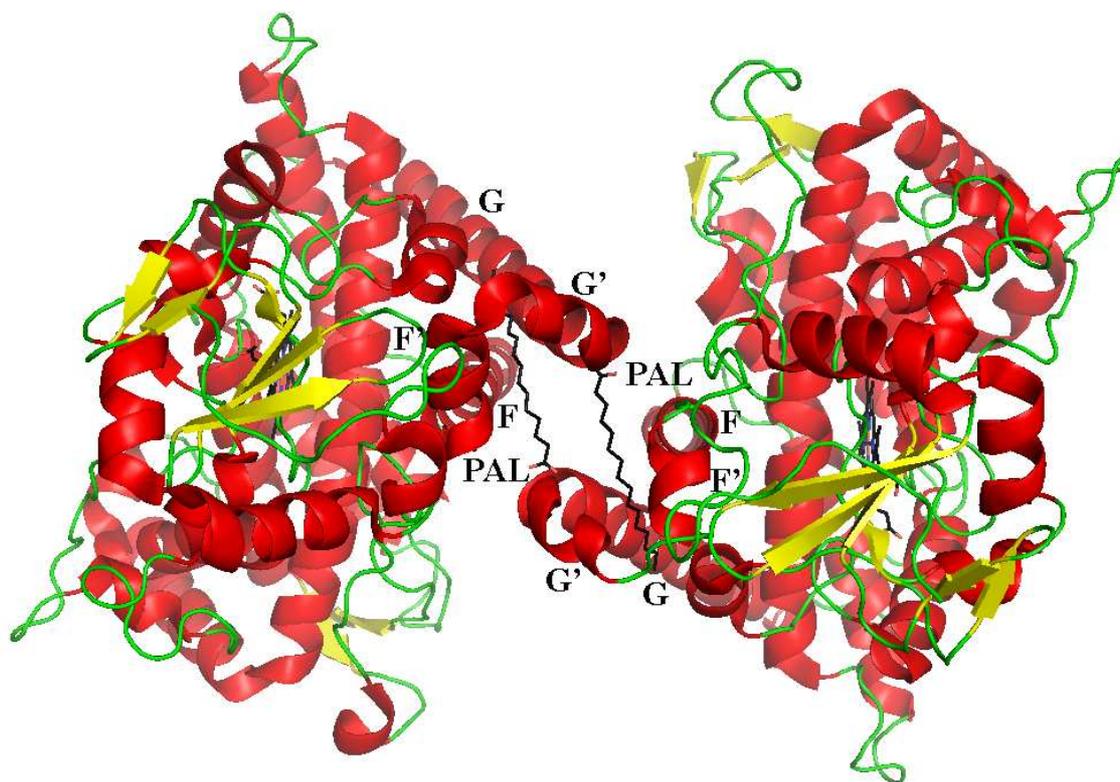
Figure 1.17 The crystal strucuture of CYP2C8 shown as a symmetric dimer. The fatty acid binding site is located at the dimerization interface. Palmitic acid (PAL).

The first crystal structure of a CYP3A enzyme was achieved by Yano et al. (2005) in the presence of the substrate erythromycin. However, electron density maps did not reveal the location of the substrate molecule and therefore could not be resolved in the final crystal structure. The CYP3A4 crystal (pdb; 1TQN) was determined to 2.05Å resolution (R-value 0.241) and, like CYP2C8 (1438Å^3), it revealed a large active-site volume (1386 Å^3), which is not surprising considering CYP3A4 oxidizes some of the largest P450 substrates (e.g. macrolide antibiotics, cyclosporin) (Polasek & Miners 2006). However, the shapes of their cavities differ markedly, due to differences in secondary and tertiary structure rearrangement. The major differences are seen in the F to G helix region that passes over the heme, creating a large cavity in CYP3A4. Close inspection of the CYP3A4 structure reveals helices F and G do

not extend over the active-site cavity (Figure 1.18) as they do in the CYP2C structures (Figures 1.16 and 1.17). In fact, the F-F' and G'-G loops are markedly elongated in the CYP3A4 structure and are comprised of residues 209-217 and 237-242, respectively. The replacement of helices (in the CYP2C proteins) with loops in CYP3A4 introduces the desired flexibility required to accommodate large or multiple substrates. Docking studies conducted by Yano et al. (2005) revealed that 7,8-benzoflavone and testosterone could bind simultaneously within the CYP3A4 active-site. Interestingly, several possible combinations of locations were observed. Similar to CYP2C9, R106 (corresponding to R108 in CYP2C9) resides at the distal end of the active-site cavity and participates in a network of hydrogen bonds linking Y53, D61, D76, R372, and E374 (Figure 1.18).

The first CYP2A enzyme to be crystallized was CYP2A6 (Yano et al. 2005). CYP2A6 is not considered a major drug metabolizing enzyme (section 1.3.3; Figure 1.6), however, this enzyme contributes to nicotine metabolism and is able to activate tobacco-specific procarcinogens (Tyndale & Sellers 2001). To better understand the mechanism of substrate and inhibitor binding in CYP2A6, Yano *et al.* (2005) generated crystal structures in both the presence of coumarin (substrate) at 1.9Å resolution (pdb; 1Z10, R-value 0.23) (Figure 1.19a), and in the presence of methoxsalen (inhibitor) at 2.05Å resolution (pdb; 1Z11, R-value 0.261) (Figure 1.19b). The volume of the CYP2A6 cavity is approximately one quarter that of CYP2C8, CYP2C9, and CYP3A4 and is well suited for small planar substrates that fit within the narrow, hydrophobic active-site. The CYP2A6 active-site shows features in common with CYP101.
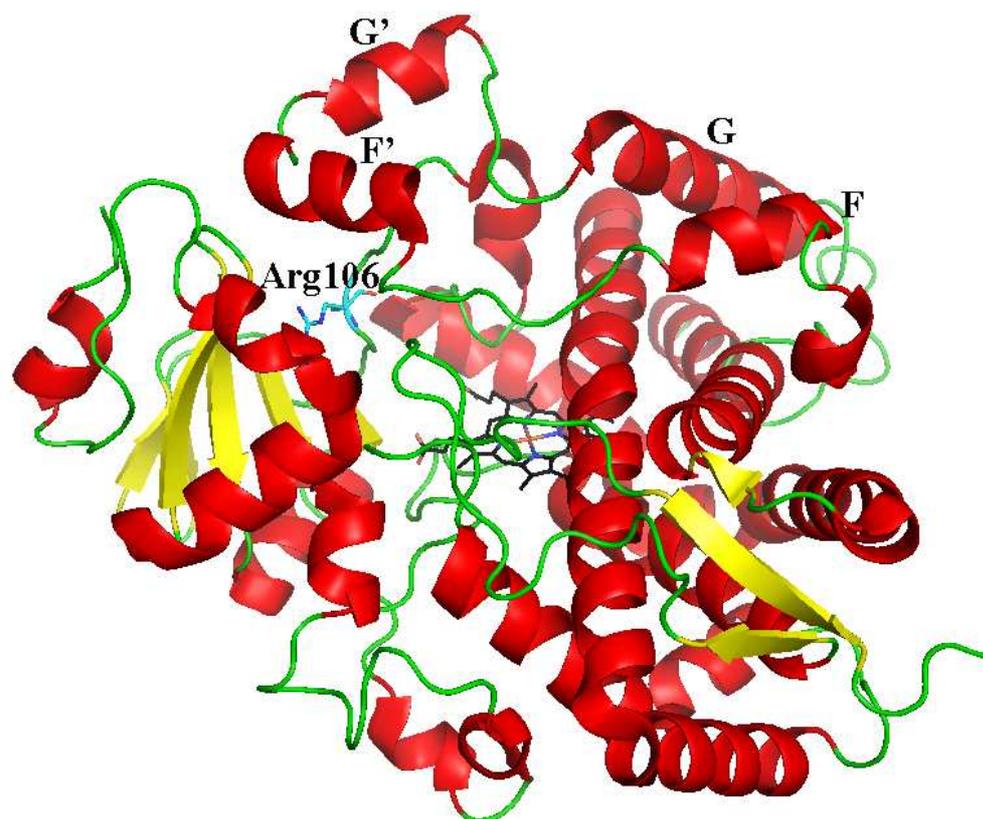
Figure 1.18 The crystal strucuture of human CYP3A4. The CYP3A4 structure identifies greater flexibility is achieved in the F through G helix region by minimizing the rigid helix strucutre. This allows the accommodation of large substrate molecules and may also contribute to heterotrophic cooperativity by allowing multiple binding sites for smaller substrate molecules. Through hydrogen bonding, R106 (cyan) is important for the orientation and stabilization of substrates in the distal end the CYP3A4 active-site.
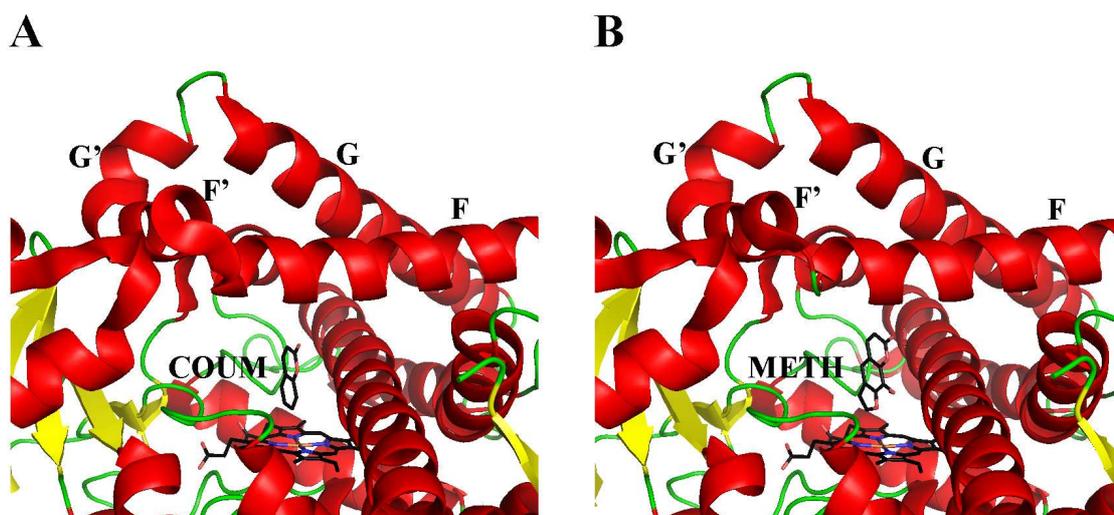
Figure 1.19 The bound and unbound crystal structures of CYP2A6. The crystal strucuture of CYP2A6 comparing the binding sites for (a) coumarin (COUM) and (b) methoxsalen (METH).

Human CYP2D6 is responsible for the metabolism of many clinical drugs (section 1.3.3; Figure 1.6) particularly those containing (protonated) basic nitrogen's and planar aromatic rings. CYP2D6 was crystallized recently by Rowland et al. (2006). The CYP2D6 structure was solved to 3.0Å (pdb; 2F9Q, R-value 0.232) and was solubilized using mutations based on its homology with CYP505 (P450$_{BM-3}$). The CYP2D6 structure (Figure 1.20) shows similarities to that of CYP2C9 (RMSD 1.16Å), but important differences occur in the F through G helix region. The F helix in CYP2D6 is two turns larger and arcs down toward the heme, thereby decreasing the active-site volume (540Å^3). There is no evidence of an F' helix, although a small G' helix was reported by the authors. However, Rowland *et al.* (2006) acknowledge that the quality of electron density maps in the F-G loop region is unsatisfactory. The crystal structure suggests that F120 is used to control the orientation of substrates relative to the heme.
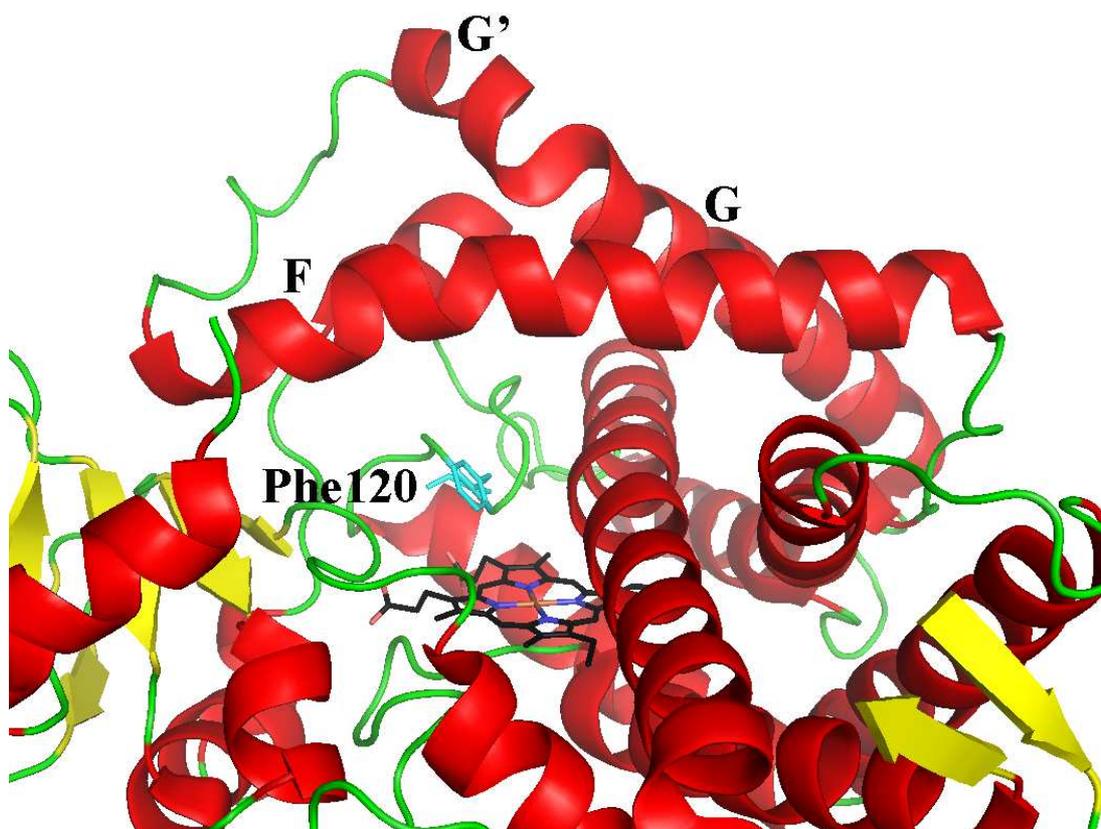
Figure 1.20 The crystal structure of CYP2D6 (2F9Q). The inward arching of the F helix is responsible for the small active-site cavity. F120 (cyan) is located in SRS1 and aids the productive alignment of aromatic substrates for catalysis.

During the course of this thesis, the human CYP1A2 crystal structure was elucidated by Sansen et al. (2007). Determination of the first CYP1 family enzyme was achieved at 1.95Å resolution (pdb; 2HI4, R-value 0.223) as a complex with the inhibitor, α-naphthoflavone. As for CYP2C5 (Wester et al. 2003b), CYP2C8 (Schoch et al. 2004), CYP2B4 (Scott et al. 2004a), and CYP2A6 (Yano et al. 2005), the N-terminal transmembrane domain of CYP1A2 was removed (at residue 42) and modified to encode the MAKKTSSKGKL (G1 construct) terminus. In addition, a further CYP1A2 3-26 deletion construct was crystallized. Both constructs were

designed to reduce protein aggregation and increase solubility. The final model comprises residues 34-513 refined from the superimposition of the derived structures. No mutations were made to increase expression yields, although the pCW-CYP1A2 expression construct was coexpressed with plasmids containing the chaperone proteins GroEL and GroES in an attempt to improve protein folding and yield (Baneyx & Mujacic 2004). Analogous with other crystallized P450 structures, the most conserved regions are the heme binding site and the proximal surface of the protein where OxR, NADPH, and cytochrome b_5 bind. Similarly, the most divergent regions appear in the substrate active-site, the B-C helix, and F-G helix regions. One notable difference in CYP1A2 is the 3^{10} F' and G' helices rather than α-helices. Even though CYP1A2 shares <29% sequence identity with CYP2A6, CYP2B4, CYP2C5, CYP2C8, CYP2C9, and CYP3A4, the secondary structure is clearly maintained (Figure 1.21a). In the presence of α-naphthoflavone the CYP1A2 structure appears in a closed conformation with the active-site cavity calculated at 363Å^3, slightly larger than CYP2A6 (260 Å^3) (Figure 1.21b). Interestingly, Sansen et al. (2007) reported the absence of water in the CYP1A2 active-site when complexed with α-naphthoflavone, which may describe the predominantly high-spin Fe state in the substrate free ferric enzyme (i.e. no water at the sixth coordination site on the heme Fe).
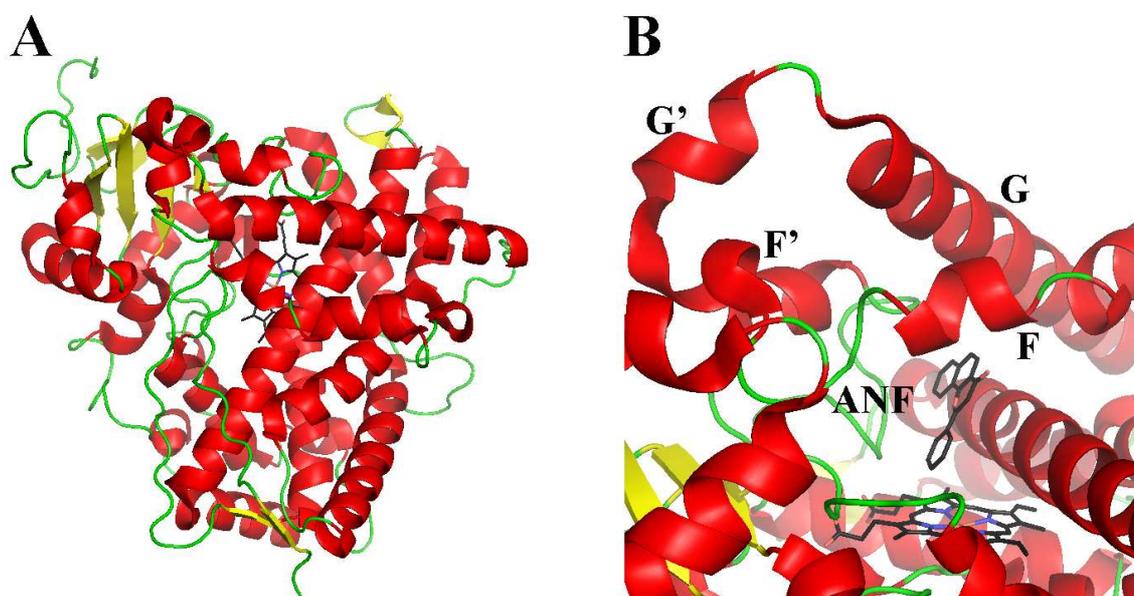
Figure 1.21 The bound and unbound crystal structures of human CYP1A2. (a)
The secondary and tertiary structure of human CYP1A2 (2HI4) closely resembles
that of other crystallised P450s. (b) The CYP1A2 inhibitor α-naphthoflavone (ANF)
is positioned in the active-site cavity in close approximation to the heme Fe. The
CYP1A2 active-site is well adapted for planar aromatic compounds.

1.5.3 Three-dimensional molecular modeling

1.5.3.1 Background

Protein modeling may be used to predict protein structure. The structure and function
of a protein are interconnected through the amino acid residues that determine the
tertiary structure. Thus, based on an understanding of a protein's function and its
primary sequence, predictions relating to protein structure can be made. Modeling is
necessary when the structure of a protein cannot be determined experimentally, or if
poor resolution is obtained with the crystallized protein. In the case of the class II
P450s, their hydrophobic and membranous nature greatly limits the solubility
required for trouble-free crystallization, highlighting the need to elucidate P450
structures by molecular modeling.

Currently, protein modeling is classified into three general categories (Krane & Raymer 2003):

(i) *ab initio* modeling, where a protein structure is generated by utilizing the primary sequence of the protein in question with no direct use of known structure(s). This is achieved by predictors that use potentials derived from distance data between atoms, rather than physical descriptors (e.g. electrostatic potential, hydrophobic contacts, etc.).

(ii) Fold recognition modeling or 'threading', where there is little similarity in the primary sequence but the target protein contains a fold that is already known. These techniques exploit the differences in the distribution of inter-residue distances for different combinations of amino acid residues.

(iii) Comparative or 'knowledge-based' modeling, where the data derived from nuclear magnetic resonance (NMR) or the X-ray crystallography of related protein structures are utilized to guide the modeling process. For comparative modeling to be successful, significant amino acid sequence homology must exist between the query and template protein(s).

Although considerable progress has been made in *ab initio* and fold recognition protein structure prediction (Lazaridis & Karplus 2000; Teodorescu et al. 2004), comparative protein modeling remains the most accurate prediction method available (Moult 1999). However, the accuracy of a protein structure derived from comparative modeling is still highly dependant on both the crystal template quality and sequence homology (Acharya & Lloyd 2005; Sippl 1993). For the purpose of this thesis, only comparative modeling will be discussed.

There are many variations in comparative modeling techniques (Wallner & Elofsson 2005). The most common technique involves the breakdown of the structure into conserved core regions and the loops that interlock them. Rigid templates that represent the conserved regions are obtained from homologous proteins of known structure and are usually found within the Brookhaven Protein Data Bank (PDB). There are currently 66478 structures in the PDB archive (last updated July, 2010). If more than one homologous template is available, then the atomic coordinates of all homologous residues (dependant on alignment) are usually averaged. All variable loop regions of the model are then searched within the PBD to resolve those fragments whose anchor regions have a good geometric fit to the template. Loop analysis usually provides a geometric tool whereby the retrieved fragment (and subsequent tertiary structure) is scored on the basis of 'fit quality' to the anchor regions sequence homology, steric interactions, and stereochemistry.

1.5.3.2 Model refinement through molecular mechanics

Homology models generated from comparative modeling usually begin as 'off-lattice' homology models, meaning that all originate as high energy protein structures that require geometry optimization and energy minimization. The potential energy functions used to relieve high energy conformations, termed Force Fields (FF), need to consider both dominant and subtle factors associated with protein stability in order to manipulate the energy of each amino acid within the protein structure. The objective of a FF is to derive an approximated energy function for which known crystallized proteins are represented in their energy minimized state. At the atomic level, a FF acts on individual residues by using statistically significant side-chain conformations (rotomers) to represent the flexibility of each amino acid

(Gordon, Marshall & Mayo 1999). A variety of stochastic and deterministic search algorithms are then used to find the combination of amino acid side-chain rotomers that complement the neighboring residue fold as ranked by a potential energy function. The potential energy terms commonly used in FFs describe the packing among non-covalently bound atoms (e.g. van der Waals forces), electrostatic interactions (e.g. distance attenuated charge), internal coordinate energies (e.g. torsional strain, dihedral angle), hydrophobic contacts, hydrogen bonding (e.g. angle dependant potential), and solvent (e.g. octanol/gas-water free energy of transfer) and entropic factors (e.g. conformational freedom) (Gordon, Marshall & Mayo 1999; Krane & Raymer 2003). A simple example of a FF is seen in equation 1.3.

$$\Delta G = \Delta G_{\text{van der Waals}} + \Delta G_{\text{H-bond}} + \Delta G_{\text{solvent}} + \Delta G_{\text{Coulomb}} \qquad \text{(eq. 1.3)}$$

where ΔG is the Gibbs free energy, with van der Waals, H-bond, solvent, and Coulomb potentials describing atom packing, sidechain-sidechain/sidechain-backbone interactions, solvation, and electrostatic energies, respectively.

In most cases, energy minimization leads to the optimization of the static atomic structure and therefore only establishes a local energy minimum for those residues in a specific configurational space (Lesyng & McCammon 1993). Hence, there is no guarantee that the local energy minimum represents that of the global minimum of the protein. Additionally, the relative weights of those descriptors making up the FF need to be correctly balanced in order to provide a more realistic refinement. However, the unrealistic weighting of electrostatic and van der Waals interactions

can generate non-optimal atom packing and unsatisfied hydrogen bond donor or acceptor sites (Linge et al. 2003).

1.5.3.3 CYP1A1 model construction

Extensive structure analysis of CYP1A1 has been accomplished by Lewis and his colleagues (Lewis, Modi & Dickins 2002; Lewis 2002a; Lewis 2002b; Lewis, Ioannides & Parke 1994a,1994b; Lewis, Ito & Lake 2006; Lewis & Lake 1996; Lewis, Lake & Dickins 2004; Lewis et al. 1999). The Lewis laboratory has primarily focused on the molecular modeling and quantitative structure-activity relationships of CYP1A enzymes and their interaction with drugs and other chemicals, particularly those associated with human drug metabolism. Utilizing the substrate bound CYP101 structure (Poulos et al. 1985), the first homology model of rat CYP1A1 was elucidated by Lewis, Ioannides & Parke (1994b) to investigate the selectivity and species differences between CYP1A1 and CYP1A2 (Lewis & Lake 1996; Lewis et al. 1999).

Using the SwissProt database sequence of CYP1A1 (accession P04798), Szklarz and Paulsen (2002) constructed the first homology model of CYP1A1 based on the coordinates of crystallized rabbit CYP2C5 (pdb entry 1DT6). In order to evaluate whether the CYP2C5 structure was the best available template, the CYP1A1 sequence was additionally aligned with those of the bacterial crystal templates CYP505 (P450$_{BM3}$), CYP101 (P450$_{cam}$), CYP108 (P450$_{terp}$), and CYP107A1 (P450$_{eryF}$). Based on this alignment, CYP1A1 showed greatest identity with CYP2C5 (28.8%) (Figure 1.22). The model was used to manually dock three substrates, 7-methoxy-, 7-ethoxy-, and 7-pentoxyresorufin into the active-site of CYP1A1. The compounds were docked in productive binding orientations, leading to their *O-*

dealkylation using the docking module of Insight-II. The docked resorufins, along with benzo[a]pyrene, were shown to be stabilized mainly by hydrophobic interactions. The key amino acids that potentially interact within 5Å of these substrates were documented as residues 111 of helix B; 122 and 123 of the B′-C loop; 224 and 228 of helix F; 312, 313, 316, 317, 320, and 321 of helix I; 381 and 382 of the loop between the K helix and sheet 1-4; 386 of sheet 1-4; 497 and 498 of sheet 4-2. These residues therefore provide candidate sites for mutagenesis as their replacement with other amino acids can be expected to alter catalytic activity. In this regard, V382 was identified computationally to affect CYP1A1's activity towards 7-methoxy- and 7-ethoxyresorufin through its interaction with the alkoxy chain of each substrate (Szklarz & Paulsen 2002). This was later confirmed by Liu et al. (2003) who showed that activity of the wild-type CYP1A1 enzyme was highest with 7-ethoxyresorufin and lowest with 7-pentoxyresorfin. Upon substitution of V382 with alanine, the activities toward 7-methoxy- and 7-ethoxyresorufin fell 10-fold, although little change in K_m values were observed (Table 1.5). The V382A mutant displayed the highest catalytic efficiency with 7-pentoxyresorufin, suggesting that replacement of Val with Ala at position 382 increases the active-site volume and allows better oxidation of the larger substrate (Figure 1.23). In contrast, generation of a V382L mutant significantly decreased oxidation of all alkoxyresorufin substrates.

```
                1                                                    50
human CYP1A1    MLFPISMSAT EFLLASVIFC LVFWVIRASR PQVPKGLKNP PGPWGWPLIG
rabbit CYP2C5   ....MDPVVV LVLGLCCLLL LSIWKQNSGR G......KLP PGPTPFPIIG

                51                                                   100
human CYP1A1    HMLTLG.KNP HLALSRMSQQ YGDVLQIRIG STPVVVLSGL DTIRQALVRQ
rabbit CYP2C5   NILQIDAKDI SKSLTKFSEC YGPVFTVYLG MKPTVVLHGY EAVKEALVDL

                101                                                  150
human CYP1A1    GDDFKGRPDL YTFTLISNGQ SMSFSPDSGP VWAARRRLAQ NGLKSFSIAS
rabbit CYP2C5   GEEFAGRGSV PILEKVSKGL GIAFS..NAK TWKEMRRFSL MTLRNFGMGK

                151                                                  200
human CYP1A1    DPASSTSCYL EEHVSKEAEV LISTLQELMA GPGHFNPYRY VVVSVTNVIC
rabbit CYP2C5   R.......SI EDRIQEEARC LVEELRKTNA SP..CDPTFI LGCAPCNVIC

                201                                                  250
human CYP1A1    AICFGRRYDH NHQELLSLVN .LNNNFGEVV GS..GNPADF IPILRYLPNP
rabbit CYP2C5   SVIFHNRFDY KDEEFLKLME SLNENVRILS SPWLQVYNNF PALLDYFPG.

                251                                                  300
human CYP1A1    SLNAFKDLNE KFYSFMQKMV KEHYKTFEKG HIRDITDSLI EHCQEKQLDE
rabbit CYP2C5   IHKTLLKNAD YIKNFIMEKV KEHQKLLDVN NPRDFIDCFL IKMEQE....

                301                                                  350
human CYP1A1    NANVQLSDEK IINIVLDLFG AGFDTVTTAI SWSLMYLVMN PRVQRKIQEE
rabbit CYP2C5   .NNLEFTLES LVIAVSDLFG AGTETTSTTL RYSLLLLKH PEVAARVQEE

                351                                                  400
human CYP1A1    LDTVIGRSRR PRLSDRSHLP YMEAFILETF RHSSFVPFTI PHSTTRDTSL
rabbit CYP2C5   IERVIGRHRS PCMQDRSRMP YTDAVIHEIQ RFIDLLPTNL PHAVTRDVRF

                401                                                  450
human CYP1A1    KGFYIPKGRC VFVNQWQINH DQKLWVNPSE FLPERFLTPD GAIDKVLSEK
rabbit CYP2C5   RNYFIPKGTD IITSLTSVLH DEKAFPNPKV FDPGHFLDES G..NFKKSDY

                451                                                  500
human CYP1A1    VIIFGMGKRK CIGETIARWE VFLFLAILLQ RVEFSVPLGV KVDMTPIYGL
rabbit CYP2C5   FMPFSAGKRM CVGEGLARME LFLFLTSILQ NFKLQS.LVE PKDLDITAVV

                501            517
human CYP1A1    TMKHACCEHF QMQLRS.
rabbit CYP2C5   NGFVSVPPSY QLCFIPI
```

Figure 1.22 Sequence alignment between human CYP1A1 and rabbit CYP2C5.
Initial alignment gave 44.3% similarity (red & green) and 28.8% identity (red only).

Table 1.5 Comparison of kinetic parameters of CYP1A1 (WT) and the V382A and V382L mutants. Taken from Liu et al. (2003).

Variant	Substrate	$K_{m \, (\mu M)}$	V_{max} (nmol. min^{-1}/nmol P450)	V_{max}/K_m
CYP1A1 (WT)	7-methoxyresorufin	1.3	2.4	1.9
	7-ethoxyresorufin	0.60	11	17.9
	7-pentoxyresorufin	0.3	0.4	1.33
CYP1A1 V382A	7-methoxyresorufin	0.8	0.2	0.23
	7-ethoxyresorufin	0.55	1.1	2.0
	7-pentoxyresorufin	0.15	0.6	4.5
CYP1A1 V382L	7-methoxyresorufin	2.0	0.15	0.07
	7-ethoxyresorufin	2.3	0.15	0.07
	7-pentoxyresorufin	1.8	0.01	0.004

Valine   Alanine   Leucine
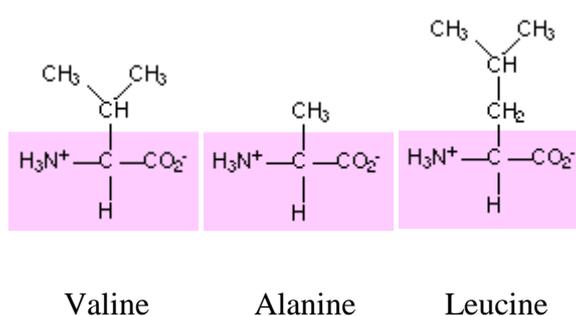
Figure 1.23 The amino acid substitutions made by Szklarz *et al.* (2003). Variablity in side-chain volume affects the kinetics of 7-methoxy-, 7-ethoxy-, and 7-pentoxyresorufin dealkylation by CYP1A1. The smaller molecular volume of alanine increases the void volume of the P450 active-site, potentially accommodating larger substrates.

One important approach to studying the structure-activity relationship of CYP1A1 for enhanced DTIC metabolism involves a combination of molecular modeling and computational chemistry. Prior to the commencement of this thesis no CYP1A1 homology model had been generated using human P450 templates. Importantly, the ways in which X-ray diffraction data are collected and refined have a strong impact on the final quality of the homology model. This being the case, the current published models might be considered inherently inaccurate, not only based on the template used, in some cases chimeras (CYP2C5; 1NR6, CYP2C9; 1OG5), but additionally in the resolution of the structural templates. By generating a model based on human templates, with inclusion of the largely homologous CYP1A2 (~80%), increased accuracy of secondary and tertiary structure organization would be achieved, therefore enabling the key residues involved in CYP1A1 substrate selectivity to be identified. From this, accurate predictions can be made as to what amino acid substitutions would enhance or inhibit enzyme/substrate affinity and DTIC activation.

1.6 Research aims

The primary aim of the *in silico* studies presented in this thesis was to generate a CYP1A1 homology model with the greatest possible accuracy. The CYP1A1 structure could then be utilized to generate modified enzymes, capable of enhanced DTIC bioactivation *in vitro*.

Specific aims were:

1. Generate a chemically and structurally valid CYP1A1 homology model based on crystallized human P450 homologs – **Chapter 3**.

2. Validate the CYP1A1 homology model by utilizing structure-activity analysis with the prototypical CYP1A1 substrate, 7-ethoxyresorufin – **Chapter 4**.

3. Characterize those residues that are critical for the alignment and orientation of DTIC in the CYP1A1 active-site and elucidate which protein mutations will enhance DTIC activation – **Chapter 5**.

4. Evaluate the bioactivation and cytotoxicity of selected CYP1A1 mutants with enhanced DTIC activation kinetics in the mammalian cell lines SK-MEL-28 and COS-7 – **Chapter 6**.