Functional effects of cytochrome P450 variants on drug metabolism and adverse drug reactions: developing and extending high throughput P450 protein technology platforms

Omesan Nair

Thesis presented for the degree of Master of Science in Medicine

(MSc MED)

In the department of Medical Biochemistry

Faculty of Health Sciences

University of Cape Town

February 2014

The financial assistance of the National Research Foundation (DAAD-NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the DAAD-NRF.

Abstract

Cytochrome P450 (CYPs) are a superfamily of heme containing enzymes that catalyse a diverse range of biological reactions. They are responsible for over 80% of primary metabolism of currently available drugs and are therefore central to its medical importance. Investigating the effects of these enzymes on drugs by metabolite detection and kinetic studies is a step forward to the vision of personalised medicine. The enzyme family is known to be associated with the development of adverse drug reactions which are usually only discovered in late stages of drug development, therefore screening for potential adverse drug reactions earlier on would aid minimising such adverse events occurring. There is therefore a need to analyse the interaction profile of new drugs with CYPs in a cost effective and high throughput manner for early stage screening, since drug discovery efforts tend to utilise large compound libraries. Recently, a novel functional CYP microarray has been developed in the Blackburn laboratory at UCT to enable label-dependent analysis of metabolism of substrates by the major CYP3A4 isoform in a high throughput manner. This thesis describes efforts involved in expanding the functional CYP microarray format to the other major CYP isoforms namely, CYP2C9 and CYP2D6 and developing a new immobilisation-free technology with label-free mass spectrometric identification and quantitation of metabolites formed. The goals of expansion of functional CYP microarrays were achieved by using microarray or confocal fluorescence scanning in conjunction with atomic force microscopy to more accurately quantitate active CYP3A4, CYP2C9 and CYP2D6 protein levels for catalytic substrate-dependent turnover rates. Finally the label- and immobilisation-free CYP technology was evaluated using probe substrates and a complex drug, rifampicin. These two platforms are primed to be a useful tool in pre-clinical drug screening for use in the drug discovery field by the academic, pharmaceutical and biotechnology industries.

Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. I have used the APA convention for citation and referencing. Each contribution to, and quotation in, this thesis from the work(s) of other people has been attributed, and has been cited and referenced.

3. This thesis is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

5. I acknowledge that copying someone else's assignment or essay, or part of it, is wrong, and declare that this is my own work.

Signature 

Acknowledgments

I would like to extend my deepest gratitude to my supervisor Professor Jonathan Blackburn for the insight and guidance throughout my masters. Thank you for the guidance and motivation through the good and bad times, it is highly appreciated. I am grateful for the gracious financial and non-financial support from the South African National Research Foundation (NRF) and German Academic Exchange Service (DAAD) joint in-country scholarship program. I also wish to give many thanks to Dr Aubrey Shoko and the Centre for proteomic and genomic research (CPGR), Cape Town, for sharing expertise and use of microarray equipment. To my lab colleagues, Lauren Coulson, Dr Andrew Nel, Jessica Duarte, Anil Pooran, Sriram Krishnan, Dr Nicolette Hendricks and the rest of the Blackburn lab, many thanks for the support and useful discussions. To Dr Brandy Young and Brandon Murugan, many thanks for sharing your expertise with mass spectrometry equipment and analysis. To Matthew Njoroge and Elizabeth Kigondu of the Chibale group, thank you for the assistance, discussions and sourcing of the microsomal preparations. I am also thankful to Dr Martina Meincken for assisting with the AFM apparatus.

Finally, I am grateful to my parents, J.G. and K. Nair, my sister – Shankari Nair, Dr Christopher Barnett and all my friends for the support and encouragement over these years. Lastly, to my colleagues in the biotechnology industry, thank you for inspiring me to pursue a career in life sciences.

Table of Contents

List of Tables and Figures

Abbreviations

6x His – Six times histidine tag

Adr – Adrenodoxin reductase

Adx – Adrenodoxin

AFM – Atomic Force Microscopy

AGC – Automatic Gain Control

ALA – Aminolevulinic acid

ANOVA – Analysis of Variance

BBSA – Biotinylated Bovine Serum Albumin

BSA – Bovine Serum Albumin

cDNA – Complementary/copy DNA

CHP – Cumene Hydroperoxide

CO – Carbon Monoxide

CPR – Cytochrome P450 reductase

Cy5 – Cyanine 5 dye

CYP – Cytochrome P450

CytB5 – Cytochrome B5

DART – Direct Analysis in Real Time

DESI – Desorption Electrospray Ionization

DMSO – Dimethyl Sulfoxide

ER – Endoplasmic Reticulum

ESI – Electrospray Ionization

FAD – Flavin Adenine Dinucleotide

Fe^{2+} - Ferrous Iron

Fe^{3+} - Ferric Iron

FMN – Flavin Mono Nucleotide

HLM – Human Liver Microsomes

HPLC – High Pressure Liquid Chromatography

HRP – Horse Radish Peroxidase

IVIVE – *In Vitro-In Vivo* Extrapolation

Kcat – Catalytic turnover rate constant

Kd – Dissociation constant

kDa – kilodaltons

Km – Michaelis-Menton kinetic constant

LB – Luria-Bertani Broth

MALDI – Matrix-Assisted Laser Desorption Ionization

mg/ml – Milligram per millilitre

MRM – Multiple Reaction Monitoring

MS – Mass Spectrometry

MS/MS – Tandem Mass Spectrometry

NADPH – Nicotinamide Adenine Dinucleotide Phosphate

nm – Nanometre

nM – Nanomolar

nM/min – Nanomolar per minute

NTA – Nickel Nitrilotriacetic Acid

PBS – Phosphate Buffered Saline

PEG – Polyethylene Glycol

RIF – Rifampicin

RLM – Rat Liver Microsomes

SDS – Sodium Dodecyl Sulphate

PAGE – Polyacrylamide Gel Electrophoresis

TB – Terrific Broth

U/ml – Units per millilitre

μM – Micromolar

um – Micrometer

v/v – Volume per volume

w/v – weight per volume

Vmax – Maximum Velocity

Chapter 1 – Introduction and Literature Review

1.1 Cytochrome P450 and drug metabolism

1.1.1 Background

Cytochrome P450s (CYPs) are heme containing proteins that catalyse many biological reactions such as oxidation, reduction, peroxidation and hydrolysis. These reactions are primarily part of first pass liver metabolism. This is the first metabolic alteration xenobiotics and drugs usually face upon entering the human body. This affects the intrinsic clearance of drugs that are administered by the oral route.

1.1.2 Role in Drug Metabolism

ADME is an abbreviation in the field of pharmacology, this stands for Administration, distribution, metabolism and excretion. Administration is the route of entry of a drug or therapeutic into the body. Various routes exist, such as oral, intravenous, intramuscular, transdermal, subcutaneous, intraperitoneal etc. Generally the oral route of administration of pharmacological entry is preferred due to its ease of administration and compliance by patients. Absorption of drugs taken through the oral route occurs in the stomach and intestine, with the exception being drugs that are administered sublingually. The principle site of absorption of a particular drug is determined largely by the drugs physicochemical properties. Drugs with acidic pK_a will be absorbed mainly through the stomach vasculature as they are likely to be protonated and uncharged in the acidic conditions of the stomach. Administration of drugs with neutral or basic pK_a will be absorbed mainly through the vasculature of lumen in the intestine as they are likely to be de-protonated and uncharged in the more alkaline conditions of the intestine; absorption through the intestine is usually more effective due to the large length of the intestine (Neal, 2002). Uncharged drug species are more likely to transverse cell membranes and be absorbed into circulation.

Distribution of a drug is another key aspect of drug action; this regards the transport of the drug following absorption. Several factors are considered including drug transporters and plasma protein binding. If the drug target is not the blood stream then the (apparent) volume of distribution which describes the distribution of the drug between plasma and the rest of the body tissue becomes a crucial measure in the drugs' efficacy, since it is not reaching the intended target and possesses a greater chance of off-target effects, potentially leading to adverse drug reactions.

Metabolism relates to the alteration of the parent or administered drug upon dosing and can be either activating or deactivating with respect to the pharmacological agent. Furthermore, numerous drugs are administered in an inactive prodrug form, with metabolism being required to liberate the active drug. Examples are the peptic ulcer drug, omeprazole, and the antiviral drug, acyclovir, used to treat herpes simplex virus (Oz & Ebersole, 2008; Thomsen, Christensen, Bagger, & Steffansen, 2004). The deactivation or breakdown of active drugs is an important aspect of pharmacology. Various enzymes have been shown to be responsible for these processes of which the cytochrome P450 family are known to play a major role. Cytochrome P450 mediated oxidation typically makes drugs more hydrophilic which facilitates excretion from the body in addition to altering the drugs' structure and activity. Overall, the drug metabolism process is conventionally divided into two steps, namely phase one (I) and phase two (II) drug metabolism. Phase I drug metabolism usually describes those reactions that are modifying reactions such as hydroxylation, reduction and oxidation reactions, whilst phase II describes the conjugation reactions such as glutathione conjugation or glycosylation (Neal, 2002).

Excretion of a drug is the end stage of pharmacokinetics; it is the elimination of the drug from the body. There are various routes of elimination including the most common urinary, biliary and faecal excretion, and the less common excretion through sweat and mammary glands, in tears, deposited into hair follicles and exhaled through the lungs. Of these, urinary excretion typically accounts for the majority of drug excretion particularly for more hydrophilic compounds such as those produced through metabolism. By contrast, the faecal route of elimination comprises drugs that have not been absorbed when administered through the oral route, drugs that have been absorbed and actively

transported back into the gut lumen or drugs that have been excreted through the biliary route of elimination.

Pharmacokinetics (PK) is the broad term describing the phenomena of ADME. Loosely it means "what the body does to a drug". The concept of PK suggests that ADME works in concert, with the various processes occurring either sequentially or sometimes simultaneously. Other concepts included under PK are drug half-life, metabolic stability and drug clearance. Drug half-life is a measure of the time it takes for a drug to reach half the maximal plasma concentration, so incorporates absorption and metabolism. Metabolic stability of a drug is all the processes of metabolism including phase I, phase II and drug transport. Drug clearance is a measure of removal of the drug from plasma and this includes drug transport, metabolism and excretion.

Pharmacodynamics (PD) describes the process of drug action and includes the mechanism of action and the dosing effect regime. An important consideration in PD studies is therefore the definition of the therapeutic window, being the balance of dose with the relative effects of desired and any unintended activities.

There are many undesired effects that occur upon drug usage. These are usually termed side effects and include various deleterious effects from genotoxicity to drug induced liver damage and cardiac arrhythmia caused by target or off-target effects of the parent drug or its metabolites. Other mechanistic origins of toxicity include co-administered drug regimes, certain herbal extracts or supplements, grapefruit juice and other xenobiotics that interact with cytochrome P450 enzymes, activating or inhibiting them leading to drug concentrations outside the therapeutic window (Guengerich, 2003).

In relation to metabolism, PK and PD are referred to as toxicokinetics and toxicodynamics and the study of these concepts is crucial to monitor, evaluate and predict the potential toxicity of new chemical entities. As the world begins to move away from reliance on animal testing in pre-clinical drug candidate evaluation, new technological solutions need to be developed in the ADME field in order to enable ethically sound and more effective high throughput screening of new chemical entities

to determine their metabolism and/or interaction profiles *in vitro*. This is the rationale for the research described in this thesis.

1.2 Cytochrome P450 isoforms

CYPs are responsible for 70-80% of phase I drug metabolism of currently used clinical drugs and catalyse oxidation and reduction reactions as well as hydrolysis (Evans & Relling, 1999). This type of metabolism serves to make drugs less reactive or to modify them for conjugation for easier elimination. CYPs are also capable of many other reactions such as hydroxylation, dehydrogenation and ring expansion as reviewed by Guengerich (Guengerich, 2001). CYPs also metabolise endogenous compounds such as steroids, fatty acids, bile acids and prostaglandins (Neal, 2002). CYPs therefore have both an endogenous and exogenous role in metabolism.

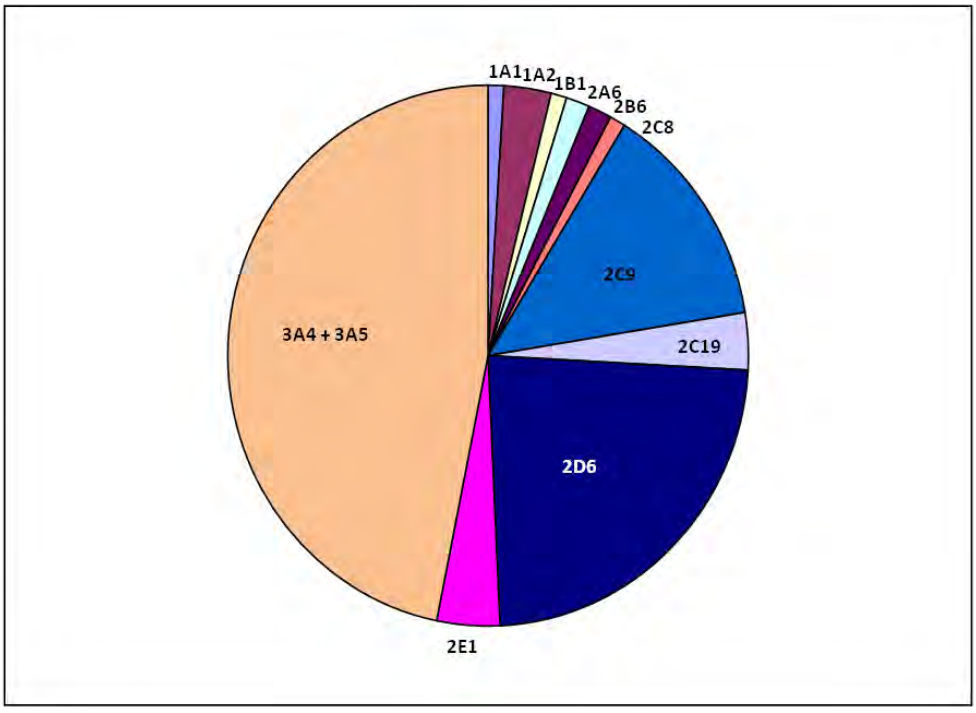


Figure1.1: Relative contribution of CYPs to drug metabolism, adapted from (Evans & Relling, 1999).

Schematic showing a representation of the relative contributions to phase I drug metabolism by the various microsomal CYP isoforms.

In humans there are 57 CYP genes and 58 pseudogenes reported in the literature from genome sequencing experiments (Nelson et al., 2004). The major CYPs involved in 70-80% of all drug and xenobiotic metabolism are the CYP3A4, CYP2D6 and the CYP2C9 isoforms (Figure 1.1).

CYP3A4 is involved in 40-45% of phase I metabolism of all marketed drugs and consists of an N-terminal domain together with a larger helical C-terminal domain that contains the heme and the active site (Zanger & Schwab, 2013). The heme is ligated by a coordination bond to the thiol of an absolutely conserved cysteine residue and the heme also interacts with the surrounding amino acid residues of the CYP active site cavity (Williams et al., 2004). The CYP3A4 substrate binding cavity and active site has been shown to be relatively large, which may contribute to multiple substrate binding and altered pharmacokinetic events and can also lead to adverse drug reactions (Yano et al., 2004). There is also the possibility of a peripheral substrate binding site which may be involved in initial binding or play a role in allosteric regulation and which may contribute to the irregular biphasic enzyme kinetic profiles known to be displayed by this CYP isoform, although this is not yet proven beyond doubt (Williams et al., 2004). CYP3A4 is involved in endogenous chemical reactions, notably: steroid and bile acid hydroxylation as well as oxidation of many xenobiotics (Rendic, 2002). These reactions typically make such compounds less toxic but can also convert them into reactive compounds.

CYP2C9 is responsible for metabolism of around 10-20% of commonly prescribed drugs (Zanger & Schwab, 2013). It has been shown that CYP2C9 metabolises some weakly acidic drugs and several non-steroidal anti-inflammatory drugs as well as the anti-coagulant drug Warfarin (Haining, Hunter, Veronese, Trager, & Rettie, 1996; Williams et al., 2003).

Cytochrome 2D6 is responsible for 20-30% of phase I drug metabolism and can accept substrates that include lipophilic bases with a protonable nitrogen atom and a planar aromatic ring (Ingelman-sundberg, 2004). In particular, CYP2D6 has a high affinity for alkaloids (Ingelman-sundberg, 2004; Rendic, 2002).

Given the importance of anti-inflammatory and anti-coagulant drugs, much is still unknown about the enzymes that metabolise them. For example, CYPs are known to have many single nucleotide polymorphisms (SNPs), the distribution of which differs between individuals in a large population and between various ethnic groups. Therefore, specific groups of people may respond differently to certain drugs or drug regimes as a result of the effects of polymorphisms on drug metabolism. One example of a polymorphism associated with altered responses is the polymorphism of CYP2D6 which can be associated with greater chance of relapse and decreased therapeutic effects of Tamoxifen in the prevention and treatment of certain types of breast cancers (Goetz et al., 2007). There is thus a need to explore and investigate more efficiently the way in which drug metabolism is affected by polymorphic variation in CYP genes, as well as by drug-induced alteration in CYP expression or activity. For example, a drug might inhibit or induce a CYP leading in turn to increased toxic effects by altering pharmacokinetic profiles of other compounds, a case in point being highly active anti-retroviral therapy (HAART) cocktails. Here, co-administration of Efavirenz together with certain protease inhibitors in treatment of HIV infection is complicated by Efavirenz causing adverse drug interactions due to the induction of CYP3A4 amongst others (Michaud et al., 2012). The induction of CYP3A4 thus alters the pharmacokinetic profile of other drugs that might be co administered and that are substrates of CYP3A4.

A drug may also induce CYPs through binding to specific nuclear receptors, such as the glucocorticoid receptor, pregnane X receptor, aryl hydrocarbon receptors and others as reviewed previously (Honkakoski & Negishi, 2000; Quattrochi & Guzelian, 2001). It is possible that these effects are also exerted indirectly through non-classical signal transduction and cross talk between the various nuclear receptors, although this remains to be proven.

Furthermore, intracellular drug concentrations can also be affected by the interaction of CYPs with drug transporters such as the p-glycoprotein or the multi drug resistance proteins which can themselves directly alter the pharmacokinetic profile of drugs (Baron, Goh, Yao, Wolf, & Friedberg, 2001; Benet, Izumi, Zhang, Silverman, & Wacher, 1999).

Prior to marketing a drug, it has to be tested for these effects. Pre-clinical drug testing is usually performed in tissue culture or in animal models (Esch, King, & Shuler, 2011). Assays carried out in immortalized cell lines can respond very differently to normal cells and animal models have different CYPs that may not show significant homology to humans and give confounding results (D. E. Johnson, 2012). Therefore there is the need for more relevant, ethically sound technology to be developed; this provided the motivation to investigate the possible platforms that are available to address this problem.

The ideal platform would allow the assessment of different human CYPs in terms of their different substrate specificities, their cross-reactivity and their differential effects based on polymorphisms in turn suggesting that many isoforms of human CYPs be investigated in a highly multiplexed, parallel approach. Microarrays have long been used for the detailed and accurate analysis of DNA (genotyping) and mRNA (transcriptomics) and have been used extensively in the pharmaceutical industry since the advent of pharmacogenomics (Hardiman, 2006).

Pharmacogenomics can be essentially defined as the study of the difference in outcome of drug action in different individuals; this is the beginning of the halcyon goal of personalized medicine whereby drug therapy is tailored dependant on the individual patient genetic profile. However, pharmacogenomic analyses currently lacks any true predictive power regarding the consequence of altered CYP expression profiles or altered CYP polymorphisms on drug metabolism or adverse drug reactions. Instead, pharmacogenomics relies very largely on statistical associations drawn across a large cohort between, for example, the prevalence of a specific CYP polymorphism and an observed adverse drug reaction in the human population being administered a given drug. In order to shift this paradigm to one where the effects of altered gene expression or polymorphic variation on drug metabolism can be predicted without administration of a new drug to a cohort of patients, a new pharmaco-proteomic approach is really needed. In a protein centric approach, multiplexed protein analysis platforms will be used to provide mechanistic information on drug metabolism in an accurate, high throughput and quantitative manner to complement data gained from conventional studies for drug toxicology.

1.3 Structure

CYPs are heme containing proteins which are membrane bound in plants, fungi and animal cells and soluble in bacteria (E. F. Johnson & Stout, 2013). The majority of the membrane-bound CYPs associate with the endoplasmic reticulum (ER) (microsomal CYPs) but a minority also associate with the inner mitochondrial membrane, and some exist between the two (Anandatheerthavarada et al., 1997). The CYPs are expressed highly in the liver but are also expressed in other extra-hepatic tissues such as the kidney, lungs, adrenal glands, gonads and brain and it is suggested that the tissue specific expression patterns might be important for their unique physiological roles (Seliskar & Rozman, 2007).

Microsomal CYPs receive electrons from NADPH via NADPH cytochrome P450 reductase (CPR) while mitochondrial CYPs receive electrons through Adrenodoxin (Adx) and Adrenodoxin reductase (Adr) (Anandatheerthavarada et al., 1997; Omura, 2006). The main difference between bacterial CYPs and others is an extension of the N-terminus that forms a membrane anchor of 20-30 amino acid residues in the membrane-bound CYPs. The protein is typically anchored in the ER membrane with the globular portion of the protein facing the cytosol and the heme lying mostly perpendicular to the membrane (Edwards, Murray, Singleton, & Boobis, 1991). Furthermore, mammalian CPR also carries an N-terminal hydrophobic peptide and thereby becomes anchored into the ER membrane. Interestingly, solution phase assays using solubilised CYPs have in general proven not successful, suggesting that the enhanced physical proximity between CYP and CPR proteins that results from their co-localisation on the same two-dimensional surface is essential for effective electron transfer to occur and furthermore suggesting a need to immobilize such solubilised enzymes onto solid supports to reconstitute activity. Usefully, there is increasing interest in immobilization of enzymes and proteins for use in the biomedical field as a means to retain activity in a cost effective manner and to lower amounts of reagents and resources used, as reviewed (Liang, Li, & Yang, 2000).

A major characteristic of all CYPs is the presence of the heme in the protein. The iron protoporphyrin IX group is axially coordinated to a thiolate ion provided by a highly conserved cysteine residue in the

heme binding site of CYPs (Munro, Girvan, Mason, Dunford, & McLean, 2013). This gives rise to the characteristic Soret absorption peak at 450 nm upon binding of carbon monoxide to CYPs by which these enzymes are spectroscopically recognised (Chottard, Schappacher, Richard, & Weiss, 1984).

CYPs share a common, largely alpha helical fold having the heme binding region located between a larger alpha helix rich region and a smaller beta sheet rich domain. The heme scaffold is coordinated equatorially by the heme pyrrole nitrogens and axially by the highly conserved cystein thiol proximally and by a water molecule (Munro et al., 2013). Various crystal structures of CYPs have been published, many of which have been engineered by truncation or other modification to the membrane anchor sequence in order to facilitate solubility for subsequent structure elucidation.

1.3.1 CYP3A4 structure

CYP3A4 is 503 amino acids in length and has a molecular weight of 57.34 kilodaltons (kDa). The gene encoding CYP3A4 is located on chromosome 7q21. Individual CYP3A4 crystal structures reveal relatively large active sites with evidence of structural change following ligand binding with an additional peripheral binding site (Figure 1.2). The structure has a β-strand N-terminal domain and the active site and heme binding site located within the C-terminal domain. The heme iron ligand is coordinated with the highly conserved cysteine residue at position 442 and there are nitrogen coordination interactions with the Arg105, Trp126, Arg130, Arg375, and Arg440 residues formed close to the active site (Williams et al., 2004). A solvent and ligand access channel is provided by loops B-C and F-G, which is thought to have a high degree of flexibility.

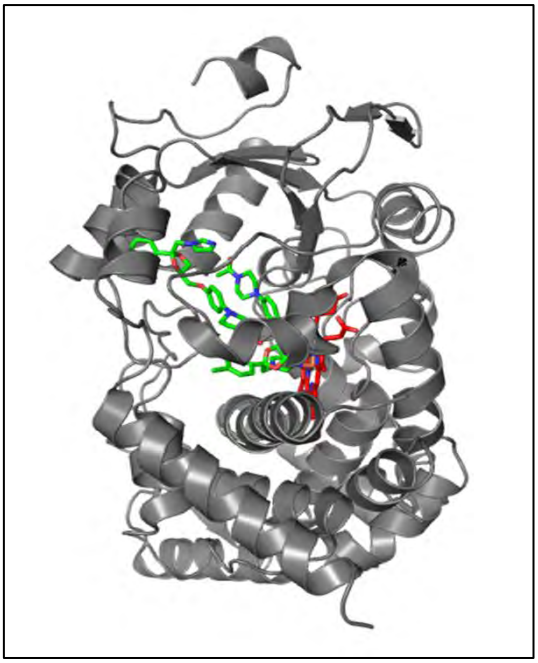


Figure 1.2: Chain and ribbon representation of CYP3A4 with bound ketoconazole (Williams et al., 2004).

Ribbon diagram represents the native protein structure of CYP3A4 protein with bound ketoconazole. Diagram depicts alpha helical regions and central heme prosthetic group (red), and ketoconazole (green).

1.3.2 CYP2C9 structure

CYP2C9 is a 490 amino acid protein with a molecular weight of 55.63 kDa; the gene encoding CYP2C9 is located in chromosome 10q24. Like CYP3A4, CYP2C9 is a two domain protein with heme axially coordinated with the Cys 435 position with the heme being stabilised by interactions with residues Trp 120, Arg 124, His 368 and Arg 433 (Williams et al., 2003). Helix B and C form part of the Substrate Recognition Site (SRS1) and this region is known to exist in open and closed forms, with a large degree of flexibility and conformational changes (Wester et al., 2004) (Figure1.3).

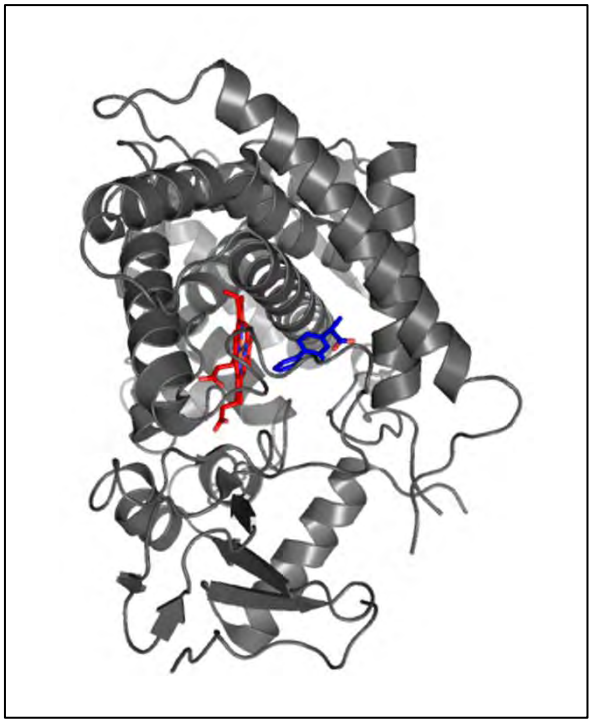


Figure 1.3 : Representation of CYP2C9 structure with bound Warfarin (Williams et al., 2003)

Figure depicts the ribbon representation of the protein structure of CYP2C9. Diagram shows alpha helical regions and location of heme prosthetic region (red) with bound Warfarin (blue).

1.3.3 CYP2D6 structure

CYP2D6 protein is 497 amino acids in length and is 55.77 kDa in molecular weight; the gene that encodes CYP2D6 is located in chromosome 22q13. The structure has some similarity with that of CYP2C9, with the exception that the F helix has two additional turns. The active site is well defined and is situated close to and above the heme pocket. The F-G loop in CYP2D6 encloses the side that remains open in 2C9 (Rowland et al., 2006) (Figure1.4).

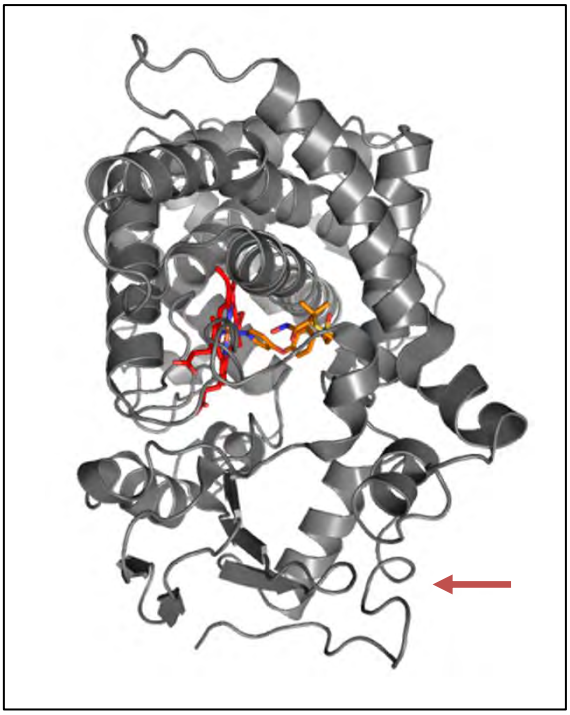


Figure 1.4 : Representation of CYP2D6 structure (Rowland et al., 2006)

Figure depicts the ribbon representation of the protein structure of CYP2D6. Alpha helical regions and heme prosthetic regions (red) shown. In addition, the extended F helix is depicted by the red arrow.

1.4 Location and membrane topology

CYPs have been shown to be bound to the membranes by either one or two transmembrane peptides that are located within the N-terminal domain of the protein, with the chance of the active site having an additional membrane contact (Black, 1992). It has been shown that there is an N-terminal approximately 29 amino acid sequence that is responsible for insertion and "stop-transfer" signal into membranes and furthermore that membrane insertion may occur co-translationally (Sato, Sakaguchi, & Mihara, 1987). The CYPs are bound with their heme ligands lying somewhere between parallel and perpendicular to the membrane (Panel 1 Figure 1.5) (Black, 1992). It is also suggested that some CYPs have membrane anchors that can transverse the membrane, binding with the heme lying parallel to the membrane (Panel 4 and 5, Figure 1.5) (Nelson & Strobel, 1988). Recently, there has been evidence to suggest that the full length membrane anchor is tilted by 17° to the lipid bilayer normal (Yamamoto et al., 2013) (Figure 1.6).

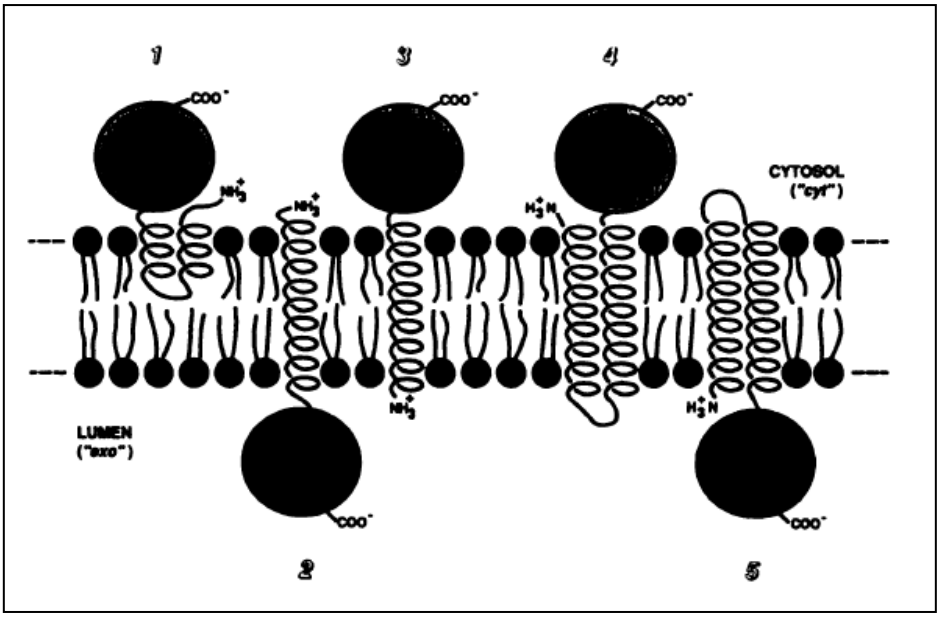


Figure 1.5: Possible membrane topologies of CYPs inserted into membranes (Black, 1992)

Figure depicting a schematic representation of possible topologies of CYP membrane anchoring, with balls representing the globular CYP structure and helical tails representing the membrane anchor signal peptide.

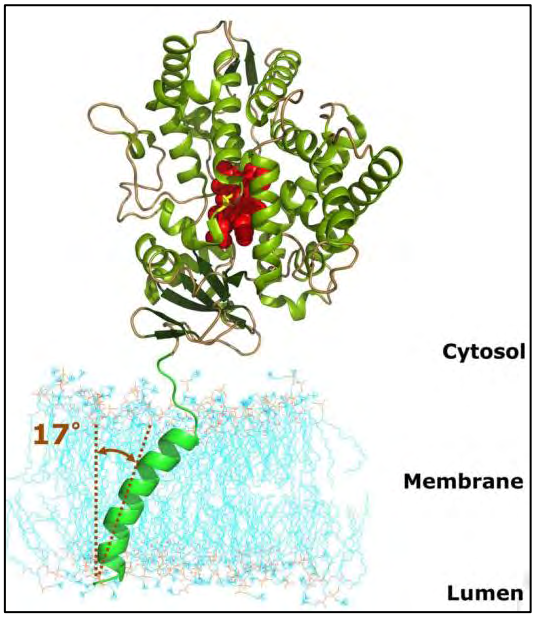


Figure 1.6: Topology of membrane binding using full length CYPs analysed by solid state NMR using artificial membranes (Yamamoto et al., 2013)

Figure depicting the topology of CYPs binding to artificial membranes and the 17° tilt to the normal.

CYPs are also known to bind to the inner mitochondrial membrane when they are expressed in the adrenal cortex, gonads and the kidney (Omura, 2006). Here, they have an endogenous steroidogenic function and are involved in several steps in aldosterone, pregnenolone and cortisol production and whilst there is evidence that these mitochondrial CYPs may have drug metabolising properties, their active sites appear more rigid and specific for their endogenous substrates compared to their microsomal counterparts (Anandatheerthavarada et al., 1997).

1.5 Accessory proteins (CPR, b5, dimers, oligomers)

CYPs function as part of a catalytic cycle involving electron transport originating from NADPH. In order for the electrons to reach the heme group during the catalytic cycle, the electrons need to be shuttled from NADPH to the CYP via CPR protein. CPR is 678 amino acids in length and is 76.93 kDa in molecular weight. It has binding sites for several ligands including the flavoprotein binding site (for cytochrome b5 binding), CYP binding interface and the nucleotide binding sites for flavin adenine dinulceotide (FAD), flavin mononucleotide (FMN) and NADPH (Figure1.7).

Cytochrome b5 (Cyb5) is a small enzyme of 134 amino acids in length, with a molecular weight of 15.33 kDa and which functions as an electron carrier for several monooxygenases. This enzyme is also membrane bound via a 40-44 amino acid membrane anchor peptide present at the C-terminus of the protein (Dürr, Yamamoto, Im, Waskell, & Ramamoorthy, 2007). CyB5 contains a heme prosthetic group that functions as an electron carrier and reduction of membrane bound CyB5 is rapid - especially compared to the relatively slow reduction in unbound form - thereby allowing efficient reaction kinetics (Rogers, Strittmatter, & Spatz, 1972).

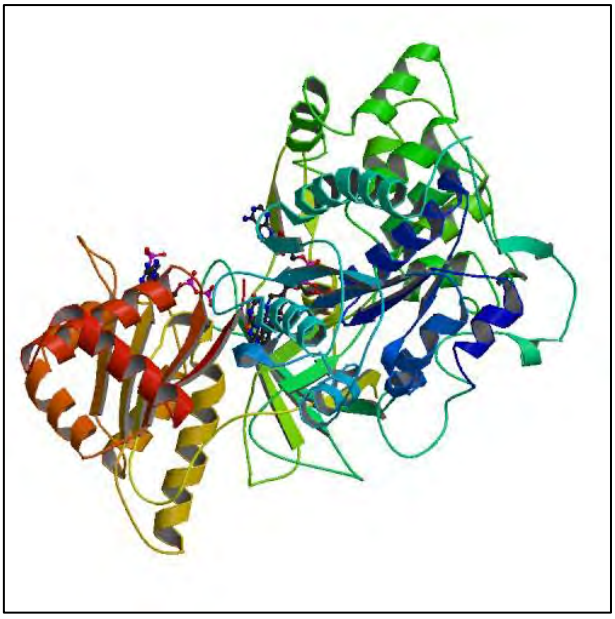


Figure 1.7: Crystal structure of CPR (Xia et al., 2011)

Figure depicts the ribbon structure of the structure of CPR protein; figure shows the alpha helical and heme substrate interaction regions

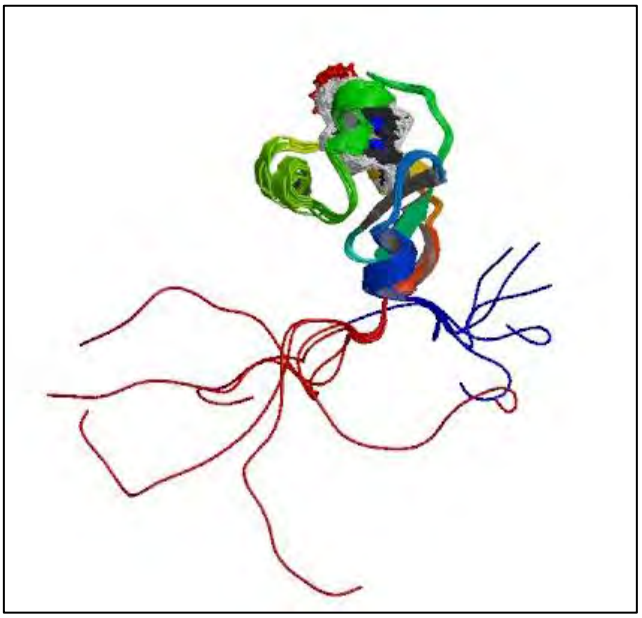


Figure 1.8: Crystal structure of cytochrome B5 http://www.rcsb.org/pdb/explore/explore.do?structureId=2I96

(currently unpublished)

Figure depicts the ribbon representation of the protein structure of CyB5, helical regions and linear unstructured regions shown.

1.6 Tissue distribution and regulation of expression

The endogenous role of CYPs varies along with their tissue distribution patterns. CYPs have been implicated in retinoic acid metabolism which has implications for the germ cells of the gonads (Seliskar & Rozman, 2007). Certain CYPs also have a role in cholesterol metabolism and steroid synthesis as well as in bile synthesis. Interestingly, there is evidence of patterned distribution in the various cells of the pancreas and it is known that some CYPs interact with the golgi body and may be cytoplasmic in the pancreas, whilst in females there is a frequent localisation of CYP2E1 with the nuclear membrane (Seliskar & Rozman, 2007). CYPs are also expressed in the skin and in the epithelial lining of the intestine; in agreement with CYPs having a second biochemical barrier function (Ahmad & Mukhtar, 2004). This may detoxify compounds entering through the skin or intestine but which could also lead to harmful small molecules being produced by biotransformation reactions. CYPs have also been shown to be expressed in the brain and lungs whilst for their endogenous steroid synthesis function, CYPs have been shown to be expressed highly in the cortex of the adrenal glands (Miksys & Tyndale, 2013).

Interestingly CYP gene expression is induced by steroid hormones and several drugs are known to induce the expression of specific CYPs, for example rifampicin induces CYP3A4 expression (Honkakoski & Negishi, 2000). However, when certain human protein chimeras were investigated in animal models using reporter constructs, the phenomenon was not witnessed suggesting species specific receptors and transcription factors are involved in CYP gene regulation (Guengerich, 1999).

Associated with the CYP3A4 gene there are several steroid response elements such as the glucocorticoid response element and the estrogen response element. These are activated by nuclear hormones and elicit a response by activating or inhibiting CYP transcription. The orphan pregnane X receptor is also known to induce CYP expression via transcription regulation mechanisms (Goodwin, Redinbo, & Kliewer, 2002).

Administration of the anti-tubercular drug rifampicin induces CYP3A4 expression in the liver by a classical Glucocorticoid Receptor (GR) signalling mechanism (Guengerich, 1999). Rifampicin acts as

a non-steroidal ligand for the GR which becomes activated upon rifampicin binding and translocates into the nucleus where it binds to response elements in the promoter of CYPs. This bound GR then recruits and stabilises co-activators to the promoter region of CYPs, thereby increasing transcriptional activity and eventually translating to an increase in CYP3A4 protein levels.

This has many implications for rifampicin administration and drugs co-administered particularly in developing nations such as South Africa and India that have a high incidence of tuberculosis and HIV, and where combination drug therapy regimes are required to combat co-infections. The poly-pharmacy of multi-drug combinations should ideally therefore be tailored to the patients' specific needs in terms of their CYP protein expression levels and could be of great value in the treatments of infectious diseases where individual drugs need to be within the therapeutic window for optimal treatment. An altered drug pharmacokinetic profile could see drug plasma concentrations drift towards the sub-therapeutic range and therefore inadvertently promote emergence of drug resistance in the infectious organism.

1.7 Function – catalytic cycle – reactions

In the resting state the iron is in the ferric state (Fe^{3+}) and has an axial thiolate ligand provided by the conserved cysteine residue. Substrate binding displaces an axial water ligand from the heme and causes a change in the spin state of the iron allowing for electron transfer to occur from a electron source or redox partner; typically the electrons originate from NADPH with electron transfer shuttled through the CPR flavins resulting in reduction of the heme iron of CYP to the ferrous (Fe^{2+}) state. This ferrous-heme species can then bind to molecular oxygen, forming the ferrous dioxy complex (Figure 1.9). A second electron transfer can then occur, also from an electron source such as NADPH via CPR and, after protonation, yields the ferric hydroperoxo complex, this is thought to be a rate limiting step in the catalytic cycle. The ferric hydroperoxide species is unstable after further protonation breaks down by homolysis to yield a radical cationic iron-IV oxo species. This is known as compound I and is the species that abstracts hydrogen from the substrate and forms a further FeIV species that rapidly breaks down to form the hydroxylated product. Finally, the product is released

and the resting ferric state is once again reached. The evidence of this mechanism has been extensively investigated and has been reviewed previously (Munro et al., 2013).

During the catalytic cycle, electrons are shuttled to CYP from NAPH via CPR. The first step in this cycle involves NADPH binding to CPR; this is followed by a two electron transfer in the form of hydride ion transfer from NADPH to FAD in CPR. The FAD becomes reduced and then transfers two electrons to the FMN in CPR. Finally, FMN reduces the heme iron in CYP by sequential single electron transfers. It is suggested that there are several conformational changes in CPR to facilitate efficient electron transfer and that the addition of glycerol may hinder conformational flexibility and therefore reduce electron transfer by CPR (Laursen, Jensen, & Møller, 2011).

This complex catalytic cycle can be short-cut by the addition of a monoxide (RO) or peroxide (ROOH) species that can react (stage 5 in Figure 1.9) directly forming the ferric hydroperoxo complex without the need for electron transfer from NADPH and CPR. This "peroxide shunt" pathway can be a useful means to reconstitute activity in the absence of membrane binding. The disadvantage of this method is that the heme can become destabilised over a period of time by peroxides or reactive oxygen species, causing oxidative degradation of the heme and/or oxidative damage to the protein. For that reason, sustained activity cannot be successfully reconstituted using this method. However, this peroxide shunt pathway is a useful tool to validate the catalytic activity of the enzymes especially when using this method in conjunction with probe substrates.

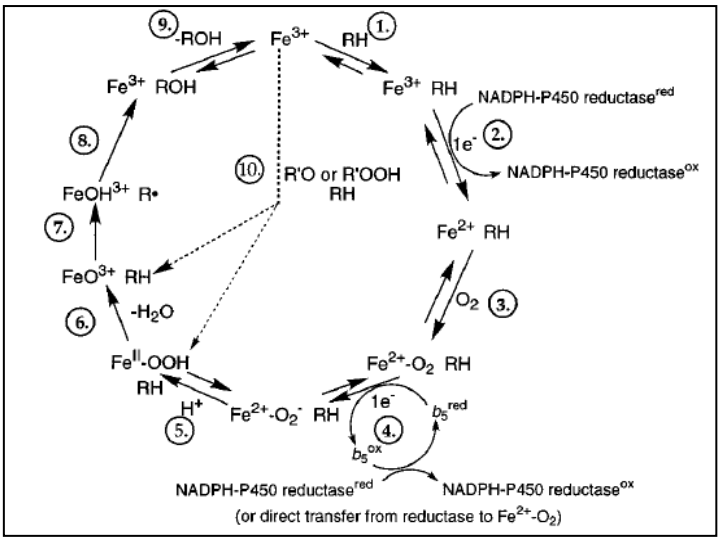


Figure 1.9: Catalytic cycle of CYP mediated hydroxylation (Guengerich, 2001)

Schematic diagram showing the CYP catalytic cycle and heme iron states with electron transfer from NADPH and showing the short-cut "peroxide shunt" pathways

While this is the classical mechanism of the CYP catalytic cycle, there are several controversial factors present. The spectroscopic proof of compound I has been elusive, but in using the hydroperoxide shunt pathway, modern molecular biology and analytical techniques, compound I has recently been verified and characterised providing direct evidence to support the mentioned catalytic mechanism (Rittle & Green, 2010). Compound I is an iron oxo species with an additional oxidising equivalent delocalised over the thiolate and porphyrin ligands (Rittle, Younker, & Green, 2010). Another controversial factor is the resting iron state, recent evidence suggests that the resting state of the iron exists in a mixture of ferric and ferrous iron, which may be counter intuitive and result in redox cycling to produce reactive oxygen species under aerobic conditions (Johnston et al., 2011).

CYPs catalyse several monooxygenase reactions including C-hydroxylations, heteroatom oxygenation, hereroatom dealklyation, epoxide formation and several others less commmon reactions (Krest et al., 2013). Surprisingly there are also some reduction reactions catalysed by certain CYPs under anerobic or hypobaric conditions, as well as non-redox chemistry involving acid hydrolysis, but mechanistic details of these remain unclear (Guengerich & Munro, 2013). Other unusual reactions include isomerisation or rearrangement reactions, notably those CYPs involved in prostagladin

synthesis and in several coupling reactions in alkaloid biosynthesis in plants (Guengerich & Munro, 2013).

Several alternate methods exist to confer electron transfer to CYPs driving catalysis which make use of natural cofactors or synthetically driving electron transfer. The most common way is to reconstitute the CYPs together with CPR and the add NADPH or an NADPH regenerating system. In order to sustain CYP catalysis for long periods without becoming rate limiting or inhibiting due to NADPH depletion. There are several NADPH regenerating systems available including the pyruvate CoA system and the glucose-6-phosphate dehydrogenase system (Munro et al., 2013). A simpler strategy would be to directly include NADPH (obtained commercially) into reconstituted CYP-CPR systems, however this may have the disadvantage of excessive NADPH inhibition of catalysis or NADPH decline as the reaction proceeds, plus NAPDH is a comparitively expensive reagent. The hydroperoxide shunt using either hydrogen peroxide or, more commonly, cumene hydroperoxide can be a useful means of driving CYP catalysis without the need for redox partners or NADPH, but has the disadvantage of being inefficient in forming catalytic intermediates and damaging to the protein stability, so is therefore not optimal for sustained catalysis. There has been some success in using the perchloro benzoic acids to transfer the oxygen and hydrogen to the CYPs directly forming compound I and this method was successful in producing and identifying active compound I in CYP experiments (Groenhof, Ehlers, & Lammertsma, 2008).

1.8 Substrate specificity

While many CYPs are known to be highly specific, certain CYPs involved in drug metabolism are also known to bind many substrates. An example of this is CYP3A4 which has a large and flexible active site capable of binding several different types of substrates as well as multiple substrate molecules simultaneously. This substrate promiscuity can lead to the enzyme being inhibited by many different compounds. It is therefore useful for there to be many routes of drug metabolism by CYPs such that if one route becomes inhibited there remain alternative routes for metabolism, justifying redundancy in CYP metabolism. This also turns out to be useful in maintaining the pharmacokinetic

profile of the drugs being administered singly or in combination therapy. Interestingly, there is evidence that supports gating mechanisms as a means to regulate the entry of substrates into the active site of certain CYPs, whilst specific active site residues also play a direct role in substrate recognition. There are thus several means by which CYPs achieve their substrate specificity. It is also useful to compare CYP similarity within and between species commonly used to model human drug metabolism. An example is the CYP2C family all having more than 80% amino acid sequence similarity yet have differing substrate specificity, for example CYP2C9 has Warfarin hydroxylation activity while CYP2C8 does not (Guengerich, 1997). Another example is the difference in activity between rabbit and human CYP2C families; CYP2C3 is the major steroid hydroxylating enzyme in rabbits but this type of activity in humans predominantly occurs in the CYP3A family (Guengerich, 1997).

A generalisation can be made in regard to the molecular, electronic and physicochemical properties of substrates and the CYPs that recognise and metabolise them (Table 1.1). This was achieved by the study of Quantitative Structure Activity Relationships (QSAR) of known chemical substrates and the relative crystal structures available, with the substrates being modelled into the CYP active sites and general characteristics identified from the docked complexes (Lewis, 2000). These generalisations can be useful in predicting which CYP can recognise a substrate but currently does not enable the *in silico* prediction of the precise site(s) of metabolism of a given compound, nor of the exact isoform responsible. Nevertheless, a general substrate profile, toxicophore or pharmacophore can be outlined for each of CYP3A4, CYP2C9 and CYP2D6 (Table 1.1). CYP3A4 allows for relatively large substrates due to its large active site capacity and lipophilicity due to amino acid residues in the substrate recognition sites of the protein, while CYP2C9 has substrate preference for more weakly acidic compounds (Figure1.10). While these generalisations apply in most cases, there are always exceptions that have to be taken into consideration during any computational predictions.

Table 1.1: General properties of substrate selectivity of CYP3A4, CYP2C9 and CYP2D6 adapted from (Lewis, 2000)

CYP	General properties displayed by most substrates
3A4	High volume, relatively lipophilic, structurally diverse with one or two hydrogen bond donor/acceptors at 5.5–7.5 Å and 8–10 Å from the site of metabolism.
2C9	Weakly acid, fairly lipophilic with one or two hydrogen bond donor/acceptors at 5–8 Å from the site of metabolism.
2D6	Basic, relatively hydrophilic, usually contain an aromatic ring and a hydrogen bond donor/acceptor, basic nitrogen at 5–7 Å from the site of metabolism.

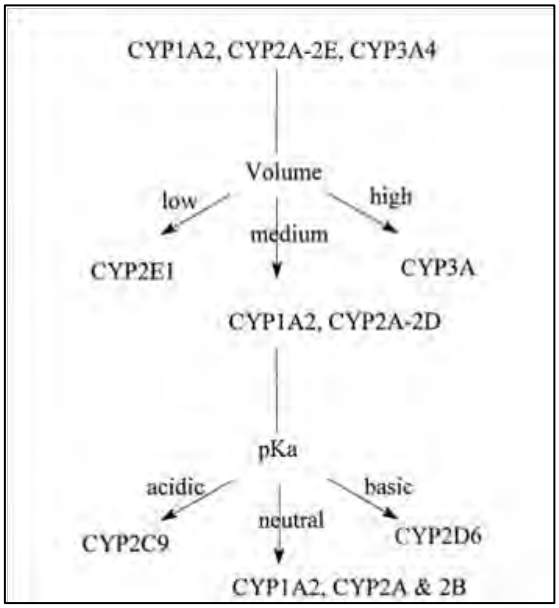


Figure 1.10: Generalisation of substrates and their physicochemical properties in relation to the CYPs that metabolise them (Lewis, 2000)

Flow diagram showing the generalised substrates with regards to volume and pKa, the physicochemical properties of drug substrates for the most common CYPs involved in drug metabolism

1.9 Enzyme kinetics behaviour and atypical kinetic plots

CYPs are known to exhibit atypical Michaelis-Menton enzyme kinetic profiles. It is widely accepted that several CYPs including CYP3A4, CYP2C9 and CYP2D6 can exhibit atypical enzyme kinetics with various substrates and that this is seen reproducibly. This is known to occur *in vitro* but does not tend to correlate with *in vivo* data; this may be accounted for by masking phenomena in the bulk effect seen *in vivo*. There are several mechanisms proposed to account for atypical kinetic profiles, as reviewed previously (Atkins, 2005). Atypical in this regard refers to non-hyperbolic profiles of substrate velocity plots (Figure 1.11). Some of the proposed mechanisms include dual or several substrate binding events occurring simultaneously within the active site of even within allosteric active sites that may exist on certain CYPs. Several other mechanisms include the substrate and/or product based inhibition (Shou et al., 2000). There have been numerous models suggested to describe this type of kinetic behaviour using multi-site descriptors or intermediates reacting with substrate or product leading to altered kinetic profiles (Houston & Galetin, 2005; Manoj et al., 2010). Nevertheless, this has implications for enzyme kinetic parameters derived from Michaelis-Menton plots and estimates of maximum rate of reaction (V_{max}) or substrate concentration at half V_{max} rate constants (K_m) must be interpreted with care, with the kinetic model used taken into consideration.

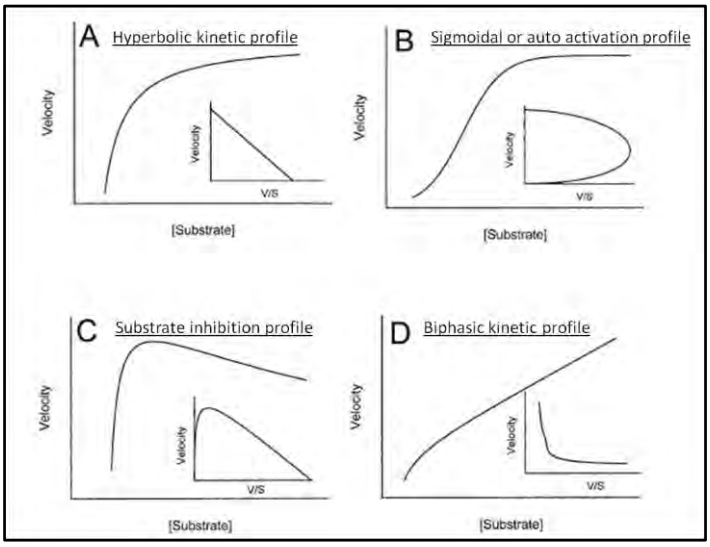


Figure 1.11: Kinetic profile of Mechaelis Menton and atypical kinetics displayed by CYPs adapted from (Hutzler & Tracy, 2002)

Representation of kinetic profiles showing classical Michaelis-Menton kinetics (A), Sigmoidal or auto-activation profiles (B), substrate inhibition kinetic profiles (C) and biphasic kinetic profiles (D). Representations show velocity-substrate plots and velocity-turnover plots (inlay).

1.10 *In vitro-in vivo* extrapolation (IVIVE)

The end point of *in vitro* CYP experiments can only be useful if the data can be applied to *in vivo* drug metabolism. The initial extrapolation factor involves the conversion of recombinant CYP activity to its equivalent in human liver microsomes, this can be achieved by applying a scaling factor dependant on the CYP and substrates relative activity factor or some intersystem extrapolation factor (Crewe, Barter, Yeo, & Rostami-Hodjegan, 2011). It is necessary to consider the impact of atypical kinetic plots in making this adjustment since it is commonplace to force or omit data regions in fitting data from atypical CYP enzyme activity plots to incorrect models, which can result in inaccurate V_{max} and K_m estimations (Houston & Kenworthy, 2000).

The *in vivo* scenario involves a more complex interaction of CYPs with other factors including drug efflux transporters and phase II drug conjugation reactions. IVIVEs should therefore be applied with caution to *in vitro* data that may be generated from high throughput screens and should subsequently be verified using *in vivo* models (Crewe et al., 2011). The simplest mode of extrapolation is the "relative activity factor" which aims to compare the activity in recombinant CYPs to microsomal CYPs. The models have to be built up depending on the method of assay: for example if human liver microsomes are assayed then the extrapolation would first be to rate of reaction per picomolar active CYP, then extrapolated to per mg liver weight and then to total liver volume proceeding to plasma concentration or bioavailability. There are several models of varying complexity from simple compartment models to physiological based multiple compartment models in existence. There are other factors that also need to be accounted for such as age and weight and volume of the liver. A one month old will have reduced CYP levels compared to an older individual. There are also other factors such as height, weight, cardiac output and body surface area that decide blood flow and will determine drug flux rates that are accounted for in the more complex models.

1.11 Interim conclusion and implications from molecular studies on CYPs

Over the past 40 years there has been a great expansion in the understanding of CYP enzymes (Figure 1.12). CYPs have been identified in all life forms from bacteria, viruses, plants and mammals (Lamb et al., 2009). CYPs have a variety of endogenous functions including steroid synthesis and fatty acid hydroxylation. These diverse enzymes are also the most important enzymes in first pass liver metabolism. CYP3A4, CYP2C9 and CYP2D6 play the primary role in phase I metabolism that serves to facilitate excretion of foreign compounds which enter the body. This is usually achieved by mono-oxygenation reactions that make compounds more hydrophilic or prime them for phase II drug metabolism conjugation reactions. Such metabolism contributes to the intrinsic clearance of a drug from the blood plasma and therefore has strong implications for the drugs pharmacokinetic profile. Therefore CYP enzymes are of considerable significance to the drug discovery and development process.

To date, the structural biology investigations of CYPs have been carried out primarily by X-ray crystallography and many crystal structures now exist both in unliganded and in ligand bound forms in the protein structure databank. These structures are revealing in terms of the details of ligand binding and have assisted somewhat in the understanding of the atypical kinetic plots that are associated with CYPs. However, the current structures exist in solubilised forms that have involved deletions and significant cloning to alter the endogenous structures, therefore further work with additional analytical techniques have an opportunity to shed light on other factors that still remain unknown, such as the influence of the membrane environment and cofactor such as CPR and CyB5 on the structures of CYPs.

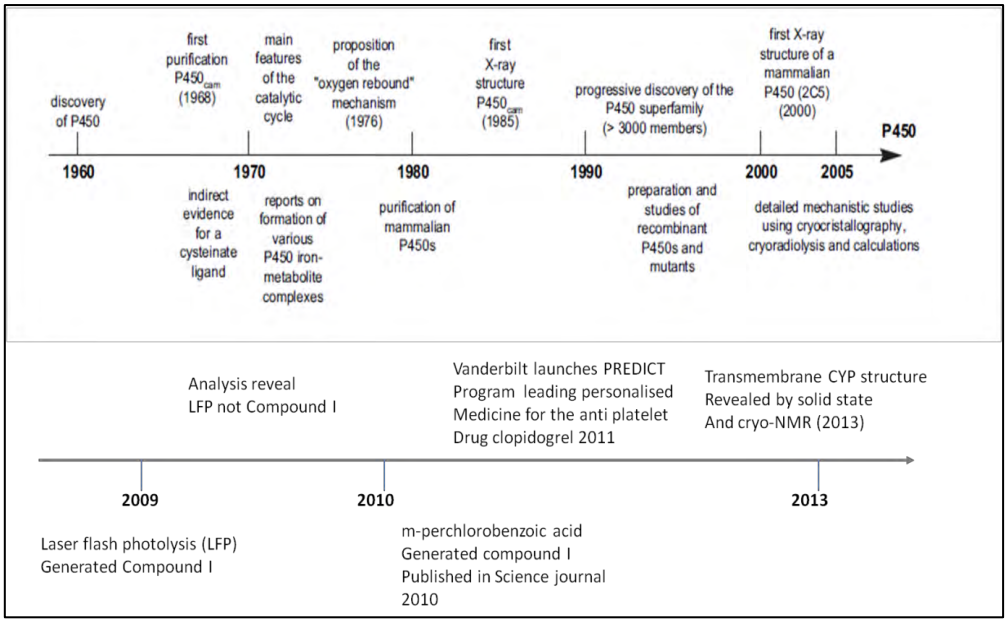


Figure 1.12: Timeline of CYP research progress during the past 40 years, extended from (Mansuy, 2007)

1.12 Role in drug metabolism: pharmacology of cytochrome P450s

Upon oral administration of a pharmaceutical agent, there are several processes that are at play before the drug reaches its target site. In more detail, the bioavailability of a compound is an essential parameter in determining a drugs therapeutic effect and the pharmacokinetic process begins with the drug being absorbed by the enterocytes of the lumen of the stomach or intestine. Several CYPs are expressed in these cells and the fractions of a given drug that becomes fully dissolute and enters the enterocytes constitute the “fraction absorbed” (Fabs). The fractions that enter the cells and avoid CYP metabolism within the epithelial luminal cells are referred to as (Fgi) (Figure 1.13); this is the unmetabolised fraction that is transported from gastro-intestinal mucosa to the liver via the hepatic portal vein. Once reaching the liver hepatocytes, there are sufficient CYP enzymes present to begin to metabolise the parent drug by mono-oxygenation reactions. The fraction remaining unmetabolised is referred to as the drugs “bioavailability” and is the component that is transported to the systemic circulation en route to the target cells where the therapeutic effect is hopefully achieved. This process is not linear and is cyclic in nature, resulting in any compounds that remain in the blood stream being recirculated to the hepatic system via the hepatic artery, which in time results in complete removal of the drug compound from the body.

The end product of first pass metabolism can be either a more soluble compound or a compound primed for phase II conjugation reactions that add on polar substituents, such as glutathione or glucuronide functional groups, that create more polar and hydrophilic compounds (Gunaratna, 2000). The increased hydrophilic nature of the compounds facilitates easier excretion via the renal excretory system. The tubules of the renal system also express relatively high amounts of CYPs (Spector, 2009). The more hydrophilic the compounds are, the higher the chance that they will remain in the kidney tubules and reach the urinary tract for micturition out of the body. By contrast, glutathione conjugated compounds are favoured for incorporation into bile salts; this occurs in the liver and bile salts are then transported to the gall bladder where they are mixed with pancreatic fluid, excreted into the duodenum and finally expelled out in faeces. Other means of excretion include excretion through the sweat glands, in saliva, in exhaled breath, through the sebaceous glands, in tears and deposited in hair, but the urinary and biliary are most common.

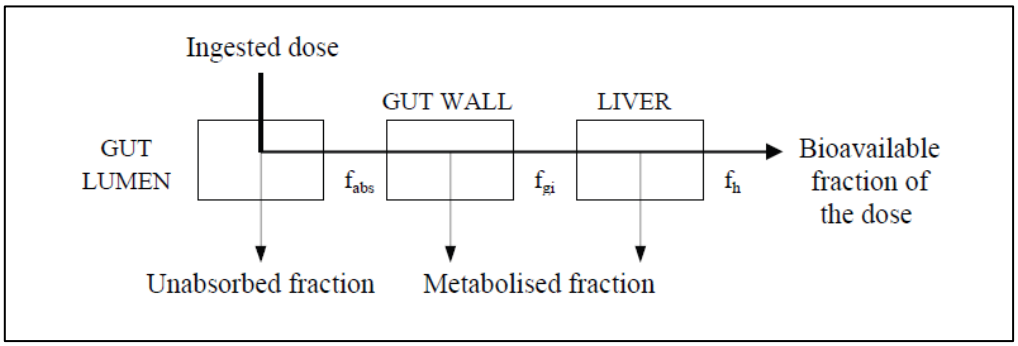


Figure 1.13: Pharmacokinetic route of metabolism and bioavailability (Gunaratna, 2000)

The figure depicts a schematic showing the fraction of drug passing through the body compartments and reaching the systemic circulation, a drugs bioavailability.

1.13 Polymorphisms and pharmacogenetics – major CYPs and accessory proteins

CYPs can exist in various forms resulting from single nucleotide polymorphisms (SNP) or alterations in genetic structure by frame shifts or epigenetic changes. Several forms of CYP SNPs exist that may result in increased CYP expression, as well as higher or lower catalytic activity. The frequency of SNPs varies in various ethnic groups and even in individuals (Ingelman-sundberg, 2004). A further understanding of the effects of these polymorphisms and their relation to drug metabolism would advance the field of personalised medicine for patient stratification and tailored drug treatments with fewer side effects.

By way of example, CYP2D6 is known to have several SNPs that can have a significant effect on its activity, causing a range of activity from poor metabolisers to efficient to ultra-rapid metabolisers. Such alterations in CYP activity can in principle affect the drugs pharmacokinetics at usual dosing and could lead to increase adverse events in poor metabolisers and no therapeutic effect in ultra-rapid metabolisers (Al-shurbaji, Bertilsson, Dahl, & Dalen, 2002; Ingelman-sundberg, 2005). CYP2C9 is also known as a highly polymorphic CYP, with CYP2C9 SNPs resulting in altered drug clearance or therapeutic properties (Zhou, Zhou, & Huang, 2010). Furthermore, CYP3A4 is also highly polymorphic, although in this case conventional wisdom suggests that the relatively high expression levels of CYP3A4 would serve to negate any effects of polymorphic variation on turn-over rates, which may be a very simplistic rationale and not be entirely correct. In addition there have been several polymorphic variants of CPR indentified, which could have wide reaching implications not only for drug metabolism but for other endogenous CYP functions and which could also cause diseases such as CPR deficiency syndrome (Chen, Pan, Naranmandura, Zeng, & Chen, 2012).

1.14 Drug-drug interactions

CYPs have important implications for drugs that are co-administered in combination therapy since they can alter the pharmacokinetic profile of drugs that may be substrates for respective CYPs. The alteration of PK plots can occur by several proposed mechanisms, including competitive inhibition. If two or more drugs are substrates for the same CYP they will compete for binding sites on the enzyme.

The drug with the higher binding affinity will have a higher chance of binding and will be metabolised at an altered rate compared to binding in the absence of a competing compound. In addition, the induction mechanism causes the deregulation of CYP gene expression causing altered PK plots. It is therefore currently preferred that any drug be metabolised by several CYPs or not metabolised by CYPs at all in order to account for CYP inhibition by commonly co-prescribed drugs.

There are several dietary phenol containing compounds that can inhibit CYP, including compounds in grapefruit juice and several herbal composites such as St John's Wart, Ginko Biloba and bergamottin (Kimura, Ito, Ohnishi, & Hatano, 2010; Yale & Glurich, 2005). Certain fruit juices such as grapefruit juice are contraindicated when taken with certain drugs especially drugs that are substrates for CYP3A4, since they contain the potent CYP3A4 inhibitor ketoconazole (Guengerich, 2003). There has also been some work investigating the interaction of drugs with flavinoids and other plant phytochemicals that are present in fruits and vegetables (Rodríguez-Fragoso & Reyes-Esparza, 2013). These interactions can potentially cause bioinactivations of the drug compound by the dietary compound or scavenger mediated, by CYP or drug transporter inhibition/ induction or more subtly by effecting gastric emptying or intestinal motility. These interactions are known to occur in humans *in vivo* both by combination therapy and by herbal or dietary interactions.

1.15 Adverse drug reaction

Adverse drug reactions can result from altered PK profiles that elevate drug plasma levels to toxic levels, thereby posing a potential for altered cellular and tissue homeostasis. In addition to altered PK properties harmful metabolites can be formed, for example by biotransformations liberating carcinogens or elevated reactive oxygen species that directly induce DNA damage (by inducing strand breaks) or constitutively activating or inhibiting signalling molecules or other enzymes. Another mechanism of adverse drug reaction is hapten formation: here, a drug or metabolite illicits an excessive immune response which can be dangerous, leading to anaphylaxis or drug induced liver damage. A mild or severe drug reaction has to be investigated before the drug can be used effectively

in the wider population, and there are several drug screens that are available to attempt to monitor this (Shu, Johnson, & Yang, 2008; Uetrecht, 2003) All these factors have to be taken into account for successful drug treatment and medicine regulatory authorities require that pharmaceutical agents show the "risk of adverse reaction" to "benefit of treatment" ratio by providing information on CYP substrate and inducing or inhibiting ability of the drug prior to approval (Zhang, Zhang, Zhao, & Huang, 2009).

1.16 Need for high throughput drug metabolism screening

Drug metabolism today remains the preserve of a limited number of major pharmaceutical companies and contract research organisations. There is therefore a need for a robust, widely available research platform to investigate the drug metabolism of new chemical entities, prior to their use in humans to treat diseases. Owing to the current drug discovery pipelines which make use of vast compound libraries which are screened for their activity, there is also the need for a format to screen for favourable drug metabolism parameters in high throughput. Miniaturised high throughput screening has the benefit of reducing reaction volumes so more compounds can be analysed in parallel with reduced reagent volumes and having cost saving potential, taking economies of scale into account. Analysis of large samples coming from large compound libraries manually is prone to errors and inaccuracies over time. This necessitates the need for automated approaches in the analysis of large libraries in order to ensure optimal data validation, integrity and reduced chance of experimental errors to ensure reproducibility (Rodrigues, 1997).

Furthermore, the current *in vivo* animal models of drug metabolism are not reflective of the human drug metabolome and are increasingly ethically complex. The challenge of modelling human complexity in animal models can lead to misinformation or inaccuracies in data interpretation (Nicholson, Holmes, Lindon, & Wilson, 2004). Modelling human drug metabolism in more accurate and reflective *in vitro* systems may therefore be advantageous for high integrity toxicity data to facilitate better decision making in risk-benefit ratios of new chemical entities, thus accelerating drug discovery and reducing costs of improved therapeutics. Furthermore, the creation of an efficient, high

throughput drug metabolism format will augment systems biology approaches to understand connections within complex pathological effects and their treatments, ultimately leading to an improved understanding of disease mechanisms and the holistic modern therapies (Hood & Perlmutter, 2004).

Recently, a functional protein CYP3A4 microarray has been developed by the Blackburn lab that has shown promise (Blackburn, Shoko, & Beeton-kempen, 2012). This protein microarray platform employs a functional protein immobilisation strategy with high throughput capabilities in principle for the modelling of human drug metabolism systems with CYP3A4. Functional protein microarray technology can make use of existing microarray workflows and automation systems, thus integrating to currently available resources. Further expanding this technology will have the benefit of developing a reflective model of human drug metabolism with the additional benefit of miniaturisation for high throughput and reagent saving, as well as highly automatable workflows for robust data integrity and reproducibility.

1.17 Conclusions and aims

A robust, cost effective and reflective drug metabolism screening platform is required for efficient elucidation of drug metabolism parameters in an academic and biotechnology environment. This information is required for the determination of the risk-benefit ratios of new chemical entities for the drug discovery process to discern better drugs with fewer side effects, as well as a means to better understand how to model human drug metabolism without recourse to animal studies. The aims of this thesis are therefore to further develop the functional protein platform by applying the CYP protein microarray technology to the other CYP isoforms, namely CYP2C9 and CYP2D6. In addition, this thesis aims to explore the use of solubilised CYP activity reconstitution in the absence of immobilisation, utilising label free detection with probe substrates and a complex drug, rifampicin, as a novel approach in the drug metabolism field. The strategies employed to obtain high yields of active protein have been developed and optimised using a heterologous bacterial expression system and simple one step purifications. The method of immobilisation used was chosen to specifically bind

proteins in a uniform distribution onto a functionalised hydrogel polymer layer tethered onto a solid glass microarray slide. The hydrogel provides the necessary aqueous environment and the flexibility required for protein and substrate diffusion and transient redox partner binding. The platform is easily amenable to medium and high throughput using existing, available and routine DNA microarray technology and equipment.

To analyse CYP activity, label-dependant fluorescent and label-free mass spectrometric detection techniques were employed. Quantitation was achieved using fluorescent standards or parent compounds recommended by drug regulatory authorities as probe substrates for CYP research. Data generated was used to construct enzyme kinetic profiles and allowed the estimation of Michaelis-Menton parameters such as V_{max} and K_m from fluorescent functional protein microarray experiments. Furthermore, label-free quantitation was used to quantitate metabolite formation using a high protein concentration without immobilisation in drug incubations.

Validation of the technology was investigated by assaying the sensitivity, limit of detection and intra-assay accuracy and precision. Furthermore, quality control was investigated to monitor reproducibility on several levels including protein quality and surface chemistry uniformity. Going forward, further cross validation studies should be done to further demonstrate the efficacy of this technology.

Chapter 2 – Cloning, expression and systematic characterisation of heterologous CYP enzymes

2.1 Introduction: Current CYP protein technology

There are several methods currently employed to investigate the effect a drug has on CYP drug metabolism systems. These all have their advantages and disadvantages and some are amenable to high throughput workflows. The importance of high throughput workflows is indispensible to drug metabolism research working with expansive compound libraries where economies of scale play a role in the preclinical drug development and with compound libraries reaching thousands to millions of chemical entities. The current drug discovery process favours the "fail fast, fail early, fail cheap" philosophy that aims to stratify massive compound libraries and identify most promising compounds early to develop medicinal chemistry efforts around only the most promising hits. This is beneficial in reducing costs of drug development and subsequent clinical analysis and can reduce time to market by more concentrated efforts.

There are three main routes for drug metabolism investigation, namely the *in vitro*, *in vivo* and *in silico* or computational. The *in vivo* assays usually require investigations in mice, rat, dog and non-human primate models prior to human clinical investigations. While there is some degree of similarity between animal models and humans there are several limiting factors present in these *in vivo* models. There is sequence similarity at the gene level but CYPs are highly polymorphic and at the protein level there could be vast differences in structure and catalysis. In addition to the difference in structure and sequence, the expression levels are not necessarily comparable to that of the human liver. Therefore while this toxicology data is still required for safety clearance to perform clinical trials in humans there is still the need for more accurate human liver models from *in vitro* systems. *In silico* prediction relies heavily on the laboratory validation which necessitates the need for high quality and reproducible CYP *in vitro* systems.

Current *in vitro* systems make use of human CYPs isolated from several sources. The initial requirement of CYP assays is the high quality, reliable and reproducible supply of CYP enzyme preparations for *in vitro* drug metabolism analysis.

2.1.1 Hepatocytes

Primary hepatocytes contain the full complement of CYPs and phase II enzymes but have to be obtained by ethical consent from patients or donor groups or from rat models (Table .2.1). Rat liver hepatocytes are usually treated with phenobarbital to induce CYP expression to ensure significant CYP yields. The advantage of these protein sources is that there have been substantial experimentations and good IVIVE intra-system factors have been derived. The disadvantage is that these hepatocytes have to be obtained from whole liver or liver biopsies and require a constant supply. This is a limiting factor as there is restricted availability of human samples and the supply can vary depending on gender, age, CYP polymorphism and so on. This leads to highly variable supply and poor reproducibility which can complicate data analysis and confound variables in IVIVE. Furthermore, primary hepatocytes can be cultured only up to a few passages as there is rapid decline in CYP expression in simple adherent monocultures. Several other culture techniques have been developed with varying success. These include co-cultures in suspension or adherent environments with extracellular matrix producing cells and even three dimensional culture conditions using synthetic scaffolds and hydrogel or matrigel matrices. There are even miniature three dimensional microfluidic chips that aim to produce viable hepatocyte cultures with higher reproducibility and ability to multiplex (Toh et al., 2009). Several other "liver equivalent" chip applications have been developed using complex micro-patterning or extracellular mimics and intricate perfusion. This research is still in the early stages of development and, while promising, is costly at present and is lacking in comparison to donor liver primary cells (Materne, Tonevitsky, & Marx, 2013).

In addition to supply issues mentioned above, the liver samples have to be further processed to isolate the CYPs. This is usually done by lysis and differential centrifugation to separate the soluble and membrane bound cellular fractions (van der Hoeven & Coon, 1974). The membrane bound fractions

are the fraction of interest as this contains the majority of CYP protein. The soluble fraction contains the abundant cellular proteins required for normal hepatocyte functioning and those responsible for phase II drug metabolism. The "S9" or microsomal fraction is the term for the isolation of CYP by differential centrifugation.

2.1.2 Microsomes

While there is relatively complex protein isolation and erratic supply issues with using microsomal protein fractions from human donors, this still remains the gold standard for CYP drug metabolism screening at present. There have been strategies employed to minimise reproducibility errors seen with heterogeneous CYP isoforms and supply. These include forming CYP donor pools of 50 to 200 donors which are age and gender matched and batch comparisons usually performed by commercial suppliers of microsomal fractions (Sjögren, Svanberg, & Kanebratt, 2012).

2.1.3 Precision cut liver slices

Another form of CYP accessibility is the precision cut liver slices. This is an *ex vivo* model touted to represent more the *in vivo* state compared to other *in vitro* bound sources. These slices contain CYPs present in their cellular compartment and located within extracellular matrix, which may retain a more natural structure and architecture. These sources contain the full complement of CYPs as well as other phase II enzymes; in addition to this the liver slices contain hepatocytes and other cell types. It has been shown that these liver slices correspond closer to the gene expression profiles to that of whole liver than hepatocytes cultures (Vanhulle et al., 2001). However the maintenance of liver slices can be challenging and complex, requiring constant perfusion, uniform and reproducible thickness requiring specialised cutting equipment.

2.1.4 Cell lines

Immortalised human derived cell lines have also been created, the most common being the HepG2 cell line and recent hepatoma derived cell lines (Gunaratna, 2000). These are human cell lines that have been modified to stably express all or some CYPs. There has been some promise with these cells

which have the capability of expressing all or individual CYPs at relatively high expression levels. The culture conditions have to be specific for hepatocyte growth and maintenance, storage is simple and cells can be flash frozen and revived. However, recent proteomic data suggests that there are wide differences in CYP expression levels between primary hepatocytes and commonly used cell lines, which further erodes confidence in data generated using cell lines (Schaefer et al., 2012).

These protein sources are useful but still have the disadvantage of protein isolation. While these protein sourcing methodologies have the advantage of isolating active membrane bound CYP and redox partners; there exists the distinct disadvantage of microsomes having a heterogeneous mix of CYPs and even trace amounts of other enzymes co-isolated. Therefore, upon assaying for a single CYP reaction there is the need to inhibit all other potentially interfering reactions. This is overcome by means of a cocktail of inhibitors that inhibit several other reaction processes while allowing the reaction of interest to proceed (Dierks et al., 2001; Pillai, Strom, Caritis, & Venkataramanan, 2012). Of course, cross-reactivity of inhibitors with the reaction of interest is a limiting factor of this type of strategy. The addition of inhibiting components also has the effect of increasing cost per assay conditions which has financial implications particularly when working with high throughput workflows.

2.1.5 cDNA expressed protein

cDNA expressed CYPs has the benefit of expressing a single CYP, or fewer CYPs, in a controlled manner that negates the use of complex inhibitor cocktails in CYP activity assays. There have been several attempts to express human CYPs in heterologous systems including yeast, bacteria, and insect and plant cells (Doty et al., 2000; C. A. Lee, Kadwell, Kost, & Serabjit-Singh, 1995; Renaud, Cullin, Pompon, Beaune, & Mansuy, 1990). The advantage of these systems is that they have relatively simple culture conditions and can improve reproducibility by expressing individual CYPs and even individual CYP isoforms and polymorphisms (Table 2.1). Another advantage of cDNA expression is that cloning can be incorporated to facilitate soluble expression and simple purification and isolation. The protein production can be highly controlled to express at certain times under certain conditions

having controlled post translational modifications and even synthetic modifications for immobilisation.

Table 2.1: Review of CYP protein technology

CYP technology	Advantages	Disadvantages	Availability
Human liver microsomes (HLM)	Highly active cyps, no addition of cofactors	Mixture of cyps, have to use complex mix of inhibitors for specific activity	Very limited, have to be harvested from donors or donor pools
Recombinant microsomes	Individual cyps can be expressed, protein can have PTMs	Can be challenging to express, specific media requirements	Easily available and more reproducible
Rat liver microsomes (RLM)	High similarity to HLMs	Limited substrate availability	More readily available than HLMs
Whole bacterial cells	Does not require additional cofactors	Limited substrate permeability	Limited correctly folded protein/PTMs
Hepatocyte cell culture	More available than HLMs	Complex culture setup and conditions required	Limited culture viability, need to be isolated regularly
Bacterial expressed protein	Individual cyps can be expressed, minimal culture requirements	Limited PTMs, requires cofactor and supplementation	Readily available simple purification

2.2 CYP protein microarrays

Previous work on creation of CYP functional protein microarrays carried out by the Blackburn group has been promising, with findings generating an initial experimental prototype using *in vitro* and *in silico* modelling of biotinylated CYP3A4 with the surface. This work had shown, for the first time, the experimental evidence of catalysis by the immobilised CYP3A4 with a single fluorogenic substrate. This work described the analysis of CYP3A4 using the peroxide shunt pathway with cumene hydroperoxide and a reconstituted system with CPR. It is data from this published work that lead to

further expansion of this format into CYP2C9 and CYP2D6 which forms the rationale of this thesis. Active and biosynthetically modified CYP3A4, CYP2C9, CYP2D6 and CPR were therefore required in sufficient quantities to facilitate expansion of the functional microarray format and this lead to the investigation of bacterial cloning, expression and characterisation.

2.3 Results: Expression of CYP2D6, CYP2C9 and CYP3A4 as soluble proteins suitable for protein microarray fabrication

2.3.1 Expression systems

The first step in extending the cytochrome P450 protein array system developed in the Blackburn lab was to obtain sufficient amounts of active protein, modified for attachment to the microarray surface and tagged for increased solubility and for efficient purification. Following this, the expression of the protein in a suitable host system warranted optimisation for improved yields and activity. For confirmation and characterisation protein purifications are performed to attain sufficiently pure protein.

Various expression strategies have been reported in the literature using various host cells including insect, yeast, mammalian, plant, bacterial cells and even cell free expression systems. Some have more advantages than others and, given the requirements of this study, the bacterial expression system was adopted. This is due to the simple cloning and expression and ease of synthetic protein modifications required for immobilisation. In addition, several species of bacteria such as E-coli do not express endogenous CYPs, which make CYP characterisation in whole cells or simply lysed cells uncomplicated as there is no background CYPs (Urlacher & Girhard, 2012).

The cloning of the CYPs and accessory proteins in this thesis has been based on the strategy developed by the Blackburn lab (Beeton-Kempen, 2010). The strategy was modelled on CYP3A4, which is believed to be most complex of the CYPs to reconstitute catalytic activity.

2.3.2 Choice of bacterial strains

The strategy employed used a custom bacterial expression system for cloning and expression of CYP enzymes in various E-coli strains. Bacterial hosts in general offer the advantage of short times required for growth and expression, as well as high yields of functional protein production. There is also the advantage of minimal culture requirements as bacteria can flourish in relatively simple media and can be easily distinguished by their antibiotic susceptibility. These cells can therefore be cultured in varying quantities such as small flask cultures and easily expanded to larger bioreactor cultures. A potential disadvantage of bacterial expression systems though is that many mammalian proteins do not fold correctly when expressed in bacterial systems.

DH5α strain is modified to lack endonuclease activity and encodes mutations in the recombination machinery to allow for higher transformation rates and stability of the insert DNA. This strain has minimal growth requirements with fast growth rates and is routinely used for high yielding protein production (Table 2.2).

Rosetta strains are adapted for the expression of rare codons not usually utilised in bacteria; this is thought to aid the expression of mammalian proteins in bacterial systems. This strain also carries components to aid in disulphide bond formation for more stable protein folding (Table 2.2).

The Arctic Express strain lacks several proteases and has components that allow for expression of protein re-folding chaperones at lower temperatures (Table 2.2). This allows for decreased protein degradation and increased protein folding for expression of highly functional protein.

Table 2.2: Comparison of E-coli strains and their characteristics for protein expression

Strain	Induction/Maintenance	Advantages	Disadvantages
DH5α	None, dependent on transformed plasmids	Simple transformation and easily manipulated	Active proteases, no internal antibiotic selection
Rosetta	Kanamycin, Tetracycline for selection of bacteria containing plasmids for rare codons	Plasmids do not have to be rare codon optimised	Additional preselection steps required for plasmid maintenance
Arctic Express	Gentamycin for selection of bacteria with plasmids for chaperone proteins	Lacks proteases, expresses re-folding chaperones, growth in low temperatures reduces protease activity	Additional preselection steps required, have to alter growth temperature during expression

These three strains were therefore selected for their potential advantages in expression of functional CYPs. All these strains can express protein from genes under the control of a T5 promoter. T5 bacteriophage promoters are known to compete for bacterial RNA polymerase and therefore can yield protein expression in any E-coli system (Gentzt & Bujard, 1985). In the expression vectors used here, the T5 promoter is itself controlled by a lac operon, thus making the expression of various CYP enzymes inducible with IPTG. The expression vector also has a pUC based origin of replication for maintenance of the vector at high copy number within the bacterial cell, as well as a transcription termination region. Each CYP gene was cloned from cDNA as N-terminal, in-frame fusions to the biotin-carboxyl carrier protein (BCCP) domain from the acetyl-CoA carboxylase (accB) gene (E-coli) essentially as described in (Beeton-Kempen, 2010). During cloning of each CYP cDNA, the DNA encoding the N-terminal membrane anchor and the stop codon were removed. The cloning was done systematically and each step quality controlled by restriction enzyme analysis. The final clones were sequenced verified to characterize the inserts.

2.3.3 Protein reconstitution strategy

Membrane binding is known to be required to retain CYP activity (Baylon, Lenov, Sligar, & Tajkhorshid, 2013). To facilitate high throughout workflows with highly active functional protein, the immobilisation strategy needs to be specific and non-denaturing. Several immobilisation strategies have been used to maintain protein activity in functional protein microarrays, with the surface chemistry determining the protein modification requirements.

For random covalent immobilisation, several surface chemistry architectures have been used including epoxy and aldehyde functionalised surfaces (Table 2.3). Random covalent binding results in efficient binding but in a non-uniform manner, this occurs at free amine or thiol groups. CYPs have an essential thiol group provided by the highly conserved cysteine residue that coordinates the heme ligand. Any covalent modifications of this thiol will destroy heme binding and result in inactive apoprotein formation.

Other types of random immobilisation surface chemistries results from adsorption or absorption forces. Binding in random orientations can result in high density protein binding but can also lead to protein misfolding by electrostatic interactions with the underlying surface. The random binding also does not favour the specific interaction with CYPs and their redox partners which can result in poor or low activity CYP-CPR complexes. This method of immobilisation might be more useful for use of CYP mono-systems where the peroxide shunt pathway is the mechanism of driving catalysis.

Affinity immobilisation is a favourable strategy for uniform oriented protein binding. Uniform oriented protein binding can in principle favour the CYP-CPR interaction that is crucial for reconstitution of catalytic activity; indeed, it is interesting to note that in the native CYPs and CPR, the N-terminal hydrophobic peptide serves to provide uniform, oriented binding to a membrane surface as a prelude to catalysis.

By way of example, the nickel-nitrilotriacetic acid (Ni-NTA) system provides weak interactions that result when nickel NTA coordinates with a histidine rich region of a protein, most commonly the 6

consecutive histidine (6x His) residues cloned as the His-tag into many recombinant proteins. Uniform protein binding can result which can favour the interaction and formation of transient CYP CPR interactions. This type of immobilisation is susceptible though to interference from imidazole containing compounds and can displace the surface protein binding. Furthermore, the relatively weak nature of the Ni-NTA-His tag interaction makes it unsuitable for protein microarray formation since arrayed proteins would rapidly dissociate from their original location.

Surface fuctionalisation with other affinity binding agents such as avidin-biotin also provide for specific interactions that result in uniform protein binding (Table 2.3). In particular, the binding of biotin to avidin or modified avidin is highly specific and is the strongest non-covalent bond known at a K_d of 10^{-15} M (Dorgan et al., 1999). This specific interaction is essentially irreversible but can be broken under extreme denaturing conditions (Holmberg et al., 2005).

For these reasons, previous protein microarray work in the Blackburn group has settled on the use of biotin-streptavidin systems as the basis for irreversible, high specificity, oriented, uniform immobilisation of the arrayed proteins; this system has been validated in CYP functional microarray in previous studies (Blackburn et al., 2012). However, this means of interaction is too strong for protein purification since it is essentially irreversible. Therefore, for protein purification purposes, the weaker interaction of nickel NTA and 6x histidine tag system was included in addition to simplify protein purifications.

Therefore, a dual method of protein modification was selected for use in this work. The protein was modified to include the 6 x histidine tag for ease of purification. In addition, to facilitate solubility, the protein membrane anchor peptide was truncated in the expression plasmid. For immobilisation the protein had to be biotinylated. This can be done in many ways including biotinylation steps post purification. However a particularly effective strategy is to clone the target protein in frame with the BCCP sequence which carries a biotinylation motif (Cull & Schatz, 2000). This sequence is recognised by bacterial biotin ligase systems for biotin incorporation. This strategy allows for specific biotinylation of BCCP-tagged recombinant proteins at a constant position in the protein sequence.

Table 2.3: Immobilisation Surface chemistry and modifications required at the protein level (Bertone & Snyder, 2005b)

Surface chemistry	Protein attachment	Protein orientation	Modifications required
Epoxy	Covalent crosslinking	Random	None
Aldehyde	Covalent crosslinking	Random	None
Poly(L-lysine)	Adsorption	Random	None
Nitrocellulose	Adsorption, absorption	Random	None
Poly(vinylidene difluoride)	Adsorption, absorption	Random	None
Avidin	Affinity binding	Random/Uniform	Biotinylation
Nickel-nitrilotriacetic acid	Affinity binding	Uniform	6x His fusion

2.3.4 Protein modifications cyps and accessory proteins

Cloning proceeded with inserting the coding region of cyps or accessory protein coding regions in frame with the transcription start site at the desired restriction site (Figure 2.1). cDNA was amplified from a cDNA library and the inserts were generated. The N-terminal membrane anchor binds the CYPs and redox partners (CPR and CytB5) to the membrane of the endoplasmic reticulum by membrane anchor signal peptides (Monier, Luc, Kreibich, Sabatini, & Kaplan, 1988). However, when purifying enzymes expressed with such N-terminal anchors present, the mode of isolation has to be optimised for isolation of membrane bound fractions. Such techniques include harsh cell lysis from inclusion bodies and differential centrifugation with membrane solubilisation. Soluble protein expression has the advantage of ease of purification, as all that is required is effective cell lysis, fractionation and affinity purification. To facilitate soluble expression of CYPs and CPR in this work, the membrane binding peptide sequence was therefore deleted during cloning (see supplementary).

Ligation reactions were then carried out using either blunt or sticky end ligations for generation of a circular plasmid. Ligation reactions were allowed to proceed and thereafter transformed into competent *E. coli* cells. The cells were grown up on solid media supplemented with antibiotic (ampicillin) and incubated overnight. Individual colonies were isolated and colony PCR was performed on several colonies to verify insert positioning in the plasmid. Primers were designed with the forward primer complementary to a sequence in the plasmid and the reverse primer sequence complementary to a sequence in the insert. Correct insert positioning was determined by a PCR product size equivalent to the size predicted by the cloning software. Desired colonies were grown overnight in liquid media supplemented with antibiotic. Glycerol stocks were prepared from this overnight culture and the remaining culture was pelleted by centrifugation and prepared for plasmid extraction. Plasmids were extracted using a plasmid extraction and purification kit. This employs lysis, alkaline denaturation and column purification to isolate the plasmids. Plasmid DNA concentration and yields were determined by spectrophotometric means using the Nanodrop spectrophotometer, with absorbance at 260nm indicative of DNA concentration and absorbance at 280nm indicative of protein contamination (absorbance at 280nm is indicative of aromatic amino acids). The plasmids were sequenced for confirmation. Sequences were analysed against predicted sequences as determined by plasmid design.

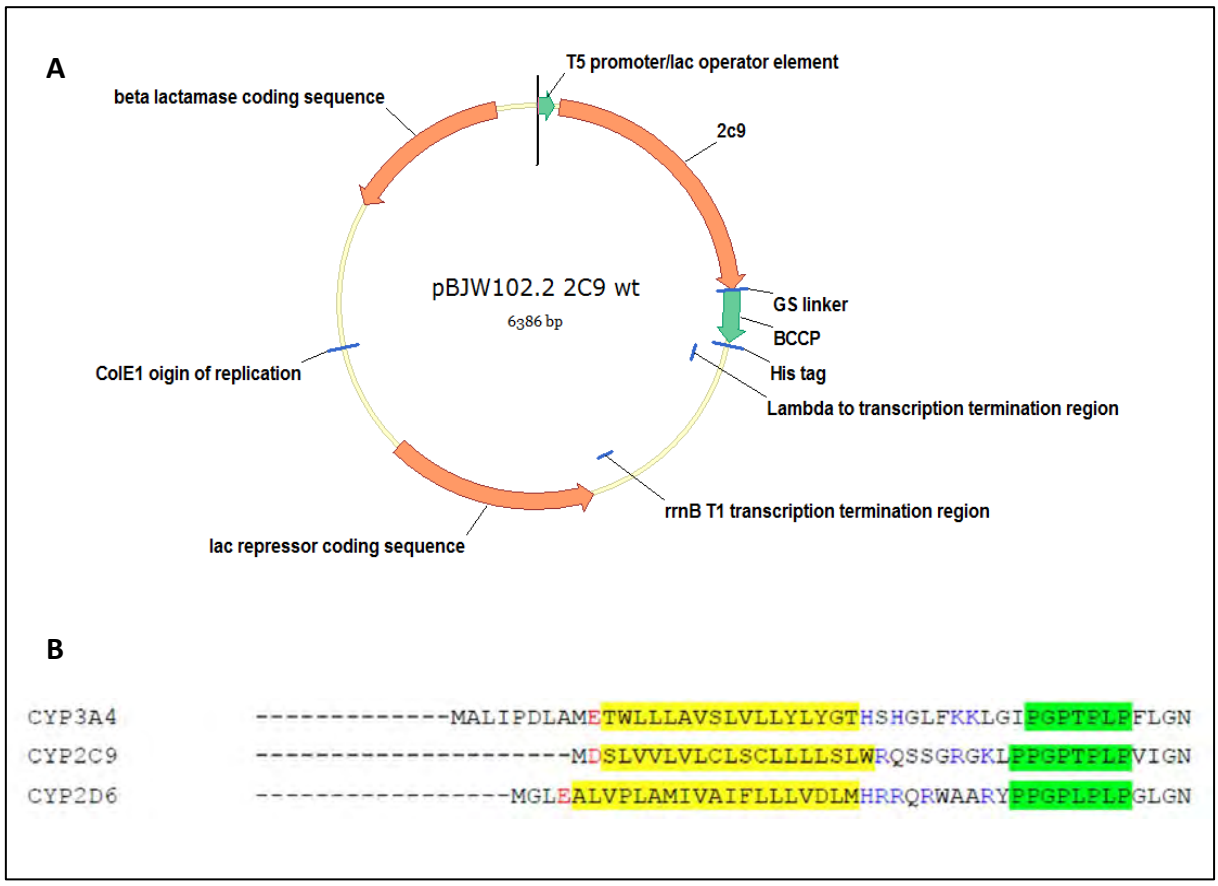


Figure 2.1: Representation of a plasmid encoding CYP2C9 and design of expression cassette with multiple sequence alignment of CYPs showing N-terminal membrane anchor region

Panel A represents plasmid may showing expression vector used in cloning *E. coli*, CYP3A4 and CYP2D6 plasmid maps are highly similar with the cDNA for the each in the location represented by CYP2C9 above (CYP3A4 and CYP2D6 plasmid maps in supplementary information).

Panel B represents multiple sequence alignments showing acidic amino acid residue (Red) usually precedes the hydrophobic membrane anchor sequence region (yellow) and a proline rich region (green) typically follows the membrane anchor sequence region, alignment adapted from (Neve & Ingelman-sundberg, 2008).

2.3.5 Recombinant expression strategy, culture conditions and supplements

Bacterial cells were transformed to ampicillin resistance, and were selected using ampicillin on agar plates. Colonies that grew were sequence verified before proceeding with protein expression. Glycerol stocks were created for storage of verified colonies. Cells were maintained in liquid culture with appropriate antibiotic and, for expression, cultures were inoculated into enriched liquid media (Terrific Broth) supplemented with thiamine, trace elements and glycerol. These conditions are optimised to account for the rare vitamin thiamine in bacteria, the requirement for metal ions, particularly iron for heme incorporation, and glycerol as an aid in protein folding as well as to prevent protein aggregation (Vagenende, Yap, & Trout, 2009). Cells were grown to mid log phase at 37°C and the media was supplemented with 5'-Aminolevulinic acid (ALA) prior to protein induction. ALA is a precursor in the heme biosynthesis pathway and results in abundant heme synthesis for a greater likelihood of proper heme protein formation (Lu et al., 2013). Free biotin was also supplemented prior to induction; the free biotin provides sufficient biotin for the *in vivo* biotinylation by *E. coli* biotin ligase. Expression was induced by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG), a non-metabolisable lactose analogue, at the required concentrations, and incubated at a lowered temperature of 30 °C for 16-18 hours; this time has been shown to be the optimal time for maximal CYP expression. Following this incubation, the cells were harvested by centrifugation and washed in phosphate buffered saline (PBS) to remove potentially interfering compounds and free supplements. Cells were thereafter resuspended in lysis buffer and subjected to a round of freeze thaw cycling from 4°C to -80°C and back to 4°C to induce cell lysis by rapid ice crystal formation.

Lysis continued with the cell suspension being incubated with a glycoside hydrolase (lysozyme) for enzymatic cleavage of the peptidoglycan layer in bacterial membranes and a nuclease (DNase I) for breakdown of long chain DNA polymers that otherwise increase viscosity by random hybridisation under lysis conditions. Following DNA and peptidoglycan digestion, a detergent mix was added to the suspension, to solubilise the membrane protein for maximal yields. The suspension was then

centrifuged to remove the cellular debris and aliquoted into appropriate storage tubes for subsequent analysis.

2.3.6 Purification strategy

The lysis strategy results in a solution consisting of soluble cellular proteins, proteins from liposomes or inclusion bodies and other biochemical components being solubilised. The purification/enrichment strategy made use of the 6x His tag cloned into the recombinant CYPs and CPR. The procedure made use of nickel affinity purification. Lysates were batch purified using a column setup for consistency and reproducibility. Lysate solutions were bound to slurry of equilibrated nickel NTA resin. Unbound materials were washed away using a buffered solution containing trace amounts of imidazole. Following sufficient washes, the recombinant CYPs were eluted off the column using higher concentrations of imidazole to out compete His-tag binding to nickel. Eluted fractions were collected and pooled and were subsequently concentrated and buffer exchanged (to remove the imidazole and other lysis components) using a molecular weight cut-off filter and several buffer washes. The final buffer that the protein was stored in was potassium phosphate buffer containing 10% v/v of glycerol to aid protein stabilisation. The protein was analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) for confirmation of protein expression, purity and presence of required modifications such as biotinylation; the results are shown in figure 2.2.

2.3.7 Confirmation of protein modification

As a quality control and assurance step, protein characterisation was carried out using several techniques. A western blot strategy was employed to confirm that the CYPs were indeed expressed and had the required synthetic modifications, namely the 6x His tag and biotinylated. Lysates and purified fractions were separated by SDS PAGE and transferred using electro-transfer onto nitrocellulose membranes. Following this transfer the membranes were blocked using bovine serum albumin (BSA) to occupy unbound regions of the membrane prior to antibody detection since proteinaceous antibodies will bind to the nitrocellulose membranes. Anti-poly histidine horse radish peroxidase (HRP) antibody was used to detect the presence of a 6x His tag, and streptavidin HRP was

used to determine the presence of biotinylation. HRP conjugated antibodies can be detected by using the luminol peroxidase chemiluminescence detection reagent. A bioluminescence is created in the presence of HRP, peroxide and luminol. This can be captured by using a film in a darkened room or using a bioluminescence signal photo detector or a CCD camera.

A similar strategy can be employed by using Nexterion H slides functionalised with streptavidin and binding CYP, CPR or a combination and blocking, probing and detecting for the presence of Anti-His HRP antibody. This is termed the dot blot and proved useful as it directly determines the two most important protein characteristics for the system, the biotinylation and the 6x His tag; the results of this are shown in figure 2.2.

2.3.8 Results: Expression of recombinant CYPs and CPR

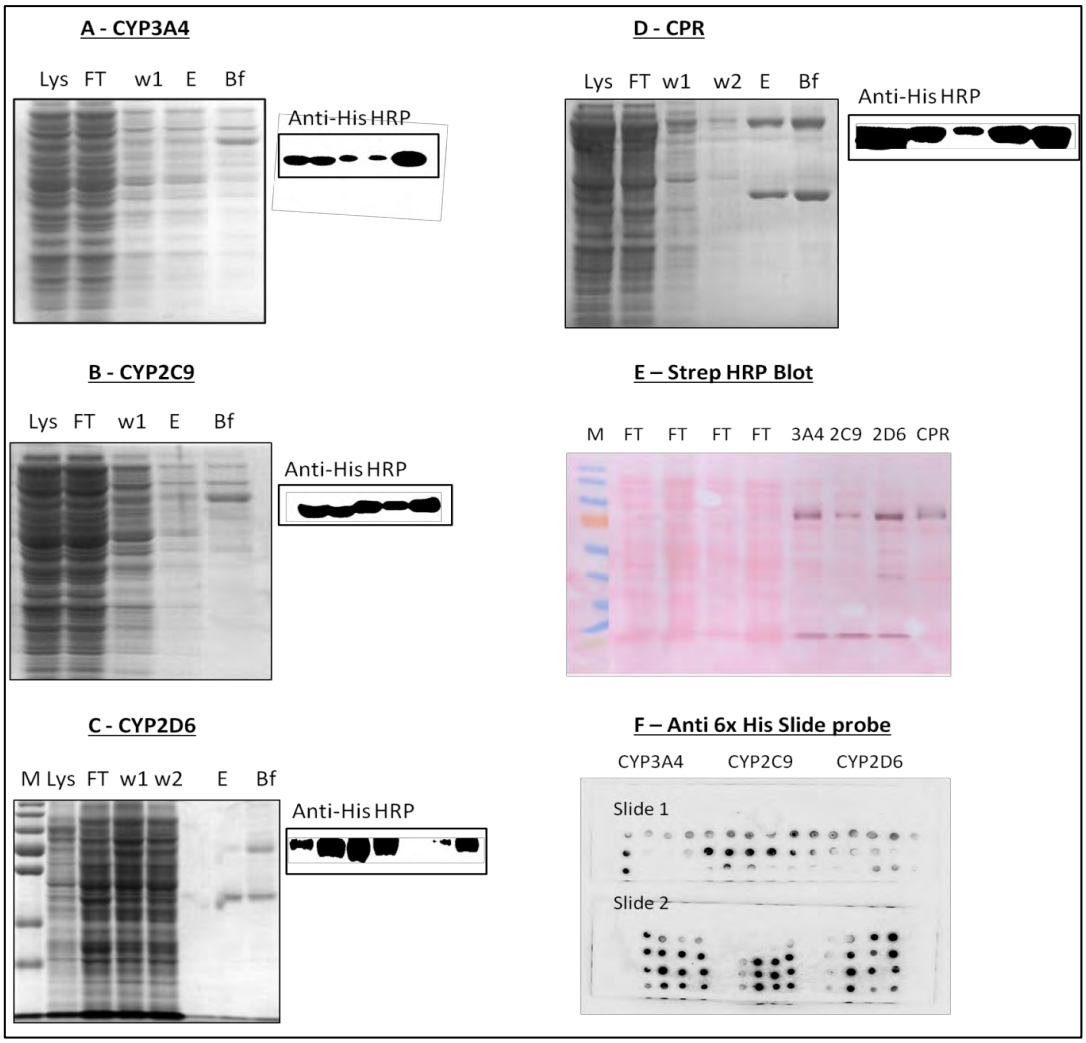


Figure 2.2: SDS PAGE and western blots performed on fractions from the purification of the expressed CYPs and CPR

M – Marker	Bf – Fraction post Buffer exchange	w1 – Wash 1
Lys – Lysate	w2 – Wash 2	
E – Elution fraction	FT – Flow through	

Panel A, B, C and D shows coomassie stained SDS PAGE gels of the purified fractions of CYP3A4, CYP2C9, CYP2D6 and CPR respectively, adjoining each is a western blot excerpt showing the result of probing with an Anti 6x His antibody conjugated to HRP. Position of the adjoining blot depicts relative position with respect to the size on the SDS PAGE gels.

Panel E shows a non-specific protein stained (Ponseau S) SDS membrane transferred from a gel of purified CYPs and CPR. This image is overlaid with the image of the same membrane probed with Streptavidin conjugated with HRP.

Panel F shows the results of Anti 6x His HRP probing of Nexterion slides derivatised with streptavidin and bound to CYPs and CPR. This image depicts the result of one step purification and immobilisation strategy in a typical hand spotting microarray type workflow.

2.4 Results: Protein characterization and reconstitution

2.4.1 CYP holoprotein quantification using CO difference spectrum

An important characteristic of CYP protein is the strong absorbance peak at 450nm in the absorbance spectrum that arises upon reducing the heme in presence of carbon monoxide (CO) (Hollenberg & Hager, 1973). The relevant extinction coeffecients have been determined and are valid for most CYPs (Guengerich, Martin, Sohl, & Cheng, 2009). The act of reducing the heme to the ferrous state allows for the binding of CO, since only the ferrous form of hemoproteins binds CO due to the stereo-electronic properties of the electrons of iron. Therefore, in this characterisation methodology the heme is artificially reduced by the reducing agent sodium dithionite ($Na_2S_2O_4$) in a CO saturated solution containing CYP protein. In the ferrous versus ferrous difference spectrum of reduced and CO bound CYP there are two predominant peaks namely at 420nm and 450nm. The 450nm peak is known to contain active CYP protein while the 420nm peak is thought to represent an inactive form of CYP. There is evidence that suggests the P420 peak is a result of ligand switching of the heme co-ordination state from the conserved cysteine to a nearby histidine that causes structural changes in the protein folding state as well as altering the redox potential of the heme-iron, therefore resulting in inactive protein (Sun et al., 2013). This methodology is therefore useful for quantifying the amount of active protein since it provides a relatively robust means to distinguish between the closely related P420 and P450 forms that would otherwise be indistinguishable.

In addition, the ratios of P450 to P420 are important as there are currently no efficient ways to physically separate these two forms of the protein prior to immobilisation for microarray fabrication. Therefore, the aim here was to express and purify protein that resulted in the highest ratio of active P450 to inactive P420 protein, thus ensuring that the majority of protein that became immobilised was in the active form for catalysis. The respective CYP proteins were thus expressed in several bacterial strains and protein amounts quantified using carbon monoxide difference spectra. The results of these experiments are shown in figure 2.3.

2.4.2 CYP catalytic characterization – cumene hydroperoxide shunt pathway

For the accurate characterisation of CYPs, specific substrates have to be selected. CYPs are specific in the substrates they recognise and therefore several strategies are available for the detection and quantification of CYP functionality (Table 2.4). A technique of functional analysis was selected based on robustness in application and ability to translate to a functional protein microarray format.

The fluorescent functional assays are the most compliant in this regard and can be integrated in current microarray fabrication and detection workflows. This is essential as the current technology aims to be compatible with routine DNA expression microarray facilities, which have the advantage of high sensitivity and availability. Furthermore, there are fewer steps required in the workflow per assay, which is advantageous for ease of operation and fewer resources required. The basis of the assay reaction is that the fluorogenic compound is derivatised to resemble a drug scaffold that is recognised and metabolised by a particular CYP (Trubetskoy, Gibson, & Marks, 2005). The substrates are modified to diminish the underlying fluorophores quantum yield, making the modified fluorophore less fluorescent that the parent. Upon CYP metabolism the drug scaffold is oxidised and the parent fluorophore is released (Di, Kerns, Li, & Carter, 2007). The increase of relative fluorescence signal over a period of time is monitored, and compared to the negative controls for validation. Furthermore, to be accurate in absolute quantitation, the fluorescence signal can be converted to a number of fluorescent molecules produced per unit time by comparison to a standard curve constructed from parent fluorophores that are commercially available. There are several

fluorophores that can be functionalised with chemical substituents that resemble drug scaffolds and minimise fluorescence, and there are several available substrates utilising various colours in the fluorescence spectrum, including several for each CYP isoform. This can be advantageous as having multiple fluorescent substrates available can be useful for cross-validation purposes.

Luminescent functional assays are limited in that the only parent compound that can be modified with drug like scaffolds is luciferin. These types of assays also have the disadvantage of additional steps that are required prior to detection, which adds complexity and cost.

While DNA- or RNA-based genotyping and gene expression studies can give useful data for pharmacokinetic and pharmacogenomics studies, these do not provide functional information directly (Table 2.4). Rather, such studies correlate previously established functional data from CYP functional studies with clinical cohorts. This type of functional assay is therefore not the focus of this thesis.

By contrast, metabolite detection by label-free technologies such as mass spectrometry is an extremely useful technique to determine CYP functionality. Label-free detection does not require any particular modified substrates and drug compounds can be assayed directly and in combinations. Furthermore, label-free detection can be used to identify multiple different metabolites that may be produced by a CYP (Table 2.4). This is a limitation of label-dependent substrate assays that can limit the information gained from fluorescent of luminescent experiments.

CYP metabolism of drugs typically causes them to have altered hydrophobicity, which can be quantified by HPLC based assays. HLPC alone relies on comparison of each metabolite with a standard curve and multiple metabolites cannot readily be directly identified using a single mode of separation. Therefore, techniques that combine HPLC with a second mode of separation such as ionisation in mass spectrometry are rich in information for both metabolite identification and quantitation. A limitation of this detection strategy is that it is not readily conductive to high throughput workflows due to the intrinsic cost of the mass spectrometry instruments.

Table 2.4: CYP assay technology routinely used in CYP drug metabolism characterisation

Type	Substrates	Mechanism/ Uses
Fluorogenic	Available as low/minimally fluorescent Vivid Substrates Resorufin, Coumarin, Fluorescein derivatives	Catalysis close to or at protecting group Becomes highly fluorescent upon substrate metabolism
Luminescent	Luciferin derivative substrates, available as P450 glo	Becomes metabolised to luciferin, requires addition of luciferase and cofactors
DNA/mRNA	For induction and SNP analysis, CYP chips	Taqman SNP arrays Illumina Veracode DMETchip
Protein	AbScieX human induction kit, ELISA several antibodies available	Peptides detected and quantified by Mass spec using peptide standards
Label Free	Probe substrates and inhibitors recommended by the FDA for each cyp	Requires metabolite standards for calibration and quantification, can become expensive many drug metabolites are unknown and may have to be chemically synthesised

The fluorescent substrate functional assay strategy was therefore selected to characterise the expressed protein, prior to further microarray and functional assays. The specific substrates chosen were both specific to one of the CYPs and were comparable with routine filters available in genomic microarray detectors. The results obtained on assay of each purified CYP isoform are shown in figure 2.5.

2.4.3 CPR characterization using MTT reduction assay

CPR reduction activity can be quantified using simple colorimetric assays routinely used to determine reduction activity in cell survival assays such as the MTT (3-(4, 5-dimethylthiazol 2-yl) -2, 5 – diphenyltetrazolium bromide) reduction assay which is similar to the classical cytochrome c reduction assay. The principal of the MTT reduction assay reduces MTT from a tetrazole to a formazan which

precipitates out of solution into a darkly coloured admixture. This colour change can be detected and quantified using absorbance spectral measurements using UV/Vis spectroscopy, of which the coloured products have been characterised and extinction coeffecients determined previously. The MTT assay has been validated for quantification of CPR activity and is robust in comparison to other types of CPR functional assays (Yim, Yun, Ahn, Jung, & Pan, 2005). The simplicity and robustness of this assay made it the ideal characterisation assay to validate CPR function *in vitro* compared to the cytochrome c reduction assay.

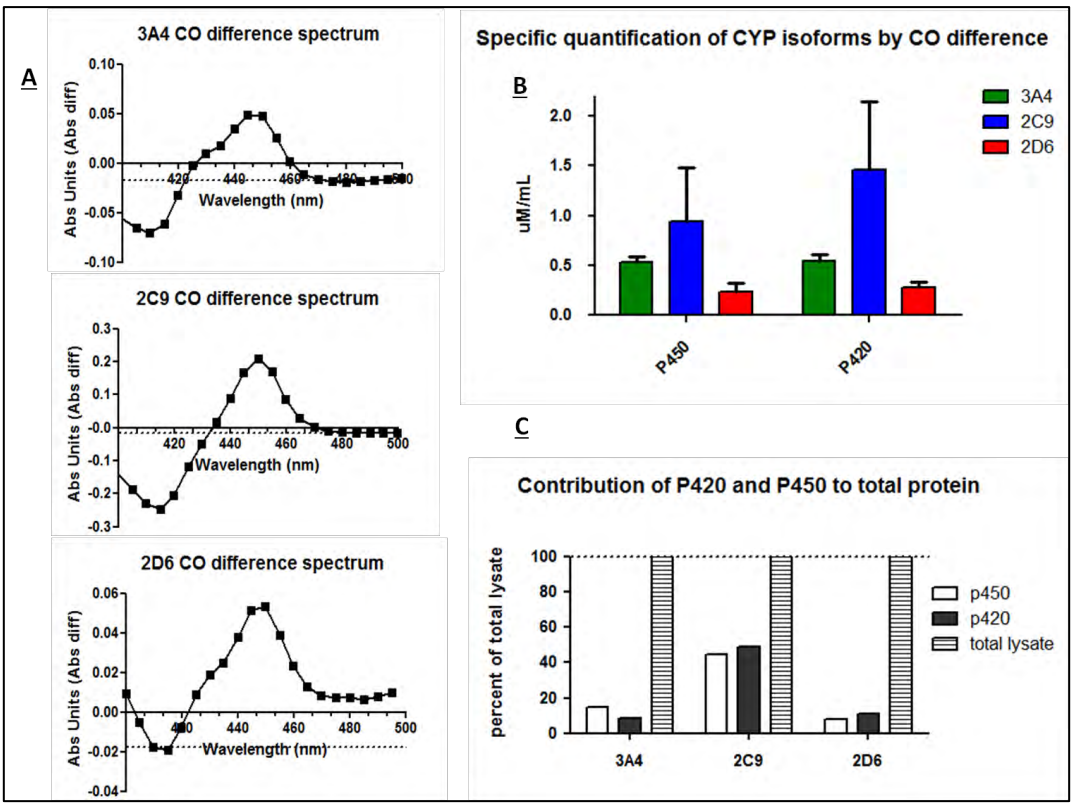


Figure 2.3: Quantification by ferrous versus ferrous difference spectrum and comparison of reproducibility between isoforms

Panel A shows a representative CO difference spectrum of each of CYP3A4, CYP2C9 and CYP2D6, dashed line indicates the average baseline. Baseline is reduced CYPs by sodium dithionite in the presence of air.

Panel B shows quantification of amount of protein absorbing at 450nm and 420nm respectively. Data represents mean and standard deviation of three independent experiments from different flask cultures grown independently.

Panel C shows percent representation of P450 and P420 to total protein concentration. Total protein concentration estimated using the Bradford protein determination assay, P450 and P420 concentrations determined using ferrous CO versus ferrous difference spectrum and concentrations extrapolated using published extinction coeffecients.

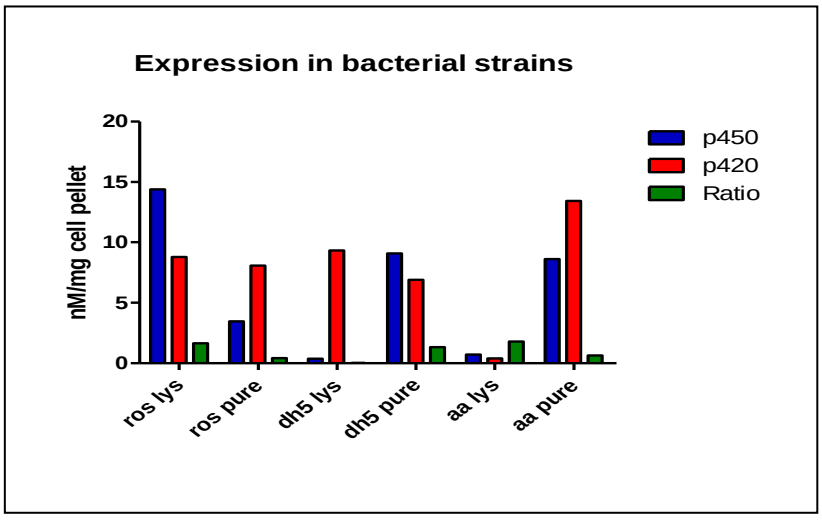


Figure 2.4: CYP3A4 quantification from expression between different bacterial strains

ROS –Rosetta strain	lys – crude lysate
DH5 – DH5α strain	pure – purified fraction
Aa – Arctic express strain	

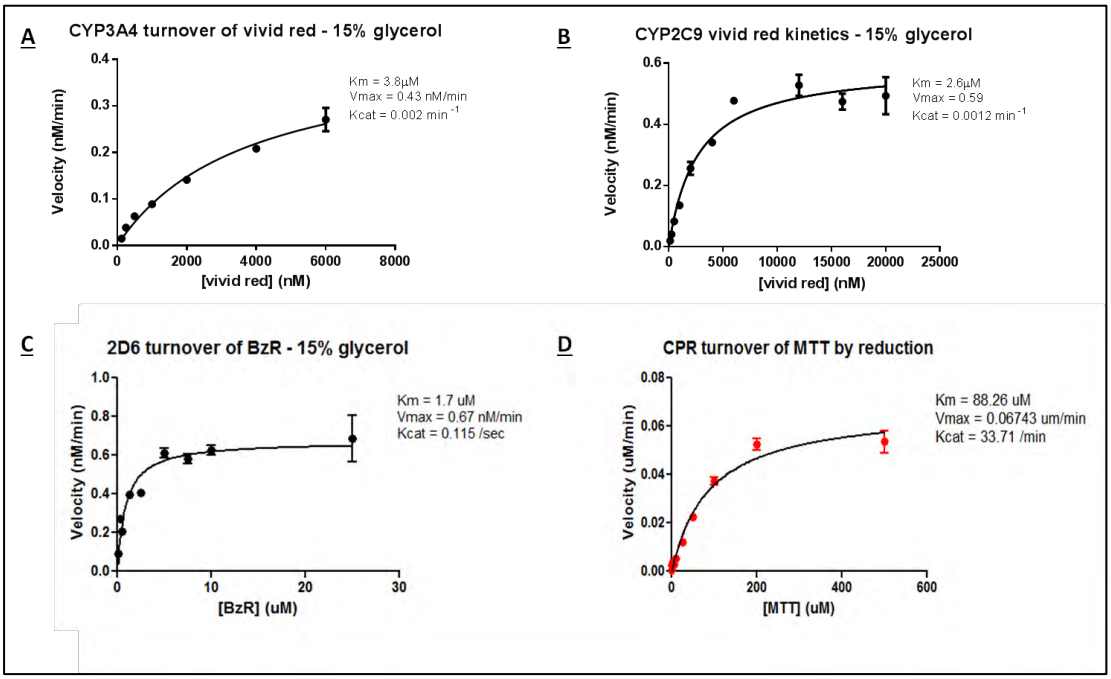
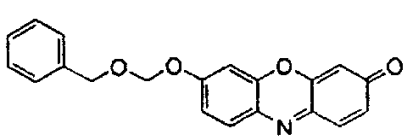
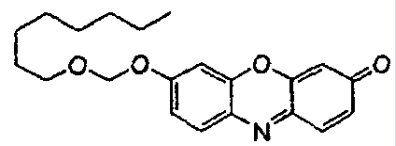
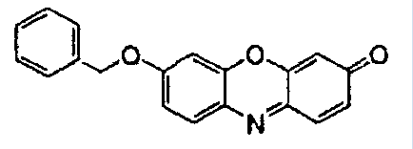


Figure 2.5: Enzyme kinetic profiles for the characterisation of expressed protein functionality

CYPs were functionally characterised using fluorescent vivid substrates specific for each isoform using the cumene hydroperoxide shunt pathway. CPR was kinetically characterised using MTT reduction and NADPH. All CYP assays were performed in K_2PO_4 buffer containing 15% glycerol and data represents the mean and standard deviation of three replicate reactions. Enzyme reactions were allowed to proceed for up to 45 minutes and data was acquired at regular intervals. Data that fell within the linear region were then fit to a Michaelis-Menton kinetic plot and kinetic parameters were estimated from best fit models. Data correlates to published data and share similar kinetic parameter ranges (Table 2.5).

Table 2.5: Comparison of published data on CYP expressed protein with the bacterial expressed protein using cumene hydroperoxide shunt pathway

CYP	Substrate	V_{max} (nm/min)	K_m (µM)	K_{cat} (per min)
3A4	BOMR 	0.55	1.4	0.0037
		0.43	*3.8*	*0.0020*
2C9	OOMR 	0.31	2.6	0.0026
		0.59	*2.6*	*0.0012*
2D6	BzR 	0.04	3.1	0.026
		0.67	*1.7*	*0.115*

Data in second column (right indent) in each field represents the data from this thesis and first column represents published figures obtained from (Makings & Zlokarnik, 1999).

2.5 Conclusion and discussion

Cloning, expression and purification of three different CYP isoforms and the redox partner CPR was carried out in this chapter and protein was obtained in sufficient yields in each case for further studies. Furthermore, the protein required biosynthetic modifications namely biotinylation as well as a tag to facilitate simple purification prior to microarray fabrication and further functional assays.

A bacterial expression system was selected to be the most suitable for expression of the four recombinant proteins as it provided all necessary requirements for efficient protein expression with

the desired synthetic modifications (Figure 2.1 and full sequences in appendix). In addition, this host system provided for simple cloning and required minimal culture and media for dense growth.

Following the cloning and expression, the presence of the expressed protein was confirmed by using SDS PAGE and western blot using both anti-His antibody and streptavidin HRP conjugates. The presence of bands corresponding to the expected molecular weight of the respective CYPs was confirmed in purified/enriched protein fractions on SDS PAGE gels, as well as on western blots (Figure 2.2). For quantification and quality control the carbon monoxide difference spectrum was determined for all expressed protein (Figure 2.3). This is an essential step as it determines the relative amount of active protein namely the P450 fraction, which is crucial for further functional assays as this is the concentration required for kinetic parameter determination. It is also interesting to note the variability in P450/P420 ratios expressed amongst the bacterial strains investigated (Figure 2.4). The different strains express the CYPs with varying ratios of P450/P420, suggesting that the expression machinery is responsible for this variation as the culture conditions were essentially identical. Furthermore, the P450/P420 ratio also varies with protein pre- and post-purification, which is seemingly paradoxical as more active CYPs would be expected in the purified fractions. Recently, a mechanism has been suggested for heme ligand switching between the conserved cysteine and a nearby histidine residue in the active site of CYPs, this could account for the seemingly paradoxical phenomena of lower P450/P420 ratio pre- and post-purification which might be a result of heme ligand switching under the slightly denaturing conditions incurred during the purification (Sun et al., 2013).

Functional validation was undertaken by utilising the CYPs peroxide shunt pathway to uncouple the catalytic cycle from CPR. Fluorescent substrates were used to confirm the CYPs were catalytically active and this activity was quantified and data processed. The validation of CPR functionality was carried out using the uncoupled MTT reduction assay. The data of both types of assays was processed by fitting the data to Michaelis-Menton kinetic plots (Figure 2.5). The best fit values were used to estimate the kinetic parameters V_{max}, K_m and K_{cat}. These values were compared to previously published data which confirmed that the four purified recombinant proteins were suitable for further

microarray fabrication and functional assays (Table 2.5). Simple Michaelis-Menton kinetic models were used for the curve fitting the experimental data in contrast to their limitations for use in CYP kinetics as mentioned in earlier discussions; however for more accurate comparison to previously published data the simple kinetic models was used. It is also interesting that the K_{cat} values for CYP2D6 obtained here is fivefold higher than of that previously published; essentially the protein concentrations were similar so the rate of reactions are likely to be the source of differentiation. Furthermore, the ratio of P450/P420 used in these comparisons could differ, resulting in altered (higher) P420 being present in these assays; with P420 having potential substrate metabolising activities which could result in the higher turnover rates obtained in this experiment, however such a mechanism is currently unconfirmed.

Thus far, suitable yields of active CYP3A4, CYP2C9, CYP2D6 and CPR have been obtained. The purified proteins corresponded in each case to the relative molecular weights of their respective CYPs and western blots suggest the CYPs are modified with the desired biotinylation and 6x His tags. CO ferrous difference spectra further suggest that the expressed proteins are indeed active CYPs and also facilitated quantification of active protein. Finally, the presence of active CYP protein was proven by functional experiments using fluorescent substrates and the peroxide shunt pathway, with the resulting kinetic data agreeing largely with the literature. These experiments therefore served to generate and functionally validate sufficient purified recombinant CYP2C9, CYP3A4, CYP2D6 and CPR for work described in further chapters.

2.6 Methods

2.6.1 Plasmid construction and cloning

Plasmid cloning was carried out according to previously published procedures (Blackburn et al., 2012). Briefly, both pBJW102.2 plasmid backbone and the various cDNAs were digested using *Eco*RI restriction enzymes and the digested fragments were purified. Sticky end ligations were performed using the T4 DNA ligase to create circular constructs with inserts in frame with the transcription start sites and terminal end site.

Restriction enzyme digestions were performed using Fast Digest enzymes (Fermentas) according to the manufacturers' recommendations. Briefly 100-200 ng of plasmid or insert DNA was added to 10 units of respective restriction enzyme in a buffered solution. The reaction was incubated at 37°C until complete restriction digestion had occurred. Following digestion, the DNA fragments were separated by gel electrophoresis using 1% Agarose gel (Sigma) and an applied voltage of 100mV. DNA fragments were visualised by staining with GR green (Biolabo) and exposing to UV light in a light box (G-box, Syngene). Desired fragments were cut out using a sterile blade, solubilised and purified using gel purification columns (GeneJet PCR purification kit, Fermentas). The concentration of purified DNA was estimated using a Nanodrop spectrometer (Nanodrop 1000, Thermo).

Sticky end ligation reactions were performed using T4 DNA ligase (Fermentas) according to the manufactures' instructions. The reaction mix consisted of 5 units T4 DNA ligase and 1:1 molar ratio of insert to vector (10-50 ng insert, 10-50 ng vector) in a buffered solution supplemented with 1 mM PEG100.

Following the ligation reaction, chemically competent E-coli strains were transformed to ampicillin resistance using heat shock treatment with the ligation reaction mix. Briefly, competent cells were thawed and incubated on ice for 30 minutes with 10ng of plasmid DNA. The cell mix was then heat shocked at 42°C for 90 seconds and transferred back onto ice for 2 minutes before spreading onto sterile bacterial agar plates containing antibiotic for selection (0.1 mg/ml Ampicillin, Sigma). Plates were spread with two different dilutions of transformed cells (a 1x and a 10x concentrate) and a negative control consisting of using untransformed cells was also plated. Plates were incubated at 37°C overnight. Following colony growth, single bacterial colonies were isolated and transferred into liquid culture (LB, 10% w/v Tryptone, 5% w/v Yeast extract, 2.5% w/v NaCl in dH_2O) and grown overnight at 37°C with agitation. Plasmid extractions were performed using the column purification of plasmid DNA (Fermentas plasmid miniprep kit) according to the manufacturer's instructions. Plasmid DNA was analysed by restriction enzyme digest for desired full length plasmid and insert orientation; DNA fragmentation patterns were visualised by agarose gel electrophoresis with GR green staining.

Plasmid DNA from suitable colonies was then subjected to DNA sequencing (dye-terminator sequencing, Applied Biosystems ABI 3130XL, Life Technologies). Sufficient coverage was achieved using primers covering the insert and partial vector sequence using the following sequencing primers; Forward primer 5'- CCC GAA AAG TGC CAC CTG -3' and Reverse primer 5'- GTT CTG AGG TCA TTA CTG G -3'. Local sequence alignment of the resultant data carried out in Vector NTI advanced (Life Technologies) using the predicted vector sequence and insert sequence obtained from DNA databases. The alignments were matched and any mutations compared to wild type were fixed using an inverse PCR strategy.

Mutation fixes was performed by initially designing primers with correcting overhangs in OligoAnalyzer software (Integrated DNA Technologies) that adjusted single point mutations. Primers were synthesised commercially (Inqaba biotech) with the 5' ends containing phosphorylation modifications for subsequent re-ligation. Inverse PCR reactions were performed using a high fidelity thermostable DNA polymerase optimised for long PCR lengths (KAPA Biosystems High Fidelity DNA polymerase). PCR conditions were as follows: Denature at 98°C for 20 seconds; Anneal at 5°C below average melting temperature of primers for 15 seconds; Elongate at 72°C for 30 seconds per kb; cycle for 15-20 cycles, with a final elongation of 72°C for 5mins. To eliminate parental re-ligation from interference, the parental plasmid was removed from the PCR product using *Dpn*I digestion (FastDigest, Fermentas) that digests methylated DNA. N.B Parental plasmid DNA is methylated in bacteria while PCR product is unmethylated.

Fragments were separated by gel electrophoresis, visualised by GR green staining on a UV box, isolated and purified using column purification (Fermentas gel extraction kit). Fragment concentrations were estimated by Nanodrop, and subject to blunt end ligation (T4 DNA ligase, Fermentas). Following incubation of the ligation reaction, a small quantity of chemically competent cells was transformed using the heat shock method with the ligation mix. The transformation mix was spread onto agar plates as before and incubated overnight for colony growth. Several single colonies were isolated and transferred to liquid suspension. Colony PCR was performed (Pyrostart Fast PCR mix, Fermentas) using primers flanking insert positions; Forward primer 5'-AAT AGG CGT ATC

ACG AGG C-3' and Reverse primer 5'-GGG AAC GTA CGA TGT GAC -3') followed by *Sph*I restriction digest (FastDigest, Fermentas) to identify colonies with the fixed mutation as opposed to those containing parental background. Colony PCR reactions were subject to restriction enzyme analysis and, following that, several colonies were sequence verified. Verified colonies were grown in liquid culture and mid-log-phase cultures of grown cells were created and stored at -80°C in a library for further use.

2.6.2 Protein expression

Cells were grown from glycerol stocks in liquid medium containing appropriate antibiotic overnight at 37°C. Cells were inoculated and grown at 37°C to mid log phase in Terrific Broth (24% w/v Tryptone, 12% w/v Yeast extract, K_2PO_4 buffer, dH_2O) supplemented with trace elements, thiamine and ampicillin. Following this 84 µg of 5'-Aminolevulinic acid (Sigma) per litre culture was added to culture in powder form and incubated at 30°C for 30 minutes. Protein expression was then induced with 1 mM IPTG (Melford) and the cultures were also supplemented with 50µM free biotin (Sigma) and incubated at 30°C for 18 hours with shaking. Cells were harvested by centrifugation at 6000g for 15 minutes at 4°C and washed once with 1x PBS. Cells pellets were subject to one cycle of freeze-thaw to accelerate lysis and, following this, were resuspended in Lysis buffer (20mM K_2PO_4, 20% v/v glycerol, 20mM β-Mercaptoethanol (Sigma) pH 7.4). Lysozyme (1mg/ml lysozyme powder from hen egg (Sigma) in 10mM Tris, pH 8.0) and DNase I (500U/ml DNase I powder (Roche), in 20mM Tris-HCl, 1mM $MgCl_2$ 50% glycerol, pH 7.5) was added and incubated on ice for 30minutes for lysis to proceed. Membranes were solubilised with a detergent mix containing 0.15% v/v IgePal, 0.1% v/v Triton X100 and 0.5% v/v CHAPS hydrate (Sigma). Cellular debris was removed by centrifugation at 10000g for 30 minutes at 4°C.

Lysate was stored for further use or subjected to affinity purification using Ni-His bind resin (Novagen) using fabricated columns plugged with silanised glass wool (Sigma) following manufacturer's instructions. Briefly, resin was resuspended and equilibrated prior to protein loading in equilibration buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM Imidazole, pH 8.0). Protein loading

was performed using batch purification prior to loading into the column and column washes were performed using wash buffer (50mM NaH_2PO_4, 300mM NaCl, 20mM Imidazole, pH 8.0). Protein was eluted in elution buffer (50mM NaH_2PO_4, 300mM NaCl, 250mM -1000mM Imidazole, pH 8.0). Elution fractions were pooled and desalted by buffer exchange with 100mM K_2PO_4 buffer (pH7.4) and then concentrated using molecular weight filter columns (Amicon centrifugal filters). Purified protein was aliquoted and frozen at -20°C for further use.

Fractions were collected throughout the affinity purification process and analysed by SDS PAGE using a molecular weight marker (PageRule Protein Ladder, Fermentas). SDS PAGE gels were visualised by (Acqua Stain, Vacutec) and photographed. For western blots, protein was separated using 10% SDS PAGE and transferred onto nitrocellulose membranes (Biotrace, Pall Corp) by electro-transfer (Bio-RAD Mini Protean Tetra cell, Bio-RAD power Pac HC). Protein concentration was estimated using the Bradford protein determination assay (Sigma) and Bio-RAD iMark microplate reader at 595nm prior to protein loading onto SDS PAGE. For slide blots, equal protein concentrations were spotted manually using a micropipette and incubated for 30mins in the dark without agitation, thereafter membranes and slides were treated in the same way. Membranes were stained with Ponseau S (sigma) for confirmation of protein transfer and washed. Once the stain was removed, the membranes were blocked in a blocking solution (BSA in TBS- Tween (20mM Tris, 150mM NaCl with 0.05% Tween 20)) for 1 hour at room temperature. Blocked membranes were probed with respective 6x His antibody conjugated with HRP or streptavidin HRP at a concentration of 1µg/ml for 1 hour at room temperature in the dark. Blots were washed 3 times for 5 minutes in TBS-Tween and then once in TBS. Membranes were immediately incubated with luminol solution (Western Bright ECL, Avansta) and blots were developed in the dark using photosensitive film (CL XPosure Film, Thermo). Films were developed, fixed, dried and aligned with the membranes and photographed or scanned for documentation.

2.6.3 CO assay

Carbon monoxide difference spectrum was used to quantify active P450 content from both crude lysates and purified protein samples. Briefly, protein concentration was estimated using the Bradford protein determination assay compared to a BSA standard curve using an iMark plate reader (Bio-RAD). Equal amounts of protein were added to two separate cuvettes containing 100 mM K_2PO_4 buffer (pH7.4). Baseline spectra were determined by a wavelength scan (Micro PS, Lasec) in the spectrometer (Varian Cary 50 UV/Vis spectrometer). Carbon monoxide (Speciality gases, 1000psi) was gently bubbled through one of the cuvettes for 1 minute at a constant pressure, followed by reduction of the heme protein using saturating sodium dithionite (sodium hydrosulphite, Sigma). Scanning spectra were recorded over a period of time until the peaks stop increasing at 450nm. Data extrapolated according to published extinction coefficients for quantification (Guengerich et al., 2009).

2.6.4 Functional assays

Vivid substrates (Invitrogen) or benzyl resorufin ether (Sigma) were resuspended in the required DMSO or acetonitrile solvent and concentrations confirmed by UV/Vis absorbance based on previously established extinction coefficients. Substrate ranges were determined as per the manufacturers' recommendations and reactions were performed in triplicate in white 96 well plates (Nunc). Reaction mixes were setup in master mixes to minimise pipette errors. 75ng of CYP protein was added to a buffered solution containing one of the various substrates of varying concentrations. The mix was pre-incubated and baseline fluorescence was determined by measuring fluorescence with excitation at 530nm and emission at 590nm for red fluorescence of resorufin using a Varian Cary eclipse fluorescence spectrophotometer. Following baseline measurements, the reactions were initiated by adding a 1 in 10 dilution of 1mM stock solution to a final concentration of 0.1mM of cumene hydroperoxide (Sigma) and mixing. The reactions were monitored continuously over a period of 45 minutes. Data was converted to molecules of resorufin per minute by comparison to standard curves constructed for each assay using a resorufin standard made fresh each time. Concentrations of

resorufin were confirmed using UV/Vis absorbance spectroscopy and known extinction coeffecients. Protein activity data was processed from the linear portion of activity curves. The average gradient was determined for the linear range and this was taken as the average rate of reaction for that substrate concentration. Data was modelled onto a Michaelis-Menton enzyme kinetic profile for enzyme kinetic parameter estimation using GraphPad Prism (GraphPad Software). Experimental data was compared to data published in the literature and/or provided by the substrate suppliers.

For functional CPR assays, MTT (3-(4,5-dimethylthiazol 2-yl)-2,5–diphenyltetrazolium bromide) reduction assays were performed as per published spectrophotometric methodology (Yim et al., 2005). Briefly, CPR protein concentration was estimated using Bradford protein determination assay as before. MTT powder (Fluka) was dissolved in dH_2O at the stock concentration and confirmed by UV/Vis spectroscopy, based on known extinction coefficients. Serial dilutions were created in cuvettes with buffered solution at the required range for substrate concentrations and CPR protein added to each cuvette. A baseline measurement was taken prior to addition of 50mM NADPH (Sigma) or regenerating solution (0.5 mM β-nicotinamide adenine dinulceotide phosphate, 5mM glucose-6-phosphate, 0.5U glucose-6-phosphate dehydrogenase, Sigma). Cuvettes were inverted to mix and scanned at 510nm continually for 10minutes. The linear portion of the increase in intensity was used for further data processing. Data was processed as above for functional CYP assays.

Chapter 3 – Functional microarray activity reconstitution of CYP activity

3.1 Introduction and initial considerations

CYPs have a requirement for membrane binding for activity and it is suggested that this is required for efficient redox partner interaction and for correct protein folding as well as potentially for access of substrates. One efficient method of high throughput study of protein functionality is the protein microarray and there are several surface chemistries and immobilisation strategies available for this technology. The question of whether protein immobilisation is in fact essential for CYP activity reconstitution is addressed in Chapter 4. Membrane binding of both CYP and CPR provides restriction in diffusion rates as well as controlled relative orientation and therefore enhances the chance of interaction of the CYPs and CPR in the optimal geometric configuration. Therefore, in order to mimic a membrane-bound, co-localised environment for CYP and CPR in a microarray format, surface chemistry that confers an aqueous environment and the required flexibility for transient interaction is necessary.

Microarray technology has well established workflows and automated setups in place from the genomic field. Microarray surfaces exist in various forms, including those that have chemistries that are adaptable for functional protein uses. Therefore, the technology is primed for efforts from the functional protein fields, but compatibility is crucial for optimal integration.

3.1.1 Immobilisation strategies

It is generally thought that in order for CYPs to retain their CPR dependent catalytic activity, there is a requirement for membrane binding. The reasons for this include rotations for redox partner binding, orientation of the access channels, flexibility and hydration levels for correct protein folding and heme binding (Gut, Richter, Cherry, Winterhalter, & Kawato, 1982). However, studies have shown that activity can be reconstituted in the absence of membranes by electrochemistry means as well as by utilising the peroxide shunt pathway (Hollenberg, 1992; Shumyantseva et al., 2004). Nevertheless,

several strategies have been developed to immobilise CYPs in membrane or artificial supports to retain activity of cDNA expressed CYPs and their redox partners (Table 3.1). One common strategy involves microsomal fractionation by differential centrifugation to yield the "S9" fraction, or microsomes. This complex sample preparation process involves several rounds of high speed differential centrifugation so another approach involves the reconstitution of soluble CYPs into lipid membrane preparations (liposomes) or synthetic membrane models, or even on tethered supports such as nanodiscs (Grinkova, Denisov, & Sligar, 2010). The complication arises here in that prior to lipid membrane anchorage, the CYPs have to be expressed in a heterologous host, and a common cloning strategy to enhance yields includes deletions in the membrane anchor region to facilitate expression and purification. Therefore only about 30% of expressed CYPs tend to be immobilised into such membrane preparations (Gut et al., 1982). Even with metallic nanodiscs, the non-specific binding process results in random, non-uniform distributions of CYP and CPR within those supports. There have been recent developments of other immobilisation strategies, including specific covalent binding onto conductive self-assembled monolayers - to facilitate electrochemically driven electron transfer to bound CYPs - and non-specific encapsulation into flexible hydrogel polymers, each having advantages and disadvantages (Cusack et al., 2013) (Table 3.1). There have also been advances in creation of CYP biosensors with nanostructured materials and multiwalled carbon nanotubes and nanowires as reviewed (Schneider & Clark, 2013). Such biosensors can be incorporated into multiplex formats and microfluidic channels having complex electrode architectures and electronic components. Such technologies are therefore promising to reconstruct CYP catalysis on the nano-scale but are still in the early stage of development and are highly complex in setup and construction, adding costs per assay. Additionally, the driven reaction is not in tandem with a redox partner and is artificially coerced using electrochemistry and so therefore requires electrically active substrates or metabolites for detection.

Table 3.1: Review of immobilisation chemistry for protein immobilisation

Immobilisation Method	Advantages	Disadvantages
Adsorption	**Simple, inexpensive, good for single use applications**	**Unstable binding, proteins denature on hydrophobic surfaces, requires extensive optimisation**
Entrapment: Polymer gel	**Has mass production potential**	**Protein denaturation by polymerisation free radicals**
Coupling to surface or polymer	**Large number of binding sites, may provide flexibility and low steric hindrance**	**Complex preparation, complex diffusion kinetics**
Capture system	**May be regenerated, good for when binding orientation is crucial**	**Costly and complex derivatisation processes, may have non-specific binding**

3.1.2 Immobilisation formats

Generally the mammalian CYPS are membrane bound through their N-terminal hydrophobic rich tail. The rationale for membrane binding can come from the CYP and CPR association being highly specific but transient so is facilitated by localized membrane binding. The CYP-CPR association is essential for electron transfer, provided by NADPH reduction and transferred to the CYP catalytic cavity via the CPR-CYP interaction surface (Laursen et al., 2011).

Microarrays are a means of high throughput analysis and are routinely used now in molecular biology studies of DNA and RNA in modern genomics. Microarray technology can simultaneously evaluate quantitatively thousands of genes from various systems in a relatively short time with multi parallel workflows. Such microarray applications are supported by expensive laboratory automation and computational data handling, as well as by data interpretation packages. To ensure inter- and intra-assay comparison, quality control and quality assurance factor guidelines have been developed by the

microarray community under the microarray quality control project (Patterson et al., 2006). Given all the resources and ease of high throughput, this microarray platform should be ideally suitable for certain proteomic workflows and as such have been developed in protein detection and in biomarker discovery and validation fields (Y. Lee et al., 2003). A number of groups have applied this technique to the more challenging fields of functional protein analysis, the goal being to analyse protein function in high throughput micro spots, providing the advantage of being highly multiplex and miniaturised, thus leading to lower reagent costs and lower costs per assay (Díaz-Mochón, Tourniaire, & Bradley, 2007; Stamos et al., 2012). A functional CYP protein microarray has been developed by the Blackburn lab and has shown promise in initial investigations (Blackburn et al., 2012). The investigations thus far have been applied to CYP3A4 reconstituted with CPR or using the peroxide shunt pathway. This method essentially employs co-immobilisation of biotinylated CYP and CPR proteins onto glass solid supports coated with a hydrogel matrix and functionalised with streptavidin.

3.1.3 Surface coatings

Glass surfaces have the tendency to interact with residues of protein by ionic interactions and can result in denaturation of the folded protein structure (Afanassiev, Hanemann, & Wölfl, 2000). For specific immobilisation of protein onto the surface, the surface chemistry therefore has to be modified to promote the desired interactions. Therefore a common strategy used is a coating of the glass surface to minimise glass interactions and to facilitate directed biomolecule attachment. To facilitate transient protein-protein interaction such as for CYP-CPR interaction there is also an additional need for restricted diffusion movements, requiring flexibility in surface architectures.

There are several strategies that have been used to coat the surface of glass for functional protein immobilisation (Table 3.2). Glass surfaces ordinarily exist with a terminal hydroxyl group when silanised. There can therefore be both ionic and non-ionic weak interactions of such hydroxyl groups to charged amino acid groups on protein chains that can cause proteins to denature upon binding to glass so silanisation is often used for binding of DNA but less often with protein.

The two dimensional coatings allow for the incorporation of spacers that can extend the atomic distance from the surface by using functionalised polymers (Table 3.2). This can allow for a significant level of flexibility and therefore not restrict transient CYP-CPR interactions. However, the microenvironment is not in an aqueous phase and can lead to rapid water loss due to evaporation. Evaporation rates are an important factor for protein microarrays as the spot sizes are typically 300 micrometers or less and hemispherical, therefore having a large surface area from which to lose water: the smaller the spot, the larger the surface area to volume ratio is so the faster the evaporation rates and the quicker the spot dehydrates (Dugas, Broutin, & Souteyrand, 2005). The loss of water from reaction spots leads to a concentration effect which results in an alteration of solute quantities in relation to evaporative loss. This complicates analysis of enzyme kinetics from microarray surfaces when chemical products are concentrated and evaporative losses are non-uniform.

Three dimensional surface coatings allow for the creation of an aqueous microenvironment by utilising hydrogel polymers that can be branched (Table 3.2). Hydrogel polymers can exist in a fully hydrated state and create a hydrophilic like environment; for added flexibility the hydrogel can be conjugated to polyethylene glycol (PEG) hydrophilic polymers. This acts as an added spacer for enhanced flexibility for protein diffusion, potentially conducive for CYP-CPR transient interactions. The hydrogel coating strategy together with the polyethylene glycol spacers was therefore selected for microarray construction in this thesis due to the aqueous microarray environments to control water loss and due to its inherent polymer flexibility. For protein attachment, the PEG polymers can be terminally functionalised with amine reactive N-hydroxysuccinamide ester groups. These groups have a high affinity for primary amine groups found in protein backbone and amino acid residues. In addition, slides with such coatings are known to have relatively good signal to noise ratios because they resist non-specific protein adsorption and reduced auto-fluorescence when assaying with fluorescent probes (Walter et al., 2010).

Table 3.2: Surface coatings with application to retaining proper protein microenvironments modified from (Kusnezow & Hoheisel, 2003)

Surface coating	Method	Advantages	Disadvantages
One dimensional	Silanisation, poly-lysine coating	Relatively simple manufacturing process	Protein denaturation by interactions with glass
Two dimensional	Polyethylene glycol polymer, Self assembled monolayers	Reduced protein denaturation, flexibility for diffusion rates	Rapid protein dehydration by evaporation from small spots
Three dimensional	Branched polymers, dendrimeric coatings	Reduced non-specific interactions, aqueous microenvironments	Complex manufacturing for uniform coatings

3.1.4 Array fabrication

In the construction of microarrays the surface, protein, enzyme reaction components and detection strategy have to be prepared in a high throughput workflow. Assaying thousands of reactions sequentially can be computationally intensive and variables can be lost, resulting in questionable data integrity. Therefore, all systems have to have checks and quality control steps to ensure high quality, reproducible results.

3.2 Results: Surface preparation

Microarray surface preparation first involved the coating of hydrogel PEG slides with a uniform layer of streptavidin. This was done using a bulk association of streptavidin protein and derivitising multiple slides per batch to reduce coating variability. Quality control steps were performed by testing slide homogeneity and spot uniformity by printing a biotinylated protein (BSA) conjugated with a fluorescent dye (Cy5), and coeffecients of variations were measured from data extrapolated from fluorescence intensities. Surface characteristics were determined by atomic force microscopy (AFM) to evaluate underlying surface uniformity and protein binding following derivatisation with streptavidin. AFM is a robust technique to evaluate surface characteristics and requires minimal

sample preparation. The principle of AFM involves surface contact of an atomic width pin attached to a flexible cantilever (Hentschel et al., 2013). This apparatus is dragged or scanned over an area of a surface. Surface-pin interactions cause deflections in the attached cantilever which is detected by laser photodiode detectors and translated to topographical images of surface roughness. There are also non-contact tapping modes AFM whereby the pin does not make direct contact with the underlying surface but is in close proximity and the pin vibrated at its resonant frequency while being swept across an area. Surface-pin interactions here lead to attraction or repulsion forces experienced by the pin and conveyed to the attached cantilever, which causes deflections that are measured and translated to topographical images. AFM in both these modes is useful for surface characterisation. Polymer orientations and protein binding and uniformity can be quantified using such a technique. AFM was therefore selected over other surface characterising technologies due to its minimal sample preparation and adaptability to microarrays.

3.3 Results: Printing

There are two means of printing multiple discrete spots namely, contact and non-contact printing. In contact printing a solid or split-pin is used to transfer a pin-sized amount of protein onto predetermined locations on the array surface robotically (McQuain, Seale, Peek, Levy, & Haselton, 2003). This produces uniform sized spots that are reproducible and can be multiplexed by using several pins in a single print run. Protein carry-over or cross contamination can be an issue here, as the pins are dipped in a solution containing protein prior to contacting the microarray surface. Printing several different proteins in the same print run can result in cross over and therefore quality control steps have to be in place to ensure that the pins are washed efficiently between each protein print. In non-contact printing, droplets are dispensed using pressurised ink-jet type propulsion or using acoustic or piezoelectric dispensing techniques (Barron et al., 2005). This type of printing requires changing of the dispensing tips or apparatus between printing different proteins and the printing tips can become clogged up with debris and can disrupt the printing process.

For the purposes of fabricating functional protein microarrays that involve multiple print runs printing in overlaid patterns, the contact mode of printing was therefore selected. This has the advantage of not having to replace the printing head prior to each print run, while maintaining the printing co-ordinates for patterning. This becomes crucial for sample tracking in high throughput workflows and when overlaying spot printing. Discrete locations are fixed prior to print setup and those locations are automatically setup in a grid manner and conducive for matrix type data extrapolations using image processing.

The printing process involved dispensing the protein followed by incubation for protein binding and washing. The microarray slide was then dried and prepared for substrate overlay, and initiation of reactions by cumene hydroperoxide or NADPH.

3.4 Detection

Using a fluorescent label based detection strategy, the microarray reactions of up to 5000 individual micro spots were monitored as a function of time to allow the reactions to progress. Quantitative detection of fluorescent reaction products with sample tracking is essential for high data integrity. For fluorescent detection, the fluorophores have to be excited by an incidence laser in the optimal wavelength range for excitation and detected in the optimal range of fluorescence wavelengths, filtering out incidence and interfering light. Several strategies can be considered for achieving quantitative detection. The standard microarray workflows optimised for DNA microarrays commonly make use of simple two colour fluorescent detection namely Cy3 and Cy5, which show fluorescence in the green (excitation at 550nm, emission at 570nm) and red (excitation at 650nm, emission at 670nm) channels respectively in microarray scanners. However, detection can be obtained by other means such as fluorescence or confocal laser scanning microscopy. This can have the advantage of a wider range of fluorophore detection in various colour ranges thereby not limiting fluorescence detection to two colour detection (Patterson et al., 2006).

3.5 Assays

3.5.1 Quality control and surface characterisation

Quality control experiments were carried out to control for slide coating homogeneity and uniformity of printed spot sizes. Cy5 conjugated biotinylated BSA was used as a probe protein to probe slides that were functionalised with streptavidin using batch processing.

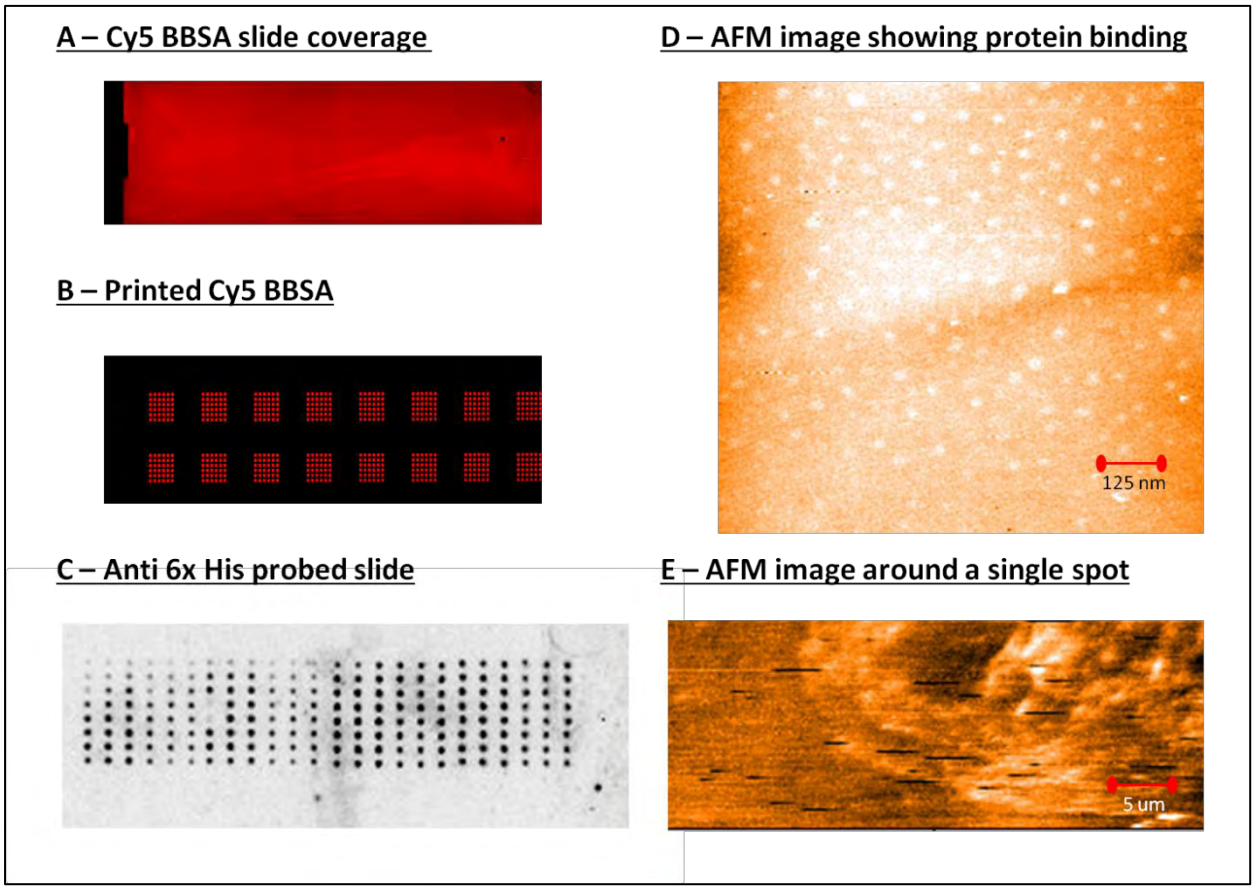


Figure 3.1: Surface characterisation and surface quality controls

Figure shows quality control checks to determine reproducibility of fabricated slides. Panel A and B show Cy5 biotin BSA (Cy5 BBSA), probed to show slide coating homogeneity and spot uniformity. Panel C shows confirmation of His-tagged CYP and CPR protein binding. Panel D and E shows scanning AFM topographical images of CYP-CPR spot centre over a 1um x 1um area (panel D) and spot edges 50um x 25um area (panel E).

3.5.2 Surface measurements

Coating homogeneity and spot uniformity displayed coefficients of variation of less than 15% of the mean spot intensity (Figure 3.1, Panel A and B). This was a good result and suggested that the surface coating was uniform and spots were being printed reproducibly. This was further confirmed using CYP and CPR printing and probing with Anti 6x His HRP (Figure 3.1, Panel C). Surface characterisation measurements revealed that the binding capacity of the functionalised PEG polymers were an average of 202.67+/-33.31 streptavidin units per micrometer squared of surface. Binding units are represented as light spots indicating protein repulsion of the AFM tip (Figure 3.1, Panel D). The light spots have an approximate diameter of 25nm, it is likely that they represent streptavidin tetramers; however the theoretical diameter of the streptavidin tetramer is 7-10nm in diameter. It is indeed probable that the light spots represent the streptavidin bound with CYP and CPR with an estimated diameter of approximately 4nm each. There can be up to 4 bound CYPs or CPR to a single streptavidin tetramer, therefore allowing for 23-25nm diameter, such evidence points to these light spots being streptavidin bound to CYP and CPR.

A single microarray spot with a diameter of 200 μm therefore has a total number of approximately 6 000 000 binding units available for CYP or CPR binding. Furthermore, taking measurements of distance between binding units reveals that the average distance between streptavidin units is 52.87+/-3.11 nm.

It is therefore possible to estimate turnover rates K_{cat} using the data from AFM images suggesting approximately 6 000 000 available binding units per spot, average spot volume as 100nl, the CYP P450/P420 ratios measured by CO difference spectra (chapter 2.4) and the V_{max} kinetic parameter from the microarray fluorescence measurements. Using the AFM technique in this way to measure total active protein present on the microarray surface is novel in the protein microarray field. See supplementary for detailed calculation of total enzyme concentration.

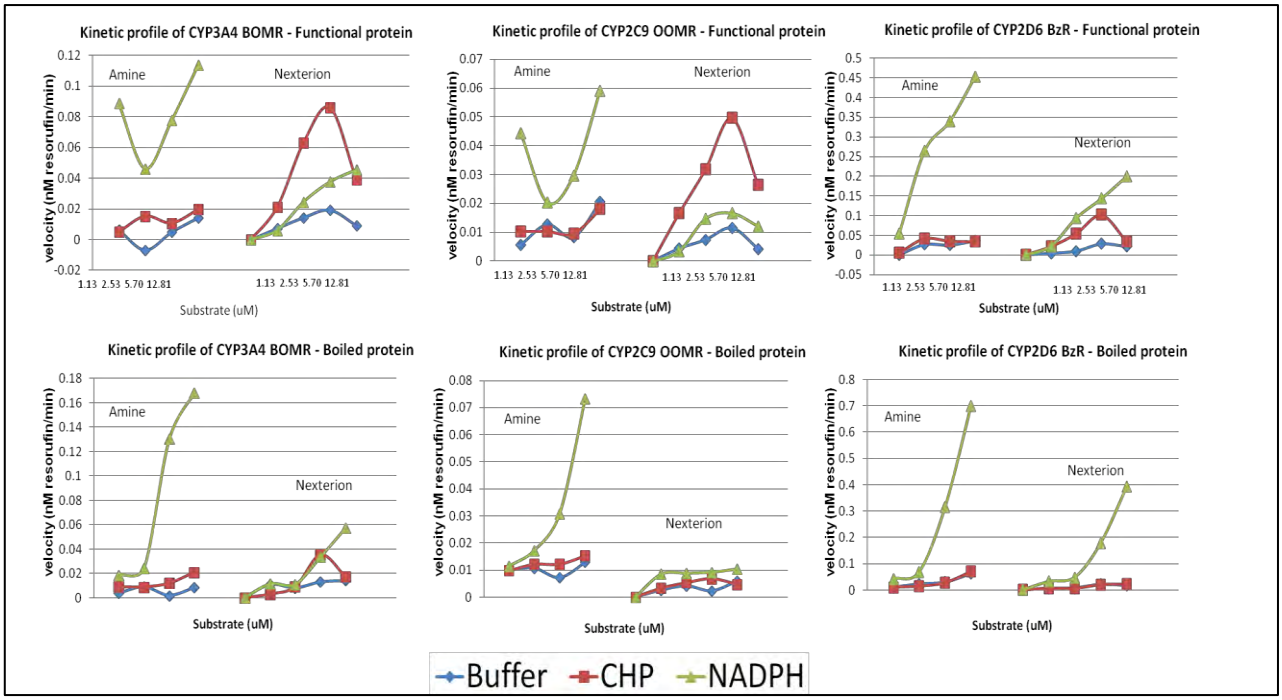


Figure 3.2: Kinetic plots of CYP kinetics without CPR showing functional protein and boiled protein controls with red fluorescent resorufin-based substrate

Figure shows robotically spotted microarray results for the CHP CYP driven reactions on both Nexterion and Amine slides with the resorufin based substrates. The reactions consisted of buffer only, CHP and NADPH additions. Activity was determined by scanning for the fluorescent product in the microarray scanner. The top three panels show CYP3A4, CYP2C9 and CYP2D6 functional protein while the bottom three panels show that of the boiled CYPs respectively. Data was extrapolated from standard curves and results fitted to Michaelis-Menton enzyme plots. Kinetic parameters were estimated using best fit models.

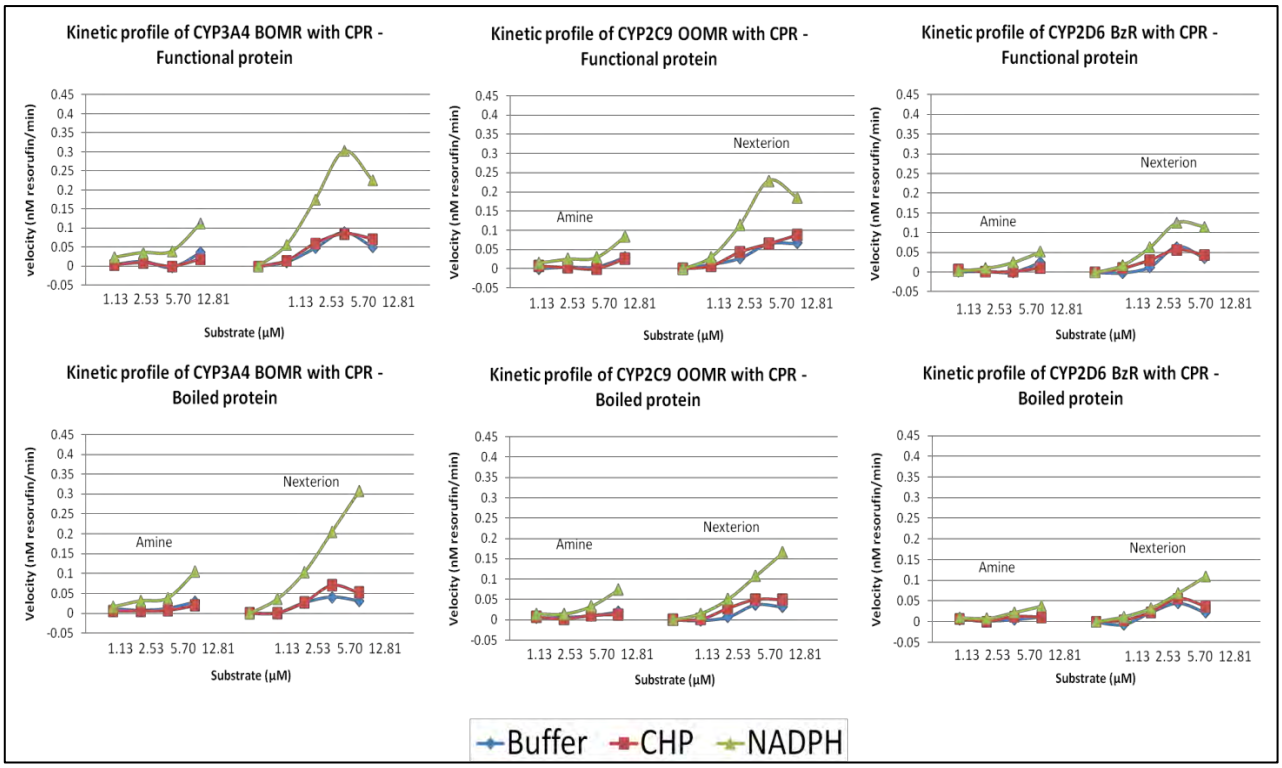


Figure 3.3: Kinetic plots showing CYP and CPR co-immobilised onto the microarray surface with red fluorescent resorufin-based substrates

Figure depicts the results of the robotically spotted reactions of CYPs with CPR on both Nexterion and Amine coated microarray slides. The reaction conditions are with buffer, CHP or NADPH additions to the reaction by overlaying the reaction mix. The top 3 panels show functional CYP3A4, CYP2C9 and CYP2D6, while the bottom three panels show the denatured boiled proteins respectively. Both the CYPs and CPR was denatured by boiling. Again data was extrapolated from standard curves and results fitted to Michaelis-Menton enzyme plots. Kinetic parameters were estimated using best fit models.

Table 3.3: Kinetic parameters obtained from solution phase with resorufin based substrates compared to those derived from microarray experiments

CYP	Microarray V_{max} (nm/min)	Microarray K_m (µm)	Microarray* K_{cat} (min^{-1})	Solution phase V_{max} (nm/min)	Solution phase K_m (µm)	Solution phase K_{cat} (min_{-1})
3A4 CHP	0.021 "-0.075 to 0.212"	2.46 "0.0 to 4.37"	**0.0026**	***0.43***	***3.8***	**0.0020**
3A4/CPR NADPH	0.092 "0.002 to 0.129"	5.03 "0.0 to 6.40"	**0.023**	-	-	-
2C9 CHP	0.041 "-0.026 to 0.108"	1.03 "0.0 to 6.539"	**0.005**	***0.59***	***2.6***	**0.0012**
2C9/CPR NADPH	0.042 "-0.011 to 0.0449"	1.39 "0.0 to 10.04"	**0.011**	-	-	-
2D6 CHP	0.071 "-0.122 to 0.264"	1.58 "0.0 to 3.17"	**0.009**	***0.67***	***1.7***	**0.115**
2D6/CPR NADPH	0.316 "0.0169 to 0.614"	5.02 "0.0 to 7.98"	**0.079**	-	-	-

Table shows figures extrapolated from velocity curves plotted from resorufin molecules formed per minute. Data modelled using Michaelis-Menton kinetic plots and K_m and V_{max} parameters estimated using best fit model. Figures represent mean and range that is within a standard deviation. The K_m values were fixed to positive values as K_m is a concentration and cannot lie in a negative range. Data represents intra array replicates of at least n=12. Red fluorescence was detected using the microarray scanner.

*Novel methodology using P450/P420 ratio and AFM image data to estimate more accurately the quantity of active protein present per reaction spot.

The V_{max} data obtained from microarray assays are approximately ten fold lower than those obtained using solution phase assays, however the K_m values are comparable. Interestingly, the K_{cat} values are similar with the CHP assays but are about tenfold larger when using CYP-CPR reconstitution with NADPH. This suggests that the substrate is being metabolised more efficiently per unit protein using the immobilised microarray format. No other adjustments were performed on CYP activities and no corrections performed for the observed activity with NADPH and boiled protein conditions.

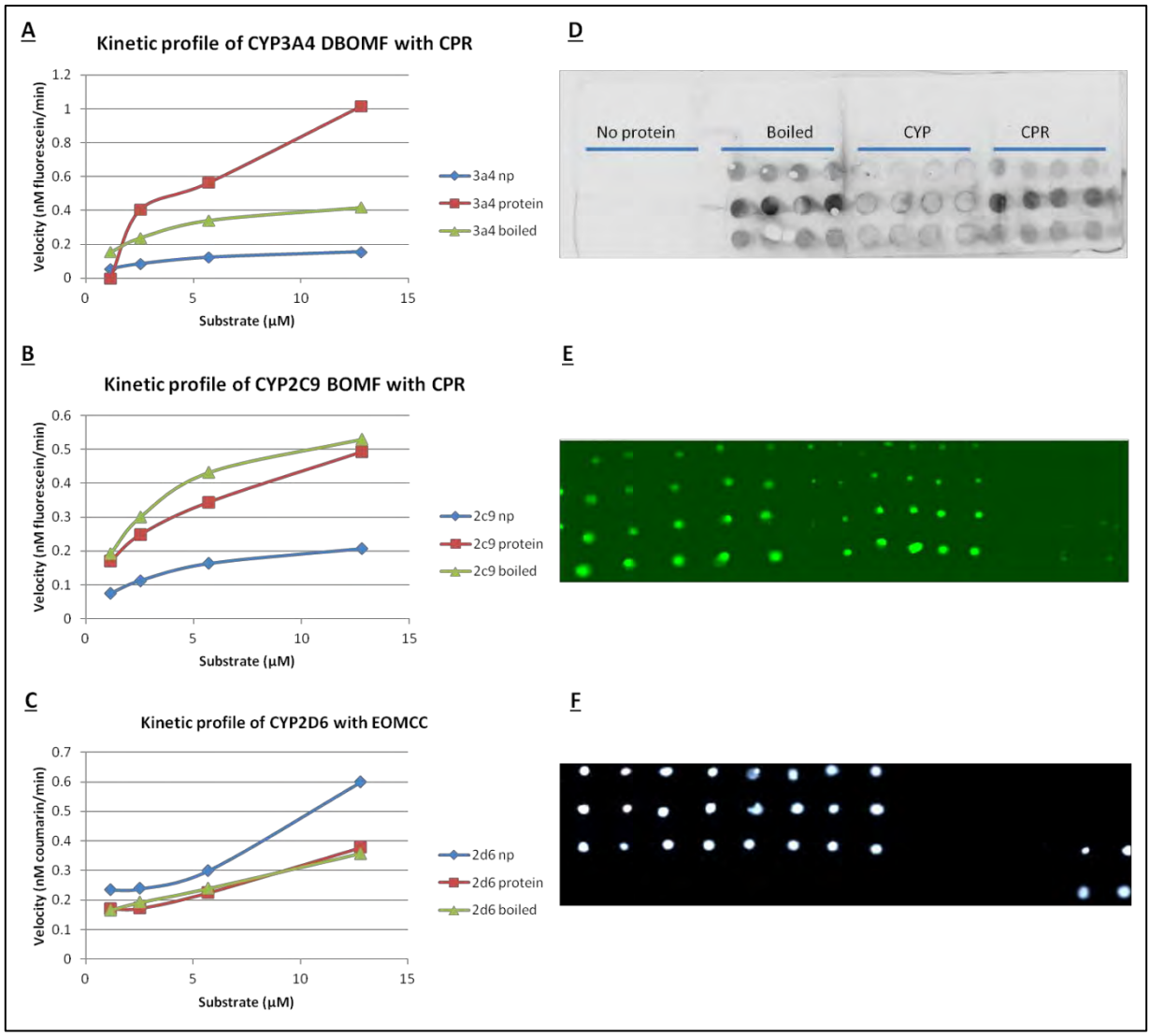
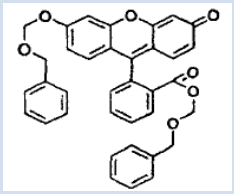
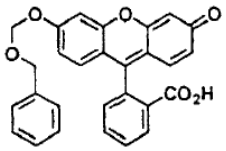
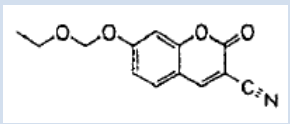


Figure 3.4: Kinetic assays with green (fluorescein) and cyan (coumarin) fluorescent substrates with microarray immobilisation

Figure shows data from microarray experiments on Nexterion slides using green and cyan fluorescent substrates detected using confocal microscopy. Data was extrapolated from standard curves and results fitted to Michaelis-Menton enzyme plots. Kinetic parameters were estimated using best fit models. Panel A-C shows kinetic plots of CYP3A4, CYP2C9 and CYP2D6 with their relative substrates and showing "no protein" ("np") and "boiled protein" controls. Panel D shows a slide that has been probed with anti 6x His HRP antibody. This slide blot shows His-tagged and immobilised protein. Panel E and F shows fluorescent images acquired by confocal scanning using the camera in tile mode.

Table 3.4: Kinetic parameters derived from microarray experiments with fluorescein or coumarin-based substrates and compared to published data

CYP/substrate	Microarray V_{max} (nM/min)	Microarray K_m (µM)	Microarray K_{cat}(min^{-1}) *	Literature V_{max}(nM/min)	Literature K_m (µM)	Lit. Kcat (min^{-1})
3A4 BDOMF 	2.49 "-4.98 to 9.95"	8.53 "0.0 to 10.28"	**0.031**	2.9	7.8	**0.143**
2C9 BOMF 	0.61 "0.32 to 0.89"	3.70 "0.0 to 8.149"	**0.008**	0.31	2.60	**0.003**
2D6 EOMCC 	0.43 "-0.09 to 0.95"	3.17 "0.0 to 13.80"	**0.005**	0.40	3.0	**0.002**

Data from manually spotted microarray experiments using purified protein. Reactions were monitored using confocal microscopy and extrapolated to standard curves of green fluorescent fluorescein and cyan fluorescent coumarin. Results represent the average and standard range of triplicate reactions. Values compared to published data and chemical structures adapted from (Makings & Zlokarnik, 1999).

***Data generated using novel AFM based protein binding measurements as described previously and utilising a 2000nm spot size as these results represent manual hand spotting assays.**

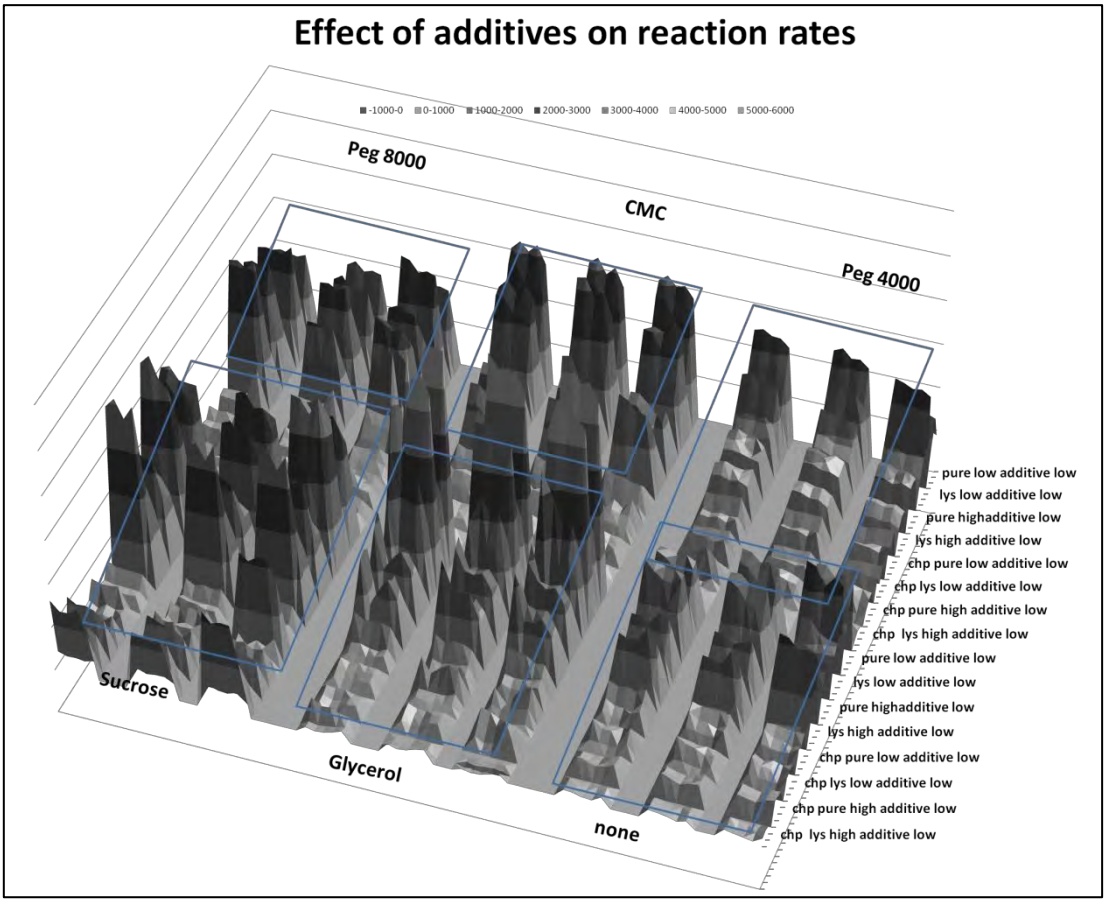


Figure 3.5: Representation of kinetic rates and effects of additives on kinetic profiles

Chp – Cumene hydroperoxide Lys – Crude lysate spotted

High Additive – Reaction in presence of 50uM additive

Low Additive – Reaction in presence of 5uM additive

Y axis represents reaction rate (nM/min)

This figure shows additive effects on CYP reaction rates using red fluorescent substrates. The microarray experiments were handled in automated workflows for consistency between additives. Reaction products were measured over a period of time that falls within the linear range of increase for quantification. Data were compared to standard curve of resorufin. Data

represents 36 replicate reactions on the microarray surface. Data shows rates of reactions with each track in a block representing CYP3A4, CYP2C9 and CYP2D6. With front of block to back representing increasing amounts of additive with both purified protein and protein lysate with CHP (CYPs only) and NADPH (CYPS and CPR).

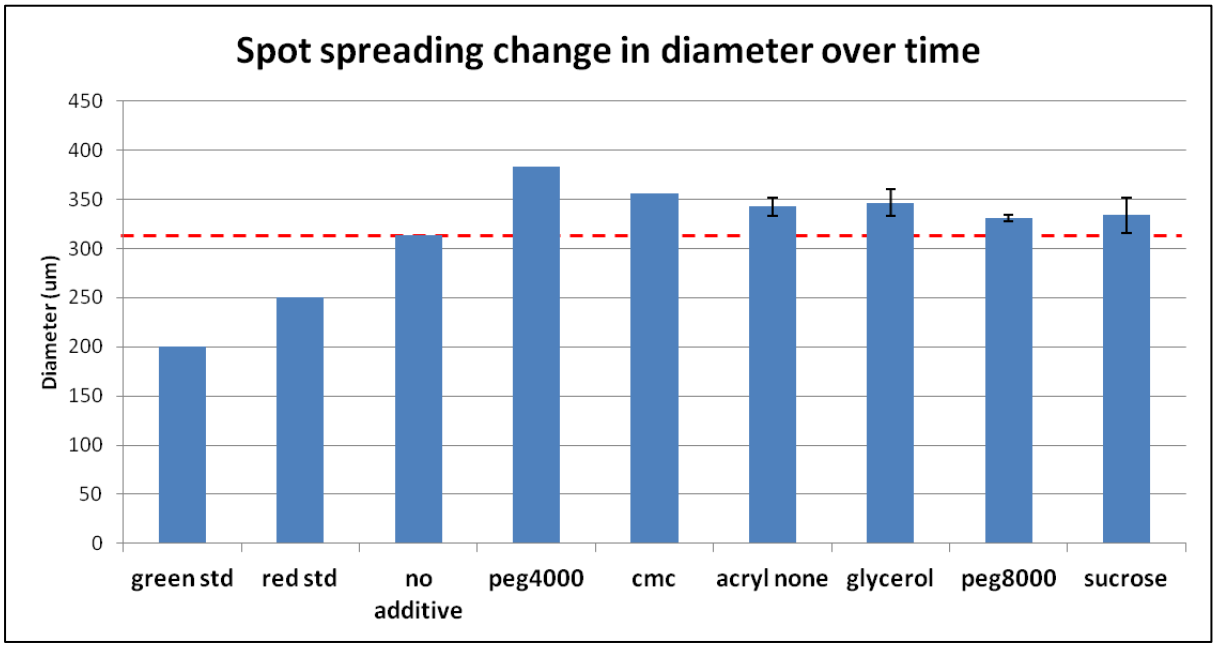


Figure 3.6: Spot spreading, judged by spot diameter, with additive supplementation

Green std – Fluorescein	cmc – carboxymethyl cellulose
Red std – Resorufin	acryl – acrylamide
peg4000 – polyethylene glycol 4000	peg8000 – polyethylene glycol 8000

This figure shows the effect of additives on the spot spreading in combination to loss of spot size by evaporation. Data bars represent the mean and standard deviation of at least 24 replicate reactions. Dashed line is in comparison to the reactions with no additive.

3.5 Discussion

This chapter focussed on CYP protein microarray fabrication and further expansion of this technology to CYP2C9 and CYP2D6. The three dimensional hydrogel coated slides with functionalised polyethylene glycol was selected for microarray construction. This was based on the desire to create an aqueous microenvironment with flexible polymers in order to control evaporation rates and protein diffusion respectively. Initially, for reproducibility and as a quality control measure the surface was characterised and validated for coating homogeneity and spot uniformity. Results shown in Figure 3.1 indicate the uniformity and spot reproducibility with coefficient of variation (CVs) less that 20%. This is acceptable and in line with published guidelines from microarray quality control standards (Patterson et al., 2006). Furthermore, the single one step purification and immobilisation of protein was confirmed by probing printed slides with anti 6x His HRP (Figure 3.1 Panel C). The binding capacity and underlying surface coating was determined using AFM in contact mode (Figure 3.1 Panel D and E). It was determined that there is a binding capacity of approximately 40 000 streptavidin units per printed microarray spot of 200μm diameter. The average distance between binding units was shown to be relatively large for direct contact at adjacent binding units. This suggests that protein-protein interactions between binding units may be hindered by steric strain. However, the PEG polymer flexibility may enable additional movement to facilitate protein interaction (Sukhishvili et al., 2002). In addition, the curved, coffee ring like patterns at the edge of spots (Figure 3.1 Panel E), suggest that the polymers maybe aligning and allowing for closer contact of binding units. The appearance of coffee-ring like patterns may suggest that at the microscopic level, this well-known effect is at play during microarray fabrication. This mere close proximity can allow for electron transfer from CPR to CYP systems (Page, Moser, Chen, & Dutton, 1999). Therefore, the reconstitution of CYP activity is favoured in the current microarray setup, and furthermore the K_{cat} kinetic turnover rate can be more accurately measured using AFM data in conjunction with CYP quantitation from earlier CO difference spectra to quantify amount of active protein bound in each spot.

Proceeding into microarray functional assays, CYP3A4, CYP2C9 and CYP2D6 were each immobilised and assayed first for activity using the peroxide shunt pathway and various red fluorescent substrates (Figure 3.2). Enzyme kinetic parameters such as the maximal velocity reveal properties of the functional enzyme that suggest rates of reactions that are important for drug metabolism and pharmacokinetics of CYPs. The K_m parameter is the concentration of substrate that yields half maximal rate of reaction and can be useful to interpret turnover rates per protein molecule. These parameters were measured and compared to literature values (Makings & Zlokarnik, 1999) (Table 3.4). The published data made use of bacterial membrane ("baculosome") immobilised enzymes and used the same substrates and peroxide shunt pathway to initiate the reaction process. This comparison is useful in contrasting the differences between a common form of lipid bilayer-based protein immobilisation and the microarray system used here. Data shown in Figures 3.2 and 3.3 reveal that CYP activity can be reconstituted both by using the peroxide shunt pathway as well as by NADPH and CPR interaction. From the profiles of the Michaelis-Menton plots the type of kinetic behaviour can be linked to substrate inhibition models in most cases. There is also a useful comparison when contrasting the protein activities on amine slides and hydrogel coated slides. Interestingly, the amine slides show significant activity even when only boiled CYP protein is immobilised, despite the boiled protein being denatured and presumably inactive. Therefore, the activity seen on the amine surfaces (boiled protein in Figures 3.2 and 3.3) probably represents an interesting surface phenomenon. Notably, this activity is seen only in the presence of NADPH, which suggests a reductive mechanism as NADPH is a mild reducing agent.

One possibility here is that NADPH reversibly reduces the resorufin moiety in the structures, creating a negatively charged species that then associates with the positively charged amine-terminal surface, promoting general acid-catalysed hydrolysis of the acetal linkages, releasing the reduced resorufin nucleus which can then re-oxidise. Further work is needed to verify this hypothesis, but it is consistent with the literature on reversible NADPH reduction of resorufin (Xu, Kong, Yeh, & Chen, 2008). It is notable that no similar effect was seen with fluorescein or coumarin-based substrates.

Nevertheless, the activity summarised in Table 3.3 suggests that the activity seen is comparable with other membrane-based immobilised techniques. Furthermore, activities in the absence of CPR and with the addition of NADPH do not yield optimal activity (Figure 3.2, functional protein activity with NADPH), as well as the activity with the addition of buffer only. This suggests that the activity seen on the immobilised array surface is CYP, CPR and NADPH dependent, and is consistent with effective CYP-CPR reconstitution. The fluorescent product is validated by fixed fluorescent excitation and emission settings and by comparing the fluorescence units to that of the standard product resorufin. The catalytic mechanism is thus presumed to be CYP dependent oxidative cleavage of protecting groups, liberating the fluorescent parent compound.

This CYP activity was also seen with other fluorescent substrates, namely the fluorescein and coumarin substrate scaffolds (Figure 3.4, Panels A-C). Again, the CYP activity was confirmed by comparing the fluorescence units to that of reactions performed identically but with boiled and denatured protein. For CYP3A4 and CYP2C9 the activity is consistent with protein reconstitution. In the case of CYP2C9, the boiled protein control (Figure 3.4, Panel B) shows a higher activity than the active protein, this is unexpected but can be reasoned by insufficient denaturation or subsequent refolding of denatured protein mediated by the microarray surface. The CYP2D6 experiments (Figure 3.4, Panel C), shows that the controlled experiments are not significantly different from each other. This can be accounted for as this fluorescent substrate (coumarin) is excited in the ultraviolet range, 410nm, which overlaps with NADPH excitation. This means that the fluorescence that is measured is pre-saturated with NADPH fluorescence, resulting in the data not being a reliable measure of CYP activity using NADPH but may be more suitable for peroxide driven CYP reactions. However, the data is promising for the green fluorescein substrates. Protein binding was confirmed by probing the slides after immobilisation and assaying with anti 6x His HRP (Figure 3.4, Panel D) which shows protein binding in the expected regions. As validation of CYP activity, the enzyme kinetic parameters for the fluorescein and coumarin-based substrates were compared to published data using immobilised baculosome preparations using the same substrates (Table 3.4). Interestingly, the data again matches closely to the published kinetic parameters. Even CYP2D6 which seemed to have NADPH

fluorescence interference matched closely to the validated parameters. This suggests that the method of immobilisation is appropriate for activity reconstitution and is amenable to high throughput workflows, independent of the exact nature of the substrates used for assay.

It is intriguing that the kinetic parameters match so closely to the validated kinetic parameters, even though the validated assays were done in very different conditions. The literature assays were done with bacterial membranes as the immobilisation strategy and in solution phase (96 well plate formats). Solution phase assays are different in that there is a constant volume and temperatures changes are negligible. Performing assays in micrometer spots on an array surface is more sensitive to changes in temperature and volume fluctuations. To counteract the effects of temperature and humidity changes various additives were applied in the printed spot assays. It is known that the evaporative loss in such droplets occur in two stages: initially, droplets flatten then proceed to shrink and this can lead to non-uniformity in drops across the surface (Dugas et al., 2005). Therefore, additives were included to reduce the seemingly small change in size and volume, since the minor effect of evaporation on the macro scale has vast implications for reactions in the micro scale. Furthermore, adding substances to micro droplets can cause coffee ring effects upon drying, with build up or increased concentration of additive at the periphery (Hook et al., 2010). Unfortunately, it is not logistically possible to keep the microarray slide in a humidified environment for the entire duration of the functional protein workflow. The microarray is kept humidified throughput the printing and overlaying procedure but is under atmospheric humidity conditions upon fluorescent scanning. This is due to the sensitivity of electronic and optical components of the fluorescent scanner that do not allow for artificial humidification. Interestingly, a number of the additives were seen to have an effect on the kinetic profiles of CYP reactions and on spot spreading (Figures 3.5 and Figure 3.6). The various additives have differing effects, with most increasing rates of reactions and some smoothing the kinetic plot. Glycerol as an additive both increased the rate of reactions and smoothed the kinetic plot, as well as kept the micro spot within the 300μm range. Interestingly, Figure 3.5 also shows the effect of spotting protein lysate and purified protein. The spotting of lysate has lower rates of reactions compared to that of the purified protein, irrespective of additive type or amount added. This leads to the understanding

that the increased rate of reaction is being facilitated by an increase in protein present. It has been known that CYP protein activity can be increased by increasing protein content and even so in the absence of immobilisation or lipid binding (Müller-Enoch, Churchill, Fleischer, & Guengerich, 1984). This intriguing phenomenon of increased CYP activity with more protein forms the basis of the investigations in the next chapter.

In conclusion, the work described in this chapter has expanded the scope of functional protein microarrays to CYP2C9 and CYP2D6 isoforms. Surface chemistry uniformity and functionalised surface binding capacity allows for uniform immobilisation of CYP protein in sufficient quantities. Data reveal that protein activity is being reconstituted in both CHP and CPR dependent assays and moreover the kinetic parameters were comparable to the literature. If the results obtained with boiled protein are considered then CYP3A4 showed activity with CHP and CPR, CYP2C9 showed activity with CHP and CYP2D6 activity was minimally demonstrated. The red fluorescent resorufin based substrates had an unusually low rate of reaction to comparable studies, while the green fluorescein and cyan coumarin substrates match more closely to other validated studies. Furthermore, this chapter has shown that the microarray format is not limited to simple two colour substrates commonly used for DNA microarray detection since other means of fluorescence detection such as scanning confocal microscopy can be used in a quantitative manner with comparable sensitivity. Studies on additives in the reaction mix revealed that compounds such as glycerol counteract spot spreading and evaporation rates, allowing for more consistent kinetic profile plots. An interesting aspect of CYP properties was revealed in that spotting protein lysate and purified protein had differing rate of reactions, purified protein having higher rates of reactions.

The AFM studies revealed that the underlying surface was uniform but that the density of streptavidin tetramers was lower than previously expected, which has certain implications for the mechanism of CYP-CPR interactions at the array surface.

Furthermore, evidence from AFM of localised distortions in the polymer surface of the printed microarrays may also influence the frequency of CYP-CPR interaction and serves to support the original choice of a flexible 3D hydrogel surface in this thesis.

3.6 Methods

3.6.1 Quality control - Surface homogeneity

All chemicals used were of molecular biology grade and buffers made up fresh before use. Water used in buffer compositions was of reagent grade with a resistivity of at least 5Ω (Purified reverse osmosis with Elix 3uv fitted with Progard 10µm filter (Merck Millipore)). Initially the Cy5 conjugated Biotinylated BSA (BBSA) was made using the Oyster amine-reactive Cy5 similar (Oyster 647, Luminartis), biotin-NHS (Sigma) and BSA (BSA fraction V, Roche). The conjugation reactions were carried out in two stages, the first being BSA biotinylation. This was done by adding biotin-NHS to BSA in equimolar concentrations, and incubating the reaction in Phosphate buffered saline (1x PBS) at room temperature for 1hr. The reaction was then confirmed for biotinylation using the HABA biotinylation calorimetric assay (Alfa Aesar) according to the manufacturer's instructions. The second stage of conjugation proceeded with adding biotinylated BSA to Cy5 dye in 1x PBS and allowing the reaction to proceed at room temperature for 1hr with regular agitation. The reaction was terminated by passing batches of the reaction through desalting columns (Sigma) to remove free biotin-NHS and unbound Cy5 dye. The dye to protein ratio was measured by UV/Vis absorption spectra and published extinction coeffecients. Approximately 4 dye molecules per protein were deemed sufficient for quality control assays.

3.6.2 Slide derivatisation

For assaying slide coating homogeneity, Nexterion H slides (Schott, Germany) were first thawed to room temperature before opening the packaging. These slides were derivatised using batch processing for consistency by coating with streptavidin (PROSPEC, pro-283) following the procedure outlined briefly below. The procedure entailed incubation of the NHS-activated Nexterion H slides with streptavidin (1mg/ml) solution in HEPES buffer (200 mM KCL, 0.02% Triton X-100, 50 mM HEPES, pH 8.5) for 1hr, followed by a blocking step in ethanolamine (50mM ethanolamine in 50 mM potassium phosphate buffer pH 8.0) for an additional hour. The slides were then washed 3 times in

100mM potassium phosphate buffer K_2PO_4 buffer (100mM K_2PO_4 (94% K_2HPO_4, 6% KH_2PO_4, pH8.0) and once in dH_2O followed by drying by centrifugation at 2000g for 5mins. Slides were then stored dry at -20°C in a desiccated environment. For slide coating homogeneity assay slides from the beginning, middle and end of batch processing were chosen and immersed in a solution of 10 µg/ml of Cy5 BBSA in HEPES buffer for 1hr at room temperature with gentle agitation. The slides were washed 3 times in 100mM potassium phosphate buffer and once in dH_2O followed by drying by centrifugation. Slides were then scanned using the microarray scanner (LS Reloaded and Array-Pro Analyser software, Tecan) in the red channel (excitation 530nm, emission 585nm) in automatic gain control (AGC) mode. Images were obtained and processed using microarray image processing software (GenePix Pro 6.0). The fluorescence intensity (background subtracted) was extrapolated from a set of dummy spots distributed across the slide surface and grouped together and data analysed for mean intensity and standard deviation. The coefficient of variation (CV) was calculated as the standard deviation as a percent of mean.

For spot homogeneity, slides were subject to derivatisation as above and 10 µg/ml of Cy5 BBSA in HEPES buffer was printed using the (QArray 2, Genetix). Cy5 BBSA was made-up in a source plate of 384 well plate high recovery plates (Genetix). Solid pins were used to print the spots in a 16 plex format. Due to single sample being printed, no pin washing steps were programmed, but spots were printed with 3 stamps per spot under 50% relative humidity at 25°C. Following printing, the slides were incubated at room temperature in the humidified environment for 1hr. The slides were then washed 3x in potassium phosphate buffer and 1x dH_2O, scanned and data extrapolated and CVs calculated as before.

For anti-His probe quality control, the slides were derivatised as before and printed with CYP isoforms with or without CPR in 100mM potassium phosphate buffer with 15% glycerol, in the same manner as the Cy5 BBSA printed spots. The slides were washed as before and probed with a solution of anti 6x His HRP antibody (9.7 µg/ml) (monoclonal anti-polyhistidine-peroxidase antibody produced in mouse; Sigma) in 1x PBS with 0.05% Tween 20 (Sigma) for 1hour at room temperature or 4°C overnight in the dark. Slides were then washed as before and incubated with Luminol peroxide

solution (Western Bright ECL, Avansta) for 5mins. The luminescence was then detected using a G-box (Syngene) CCD camera.

For surface characterisation, AFM analysis was performed. The AFM (Easyscan 2 AFM, Nanosurf) was used in contact mode with AFM pin of thickness 2μm (Nanosensors, Silicon SPM sensor). Slides were derivatised and protein immobilised as before. The slides were then dried by centrifugation at 2000g and kept under a gentle air stream for 5mins to ensure complete drying. Slides were fixed onto the piezoelectric stage using double sided tape and ensured flat using a spirit level. Areas were selected and approached with the tip; tip scans were performed using either 50μm^2 or 1μm^2 areas. Images were analysed using the Easy scan2 software to measure distance between streptavidin binding units.

3.6.3 Assaying

3.6.3.1 Manual functional assays

Manual assays were carried out by using a derivatised Nexterion H slide (Schott, Germany) or on a superamine slide (Arrayit, USA) after equilibration at room temperature for 1 hr. CYP protein mixes were made up in 100mM potassium phosphate buffer with 15% glycerol and stored on ice prior to immobilisation. Boiled protein mixes consisted of CYP protein boiled at 99°C for up to 15mins and produced no noticeable peak at 450nm in the CO difference spectra. Protein was spotted using a micropipette dispensing 0.5μl of volume onto a defined spot on the surface using a 384 well plate as a backing mask. Spots were incubated for 1-2 hours at room temperature for immobilisation. The slides were then washed in 3 times of 5min each in potassium phosphate buffer and once in dH_2O for 5mins. The Vivid substrate (Invitrogen) mixes were prepared and stored on ice protected from light. Once the slides had been centrifuged dry at 2000g, 0.5μl of substrate mixes were dispensed onto the region where the protein spot was spotted again using the 384 well plates as a backing mask, and 0.5μl of cumene hydroperoxide (Sigma) or NADPH (Sigma) was then further dispensed or overlaid with a micropipette for reaction initiation. A standard curve of a serial dilution of fluorescent standards

(fluorescein sodium salt, 3' cyanoumbelliferone, resorufin sodium salt, Sigma) was created in 100mM potassium phosphate buffer and spotted in duplicate on each respective array surface. Once the reactions were initiated, the slide was repeatedly scanned in the appropriate fluorescent channels, either in the LS Reloaded microarray scanner in AGC mode or using a confocal camera (AxioVert LSM 510, Zeiss) in tile scanning mode. Data was extrapolated as before from raw images and florescence units converted to concentration using the standard curves (GenePix Pro 6.0 or AxioVision 4.7). Fluorescence was plotted against time and the average gradients representing the reaction velocity were determined. Velocity and substrate concentration was imported into Graphpad Prism and fitted to the Michaelis-Menton plots. Enzyme kinetic parameters were estimated using the best fit model.

3.6.3.2 Automated functional assays

Nexterion and amine slides were equilibrated to room temperature as before and placed in the printing chamber, maintained at 24°C and 50% relative humidity. CYP protein samples were prepared in 100mM potassium phosphate buffer with 15% glycerol in 384 well source plates (Genetix). The source plates were placed in the printing plate and the printing programme was set to print spots in the defined locations, with 3 stamps per spot. Since several protein samples were printed on each microarray, a wash cycle was conducted after each source plate as follows: 1x at water wash; 1x 70% ethanol in dH_2O wash; 1x water wash; and final air dry. Slides were maintained in the printing chamber for an additional hour to allow for immobilisation. Printed slides were then removed and washed as before and dried by centrifugation at 2000g, before being placed back into the printing chamber. Thereafter, substrate mixes were introduced into a fresh source plate and printed in the desired printing pattern with a single stamp per spot and the same wash cycle as before. Immediately following the substrate printing cycle, the reactions were initiated by printing cumene hydroperoxide or NAPDH in an overlaying manner. Slides were thereafter removed and placed in the scanner carefully. Reactions were monitored by repeated scanning in the relevant channels; data was extrapolated and kinetic parameters established as before.

3.6.4 Post assay quality control

Following assaying, the slides were washed as before and probed with anti-His HRP in PBS-Tween for 1hr at room temperature or at 4°C overnight. The slides were incubated with luminol peroxide solution for 5mins and thereafter placed between transparency films and detected using a Syngene G-box.

Chapter 4 – High protein activity reconstitution

4.1 Introduction

Membrane binding of proteins provides control over relative geometric orientation and also provides restriction in diffusion rates to two dimensions and therefore greatly enhances the chance of protein interaction such as the CYP-CPR interaction. An alternate method of activity reconstitution can in principle be achieved by simply assaying CYP functionality in high protein concentrations, as this in essence will also increase the probability of protein interaction, not by restricting diffusion rates but by providing a greater physical number of interaction surfaces. Results from the functional protein microarrays experiments described in the previous chapter reveal that high relative protein concentrations provided by the printing process, may be driving reaction rates (Chapter 3). Previous published work has shown that CYP activity can be reconstituted in solution using high protein concentrations of CYP and CPR in the absence of immobilisation on surfaces or bound in membranes (Müller-Enoch et al., 1984). Other works have shown that in order to increase the probability of CYP-CPR interactions the ionic strength of the buffer is important: the higher ionic strength buffers stabilise the charged interactions of the CYP-CPR complex (Voznesensky & Schenkman, 1994). Furthermore, use of non-ionic detergent systems at concentrations below the values for micelle formation can activate CYP function in reconstituted systems, possibly by creating hydrophobic conditions and favouring substrate binding close to the heme (Myasoedova, Arutyunyan, & Magretova, 2006). Therefore, taking these reports in combination, a system of reconstitution was developed including using high amounts of protein in the absence of lipid or immobilisation, together with high ionic strength buffers and trace amounts of mild, non-ionic detergent. In principle this combination should increase the probability of CYP-CPR interaction and therefore reconstitute sustained, optimal activity. To measure and detect CYP activity in solution, several strategies can be employed; of these, the label-free strategy is most robust and can be cross validated with other models of *in vitro* and *in vivo* CYP metabolism. A limitation to label-based activity probes, such as fluorescent probes, is the requirements required imposed by the chemical scaffolds needed for optimal

detection, meaning that such label-based approaches can only be used to determine drug metabolism parameters through use of competition assays and cannot directly determine the identity of drug metabolites formed.

Therefore, the aim of this chapter is to further investigate high protein activity reconstitution using label-free metabolite detection techniques, with probe substrates and a complex drug rifampicin.

4.2 Label-free detection

The advantage of label-free detection in CYP functional assays is that the assays are not limited by substrates as in fluorescent or luminescent assays. Using label-free detection can reveal the identity of multiple metabolites simultaneously and can be quantitative and multiplexed. There are several label-free technologies available for detection and identification of CYP metabolites; these have varying sensitivities and some are amenable to high throughput workflows.

4.2.1 Mass spectrometry

Mass spectrometry (MS) is a robust and very sensitive technique to detect and quantify various molecules. This technique is base on the principle of ionisation and accurate mass and high resolution molecular weight detection. This technique determines the mass to charge ratio of individual ionised molecules that shows the relative abundance of that ion in a sample. Samples can be ionised using various ionisation strategies and molecular weight can be determined using various mass analysing strategies. MS techniques have been widely used and validated for the identification of drug metabolites in CYP label-free assays (Shumyantseva, Bulko, & Archakov, 2005). Due to the versatility and robustness of MS, this technique was selected for label-free detection of the CYP metabolites from probe substrates and a complex drug.

4.2.2 Tandem MS

MS is a single round of analysis; this involves ionisation and detection of ions that can yield mass to charge ratios and identification of various charged species. This single dimension of mass

spectrometry can reveal molecular weight of molecules at atomic precision; however molecules having the same molecular weight (such as hydroxylation at different positions in drug molecules) cannot be differentiated. Therefore, a round of fragmentation can be introduced since molecules of the same molecular weight will likely have unique fragmentation patterns. MS/MS is a dual or tandem MS setup which involves two rounds of MS analysis. Fragmentation can occur by several methods including collision of accelerated ions with an inert gas such as argon, nitrogen or helium (Oppermann et al., 2012). The kinetic energy is converted to energy that can break bonds and results in smaller fragmented ions that are then detected. Fragmentation is useful to determine the identity of the ions as well as infer some structural information. MS^n is the n rounds of fragmentation and detection of ions that yields high structural information (Want, Cravatt, & Siuzdak, 2005). This type of MS is done in trap type instruments. Mass spectrometry fragmentation is usually done using the collision induced dissociation, although high collision dissociation yields more efficient fragmentation.

4.2.3 Ionisation strategy

There are several strategies used to induce ionisation in chemical compounds. The ionisation technique required for CYP metabolite identification and quantitation is required to be robust enough to ionise various small molecules without imparting high energies that can result in fragmentation. The various soft ionisation techniques include matrix-assisted laser desorption ionisation (MALDI) and electrospray ionisation (ESI).

4.2.4 MALDI

MALDI is an ionisation technique that involves the laser induced ionisation of a matrix consisting of crystallised molecules. The matrix absorbs laser light and vaporises into a hot plume of charged matrix species (Bellon et al., 2009). This charged matrix can interact with substrate molecules and ionise them by proton transfer or deprotonation.

4.2.5 ESI

ESI involves the spray of the liquid sample in solvent - usually acidified acetonitrile - for ionisation by protonation. The solvent with the chemical sample vaporises, leaving behind the protonated chemical compound species. The aerosol can be heated to increase the evaporation rate of solvent from the samples. There are also several adaptations of this technique such as the desorption ESI; which uses an electrically charged mist that is directed to the surface of the chemical sample; this charged mist ionises by desorbing and liberating the chemical substance from the surface and the ionised species are then detected (Wiseman & Laughlin, 2012). Similar to this is the direct analysis in real time (DART) ion source which ionises by means of firing ionised gas species at surfaces carrying chemical compounds of interest, molecules are desorbed and are detected (Hopfgartner, 2013).

4.2.6 Mass analysers

4.2.6.1 Sector mass analysers

This technique is based on separation of charged ions by mass when accelerated through a magnetic or electric field (Korfmacher, 2005). The ions are deflected in their trajectories and this deflection is proportional to the magnitude of the applied field and the voltage the ions were accelerated by. The angle of deflection reveals the charge state and mass of the ions. Both these are focussing ion instruments, magnetic sector techniques focus angular dispersions of ions while electric sectors focus ions in beams based on kinetic energies.

4.2.6.2 Time-of-flight mass analysers

This technique analyses ions without the use of electric of magnetic fields. The ions are separated based on their kinetic energies and velocities through a flight tube (Lacorte & Fernandez-Alba, 2006). The particles are ionised by an ion source and allowed to move through a flight tube, the time it takes for the ions to reach the detector is a representation of the ions mass and charge. Heavier ions take a longer time to reach the detector than lighter ions.

4.2.6.3 Ion cyclotron resonance analysers

This technique is a based on trapping ions by means of applying a high magnetic field since charged ions in a magnetic field will exhibit oscillating motions. The frequency of this orbit gives rise to the mass spectra following Fourier transform of the frequency data (Hopfgartner, 2013). Larger ions will orbit at a slower frequency than smaller ions. This is proportional to the ions mass to charge ratios and this is the basis of this type of analyser.

4.2.6.4 Orbitrap mass analysers

Being similar to ion cyclotron resonance, this is a "trap" type instrumentation that traps ions in orbit electrostatically. Oscillation occurs around a spindle shaped electrode with an axis. The frequency of ion orbits are related to their mass to charge ratios and these ion frequencies gives rise to a frequency spectrum or image current which is then subject to Fourier transform to yield ion spectra (Hopfgartner, 2013).

4.2.6.5 Quadrupole Mass Analyser

The principle of detection the technique uses is based on mass filtering of ions. Ions pass through rods at a set frequency range and ions are selected for based on their stable trajectories through the quadrupole rods. Ions with unstable trajectories collide with the rods and do not reach the detector. Trajectories are based on complex motion that is related to each ion mass (Hopfgartner, 2013). Therefore this is an efficient means of filtering and detecting differentially weighted ions. This type of mass analyser was selected for this study due to its application in metabolite identification, selection and quantitation.

4.2.7 Sample preparation

In order to reduce noise or background levels in mass spectrometry detection, sample preparation has to be employed to reduce complexity and separate compounds. Usually liquid chromatography is employed in the separation of compounds for reduced complexity. There are various methods of

sample preparation including both offline and online separations. Online sample preparation involves pre-preparation of the biological sample prior to separation by liquid chromatography (Kostiainen et al., 2003). Online techniques include column switching, using a separation or extraction column as well as an analytical column (Meyer & Maurer, 2012). Here, the biological sample is initially fractionated using the extraction column and eluted into a path directed to the analytical column. Following separation in the analytical column the molecules can enter the mass spectrometer for ionisation and detection.

Offline separation involves techniques such as solid-phase extraction, where extraction of metabolites occurs onto a solid support such as silica which is then washed and eluted in appropriate solvent (Kostiainen et al., 2003). Alternatively, the liquid partitioning extraction methods involve separation of compounds into aqueous and organic phases (Hopfgqrtner & Bourgogne, 2003). The protein precipitation methods reduce complexity by precipitating the protein out of solution which can be filtered or removed by sedimentation to reduce complexity (Kostiainen et al., 2003). These methods can all be automated for high throughput applications, especially since a reduction in sample complexity can result in an increase in sensitivity. For ease of use and wide applicability, the offline sample preparation was used in this study using both protein precipitation and liquid chromatography.

4.2.8 Metabolite identification

The identification of metabolites is used in cases where some prior knowledge about the metabolite exists from predicted or previously determined masses (Prakash, Shaffer, & Nedderman, 2007). The challenge of detection is then reduced to identifying significant peaks and spectra from background spectrum arising from other interfering biological compounds.

4.2.9 Full scan

In the absence of knowledge, a common technique to identify all metabolites is using full scan mass spectrometry whereby samples created by incubating a drug substrate with CYPs are analysed by full scanning in the mass spectrometer (Kostiainen et al., 2003). Sample peaks are compared to controls

and blanks and the molecular apparent changes in molecular masses are analysed for metabolism according to the typical routes of metabolism such as hydroxylation (+16 mass units), acetylation (+42 mass units) and so on (Table 4.1).

4.2.10 Tandem MS (MS/MS)

Another method for metabolite identification is the tandem mass spectrometry method. This technique is useful to identify metabolites based on their fragmentation pattern. Different metabolites will fragment differently depending on their route of metabolism. Different fragmentation patterns will lead to different MS/MS fragmentation spectra and this can be used to determine the identification of metabolites quantitatively (Lin, Pan, Mordenti, & Pan, 2007).

Table 4.1: Possible metabolic reactions and associated changes in molecular mass adapted from (Kostiainen et al., 2003)

Metabolic Reaction	Change in mass (u)
Demethylation	-14
Methylation	+14
Carbon hydroxylation	+16
N- hydroxylation	+16
N- oxidation	+16
Epoxidation	+16
Acetylation	+42

4.2.11 Metabolite quantitation

In order for mass spectrometry to be quantitative, various strategies can be employed. The initial method of quantitation is calibrating metabolites via a metabolite standard curve. This can be limiting since in the case of multiple metabolites, there might just be no commercially available standards. Spike in methods use a similar compound scaffold which would be essentially the same in all reaction conditions (Kostiainen et al., 2003). Quantitation is done by calibrating and using the absolute intensity or area under the curves of peaks and transitions generated in the mass spec. In the case of multiple metabolites, the formed metabolites can be compared to a percentage of parent drug or as a fold difference indicative of relative quantitation. Calibration using a standard curve method and area under the curves of intensity versus time in extracted ion chromatograms was selected for this study.

4.3 High protein CYP activity reconstitution

Lipid binding or immobilisation is ordinarily required to facilitate the interaction of CYP and CPR, but by simply increasing the amount of protein in a reaction the probability of CYP-CPR interactions in favourable states can be increased (Müller-Enoch et al., 1984). Using cost effective protein production in bacterial systems such as the *E. coli* expression system used in this thesis facilitates abundant protein production for this purpose. In order to validate the protein activity reconstitution, functional protein reactions were carried out on probe substrates for each of CYP3A4, CYP2C9 and CYP2D6. The probe substrates were selected as recommended by the US FDA due to their selectivity for the mentioned CYPs (Zhang et al., 2009). Furthermore, the metabolites in each case have been well documented using mass spectrometry. The strategy used in these experiments involved reconstituting CYP activities and testing the activity using the probe substrates. Upon validation of CYP activity reconstitution, the CYPs were analysed for activity on a complex drug, rifampicin (Rif). The presence of Rif metabolites was analysed and compared between "high protein" reconstituted activity and activity from immobilised protein in the form of human liver microsomes. The metabolites were identified using triple quadrupole mass spectrometry and quantified by calibration

against a standard curve. Areas under the curve of transition peaks in “multiple reaction monitoring” mode were calculated.

4.4 Results: Label-free mass spectrometry based quantitation

Table 4.2: CYP probe substrates and identified primary metabolites from “high protein” activity reconstitution

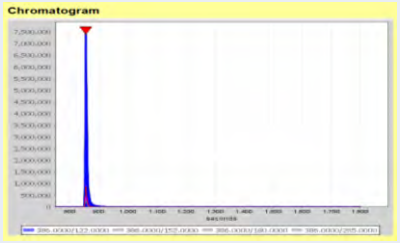
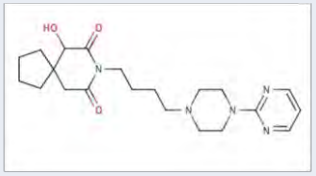
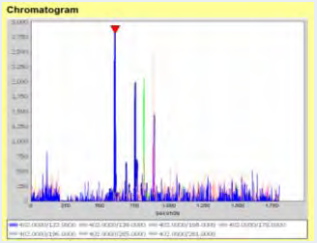
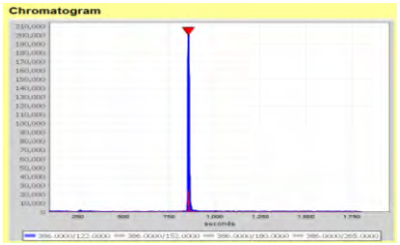
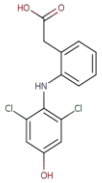
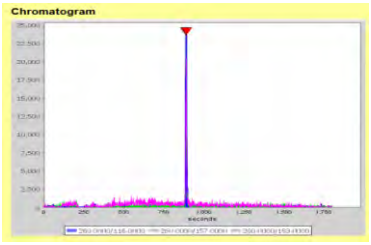
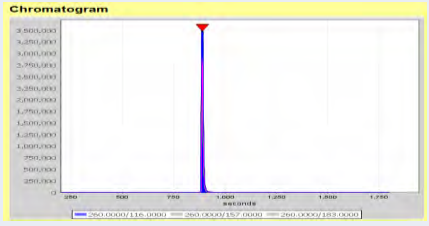
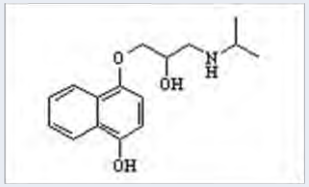
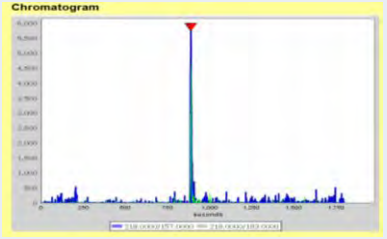
CYP/substrate	Parent chromatogram	Primary metabolite	Metabolite chromatogram
CYP3A4 Buspirone		 6-Hydroxy Buspirone	
CYP2C9 Diclofenac		 4-Hydroxy Diclofenac	
CYP2D6 Propanolol		 4-Hydroxy-propanolol	

Table 4.3: Identified Rifampicin metabolites in CYP incubations using expressed “high protein” reconstitution or human liver microsomes

HLM Rif	Rifampicin quinone	Demethyl-desacetylrifampicin	Desacetylrifampcin	Desacetyl-3-formylrifampicin	3-Formylrifampicin	Mono-oxygenated rifampicin-2	Acetylated desacetylrifampicin	N-Demethyl rifampicin	Mono-oxygenated desacetylrifampicin-2	Mono-oxygenated desacetylrifampicin-1	Mono-oxygenated rifampicin-1	Other	Other2
	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
RLM Rif		✓	✓			✓		✓		✓			
Rif 3A4 pure	✓	✓	✓	✓			✓						
Rif 3A4 Lys	✓	✓	✓			✓	✓			✓		Mono-oxygenated rifampicin-4	Mono-oxygenated rifampicin-6
Rif 2C9 pure	✓	✓	✓			✓		✓		✓			
Rif 2C9 Lys	✓	✓	✓	✓		✓	✓						Mono-oxygenated rifampicin-6
Rif 2D6 pure	✓	✓	✓	✓		✓	✓	✓					
Rif 2D6 Lys	✓	✓	✓		✓	✓		✓		✓	✓		
DH5α Rif	✓					✓							

HLM – Human Liver Microsomes

RLM – Rat Liver Microsomes

Rif – Rifampicin

Pure – Purified CYP protein

Lys – Crude lysate

The data from Table 4.3 indicate the presence of Rif metabolites in CYP reactions, with "high protein" conditions representing the incubations with a the relative concentration required, the HLM and RLM protein concentrations used were the routine recommended protein concentrations for comparison. Circled are the metabolites of particular interest, as they appear in incubations with crude lysate and not with purified CYP protein. This is counterintuitive as this purified protein reactions would be expected to produce the majority of the metabolites under the rationale discussed previously. It is however interesting, that these metabolites appear in CYP crude lysate incubations, while the presence of these metabolites are not present in the wild type E-coli crude lysate. This suggests that other bacterial proteins might be acting in conjunction with CYP proteins, possibly through a similar mechanism to CytB5. CytB5 is thought to potentially have both allosteric effects on CYP protein affecting catalysis, and potential redox activity itself (Pompon & Coon, 1984; Porter, 2002).

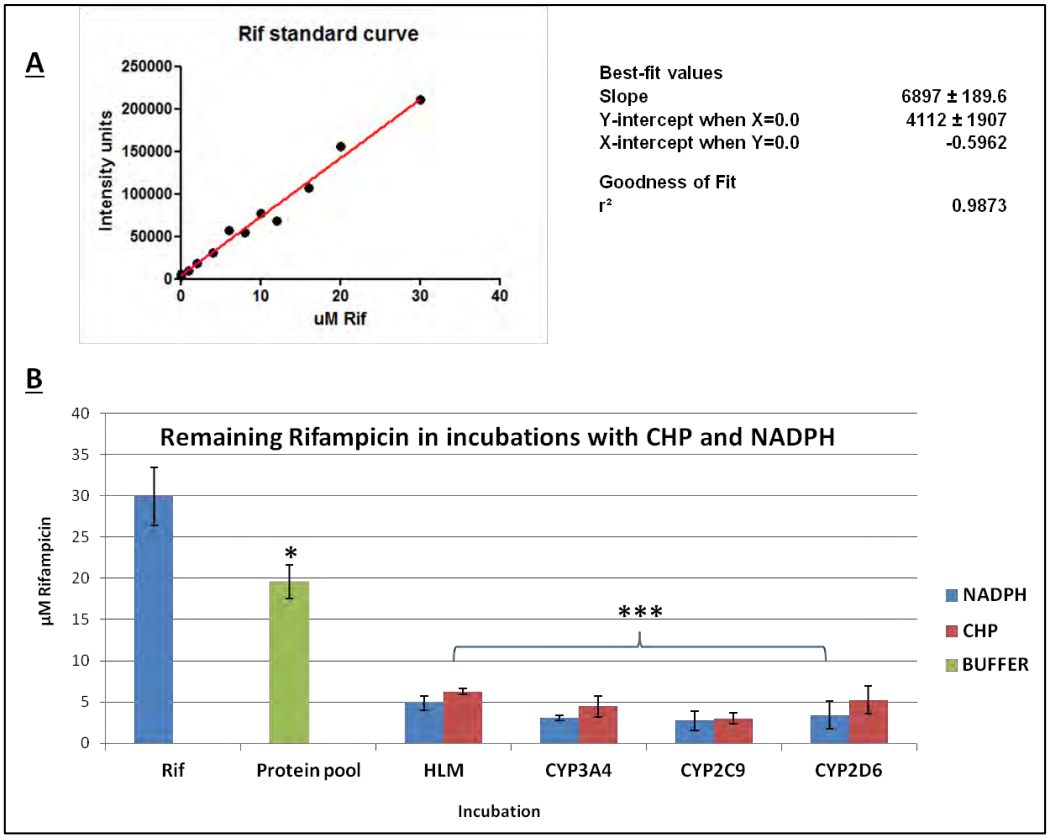


Figure 4.1: Data showing quantification of Rifampicin and comparison between incubations

Rif – Rifampicin incubation without CYP or HLM

Protein pool – Reactions with CYP3A4, CYP2C9 and CYP2D6 pooled together with HLM, without NADPH or CHP

HLM – Human Liver Microsomes

CHP – Cumene hydroperoxide as a control to indicate the relative Rif metabolism

The data from the above figure shows construction of a rifampicin standard curve and linearity of mass spectrometry based quantitation of serial dilutions of rifampicin in Panel A. The limit of detection was found to be 10pmol rifampicin and the limit of quantitation at 100nM rifampicin. Good correlations with a R^2 value of 0.9873 were achieved. Panel B shows the quantitation of rifampicin in each of the sample incubations and controls. The "Rif" bar indicates incubations with Rif and NADPH only, while the "Protein pool" bar indicates a buffer only control in which a pooled protein mix of HLMs and purified recombinant CYPs was incubated with rifampicin in the absence of both NADPH and CHP. One way ANOVA statistical analysis done which revealed that the means are

significantly different (P<0.001), and following a Bonferroni test for multiple comparisons the Rif bar was found to be significantly different to the protein pool (*, P<0.05) and to the HLM, CYP3A4, CYP2C9 and CYP2D6 groups (***, P<0.01).

4.5 Discussion and conclusion

This chapter focused on investigation of a high protein activity reconstitution strategy for CYP and CPR, using label-free detection of the metabolites of probe substrates and a complex drug rifampicin. A mass spectrometric detection strategy was employed using liquid chromatography coupled with electrospray ionisation with and triple quadrupole mass analyser. Tandem mass spectrometry was used to identify and quantify metabolites produced. Tandem mass spectrometry data was validated by comparison to published MS/MS spectra in each case.

Initially, CYP3A4 functional reconstitution was studied using buspirone as a probe substrate, monitoring formation of the major metabolite 6'-hydroxybuspirone (Table 4.2). This resultant data was characterised and validated using published MS/MS spectra of buspirone (Chew, Xu, Cordova, & Chow, 2006), which demonstrated that the expected metabolite is indeed produced by the high protein reconstitution of CYP3A4. The CYP2C9 isoform activity was investigated using diclofenac as probe substrate and 4'-hydroxydiclofenac was identified as the major metabolite (Table 4.2), again the resultant spectra were compared to published spectra of this metabolite (Yan et al., 2005). Finally, the functionality of the CYP2D6 isoform was investigated using propanolol as a probe substrate and 4'-hydroxypropanolol was identified as the major metabolite. This resultant data again agreed with previously published mass spectra for this major metabolite (Upthagrove & Nelson, 2001). Overall, the data on metabolism of probe substrates is consistent with the hypothesis that the functionality of each of the three CYPs was successfully reconstituted using the "high protein" strategy.

Further validation of this approach was carried out using the complex drug rifampicin, determining and quantifying metabolite formation using tandem mass spectrometry. There are approximately 20 known Rif metabolites (Prasad & Singh, 2009), the majority of which (Table 4.3) were identified in the current study. This is promising as these metabolites were generated using the high protein

reconstitution strategy and compared to metabolites produced by human and rat liver microsomes. Interestingly, not all the metabolites were identified, possibly due to there being isoform specific requirements of the other less abundant CYPs. However these "missing metabolites" also did not appear in the human liver microsome incubations so it is possible that these metabolites are low abundance and were below the limit of detection under the reaction conditions used here (Velde, Alffenaar, Wessels, Greijdanus, & Uges, 2009).

In order to accurately quantify the metabolites formed, a rifampicin standard curve was generated (Figure 4.1, Panel A). The standard curve yielded good linearity with an R^2 value close to 0.99. The standard curve was used to determine the overall metabolites formed judged by a decrease in parent rifampicin drug. The area under the curve of extracted chromatograms (intensity verses time eluted) for rifampicin (identity confirmed using tandem MS and multiple reaction monitoring) was converted to a concentration of rifampicin using the standard curve. The relative decline in remaining rifampicin was seen in the incubations of human liver microsomes and in each of the high protein reconstituted CYP incubations. This was compared to the rifampicin without protein (Figure 4.1, Panel B) as well as to a protein pool including all three recombinant CYPs as well as human liver microsomes but without the addition of NADPH – i.e. essentially "buffer only" conditions with drug and protein (Protein pool bar, Figure 4.1, Panel B). This analysis showed that there were statistically significantly different levels of remaining rifampicin between the Rif only control and each of the recombinant CYPs ($P<0.01$ using ANOVA with Bonferroni post comparison tests). There was also a significant difference between the rifampicin only cohort and the protein pool cohort ($P<0. 05$), which can be accounted for by residual levels of metabolism in the microsomal preparations until NADPH becomes exhausted. Comparing the activity levels within groups, there is no significant difference between CHP and NADPH conditions, suggesting that there is good reconstitution of CYP activity in this system.

This data provides evidence in support of the high protein reconstitution strategy and suggest that this approach can be useful for functional CYP assays since comparable metabolite production was observed compared to the gold standard for CYP functional assays (human liver microsomes). The

use of high amounts of protein (1-4μM CYP and CPR in a 1:1 ratio) has been previously shown to reconstitute CYP activity when applied to probe substrates but the demonstration of such reconstituted activity on a complex drug such as rifampicin, with characterisation of all metabolites formed is new (Müller-Enoch et al., 1984). These results are therefore promising for the development of drug metabolism screening platforms and further work is warranted to study the interaction of complex mixes, including enzyme inhibitors and co-incubations for the further validation and characterisation of such a technological approach.

4.6 Methods

4.6.1 Probe substrate incubations

Probe substrates (buspirone, diclofenac, propanolol and rifampicin) were obtained from Sigma as molecular biology grade reagents. The substrates were solubilised in DMSO as a stock solution and diluted into the incubation mix containing 100mM potassium phosphate buffer. NADPH was generated using a regeneration mix containing (0.5 mM β-nicotinamide adenine dinucleotide phosphate, 5mM glucose-6-phosphate, 0.5U glucose-6-phosphate dehydrogenase; all Sigma). Reactions were set up by adding either the human liver microsomes (20μg) (pooled human mixed gender, Xenotech) or rat liver microsomes (25μg) (male rat IGS, Xenotech) or expressed recombinant CYP and CPR protein (4μM) to potassium phosphate buffer and pre-incubating with drug or probe substrate for 2 mins at room temperature. Reaction was initiated with the addition of regenerating mix and incubated at room temperature for 1 hour.

For standard curve generation and identification of limits of detection and quantitation, a dilution series of rifampicin was made up from a stock solution into potassium phosphate buffer, ranging from atomolar to micromolar concentrations.

4.6.2 Sample preparation for mass spectrometry

Samples were quenched with the addition of 50% v/v ice-cold acetonitrile (HPLC grade, Sigma) to precipitate the protein. The sample was then centrifuged at 14000g for 15mins to sediment the precipitated protein and the supernatant was filtered through a sterile 10μm acetonitrile resistant nylon syringe filter (Agela Technologies) to remove any particulates. Samples were transferred into HPLC screw cap vials for storage at -20°C and later analysis.

4.6.3 Analysis

4.6.3.1 Liquid chromatography

Sample of 10µl was injected into the HPLC column (C18 Hydro Reverse phase HPLC column 150x4.6 mm, Synergi) using an autosampler (Accela Autosampler, Thermo) at a flow rate of 8ul/second. The HPLC mobile phase consisted solvent A (H_20 95%, 5% Formic acid (Acros) v/v) and solvent B (Acetonitrile (Fisher) 95%, Formic acid 5% v/v). The HPLC routine used was 95% Solvent A/5% Solvent B for 1 min followed by 85% Solvent A/15% Solvent B for 15min; a linear gradient was then run to 50% Solvent A/50% Solvent B over 10mins, and to 15% Solvent A/85% Solvent B over 3mins. The column was then washed back to 95% Solvent A/5% Solvent B from at a flow rate of 300µl/min over the gradient and washing.

4.6.3.2 Mass spectrometry

The triple quadrupole TSQ Vantage (Thermo Scientific) was used in MRM mode and operated in positive ion mode. The ion source was used in ESI mode with the capillary temperature at 350°C, spray voltage at 3000V, the sheath gas and auxiliary gas pressures at 35 and 10 Bar. Xcalibur software version 2.1.0.1139 (Thermo Scientific) was used for peak height integration and MRMer (McIntosh laboratory, (Martin et al., 2008)) was used for extracting ion chromatograms and area under the curve measurements. ReadW (IonSource) was used for file type conversions from propriety raw Xcalibur native files to mzXML open source file types.

4.6.4 Data analysis

Data for the standard curve was exported from MRMer to GraphPad Prism (Version 5, GraphPad Software). The data points were fitted using linear regression and trend line parameters were estimated using best fit models. For statistical analysis, data was imported into GraphPad Prism and ANOVA analysis was carried out for difference of means throughout the experiments; a significance

level of 0.05 was selected and Bonferroni post tests were performed for multiple comparisons within the experiment.

Chapter 5 – Conclusions and Future work

5.1 Conclusions

The aims of this thesis were to further develop functional CYP technology platforms that are adaptable to high throughput workflows. The initial goal was to adapt the functional protein microarray previously developed for CYP3A4 (Blackburn et al., 2012), to other human CYPs. Work was carried out using the other two most important CYPs for drug metabolism, namely CYP2C9 and CYP2D6. As a second goal, an alternate means of functional activity reconstitution was investigated, namely the "high protein" activity reconstitution technology, again for CYP3A4, CYP2C9 and CYP2D6 (Müller-Enoch et al., 1984).

Protein was expressed under low cost high volume protein expression conditions in a bacterial host, which also facilitated simple isolation and purification. Sufficient active protein was obtained using this technique and functional assays were performed. Both label-based and label-free detection and quantification of activity were investigated. The label-based detection made use of fluorogenic substrates that became highly fluorescent upon CYP metabolism and these were amenable to high throughput workflows (Trubetskoy et al., 2005). It was also shown that the limitation of simple two colour detection using microarray scanners can be overcome by using other techniques such as confocal detection. Furthermore, surface characterisation using AFM has revealed the lower than expected surface densities of immobilised streptavidin, which has implications for microarray-based reconstitution of CYP-CPR activity given that relatively large spacing between streptavidin tetramers. Interestingly, the measured kinetic data from reconstituted CYP and CPR from microarray experiments are comparable with the literature, which suggests that CYP activity is indeed occurring on the surface. Therefore, the relatively large spacing on the microarray surface is not limiting catalysis, implying that there are likely surface interactions between CYP and CPR in "cis" or on the same streptavidin tetramer, and/or the underlying polymer flexibility with the possibility of "coffee-ring" effects is facilitating transient CYP-CPR interaction between adjacent streptavidin tetramers.

The label-free approach was undertaken using probe substrates and a complex drug rifampicin with mass spectrometric detection and quantification. Both strategies of activity reconstitution were shown to yield functional CYP protein systems for all CYPs studied here, suggesting that they are likely both to prove generalisable to other CYPs, and can be amenable to high throughput workflows. The work described in this thesis has therefore moved forward the potential of both these reconstitution systems towards wider application in drug discovery applications and has therefore fulfilled the original goals of this study.

5.2 Future Work

Going forward, it can be seen that optimal CYP activity can be reconstituted in the absence of immobilisation by simply increasing protein concentrations. Using the low cost and minimal requirement bacterial protein expression system, sufficient protein can be obtained for high throughput assays using this type of workflow. Future work is envisioned to optimise this activity reconstitution strategy for miniaturisation and high throughput. Miniaturisation has the advantage of lower protein and reagent requirement as well as simpler, more reproducible automatable sample dispensing and assay workflows. Work is also envisioned for the analysis of inhibitor interactions in these systems. Inhibitor interactions may present differently in miniaturised conditions where evaporative and substrate concentrating effects can play a role in binding dynamics of inhibitors (Uttamchandani, Huang, Chen, & Yao, 2005). In both these systems the IVIVE extrapolations to whole body scaling factors will need to be determined once sufficient data is available. This would result in further validity of this type of drug metabolism assay for preclinical screening.

Investigating the individual CYP polymorphisms in either CYP assay technique explored in this thesis would be beneficial for the progress towards individualised medicine. Assaying the functional relevance of CYP polymorphisms and their correlations to groups of genetically distinct individuals would lead to a better understanding of how and why patient groups responds differently to drug therapy. Furthermore, responders could then be distinguished from non-responders for a particular

drug and potential adverse effects could be limited through tailoring treatment of disease to personal CYP metabotype for optimal pharmaceutical therapy (Ingelman-sundberg, 2004).

Other avenues opened by the work in this thesis include merging the label-free detection and the high concentration protein reconstitution with the microarray format for a truly miniaturised functional protein assay. The mass spectrometric detection would have to be adapted to a surface based format, such as desorption electrospray ionisation (DESI) or matrix-assisted laser desorption ionisation (MALDI) (Díaz-Mochón et al., 2007; Wiseman & Laughlin, 2012). Linking the solution phase high protein reconstitution strategy to microfluidic chips for miniaturisation and on-chip liquid chromatography sample preparation (Sung, Choi, Kim, & Shuler, 2009), could be advantageous to separate the ionisation matrix from the reaction chambers prior to mass spectrometry as a means required for surface ionisation strategies.

The array format itself can be further optimised using other surfaces or surface modifications, including 3 dimensional surface architectures for higher binding capacity and/or encapsulation in order to gain greater control over aqueous micro-environments and to reduce evaporation rates. In addition, protein production "on chip" in cell free expression systems might be cost effective when working on miniature scales such as that of the microarray spot (Chandra & Srivastava, 2010), although it remains to be seen whether active CYP and CPR protein would be produced by this route. There can also be further gains by investigating other microarray dispensing strategies, such as acoustic dispensing for faster more efficient workflows and to reduce time to detection (Ekins, Olechno, & Williams, 2013).

Going forward, these miniaturised assay types can be expanded to incorporate phase II drug metabolism conjugation reactions and even bacterial CYPs. Bacterial CYPs can play a role in the total human gut metabolism of orally administered drugs since the bacteria that colonise human guts can metabolise drugs and alter pharmacokinetics (Oz & Ebersole, 2008). Certain bacterial CYPs may even prove to be drug targets themselves, such as those found in *Mycobacterium tuberculosis* (Ouellet, Johnston, & Ortiz de Montellano, 2010).

In summary, this thesis provides evidence to support the further development and application of miniaturised functional protein assays that reconstitute CYP activity for CYP3A4, CYP2C9 and CYP2D6 in two different formats. Label-dependent fluorescent and label-free detection strategies were employed and drug metabolites were identified. Further developments of these functional reconstitution tools should therefore aid the utility and application of this innovative technology.

Bibliography

Afanassiev, V., Hanemann, V., & Wölfl, S. (2000). Preparation of DNA and protein micro arrays on glass slides coated with an agarose film. *Nucleic Acids Research*, *28*(12).

Ahmad, N., & Mukhtar, H. (2004). Cytochrome p450: a target for drug development for skin diseases. *The Journal of Investigative Dermatology*, *123*(3), 417–425.

Al-shurbaji, A., Bertilsson, L., Dahl, M.-L., & Dalen, P. (2002). Molecular genetics of CYP2D6 : Clinical relevance with focus on psychotropic drugs. *Journal of Clinical Pharmacology*, *53*, 111–122.

Anandatheerthavarada, H. K., Addya, S., Dwivedi, R. S., Biswas, G., Mullick, J., & Avadhani, N. G. (1997). Localization of Multiple Forms of Inducible Cytochromes P450 in Rat Liver Mitochondria : Immunological Characteristics and Patterns of Xenobiotic Substrate Metabolism. *Archives of Biochemistry and Biophysics*, *339*(1), 136–150.

Atkins, W. M. (2005). Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annual Review of Pharmacology and Toxicology*, *45*, 291–310.

Baron, J. M., Goh, L. B., Yao, D., Wolf, R. C., & Friedberg, T. (2001). Modulation of P450 CYP3A4-Dependent Metabolism by P-glycoprotein : Implications for P450 Phenotyping. *Pharmacology*, *296*(2), 351–358.

Barron, J. A., Young, H. D., Dlott, D. D., Darfler, M. M., Krizman, D. B., & Ringeisen, B. R. (2005). Printing of protein microarrays via a capillary-free fluid jetting mechanism. *Proteomics*, *5*(16), 4138–4144.

Baylon, J. L., Lenov, I. L., Sligar, S. G., & Tajkhorshid, E. (2013). Characterizing the Membrane-Bound State of Cytochrome P450 3A4: Structure, Depth of Insertion, and Orientation. *Journal of the American Chemical Society*, *135*(23), 8542–8551.

Beeton-Kempen, N. (2010). *P450 Biochips: Development of a protein microarray platform for investigating cytochrome P450 clinical drug metabolism*. University of Cape Town.

Bellon, S., Buchmann, W., Gonnet, F., Jarroux, N., Anger-Leroy, M., Guillonneau, F., & Daniel, R. (2009). Hyphenation of surface plasmon resonance imaging to matrix-assisted laser desorption ionization mass spectrometry by on-chip mass spectrometry and tandem mass spectrometry analysis. *Analytical Chemistry*, *81*(18), 7695–7702.

Benet, L. Z., Izumi, T., Zhang, Y., Silverman, J. A., & Wacher, V. J. (1999). Intestinal MDR transport proteins and P-450 enzymes as barriers to oral drug delivery. *Journal of Controlled Release*, *62*, 25–31.

Bertone, P., & Snyder, M. (2005a). Advances in functional protein microarray technology. *The FEBS Journal*, *272*(21), 5400–11.

Bertone, P., & Snyder, M. (2005b). Advances in functional protein microarray technology. *The FEBS Journal*, *272*(21), 5400–5411.

Black, S. D. (1992). Membrane topology of the mammalian P450 cytochromes. *The FASEB Journal*, *6*, 680–685.

Blackburn, J. M., Shoko, A., & Beeton-kempen, N. (2012). Miniaturized, Microarray-Based Assays for Chemical Proteomic Studies of Protein Function. In *Chemical Genomics and Proteomics: Reviews and Protocols* (Vol. 800, pp. 133–162).

Chandra, H., & Srivastava, S. (2010). Cell-free synthesis-based protein microarrays and their applications. *Proteomics*, *10*(4), 717–30.

Chen, X., Pan, L. Q., Naranmandura, H., Zeng, S., & Chen, S. Q. (2012). Influence of Various Polymorphic Variants of Cytochrome P450 Oxidoreductase (POR) on Drug Metabolic Activity of CYP3A4 and CYP2B6. *PloS One*, *7*(6), 38495–38502.

Chew, W. M., Xu, M., Cordova, C. A., & Chow, H.-H. S. (2006). Quantification of a cytochrome P450 3A4 substrate, buspirone, in human plasma by liquid chromatography – tandem mass spectrometry. *Journal of Chromatography B*, *844*, 235–239.

Chottard, G., Schappacher, M., Richard, L., & Weiss, R. (1984). Resonance Raman Spectra of Iron(II) Cytochrome P450 Model Complexes: Influence of the Thiolate Ligand. *Inorganic Chemistry*, *23*, 4557–4561.

Crewe, H. K., Barter, Z. E., Yeo, K. R., & Rostami-Hodjegan, A. (2011). Are there differences in the catalytic activity per unit enzyme of recombinantly expressed and human liver microsomal cytochrome P450 2C9? A systematic investigation into inter-system extrapolation factors. *Biopharmaceutics & Drug Disposition*, *32*, 303–318.

Cull, M. G., & Schatz, P. J. (2000). Biotinylation of Protein In Vivo and in Vitro Using Small Peptide Tags. *Methods in Enzymology*, *326*, 430–440.

Cusack, K. P., Koolman, H. F., Lange, U. E. W., Peltier, H. M., Piel, I., & Vasudevan, A. (2013). Emerging technologies for metabolite generation and structural diversification. *Bioorganic & Medicinal Chemistry Letters*.

Di, L., Kerns, E. H., Li, S. Q., & Carter, G. T. (2007). Comparison of cytochrome P450 inhibition assays for drug discovery using human liver microsomes with LC-MS, rhCYP450 isozymes with fluorescence, and double cocktail with LC-MS. *International Journal of Pharmaceutics*, *335*, 1–11.

Díaz-Mochón, J. J., Tourniaire, G., & Bradley, M. (2007). Microarray platforms for enzymatic and cell-based assays. *Chemical Society Reviews*, *36*(3), 449–457.

Dierks, E. A., Stams, K. R., Lim, H., Cornelius, G., Zhang, H., & Ball, S. E. (2001). A Method for the Simultaneous Evaluation of the Activities of Seven Major Human Drug-Metabolizing Cytochrome P450s Using an In Vitro Cocktail of Probe Substrates and Fast Gradient Liquid Chromatography Tandem Mass Spectrometry. *Pharmacology*, *29*(1), 23–29.

Dorgan, L., Magnotti, R., Hou, J., Engle, T., Ruley, K., & Shull, B. (1999). Methods to determine biotin-binding capacity of streptavidin-coated magnetic particles. *Journal of Magnetism and Magnetic Materials*, *194*, 69–75.

Doty, S. L., Shang, T. Q., Wilson, A. M., Tangen, J., Westergreen, A. D., Newman, L. A., Gordon, M. P. (2000). Enhanced metabolism of halogenated hydrocarbons in transgenic plants containing mammalian cytochrome P450 2E1. *Proceedings of the National Academy of Sciences*, *97*(12), 6287–6291.

Dugas, V., Broutin, J., & Souteyrand, E. (2005). Droplet evaporation study applied to DNA chip manufacturing. *Langmuir*, *21*, 9130–9136.

Dürr, U. H. N., Yamamoto, K., Im, S.-C., Waskell, L., & Ramamoorthy, A. (2007). Solid-state NMR reveals structural and dynamical properties of a membrane-anchored electron-carrier protein, cytochrome b5. *Journal of the American Chemical Society*, *129*(21), 6670–6671.

Edwards, R. J., Murray, B. P., Singleton, A. M., & Boobis, A. R. (1991). Orientation of Cytochromes P450 in the Endoplasmic Reticulum. *Biochemistry*, *30*(1), 71–76.

Ekins, S., Olechno, J., & Williams, A. J. (2013). Dispensing processes impact apparent biological activity as determined by computational and statistical analyses. *PloS One*, *8*(5).

Esch, M. B., King, T. L., & Shuler, M. L. (2011). The role of body-on-a-chip devices in drug and toxicity studies. *Annual Review of Biomedical Engineering*, *13*, 55–72.

Evans, W. E., Evans, W. E., & Relling, M. V. (1999). Pharmacogenomics : Translating Functional Genomics into Rational Therapeutics. *Science*, *286*, 487–491.

Evans, W. E., & Relling, M. V. (1999). Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics. *Science*, *286*, 487–491.

Gentzt, R., & Bujard, H. (1985). Promoters Recognized by Escherichia coli RNA Polymerase Selected by Function : Highly Efficient Promoters from Bacteriophage T5. *Journal of Bacteriology*, *164*(1), 70–77.

Goetz, M. P., Knox, S. K., Suman, V. J., Rae, J. M., Safgren, S. L., Ames, M. M., Ingle, J. N. (2007). The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Research and Treatment*, *101*(1), 113–121.

Goodwin, B., Redinbo, M. R., & Kliewer, S. A. (2002). Regulation of CYP3A Gene Transcription by the Pregnane X Receptor. *Annual Review of Pharmacology and Toxicology*, *42*, 1–23.

Grinkova, Y. V, Denisov, I. G., & Sligar, S. G. (2010). Functional reconstitution of monomeric CYP3A4 with multiple cytochrome P450 reductase molecules in Nanodiscs. *Biochemical and Biophysical Research Communications*, *398*, 194–198.

Groenhof, A. R., Ehlers, A. W., & Lammertsma, K. (2008). Alkane Hydroxylation by Peroxy Acids: A Comparison with the Cytochrome P450 Hydroxylation. *Journal of Physical Chemistry*, *112*, 12855–12861.

Guengerich, F. P. (1997). Comparisons of catalytic selectivity of cytochrome P450 subfamily enzymes from different species. *Chemico-Biological Interactions*, *106*, 161–182.

Guengerich, F. P. (1999). Cytochrome P-450 3A4: Regulation and Role in Drug Metabolism. *Annual Review of Pharmacology and Toxicology*, *39*, 1–17.

Guengerich, F. P. (2001). Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chemical Research in Toxicology*, *14*(6), 612–650.

Guengerich, F. P. (2003). Cytochromes P450, Drugs and Diseases. *Molecular Interventions*, *3*(4), 194–204.

Guengerich, F. P., Martin, M. V, Sohl, C. D., & Cheng, Q. (2009). Measurement of cytochrome P450 and NADPH – cytochrome P450 reductase. *Nature Protocols*, *4*(9), 1245–1251.

Guengerich, F. P., & Munro, A. W. (2013). Unusual Cytochrome P450 Enzymes and Reactions. *The Journal of Biological Chemistry*, *288*(24), 17065–17073.

Gunaratna, C. (2000). Drug Metabolism & Pharmacokinetics in Drug Discovery: A Primer for Bioanalytical Chemists , Part I. *Current Separations*, *19*(1), 17–23.

Gut, J., Richter, C., Cherry, R. J., Winterhalter, K. H., & Kawato, S. (1982). Rotation of cytochrome P-450. II. Specific interactions of cytochrome P-450 with NADPH-cytochrome P-450 reductase in phospholipid vesicles. *Journal of Biological Chemistry*, *257*(12), 7030–7036.

Haining, R. L., Hunter, A. P., Veronese, M. E., Trager, W. F., & Rettie, A. E. (1996). Allelic Variants of Human Cytochrome P450 2C9: Baculovirus-Mediated Expression, Purification, Structural Characterization, Substrate Stereoselectivity and Prochiral Selectivity of the Wild-Type and I359L Mutant Forms. *Archives of Biochemistry and Biophysics*, *333*(2), 447–458.

Hardiman, G. (2006). Applications of Microarrays and Biochips in Pharmacogenomics. In Q. Yan (Ed.), *Methods in Molecular Biology* (Vol. 448, pp. 21–30).

Hentschel, C., Wagner, H., Smiatek, J., Heuer, A., Fuchs, H., Zhang, X., Chi, L. (2013). AFM-based force spectroscopy on polystyrene brushes: Effect of brush thickness on protein adsorption. *Langmuir*, *29*(6), 1850–1856.

Hollenberg, P. F. (1992). Mechanisms of cytochrome P450 and peroxidase-catalysed xenobiotic metabolism. *The FASEB Journal*, *6*, 686–694.

Hollenberg, P. F., & Hager, L. P. (1973). The P-450 Nature of the Carbon Monoxide Complex of Ferrous Chloroperoxidase. *Journal of Biological Chemistry*, *248*(7), 2630–2633.

Holmberg, A., Blomstergren, A., Nord, O., Lukacs, M., Lundeberg, J., & Uhlén, M. (2005). The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis*, *26*(3), 501–10.

Honkakoski, P., & Negishi, M. (2000). Regulation of cytochrome P450 (CYP) genes by nuclear receptors. *Biochemistry*, *347*, 321–337.

Hood, L., & Perlmutter, R. M. (2004). The impact of systems approaches on biological problems in drug discovery. *Nature Biotechnology*, *22*(10), 1215–1217.

Hook, A. L., Anderson, D. G., Langer, R., Williams, P., Davies, M. C., & Alexander, M. R. (2010). High throughput methods applied in biomaterial development and discovery. *Biomaterials*, *31*(2), 187–98.

Hopfgartner, G. (2013). Overview of the Various Types of Mass Spectrometers that are used in Drug Discovery and Drug Development. *Mass Spectrometry for Drug Discovery and Development*.

Hopfgqrtner, G., & Bourgogne, E. (2003). Quantitative High-Throughput Analysis of Drugs in Biological Matrices by Mass Spectrometry. *Mass Spectrometry Reviews*, *22*, 195–214.

Houston, J. B., & Galetin, A. (2005). Modelling atypical CYP3A4 kinetics: principles and pragmatism. *Archives of Biochemistry and Biophysics*, *433*(2), 351–60.

Houston, J. B., & Kenworthy, K. E. (2000). In Vitro-In Vivo scaling of CYP Kinetic Data Not Consistent with the Classical Michaelis Menton Model. *Drug Metabolism and Disposition*, *28*(3), 246–254.

Hutzler, J. M., & Tracy, T. S. (2002). Atypical kinetic profiles in drug metabolism reactions. *Drug Metabolism and Disposition*, *30*(4), 355–362.

Ingelman-sundberg, M. (2004). Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends in Pharmacological Sciences*, *25*(4), 193–200.

Ingelman-sundberg, M. (2005). Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics*, *5*, 6–13.

Johnson, D. E. (2012). Predicting Drug Safety: Next Generation Solutions. *Journal of Drug Metabolism & Toxicology*, *3*(2), 106–110.

Johnson, E. F., & Stout, C. D. (2013). Structural diversity of eukaryotic membrane cytochrome p450s. *The Journal of Biological Chemistry*, *288*(24), 17082–17090.

Johnston, W. A., Hunter, D. J. B., Noble, C. J., Hanson, G. R., Stok, J. E., Hayes, M. A., Gillam, E. M. J. (2011). Cytochrome P450 is present in both ferrous and ferric forms in the resting state within intact Escherichia coli and hepatocytes. *The Journal of Biological Chemistry*, *286*(47), 40750–40759.

Kimura, Y., Ito, H., Ohnishi, R., & Hatano, T. (2010). Inhibitory effects of polyphenols on human cytochrome P450 3A4 and 2C9 activity. *Food and Chemical Toxicology*, *48*(1), 429–435.

Korfmacher, W. A. (2005). Principles and applications of LC – MS in new drug discovery. *Drug Discovery Today*, *10*(20), 1357–1367.

Kostiainen, R., Kotiaho, T., Kuuranne, T., & Auriola, S. (2003). Liquid chromatography/atmospheric pressure ionization – mass spectrometry in drug metabolism studies. *Journal of Mass Spectrometry*, *38*, 357–372.

Krest, C. M., Onderko, E. L., Yosca, T. H., Calixto, J. C., Karp, R. F., Livada, J., Green, M. T. (2013). Reactive intermediates in cytochrome p450 catalysis. *The Journal of Biological Chemistry*, *288*(24), 17074–17081.

Kusnezow, W., & Hoheisel, J. D. (2003). Solid supports for microarray immunoassays. *Journal of Molecular Recognition*, *16*, 165–176.

Lacorte, S., & Fernandez-Alba, A. R. (2006). Time of Flight Mass Spectrometry Applied to the Liquid Chromatographic Analysis of Pesticides in Water and Food. *Mass Spectrometry Reviews*, *25*, 866– 880.

Lamb, D. C., Lei, L., Warrilow, A. G. S., Lepesheva, G. I., Mullins, J. G. L., Waterman, M. R., & Kelly, S. L. (2009). The First Virally Encoded Cytochrome P450. *Journal of Virology*, *83*(16), 8266–8269.

Laursen, T., Jensen, K., & Møller, B. L. (2011). Conformational changes of the NADPH-dependent cytochrome P450 reductase in the course of electron transfer to cytochromes P450. *Biochimica et Biophysica Acta*, *1814*(1), 132–8.

Lee, C. A., Kadwell, S. K., Kost, T. A., & Serabjit-Singh, C. J. (1995). CYP3A4 Expressed by Insect Cells Infected with a Recombinant Baculovirus Containing Both CYP3A4 and Human NADPH-Cytochrome P450 Reductase Is Catalytically Similar to Human Liver Microsomal CYP3A4. *Archives of Biochemistry and Biophysics*, *319*(1), 157–167.

Lee, Y., Lee, E. K., Cho, Y. W., Matsui, T., Kang, I.-C., Kim, T.-S., & Han, M. H. (2003). ProteoChip: a highly sensitive protein microarray prepared by a novel method of protein immobilization for application of protein-protein interaction studies. *Proteomics*, *3*(12), 2289–2304.

Lewis, D. F. V. (2000). On the Recognition of Mammalian Microsomal Cytochrome P450 Substrates and Their Characteristics. *Biochemical Pharmacology*, *60*, 293–306.

Liang, J. F., Li, Y. T., & Yang, V. C. (2000). Biomedical Application of Immobilized Enzymes. *Journal of Pharmaceutical Sciences*, *89*(8), 979–990.

Lin, T., Pan, K., Mordenti, J., & Pan, L. (2007). In Vitro Assessment of Cytochrome P450 Inhibition: Strategies for Increasing LC/MS-Based Assay Throughput Using a One-Point IC50 Method and Multiplexing High-Performance Liquid Chromatography. *Journal of Pharmaceutical Sciences*, *96*(9), 2485–2493.

Lu, H.-Y., Qiu, L.-L., Yang, X.-J., Zhang, X.-M., Zhang, Z., & Wang, S.-L. (2013). Optimization of heme precursors for the expression of human cytochrome P450 2A13 and its co-expression with oxidoreductase in baculovirus/sf9 system. *Journal of Biochemistry*, *153*(6), 555–563.

Makings, L. R., & Zlokarnik, G. (1999). Optical Molecular Sensors for Cytochrome P450 Activity. United States: European Patent Office.

Manoj, K. M., Baburaj, A., Ephraim, B., Pappachan, F., Maviliparambathu, P. P., Vijayan, U. K., Mathew, L. T. (2010). Explaining the atypical reaction profiles of heme enzymes with a novel mechanistic hypothesis and kinetic treatment. *PloS One*, *5*(5).

Mansuy, D. (2007). A brief history of the contribution of metalloporphyrin models to cytochrome P450 chemistry and oxidation catalysis. *Comptes Rendus Chimie*, *10*, 392–413.

Martin, D. B., Holzman, T., May, D., Peterson, A., Eastham, A., Eng, J., & McIntosh, M. (2008). MRMer, an interactive open source and cross-platform system for data extraction and visualization of multiple reaction monitoring experiments. *Molecular & Cellular Proteomics*, *7*(11), 2270–2278.

Materne, E.-M., Tonevitsky, A. G., & Marx, U. (2013). Chip-based liver equivalents for toxicity testing - organotypicalness versus cost-efficient high throughput. *Lab on a Chip*.

McQuain, M. K., Seale, K., Peek, J., Levy, S., & Haselton, F. R. (2003). Effects of relative humidity and buffer additives on the contact printing of microarrays by quill pins. *Analytical Biochemistry*, *320*(2), 281–291.

Meyer, M. R., & Maurer, H. H. (2012). Current applications of high-resolution mass spectrometry in drug metabolism studies. *Analytical and Bioanalytical Chemistry*, *403*, 1221–1231.

Michaud, V., Ogburn, E., Thong, N., Aregbe, A. O., Quigg, T. C., Flockhart, D. A., & Desta, Z. (2012). Induction of CYP2C19 and CYP3A activity following repeated administration of efavirenz in healthy volunteers. *Clinical Pharmacology and Therapeutics*, *91*(3), 475–482.

Miksys, S., & Tyndale, R. F. (2013). Cytochrome P450-mediated drug metabolism in the brain. *Journal of Psychiatry & Neuroscience*, *38*(3), 152–163.

Monier, S., Luc, P. Van, Kreibich, G., Sabatini, D. D., & Kaplan, T. (1988). Signals for the Incorporation and Orientation of Cytochrome P450 in the Endoplasmic Reticulum Membrane. *Journal of Cell Biology*, *107*, 457–470.

Müller-Enoch, D., Churchill, P., Fleischer, S., & Guengerich, F. P. (1984). Interaction of liver microsomal cytochrome P-450 and NADPH-cytochrome P-450 reductase in the presence and absence of lipid. *The Journal of Biological Chemistry*, *259*(13), 8174–8182.

Munro, A. W., Girvan, H. M., Mason, A. E., Dunford, A. J., & McLean, K. J. (2013). What makes a P450 tick? *Trends in Biochemical Sciences*, *38*(3), 140–150.

Myasoedova, K. N., Arutyunyan, A. M., & Magretova, N. N. (2006). Phenomenon of activation of cytochrome P450 by nonionic detergents. *Bioscience Reports*, *26*(1), 69–78.

Neal, M. J. (2002). *Medical Pharmacology at a Glance* (Fourth.). London: Blackwell Science.

Nelson, D. R., & Strobel, H. W. (1988). On the Membrane Topology of Vertebrate Cytochrome P-450 Proteins. *Journal of Biological Chemistry*, *263*(13), 6038–6050.

Nelson, D. R., Zeldin, D. C., Hoffman, S. M. G., Maltais, L. J., Wain, H. M., & Nebert, D. W. (2004). Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics*, *14*, 1–18.

Neve, E. P. A., & Ingelman-sundberg, M. (2008). Intracellular transport and localization of microsomal cytochrome P450. *Analytical and Bioanalytical Chemistry*, *392*, 1075–1084.

Nicholson, J. K., Holmes, E., Lindon, J. C., & Wilson, I. D. (2004). The challenges of modeling mammalian biocomplexity. *Nature Biotechnology*, *22*(10), 1268–1274.

Omura, T. (2006). Mitochondrial P450s. *Chemico-Biological Interactions*, *163*, 86–93.

Oppermann, M., Damoc, N. E., Crone, C., Moehring, T., Muenster, H., & Hornshaw, M. (2012). High Precision Measurement and Fragmentation Analysis for Metabolite Identification. *Methods in Molecular Biology*, *860*, 145–156.

Ouellet, H., Johnston, J. B., & Ortiz de Montellano, P. R. (2010). The Mycobacterium tuberculosis cytochrome P450 system. *Archives of Biochemistry and Biophysics*, *493*, 82–95.

Oz, H. S., & Ebersole, J. L. (2008). Application of prodrugs to inflammatory diseases of the gut. *Molecules*, *13*(2), 452–74.

Page, C. C., Moser, C. C., Chen, X., & Dutton, P. L. (1999). Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature*, *402*, 47–52.

Patterson, T. A., Lobenhofer, E. K., Fulmer-Smentek, S. B., Collins, P. J., Chu, T.-M., Bao, W., Wolfinger, R. D. (2006). Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nature Biotechnology*, *24*(9), 1140–50.

Pillai, V. C., Strom, S. C., Caritis, S. N., & Venkataramanan, R. (2012). A sensitive and specific CYP cocktail assay for the simultaneous assessment of human cytochrome P450 activities in Primary cultures of Human hepatocytes using LC-MS/MS. *Journal of Pharmaceutical and Biomedical Analysis*, *12*.

Pompon, D., & Coon, J. (1984). On the Mechanism of Action of Cytochrome P-450, oxidation and reduction of the ferrous dioxygen complex of liver microsomal cytochrome P-450 by cytochrome b5. *Journal of Biological Chemistry*, *259*, 15377–15385.

Porter, T. D. (2002). The Roles of Cytochrome b5 in Cytochrome P450 Reactions. *Journal of Biochemistry and Molecular Toxicology*, *16*(6), 311–316.

Prakash, C., Shaffer, C. L., & Nedderman, A. (2007). Analytical Strategies for Identifying Drug Metabolites. *Mass Spectrometry Reviews*, *26*, 340– 369.

Prasad, B., & Singh, S. (2009). In vitro and in vivo investigation of metabolic fate of rifampicin using an optimized sample preparation approach and modern tools of liquid chromatography-mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis*, *50*(3), 475–490.

Quattrochi, L. C., & Guzelian, P. S. (2001). CYP3A Regulation: From Pharmacology to Nuclear receptors. *Pharmacology*, *29*(5), 615–622.

Renaud, J. P., Cullin, C., Pompon, D., Beaune, P., & Mansuy, D. (1990). Expression of human liver cytochrome P450 IIIA4 in yeast, A functional model for the hepatic enzyme. *European Journal of Biochemistry*, *194*(3), 889–896.

Rendic, S. (2002). Summary of Information on Human CYP Enzymes: Human P450 Metabolism Data. *Drug Metabolism Reviews*, *34*, 83–448.

Rittle, J., & Green, M. T. (2010). Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science*, *330*, 933–937.

Rittle, J., Younker, J. M., & Green, M. T. (2010). Cytochrome P450: the active oxidant and its spectrum. *Inorganic Chemistry*, *49*(8), 3610–3617.

Rodrigues, A. D. (1997). Preclinical Drug Metabolism in the Age of High-Throughput Screening: An Industrial Perspective. *Pharmaceutical Research*, *14*(11), 1504–1510.

Rodríguez-Fragoso, L., & Reyes-Esparza, J. (2013). *Fruit/Vegetable-Drug Interactions : Effects on Drug Metabolizing Enzymes and Drug Transporters* (pp. 1–34). InTech Drug Discovery.

Rogers, M. J., Strittmatter, P., & Spatz, L. (1972). The binding of Cytochrome b5 to Liver Microsomes. *Journal of Biological Chemistry*, *247*(22), 7188–7194.

Rowland, P., Blaney, F. E., Smyth, M. G., Jones, J. J., Leydon, V. R., Oxbrow, A. K., Bridges, A. M. (2006). Crystal structure of human cytochrome P450 2D6. *The Journal of Biological Chemistry*, *281*(11), 7614–7622.

Sato, R., Sakaguchi, M., & Mihara, K. (1987). A short amino-terminal segment of microsomal cytochrome P-450 functions both as an insertion signal and as a stop-transfer. *EMBO Journal*, *6*(8), 2425–2431.

Schaefer, O., Ohtsuki, S., Kawakami, H., Inoue, T., Liehner, S., Saito, A., Ebner, T. (2012). Absolute Quantification and Differential Expression of Drug Transporters, Cytochrome P450 Enzymes, and UDP- Glucuronosyltransferases in Cultured Primary Human Hepatocytes. *Drug Metabolism and Disposition*, *40*, 93–103.

Schneider, E., & Clark, D. S. (2013). Cytochrome P450 (CYP) enzymes and the development of CYP biosensors. *Biosensors & Bioelectronics*, *39*, 1–13.

Seliskar, M., & Rozman, D. (2007). Mammalian cytochromes P450-Importance of tissue specificity. *Biochemical and Biophysical Research Communications*, *1770*, 458 – 466.

Shou, M., Lin, Y., Lu, P., Tang, C., Mei, Q., Cui, D., Rushmore, T. H. (2000). Enzyme kinetics of cytochrome P450-mediated reactions. *Current Drug Metabolism*, *2*, 17–36.

Shu, Y.-Z., Johnson, B. M., & Yang, T. J. (2008). Role of biotransformation studies in minimizing metabolism-related liabilities in drug discovery. *The AAPS Journal*, *10*, 178–192.

Shumyantseva, V. V, Bulko, T. V, & Archakov, A. I. (2005). Electrochemical reduction of cytochrome P450 as an approach to the construction of biosensors and bioreactors. *Journal of Inorganic Biochemistry*, *99*, 1051–1063.

Shumyantseva, V. V, Ivanov, Y. D., Bistolas, N., Scheller, F. W., Archakov, A. I., & Wollenberger, U. (2004). Direct Electron Transfer of Cytochrome P450 2B4 at Electrodes Modified with Nonionic Detergent and Colloidal Clay Nanoparticles. *Analytical Chemistry*, *76*(20), 6046–6052.

Sjögren, E., Svanberg, P., & Kanebratt, K. P. (2012). Optimized experimental design for the estimation of enzyme kinetic parameters: an experimental evaluation. *Pharmacology*, 112–145.

Spector, A. A. (2009). Arachidonic acid cytochrome P450 epoxygenase pathway. *Journal of Lipid Research*, *450*, s52–s56.

Stamos, B., Loredo, L., Chand, S., Phan, T. V, Zhang, Y., Mohapatra, S., Perera, R. (2012). Biosynthetic approach for functional protein microarrays. *Analytical Biochemistry*, *424*(2), 114–123.

Sukhishvili, S. A., Chen, Y., Mu, J. D., Gratton, E., Schweizer, K. S., & Granick, S. (2002). Surface Diffusion of Poly (ethylene glycol). *Macromolecules*, *35*, 1776–1784.

Sun, Y., Zeng, W., Benabbas, A., Ye, X., Denisov, I., Sligar, S. G., Champion, P. M. (2013). Investigations of heme ligation and ligand switching in cytochromes p450 and p420. *Biochemistry*, *52*(34), 5941–5951.

Sung, J. H., Choi, J., Kim, D., & Shuler, M. L. (2009). Fluorescence optical detection in situ for real-time monitoring of cytochrome P450 enzymatic activity of liver cells in multiple microfluidic devices. *Biotechnology and Bioengineering*, *104*(3), 516–525.

Thomsen, A. E., Christensen, M. S., Bagger, M. A., & Steffansen, B. (2004). Acyclovir prodrug for the intestinal di/tri-peptide transporter PEPT1: comparison of in vivo bioavailability in rats and transport in Caco-2 cells. *European Journal of Pharmaceutical Sciences*, *23*, 319–325.

Toh, Y.-C., Lim, T. C., Tai, D., Xiao, G., van Noort, D., & Yu, H. (2009). A microfluidic 3D hepatocyte chip for drug toxicity testing. *Lab on a Chip*, *9*(14), 2026–2035.

Trubetskoy, O. V, Gibson, J. R., & Marks, B. D. (2005). Highly miniaturized formats for in vitro drug metabolism assays using vivid fluorescent substrates and recombinant human cytochrome P450 enzymes. *Journal of Biomolecular Screening*, *10*(1), 56–66.

Uetrecht, J. (2003). Screening for the potential of a drug candidate to cause idiosyncratic drug reactions. *Drug Discovery Today*, *8*(18), 832–837.

Upthagrove, A. L., & Nelson, W. L. (2001). Importance of amine pKa and distribution coefficient in the metabolism of fluorinated propranolol derivatives. Preparation, identification of metabolite regioisomers, and metabolism by CYP2D6. *Drug Metabolism and Disposition*, *29*(11), 1377–1388.

Urlacher, V. B., & Girhard, M. (2012). Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends in Biotechnology*, *30*, 26–36.

Uttamchandani, M., Huang, X., Chen, G. Y. J., & Yao, S. Q. (2005). Nanodroplet profiling of enzymatic activities in a microarray. *Bioorganic & Medicinal Chemistry Letters*, *15*(8), 2135–2139.

Vagenende, V., Yap, M. G. S., & Trout, B. L. (2009). Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry*, *48*, 11084–11096.

Van der Hoeven, T. A., & Coon, M. J. (1974). Preparation and Properties of Partially Purified Cytochrome P-450 and Reduced Nicotinamide Adenine Dinucleotide P-450 Reductase from Rabbit Liver Microsomes. *Journal of Biological Chemistry*, *19*, 6302–6310.

Vanhulle, V. P., Martiat, G. A., Verbeeck, R. K., Horsmans, Y., Calderon, P. B., Eeckhoudt, S. L., Delzenne, N. (2001). Cryopreservation of rat precision-cut liver slices by ultrarapid freezing: influence on phase I and II metabolism and on cell viability upon incubation for 24 hours. *Life Sciences*, *68*, 2391–2403.

Velde, F. De, Alffenaar, J. C., Wessels, A. M. A., Greijdanus, B., & Uges, D. R. A. (2009). Simultaneous determination of clarithromycin , rifampicin and their main metabolites in human plasma by liquid chromatography – tandem mass spectrometry. *Journal of Chromatography B*, *877*, 1771–1777.

Voznesensky, A. I., & Schenkman, J. B. (1994). Quantitative analyses of electrostatic interactions between NADPH-cytochrome P450 reductase and cytochrome P450 enzymes. *The Journal of Biological Chemistry*, *269*(22), 15724–15731.

Walter, J.-G., Stahl, F., Reck, M., Praulich, I., Nataf, Y., Hollas, M., Scheper, T. (2010). Protein microarrays: Reduced autofluorescence and improved LOD. *Engineering in Life Sciences*, *10*(2), 103–108.

Want, E. J., Cravatt, B. F., & Siuzdak, G. (2005). The expanding role of mass spectrometry in metabolite profiling and characterization. *Chembiochem*, *6*(11), 1941–1951.

Wester, M. R., Yano, J. K., Schoch, G. a, Yang, C., Griffin, K. J., Stout, C. D., & Johnson, E. F. (2004). The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *The Journal of Biological Chemistry*, *279*(34), 35630–35637.

Williams, P. A., Cosme, J., Vinkovic, D. M., Ward, A., Angove, H. C., Day, P. J., Jhoti, H. (2004). Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science*, *305*, 683–686.

Williams, P. A., Cosme, J., Ward, A., Angove, H. C., & Jhoti, H. (2003). Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature*, *424*, 464–468.

Wiseman, J. M., & Laughlin, B. C. (2012). Desorption Electrospray Ionization (DESI) Mass Spectrometry: A brief introduction and overview. *Current Separations*, 11–14.

Xia, C., Panda, S. P., Marohnic, C. C., Martásek, P., Masters, B. S., & Kim, J.-J. P. (2011). Structural basis for human NADPH-cytochrome P450 oxidoreductase deficiency. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13486–13491.

Xu, W., Kong, J. S., Yeh, Y.-T. E., & Chen, P. (2008). Single-molecule nanocatalysis reveals heterogeneous reaction pathways and catalytic dynamics. *Nature Materials*, *7*(12), 992–996.

Yale, S. H., & Glurich, I. (2005). Analysis of the Inhibitory Potential of Ginkgo biloba, Echinacea purpurea, and Serenoa repens on the Metabolic Activity of Cytochrome P450 3A4, 2D6, and 2C9. *Journal of Alternate and Complementary Medicine*, *11*(3), 433–439.

Yamamoto, K., Gildenberg, M., Ahuja, S., Im, S.-C., Pearcy, P., Waskell, L., & Ramamoorthy, A. (2013). Probing the transmembrane structure and topology of microsomal cytochrome-p450 by solid-state NMR on temperature-resistant bicelles. *Scientific Reports*, *3*, 2551–2556.

Yan, Z., Li, J., Huebert, N., Caldwell, G. W., Du, Y., & Zhong, H. (2005). Dection of a novel reactive metabolite of Diclofenac: Evidence for CYP2C9-mediated bioactivation via arene oxides. *Drug Metabolism and Disposition*, *33*(6), 706–713.

Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D., & Johnson, E. F. (2004). The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. *The Journal of Biological Chemistry*, *279*(37), 38091–38094.

Yim, S., Yun, C., Ahn, T., Jung, H., & Pan, J. (2005). A Continuous Spectrophotometric Assay for NADPH-cytochrome P450 Reductase Activity Using 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium Bromide. *Journal of Biochemistry and Molecular Biology*, *38*(3), 366–369.

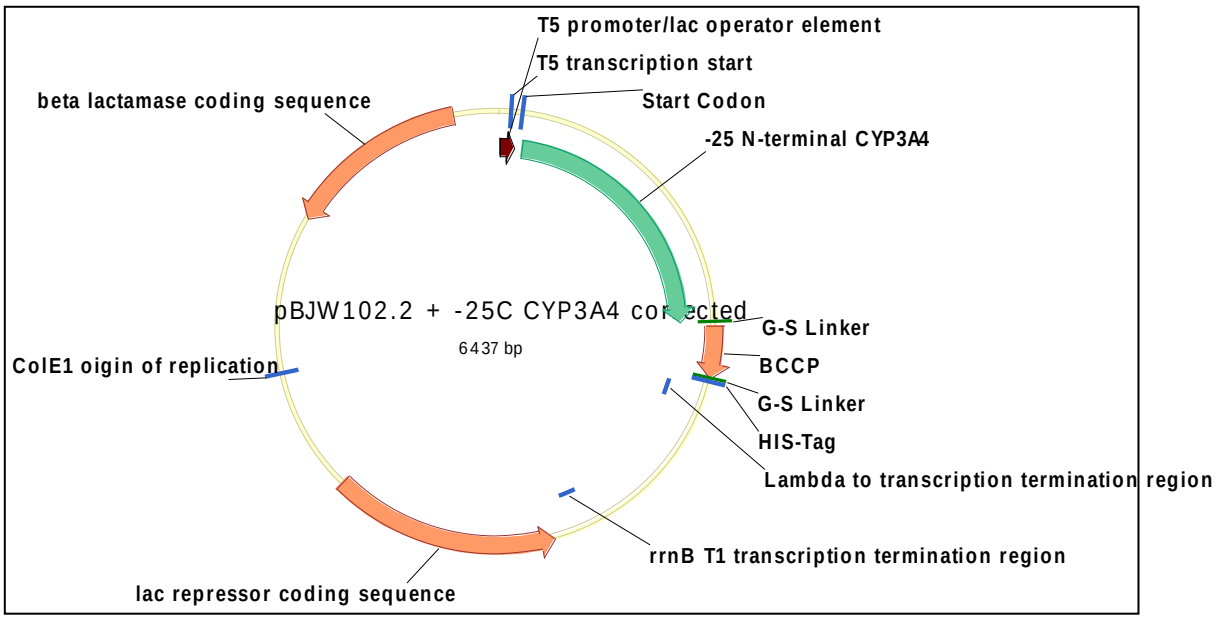
Zanger, U. M., & Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics*, *138*, 103–141.

Zhang, L., Zhang, Y. D., Zhao, P., & Huang, S. (2009). Predicting Drug – Drug Interactions: An FDA Perspective. *The AAPS Journal*, *11*(2), 300–306.

Zhou, S., Zhou, Z., & Huang, M. (2010). Polymorphisms of human cytochrome P450 2C9 and the functional relevance. *Toxicology*, *278*(2), 165–188.

Appendix

CYP3A4 vector diagram



CYP3A4 vector sequence

1 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG

GAGCTCTTTA GTATTTTTTA AATAAACGAA ACACTCGCCT ATTGTTAATA TTATCTAAGT TAACACTCGC CTATTGTTAA AGTGTGTCTT AAGTAATTTC

101 AGGAGAAATT AACTATGGCA CTTACGGCCT ATGGAACCCA TTCACATGGA CTTTTTAAGA AGCTTGGAAT TCCAGGGCCC ACACCTCTGC CTTTTTTGGG

TCCTCTTTAA TTGATACCGT GAATGCCGGA TACCTTGGGT AAGTGTACCT GAAAAATTCT TCGAACCTTA AGGTCCCGGG TGTGGAGACG GAAAAAACCC

201 AAATATTTTG TCCTACCATA AGGGCTTTTG TATGTTTGAC ATGGAATGTC ATAAAAAGTA TGGAAAAGTG TGGGGCTTTT ATGATGGTCA ACAGCCTGTG

TTTATAAAAC AGGATGGTAT TCCCGAAAAC ATACAAACTG TACCTTACAG TATTTTTCAT ACCTTTTCAC ACCCCGAAAA TACTACCAGT TGTCGGACAC

301 CTGGCTATCA CAGATCCTGA CATGATCAAA ACAGTGCTAG TGAAAGAATG TTATTCTGTC TTCACAAACC GGAGGCCTTT TGGTCCAGTG GGATTTATGA

GACCGATAGT GTCTAGGACT GTACTAGTTT TGTCACGATC ACTTTCTTAC AATAAGACAG AAGTGTTTGG CCTCCGGAAA ACCAGGTCAC CCTAAATACT

401 AAAGTGCCAT CTCTATAGCT GAGGATGAAG AATGGAAGAG ATTACGATCA TTGCTGTCTC CAACCTTCAC CAGTGGAAAA CTCAAGGAGA TGGTCCCTAT

TTTCACGGTA GAGATATCGA CTCCTACTTC TTACCTTCTC TAATGCTAGT AACGACAGAG GTTGGAAGTG GTCACCTTTT GAGTTCCTCT ACCAGGGATA

501 CATTGCCCAG TATGGAGATG TGTTGGTGAG AAATCTGAGG CGGGAAGCAG AGACAGGCAA GCCTGTCACC TTGAAAGACA TCTTTGGGGC CTACAGCATG

GTAACGGGTC ATACCTCTAC ACAACCACTC TTTAGACTCC GCCCTTCGTC TCTGTCCGTT CGGACAGTGG AACTTTCTGT AGAAACCCCG GATGTCGTAC

601 GATGTGATCA CTAGCACATC ATTTGGAGTG AACATCGACT CTCTCAACAA TCCACAAGAC CCCTTTGTGG AAAACACCAA GAAGCTTTTA AGATTTGATT

CTACACTAGT GATCGTGTAG TAAACCTCAC TTGTAGCTGA GAGAGTTGTT AGGTGTTCTG GGGAAACACC TTTTGTGGTT CTTCGAAAAT TCTAAACTAA

701 TTTTGGATCC ATTCTTTCTC TCAATAACAG TCTTTCCATT CCTCATCCCA ATTCTTGAAG TATTAAATAT CTGTGTGTTT CCAAGGGAAG TTACAAATTT

AAAACCTAGG TAAGAAAGAG AGTTATTGTC AGAAAGGTAA GGAGTAGGGT TAAGAACTTC ATAATTTATA GACACACAAA GGTTCCCTTC AATGTTTAAA

801 TTTAAGAAAA TCTATAAAAA GGATGAAAGA AAGTCGCCTC GAAGATACAC AAAAGCACCG AGTGGATTTC CTTCAGCTGA TGATTGACTC TCAGAATTCA

AAATTCTTTT AGATATTTTT CCTACTTTCT TTCAGCGGAG CTTCTATGTG TTTTCGTGGC TCACCTAAAG GAAGTCGACT ACTAACTGAG AGTCTTAAGT

901 AAAGAAACTG AGTCCCACAA AGCTCTGTCC GATCTGGAGC TCGTGGCCCA ATCAATTATC TTTATTTTTG CTGGCTATGA AACCACGAGC AGTGTTCTCT

TTTCTTTGAC TCAGGGTGTT TCGAGACAGG CTAGACCTCG AGCACCGGGT TAGTTAATAG AAATAAAAAC GACCGATACT TTGGTGCTCG TCACAAGAGA

1001 CCTTCATTAT GTATGAACTG GCCACTCACC CTGATGTCCA GCAGAAACTG CAGGAGGAAA TTGATGCAGT TTTACCCAAT AAGGCACCAC CCACCTATGA

GGAAGTAATA CATACTTGAC CGGTGAGTGG GACTACAGGT CGTCTTTGAC GTCCTCCTTT AACTACGTCA AAATGGGTTA TTCCGTGGTG GGTGGATACT

1101 TACTGTGCTA CAGATGGAGT ATCTTGACAT GGTGGTGAAT GAAACGCTCA GATTATTCCC AATTGCTATG AGACTTGAGA GGGTCTGCAA AAAAGATGTT

ATGACACGAT GTCTACCTCA TAGAACTGTA CCACCACTTA CTTTGCGAGT CTAATAAGGG TTAACGATAC TCTGAACTCT CCCAGACGTT TTTTCTACAA

1201 GAGATCAATG GGATGTTCAT TCCCAAAGGG GTGGTGGTGA TGATTCCAAG CTATGCTCTT CACCGTGACC CAAAGTACTG GACAGAGCCT GAGAAGTTCC

CTCTAGTTAC CCTACAAGTA AGGGTTTCCC CACCACCACT ACTAAGGTTC GATACGAGAA GTGGCACTGG GTTTCATGAC CTGTCTCGGA CTCTTCAAGG

1301 TCCCTGAAAG ATTCAGCAAG AAGAACAAGG ACAACATAGA TCCTTACATA TACACACCCT TTGGAAGTGG ACCCAGAAAC TGCATTGGTA TGAGGTTTGC

AGGGACTTTC TAAGTCGTTC TTCTTGTTCC TGTTGTATCT AGGAATGTAT ATGTGTGGGA AACCTTCACC TGGGTCTTTG ACGTAACCAT ACTCCAAACG

1401 TCTCATGAAC ATGAAACTTG CTCTAATCAG AGTCCTTCAG AACTTCTCCT TCAAACCTTG TAAAGAAACA CAGATCCCCC TGAAATTAAG CTTAGGAGGA

AGAGTACTTG TACTTTGAAC GAGATTAGTC TCAGGAAGTC TTGAAGAGGA AGTTTGGAAC ATTTCTTTGT GTCTAGGGGG ACTTTAATTC GAATCCTCCT

1501 CTTCTTCAAC CAGAAAAACC CGTTGTTCTA AAGGTTGAGT CAAGGGATGG CACCGTAAGT GGAGCCGGGG GTGGCAGCGG TTCTGGCGCA GCAGCGGAAA

GAAGAAGTTG GTCTTTTTGG GCAACAAGAT TTCCAACTCA GTTCCCTACC GTGGCATTCA CCTCGGCCCC CACCGTCGCC AAGACCGCGT CGTCGCCTTT

1601 TCAGTGGTCA CATCGTACGT TCCCCGATGG TTGGTACTTT CTACCGCACC CCAAGCCCGG ACGCAAAAGC GTTCATCGAA GTGGGTCAGA AAGTCAACGT

AGTCACCAGT GTAGCATGCA AGGGGCTACC AACCATGAAA GATGGCGTGG GGTTCGGGCC TGCGTTTTCG CAAGTAGCTT CACCCAGTCT TTCAGTTGCA

1701 GGGCGATACC CTGTGCATCG TTGAAGCCAT GAAAATGATG AACCAGATCG AAGCGGACAA ATCCGGTACC GTGAAAGCAA TTCTGGTCGA AAGTGGACAA

CCCGCTATGG GACACGTAGC AACTTCGGTA CTTTTACTAC TTGGTCTAGC TTCGCCTGTT TAGGCCATGG CACTTTCGTT AAGACCAGCT TTCACCTGTT

1801 CCGGTAGAAT TTGACGAGCC GCTGGTCGTC ATCGAGGGTG GCAGCGGTTC TGGCCACCAT CACCATCACC ATAAGCTTAA TTAGCTGAGC TTGGACTCCT

GGCCATCTTA AACTGCTCGG CGACCAGCAG TAGCTCCCAC CGTCGCCAAG ACCGGTGGTA GTGGTAGTGG TATTCGAATT AATCGACTCG AACCTGAGGA

1901 GTTGATAGAT CCAGTAATGA CCTCAGAACT CCATCTGGAT TTGTTCAGAA CGCTCGGTTG CCGCCGGGCG TTTTTTATTG GTGAGAATCC AAGCTAGCTT

CAACTATCTA GGTCATTACT GGAGTCTTGA GGTAGACCTA AACAAGTCTT GCGAGCCAAC GGCGGCCCGC AAAAAATAAC CACTCTTAGG TTCGATCGAA

2001 GGCGAGATTT TCAGGAGCTA AGGAAGCTAA AATGGAGAAA AAAATCACTG GATATACCAC CGTTGATATA TCCCAATGGC ATCGTAAAGA ACATTTTGAG

CCGCTCTAAA AGTCCTCGAT TCCTTCGATT TTACCTCTTT TTTTAGTGAC CTATATGGTG GCAACTATAT AGGGTTACCG TAGCATTTCT TGTAAAACTC

2101 GCATTTCAGT CAGTTGCTCA ATGTACCTAT AACCAGACCG TTCAGCTGGA TATTACGGCC TTTTTAAAGA CCGTAAAGAA AAATAAGCAC AAGTTTTATC

CGTAAAGTCA GTCAACGAGT TACATGGATA TTGGTCTGGC AAGTCGACCT ATAATGCCGG AAAAATTTCT GGCATTTCTT TTTATTCGTG TTCAAAATAG

2201 CGGCCTTTAT TCACATTCTT GCCCGCCTGA TGAATGCTCA TCCGGAATTT CGTATGGCAA TGAAAGACGG TGAGCTGGTG ATATGGGATA GTGTTCACCC

GCCGGAAATA AGTGTAAGAA CGGGCGGACT ACTTACGAGT AGGCCTTAAA GCATACCGTT ACTTTCTGCC ACTCGACCAC TATACCCTAT CACAAGTGGG

2301 TTGTTACACC GTTTTCCATG AGCAAACTGA AACGTTTTCA TCGCTCTGGA GTGAATACCA CGACGATTTC CGGCAGTTTC TACACATATA TTCGCAAGAT

AACAATGTGG CAAAAGGTAC TCGTTTGACT TTGCAAAAGT AGCGAGACCT CACTTATGGT GCTGCTAAAG GCCGTCAAAG ATGTGTATAT AAGCGTTCTA

2401 GTGGCGTGTT ACGGTGAAAA CCTGGCCTAT TTCCCTAAAG GGTTTATTGA GAATATGTTT TTCGTCTCAG CCAATCCCTG GGTGAGTTTC ACCAGTTTTG

CACCGCACAA TGCCACTTTT GGACCGGATA AAGGGATTTC CCAAATAACT CTTATACAAA AAGCAGAGTC GGTTAGGGAC CCACTCAAAG TGGTCAAAAC

2501 ATTTAAACGT GGCCAATATG GACAACTTCT TCGCCCCCGT TTTCACCATG GGCAAATATT ATACGCAAGG CGACAAGGTG CTGATGCCGC TGGCGATTCA

TAAATTTGCA CCGGTTATAC CTGTTGAAGA AGCGGGGGCA AAAGTGGTAC CCGTTTATAA TATGCGTTCC GCTGTTCCAC GACTACGGCG ACCGCTAAGT

2601 GGTTCATCAT GCCGTTTGTG ATGGCTTCCA TGTCGGCAGA ATGCTTAATG AATTACAACA GTACTGCGAT GAGTGGCAGG GCGGGGCGTA ATTTTTTTAA

CCAAGTAGTA CGGCAAACAC TACCGAAGGT ACAGCCGTCT TACGAATTAC TTAATGTTGT CATGACGCTA CTCACCGTCC CGCCCCGCAT TAAAAAAATT

2701 GGCAGTTATT GGTGCCCTTA AACGCCTGGG GTAATGACTC TCTAGCTTGA GGCATCAAAT AAAACGAAAG GCTCAGTCGA AAGACTGGGC CTTTCGTTTT

CCGTCAATAA CCACGGGAAT TTGCGGACCC CATTACTGAG AGATCGAACT CCGTAGTTTA TTTTGCTTTC CGAGTCAGCT TTCTGACCCG GAAAGCAAAA

2801 ATCTGTTGTT TGTCGGTGAA CGCTCTCCTG AGTAGGACAA ATCCGCCCTC TAGATTACGT GCAGTCGATG ATAAGCTGTC AAACATGAGA ATTGTGCCTA

TAGACAACAA ACAGCCACTT GCGAGAGGAC TCATCCTGTT TAGGCGGGAG ATCTAATGCA CGTCAGCTAC TATTCGACAG TTTGTACTCT TAACACGGAT

2901 ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG

TACTCACTCG ATTGAATGTA ATTAACGCAA CGCGAGTGAC GGGCGAAAGG TCAGCCCTTT GGACAGCACG GTCGACGTAA TTACTTAGCC GGTTGCGCGC

3001 GGGAGAGGCG GTTTGCGTAT TGGGCGCCAG GGTGGTTTTT CTTTTCACCA GTGAGACGGG CAACAGCTGA TTGCCCTTCA CCGCCTGGCC CTGAGAGAGT

CCCTCTCCGC CAAACGCATA ACCCGCGGTC CCACCAAAAA GAAAAGTGGT CACTCTGCCC GTTGTCGACT AACGGGAAGT GGCGGACCGG GACTCTCTCA

3101 TGCAGCAAGC GGTCCACGCT GGTTTGCCCC AGCAGGCGAA AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTCG GTATCGTCGT

ACGTCGTTCG CCAGGTGCGA CCAAACGGGG TCGTCCGCTT TTAGGACAAA CTACCACCAA TTGCCGCCCT ATATTGTACT CGACAGAAGC CATAGCAGCA

3201 ATCCCACTAC CGAGATATCC GCACCAACGC GCAGCCCGGA CTCGGTAATG GCGCGCATTG CGCCCAGCGC CATCTGATCG TTGGCAACCA GCATCGCAGT

TAGGGTGATG GCTCTATAGG CGTGGTTGCG CGTCGGGCCT GAGCCATTAC CGCGCGTAAC GCGGGTCGCG GTAGACTAGC AACCGTTGGT CGTAGCGTCA

3301 GGGAACGATG CCCTCATTCA GCATTTGCAT GGTTTGTTGA AAACCGGACA TGGCACTCCA GTCGCCTTCC CGTTCCGCTA TCGGCTGAAT TTGATTGCGA

CCCTTGCTAC GGGAGTAAGT CGTAAACGTA CCAAACAACT TTTGGCCTGT ACCGTGAGGT CAGCGGAAGG GCAAGGCGAT AGCCGACTTA AACTAACGCT

3401 GTGAGATATT TATGCCAGCC AGCCAGACGC AGACGCGCCG AGACAGAACT TAATGGGCCC GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT

CACTCTATAA ATACGGTCGG TCGGTCTGCG TCTGCGCGGC TCTGTCTTGA ATTACCCGGG CGATTGTCGC GCTAAACGAC CACTGGGTTA CGCTGGTCTA

3501 GCTCCACGCC CAGTCGCGTA CCGTCTTCAT GGGAGAAAAT AATACTGTTG ATGGGTGTCT GGTCAGAGAC ATCAAGAAAT AACGCCGGAA CATTAGTGCA

CGAGGTGCGG GTCAGCGCAT GGCAGAAGTA CCCTCTTTTA TTATGACAAC TACCCACAGA CCAGTCTCTG TAGTTCTTTA TTGCGGCCTT GTAATCACGT

3601 GGCAGCTTCC ACAGCAATGG CATCCTGGTC ATCCAGCGGA TAGTTAATGA TCAGCCCACT GACGCGTTGC GCGAGAAGAT TGTGCACCGC CGCTTTACAG

CCGTCGAAGG TGTCGTTACC GTAGGACCAG TAGGTCGCCT ATCAATTACT AGTCGGGTGA CTGCGCAACG CGCTCTTCTA ACACGTGGCG GCGAAATGTC

3701 GCTTCGACGC CGCTTCGTTC TACCATCGAC ACCACCACGC TGGCACCCAG TTGATCGGCG CGAGATTTAA TCGCCGCGAC AATTTGCGAC GGCGCGTGCA

CGAAGCTGCG GCGAAGCAAG ATGGTAGCTG TGGTGGTGCG ACCGTGGGTC AACTAGCCGC GCTCTAAATT AGCGGCGCTG TTAAACGCTG CCGCGCACGT

3801 GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG TTTGCCCGCC AGTTGTTGTG CCACGCGGTT GGGAATGTAA TTCAGCTCCG CCATCGCCGC

CCCGGTCTGA CCTCCACCGT TGCGGTTAGT CGTTGCTGAC AAACGGGCGG TCAACAACAC GGTGCGCCAA CCCTTACATT AAGTCGAGGC GGTAGCGGCG

3901 TTCCACTTTT TCCCGCGTTT TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGGAAAC GGTCTGATAA GAGACACCGG CATACTCTGC GACATCGTAT

AAGGTGAAAA AGGGCGCAAA AGCGTCTTTG CACCGACCGG ACCAAGTGGT GCGCCCTTTG CCAGACTATT CTCTGTGGCC GTATGAGACG CTGTAGCATA

4001 AACGTTACTG GTTTCACATT CACCACCCTG AATTGACTCT CTTCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTGCA CCATTCGATG GTGTCGGAAT

TTGCAATGAC CAAAGTGTAA GTGGTGGGAC TTAACTGAGA GAAGGCCCGC GATAGTACGG TATGGCGCTT TCCAAAACGT GGTAAGCTAC CACAGCCTTA

4101 TTCGGGCAGC GTTGGGTCCT GGCCACGGGT GCGCATGATC TAGAGCTGCC TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG

AAGCCCGTCG CAACCCAGGA CCGGTGCCCA CGCGTACTAG ATCTCGACGG AGCGCGCAAA GCCACTACTG CCACTTTTGG AGACTGTGTA CGTCGAGGGC

4201 GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCGCA GCCATGACCC

CTCTGCCAGT GTCGAACAGA CATTCGCCTA CGGCCCTCGT CTGTTCGGGC AGTCCCGCGC AGTCGCCCAC AACCGCCCAC AGCCCCGCGT CGGTACTGGG

4301 AGTCACGTAG CGATAGCGGA GTGTATACTG GCTTAACTAT GCGGCATCAG AGCAGATTGT ACTGAGAGTG CACCATATGC GGTGTGAAAT ACCGCACAGA

TCAGTGCATC GCTATCGCCT CACATATGAC CGAATTGATA CGCCGTAGTC TCGTCTAACA TGACTCTCAC GTGGTATACG CCACACTTTA TGGCGTGTCT

4401 TGCGTAAGGA GAAAATACCG CATCAGGCGC TCTTCCGCTT CCTCGCTCAC TGACTCGCTG CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT

ACGCATTCCT CTTTTATGGC GTAGTCCGCG AGAAGGCGAA GGAGCGAGTG ACTGAGCGAC GCGAGCCAGC AAGCCGACGC CGCTCGCCAT AGTCGAGTGA

4501 CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC AGGAACCGTA AAAAGGCCGC

GTTTCCGCCA TTATGCCAAT AGGTGTCTTA GTCCCCTATT GCGTCCTTTC TTGTACACTC GTTTTCCGGT CGTTTTCCGG TCCTTGGCAT TTTTCCGGCG

4601 GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC

CAACGACCGC AAAAAGGTAT CCGAGGCGGG GGGACTGCTC GTAGTGTTTT TAGCTGCGAG TTCAGTCTCC ACCGCTTTGG GCTGTCCTGA TATTTCTATG

4701 CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCT GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC

GTCCGCAAAG GGGGACCTTC GAGGGAGCAC GCGAGAGGAC AAGGCTGGGA CGGCGAATGG CCTATGGACA GGCGGAAAGA GGGAAGCCCT TCGCACCGCG

4801 TTTCTCATAG CTCACGCTGT AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC

AAAGAGTATC GAGTGCGACA TCCATAGAGT CAAGCCACAT CCAGCAAGCG AGGTTCGACC CGACACACGT GCTTGGGGGG CAAGTCGGGC TGGCGACGCG

4901 CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA

GAATAGGCCA TTGATAGCAG AACTCAGGTT GGGCCATTCT GTGCTGAATA GCGGTGACCG TCGTCGGTGA CCATTGTCCT AATCGTCTCG CTCCATACAT

5001 GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA

CCGCCACGAT GTCTCAAGAA CTTCACCACC GGATTGATGC CGATGTGATC TTCCTGTCAT AAACCATAGA CGCGAGACGA CTTCGGTCAA TGGAAGCCTT

5101 AAAGAGTTGG TAGCTCTTGA TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA

TTTCTCAACC ATCGAGAACT AGGCCGTTTG TTTGGTGGCG ACCATCGCCA CCAAAAAAAC AAACGTTCGT CGTCTAATGC GCGTCTTTTT TTCCTAGAGT

5201 AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC

TCTTCTAGGA AACTAGAAAA GATGCCCCAG ACTGCGAGTC ACCTTGCTTT TGAGTGCAAT TCCCTAAAAC CAGTACTCTA ATAGTTTTTC CTAGAAGTGG

5301 TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA

ATCTAGGAAA ATTTAATTTT TACTTCAAAA TTTAGTTAGA TTTCATATAT ACTCATTTGA ACCAGACTGT CAATGGTTAC GAATTAGTCA CTCCGTGGAT

5401 TCTCAGCGAT CTGTCTATTT CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GGAGGGCTTA CCATCTGGCC CCAGTGCTGC

AGAGTCGCTA GACAGATAAA GCAAGTAGGT ATCAACGGAC TGAGGGGCAG CACATCTATT GATGCTATGC CCTCCCGAAT GGTAGACCGG GGTCACGACG

5501 AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC

TTACTATGGC GCTCTGGGTG CGAGTGGCCG AGGTCTAAAT AGTCGTTATT TGGTCGGTCG GCCTTCCCGG CTCGCGTCTT CACCAGGACG TTGAAATAGG

5601 GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG

CGGAGGTAGG TCAGATAATT AACAACGGCC CTTCGATCTC ATTCATCAAG CGGTCAATTA TCAAACGCGT TGCAACAACG GTAACGATGT CCGTAGCACC

5701 TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG TGCAAAAAAG CGGTTAGCTC

ACAGTGCGAG CAGCAAACCA TACCGAAGTA AGTCGAGGCC AAGGGTTGCT AGTTCCGCTC AATGTACTAG GGGGTACAAC ACGTTTTTTC GCCAATCGAG

5801 CTTCGGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA

GAAGCCAGGA GGCTAGCAAC AGTCTTCATT CAACCGGCGT CACAATAGTG AGTACCAATA CCGTCGTGAC GTATTAAGAG AATGACAGTA CGGTAGGCAT

5901 AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA

TCTACGAAAA GACACTGACC ACTCATGAGT TGGTTCAGTA AGACTCTTAT CACATACGCC GCTGGCTCAA CGAGAACGGG CCGCAGTTAT GCCCTATTAT

6001 CCGCGCCACA TAGCAGAACT TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT

GGCGCGGTGT ATCGTCTTGA AATTTTCACG AGTAGTAACC TTTTGCAAGA AGCCCCGCTT TTGAGAGTTC CTAGAATGGC GACAACTCTA GGTCAAGCTA

6101 GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA

CATTGGGTGA GCACGTGGGT TGACTAGAAG TCGTAGAAAA TGAAAGTGGT CGCAAAGACC CACTCGTTTT TGTCCTTCCG TTTTACGGCG TTTTTTCCCT

6201 ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG

TATTCCCGCT GTGCCTTTAC AACTTATGAG TATGAGAAGG AAAAAGTTAT AATAACTTCG TAAATAGTCC CAATAACAGA GTACTCGCCT ATGTATAAAC

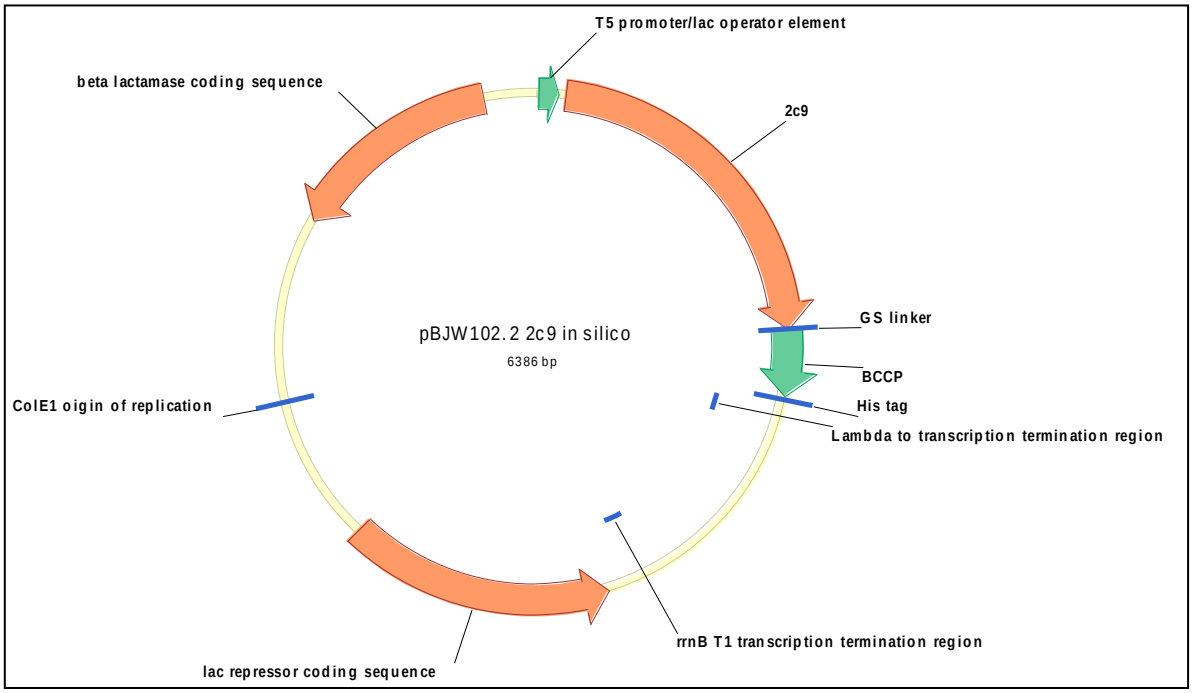
6301 AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA

TTACATAAAT CTTTTTATTT GTTTATCCCC AAGGCGCGTG TAAAGGGGCT TTTCACGGTG GACTGCAGAT TCTTTGGTAA TAATAGTACT GTAATTGGAT

6401 TAAAAATAGG CGTATCACGA GGCCCTTTCG TCTTCAC

ATTTTTATCC GCATAGTGCT CCGGGAAAGC AGAAGTG

CYP2C9 vector diagram



CYP2C9 vector sequence

```
1     CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG
      GAGCTCTTTA GTATTTTTTA AATAAACGAA ACACTCGCCT ATTGTTAATA TTATCTAAGT TAACACTCGC CTATTGTTAA AGTGTGTCTT AAGTAATTTC
101   AGGAGAAATT AACTATGGCA CTTACGGCCC TCCCTCCTGG CCCCACTCCT CTCCCAGTGA TTGGAAATAT CCTACAGATA GGTATTAAGG ACATCAGCAA
      TCCTCTTTAA TTGATACCGT GAATGCCGGG AGGGAGGACC GGGGTGAGGA GAGGGTCACT AACCTTTATA GGATGTCTAT CCATAATTCC TGTAGTCGTT
201   ATCCTTAACC AATCTCTCAA AGGTCTATGG CCCTGTGTTC ACTCTGTATT TTGGCCTGAA ACCCATAGTG GTGCTGCATG GATATGAAGC AGTGAAGGAA
      TAGGAATTGG TTAGAGAGTT TCCAGATACC GGGACACAAG TGAGACATAA AACCGGACTT TGGGTATCAC CACGACGTAC CTATACTTCG TCACTTCCTT
301   GCCCTGATTG ATCTTGGAGA GGAGTTTTCT GGAAGAGGCA TTTTCCCACT GGCTGAAAGA GCTAACAGAG GATTTGGAAT TGTTTTCAGC AATGGAAAGA
      CGGGACTAAC TAGAACCTCT CCTCAAAAGA CCTTCTCCGT AAAAGGGTGA CCGACTTTCT CGATTGTCTC CTAAACCTTA ACAAAAGTCG TTACCTTTCT
401   AATGGAAGGA GATCCGGCGT TTCTCCCTCA TGACGCTGCG GAATTTTGGG ATGGGGAAGA GGAGCATTGA GGACCGTGTT CAAGAGGAAG CCCGCTGCCT
      TTACCTTCCT CTAGGCCGCA AAGAGGGAGT ACTGCGACGC CTTAAAACCC TACCCCTTCT CCTCGTAACT CCTGGCACAA GTTCTCCTTC GGGCGACGGA
501   TGTGGAGGAG TTGAGAAAAA CCAAGGCCTC ACCCTGTGAT CCCACTTTCA TCCTGGGCTG TGCTCCCTGC AATGTGATCT GCTCCATTAT TTTCCATAAA
      ACACCTCCTC AACTCTTTTT GGTTCCGGAG TGGGACACTA GGGTGAAAGT AGGACCCGAC ACGAGGGACG TTACACTAGA CGAGGTAATA AAAGGTATTT
601   CGTTTTGATT ATAAAGATCA GCAATTTCTT AACTTAATGG AAAAGTTGAA TGAAAACATC AAGATTTTGA GCAGCCCCTG GATCCAGATC TGCAATAATT
      GCAAAACTAA TATTTCTAGT CGTTAAAGAA TTGAATTACC TTTTCAACTT ACTTTTGTAG TTCTAAAACT CGTCGGGGAC CTAGGTCTAG ACGTTATTAA
701   TTTCTCCTAT CATTGATTAC TTCCCGGGAA CTCACAACAA ATTACTTAAA AACGTTGCTT TTATGAAAAG TTATATTTTG GAAAAAGTAA AAGAACACCA
      AAAGAGGATA GTAACTAATG AAGGGCCCTT GAGTGTTGTT TAATGAATTT TTGCAACGAA AATACTTTTC AATATAAAAC CTTTTTCATT TTCTTGTGGT
801   AGAATCAATG GACATGAACA ACCCTCAGGA CTTTATTGAT TGCTTCCTGA TGAAAATGGA GAAGGAAAAG CACAACCAAC CATCTGAATT TACTATTGAA
      TCTTAGTTAC CTGTACTTGT TGGGAGTCCT GAAATAACTA ACGAAGGACT ACTTTTACCT CTTCCTTTTC GTGTTGGTTG GTAGACTTAA ATGATAACTT
901   AGCTTGGAAA ACACTGCAGT TGACTTGTTT GGAGCTGGGA CAGAGACGAC AAGCACAACC CTGAGATATG CTCTCCTTCT CCTGCTGAAG CACCCAGAGG
      TCGAACCTTT TGTGACGTCA ACTGAACAAA CCTCGACCCT GTCTCTGCTG TTCGTGTTGG GACTCTATAC GAGAGGAAGA GGACGACTTC GTGGGTCTCC
1001  TCACAGCTAA AGTCCAGGAA GAGATTGAAC GTGTGATTGG CAGAAACCGG AGCCCCTGCA TGCAAGACAG GAGCCACATG CCCTACACAG ATGCTGTGGT
      AGTGTCGATT TCAGGTCCTT CTCTAACTTG CACACTAACC GTCTTTGGCC TCGGGGACGT ACGTTCTGTC CTCGGTGTAC GGGATGTGTC TACGACACCA
1101  GCACGAGGTC CAGAGATACA TTGACCTTCT CCCCACCAGC CTGCCCCATG CAGTGACCTG TGACATTAAA TTCAGAAACT ATCTCATTCC CAAGGGCACA
      CGTGCTCCAG GTCTCTATGT AACTGGAAGA GGGGTGGTCG GACGGGGTAC GTCACTGGAC ACTGTAATTT AAGTCTTTGA TAGAGTAAGG GTTCCCGTGT
1201  ACCATATTAA TTTCCCTGAC TTCTGTGCTA CATGACAACA AAGAATTTCC CAACCCAGAG ATGTTTGACC CTCATCACTT TCTGGATGAA GGTGGCAATT
      TGGTATAATT AAAGGGACTG AAGACACGAT GTACTGTTGT TTCTTAAAGG GTTGGGTCTC TACAAACTGG GAGTAGTGAA AGACCTACTT CCACCGTTAA
1301  TTAAGAAAAG TAAATACTTC ATGCCTTTCT CAGCAGGAAA ACGGATTTGT GTGGGAGAAG CCCTGGCCGG CATGGAGCTG TTTTTATTCC TGACCTCCAT
      AATTCTTTTC ATTTATGAAG TACGGAAAGA GTCGTCCTTT TGCCTAAACA CACCCTCTTC GGGACCGGCC GTACCTCGAC AAAAATAAGG ACTGGAGGTA
1401  TTTACAGAAC TTTAACCTGA AATCTCTGGT TGACCCAAAG AACCTTGACA CCACTCCAGT TGTCAATGGA TTTGCCTCTG TGCCGCCCTT CTACCAGCTG
      AAATGTCTTG AAATTGGACT TTAGAGACCA ACTGGGTTTC TTGGAACTGT GGTGAGGTCA ACAGTTACCT AAACGGAGAC ACGGCGGGAA GATGGTCGAC
1501  TGCTTCATTC CTGTCGGGGG TGGCAGCGGT TCTGGCGCAG CAGCGGAAAT CAGTGGTCAC ATCGTACGTT CCCCGATGGT TGGTACTTTC TACCGCACCC
      ACGAAGTAAG GACAGCCCCC ACCGTCGCCA AGACCGCGTC GTCGCCTTTA GTCACCAGTG TAGCATGCAA GGGGCTACCA ACCATGAAAG ATGGCGTGGG
1601  CAAGCCCGGA CGCAAAAGCG TTCATCGAAG TGGGTCAGAA AGTCAACGTG GGCGATACCC TGTGCATCGT TGAAGCCATG AAAATGATGA ACCAGATCGA
      GTTCGGGCCT GCGTTTTCGC AAGTAGCTTC ACCCAGTCTT TCAGTTGCAC CCGCTATGGG ACACGTAGCA ACTTCGGTAC TTTTACTACT TGGTCTAGCT
1701  AGCGGACAAA TCCGGTACCG TGAAAGCAAT TCTGGTCGAA AGTGGACAAC CGGTAGAATT TGACGAGCCG CTGGTCGTCA TCGAGGGTGG CAGCGGTTCT
      TCGCCTGTTT AGGCCATGGC ACTTTCGTTA AGACCAGCTT TCACCTGTTG GCCATCTTAA ACTGCTCGGC GACCAGCAGT AGCTCCCACC GTCGCCAAGA
1801  GGCCACCATC ACCATCACCA TAAGCTTAAT TAGCTGAGCT TGGACTCCTG TTGATAGATC CAGTAATGAC CTCAGAACTC CATCTGGATT TGTTCAGAAC
      CCGGTGGTAG TGGTAGTGGT ATTCGAATTA ATCGACTCGA ACCTGAGGAC AACTATCTAG GTCATTACTG GAGTCTTGAG GTAGACCTAA ACAAGTCTTG
1901  GCTCGGTTGC CGCCGGGCGT TTTTTATTGG TGAGAATCCA AGCTAGCTTG GCGAGATTTT CAGGAGCTAA GGAAGCTAAA ATGGAGAAAA AAATCACTGG
      CGAGCCAACG GCGGCCCGCA AAAAATAACC ACTCTTAGGT TCGATCGAAC CGCTCTAAAA GTCCTCGATT CCTTCGATTT TACCTCTTTT TTTAGTGACC
2001  ATATACCACC GTTGATATAT CCCAATGGCA TCGTAAAGAA CATTTTGAGG CATTTCAGTC AGTTGCTCAA TGTACCTATA ACCAGACCGT TCAGCTGGAT
      TATATGGTGG CAACTATATA GGGTTACCGT AGCATTTCTT GTAAAACTCC GTAAAGTCAG TCAACGAGTT ACATGGATAT TGGTCTGGCA AGTCGACCTA
2101  ATTACGGCCT TTTTAAAGAC CGTAAAGAAA AATAAGCACA AGTTTTATCC GGCCTTTATT CACATTCTTG CCCGCCTGAT GAATGCTCAT CCGGAATTTC
      TAATGCCGGA AAAATTTCTG GCATTTCTTT TTATTCGTGT TCAAAATAGG CCGGAAATAA GTGTAAGAAC GGGCGGACTA CTTACGAGTA GGCCTTAAAG
2201  GTATGGCAAT GAAAGACGGT GAGCTGGTGA TATGGGATAG TGTTCACCCT TGTTACACCG TTTTCCATGA GCAAACTGAA ACGTTTTCAT CGCTCTGGAG
```

CATACCGTTA CTTTCTGCCA CTCGACCACT ATACCCTATC ACAAGTGGGA ACAATGTGGC AAAAGGTACT CGTTTGACTT TGCAAAAGTA GCGAGACCTC

2301 TGAATACCAC GACGATTTCC GGCAGTTTCT ACACATATAT TCGCAAGATG TGGCGTGTTA CGGTGAAAAC CTGGCCTATT TCCCTAAAGG GTTTATTGAG

ACTTATGGTG CTGCTAAAGG CCGTCAAAGA TGTGTATATA AGCGTTCTAC ACCGCACAAT GCCACTTTTG GACCGGATAA AGGGATTTCC CAAATAACTC

2401 AATATGTTTT TCGTCTCAGC CAATCCCTGG GTGAGTTTCA CCAGTTTTGA TTTAAACGTG GCCAATATGG ACAACTTCTT CGCCCCCGTT TTCACCATGG

TTATACAAAA AGCAGAGTCG GTTAGGGACC CACTCAAAGT GGTCAAAACT AAATTTGCAC CGGTTATACC TGTTGAAGAA GCGGGGGCAA AAGTGGTACC

2501 GCAAATATTA TACGCAAGGC GACAAGGTGC TGATGCCGCT GGCGATTCAG GTTCATCATG CCGTTTGTGA TGGCTTCCAT GTCGGCAGAA TGCTTAATGA

CGTTTATAAT ATGCGTTCCG CTGTTCCACG ACTACGGCGA CCGCTAAGTC CAAGTAGTAC GGCAAACACT ACCGAAGGTA CAGCCGTCTT ACGAATTACT

2601 ATTACAACAG TACTGCGATG AGTGGCAGGG CGGGGCGTAA TTTTTTTAAG GCAGTTATTG GTGCCCTTAA ACGCCTGGGG TAATGACTCT CTAGCTTGAG

TAATGTTGTC ATGACGCTAC TCACCGTCCC GCCCCGCATT AAAAAAATTC CGTCAATAAC CACGGGAATT TGCGGACCCC ATTACTGAGA GATCGAACTC

2701 GCATCAAATA AAACGAAAGG CTCAGTCGAA AGACTGGGCC TTTCGTTTTA TCTGTTGTTT GTCGGTGAAC GCTCTCCTGA GTAGGACAAA TCCGCCCTCT

CGTAGTTTAT TTTGCTTTCC GAGTCAGCTT TCTGACCCGG AAAGCAAAAT AGACAACAAA CAGCCACTTG CGAGAGGACT CATCCTGTTT AGGCGGGAGA

2801 AGATTACGTG CAGTCGATGA TAAGCTGTCA AACATGAGAA TTGTGCCTAA TGAGTGAGCT AACTTACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA

TCTAATGCAC GTCAGCTACT ATTCGACAGT TTGTACTCTT AACACGGATT ACTCACTCGA TTGAATGTAA TTAACGCAAC GCGAGTGACG GGCGAAAGGT

2901 GTCGGGAAAC CTGTCGTGCC AGCTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT GGGCGCCAGG GTGGTTTTTC TTTTCACCAG

CAGCCCTTTG GACAGCACGG TCGACGTAAT TACTTAGCCG GTTGCGCGCC CCTCTCCGCC AAACGCATAA CCCGCGGTCC CACCAAAAAG AAAAGTGGTC

3001 TGAGACGGGC AACAGCTGAT TGCCCTTCAC CGCCTGGCCC TGAGAGAGTT GCAGCAAGCG GTCCACGCTG GTTTGCCCCA GCAGGCGAAA ATCCTGTTTG

ACTCTGCCCG TTGTCGACTA ACGGGAAGTG GCGGACCGGG ACTCTCTCAA CGTCGTTCGC CAGGTGCGAC CAAACGGGGT CGTCCGCTTT TAGGACAAAC

3101 ATGGTGGTTA ACGGCGGGAT ATAACATGAG CTGTCTTCGG TATCGTCGTA TCCCACTACC GAGATATCCG CACCAACGCG CAGCCCGGAC TCGGTAATGG

TACCACCAAT TGCCGCCCTA TATTGTACTC GACAGAAGCC ATAGCAGCAT AGGGTGATGG CTCTATAGGC GTGGTTGCGC GTCGGGCCTG AGCCATTACC

3201 CGCGCATTGC GCCCAGCGCC ATCTGATCGT TGGCAACCAG CATCGCAGTG GGAACGATGC CCTCATTCAG CATTTGCATG GTTTGTTGAA AACCGGACAT

GCGCGTAACG CGGGTCGCGG TAGACTAGCA ACCGTTGGTC GTAGCGTCAC CCTTGCTACG GGAGTAAGTC GTAAACGTAC CAAACAACTT TTGGCCTGTA

3301 GGCACTCCAG TCGCCTTCCC GTTCCGCTAT CGGCTGAATT TGATTGCGAG TGAGATATTT ATGCCAGCCA GCCAGACGCA GACGCGCCGA GACAGAACTT

CCGTGAGGTC AGCGGAAGGG CAAGGCGATA GCCGACTTAA ACTAACGCTC ACTCTATAAA TACGGTCGGT CGGTCTGCGT CTGCGCGGCT CTGTCTTGAA

3401 AATGGGCCCG CTAACAGCGC GATTTGCTGG TGACCCAATG CGACCAGATG CTCCACGCCC AGTCGCGTAC CGTCTTCATG GGAGAAAATA ATACTGTTGA

TTACCCGGGC GATTGTCGCG CTAAACGACC ACTGGGTTAC GCTGGTCTAC GAGGTGCGGG TCAGCGCATG GCAGAAGTAC CCTCTTTTAT TATGACAACT

3501 TGGGTGTCTG GTCAGAGACA TCAAGAAATA ACGCCGGAAC ATTAGTGCAG GCAGCTTCCA CAGCAATGGC ATCCTGGTCA TCCAGCGGAT AGTTAATGAT

ACCCACAGAC CAGTCTCTGT AGTTCTTTAT TGCGGCCTTG TAATCACGTC CGTCGAAGGT GTCGTTACCG TAGGACCAGT AGGTCGCCTA TCAATTACTA

3601 CAGCCCACTG ACGCGTTGCG CGAGAAGATT GTGCACCGCC GCTTTACAGG CTTCGACGCC GCTTCGTTCT ACCATCGACA CCACCACGCT GGCACCCAGT

GTCGGGTGAC TGCGCAACGC GCTCTTCTAA CACGTGGCGG CGAAATGTCC GAAGCTGCGG CGAAGCAAGA TGGTAGCTGT GGTGGTGCGA CCGTGGGTCA

3701 TGATCGGCGC GAGATTTAAT CGCCGCGACA ATTTGCGACG GCGCGTGCAG GGCCAGACTG GAGGTGGCAA CGCCAATCAG CAACGACTGT TTGCCCGCCA

ACTAGCCGCG CTCTAAATTA GCGGCGCTGT TAAACGCTGC CGCGCACGTC CCGGTCTGAC CTCCACCGTT GCGGTTAGTC GTTGCTGACA AACGGGCGGT

3801 GTTGTTGTGC CACGCGGTTG GGAATGTAAT TCAGCTCCGC CATCGCCGCT TCCACTTTTT CCCGCGTTTT CGCAGAAACG TGGCTGGCCT GGTTCACCAC

CAACAACACG GTGCGCCAAC CCTTACATTA AGTCGAGGCG GTAGCGGCGA AGGTGAAAAA GGGCGCAAAA GCGTCTTTGC ACCGACCGGA CCAAGTGGTG

3901 GCGGGAAACG GTCTGATAAG AGACACCGGC ATACTCTGCG ACATCGTATA ACGTTACTGG TTTCACATTC ACCACCCTGA ATTGACTCTC TTCCGGGCGC

CGCCCTTTGC CAGACTATTC TCTGTGGCCG TATGAGACGC TGTAGCATAT TGCAATGACC AAAGTGTAAG TGGTGGGACT TAACTGAGAG AAGGCCCGCG

4001 TATCATGCCA TACCGCGAAA GGTTTTGCAC CATTCGATGG TGTCGGAATT TCGGGCAGCG TTGGGTCCTG GCCACGGGTG CGCATGATCT AGAGCTGCCT

ATAGTACGGT ATGGCGCTTT CCAAAACGTG GTAAGCTACC ACAGCCTTAA AGCCCGTCGC AACCCAGGAC CGGTGCCCAC GCGTACTAGA TCTCGACGGA

4101 CGCGCGTTTC GGTGATGACG GTGAAAACCT CTGACACATG CAGCTCCCGG AGACGGTCAC AGCTTGTCTG TAAGCGGATG CCGGGAGCAG ACAAGCCCGT

GCGCGCAAAG CCACTACTGC CACTTTTGGA GACTGTGTAC GTCGAGGGCC TCTGCCAGTG TCGAACAGAC ATTCGCCTAC GGCCCTCGTC TGTTCGGGCA

4201 CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT CGGGGCGCAG CCATGACCCA GTCACGTAGC GATAGCGGAG TGTATACTGG CTTAACTATG CGGCATCAGA

GTCCCGCGCA GTCGCCCACA ACCGCCCACA GCCCCGCGTC GGTACTGGGT CAGTGCATCG CTATCGCCTC ACATATGACC GAATTGATAC GCCGTAGTCT

4301 GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCT CTTCCGCTTC CTCGCTCACT

CGTCTAACAT GACTCTCACG TGGTATACGC CACACTTTAT GGCGTGTCTA CGCATTCCTC TTTTATGGCG TAGTCCGCGA GAAGGCGAAG GAGCGAGTGA

4401 GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA

CTGAGCGACG CGAGCCAGCA AGCCGACGCC GCTCGCCATA GTCGAGTGAG TTTCCGCCAT TATGCCAATA GGTGTCTTAG TCCCCTATTG CGTCCTTTCT

4501 ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA

TGTACACTCG TTTTCCGGTC GTTTTCCGGT CCTTGGCATT TTTCCGGCGC AACGACCGCA AAAAGGTATC CGAGGCGGGG GGACTGCTCG TAGTGTTTTT

4601 TCGACGCTCA AGTCAGAGGT GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG

AGCTGCGAGT TCAGTCTCCA CCGCTTTGGG CTGTCCTGAT ATTTCTATGG TCCGCAAAGG GGGACCTTCG AGGGAGCACG CGAGAGGACA AGGCTGGGAC

4701 CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT

GGCGAATGGC CTATGGACAG GCGGAAAGAG GGAAGCCCTT CGCACCGCGA AAGAGTATCG AGTGCGACAT CCATAGAGTC AAGCCACATC CAGCAAGCGA

4801 CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC

GGTTCGACCC GACACACGTG CTTGGGGGGC AAGTCGGGCT GGCGACGCGG AATAGGCCAT TGATAGCAGA ACTCAGGTTG GGCCATTCTG TGCTGAATAG

4901 GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA

CGGTGACCGT CGTCGGTGAC CATTGTCCTA ATCGTCTCGC TCCATACATC CGCCACGATG TCTCAAGAAC TTCACCACCG GATTGATGCC GATGTGATCT

5001 AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG

TCCTGTCATA AACCATAGAC GCGAGACGAC TTCGGTCAAT GGAAGCCTTT TTCTCAACCA TCGAGAACTA GGCCGTTTGT TTGGTGGCGA CCATCGCCAC

5101 GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAA

CAAAAAAACA AACGTTCGTC GTCTAATGCG CGTCTTTTTT TCCTAGAGTT CTTCTAGGAA ACTAGAAAAG ATGCCCCAGA CTGCGAGTCA CCTTGCTTTT

5201 CTCACGTTAA GGGATTTTGG TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT

GAGTGCAATT CCCTAAAACC AGTACTCTAA TAGTTTTTCC TAGAAGTGGA TCTAGGAAAA TTTAATTTTT ACTTCAAAAT TTAGTTAGAT TTCATATATA

5301 GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTC GTTCATCCAT AGTTGCCTGA CTCCCCGTCG

CTCATTTGAA CCAGACTGTC AATGGTTACG AATTAGTCAC TCCGTGGATA GAGTCGCTAG ACAGATAAAG CAAGTAGGTA TCAACGGACT GAGGGGCAGC

5401 TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG CTCACCGGCT CCAGATTTAT CAGCAATAAA

ACATCTATTG ATGCTATGCC CTCCCGAATG GTAGACCGGG GTCACGACGT TACTATGGCG CTCTGGGTGC GAGTGGCCGA GGTCTAAATA GTCGTTATTT

5501 CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGGTCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG

GGTCGGTCGG CCTTCCCGGC TCGCGTCTTC ACCAGGACGT TGAAATAGGC GGAGGTAGGT CAGATAATTA ACAACGGCCC TTCGATCTCA TTCATCAAGC

5601 CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT

GGTCAATTAT CAAACGCGTT GCAACAACGG TAACGATGTC CGTAGCACCA CAGTGCGAGC AGCAAACCAT ACCGAAGTAA GTCGAGGCCA AGGGTTGCTA

5701 CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT

GTTCCGCTCA ATGTACTAGG GGGTACAACA CGTTTTTTCG CCAATCGAGG AAGCCAGGAG GCTAGCAACA GTCTTCATTC AACCGGCGTC ACAATAGTGA

5801 CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG

GTACCAATAC CGTCGTGACG TATTAAGAGA ATGACAGTAC GGTAGGCATT CTACGAAAAG ACACTGACCA CTCATGAGTT GGTTCAGTAA GACTCTTATC

5901 TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC GGGATAATAC CGCGCCACAT AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT

ACATACGCCG CTGGCTCAAC GAGAACGGGC CGCAGTTATG CCCTATTATG GCGCGGTGTA TCGTCTTGAA ATTTTCACGA GTAGTAACCT TTTGCAAGAA

6001 CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG

GCCCCGCTTT TGAGAGTTCC TAGAATGGCG ACAACTCTAG GTCAAGCTAC ATTGGGTGAG CACGTGGGTT GACTAGAAGT CGTAGAAAAT GAAAGTGGTC

6101 CGTTTCTGGG TGAGCAAAAA CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT

GCAAAGACCC ACTCGTTTTT GTCCTTCCGT TTTACGGCGT TTTTTCCCTT ATTCCCGCTG TGCCTTTACA ACTTATGAGT ATGAGAAGGA AAAAGTTATA

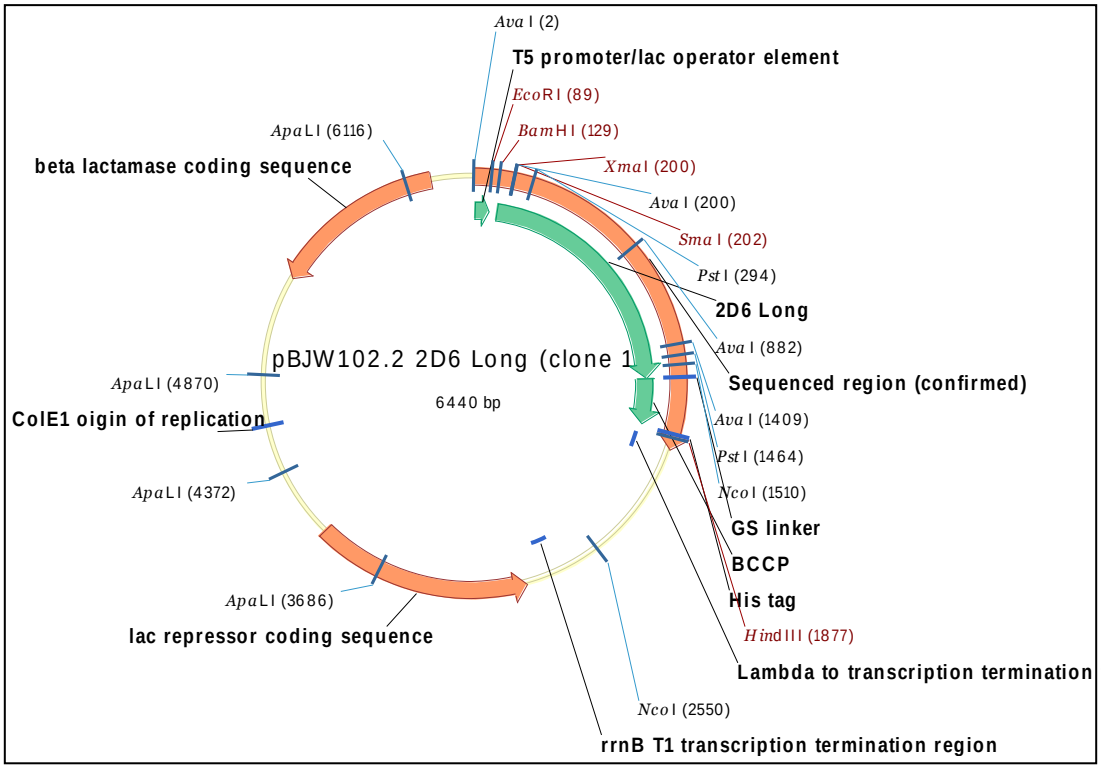
6201 TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA

ATAACTTCGT AAATAGTCCC AATAACAGAG TACTCGCCTA TGTATAAACT TACATAAATC TTTTTATTTG TTTATCCCCA AGGCGCGTGT AAAGGGGCTT

6301 AAGTGCCACC TGACGTCTAA GAAACCATTA TTATCATGAC ATTAACCTAT AAAAATAGGC GTATCACGAG GCCCTTTCGT CTTCAC

TTCACGGTGG ACTGCAGATT CTTTGGTAAT AATAGTACTG TAATTGGATA TTTTTATCCG CATAGTGCTC CGGGAAAGCA GAAGTG

CYP2D6 vector diagram



CYP2D6 vector sequence

1 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG

GAGCTCTTTA GTATTTTTTA AATAAACGAA ACACTCGCCT ATTGTTAATA TTATCTAAGT TAACACTCGC CTATTGTTAA AGTGTGTCTT AAGTAATTTC

101 AGGAGAAATT ACATATGGCA CTTAGTGGGA TCCCCGTGGA CCTGATGCAC CGGCGCCAAC GCTGGGCTGC ACGCTACCCA CCAGGCCCCC TGCCACTGCC

TCCTCTTTAA TGTATACCGT GAATCACCCT AGGGGCACCT GGACTACGTG GCCGCGGTTG CGACCCGACG TGCGATGGGT GGTCCGGGGG ACGGTGACGG

201 CGGGCTGGGC AACCTGCTGC ATGTGGACTT CCAGAACACA CCATACTGCT TCGACCAGTT GCGGCGCCGC TTCGGGGACG TGTTCAGCCT GCAGCTGGCC

GCCCGACCCG TTGGACGACG TACACCTGAA GGTCTTGTGT GGTATGACGA AGCTGGTCAA CGCCGCGGCG AAGCCCCTGC ACAAGTCGGA CGTCGACCGG

301 TGGACGCCGG TGGTCGTGCT CAATGGGCTG GCGGCCGTGC GCGAGGCGCT GGTGACCCAC GGCGAGGACA CCGCCGACCG CCCGCCTGTG CCCATCACCC

ACCTGCGGCC ACCAGCACGA GTTACCCGAC CGCCGGCACG CGCTCCGCGA CCACTGGGTG CCGCTCCTGT GGCGGCTGGC GGGCGGACAC GGGTAGTGGG

401 AGATCCTGGG TTTCGGGCCG CGTTCCCAAG GGGTGTTCCT GGCGCGCTAT GGGCCCGCGT GGCGCGAGCA GAGGCGCTTC TCCGTGTCCA CCTTGCGCAA

TCTAGGACCC AAAGCCCGGC GCAAGGGTTC CCCACAAGGA CCGCGCGATA CCCGGGCGCA CCGCGCTCGT CTCCGCGAAG AGGCACAGGT GGAACGCGTT

501 CTTGGGCCTG GGCAAGAAGT CGCTGGAGCA GTGGGTGACC GAGGAGGCCG CCTGCCTTTG TGCCGCCTTC GCCAACCACT CCGGACGCCC CTTTCGCCCC

GAACCCGGAC CCGTTCTTCA GCGACCTCGT CACCCACTGG CTCCTCCGGC GGACGGAAAC ACGGCGGAAG CGGTTGGTGA GGCCTGCGGG GAAAGCGGGG

601 AACGGTCTCT TGGACAAAGC CGTGAGCAAC GTGATCGCCT CCCTCACCTG CGGGCGCCGC TTCGAGTACG ACGACCCTCG CTTCCTCAGG CTGCTGGACC

TTGCCAGAGA ACCTGTTTCG GCACTCGTTG CACTAGCGGA GGGAGTGGAC GCCCGCGGCG AAGCTCATGC TGCTGGGAGC GAAGGAGTCC GACGACCTGG

701 TAGCTCAGGA GGGACTGAAG GAGGAGTCGG GCTTTCTGCG CGAGGTGCTG AATGCTGTCC CCGTCCTCCT GCATATCCCA GCGCTGGCTG GCAAGGTCCT

ATCGAGTCCT CCCTGACTTC CTCCTCAGCC CGAAAGACGC GCTCCACGAC TTACGACAGG GGCAGGAGGA CGTATAGGGT CGCGACCGAC CGTTCCAGGA

801 ACGCTTCCAA AAGGCTTTCC TGACCCAGCT GGATGAGCTG CTAACTGAGC ACAGGATGAC CTGGGACCCA GCCCAGCCAC CCCGAGACCT GACTGAGGCC

TGCGAAGGTT TTCCGAAAGG ACTGGGTCGA CCTACTCGAC GATTGACTCG TGTCCTACTG GACCCTGGGT CGGGTCGGTG GGGCTCTGGA CTGACTCCGG

901 TTCCTGGCAG AGATGGAGAA GGCCAAGGGG AACCCTGAGA GCAGCTTCAA TGATGAGAAC CTGCGCATAG TGGTGGCTGA CCTGTTCTCT GCCGGGATGG

AAGGACCGTC TCTACCTCTT CCGGTTCCCC TTGGGACTCT CGTCGAAGTT ACTACTCTTG GACGCGTATC ACCACCGACT GGACAAGAGA CGGCCCTACC

1001 TGACCACCTC GACCACGCTG GCCTGGGGCC TCCTGCTCAT GATCCTACAT CCGGATGTGC AGCGCCGTGT CCAACAGGAG ATCGACGACG TGATAGGGCA

ACTGGTGGAG CTGGTGCGAC CGGACCCCGG AGGACGAGTA CTAGGATGTA GGCCTACACG TCGCGGCACA GGTTGTCCTC TAGCTGCTGC ACTATCCCGT

1101 GGTGCGGCGA CCAGAGATGG GTGACCAGGC TCACATGCCC TACACCACTG CCGTGATTCA TGAGGTGCAG CGCTTTGGGG ACATCGTCCC CCTGGGTGTG

CCACGCCGCT GGTCTCTACC CACTGGTCCG AGTGTACGGG ATGTGGTGAC GGCACTAAGT ACTCCACGTC GCGAAACCCC TGTAGCAGGG GGACCCACAC

1201 ACCCATATGA CATCCCGTGA CATCGAAGTA CAGGGCTTCC GCATCCCTAA GGGAACGACA CTCATCACCA ACCTGTCATC GGTGCTGAAG GATGAGGCCG

TGGGTATACT GTAGGGCACT GTAGCTTCAT GTCCCGAAGG CGTAGGGATT CCCTTGCTGT GAGTAGTGGT TGGACAGTAG CCACGACTTC CTACTCCGGC

1301 TCTGGGAGAA GCCCTTCCGC TTCCACCCCG AACACTTCCT GGATGCCCAG GGCCACTTTG TGAAGCCGGA GGCCTTCCTG CCTTTCTCAG CAGGCCGCCG

AGACCCTCTT CGGGAAGGCG AAGGTGGGGC TTGTGAAGGA CCTACGGGTC CCGGTGAAAC ACTTCGGCCT CCGGAAGGAC GGAAAGAGTC GTCCGGCGGC

1401 TGCATGCCTC GGGGAGCCCC TGGCCCGCAT GGAGCTCTTC CTCTTCTTCA CCTCCCTGCT GCAGCACTTC AGCTTCTCGG TGCCCACTGG ACAGCCCCGG

ACGTACGGAG CCCCTCGGGG ACCGGGCGTA CCTCGAGAAG GAGAAGAAGT GGAGGGACGA CGTCGTGAAG TCGAAGAGCC ACGGGTGACC TGTCGGGGCC

1501 CCCAGCCACC ATGGTGTCTT TGCTTTCCTG GTGAGCCCAT CCCCCTATGA GCTTTGTGCT GTGCCCCGCG GGGGTGGCAG CGGTTCTGGC GCAGCAGCGG

GGGTCGGTGG TACCACAGAA ACGAAAGGAC CACTCGGGTA GGGGGATACT CGAAACACGA CACGGGGCGC CCCCACCGTC GCCAAGACCG CGTCGTCGCC

1601 AAATCAGTGG TCACATCGTA CGTTCCCCGA TGGTTGGTAC TTTCTACCGC ACCCCAAGCC CGGACGCAAA AGCGTTCATC GAAGTGGGTC AGAAAGTCAA

TTTAGTCACC AGTGTAGCAT GCAAGGGGCT ACCAACCATG AAAGATGGCG TGGGGTTCGG GCCTGCGTTT TCGCAAGTAG CTTCACCCAG TCTTTCAGTT

1701 CGTGGGCGAT ACCCTGTGCA TCGTTGAAGC CATGAAAATG ATGAACCAGA TCGAAGCGGA CAAATCCGGT ACCGTGAAAG CAATTCTGGT CGAAAGTGGA

GCACCCGCTA TGGGACACGT AGCAACTTCG GTACTTTTAC TACTTGGTCT AGCTTCGCCT GTTTAGGCCA TGGCACTTTC GTTAAGACCA GCTTTCACCT

1801 CAACCGGTAG AATTTGACGA GCCGCTGGTC GTCATCGAGG GTGGCAGCGG TTCTGGCCAC CATCACCATC ACCATAAGCT TAATTAGCTG AGCTTGGACT

GTTGGCCATC TTAAACTGCT CGGCGACCAG CAGTAGCTCC CACCGTCGCC AAGACCGGTG GTAGTGGTAG TGGTATTCGA ATTAATCGAC TCGAACCTGA

1901 CCTGTTGATA GATCCAGTAA TGACCTCAGA ACTCCATCTG GATTTGTTCA GAACGCTCGG TTGCCGCCGG GCGTTTTTTA TTGGTGAGAA TCCAAGCTAG

GGACAACTAT CTAGGTCATT ACTGGAGTCT TGAGGTAGAC CTAAACAAGT CTTGCGAGCC AACGGCGGCC CGCAAAAAAT AACCACTCTT AGGTTCGATC

2001 CTTGGCGAGA TTTTCAGGAG CTAAGGAAGC TAAAATGGAG AAAAAAATCA CTGGATATAC CACCGTTGAT ATATCCCAAT GGCATCGTAA AGAACATTTT

GAACCGCTCT AAAAGTCCTC GATTCCTTCG ATTTTACCTC TTTTTTTAGT GACCTATATG GTGGCAACTA TATAGGGTTA CCGTAGCATT TCTTGTAAAA

2101 GAGGCATTTC AGTCAGTTGC TCAATGTACC TATAACCAGA CCGTTCAGCT GGATATTACG GCCTTTTTAA AGACCGTAAA GAAAAATAAG CACAAGTTTT

CTCCGTAAAG TCAGTCAACG AGTTACATGG ATATTGGTCT GGCAAGTCGA CCTATAATGC CGGAAAAATT TCTGGCATTT CTTTTTATTC GTGTTCAAAA

2201 ATCCGGCCTT TATTCACATT CTTGCCCGCC TGATGAATGC TCATCCGGAA TTTCGTATGG CAATGAAAGA CGGTGAGCTG GTGATATGGG ATAGTGTTCA

TAGGCCGGAA ATAAGTGTAA GAACGGGCGG ACTACTTACG AGTAGGCCTT AAAGCATACC GTTACTTTCT GCCACTCGAC CACTATACCC TATCACAAGT

2301 CCCTTGTTAC ACCGTTTTCC ATGAGCAAAC TGAAACGTTT TCATCGCTCT GGAGTGAATA CCACGACGAT TTCCGGCAGT TTCTACACAT ATATTCGCAA

GGGAACAATG TGGCAAAAGG TACTCGTTTG ACTTTGCAAA AGTAGCGAGA CCTCACTTAT GGTGCTGCTA AAGGCCGTCA AAGATGTGTA TATAAGCGTT

2401 GATGTGGCGT GTTACGGTGA AAACCTGGCC TATTTCCCTA AAGGGTTTAT TGAGAATATG TTTTTCGTCT CAGCCAATCC CTGGGTGAGT TTCACCAGTT

CTACACCGCA CAATGCCACT TTTGGACCGG ATAAAGGGAT TTCCCAAATA ACTCTTATAC AAAAAGCAGA GTCGGTTAGG GACCCACTCA AAGTGGTCAA

2501 TTGATTTAAA CGTGGCCAAT ATGGACAACT TCTTCGCCCC CGTTTTCACC ATGGGCAAAT ATTATACGCA AGGCGACAAG GTGCTGATGC CGCTGGCGAT

AACTAAATTT GCACCGGTTA TACCTGTTGA AGAAGCGGGG GCAAAAGTGG TACCCGTTTA TAATATGCGT TCCGCTGTTC CACGACTACG GCGACCGCTA

2601 TCAGGTTCAT CATGCCGTTT GTGATGGCTT CCATGTCGGC AGAATGCTTA ATGAATTACA ACAGTACTGC GATGAGTGGC AGGGCGGGGC GTAATTTTTT

AGTCCAAGTA GTACGGCAAA CACTACCGAA GGTACAGCCG TCTTACGAAT TACTTAATGT TGTCATGACG CTACTCACCG TCCCGCCCCG CATTAAAAAA

2701 TAAGGCAGTT ATTGGTGCCC TTAAACGCCT GGGGTAATGA CTCTCTAGCT TGAGGCATCA AATAAAACGA AAGGCTCAGT CGAAAGACTG GGCCTTTCGT

ATTCCGTCAA TAACCACGGG AATTTGCGGA CCCCATTACT GAGAGATCGA ACTCCGTAGT TTATTTTGCT TTCCGAGTCA GCTTTCTGAC CCGGAAAGCA

2801 TTTATCTGTT GTTTGTCGGT GAACGCTCTC CTGAGTAGGA CAAATCCGCC CTCTAGATTA CGTGCAGTCG ATGATAAGCT GTCAAACATG AGAATTGTGC

AAATAGACAA CAAACAGCCA CTTGCGAGAG GACTCATCCT GTTTAGGCGG GAGATCTAAT GCACGTCAGC TACTATTCGA CAGTTTGTAC TCTTAACACG

2901 CTAATGAGTG AGCTAACTTA CATTAATTGC GTTGCGCTCA CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC

GATTACTCAC TCGATTGAAT GTAATTAACG CAACGCGAGT GACGGGCGAA AGGTCAGCCC TTTGGACAGC ACGGTCGACG TAATTACTTA GCCGGTTGCG

3001 GCGGGGAGAG GCGGTTTGCG TATTGGGCGC CAGGGTGGTT TTTCTTTTCA CCAGTGAGAC GGGCAACAGC TGATTGCCCT TCACCGCCTG GCCCTGAGAG

CGCCCCTCTC CGCCAAACGC ATAACCCGCG GTCCCACCAA AAAGAAAAGT GGTCACTCTG CCCGTTGTCG ACTAACGGGA AGTGGCGGAC CGGGACTCTC

3101 AGTTGCAGCA AGCGGTCCAC GCTGGTTTGC CCCAGCAGGC GAAAATCCTG TTTGATGGTG GTTAACGGCG GGATATAACA TGAGCTGTCT TCGGTATCGT

TCAACGTCGT TCGCCAGGTG CGACCAAACG GGGTCGTCCG CTTTTAGGAC AAACTACCAC CAATTGCCGC CCTATATTGT ACTCGACAGA AGCCATAGCA

3201 CGTATCCCAC TACCGAGATA TCCGCACCAA CGCGCAGCCC GGACTCGGTA ATGGCGCGCA TTGCGCCCAG CGCCATCTGA TCGTTGGCAA CCAGCATCGC

GCATAGGGTG ATGGCTCTAT AGGCGTGGTT GCGCGTCGGG CCTGAGCCAT TACCGCGCGT AACGCGGGTC GCGGTAGACT AGCAACCGTT GGTCGTAGCG

3301 AGTGGGAACG ATGCCCTCAT TCAGCATTTG CATGGTTTGT TGAAAACCGG ACATGGCACT CCAGTCGCCT TCCCGTTCCG CTATCGGCTG AATTTGATTG

TCACCCTTGC TACGGGAGTA AGTCGTAAAC GTACCAAACA ACTTTTGGCC TGTACCGTGA GGTCAGCGGA AGGGCAAGGC GATAGCCGAC TTAAACTAAC

3401 CGAGTGAGAT ATTTATGCCA GCCAGCCAGA CGCAGACGCG CCGAGACAGA ACTTAATGGG CCCGCTAACA GCGCGATTTG CTGGTGACCC AATGCGACCA

GCTCACTCTA TAAATACGGT CGGTCGGTCT GCGTCTGCGC GGCTCTGTCT TGAATTACCC GGGCGATTGT CGCGCTAAAC GACCACTGGG TTACGCTGGT

3501 GATGCTCCAC GCCCAGTCGC GTACCGTCTT CATGGGAGAA AATAATACTG TTGATGGGTG TCTGGTCAGA GACATCAAGA AATAACGCCG GAACATTAGT

CTACGAGGTG CGGGTCAGCG CATGGCAGAA GTACCCTCTT TTATTATGAC AACTACCCAC AGACCAGTCT CTGTAGTTCT TTATTGCGGC CTTGTAATCA

3601 GCAGGCAGCT TCCACAGCAA TGGCATCCTG GTCATCCAGC GGATAGTTAA TGATCAGCCC ACTGACGCGT TGCGCGAGAA GATTGTGCAC CGCCGCTTTA

CGTCCGTCGA AGGTGTCGTT ACCGTAGGAC CAGTAGGTCG CCTATCAATT ACTAGTCGGG TGACTGCGCA ACGCGCTCTT CTAACACGTG GCGGCGAAAT

3701 CAGGCTTCGA CGCCGCTTCG TTCTACCATC GACACCACCA CGCTGGCACC CAGTTGATCG GCGCGAGATT TAATCGCCGC GACAATTTGC GACGGCGCGT

GTCCGAAGCT GCGGCGAAGC AAGATGGTAG CTGTGGTGGT GCGACCGTGG GTCAACTAGC CGCGCTCTAA ATTAGCGGCG CTGTTAAACG CTGCCGCGCA

3801 GCAGGGCCAG ACTGGAGGTG GCAACGCCAA TCAGCAACGA CTGTTTGCCC GCCAGTTGTT GTGCCACGCG GTTGGGAATG TAATTCAGCT CCGCCATCGC

CGTCCCGGTC TGACCTCCAC CGTTGCGGTT AGTCGTTGCT GACAAACGGG CGGTCAACAA CACGGTGCGC CAACCCTTAC ATTAAGTCGA GGCGGTAGCG

3901 CGCTTCCACT TTTTCCCGCG TTTTCGCAGA AACGTGGCTG GCCTGGTTCA CCACGCGGGA AACGGTCTGA TAAGAGACAC CGGCATACTC TGCGACATCG

GCGAAGGTGA AAAAGGGCGC AAAAGCGTCT TTGCACCGAC CGGACCAAGT GGTGCGCCCT TTGCCAGACT ATTCTCTGTG GCCGTATGAG ACGCTGTAGC

4001 TATAACGTTA CTGGTTTCAC ATTCACCACC CTGAATTGAC TCTCTTCCGG GCGCTATCAT GCCATACCGC GAAAGGTTTT GCACCATTCG ATGGTGTCGG

ATATTGCAAT GACCAAAGTG TAAGTGGTGG GACTTAACTG AGAGAAGGCC CGCGATAGTA CGGTATGGCG CTTTCCAAAA CGTGGTAAGC TACCACAGCC

4101 AATTTCGGGC AGCGTTGGGT CCTGGCCACG GGTGCGCATG ATCTAGAGCT GCCTCGCGCG TTTCGGTGAT GACGGTGAAA ACCTCTGACA CATGCAGCTC

TTAAAGCCCG TCGCAACCCA GGACCGGTGC CCACGCGTAC TAGATCTCGA CGGAGCGCGC AAAGCCACTA CTGCCACTTT TGGAGACTGT GTACGTCGAG

4201 CCGGAGACGG TCACAGCTTG TCTGTAAGCG GATGCCGGGA GCAGACAAGC CCGTCAGGGC GCGTCAGCGG GTGTTGGCGG GTGTCGGGGC GCAGCCATGA

GGCCTCTGCC AGTGTCGAAC AGACATTCGC CTACGGCCCT CGTCTGTTCG GGCAGTCCCG CGCAGTCGCC CACAACCGCC CACAGCCCCG CGTCGGTACT

4301 CCCAGTCACG TAGCGATAGC GGAGTGTATA CTGGCTTAAC TATGCGGCAT CAGAGCAGAT TGTACTGAGA GTGCACCATA TGCGGTGTGA AATACCGCAC

GGGTCAGTGC ATCGCTATCG CCTCACATAT GACCGAATTG ATACGCCGTA GTCTCGTCTA ACATGACTCT CACGTGGTAT ACGCCACACT TTATGGCGTG

4401 AGATGCGTAA GGAGAAAATA CCGCATCAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG TCGTTCGGCT GCGGCGAGCG GTATCAGCTC

TCTACGCATT CCTCTTTTAT GGCGTAGTCC GCGAGAAGGC GAAGGAGCGA GTGACTGAGC GACGCGAGCC AGCAAGCCGA CGCCGCTCGC CATAGTCGAG

4501 ACTCAAAGGC GGTAATACGG TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG CCAGCAAAAG GCCAGGAACC GTAAAAAGGC

TGAGTTTCCG CCATTATGCC AATAGGTGTC TTAGTCCCCT ATTGCGTCCT TTCTTGTACA CTCGTTTTCC GGTCGTTTTC CGGTCCTTGG CATTTTTCCG

4601 CGCGTTGCTG GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG AGGTGGCGAA ACCCGACAGG ACTATAAAGA

GCGCAACGAC CGCAAAAAGG TATCCGAGGC GGGGGGACTG CTCGTAGTGT TTTTAGCTGC GAGTTCAGTC TCCACCGCTT TGGGCTGTCC TGATATTTCT

4701 TACCAGGCGT TTCCCCCTGG AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG

ATGGTCCGCA AAGGGGGACC TTCGAGGGAG CACGCGAGAG GACAAGGCTG GGACGGCGAA TGGCCTATGG ACAGGCGGAA AGAGGGAAGC CCTTCGCACC

4801 CGCTTTCTCA TAGCTCACGC TGTAGGTATC TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC CCCGTTCAGC CCGACCGCTG

GCGAAAGAGT ATCGAGTGCG ACATCCATAG AGTCAAGCCA CATCCAGCAA GCGAGGTTCG ACCCGACACA CGTGCTTGGG GGGCAAGTCG GGCTGGCGAC

4901 CGCCTTATCC GGTAACTATC GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT GGCAGCAGCC ACTGGTAACA GGATTAGCAG AGCGAGGTAT

GCGGAATAGG CCATTGATAG CAGAACTCAG GTTGGGCCAT TCTGTGCTGA ATAGCGGTGA CCGTCGTCGG TGACCATTGT CCTAATCGTC TCGCTCCATA

5001 GTAGGCGGTG CTACAGAGTT CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGGACA GTATTTGGTA TCTGCGCTCT GCTGAAGCCA GTTACCTTCG

CATCCGCCAC GATGTCTCAA GAACTTCACC ACCGGATTGA TGCCGATGTG ATCTTCCTGT CATAAACCAT AGACGCGAGA CGACTTCGGT CAATGGAAGC

5101 GAAAAAGAGT TGGTAGCTCT TGATCCGGCA AACAAACCAC CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGCAGATT ACGCGCAGAA AAAAAGGATC

CTTTTTCTCA ACCATCGAGA ACTAGGCCGT TTGTTTGGTG GCGACCATCG CCACCAAAAA AACAAACGTT CGTCGTCTAA TGCGCGTCTT TTTTTCCTAG

5201 TCAAGAAGAT CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG TTAAGGGATT TTGGTCATGA GATTATCAAA AAGGATCTTC

AGTTCTTCTA GGAAACTAGA AAAGATGCCC CAGACTGCGA GTCACCTTGC TTTTGAGTGC AATTCCCTAA AACCAGTACT CTAATAGTTT TTCCTAGAAG

5301 ACCTAGATCC TTTTAAATTA AAAATGAAGT TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTACCA ATGCTTAATC AGTGAGGCAC

TGGATCTAGG AAAATTTAAT TTTTACTTCA AAATTTAGTT AGATTTCATA TATACTCATT TGAACCAGAC TGTCAATGGT TACGAATTAG TCACTCCGTG

5401 CTATCTCAGC GATCTGTCTA TTTCGTTCAT CCATAGTTGC CTGACTCCCC GTCGTGTAGA TAACTACGAT ACGGGAGGGC TTACCATCTG GCCCCAGTGC

GATAGAGTCG CTAGACAGAT AAAGCAAGTA GGTATCAACG GACTGAGGGG CAGCACATCT ATTGATGCTA TGCCCTCCCG AATGGTAGAC CGGGGTCACG

5501 TGCAATGATA CCGCGAGACC CACGCTCACC GGCTCCAGAT TTATCAGCAA TAAACCAGCC AGCCGGAAGG GCCGAGCGCA GAAGTGGTCC TGCAACTTTA

ACGTTACTAT GGCGCTCTGG GTGCGAGTGG CCGAGGTCTA AATAGTCGTT ATTTGGTCGG TCGGCCTTCC CGGCTCGCGT CTTCACCAGG ACGTTGAAAT

5601 TCCGCCTCCA TCCAGTCTAT TAATTGTTGC CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT AATAGTTTGC GCAACGTTGT TGCCATTGCT ACAGGCATCG

AGGCGGAGGT AGGTCAGATA ATTAACAACG GCCCTTCGAT CTCATTCATC AAGCGGTCAA TTATCAAACG CGTTGCAACA ACGGTAACGA TGTCCGTAGC

5701 TGGTGTCACG CTCGTCGTTT GGTATGGCTT CATTCAGCTC CGGTTCCCAA CGATCAAGGC GAGTTACATG ATCCCCCATG TTGTGCAAAA AAGCGGTTAG

ACCACAGTGC GAGCAGCAAA CCATACCGAA GTAAGTCGAG GCCAAGGGTT GCTAGTTCCG CTCAATGTAC TAGGGGGTAC AACACGTTTT TTCGCCAATC

5801 CTCCTTCGGT CCTCCGATCG TTGTCAGAAG TAAGTTGGCC GCAGTGTTAT CACTCATGGT TATGGCAGCA CTGCATAATT CTCTTACTGT CATGCCATCC

GAGGAAGCCA GGAGGCTAGC AACAGTCTTC ATTCAACCGG CGTCACAATA GTGAGTACCA ATACCGTCGT GACGTATTAA GAGAATGACA GTACGGTAGG

5901 GTAAGATGCT TTTCTGTGAC TGGTGAGTAC TCAACCAAGT CATTCTGAGA ATAGTGTATG CGGCGACCGA GTTGCTCTTG CCCGGCGTCA ATACGGGATA

CATTCTACGA AAAGACACTG ACCACTCATG AGTTGGTTCA GTAAGACTCT TATCACATAC GCCGCTGGCT CAACGAGAAC GGGCCGCAGT TATGCCCTAT

6001 ATACCGCGCC ACATAGCAGA ACTTTAAAAG TGCTCATCAT TGGAAAACGT TCTTCGGGGC GAAAACTCTC AAGGATCTTA CCGCTGTTGA GATCCAGTTC

TATGGCGCGG TGTATCGTCT TGAAATTTTC ACGAGTAGTA ACCTTTTGCA AGAAGCCCCG CTTTTGAGAG TTCCTAGAAT GGCGACAACT CTAGGTCAAG

6101 GATGTAACCC ACTCGTGCAC CCAACTGATC TTCAGCATCT TTTACTTTCA CCAGCGTTTC TGGGTGAGCA AAAACAGGAA GGCAAAATGC CGCAAAAAAG

CTACATTGGG TGAGCACGTG GGTTGACTAG AAGTCGTAGA AAATGAAAGT GGTCGCAAAG ACCCACTCGT TTTTGTCCTT CCGTTTTACG GCGTTTTTTC

6201 GGAATAAGGG CGACACGGAA ATGTTGAATA CTCATACTCT TCCTTTTTCA ATATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC GGATACATAT

CCTTATTCCC GCTGTGCCTT TACAACTTAT GAGTATGAGA AGGAAAAAGT TATAATAACT TCGTAAATAG TCCCAATAAC AGAGTACTCG CCTATGTATA

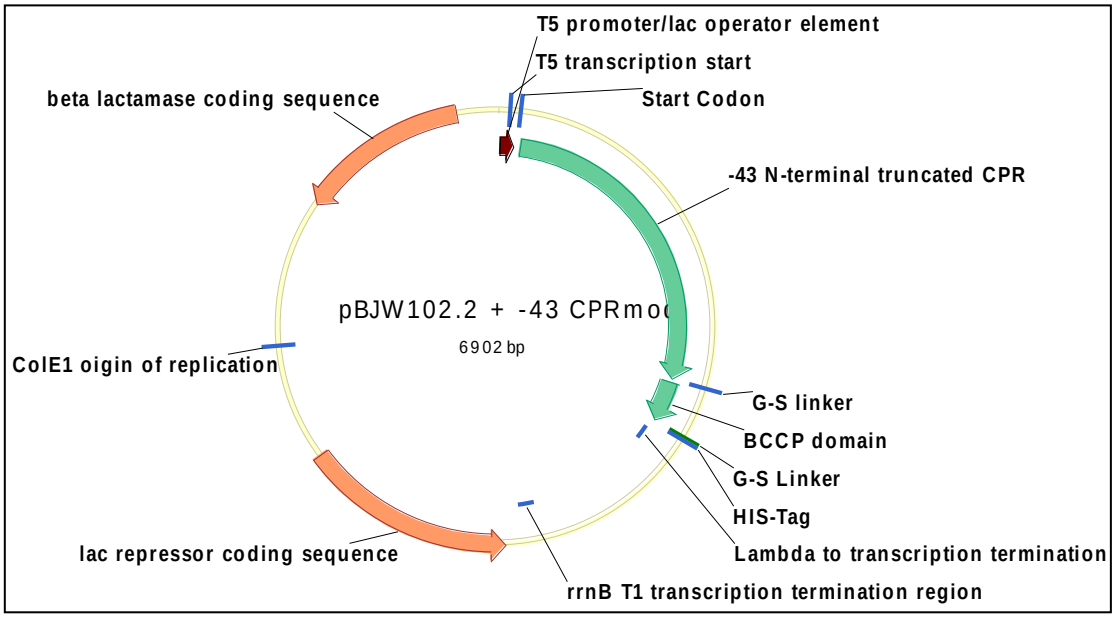
6301 TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC

AACTTACATA AATCTTTTTA TTTGTTTATC CCCAAGGCGC GTGTAAAGGG GCTTTTCACG GTGGACTGCA GATTCTTTGG TAATAATAGT ACTGTAATTG

6401 CTATAAAAAT AGGCGTATCA CGAGGCCCTT TCGTCTTCAC

GATATTTTTA TCCGCATAGT GCTCCGGGAA AGCAGAAGTG

CPR vector diagram



CPR vector sequence

1 CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG

GAGCTCTTTA GTATTTTTTA AATAAACGAA ACACTCGCCT ATTGTTAATA TTATCTAAGT TAACACTCGC CTATTGTTAA AGTGTGTCTT AAGTAATTTC

101 AGGAGAAATT AACTATGGCA CTTACGGCCT TCAGAAAGAA AAAAGAAGAA GTCCCCGAGT TCACCAAAAT TCAGACATTG ACCTCCTCTG TCAGAGAGAG

TCCTCTTTAA TTGATACCGT GAATGCCGGA AGTCTTTCTT TTTTCTTCTT CAGGGGCTCA AGTGGTTTTA AGTCTGTAAC TGGAGGAGAC AGTCTCTCTC

201 CAGCTTTGTG GAAAAGATGA AGAAAACGGG GAGGAACATC ATCGTGTTCT ACGGCTCCCA GACGGGGACT GCAGAGGAGT TTGCCAACCG CCTGTCCAAG

GTCGAAACAC CTTTTCTACT TCTTTTGCCC CTCCTTGTAG TAGCACAAGA TGCCGAGGGT CTGCCCCTGA CGTCTCCTCA AACGGTTGGC GGACAGGTTC

301 GACGCCCACC GCTACGGGAT GCGAGGCATG TCAGCGGACC CTGAGGAGTA TGACCTGGCC GACCTGAGCA GCCTGCCAGA GATCGACAAC GCCCTGGTGG

CTGCGGGTGG CGATGCCCTA CGCTCCGTAC AGTCGCCTGG GACTCCTCAT ACTGGACCGG CTGGACTCGT CGGACGGTCT CTAGCTGTTG CGGGACCACC

401 TTTTCTGCAT GGCCACCTAC GGTGAGGGAG ACCCCACCGA CAATGCCCAG GACTTCTACG ACTGGCTGCA GGAGACAGAC GTGGATCTCT CTGGGGTCAA

AAAAGACGTA CCGGTGGATG CCACTCCCTC TGGGGTGGCT GTTACGGGTC CTGAAGATGC TGACCGACGT CCTCTGTCTG CACCTAGAGA GACCCCAGTT

501 GTTCGCGGTG TTTGGTCTTG GGAACAAGAC CTACGAGCAC TTCAATGCCA TGGGCAAGTA CGTGGACAAG CGGCTGGAGC AGCTCGGCGC CCAGCGCATC

CAAGCGCCAC AAACCAGAAC CCTTGTTCTG GATGCTCGTG AAGTTACGGT ACCCGTTCAT GCACCTGTTC GCCGACCTCG TCGAGCCGCG GGTCGCGTAG

601 TTTGAGCTGG GGTTGGGCGA CGACGATGGG AACTTGGAGG AGGACTTCAT CACCTGGCGA GAGCAGTTCT GGCCGGCCGT GTGTGAACAC TTTGGGGTGG

AAACTCGACC CCAACCCGCT GCTGCTACCC TTGAACCTCC TCCTGAAGTA GTGGACCGCT CTCGTCAAGA CCGGCCGGCA CACACTTGTG AAACCCCACC

701 AAGCCACTGG CGAGGAGTCC AGCATTCGCC AGTACGAGCT TGTGGTCCAC ACCGACATAG ATGCGGCCAA GGTGTACATG GGGGAGATGG GCCGGCTGAA

TTCGGTGACC GCTCCTCAGG TCGTAAGCGG TCATGCTCGA ACACCAGGTG TGGCTGTATC TACGCCGGTT CCACATGTAC CCCCTCTACC CGGCCGACTT

801 GAGCTACGAG AACCAGAAGC CCCCCTTTGA TGCCAAGAAT CCGTTCCTGG CTGCAGTCAC CACCAACCGG AAGCTGAACC AGGGAACCGA GCGCCACCTC

CTCGATGCTC TTGGTCTTCG GGGGGAAACT ACGGTTCTTA GGCAAGGACC GACGTCAGTG GTGGTTGGCC TTCGACTTGG TCCCTTGGCT CGCGGTGGAG

901 ATGCACCTGG AATTGGACAT CTCGGACTCC AAAATCAGGT ATGAATCTGG GGACCACGTG GCTGTGTACC CAGCCAACGA CTCTGCTCTC GTCAACCAGC

TACGTGGACC TTAACCTGTA GAGCCTGAGG TTTTAGTCCA TACTTAGACC CCTGGTGCAC CGACACATGG GTCGGTTGCT GAGACGAGAG CAGTTGGTCG

1001 TGGGCAAAAT CCTGGGTGCC GACCTGGACG TCGTCATGTC CCTGAACAAC CTGGATGAGG AGTCCAACAA GAAGCACCCA TTCCCGTGCC CTACGTCCTA

ACCCGTTTTA GGACCCACGG CTGGACCTGC AGCAGTACAG GGACTTGTTG GACCTACTCC TCAGGTTGTT CTTCGTGGGT AAGGGCACGG GATGCAGGAT

1101 CCGCACGGCC CTCACCTACT ACCTGGACAT CACCAACCCG CCGCGTACCA ACGTGCTGTA CGAGCTGGCG CAGTACGCCT CGGAGCCCTC GGAGCAGGAG

GGCGTGCCGG GAGTGGATGA TGGACCTGTA GTGGTTGGGC GGCGCATGGT TGCACGACAT GCTCGACCGC GTCATGCGGA GCCTCGGGAG CCTCGTCCTC

1201 CTGCTGCGCA AGATGGCCTC CTCCTCCGGC GAGGGCAAGG AGCTGTACCT GAGCTGGGTG GTGGAGGCCC GGAGGCACAT CCTGGCCATC CTGCAGGACT

GACGACGCGT TCTACCGGAG GAGGAGGCCG CTCCCGTTCC TCGACATGGA CTCGACCCAC CACCTCCGGG CCTCCGTGTA GGACCGGTAG GACGTCCTGA

1301 GCCCGTCCCT GCGGCCCCCC ATCGACCACC TGTGTGAGCT GCTGCCGCGC CTGCAGGCCC GCTACTACTC CATCGCCTCA TCCTCCAAGG TCCACCCCAA

CGGGCAGGGA CGCCGGGGGG TAGCTGGTGG ACACACTCGA CGACGGCGCG GACGTCCGGG CGATGATGAG GTAGCGGAGT AGGAGGTTCC AGGTGGGGTT

1401 CTCTGTGCAC ATCTGTGCGG TGGTTGTGGA GTACGAGACC AAGGCCGGCC GCATCAACAA GGGCGTGGCC ACCAACTGGC TGCGGGCCAA GGAGCCTGCC

GAGACACGTG TAGACACGCC ACCAACACCT CATGCTCTGG TTCCGGCCGG CGTAGTTGTT CCCGCACCGG TGGTTGACCG ACGCCCGGTT CCTCGGACGG

1501 GGGGAGAACG GCGGCCGTGC GCTGGTGCCC ATGTTCGTGC GCAAGTCCCA GTTCCGCCTG CCCTTCAAGG CCACCACGCC TGTCATCATG GTGGGCCCCG

CCCCTCTTGC CGCCGGCACG CGACCACGGG TACAAGCACG CGTTCAGGGT CAAGGCGGAC GGGAAGTTCC GGTGGTGCGG ACAGTAGTAC CACCCGGGGC

1601 GCACCGGGGT GGCACCCTTC ATAGGCTTCA TCCAGGAGCG GGCCTGGCTG CGACAGCAGG GCAAGGAGGT GGGGGAGACG CTGCTGTACT ACGGCTGCCG

CGTGGCCCCA CCGTGGGAAG TATCCGAAGT AGGTCCTCGC CCGGACCGAC GCTGTCGTCC CGTTCCTCCA CCCCCTCTGC GACGACATGA TGCCGACGGC

1701 CCGCTCGGAT GAGGACTACC TGTACCGGGA GGAGCTGGCG CAGTTCCACA GGGACGGTGC GCTCACCCAG CTCAACGTGG CCTTCTCCCG GGAGCAGTCC

GGCGAGCCTA CTCCTGATGG ACATGGCCCT CCTCGACCGC GTCAAGGTGT CCCTGCCACG CGAGTGGGTC GAGTTGCACC GGAAGAGGGC CCTCGTCAGG

1801 CACAAGGTCT ACGTCCAGCA CCTGCTAAAG CAAGACCGAG AGCACCTGTG GAAGTTGATC GAAGGCGGTG CCCACATCTA CGTCTGTGGG GATGCACGGA

GTGTTCCAGA TGCAGGTCGT GGACGATTTC GTTCTGGCTC TCGTGGACAC CTTCAACTAG CTTCCGCCAC GGGTGTAGAT GCAGACACCC CTACGTGCCT

1901 ACATGGCCAG GGATGTGCAG AACACCTTCT ACGACATCGT GGCTGAGCTC GGGGCCATGG AGCACGCGCA GGCGGTGGAC TACATCAAGA AACTGATGAC

TGTACCGGTC CCTACACGTC TTGTGGAAGA TGCTGTAGCA CCGACTCGAG CCCCGGTACC TCGTGCGCGT CCGCCACCTG ATGTAGTTCT TTGACTACTG

2001 CAAGGGCCGC TACTCCCTGG ACGTGTGGAG CGGGGGTGGC AGCGGTTCTG GCGCAGCAGC GGAAATCAGT GGTCACATCG TACGTTCCCC GATGGTTGGT

GTTCCCGGCG ATGAGGGACC TGCACACCTC GCCCCCACCG TCGCCAAGAC CGCGTCGTCG CCTTTAGTCA CCAGTGTAGC ATGCAAGGGG CTACCAACCA

2101 ACTTTCTACC GCACCCCAAG CCCGGACGCA AAAGCGTTCA TCGAAGTGGG TCAGAAAGTC AACGTGGGCG ATACCCTGTG CATCGTTGAA GCCATGAAAA

TGAAAGATGG CGTGGGGTTC GGGCCTGCGT TTTCGCAAGT AGCTTCACCC AGTCTTTCAG TTGCACCCGC TATGGGACAC GTAGCAACTT CGGTACTTTT

2201 TGATGAACCA GATCGAAGCG GACAAATCCG GTACCGTGAA AGCAATTCTG GTCGAAAGTG GACAACCGGT AGAATTTGAC GAGCCGCTGG TCGTCATCGA

ACTACTTGGT CTAGCTTCGC CTGTTTAGGC CATGGCACTT TCGTTAAGAC CAGCTTTCAC CTGTTGGCCA TCTTAAACTG CTCGGCGACC AGCAGTAGCT

2301 GGGTGGCAGC GGTTCTGGCC ACCATCACCA TCACCATAAG CTTAATTAGC TGAGCTTGGA CTCCTGTTGA TAGATCCAGT AATGACCTCA GAACTCCATC

CCCACCGTCG CCAAGACCGG TGGTAGTGGT AGTGGTATTC GAATTAATCG ACTCGAACCT GAGGACAACT ATCTAGGTCA TTACTGGAGT CTTGAGGTAG

2401 TGGATTTGTT CAGAACGCTC GGTTGCCGCC GGGCGTTTTT TATTGGTGAG AATCCAAGCT AGCTTGGCGA GATTTTCAGG AGCTAAGGAA GCTAAAATGG

ACCTAAACAA GTCTTGCGAG CCAACGGCGG CCCGCAAAAA ATAACCACTC TTAGGTTCGA TCGAACCGCT CTAAAAGTCC TCGATTCCTT CGATTTTACC

2501 AGAAAAAAAT CACTGGATAT ACCACCGTTG ATATATCCCA ATGGCATCGT AAAGAACATT TTGAGGCATT TCAGTCAGTT GCTCAATGTA CCTATAACCA

TCTTTTTTTA GTGACCTATA TGGTGGCAAC TATATAGGGT TACCGTAGCA TTTCTTGTAA AACTCCGTAA AGTCAGTCAA CGAGTTACAT GGATATTGGT

2601 GACCGTTCAG CTGGATATTA CGGCCTTTTT AAAGACCGTA AAGAAAAATA AGCACAAGTT TTATCCGGCC TTTATTCACA TTCTTGCCCG CCTGATGAAT

CTGGCAAGTC GACCTATAAT GCCGGAAAAA TTTCTGGCAT TTCTTTTTAT TCGTGTTCAA AATAGGCCGG AAATAAGTGT AAGAACGGGC GGACTACTTA

2701 GCTCATCCGG AATTTCGTAT GGCAATGAAA GACGGTGAGC TGGTGATATG GGATAGTGTT CACCCTTGTT ACACCGTTTT CCATGAGCAA ACTGAAACGT

CGAGTAGGCC TTAAAGCATA CCGTTACTTT CTGCCACTCG ACCACTATAC CCTATCACAA GTGGGAACAA TGTGGCAAAA GGTACTCGTT TGACTTTGCA

2801 TTTCATCGCT CTGGAGTGAA TACCACGACG ATTTCCGGCA GTTTCTACAC ATATATTCGC AAGATGTGGC GTGTTACGGT GAAAACCTGG CCTATTTCCC

AAAGTAGCGA GACCTCACTT ATGGTGCTGC TAAAGGCCGT CAAAGATGTG TATATAAGCG TTCTACACCG CACAATGCCA CTTTTGGACC GGATAAAGGG

2901 TAAAGGGTTT ATTGAGAATA TGTTTTTCGT CTCAGCCAAT CCCTGGGTGA GTTTCACCAG TTTTGATTTA AACGTGGCCA ATATGGACAA CTTCTTCGCC

ATTTCCCAAA TAACTCTTAT ACAAAAAGCA GAGTCGGTTA GGGACCCACT CAAAGTGGTC AAAACTAAAT TTGCACCGGT TATACCTGTT GAAGAAGCGG

3001 CCCGTTTTCA CCATGGGCAA ATATTATACG CAAGGCGACA AGGTGCTGAT GCCGCTGGCG ATTCAGGTTC ATCATGCCGT TTGTGATGGC TTCCATGTCG

GGGCAAAAGT GGTACCCGTT TATAATATGC GTTCCGCTGT TCCACGACTA CGGCGACCGC TAAGTCCAAG TAGTACGGCA AACACTACCG AAGGTACAGC

3101 GCAGAATGCT TAATGAATTA CAACAGTACT GCGATGAGTG GCAGGGCGGG GCGTAATTTT TTTAAGGCAG TTATTGGTGC CCTTAAACGC CTGGGGTAAT

CGTCTTACGA ATTACTTAAT GTTGTCATGA CGCTACTCAC CGTCCCGCCC CGCATTAAAA AAATTCCGTC AATAACCACG GGAATTTGCG GACCCCATTA

3201 GACTCTCTAG CTTGAGGCAT CAAATAAAAC GAAAGGCTCA GTCGAAAGAC TGGGCCTTTC GTTTTATCTG TTGTTTGTCG GTGAACGCTC TCCTGAGTAG

CTGAGAGATC GAACTCCGTA GTTTATTTTG CTTTCCGAGT CAGCTTTCTG ACCCGGAAAG CAAAATAGAC AACAAACAGC CACTTGCGAG AGGACTCATC

3301 GACAAATCCG CCCTCTAGAT TACGTGCAGT CGATGATAAG CTGTCAAACA TGAGAATTGT GCCTAATGAG TGAGCTAACT TACATTAATT GCGTTGCGCT

CTGTTTAGGC GGGAGATCTA ATGCACGTCA GCTACTATTC GACAGTTTGT ACTCTTAACA CGGATTACTC ACTCGATTGA ATGTAATTAA CGCAACGCGA

3401 CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCCAGGGTGG

GTGACGGGCG AAAGGTCAGC CCTTTGGACA GCACGGTCGA CGTAATTACT TAGCCGGTTG CGCGCCCCTC TCCGCCAAAC GCATAACCCG CGGTCCCACC

3501 TTTTTCTTTT CACCAGTGAG ACGGGCAACA GCTGATTGCC CTTCACCGCC TGGCCCTGAG AGAGTTGCAG CAAGCGGTCC ACGCTGGTTT GCCCCAGCAG

AAAAAGAAAA GTGGTCACTC TGCCCGTTGT CGACTAACGG GAAGTGGCGG ACCGGGACTC TCTCAACGTC GTTCGCCAGG TGCGACCAAA CGGGGTCGTC

3601 GCGAAAATCC TGTTTGATGG TGGTTAACGG CGGGATATAA CATGAGCTGT CTTCGGTATC GTCGTATCCC ACTACCGAGA TATCCGCACC AACGCGCAGC

CGCTTTTAGG ACAAACTACC ACCAATTGCC GCCCTATATT GTACTCGACA GAAGCCATAG CAGCATAGGG TGATGGCTCT ATAGGCGTGG TTGCGCGTCG

3701 CCGGACTCGG TAATGGCGCG CATTGCGCCC AGCGCCATCT GATCGTTGGC AACCAGCATC GCAGTGGGAA CGATGCCCTC ATTCAGCATT TGCATGGTTT

GGCCTGAGCC ATTACCGCGC GTAACGCGGG TCGCGGTAGA CTAGCAACCG TTGGTCGTAG CGTCACCCTT GCTACGGGAG TAAGTCGTAA ACGTACCAAA

3801 GTTGAAAACC GGACATGGCA CTCCAGTCGC CTTCCCGTTC CGCTATCGGC TGAATTTGAT TGCGAGTGAG ATATTTATGC CAGCCAGCCA GACGCAGACG

CAACTTTTGG CCTGTACCGT GAGGTCAGCG GAAGGGCAAG GCGATAGCCG ACTTAAACTA ACGCTCACTC TATAAATACG GTCGGTCGGT CTGCGTCTGC

3901 CGCCGAGACA GAACTTAATG GGCCCGCTAA CAGCGCGATT TGCTGGTGAC CCAATGCGAC CAGATGCTCC ACGCCCAGTC GCGTACCGTC TTCATGGGAG

GCGGCTCTGT CTTGAATTAC CCGGGCGATT GTCGCGCTAA ACGACCACTG GGTTACGCTG GTCTACGAGG TGCGGGTCAG CGCATGGCAG AAGTACCCTC

4001 AAAATAATAC TGTTGATGGG TGTCTGGTCA GAGACATCAA GAAATAACGC CGGAACATTA GTGCAGGCAG CTTCCACAGC AATGGCATCC TGGTCATCCA

TTTTATTATG ACAACTACCC ACAGACCAGT CTCTGTAGTT CTTTATTGCG GCCTTGTAAT CACGTCCGTC GAAGGTGTCG TTACCGTAGG ACCAGTAGGT

4101 GCGGATAGTT AATGATCAGC CCACTGACGC GTTGCGCGAG AAGATTGTGC ACCGCCGCTT TACAGGCTTC GACGCCGCTT CGTTCTACCA TCGACACCAC

CGCCTATCAA TTACTAGTCG GGTGACTGCG CAACGCGCTC TTCTAACACG TGGCGGCGAA ATGTCCGAAG CTGCGGCGAA GCAAGATGGT AGCTGTGGTG

4201 CACGCTGGCA CCCAGTTGAT CGGCGCGAGA TTTAATCGCC GCGACAATTT GCGACGGCGC GTGCAGGGCC AGACTGGAGG TGGCAACGCC AATCAGCAAC

GTGCGACCGT GGGTCAACTA GCCGCGCTCT AAATTAGCGG CGCTGTTAAA CGCTGCCGCG CACGTCCCGG TCTGACCTCC ACCGTTGCGG TTAGTCGTTG

4301 GACTGTTTGC CCGCCAGTTG TTGTGCCACG CGGTTGGGAA TGTAATTCAG CTCCGCCATC GCCGCTTCCA CTTTTTCCCG CGTTTTCGCA GAAACGTGGC

CTGACAAACG GGCGGTCAAC AACACGGTGC GCCAACCCTT ACATTAAGTC GAGGCGGTAG CGGCGAAGGT GAAAAAGGGC GCAAAAGCGT CTTTGCACCG

4401 TGGCCTGGTT CACCACGCGG GAAACGGTCT GATAAGAGAC ACCGGCATAC TCTGCGACAT CGTATAACGT TACTGGTTTC ACATTCACCA CCCTGAATTG

ACCGGACCAA GTGGTGCGCC CTTTGCCAGA CTATTCTCTG TGGCCGTATG AGACGCTGTA GCATATTGCA ATGACCAAAG TGTAAGTGGT GGGACTTAAC

4501 ACTCTCTTCC GGGCGCTATC ATGCCATACC GCGAAAGGTT TTGCACCATT CGATGGTGTC GGAATTTCGG GCAGCGTTGG GTCCTGGCCA CGGGTGCGCA

TGAGAGAAGG CCCGCGATAG TACGGTATGG CGCTTTCCAA AACGTGGTAA GCTACCACAG CCTTAAAGCC CGTCGCAACC CAGGACCGGT GCCCACGCGT

4601 TGATCTAGAG CTGCCTCGCG CGTTTCGGTG ATGACGGTGA AAACCTCTGA CACATGCAGC TCCCGGAGAC GGTCACAGCT TGTCTGTAAG CGGATGCCGG

ACTAGATCTC GACGGAGCGC GCAAAGCCAC TACTGCCACT TTTGGAGACT GTGTACGTCG AGGGCCTCTG CCAGTGTCGA ACAGACATTC GCCTACGGCC

4701 GAGCAGACAA GCCCGTCAGG GCGCGTCAGC GGGTGTTGGC GGGTGTCGGG GCGCAGCCAT GACCCAGTCA CGTAGCGATA GCGGAGTGTA TACTGGCTTA

CTCGTCTGTT CGGGCAGTCC CGCGCAGTCG CCCACAACCG CCCACAGCCC CGCGTCGGTA CTGGGTCAGT GCATCGCTAT CGCCTCACAT ATGACCGAAT

4801 ACTATGCGGC ATCAGAGCAG ATTGTACTGA GAGTGCACCA TATGCGGTGT GAAATACCGC ACAGATGCGT AAGGAGAAAA TACCGCATCA GGCGCTCTTC

TGATACGCCG TAGTCTCGTC TAACATGACT CTCACGTGGT ATACGCCACA CTTTATGGCG TGTCTACGCA TTCCTCTTTT ATGGCGTAGT CCGCGAGAAG

4901 CGCTTCCTCG CTCACTGACT CGCTGCGCTC GGTCGTTCGG CTGCGGCGAG CGGTATCAGC TCACTCAAAG GCGGTAATAC GGTTATCCAC AGAATCAGGG

GCGAAGGAGC GAGTGACTGA GCGACGCGAG CCAGCAAGCC GACGCCGCTC GCCATAGTCG AGTGAGTTTC CGCCATTATG CCAATAGGTG TCTTAGTCCC

5001 GATAACGCAG GAAAGAACAT GTGAGCAAAA GGCCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT CCATAGGCTC CGCCCCCCTG

CTATTGCGTC CTTTCTTGTA CACTCGTTTT CCGGTCGTTT TCCGGTCCTT GGCATTTTTC CGGCGCAACG ACCGCAAAAA GGTATCCGAG GCGGGGGGAC

5101 ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC

TGCTCGTAGT GTTTTTAGCT GCGAGTTCAG TCTCCACCGC TTTGGGCTGT CCTGATATTT CTATGGTCCG CAAAGGGGGA CCTTCGAGGG AGCACGCGAG

5201 TCCTGTTCCG ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT GGCGCTTTCT CATAGCTCAC GCTGTAGGTA TCTCAGTTCG

AGGACAAGGC TGGGACGGCG AATGGCCTAT GGACAGGCGG AAAGAGGGAA GCCCTTCGCA CCGCGAAAGA GTATCGAGTG CGACATCCAT AGAGTCAAGC

5301 GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGCGCCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG

CACATCCAGC AAGCGAGGTT CGACCCGACA CACGTGCTTG GGGGGCAAGT CGGGCTGGCG ACGCGGAATA GGCCATTGAT AGCAGAACTC AGGTTGGGCC

5401 TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC AGAGCGAGGT ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCCTAA

ATTCTGTGCT GAATAGCGGT GACCGTCGTC GGTGACCATT GTCCTAATCG TCTCGCTCCA TACATCCGCC ACGATGTCTC AAGAACTTCA CCACCGGATT

5501 CTACGGCTAC ACTAGAAGGA CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA GTTGGTAGCT CTTGATCCGG CAAACAAACC

GATGCCGATG TGATCTTCCT GTCATAAACC ATAGACGCGA GACGACTTCG GTCAATGGAA GCCTTTTTCT CAACCATCGA GAACTAGGCC GTTTGTTTGG

5601 ACCGCTGGTA GCGGTGGTTT TTTTGTTTGC AAGCAGCAGA TTACGCGCAG AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGGTCTGACG

TGGCGACCAT CGCCACCAAA AAAACAAACG TTCGTCGTCT AATGCGCGTC TTTTTTTCCT AGAGTTCTTC TAGGAAACTA GAAAAGATGC CCCAGACTGC

5701 CTCAGTGGAA CGAAAACTCA CGTTAAGGGA TTTTGGTCAT GAGATTATCA AAAAGGATCT TCACCTAGAT CCTTTTAAAT TAAAAATGAA GTTTTAAATC

GAGTCACCTT GCTTTTGAGT GCAATTCCCT AAAACCAGTA CTCTAATAGT TTTTCCTAGA AGTGGATCTA GGAAAATTTA ATTTTTACTT CAAAATTTAG

5801 AATCTAAAGT ATATATGAGT AAACTTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC ACCTATCTCA GCGATCTGTC TATTTCGTTC ATCCATAGTT

TTAGATTTCA TATATACTCA TTTGAACCAG ACTGTCAATG GTTACGAATT AGTCACTCCG TGGATAGAGT CGCTAGACAG ATAAAGCAAG TAGGTATCAA

5901 GCCTGACTCC CCGTCGTGTA GATAACTACG ATACGGGAGG GCTTACCATC TGGCCCCAGT GCTGCAATGA TACCGCGAGA CCCACGCTCA CCGGCTCCAG

CGGACTGAGG GGCAGCACAT CTATTGATGC TATGCCCTCC CGAATGGTAG ACCGGGGTCA CGACGTTACT ATGGCGCTCT GGGTGCGAGT GGCCGAGGTC

6001 ATTTATCAGC AATAAACCAG CCAGCCGGAA GGGCCGAGCG CAGAAGTGGT CCTGCAACTT TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC

TAAATAGTCG TTATTTGGTC GGTCGGCCTT CCCGGCTCGC GTCTTCACCA GGACGTTGAA ATAGGCGGAG GTAGGTCAGA TAATTAACAA CGGCCCTTCG

6101 TAGAGTAAGT AGTTCGCCAG TTAATAGTTT GCGCAACGTT GTTGCCATTG CTACAGGCAT CGTGGTGTCA CGCTCGTCGT TTGGTATGGC TTCATTCAGC

ATCTCATTCA TCAAGCGGTC AATTATCAAA CGCGTTGCAA CAACGGTAAC GATGTCCGTA GCACCACAGT GCGAGCAGCA AACCATACCG AAGTAAGTCG

6201 TCCGGTTCCC AACGATCAAG GCGAGTTACA TGATCCCCCA TGTTGTGCAA AAAAGCGGTT AGCTCCTTCG GTCCTCCGAT CGTTGTCAGA AGTAAGTTGG

AGGCCAAGGG TTGCTAGTTC CGCTCAATGT ACTAGGGGGT ACAACACGTT TTTTCGCCAA TCGAGGAAGC CAGGAGGCTA GCAACAGTCT TCATTCAACC

6301 CCGCAGTGTT ATCACTCATG GTTATGGCAG CACTGCATAA TTCTCTTACT GTCATGCCAT CCGTAAGATG CTTTTCTGTG ACTGGTGAGT ACTCAACCAA

GGCGTCACAA TAGTGAGTAC CAATACCGTC GTGACGTATT AAGAGAATGA CAGTACGGTA GGCATTCTAC GAAAAGACAC TGACCACTCA TGAGTTGGTT

6401 GTCATTCTGA GAATAGTGTA TGCGGCGACC GAGTTGCTCT TGCCCGGCGT CAATACGGGA TAATACCGCG CCACATAGCA GAACTTTAAA AGTGCTCATC

CAGTAAGACT CTTATCACAT ACGCCGCTGG CTCAACGAGA ACGGGCCGCA GTTATGCCCT ATTATGGCGC GGTGTATCGT CTTGAAATTT TCACGAGTAG

6501 ATTGGAAAAC GTTCTTCGGG GCGAAAACTC TCAAGGATCT TACCGCTGTT GAGATCCAGT TCGATGTAAC CCACTCGTGC ACCCAACTGA TCTTCAGCAT

TAACCTTTTG CAAGAAGCCC CGCTTTTGAG AGTTCCTAGA ATGGCGACAA CTCTAGGTCA AGCTACATTG GGTGAGCACG TGGGTTGACT AGAAGTCGTA

6601 CTTTTACTTT CACCAGCGTT TCTGGGTGAG CAAAAACAGG AAGGCAAAAT GCCGCAAAAA AGGGAATAAG GGCGACACGG AAATGTTGAA TACTCATACT

GAAAATGAAA GTGGTCGCAA AGACCCACTC GTTTTTGTCC TTCCGTTTTA CGGCGTTTTT TCCCTTATTC CCGCTGTGCC TTTACAACTT ATGAGTATGA

6701 CTTCCTTTTT CAATATTATT GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT ATTTAGAAAA ATAAACAAAT AGGGGTTCCG

GAAGGAAAAA GTTATAATAA CTTCGTAAAT AGTCCCAATA ACAGAGTACT CGCCTATGTA TAAACTTACA TAAATCTTTT TATTTGTTTA TCCCCAAGGC

6801 CGCACATTTC CCCGAAAAGT GCCACCTGAC GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCGTAT CACGAGGCCC TTTCGTCTTC

GCGTGTAAAG GGGCTTTTCA CGGTGGACTG CAGATTCTTT GGTAATAATA GTACTGTAAT TGGATATTTT TATCCGCATA GTGCTCCGGG AAAGCAGAAG

6901 AC

TG

Deleted regions showing membrane anchorage sequence

```
CYP3A4          -------------MALIPDLAMETWLLLAVSLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGN
CYP2C9          ----------------------MDSLVVLVLCLSCLLLLSLWRQSSGRGKLPPGPTPLPVIGN
CYP2D6          -----------------MGLEALVPLAMIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGN
```

AFM calculation details

Spot radius = 100μm

Spot area = $\pi r^2 = 31\ 416\ \mu m^2$

Streptavidin diameter= 25nm +/- 7nm area measured from AFM images

Streptavidin space diameter= 52.87nm +/- 3.11nm

Streptavidin total diameter= streptavidin + spaces

Streptavidin diameter ≈ 78 nm

Streptavidin area ≈$4778nm^2$

Streptavidin units per μm^2 of surface = 202.67 +/- 33.31 units as per AFM image

Streptavidin units per μm^2 x spot area = 202 units x 31416 μm^2 spot = 6 575 130 strep units per spot

Average ratio of P450 to P420 = 1:1

∴ There are approximately 3 250 000 streptavidin units per spot

Streptavidin exists as a tetramer so 13 000 000 binding sites available per spot,

Assuming complete binding, there are 13 000 000 binding sites / N_A = n ,V = 100nL

c = n/V

$= 2.16x10^{-17}/100x10^{-9} = 2.16*10^{-10}$ mol/L ≈ 0.22 nM

$E_T \approx 0.2\text{-}4$ nM