P450 BIOCHIPS:

Development of a protein microarray platform for investigating cytochrome P450 clinical drug metabolism

Natasha Beeton-Kempen

Thesis presented for the degree of Doctor of Philosophy
In the Department of Molecular & Cell Biology
University of Cape Town
May 2010

Abstract

The Cytochrome P450 enzyme superfamily is responsible for the primary metabolism of over 80% of currently available pharmaceutical drugs. These enzymes are highly polymorphic and this large degree of genetic variation is increasingly understood to play a significant role in the variability with which patients respond to drugs in terms of efficacy and toxicity. Currently, the ability of the general population to metabolise novel drugs is extrapolated from solution-phase assays performed on wild-type cytochrome P450 enzymes - no assay exists with the high-throughput capacity to accurately and quantitatively compare the effect of single nucleotide polymorphisms on P450 enzymatic activity. Drugs candidates are being pushed through incomplete pre-clinical testing to the costly human clinical trial stage or even to the market, where failure carries a high financial penalty. Of those that are successfully brought to market, most are typically only effective in around 40% of the patients. In addition, adverse drug reactions are contributing significantly to health care expenditure and causing thousands of fatalities each year.

There is therefore a clear need for a high-throughput platform that allows quick, cost-effective and accurate measurement of the metabolism of substrates by the cytochrome P450s, as well as a means to rapidly assess the effect of polymorphisms on drug metabolism and drug-drug interactions. Such a platform would provide an important addition to the current repertoire of tools used by pharmaceutical companies and others for pre-clinical drug screening. This thesis describes the development of a novel cytochrome P450 array format, the P450 Biochip, that allows quantitative and truly high-throughput measurement of cytochrome P450-mediated turnover reactions in sub-nanolitre volumes. The method for fabricating these arrays is significantly faster than that allowed by the membrane encapsulation methods used thus far and the array format is capable of > 329 reactions/cm^2. Both of these attributes are significant improvements in terms of high-throughput capacity. This format also provides greater ease of multiplexing and platform extension to other detection methods. With the basic format in hand, we were able to generate large numbers of high quality data points over several different parameters within a single experiment. This array format can now be extended to incorporate the various isoforms and clinically relevant polymorphic variants.

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 …… convention for citation and referencing. Each contribution to, and quotation in, this essay/report/project/…… from the work(s) of other people has been attributed, and has been cited and referenced.

3. This essay/report/project/…….. 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 _____

Acknowledgements

I would like to thank my supervisor, Professor Jonathan Blackburn, for putting up with my never-ending questions and general block-headedness. He gave me the space and resources I needed to figure things out on my own. I also wish to thank Rachel van Dyk and Dr Aubrey Shoko from the Centre for Proteomics and Genomics Research, Cape Town. Their expertise and patience were a great help in the development of the microarray platform. Other people that have contributed significantly in various ways have been Dr Siddharth Sharma, Dr Alexander Zawaira and Nicolette Hendrikse. To the Blackburn lab members, both past and present, thank you for your assistance and encouragement (as well as good company and much laughter!) over the years. There are so many people that have assisted me along the way that it is impossible to do justice to them all.

This thesis is dedicated to my husband and my parents - words cannot begin to qualify everything they have done for me. My parents have always been my cheerleaders and confidantes, with whom I could entrust every aspiration both lofty and humble. My husband, Benedict Kempen, was a constant source of support and patience throughout. To all three of you, thank you.

Contents

Abbreviations

ADME	Adsorption, distribution, metabolism and excretion
ADR	Adverse drug reaction
ALA	Aminolevulinic acid
Amp	Ampicillin
Anti-His Ab	Anti-His antibody (conjugated to horse radish peroxidase)
BCCP	Biotin carboxyl carrier protein
bp	Base pairs
BSA	Bovine serum albumin
C-	Carboxy-terminal
CD	Circular dichroism
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CAS No. 75621-03-3)
CPR	NADPH cytochrome P450 Oxidoreductase
CuOOH	Cumene hydroperoxide
CYP	Cytochrome P450
DME	Drug metabolising enzyme
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
FAD	Flavin adenine dinucleotide
FDA	Food and Drug Administration (USA)
FMN	Flavin mononucleotide
FL	Full-length protein
His_6	Hexahistidine tag
HTP	High through-put
ITPG	Isopropyl-β-D-galactoside
kDa	Kilo Daltons (1000 Daltons)
LB	Luria Bertani media
LC	Liquid chromatography
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
N-	Amino-terminal
$NADP^+$	Nicotinamide dinucleotide phosphate
NADPH	Reduced nicotinamide dinucleotide phosphate
NHS	N-hydroxysuccinimidyl ester
nm	Nanometres (1×10^{-9} m)
nt	Nucleotides
OD	Optical density
PBS	Phosphate-buffered saline

PCR	Polymerase chain reaction
PD	Pharmacodynamics
PEG	Poly(ethylene glycol)
P-gp	P-glycoprotein
PGt	Pharmacogenetics
PGx	Pharmacogenomics
PK	Pharmacokinetics
PMT	Photomultiplier tube
PTM	Post-translational modification
RBS	Ribosome binding site
RE	Restriction enzyme
RFU	Relative fluorescence units
RMSD	Root mean square distances
ROS	Reactive oxygen species
rpm	Revolutions per minute
SAB	Sample application buffer
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNP	Single nucleotide polymorphism
SPR	Surface plasmon resonance
SRS	Substrate recognition site
Strep-HRP	Streptavidin conjugated to horse radish peroxidase
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
Δ	Deletion

CHAPTER ONE

Introduction

1.1 THE CYTOCHROME P450s

1.1.1 Preface

By the 1950s it was already widely known that non-polar xenobiotics are metabolised by mammalian tissues, however, the specific enzymes responsible for this were still unidentified (Estabrook, 2003). The first experimental evidence of the cytochromes P450 (CYPs) was discovered in 1955 when an enzyme system oxidising xenobiotic compounds was identified in liver microsomes (Axelrod, 1955;Brodie *et al.*, 1955), for which Axelrod won the Nobel Prize in 1970. Then, in 1958**,** it was observed that this membrane-bound pigment exhibits a peak at 450nm (called a Soret peak) upon binding to carbon monoxide, after pre-treatment with a reducing agent (Klingenberg, 1958). This pigment was later identified by Omura and Sato to be a hemoprotein containing Fe-protoporphyrin IX, which they named cytochrome P450 after the pigment's distinctive Soret peak (Omura and Sato, 1964a;Omura and Sato, 1964b). The cytochrome P450s are now known to be a superfamily of heme-containing enzymes present in virtually all organisms, responsible for the metabolism of a huge variety of endogenous and exogenous compounds.

1.1.2 Evolution and diversity

The cytochrome P450 family is one of the largest and most functionally diverse of the enzyme superfamilies, with representatives in almost all organisms of every domain: *Archaea*, *Bacteria* and *Eukarya* (Danielson, 2002). The ancestral gene is thought to have existed over 3 billion years ago, giving rise to the present-day superfamily mainly through multiple gene duplications. Currently, there are over 7200 identified CYPs, divided into a total of 781 different families (see the cytochrome P450 Homepage: http://drnelson.utmem.edu/CytochromeP450.html).

The origin of the first CYP gene is thought to have occurred in prokaryotes, before the advent of eukaryotes and an oxygen-rich atmosphere and possibly in response to the anaerobic conditions that existed then. About 2.8 billion years ago the atmosphere started accumulating oxygen, during which time the CYPs may have been recruited to help early life forms survive oxygen-related toxicity.

Despite the critical role CYPs play in the detoxification of xenobiotics, this is probably a minority function that evolved as an extension of their existing roles in metabolising endogenous

substrates. This extension of roles is thought to have begun around 0.5-1 billion years ago during the tremendous adaptation of life forms to their new terrestrial environment, and the coevolution of herbivorous animals and the plants that acted as their food source (Gonzalez and Nebert, 1990). This coevolution led to the phenomenon of 'plant-animal warfare': plants produced toxins to make themselves less attractive as a food source to mammals and insects, and these in turn evolved the ability to metabolise these plant toxins such that they could still utilise the plants as food. This phenomenon continues and the method by which animals carry out their side of the battle has always been primarily with their artillery of CYPs.

More 'modern' cytochrome P450s function not only to catabolise compounds for carbon sources and to detoxify xenobiotics, but also to synthesise biologically active compounds such as steroids and prostaglandins. For example, they are involved in antibiotic biosynthesis in bacteria, chemical defence in fungi, pigment production in plants, and the regulation of development and reproduction in arthropods - to name just a few (Danielson, 2002). Importantly, as a result of animals evolving the ability to metabolise plant toxins, modern day humans are also able to metabolise other xenobiotics such as pharmaceutical drugs. In fact, the cytochrome P450s are involved in over 80% of all phase I drug metabolism (Ingelman-Sundberg, 2005).

Phylogenetic analyses of the CYPs suggest that this superfamily is one of the fastest-evolving (Danielson, 2002). The time required for a 1% change to occur in the amino acid of the CYP protein sequence (called the unit evolutionary period) is only 2 to 4 million years, whereas the unit period of the highly conserved histone proteins is 400 million years. This high rate of evolution would certainly facilitate organisms' needs to constantly respond to an ever-changing array of toxins, as well as to harness new food sources or expand into environments that were previously toxic.

The number of CYP genes per species is highly variable without correlation to any perceived 'complexity' of the organism (see Table 1.1), however there is a trend for higher eukaryotes to possess larger numbers of paralogous sequences (closely related sequences that arise through gene duplication within a single species), as well as pseudogenes. Pseudogenes are defective genes that no longer produce a functional protein and are often relics of gene duplications where one of the copies has lost its function – however they may still play a role through e.g. RNAi-like regulation.

Species	Total number of CYP genes	Number of predicted pseudogenes
Escherichia coli	0	0
Mycobacterium tuberculosis	20	0
Saccharomyces cerevisiae (yeast)	3	0
Drosophila melanogaster (fruit fly)	86	4
Caenorhabditis elegans	76	9
Mus musculus (mouse)	102	88
Homo sapiens	57	58
Oryza sativa (rice)	356	99
Arabidopsis thaliana	246	26

Table 1.1: The total number of predicted or known CYP genes and pseudogenes for selected species (see the Cytochrome P450 Homepage: http://drnelson.utmem.edu/CytochromeP450.html).

1.1.3 Nomenclature

The nomenclature of the CYPs was originally based on their physiological function (e.g. P450cam for camphor-metabolising) and then in the 1990s, CYPs were named according to serial numbers based on amino acid sequence similarity (Danielson, 2002). The CYP nomenclature system is now based on amino-acid sequence identity, phylogenetic association, and gene organisation, the rules of which are determined by the P450 Nomenclature Committee (Nelson *et al.*, 1993). (See also http://drnelson.utmem.edu/Nomenclature.html and www.cypalleles.ki.se.)

The root for all cytochrome P450 gene names is *CYP* (*Cyp* is the convention for mouse), with the individual family designated by an Arabic numeral, and a capital letter indicating the subfamily. Member sequences within a subfamily are numbered consecutively as they are reported to the nomenclature committee. Pseudogenes are designated with a letter 'P' following the Arabic numeral. The same nomenclature is used for mRNA and protein sequences, except that the latter designations are not italicised. According to these rules, the first officially named CYP was *CYP1A1*, or CYP1A1 for the protein. In general, CYP sequences displaying >40% amino acid identity are placed in the same family. Those with >55% sequence identity are placed in the same subfamily, while those that are more than 97% identical are considered to be allelic variants unless there is evidence to the contrary.

CYPs are so ubiquitous amongst living organisms that blocks of CYP family numbers have been reserved, awaiting the classification of new CYP sequences (Danielson, 2002). I.e., *CYP1-CYP49* and *CYP301-499* are for animals, *CYP51-CYP69* and *CYP501-CYP699* are for lower eukaryotes, *CYP71-CYP99* and *CYP701-CYP999* are for plants, while *CYP101-CYP299* are for bacteria. Importantly however, the CYP family numbers do not in any way reflect their phylogenetic

relatedness. In recent years, the increasing rate of gene discovery has resulted in a flood of new CYP sequences that have had to be classified and named. Of the many CYP families, only one family, CYP51, has been conserved across plants, animals, fungi and bacteria and is known to produce sterols and steroids that activate meiosis (Ning *et al.*, 2008).

1.1.4 Membrane targeting and sub-cellular localisation

The cytochrome P450s generally range between 480 to 560 amino acids in length and are grouped into three main categories based on their sub-cellular localisation (Neve and Ingelman-Sundberg, 2008). The majority of the eukaryotic CYPs are bound to the membranes of the endoplasmic reticulum (microsomal type), where they metabolise xenobiotics and some endogenous substrates. The second group of CYPs are bound to the inner membrane of the mitochondria (mitochondrial type), whereas the third minority group consists of the soluble CYPs (cytosolic type). Whilst all prokaryotic CYPs are cytosolic, this soluble CYP form is rare in eukaryotes. Mitochondrial CYPs have only been found in animals, insects and nematodes, and are thought to have originated from a microsomal CYP that accumulated mutations in the endoplasmic reticulum (ER)-targeting sequence resulting in mis-targeting to the mitochondria.

Microsomal type

Microsomal CYPs have a characteristic N-terminal hydrophobic domain of 20-25 residues, which acts as a signal-anchor sequence for insertion into the membrane of the endoplasmic reticulum (see Fig. 1.1 below) (Neve and Ingelman-Sundberg, 2008). This association is facilitated by less well-defined hydrophobic regions on other areas of the protein surface, which are thought to be localised in the F and G helices. The signal-anchor sequence is usually preceded by a negatively charged amino acid and followed by a short stretch rich in basic residues. These flanking charged residues are thought to ensure the correct topological orientation of the nascent CYP in the ER membrane, by providing a halt-transfer signal between the hydrophobic membrane anchor and the globular part of the protein. A similar targeting sequence is also present at the N-terminus of the mitochondrial CYPs; cytosolic CYPs of course do not contain these signal-anchor regions. C-terminal to the basic signal-anchor/targeting sequence of both the microsomal and the mitochondrial CYPs, there is usually a proline-rich region with the conserved core motif of Pro-Pro-Gly-Pro. This region is thought to act as a flexible 'hinge', while also facilitating proper

folding of the nascent protein (Kusano *et al.*, 2001b;Kusano *et al.*, 2001a). The majority of the protein however, including the catalytic domain, is exposed to the cytoplasm.

```
CYP1A1          ----------MLFPISMSATEFLLASVIFCLVFWVIRASRPQVPKGLKNPPGPWGWPLIGH
CYP1A2          ---------MALSQSVPFSATELLLLASAIFCLVFWVLKGLRPRVPKGLKSPPEPWGWPLLGH
CYP1B1          MGTSLSPNDPWPLNPLSIQQTTLLLLLSVLATVHVGQRLLRQRRRQLRSAPPGPFAWPLIGN
CYP2A6          ----------------MLASGMLLVALLVCLTVMVLMSVWQQRKSKGKLPPGPTPLPFIGN
CYP2B6          --------------------MELSVLLFLALLTGLLLLLVQRHPNTHDRLPPGPRPLPLLGN
CYP2C9          --------------------MDSLVVLVLCLSCLLLLSLWRQSSGRGKLPPGPTPLPVIGN
CYP2C19         --------------------MDPFVVLVLCLSCLLLLSIWRQSSGRGKLPPGPTPLPVIGN
CYP2D6          ----------------MGLEALVPLAMIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGN
CYP2E1          ----------------MSALGVTVALLVWAAFLLLVSMWRQVHSSWNLPPGPFPLPIIGN
CYP3A4          ------------MALIPDLAMETWLLLAVSLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGN
```

Figure 1.1: Alignment of several human microsomal CYP N-terminal signal-anchor regions. An acidic amino acid residue (*red*) often flanks the hydrophobic signal-anchor sequence (*yellow*), which targets the protein to the endoplasmic reticulum. C-terminal to this sequence is a region rich in basic amino acids (*blue*), followed by a proline-rich region (*green*). [Adapted from (Neve and Ingelman-Sundberg, 2008).]

The hydrophobic N-terminus is not only responsible for anchoring the protein, but also for the co-translational targeting of the protein into the ER membrane in a signal recognition particle (SRP) dependent manner (Neve and Ingelman-Sundberg, 2008). CYPs are translated on membrane-free ribosomes in the cytoplasm and as soon as the N-terminus emerges from the ribosome, translation is halted through the binding of the SRP. Translation is only resumed after this complex binds to the SRP receptor in the ER membrane. After binding, the SRP is released, the N-terminus is translocated through the membrane, and the rest of the protein is then translated on the cytoplasmic side of the ER membrane. This results in an N_{lum}/C_{cyt} orientation.

Mitochondrial type

Mitochondrial CYPs are distinct from the microsomal CYPs, not only with respect to localisation, but also regarding substrate specificity and structural characteristics (Neve and Ingelman-Sundberg, 2008). Mitochondrial CYPs are encoded by nuclear genes, which are translated in the cytoplasm and then post-translationally imported into the mitochondria. While microsomal CYPs are mainly involved in the metabolism of exogenous compounds, mitochondrial CYPs predominantly metabolise endogenous compounds such as those involved in the synthesis of vitamin D and steroid hormones. Mitochondrial CYPs also differ in their N-terminal membrane-targeting domains. Instead of an ER-targeting sequence they contain a typical mitochondrial-targeting sequence: a stretch of neutral and hydrophobic amino acids containing a few positively

charged amino acids. After translocation of the entire protein into the mitochondrial matrix the targeting sequence is removed and the mature protein is then incorporated into the inner membrane, leaving the catalytic domain exposed to the matrix space. Unlike microsomal CYPs, mitochondrial CYPs rely on adrenodoxin and NADPH adrenodoxin reductase, two soluble matrix proteins, for their supply of electrons.

1.1.5 Structure

The first CYP to be crystallised was the camphor-metabolising CYP, CYP101 (also known as P450cam), from the bacterium *Pseudomonas putida* in 1985 by Poulos and co-workers (Poulos *et al.*, 1985). The extensive studies carried out on the structure of this CYP provided the framework for predictions on the structure of the eukaryotic CYPs. Since then, several prokaryotic and eukaryotic CYPs have been crystallised, some bound to various ligands; today there are 287, mostly non-redundant CYP structures in the Protein Data Bank (http://www.rcsb.org). The first eukaryotic CYP structure to be determined was CYP2C5 (Williams *et al.*, 2000). Eukaryotic CYPs have had to be crystallised in the soluble form where the hydrophobic N-terminal domain has been deleted through cloning techniques, since the detergents required to solubilise membrane-bound CYPs are typically incompatible with the formation of high quality crystals.

A striking finding from the comparison of the crystal structures of eukaryotic and prokaryotic CYPs is that even where the global sequence identity between any two CYPs is less than 20%, the overall fold topography (i.e. the three-dimensional structure) is highly conserved, with the majority of secondary structures found in prokaryotic CYPs also observed in eukaryotic CYPs (Otyepka *et al.*, 2007).

Other key features of CYP structure that have emerged are that within the CYP's common globular to triangular three-dimensional structure, there is an α-helix-rich C-terminal half and a β-sheet-rich N-terminal half (Graham and Peterson, 1999;Werck-Reichhart and Feyereisen, 2000). The first domain is associated with the catalytic centre, the second domain with substrate recognition and the access channel. There are usually around 4 β-sheets and 13 α-helices in total. Sequence alignments have yielded several conserved sequence motifs, with the majority located in the C-terminal region (around the core heme group), as well as three absolutely conserved amino acids (shown in bold below). The highly conserved core consists of a four-helix bundle

(parallel D, I, L and E helices), the J and K helices, β-sheets β1 and β2, and a coil called the 'meander' (see Fig. 1.2). These two β-sheets are involved in forming a hydrophobic substrate access channel.

Situated within the I-helix near the protein centre, is the conserved consensus motif (Gly/Ala)-Gly-X-(Asp/Glu)-Thr, which forms part of the proton transfer groove on the distal side of the heme. The terminal threonine residue in this motif forms a critical part of the CYP's oxygen-binding pocket. In the K-helix, there is an invariant **Glu**-X-X-**Arg** motif on the proximal side of the heme group. Any mutation of this motif results in complete loss-of-function and it is thought to stabilise the entire core structure. This motif also interacts with residues in the meander, which is a loop of approximately 14 residues N-terminal to the heme-binding loop. The heme-binding loop lies N-terminal to the L-helix, and contains the most characteristic CYP consensus sequence: Phe-X-X-Gly-X-X-X-**Cys**-X-Gly. The heme group is confined within the inner core between the distal I-helix and the proximal L-helix, and is attached *via* a dative bond to the cysteine residue in the proximal heme-binding loop. This cysteine is crucial as it forms the fifth ligand to the heme group. The I-helix is the longest helix in the protein and spans the entire molecule, running over the distal surface of the heme group (i.e. the protein surface furthest from the heme).

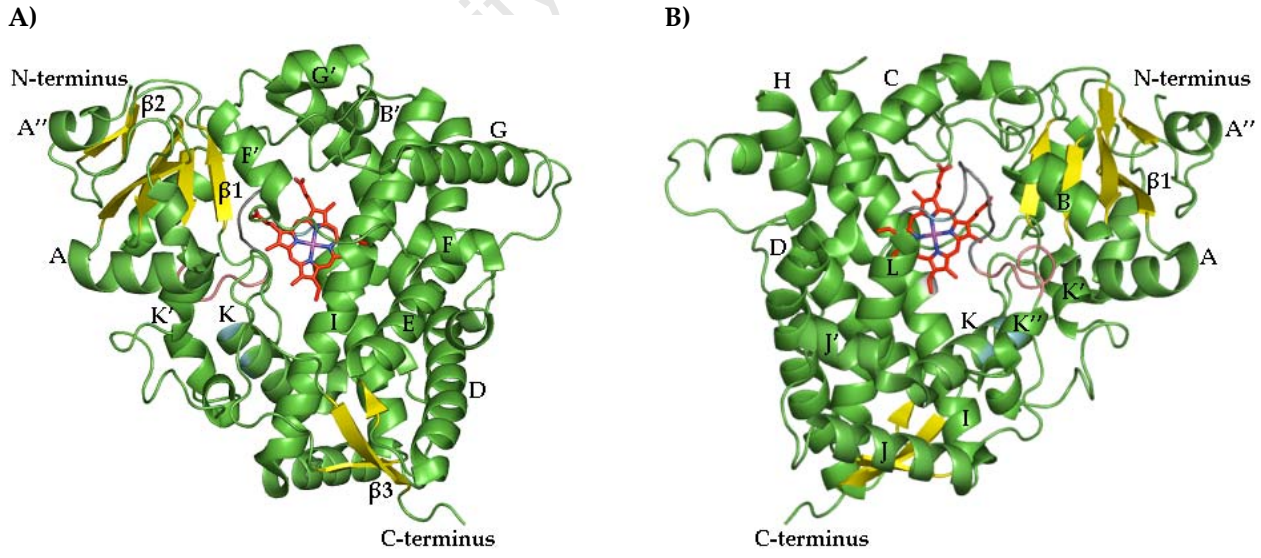


Figure 1.2: A PyMOL representation of the distal face (A) and proximal face (B) of human CYP3A4. The various secondary structure elements indicated are the α-helices (*green*), the β-sheets (*yellow*), the heme-binding loop (*grey*), and the meander loop (*pink*). The three absolutely conserved residues, Glu362, Arg365 and Cys442 (*cyan*), and the conserved residue, Thr309 (*white*), are also highlighted. The heme group (*red*) lies within the core bundle. [Based on the ligand-free crystal structure 1TQN resolved in (Yano *et al.*, 2004). Nomenclature from (Graham and Peterson, 1999)].

1.1.6 Substrate specificity

The CYPs are well known for the diverse reactions they catalyse, as well as the variety of substrates upon which they act. Although the CYP fold is highly conserved, CYPs have enough structural diversity to metabolise widely dissimilar substrates, bind to different redox partners, and get targeted to the correct cellular location (Graham and Peterson, 1999). The most variable structural elements in the CYPs are the helices A, B, B', F, G, H, K', the β-sheets β3 and β4, with the presence of β5 itself being variable. These variable regions appear to be associated with substrate recognition, substrate binding and redox partner binding. Residues associated with substrate recognition and binding comprise six substrate recognition sequences (SRSs) [see Fig.1.3 (Gotoh, 1992)]. Loops B-B' and B'-C line the active site (SRS1) whilst helices F and G and their joining loop form part of the access channel and ceiling of the active site (SRS2 and SRS3). A part of the I-helix (SRS4), the K-helix-β2 connecting region (SRS6), and the β4 hairpin (SRS5) also line the CYP active site.

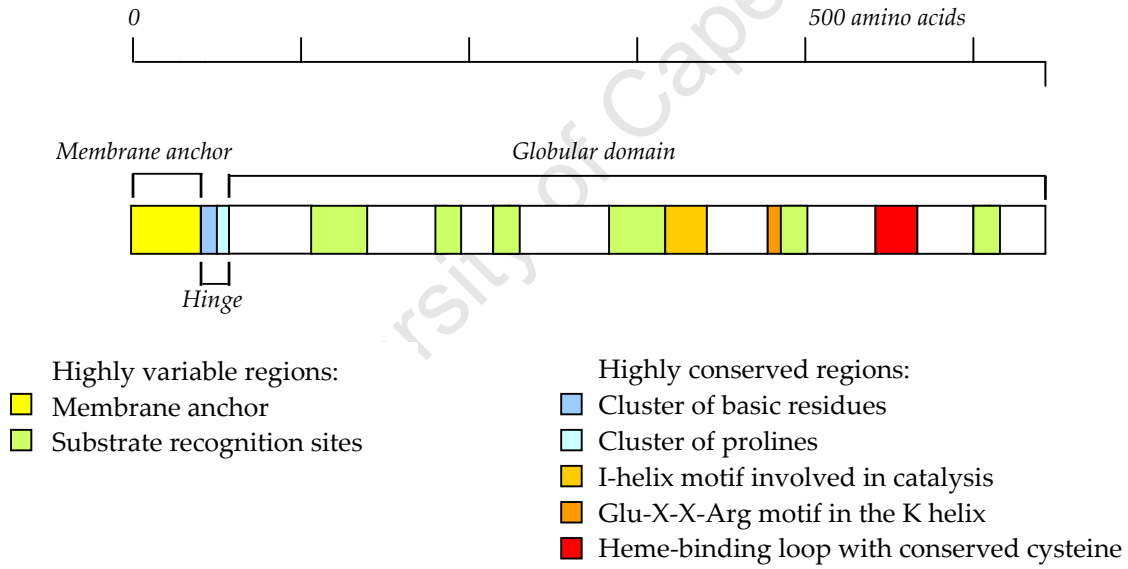


Figure 1.3: Diagram depicting the various conserved and variable regions along a typical microsomal cytochrome P450 sequence. [Adapted from (Werck-Reichhart and Feyereisen, 2000).]

The SRSs are thought to determine substrate specificity and any point mutations in these sequences therefore have the potential to significantly change this specificity. In many cases, the perfect molecular fit between the substrate and the CYP only occurs after binding and the flexibility of the SRSs are thought to play a role in this 'induced-fit' model of substrate binding. These sites are thus an area of much interest, as the enormous catalytic diversity of the CYPs is

coupled to such narrow sequence diversity. Two of the most productive approaches used so far to determine SRSs have been site-directed mutagenesis and homology modelling (Danielson, 2002).

One powerful example of these approaches lies in studies of CYP2A4 (androgen 15α-hydroxylase) and CYP2A5 (coumarin 7-hydroxylase). Although these two CYPs share 97.8% sequence identity, neither enzyme is able to catalyse the reaction of the other. Using site-directed mutagenesis, Lindberg and Negishi showed that coumarin 7-hydroxylase activity could be conferred on CYP2A4 and androgen 15α-hydroxylase activity could be conferred on CYP2A5 by reciprocal substitution of a single amino acid, Phe to Leu209 (Lindberg and Negishi, 1989). These two CYPs were observed to differ by only 11 amino acids that were confined to short stretches along the protein sequences. These regions were later named substrate recognition sites, and as mentioned above, a total of six SRS regions have since been identified in eukaryotic CYPs (Gotoh, 1992).

Differences in reaction rates between CYP isoforms that metabolise the same substrate have also been linked to amino acid differences within these SRSs. For example, both CYP2C4 and CYP2C5 catalyse the 21-hydroxylation of progesterone, but the apparent K_m of CYP2C4 for this reaction is more than 10-fold greater than that of CYP2C5, despite a >95% sequence identity between these two CYPs (Kronbach and Johnson, 1991). This disparity was linked to a single amino acid difference in SRS1 (Johnson *et al.*, 1992).

Finally, the amino acid composition of these SRSs can also cause the regioselectivity of metabolism of the same substrate to differ between two CYPs, as in the metabolism of aflatoxin B1 by human CYP3A4 and CYP3A5. CYP3A4 prefers 3α-hydroxylation to exo-8,9-oxidation, while CYP3A5 can only perform exo-8,9-oxidation. This difference in regioselectivity has important physiological consequences, as metabolism by CYP3A4 results in detoxification of aflatoxin B_1, whereas metabolism by CYP3A5 results in the production of a potent genotoxin. This effect has been linked to residues in SRS2 (Wang *et al.*, 1998).

The selectivity of a CYP for any particular substrate is generally governed by several properties of the substrate and how it 'matches' the substrate-binding site of a particular CYP. The molecular weight (volume), planarity (area/depth^2), pKa, and lipophilicity of the substrate, as well as the distance between the hydrogen bond donor/acceptor atoms on the substrate and the preferred site

of metabolism by the specific CYP, are all important physicochemical factors. Ioannides and Lewis have devised a decision tree for determining which CYP isoforms are likely to metabolise a particular substrate given these parameters [see (Ioannides and Lewis, 2004;Lewis, 2000)]. Despite individual cytochrome P450s having different substrate specificities, there is however also considerable overlap in specificities. Thus, a single CYP enzyme may be responsible for all the CYP-mediated metabolism of a given drug, or a variety of CYPs may contribute.

Due to this substrate overlap and the ability of individual CYPs to metabolise compounds that vary greatly in size and shape, these enzymes are often termed 'promiscuous'. This ability is thought to be due to the large degree of conformational flexibility of the CYPs (Guengerich, 2006). Of the CYP isoforms, CYP3A4 exhibits the most ligand promiscuity: it is involved in the metabolism of ~50% of marketed drugs, it is the isoform most often implicated in drug-drug interactions, and it is able to metabolise substrates ranging in size from acetaminophen (M_r 151) to cyclosporin A (M_r 1201). This demonstrates remarkable malleability, especially considering the small size of the CYP3A4 active site relative to other promiscuous CYPs such as CYP2C9 and CYP2C8 (Williams *et al.*, 2004). In the most recent published CYP3A4 crystal structures, the binding of ketoconazole (M_r 531) and erythromycin (M_r 734) was shown to increase the CYP3A4 active site size by >80%, while ligand-bound CYP3A4 crystal structures for these substrates differ from the ligand-free structure by >1 Å RMSD (root mean square distances) for all ordered C^{α} atoms (Ekroos and Sjörgren, 2006).

In the above ketoconazole-bound structure described by Ekroos *et al.*, two ketoconazole molecules were observed in the CYP3A4 active site demonstrating that it can accommodate more than one substrate molecule at a time. In a structure published earlier by another group (Williams *et al.*, 2004), progesterone was not found binding within the active site but was instead observed at a peripheral site, which the authors suggested could be an effector site. This multiple ligand binding ability probably adds further to ligand promiscuity as the active site shape and charge distribution may be modulated by the second ligand. Multiple ligand binding has also been suggested as a likely explanation for the non-Michaelis-Menten kinetics displayed by different CYPs using various substrates, as both homo- and heterotropic cooperativity have often been observed (Guengerich, 2006).

Further evidence of CYP conformational flexibility is that different ligands induce different conformational states and it is therefore unlikely that current ligand-bound structures are fully representative of the CYP conformational space (Ekroos and Sjörgren, 2006). This, in addition to the complexities added by possible multiple ligand binding by certain CYPs (including CYP2B6, 2C8 and 2C9), greatly complicates attempts to develop general methods for predicting binding modes and sites of metabolism of uncharacterised substrates/substrate combinations, let alone the likely ligand-bound CYP structures. A clinically significant implication of all this is that it further confounds attempts to use available crystal structures to predict the effect of CYP genetic variation on substrate metabolism, an activity that is still difficult for proteins in general (Bromberg and Rost, 2009).

1.1.7 Catalytic Mechanisms

The cytochrome P450s are traditionally referred to as hydroxylases, mixed function oxidases and monooxygenases (Danielson, 2002). They are best known for catalysing the oxygenation of a large variety of endogenous and xenobiotic substrates in both a regiospecific and stereospecific manner. However, most of these reactions are variations on a theme and can therefore be broadly classified into four categories: hydroxylations (an oxygen atom is introduced into a C-H bond to yield a hydroxyl group), epoxidations (an oxygen is introduced into a carbon-carbon π-bond), heteroatom oxidations (an oxygen is added to a nitrogen, sulphur, or other heteroatom), or reduction reactions (these take place under conditions of limited oxygen where an alternate electron acceptor is available).

The main function of the CYPs is however in the conversion of unreactive, lipid-soluble substrates into more reactive, water-soluble products, which facilitates the elimination of foreign compounds through organisms' waste systems. This is predominantly achieved through the insertion of an oxygen molecule derived from molecular oxygen, which, when activated, yields a reactive species that can attack and insert an oxygen atom into a relatively inert C-H bond. By providing foreign compounds with reactive functional groups, the CYPs pave the way for the next phase of metabolism in which these groups are required for conjugation.

During the CYP-catalysed hydroxylation reaction, molecular oxygen (O_2) is split: one atom is incorporated into the substrate while the other is released as part of a water molecule. The following equation depicts the basic stoichiometry of this reaction:

$$RH + 2NAD(P)H + O_2 + 2H^+ \longrightarrow ROH + 2NAD(P)^+ + H_2O$$

The CYPs can be divided into four classes depending on how the electrons are delivered to the active site from NAD(P)H (Werck-Reichhart and Feyereisen, 2000). Class I proteins include the mitochondrial CYPs and some bacterial CYPs: these CYPs require both an FAD-containing redoxin reductase and an iron-sulphur-containing protein called redoxin. Electrons from NAD(P)H are shuttled from FAD to Fe_2S_2 to the CYP heme. Class II proteins include eukaryotic microsomal CYPs as well as some bacterial CYPs. These proteins use an FAD/FMN-containing cytochrome P450 reductase and electrons from NADPH are shuttled from FAD to FMN to the heme. In microsomal CYP systems, different CYPs use the same reductase (for which there is only one gene in mammals), whereas for soluble, bacterial CYPs of this class, the CYP and the reductase are present on the same polypeptide chain (e.g. P450 BM3). For some of the Class I and II CYPs, the catalytic efficiency of certain reactions is increased in the presence of a second accessory enzyme, cytochrome b_5. Class III proteins are self-sufficient and do not require any molecular oxygen or electron donors. In mammals, these CYPs are involved in the isomerisation of signalling molecules such as prostaglandins. Class IV enzymes receive electrons directly from NADH. These CYPs are unique to fungi and function to reduce NO levels during denitrification; only one representative of this class, P450nor/CYP55A1, has been found so far. The latter two classes appear to be the most ancestral forms of CYPs and were probably involved in the detoxification of harmful activated oxygen species.

More than 40 years after the first isolation and characterisation of CYP enzymes, the exact nature of the active species responsible for the oxygen insertion step was still a matter of intense debate (Meunier *et al.*, 2004). Studies of active CYP species have greatly benefited from various spectroscopic techniques and recent developments in methods using quantum and molecular mechanics. These methods predicted a high-valent iron(IV)-oxo species as the key reactive intermediate in the hydroxylation catalytic cycle of CYPs (de Visser, 2009;Nam, 2007). In 2006, this species was physically observed (Makris *et al.*, 2006), while several reports have also identified non-heme iron(IV)-oxo intermediates as the active oxidising species for various enzymes in *E. coli* (Bollinger and Krebs, 2006;Galonic *et al.*, 2007;Hoffart *et al.*, 2006). This

intermediate is generally depicted as either FeO^{3+}, which avoids the specifics of charge localisation; $Fe^{V}=O$; or $Fe^{IV}=OP^{+\bullet}$, where P is the heme porphyrin (Guengerich, 2004). The latter depiction avoids the high charge on the iron and is probably a better representation of the true electronic state. While the consensus is that the high-valent iron-oxene species is the predominant oxidant, there are alternative oxygenating intermediates that have been postulated for certain reactions, including Fe(III)-OOH (iron-oxenoid) or Fe(III)-H_2O_2 [(hydro)peroxo-iron], with some preliminary physical evidence available (Hlavica, 2004). These alternative oxidants may contribute to the remarkable variety of CYP substrate modifications observed.

The predominant CYP catalytic cycle that results in hydroxylation is illustrated in Fig. 1.4 (Nam, 2007;Pylypenko and Schlichting, 2004). In its resting state, the heme group exists as a mixture of low-spin Fe(III) with a water molecule bound (a hexacoordinate state), and high-spin Fe(III) with cysteine as the only ligand (a pentacoordinate state). The binding of the substrate (R-H) in the active site shifts this equilibrium towards the high-spin pentacoordinate state. This shift between a low-spin state to a high-spin state induces a change in the redox potential of the iron centre from –300 to -170 mV, which triggers the transfer of a single electron to the heme iron *via* the reductase, reducing the ferric iron [Fe(III)] to the ferrous [Fe(II)] form. The CYP-substrate complex then binds and reacts with molecular oxygen to produce an unstable ferric dioxygen-bound complex (also called the superoxide complex) that accepts the transfer of a second electron. The resulting iron(III)-peroxo complex then accepts a proton and by doing so, increases the oxidising potential of the peroxide.

Subsequent heterolysis of the peroxo bond results in the loss of hydroxide and is accompanied by two sequential single electron oxidations of Fe(III) to Fe(IV) and of the porphyrin to a π-cation radical. The active oxidants in CYP are thus the iron(IV)-oxo intermediate, accompanied by a porphyrin π-cation radical; together, these make up the two oxidising equivalents required to oxidise substrates. Of the two electrons from the oxidised substrate, one electron reduces iron(IV)-oxo back to iron(III) with the insertion of the oxygen atom into the C-H bond of the substrate, whilst the other electron reduces the cation radical back to stable porphyrin. Finally, the product (R-OH) is released, returning the enzyme to its resting state.

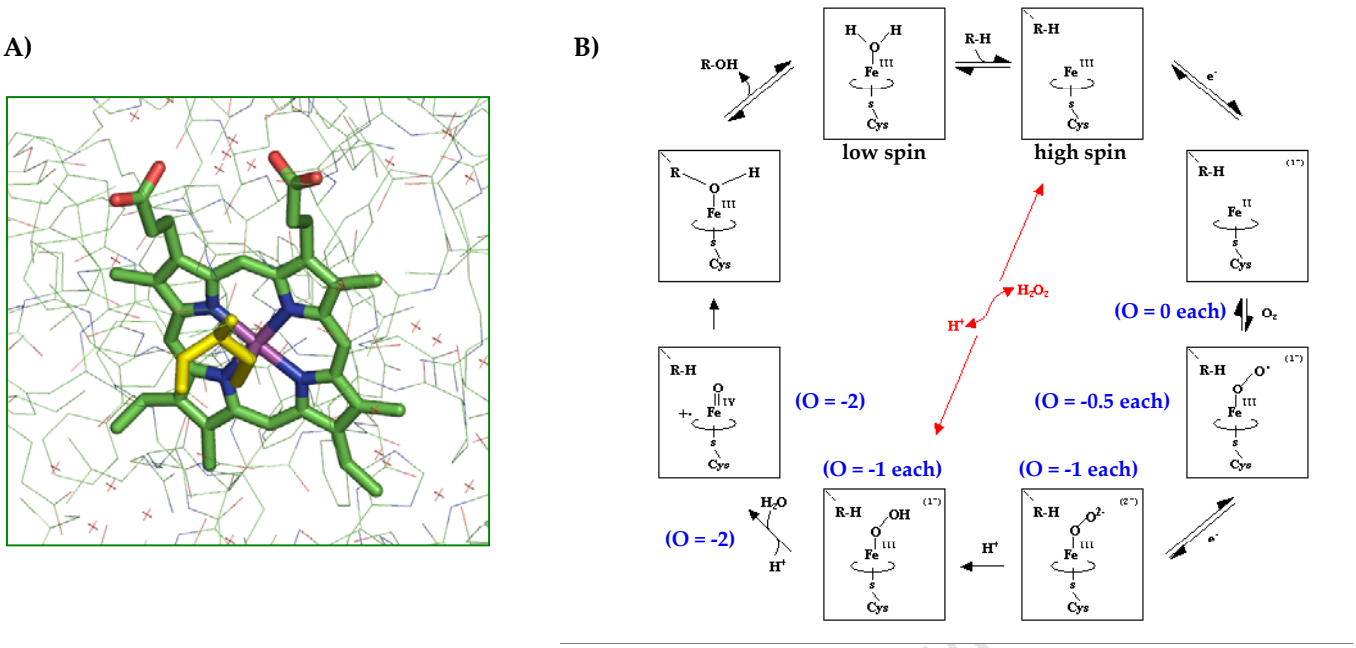


Figure 1.4: A) PyMOL representation of the CYP iron(III) protoporphyrin-IX group (C atoms *green*, O atoms *red*, N atoms *blue*, Fe atom *purple*) bound to the conserved cysteine residue (*yellow*) within the heme-binding loop. B) The catalytic cycle of CYP-catalysed hydroxylation. The number of electrons above resting state (*brackets*) is shown where relevant and the shunt pathway is depicted (*red*). The oxidation states of the oxygen atoms are also shown (*blue*). [Adapted from (Nam, 2007;Pylypenko and Schlichting, 2004).]

The short-lived and highly reactive iron(IV)-oxo intermediate results in oxidation that is intrinsically non-selective from a chemical perspective (Poulos, 2005). Hence, specificity is controlled entirely by substrate access, binding and orientation in the CYP active site (see Section 1.1.6 above).

1.1.8 Regulation

Several levels of regulation, including transcription, translation and post-translational modification, participate in maintaining the proper function of CYPs (Aguiar *et al.*, 2005).

Interest in the regulation of cytochrome P450s was initially sparked in the 1960s by the finding that phenobarbital and related barbiturates were able to induce CYP expression and thereby increase their own metabolism (Alvares *et al.*, 1967). Since then, transcriptional regulation *via* several steroid receptors has been observed in response to specific endogenous and xenobiotic compounds. The CYP1 proteins were found to be commonly inducible by polycyclic aromatic hydrocarbons through the aryl hydrocarbon receptor (AhR)/ARNT pathway. In contrast, the

CYP2 and CYP3 families are generally inducible, but to various extents by diverse, structurally unrelated xenobiotics, which are usually ligands of the orphan nuclear receptors pregnane X receptor (PXR) and constitutive androstane receptor (CAR) (Zanger *et al.*, 2008). This compound-mediated regulation of CYP protein levels almost always occurs at the point of transcription, i.e. more mRNA transcripts are produced in response to the xenobiotic. The induction of CYP2E1 by ethanol is one of few known instances where mRNA and protein stabilisation appear to play a role.

Another mechanism that acts at the level of transcription is that of epigenetic regulation (Hirota *et al.*, 2008). The most common forms of epigenetic regulation in mammalian systems are DNA methylation and histone acetylation. Neither of these change the DNA sequence but both can alter gene expression levels. Epigenetic regulation has been observed so far for several of the xenobiotic-metabolising CYPs, particularly CYP1A2 and CYP2E1.

Many CYP enzymes have yet to be fully characterised in terms of their post-translational regulation and what knowledge there is has been derived from studies with relatively few enzymes (Aguiar *et al.*, 2005). Post-translational modifications (PTMs) such as phosphorylation, glycosylation, nitration, and ubiquitination have all been described for CYPs and are known to modulate CYP activity, sub-cellular localisation and degradation. Of these, phosphorylation and ubiquitination are the best studied so far.

Phosphorylation has been reported for a number of CYP enzymes (Aguiar *et al.*, 2005;Neve and Ingelman-Sundberg, 2008). Phosphorylation regulates the CYPs in a variety of ways including the modulation of catalytic activity, substrate binding, binding of redox partners and substrate specificity. In addition, several CYP enzymes are thought to be re-directed to the mitochondrial inner membrane in response to phosphorylation during translation and for certain CYPs phosphorylation results in faster degradation (possibly through the destruction of the heme group).

In the course of CYP reactions, reactive oxygen species and/or reactive metabolic products are generated that can attack the heme group or the protein itself, leading to structural and functional damage. The structural unravelling of a damaged CYP signals its rapid removal *via* either of two degradation pathways: lysosomal and ubiquitin-dependent (Correia *et al.*, 2005;Neve and

Ingelman-Sundberg, 2008). Individual CYPs can be targeted to either of these two pathways depending on their conformational state: the lysosomal pathway is non-specific and not tightly regulated, while the ubiquitin-dependent pathway is highly regulated and specific, branding proteins for removal by covalently modifying them with a poly-ubiquitin tag. Microsomal CYPs vary considerably with respect to their protein stability, with reported half-lives ranging from 5 to 37 hours. These differences in stability reflect the different cellular pathways by which these CYPs are degraded: lysosomal pathway for long-lived CYPs, ubiquitin-dependent pathway for short-lived CYPs.

1.1.9 Role in clinical drug metabolism

The human genome encodes 57 cytochrome P450s and around 58 pseudogenes, organised into 18 families and 43 subfamilies. These CYPs play a critical role in human health (Danielson, 2002). Not only do they metabolise endogenous and xenobiotic substrates ranging from environmental pollutants, plant toxins, fatty acids, to prostaglandins, they also carry out functions such hormone synthesis and cholesterol homeostasis. Their importance is also emphasised by the several diseases associated with defects in CYP genes, including glaucoma (CYP1B1) and adrenal hyperplasia (CYP11A1, CYP21A2) (Guengerich, 2003).

Fifteen of the human CYP families, CYP4-51, are primarily involved in the metabolism and biosynthesis of endogenous compounds (see Table 1.2 below). However, the CYPs are best known for the role the CYP1-3 families play in the metabolism of pharmaceutical drugs, and their contribution to two of the most challenging issues faced in clinical pharmacology: adverse drug reactions and inter-individual differences in therapeutic efficacy. These enzymes are predominantly expressed in the liver, whilst expression in a variety of extrahepatic tissues such as the intestine, skin, brain, lung and placenta has also been observed (Danielson, 2002). Although these extrahepatic CYPs play a minor role in total drug clearance, they are thought to play a significant role in the local disposition of drugs and hence the therapeutic and toxicological effects observed in the peripheral tissues (Pavek and Dvorak, 2008).

CYP Family	Functions	Subfamilies	Genes
1	Drug metabolism mainly (also a few steroids such as estrogen)	3	3
2	Drug and steroid metabolism	13	16
3	Drug metabolism mainly (also a few steroids such as testosterone)	1	4
4	Fatty acid metabolism (including arachidonic acid)	5	12
5	Hemostasis and vasoconstriction through thromboxane A_2 synthesis	1	1
7	Cholesterol homeostasis by converting cholesterol into bile acids	2	2
8	Hemostasis and vasodilation through prostacyclin synthesis; indirect cholesterol homeostasis through regulation of bile acid biosynthesis	2	2
11	Steroid biosynthesis	2	3
17	Steroid biosynthesis	1	1
19	Steroid biosynthesis (conversion of androgens into estrogens)	1	1
20	Function not determined	1	1
21	Steroid biosynthesis	1	1
24	Vitamin D3 degradation	1	1
26	Regulation of retinoic acid levels, important in vertebrate development	3	3
27	Cholesterol homeostasis through bile acid biosynthesis, Vitamin D3 biosynthesis pathway	3	3
39	Cholesterol homeostasis through bile acid biosynthesis	1	1
46	: Cholesterol homeostasis through bile acid biosynthesis	1	1
51	Cholesterol homeostasis through cholesterol synthesis from lanosterol	1	1

Table 1.2: Human cytochrome P450 families and their main functions [(Danielson, 2002), Cytochrome P450 Homepage: http://drnelson.utmem.edu/CytochromeP450.html].

Drugs are metabolised by a variety of sequential or competitive chemical processes involving oxidation, reduction, and hydrolysis (phase I reactions) or glucoronidation, sulfation, acetylation, and methylation (phase II reactions) (Wilkinson, 2005). The phase I drug metabolising enzymes (DMEs) include the cytochrome P450s, flavin monooxygenases, monoamine oxidases, epoxide hydrolases, dehydrogenases and esterases. The phase II DMEs include the sulfotransferases, UDP-glucuronosyl transferases, N-acetyltransferases and glutathione S-transferases. The CYPs in the CYP1-3 families alone are responsible for around 80% of phase I-dependent clinical drug metabolism (see Fig. 1.5) (Ingelman-Sundberg, 2005).

The human body's first line of defence against drugs and other foreign compounds consists of the drug transporters such as P-glycoprotein (P-gp). These proteins prevent drug absorption through the gastrointestinal tract by facilitating their efflux from the enterocytes back into the lumen. They repeatedly transport drugs out of the mucosa and back into the lumen, from which are then passively reabsorbed again. This process is thought to increase the efficiency of drug metabolism by preventing CYP enzymes from being overwhelmed by high drug concentrations within the enterocytes (Christians, 2004). Around 25% of the drugs that evade this obstacle and enter the

body are then excreted unchanged into the urine *via* the kidneys (Zanger *et al.*, 2008). However, most xenobiotics are lipophilic, which is essential for them to permeate cell membranes. Lipophilic compounds are very difficult to eliminate through direct excretion in the bile and urine, and need to first be metabolised into hydrophilic entities before excretion can occur. This metabolism occurs mainly in the liver.

The liver is therefore, unsurprisingly, the major site of cytochrome P450-mediated metabolism (see Fig. 1.5A) (Wilkinson, 2005). The enterocytes in the epithelium of the small intestine are also an important site, with high levels of CYP3A present. Together, P-gp and CYP3A form a cooperative barrier against the absorption of xenobiotics (and are both regulated by the PXR nuclear receptor). After oral administration of a drug, cytochrome P450 enzymes located in the intestine and the liver reduce the dose before it reaches the systemic circulation, thereby decreasing the drug's 'bioavailability'. This phenomenon of pre-systemic metabolism, termed 'first-pass metabolism', consequently affects the therapeutic efficacy of a drug as it lowers the concentration at the drug target. Less than half the administered oral dose of about 40% of commonly used drugs is bioavailable because of limited absorption, first-pass metabolism, or both. Drug interactions resulting in either inhibition or induction of the involved enzymes can therefore markedly affect bioavailability and, as a result, drug efficacy and toxicity.

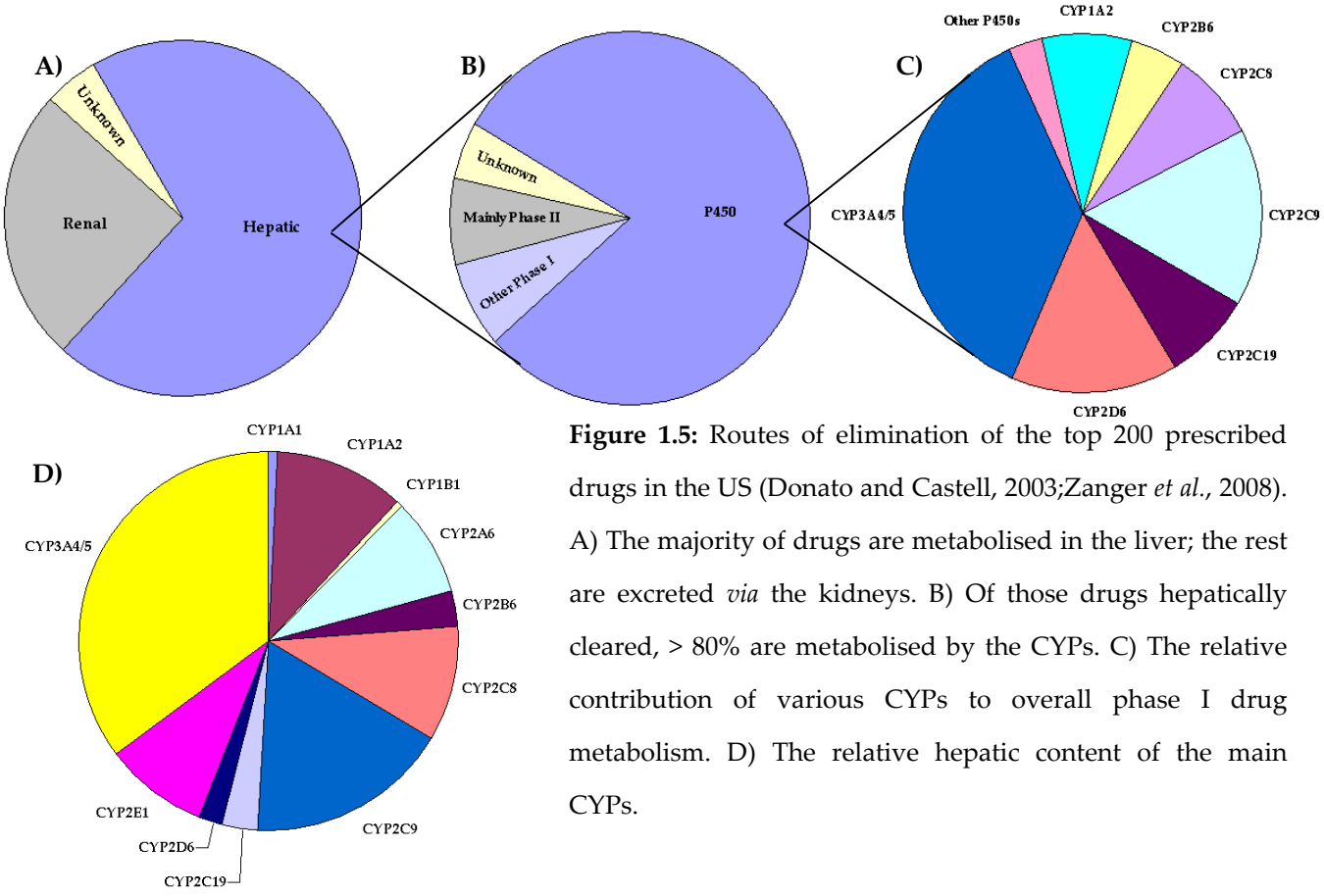


Figure 1.5: Routes of elimination of the top 200 prescribed drugs in the US (Donato and Castell, 2003;Zanger *et al.*, 2008). A) The majority of drugs are metabolised in the liver; the rest are excreted *via* the kidneys. B) Of those drugs hepatically cleared, > 80% are metabolised by the CYPs. C) The relative contribution of various CYPs to overall phase I drug metabolism. D) The relative hepatic content of the main CYPs.

CYPs are also involved in processes detrimental to human health. The poor coupling efficiency of the CYP catalytic cycle leads to constant production of reactive oxygen species (ROS) that can in turn lead to lipid peroxidation, cellular toxicity, and eventually cell death (Zangar *et al.*, 2004). Furthermore, despite the metabolism of foreign compounds usually being a detoxification process, in some instances foreign compounds are converted into products with much greater cytotoxicity, mutagenecity and carcinogenecity, a phenomenon termed 'bioactivation' (Ioannides and Lewis, 2004). These deleterious products are typically electrophiles that interact covalently with DNA and proteins causing mutations, cell death and immunotoxicity through stimulation of autoantibody production. Electrophiles can also interact with O_2 to form ROS, adding to that already produced during the catalytic cycle. Bioactivation is thought to be carried out by a limited number of CYPs, usually CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2B6, CYP2E1 and CYP3A4.

The CYP1 family is undoubtedly the dominant CYP family involved in the bioactivation of chemicals including many carcinogens such as air pollution, cigarette smoke, and dietary contaminants such as charred food (Ioannides and Lewis, 2004). In fact, the CYP1 family is believed to be responsible for the bioactivation of >90% of known carcinogenic chemicals into genotoxic metabolites. Most of these carcinogens are planar molecules, which are favoured substrates and inducers of these enzymes. This CYP family has therefore been the focus of considerable research due to their association with the etiology of certain cancers. Indeed, polymorphisms that impact the transcription levels or catalytic activities of these CYPs could prove to be useful markers for the predisposition of affected individuals to these cancers.

The main human xenobiotic-metabolising CYPs have been well studied and characterised. Table 1.3 below lists of some of the well-known substrates, inhibitors and inducers for each of these CYPs, together with their localisation, level of inducibility and the extent of their role in bioactivation. The compounds indicated are common clinical drugs, dietary constituents or environmental contaminants, which provides a glimpse into the origin of many adverse drug reactions.

Isoform	Main localisations	Inducibility	Role in bio-activation	Substrates	Inhibitors	Inducers
CYP1A2	Liver	+++++	+++++	Caffeine Ethinylestradiol Paracetomol Phenacetin Theophylline	Fluvoxamine Oral contraceptives	Charred meat Cigarette smoke Cruciferous vegetables Omeprazole
CYP2B6	Liver	++++	+++	Bupropion Cyclo-phosphamide Efavirenz Nevirapine Nicotine Propofol Selegiline S-mephenytoin	Ticlopidine	Cyclo-phosphamide Efavirenz Phenobarbital Rifampicin Statins Vitamin D
CYP2C8	Liver	+++	++	Cerivastatin Chloroquine Paclitaxel Replaginide Retinoic acid Rosiglitazone	Efavirenz Gemfibrozil Montelukast Trimethoprim	Barbituates Rifampicin
CYP2C9	Liver	+++	++	Diclofenac Ethinylestradiol Ibuprofen NSAIDS Progesterone S-mephenytoin Tolbutamide Warfarin	Fluconazole Sulphaphenazole	Barbituates Rifampicin
CYP2C19	Liver	+++	++	Diazepam Hexobarbitol Omeprazole Phenobarbital	Fluvoxamine	Barbituates Rifampicin
CYP2D6	Mainly liver, but also duodenum and brain	-	+	Bufuralol Codeine Desipramine Dextromethorphan Haloperidol Metoprolol, Propafenone	Fluoxetine Qunidine	No major inducers identified
CYP2E1	Mainly liver, but also lungs and brain	++++	++++	Chlorzoxazone Halothane Nitrosamines Paracetomol	Disulfirum	Alcohol Isoniazid Organic solvents
CYP3A4	Most abundant CYP in liver and intestine	++++	+++	Aflatoxin B1 Caffeine Cyclosporin Erythromycin Lovastatin Midazolam Nifedipine Progesterone Quinidine Ritonavir Testosterone Verapamil	Ethinylestradiol Ketoconazole Grape fruit juice Nefazodone	Barbituates Efavirenz Glucocorticoids Nevirapine Phenobarbital Rifampicin Statins St. John's wort Troglitazone
CYP3A5	Mainly extrahepatic;	++++	+++	Same as CYP3A4 (usually at lower	Likely to be the same as CYP3A4	Phenobarbital Rifampicin

Isoform	Main localisations	Inducibility	Role in bio-activation	Substrates	Inhibitors	Inducers
	liver expression highly variable in adults, although major CYP in fetal liver			turnover rates)		

Table 1.3: Characteristics of the main xenobiotic-metabolising human cytochrome P450s (Danielson, 2002;Ioannides and Lewis, 2004;Zanger *et al.*, 2008).

1.2 PHARMACOGENOMICS, DRUG DEVELOPMENT AND THE IDEAL OF PERSONALISED MEDICINE

1.2.1 Overview of the drug development process

In order to appreciate the impact of the cytochrome P450s on drug development, one needs to have an understanding of the process of drug development itself. This process consists of two principal stages: drug discovery and drug development.

1.2.1.1 The drug discovery stage

The drug discovery stage includes target selection, lead identification and pre-clinical studies (Kang *et al.*, 2008). Libraries of compounds are initially synthesised with the drug target in mind and leads are identified by screening thousands of similar compounds for specific characteristics such as desirable physicochemical parameters and target binding affinity. The more promising of these are then tested in pre-clinical studies using a variety of *in vitro* assays and animal models. Drug companies currently perform a fairly standard package of pre-clinical studies before commencing with 'First-In-Man' clinical trial investigations, comprising kinetic, safety pharmacology, and toxicology studies (Baldrick, 2008b). These studies are necessarily in agreement with regulatory guidance issued by regional bodies (such as the FDA and the EMEA) as well as that issued internationally *via* the International Conference on Harmonisation (Huang *et al.*, 2007).

Kinetic studies:

Early investigations explore whether a drug is eliminated primarily through excretion or metabolism and, if the latter, what the major metabolic routes are (Baldrick, 2008b;Wishart, 2007).

A key *in vitro* kinetic study is examining the metabolism of the drug using hepatocytes or liver microsomes from a range of species (often mouse, rat, dog and monkey), including human. In many studies, cytochrome P450 enzyme activity is used not only for isoform mapping (i.e. determining which CYPs are responsible for metabolism of the drug), but also as a marker for enzyme induction, inhibition and drug interactions. These CYP studies use various model systems such as hepatocytes, liver slices and recombinant enzymes (see Section 1.2.5). In some cases, other *in vitro* kinetic studies are undertaken such as serum and plasma protein binding, blood-brain barrier penetration (using bovine brain microvessel cells), and gastrointestinal adsorption (using Caco-2 or MDCK cells).

In vivo kinetic studies usually use single dose pharmacokinetic studies in a rodent and non-rodent model, with separate oral and intravenous studies (Baldrick, 2008b). Most of these studies calculate exposure [maximum plasma level (C_{max}) and area under the curve (AUC)], as well as time to maximum plasma level (T_{max}). In some cases, mass balance examinations in rodents are carried out, which involve investigating the levels of the parent drug and its metabolites in the faeces and urine. Whole body autoradiography in rodents is also sometimes performed to determine systemic drug distribution.

Safety pharmacology studies:

Safety pharmacology studies investigate the potential undesirable effects of a drug on physiological functions using drug exposure in the therapeutic range and above (Baldrick, 2008b). A 'core battery' comprising central nervous system (CNS), cardiovascular function, and respiratory function studies is critical. An *in vivo* cardiovascular and respiratory study is usually performed in the non-rodent (dog or monkey) using anaesthetised animals. Commonly measured parameters include heart rate, blood pressure, systemic resistance, and respiration rate. CNS function is assessed *in vivo* using a general activity and behaviour study in rodents. In addition, follow-up safety studies (e.g. renal/urinary or gastrointestinal function) may also be required dependent on the properties of the drug.

Developers are now expected to perform *in vitro* cardiovascular safety studies as well, to further investigate a novel drug's potential to delay ventricular repolarisation (Baldrick, 2008b;Kramer *et al.*, 2007). This is carried out using electrophysiology studies, the most common of which is the hERG (human ether-à-go-go-related gene) assay. Potential hERG channel blocking is assessed

with either a competitive binding assay or a patch-clamp assay (using a microelectrode to study the activity of the hERG ion channel in transfected cells). Blocking of the hERG potassium channel may be predictive for QT interval prolongation and ultimately for the potential to cause Torsades de Pointes, a rare but potentially life-threatening drug-induced ventricular tachyarrhythmia. When an effect is observed in the hERG assay, it is often accompanied by a non-rodent Purkinje fibre or papillary muscle assay.

Toxicology studies:

Toxicology studies involve both *in vivo* and *in vitro* assays (Baldrick, 2008a;Kramer *et al.*, 2007).
In vitro toxicology assays include assays for cytotoxicity, genotoxicity and metabolite-mediated toxicity. Of these, cytotoxicity assays using primary cells or tissue slices are often among the earliest toxicity assays to be conducted. The most commonly used genotoxicity assays for predicting potential carcinogenicity are the bacterial Ames test, the *in vitro* chromosome aberration or mouse lymphoma assays, and the *in vivo* rodent bone marrow micronucleus test. The Ames test detects frameshifts and point mutations in bacteria, while the latter tests detect chromosomal aberrations *in vitro* and *in vivo*. Metabolite-mediated toxicity studies are performed using the above methods to determine whether the drug's metabolites are reactive or electrophilic, as these too have the potential for genotoxicity and cytotoxicity.

In vivo assays performed include single dose toxicity studies in the rodent of up to 2000 mg/kg (the IC_{50} parameter was abolished in 2001), and range-finding toxicity and 2-4 week repeat dose toxicity studies in the rodent and non-rodent (Baldrick, 2008a). Common parameters measured are body weight, number of deaths, food consumption, clinical pathology (haematology, clinical chemistry and urinalysis), organ weights, macroscopic examinations, and histopathology.

Many computational models have been developed to predict the ADME/Tox properties of drug candidates, such as bioavailability, aqueous solubility, intestinal absorption, blood-brain barrier penetration, drug-drug interactions, transporter interactions, plasma-protein binding and toxicity (Hou *et al.*, 2006;Wishart, 2007). However, these *in silico* models still need to be augmented with the above *in vitro* and *in vivo* methods before compounds can proceed to clinical trials.

1.2.1.2 The drug development stage

The drug development stage includes clinical trials, manufacturing and product lifecycle management (Kang *et al.*, 2008;Pharmaceutical Research and Manufacturers of America, 2009). Clinical trials are divided into three phases. Phase I consists of a small number of healthy volunteers (20-100 people) and is mainly focused on ascertaining safe and tolerable doses of the drug culminating in an initial estimate of the optimal dose and regimen. In this phase, pharmacokinetic (PK) parameters such as adsorption, distribution, metabolism and excretion (ADME) are first investigated and any adverse reactions are noted. During Phase II a larger group is recruited (100-500 people), consisting of individuals presenting the target disease or condition. In this phase the most effective dose regimen is verified and adverse reactions are monitored using increasing doses for longer periods. Phase III is similar, except that it is randomised and controlled with hundreds and sometimes thousands of patients involved (usually 1000-5000 people), for final confirmation of drug efficacy and safety. It is in Phases II and III that the pharmacodynamic (PD) parameters of the drug are mainly investigated, such as the drug's effect on its targets and downstream signalling events, as well as the overall pharmacological response. PD parameters vary with different types of diseases; for example, the level of glucose in circulating blood is a PD parameter that can be monitored for the treatment of diabetes.

1.2.2 Problems facing the pharmaceutical industry

The pharmaceutical industry is arguably facing a future financial crisis due to an unsustainable increase in R&D spending coupled with a concomitant decrease in productivity. In the past decade, there has been a dramatic increase in the number of new chemical entities and screenable drug targets as a result of combinatorial chemistry and rapid advances in fields such as genomics, proteomics, chemistry and nanotechnology. However, this era has seen fewer drug approvals, several high-profile drug withdrawals and escalating development costs (Caskey, 2007;Woodcock, 2007).

1.2.2.1 Drug efficacy and toxicity

The failure rate from unacceptable pharmacokinetics (ADME) has been greatly reduced over the last several years due to improved pre-clinical screening of drug candidates (Blake, 2005;Peck, 2007). However, the increasingly large number of compounds in the lead discovery stage currently precludes anything more than limited screening and the three leading causes of attrition

in drug discovery are now attributed to problems with drug efficacy (30%), safety and toxicology (33%), and commercial reasons (20%).

However, even if a drug succeeds through each R&D stage and eventually receives approval, there is still no guarantee that the drug will deliver on its promises to the public. Despite escalating costs, statistics show that the great majority of drugs on the market are only effective for around 30 - 70% of the patients to whom they are prescribed; in the anti-cancer field, this figure drops to around 20% (Sadée and Dai, 2005;Wilkinson, 2005). Furthermore, even some of the most advanced drugs still induce adverse side effects in a significant portion of the target population.

Statistics from the US show that up to 7% of all hospital admissions are related to adverse drug reactions (ADRs) (Ingelman-Sundberg, 2008), and of the >2 million cases of ADRs that are reported annually, 100,000 result in deaths (Lazarou *et al.*, 1998). This obviously places tremendous pressure on a country's healthcare system. For example, costs associated with drug-related problems in the US have more than doubled from 1995 to 2000 - from $76.6 billion to $177.4 billion (Güzey and Spigset, 2004).

Many drug toxicities are only recognised at a late phase in clinical trials after an already enormous investment in the drug candidate (Caskey, 2007). However, previously unobserved adverse drug reactions can also appear once a drug has been approved for use in the general population, since, prior to approval, the drug has usually only been tested in ~1000-2000 individuals. Groups of this size are only large enough to detect an ADR with an occurrence of 1 in every 250 people (Hodges, 2008). In fact, around 4% of all new approved drugs are withdrawn from the market due to this problem (Ingelman-Sundberg, 2008). During the period of 1995-2005 alone, 34 drugs were withdrawn mainly as a result of hepatoxic or cardiotoxic effects, e.g. Baycol, Serzone, Vioxx, Triludan, and Rezulin. Vioxx, a multibillion dollar blockbuster analgesic drug, was the biggest ever withdrawal in the US wiping more than $25 billion off Merck's market value that day (Kolata, 2004;Merck & Co., 2004). It was withdrawn worldwide in 2004 after five years on the market, due to cardiovascular complications in those who took it for more than 18 months. Other than causing adverse reactions, withdrawals are also sometimes due to an elevation of risk for a disease not related to the drug's intended usage, or interactions with other drugs often used in combination therapy (Caskey, 2007).

Among the most common causes of pre-clinical toxicity-related attrition is target organ toxicity, which can be difficult to predict *in vitro*. Even for target organs like the liver, for which primary cells and tissue slices are available, *in vitro* assays often cannot reliably predict safety margins; hence the continuing use of animal models. However, animal models are currently only ~70% efficient in predicting toxicity in the human liver, heart and during human development: cardiotoxicity and hepatotoxicity remain the highest causes of safety-related drug attrition both during clinical trials and after market launch (Mayne *et al.*, 2006;Olson *et al.*, 2000;Ukelis *et al.*, 2008).

For these reasons, drug safety is continuously re-evaluated throughout the marketing lifetime of the product. This is done primarily by monitoring reports of adverse reactions through the 'pharmacovigilance' system in which pharmaceutical companies are required to collate such information and report urgent or periodic data to the public authorities (Hodges, 2008). However, a retrospective review has suggested that less than 5% of all ADRs are ever reported (Fletcher, 1991).

1.2.2.2 The costs of drug development

The average cost of taking a drug through to market using data from 1983 to 2000 was found to be approximately $802 million (DiMasi *et al.*, 2003). This figure is up from $54 million in 1976 and $231 million in 1987, representing an annual inflation rate of 7.4% above general price inflation (see Fig. 1.6). Refined estimates using more recent data suggest that each successful drug candidate now has to recoup over $1.3 billion, most of which is R&D costs absorbed from failed drug candidates (Pharmaceutical Research and Manufacturers of America, 2009). The average return on investment is down from 9% to 5%. In addition to this cost escalation, 4% of all approved drugs are later withdrawn leading to lost income and potentially to expensive, high-profile litigation (Ingelman-Sundberg, 2008).

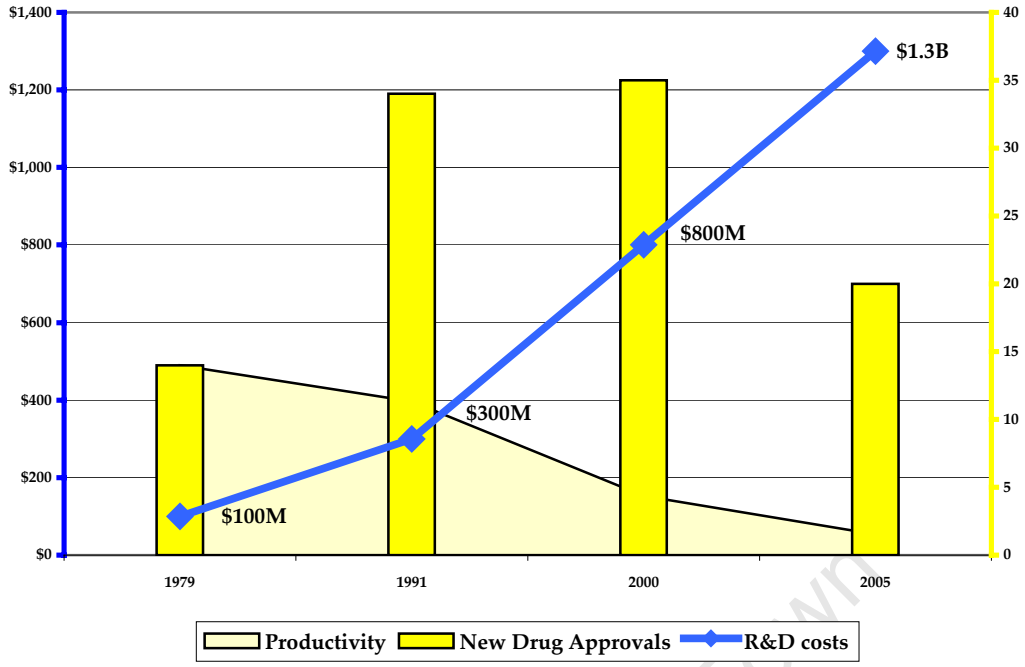


Figure 1.6: Within the last three decades, the average cost to develop a drug has risen from around $100 million to more than $1.3 billion. This increase in R&D expenditure has been accompanied by a decrease in productivity, which is expressed here as the number of drugs approved as a percentage of total R&D costs. [Adapted from (Cockburn, 2006;Pharmaceutical Research and Manufacturers of America, 2009).]

Even when a drug is successful, it is not necessarily profitable. A mere two out of ten approved compounds generate sufficient revenue to recover their average R&D costs (Pharmaceutical Research and Manufacturers of America, 2009). To make matters worse, the development timeline has stretched to 10-15 years, while a typical patent length is only 20 years (Hodges, 2008). This leaves drug manufacturers with less and less time within which to recover their costs before their patent expires and generic copies are permitted to enter the market. Over and above the pre-marketing development costs, a Phase IV stage is also increasingly being implemented, to carry out post-approval research and post-marketing surveillance (Pharmaceutical Research and Manufacturers of America, 2009).

In addition, a large number of the blockbuster drugs from the 1970s and 1980s are approaching or have reached their patent maturity and will no longer serve as a major source of funds for further drug innovation (Sollana *et al.*, 2008). These blockbuster drugs are not being replaced by new products - in fact, only around 6% of new drugs represent significant clinical improvements, rather than being 'me too' drugs (Hodges, 2008). This has affected investors' confidence leading to significant falls in share prices.

1.2.2.3 Increasing attrition rates in the development process

According to the FDA, of those drugs entering Phase I clinical trials, only 1 in 13 are successfully making it to market compared to 1 in 8 ten years ago (FDA, 2006a). The attrition rate in phase II is now over 70% and rising, and even in phase III trials products fail 50% of time, compared to 20% ten years ago (see Fig. 1.7) (Woodcock, 2007). This increasing attrition rate of lead compounds is a significant contributing factor to the escalation of R&D costs, as the clinical trial stage of the R&D process absorbs over 50% of the total costs (Pharmaceutical Research and Manufacturers of America, 2009). The majority of these funds goes towards supporting the Phase III trials, the average size of which has tripled in the past 20 years (Lesko and Woodcock, 2004).

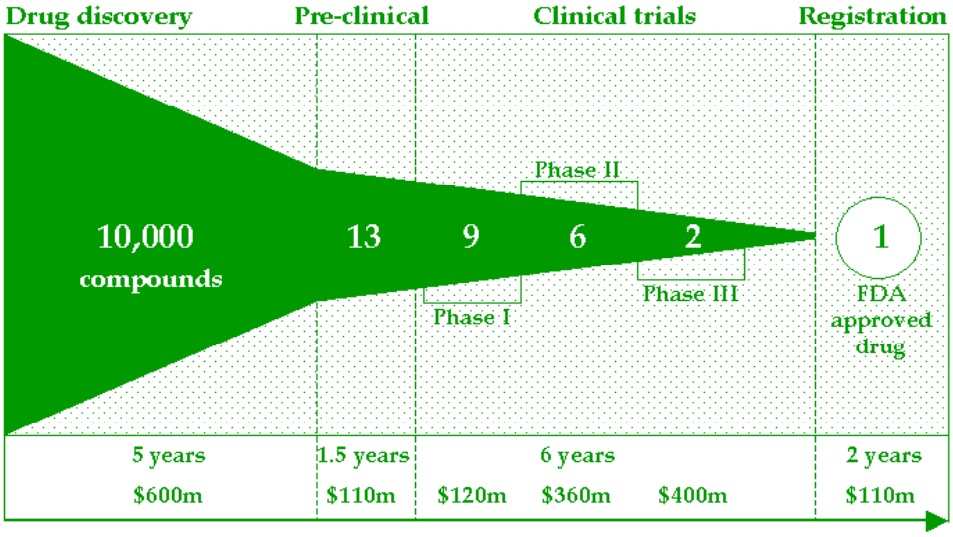


Figure 1.7: The drug development pipeline showing the number of lead compounds progressing through each stage. A typical pipeline now spans 10-15 years with an average cost of $1.3 billion. [Adapted from (Gilbert *et al.*, 2003;Pharmaceutical Research and Manufacturers of America, 2009;Woodcock, 2007).]

As mentioned earlier, the main causes of attrition in drug development are attributable to problems with drug efficacy and safety. This indicates that current approaches to understanding drug efficacy and safety before a compound enters clinical trials or even the market are limited, as the development process is not as effective as it should be in filtering out unsuitable drug candidates. There have been several suggestions as to how this may be improved.

One suggestion is that, ironically, the problem of increasing attrition due to lack of drug efficacy is a result of a shift towards target-based discovery following the remarkable scientific and technological advances over the last 10-15 years. It is thought that this shift may have contributed to increasing attrition rates by increasing the effects of bias on pre-clinical assessments of

potential efficacy (Lindner, 2007). The effects of bias in pre-clinical studies are thought to be at least as large as that in clinical studies, however bias is not usually controlled in pre-clinical studies and compounds that actually have little or no efficacy may often be advanced into clinical trials. It is unlikely that this phenomenon is due to limitations in the predictive ability of pre-clinical models, as these models are still at least as predictive as they were 30-40 years ago.

It is also suggested that the downgrading or termination of natural product research programmes by the majority of large pharmaceutical companies may also have contributed to the decline in the number of successful chemical entities in recent years (Lam, 2007). The late 1980s and early 1990s saw the introduction of combinatorial chemistry and dramatically increased assay speeds due to high-throughput screening technology. The promise of a ready supply of large synthetic compound libraries led pharmaceutical companies to discard or drastically reduce the natural product screening paradigm that already existed, as it was not compatible with this 'blitz' screening approach with a timeline of three months.

While these issues may well factor into the reasons for increasing attrition rates, they cannot account for the underlying problems with drug efficacy and toxicity that have always been present in drug development. Drugs have always offered a very poor risk/benefit ratio for patients. Added to this, recent innovations in drug discovery have come at a price that is unsustainable for industry as well as for patients and healthcare systems. The challenge is therefore for pharmaceutical companies to design new strategies that strengthen innovation *and* rigorously evaluate drugs throughout their development. The ultimate challenge is to do this in a financially sustainable manner in the face of shorter patent periods as well as tighter price controls and reimbursement criteria (Lichtenburg, 2005).

1.2.3 Pharmacogenetics and pharmacogenomics

One underlying problem causing the failure of drugs in development and on the market is believed to be a lack of understanding of the variability in drug response between patients (Lesko and Woodcock, 2004). It has long been known that patients vary greatly in their response to pharmaceuticals: some metabolise a drug rapidly, while others metabolise the same drug relatively slowly; some experience serious side affects while others only mild reactions; some show excellent therapeutic response while for others little or no response is observed (Danielson,

2002). In the effort to extricate the underlying causes of this variability, the fields of pharmacogenetics and pharmacogenomics are receiving ever-increasing attention and for the past year alone, over 1,200 papers result from a PubMed search for either 'pharmacogenetics' or 'pharmacogenomics' - over 15% of the total number published so far (see http://www.ncbi.nlm.nih.gov/sites/entrez).

The terms pharmacogenetics (PGt) and pharmacogenomics (PGx) are often used interchangeably (Lesko and Woodcock, 2004;McLeod, 2003). Pharmacogenetics is defined as the study of genetic variation that results in differing drug response between individuals. Such studies usually examine only a few genes at a time, and PGt can therefore be viewed as a 'subset' of pharmacogenomics. Pharmacogenomics looks at the variation across the entire genome, either at the cellular, tissue, individual or population level, thereby integrating pharmacology with modern advances in genome analysis. It therefore uses a polygenic or genome-wide approach to identifying genetic determinants of drug response.

It is estimated that there are approximately 170 genes whose products affect drug disposition and that over half of these are polymorphic (Bhathena and Spear, 2008). Many of these have consistently replicated associations between genetic variants and the clinical PK of at least one drug. Although a number of non-genetic factors are known to influence clinical outcome (e.g. age, sex, body weight, diet, organ function, disease, co-morbidity, co-medications, poor diagnosis and lifestyle), genetic variation has been shown to underlie many of these inter-individual differences (Koch, 2004;Wilkinson, 2005). A study based on twins found that genetic variability is expected to account for 20-95% of the variability depending on the drug and the condition being treated (Kalow *et al.*, 1998).

The ideal of PGx is not only to understand the basis of the variation observed for drug efficacy and toxicity, but also to ultimately realise the goal of personalised medicine. The knowledge gained in this field has the potential to revolutionise both the use of many current medications, as well as the development of future drugs (McLeod, 2003;Sadée and Dai, 2005). Physicians will be able to prospectively identify patients at risk for severe toxicity or poor efficacy from a particular drug. This can be achieved using genotyping kits to determine which alleles the patient has of genes important in the drug's metabolism and response. Genotyping is simple to perform and results can be gained within 24 hours. A small sample of urine, blood or tissue is required and the

test needs to be done only once in a lifetime. PGx therefore promises to provide a paradigm shift from the 'one-drug-fits-all' mentality to that of 'the right drug for the right patient at the right dose'.

Since the genomes of two unrelated persons share over 99.9% DNA sequence identity, it is the remaining 0.1% of DNA that gives rise to the observed differences in drug response and disease susceptibility (Vizirianakis, 2007). The many types of genetic variation within the human genome (which consists of >3 billion base pairs) include insertions, deletions, multiplications, rearrangements and point mutations (see Table 1.4). These events can affect millions of base pairs or cause variations in single nucleotides. There are also epigenetic sources of variation, such as differences in DNA methylation, histone acetylation and regulatory microRNA levels.

Type of variation	Abbreviation	Frequency in human genome
Single nucleotide polymorphisms	SNP	> 12,000,000
Deletions/Insertions	InDel	> 1,000,000
Varying number of tandem repeats	VNTR	> 500,000
Copy number variation	CNV	> 1500 loci covering 12% of the genome
DNA methylation	-	> 20% of all genes

Table 1.4: Types and frequencies of human genetic and epigenetic variation. [Adapted from (Brockmöller and Tzvetkov, 2008).]

Variations of single nucleotides, called single-nucleotide polymorphisms (SNPs), probably hold the most promise for PGx analyses, as they are the most common class of genetic variation. Currently, there are >24 million unique SNPs catalogued in the human SNP database, dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/index.html), including 12 million common SNPs (i.e. SNPs that occur at a frequency of >5%). SNPs within the protein-coding region of genes can either be silent (i.e. the amino acid sequence is unaffected), or result in amino acid substitutions, the introduction of stop codons, or the alteration of mRNA splice sites - some of which can have a dramatic effect on protein expression, stability and function (Brockmöller and Tzvetkov, 2008). In addition, genetic variation within the non-coding portions of the gene such as the introns or the promoter region, can affect transcription levels, RNA splicing and mRNA stability.

1.2.4 Pharmacogenetics of the cytochrome P450s

Inter-individual differences in drug response were probably first noted by Pythagoras in 510 B.C. when he observed that some, but not all, individuals fell ill after ingesting uncooked fava beans (Ingelman-Sundberg, 2004). It was only in the 1950s that the association was found between fava beans, haemolytic anaemia, and a deficiency in glucose-6-phosphate dehydrogenase (Carson *et al.*, 1956). Genetic variation has since been discovered in all the principal effectors of ADME/Tox and drug response - drug transporters, drug receptors, drug metabolising enzymes, and drug targets - as well as in the promoter elements regulating the expression of these proteins (Koch, 2004).

Of these, genetic variation in the cytochrome P450s is thought to play a significant role in variable drug response. Among the top 200 drugs sold in the greatest quantities in the US, a 2008 study showed that 78% of the hepatically cleared drugs are primarily cleared by the cytochrome P450s, while 48% of these are metabolised by CYPs with known functional variants (Zanger *et al.*, 2008). In addition, of all the drugs cited in studies as responsible for adverse drug reactions, only 20% of these are metabolised by non-polymorphic enzymes, while 48% of them are metabolised by the polymorphic CYPs (Phillips *et al.*, 2001).

The three CYPs responsible for the primary metabolism of the majority of drugs are CYP3A4, CYP2D6 and CYP2C9. These CYPs are highly polymorphic, with over 100 different alleles for CYP2D6 alone [PharmGKB Database: http://www.pharmgkb.org/ (Hernandez-Boussard *et al.*, 2008)]. Site-directed mutagenesis studies have shown that coding region mutations in the CYP enzymes can affect substrate specificity, enantioselectivity, regioselectivity, and turnover rates, with the latter differing as much as 1000-fold between polymorphic variants (Ingelman-Sundberg, 2004). The effect of these polymorphisms on CYP-drug interactions can therefore be at a number of different levels, such as the affinity of the drug for the CYP substrate-binding site, as well as the site of metabolism and therefore potentially the metabolites produced. However, even when the crystal structure of a protein is known, we still lack the ability to accurately predict the effect of a mutation on the folding and function of that protein (Bromberg and Rost, 2009). To add to this complexity, the effect of a specific mutation on substrate turnover is not necessarily the same for all substrates; for example, allele CYP3A4*7 has a G56D mutation and shows decreased clearance for nifedipine and midazolam compared to wildtype, while testosterone turnover is unaffected (Miyazaki *et al.*, 2008). The same is true for the effect of inhibitors.

Genotype-phenotype associations within the drug metabolising enzymes such as the CYPs are normally classified into the following groups: poor metaboliser (PM), intermediate metaboliser (IM), extensive metaboliser (EM), and ultra-rapid metaboliser (UM). Poor metabolisers usually lack any functional enzyme; intermediate metabolisers are heterozygous with one non-functional allele, or homozygous for two partially functioning alleles; extensive metabolisers have two normal, functional copies of the allele; and ultra-rapid metabolisers have multiple functional gene copies (see Fig. 1.8) (Ingelman-Sundberg, 2004). Poor metabolisers are at risk of experiencing adverse reactions due to prolonged exposure, while ultra-rapid metabolisers are at risk of poor therapeutic response due to sub-therapeutic concentrations of the drug reaching its target (as well as possible accusations of non-compliance or addiction-seeking behaviour). The converse is true when the compound is a 'prodrug' such as codeine (metabolised into morphine); i.e. a drug that is therapeutically inactive until it is metabolised.

However, as alluded to above, these traditional PM, IM, EM and UM designations do not necessarily hold true across all drug substrates for a particular CYP allele. A polymorphism present on a specific CYP allele might decrease the turnover of one particular substrate quite dramatically, while the turnover of another substrate is not significantly affected. In other words, these genotype-phenotype associations are not broadly applicable and are instead drug-specific. This implies that knowledge of a patient's genotype is not sufficient in itself: unless prior studies have been carried out to determine the effect of the genotype on the particular drug being prescribed, the clinical outcome will remain unclear.

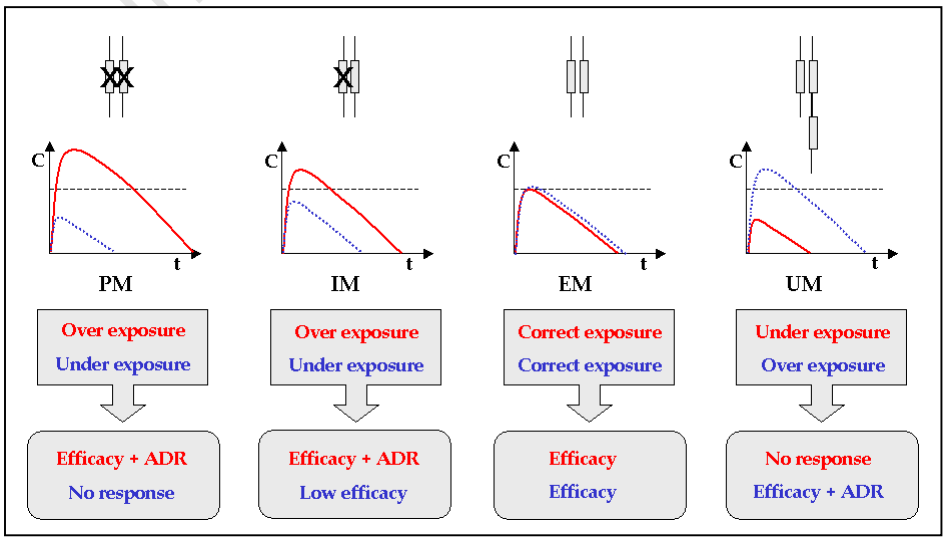


Figure 1.8: An overview of DME genotype-based drug response for a drug (*red*) and a prodrug (*blue*). *C* is the plasma drug concentration and *t* is time. (Bhathena and Spear, 2008;Roots *et al.*, 2004)

Below follow several examples illustrating how the CYPs contribute to variable drug response.

1.2.4.1 Variable efficacy and toxicity

An often-cited example of the importance of pharmacogenetics in drug metabolism is warfarin, one of the most prescribed oral anticoagulants. Warfarin is used in the treatment and prophylaxis of myocardial infarction, thromboembolism and stroke and is mainly metabolised by CYP2C9 (Dervieux *et al.*, 2005;Thomas *et al.*, 2004). The wild-type alleles are predominant; however there are several polymorphic alleles associated with either the risk of bleeding and haemorrhage (when the dose is too high), or of the formation of blood clots (when the dose is sub-therapeutic). Both of these events can be life threatening. In addition to the effect of CYP2C9 variation, polymorphisms in CYP4F2 and the target of warfarin, vitamin K epoxide reductase complex 1 (*VKORC1*), are associated with up to 55% of the variability observed between patients and consequently with lower or higher warfarin dose requirements (Glurich *et al.*, 2008). Despite the large genetic contribution to clinical outcome, warfarin dosing is still typically done by trial-and-error, where an initial dose ranging from 2-10 mg is selected based on therapeutic indication and other factors (Bhathena and Spear, 2008). The maintenance dose is then obtained by adjusting the dose while monitoring the International Normalised Ratio until it reaches and stabilises in the target range. A wide inter-individual variability in dose-response is observed, and weekly maintenance doses range from 4-80mg.

Warfarin remains the drug of choice for long-term anticoagulation management with no effective alternative therapeutic approaches available as yet, however it is under-utilised due to the high risks of serious adverse events (Glurich *et al.*, 2008). This under-utilisation is mainly attributable to the fact that warfarin is ranked among the top ten drugs associated with serious adverse drug events over the past two decades, with 10% of warfarin-related bleeding events between 1993 and 2006 resulting in death. Other problems include the lack of precise dosing parameters and the need to monitor patients on a regular basis. This under-utilisation is placing a heavy burden on healthcare - the morbidity, mortality and cost of care associated with preventable stroke are staggering, with the Centres for Disease Control in the US estimating costs related to stroke at $62.7 billion in 2007 alone.

On the other hand, there are an estimated two million Americans taking the drug each year (Woodcock, 2007) and adverse events due to incorrect warfarin dosing are estimated to cost an

average of $800/patient/year (Hamby *et al.*, 2000). Amongst current patients, warfarin PGt testing could therefore save an estimated $1.1 billion in US healthcare spending each year, while preventing 17,000 strokes and 85,000 serious bleeding events (McWilliam *et al.*, 2006).

In an attempt to start addressing the above issues, in 2007 the FDA updated the warfarin label with pharmacogenetics information (http://packageinserts.bms.com/pi/pi_coumadin.pdf), however genetic testing is still not mandatory for patients initiating warfarin therapy due to persistent unresolved questions (Glurich *et al.*, 2008). Large, prospective trials that provide statistically significant associations between genetic variants and phenotypic outcomes still need to be completed, as well as studies to determine the cost-benefit ratio of warfarin PGt testing. Once these are concluded, it is likely that testing will become mandatory. The outcome of such studies should be a rigorous yet easily applied algorithm that can be used across populations, taking both genetic and non-genetic factors into account.

Another prominent example of CYP variation resulting in variable drug efficacy is that of the metabolism of the antiestrogen agent, tamoxifen, by CYP2D6. Tamoxifen is used in the treatment of estrogen receptor-positive breast cancer and as a chemopreventative agent for patients at risk of developing breast cancer. This drug is known to significantly reduce recurrence and mortality rates, however 30-50% of patients on tamoxifen therapy relapse or die (Yong and Innocenti, 2007). Studies have since demonstrated a strong association between CYP2D6 PM and IM genotypes and a non-favourable treatment outcome in breast cancer patients (Schroth *et al.*, 2007). In 2006 the FDA recommended a label change for tamoxifen, mentioning the option of genetic testing before commencing tamoxifen treatment. With this information, clinicians could offer aromatase inhibitors instead of tamoxifen to postmenopausal woman homozygous for non-functioning CYP2D6 alleles.

CYP enzyme	Allele	Examples of associated ADRs
CYP2C9	CYP2C9*2 and CYP2C9*3	Warfarin: haemorrhage, stroke Phenytoin: phenytoin toxicity Tolbutamide: hypoglycemia
CYP2D6	CYP2D6*10, CYP2D6*17	Propafenone: arrhythmias Metoprolol: bradycardia Nortriptyline: confusion, hallucinations, hypertension Opioids: dependence Perhexilene: hepatoxicity Oxycodone: nausea, vomiting, CNS and respiratory depression

Table 1.5: Further examples of adverse drug reactions associated with cytochrome P450 coding region variants. [Adapted from (Pirmohamed and Park, 2003).]

1.2.4.2 Variable drug-drug interactions

CYPs are also implicated in adverse drug reactions when co-administered drugs compete for metabolism by the same CYP, or when one drug affects the CYP activity necessary for the other drug's metabolism. Observed changes arising from drug-drug interactions can be substantial – an order of magnitude or more in the blood and tissue concentrations of the drug and metabolite (Huang *et al.*, 2007). The clinical effect of these CYP-mediated drug-drug interactions will of course be influenced by genetic variation within the CYPs involved, as different genotypes could either mitigate or exacerbate the outcome.

ADRs that are likely to increase, particularly in the developing world, are those resulting from the ingestion of non-prescribed substances that are freely available and are widely perceived as 'safe'. Herbal medicines such as St. John's Wort and ginseng and foodstuffs such as garlic and grapefruit have produced serious clinical interactions when co-administered with prescription drugs (Güzey and Spigset, 2004). For example, pregnancies have been reported after concomitant treatment with St. John's wort and oral contraceptives, whilst this herb has also been implicated in organ rejection due to interactions with the immune suppressant cyclosporin.

Herbal and traditional medicines are not always well researched, are poorly regulated and can contain adulterated products, all of which pose serious health risks (Mills *et al.*, 2005;Patel *et al.*, 2004). An example particularly important in the African context is the widespread use of traditional medicines such as African Potato for the treatment of HIV/AIDS. Compounds present in African Potato have been shown to inhibit normal CYP3A4 activity by almost 90%. When this traditional remedy is used in conjunction with anti-HIV drugs that are CYP3A4 substrates (e.g. nifedipine and ritonavir), these drugs can build up to toxic levels in the patient's system potentially causing severe adverse reactions.

1.2.4.3 Variable CYP-drug transporter interactions

Only over the last 15 years have pharmacologists become fully aware of the importance of the transmembrane drug transporters in the uptake, circulation and elimination of drugs (Brockmöller and Tzvetkov, 2008), and for some time it was thought that drug distribution was mainly due to passive diffusion (Roots *et al.*, 2004). It is now clear that to completely understand the pharmacokinetics of drugs, phase 0 (influx transport), phase I (mostly oxidative or reductive biotransformation), phase 2 (conjugations) and phase 3 (efflux transport), all have to be

considered. There are only a few drugs for which all the transporter proteins responsible for their uptake and elimination have been identified. One of the human multiple drug resistance (MDR) conferring transporters, P-glycoprotein/MDR1 (P-gp), was among the first transporters extensively screened for genetic polymorphisms. A number of functionally significant variants have been found for this protein, although these known variants do not correlate well with phenotypic variation (Hirota *et al.*, 2008) and in general, P-gp diminishes drug absorption and enhances drug excretion (Christians, 2004).

With regards to the effects of CYP polymorphisms on drug ADME/Tox, these effects can be exaggerated, diminished or counter-balanced in combination with differing levels of activity of the drug transporter enzymes, which occur in several organs including the intestine, liver and kidneys and play an important role in the blood-brain and blood-placenta barriers (Roots *et al.*, 2004). For example, in the human intestine where drugs encounter P-gp before CYP3A4, compounds that are substrates for both P-gp and CYP3A4 will exhibit decreased metabolism when a P-gp inhibitor is present (Benet *et al.*, 2004). In the liver where P-gp and CYP3A4 orientation is reversed, metabolism would increase. If both P-gp and CYP3A4 are inhibited, the intestine will show decreased metabolism, while in the liver the effects will counterbalance each other.

1.2.4.4 Variable expression

In some cases, the differing rates of drug metabolism observed are due to inter-individual differences in CYP expression level (Danielson, 2002). In the worst extreme, no protein is expressed at all, or that which is expressed is completely inactive (such alleles are termed 'null' alleles). However, most expression variability is due to the characteristic inducibility of many CYP enzymes by xenobiotics and, in particular, by drugs (see Table 1.3 above). From a biological point of view, this is an adaptative response to increase levels of detoxification of the substrate (although induction can also lead to increasing levels of toxicity if reactive metabolites are formed). This phenomenon of induction can contribute significantly to inter-individual variation in drug response.

This is particularly true of CYP3A4 where hepatic and intestinal levels can differ more than 40-fold between individuals (Lamba *et al.*, 2002). Unlike other human drug-metabolising CYPs, there are no 'null' alleles for CYP3A4. Variants in the coding region generally occur at low population

frequencies (<5%), heterozygous with a wild type allele, and generally do not significantly affect expression or catalytic activity when compared to the wild type. Instead, the high degree of CYP3A4-related variability is thought to be due to variation in the flanking and intronic regions effecting levels of expression in response to induction. In fact, most human drug-metabolising CYPs are either polymorphic or inducible, which accounts for the large contribution of the CYPs to variable drug response (Donato and Castell, 2003).

However, the pharmaceutical industry should not ignore polymorphic variation in CYP3A4 as mutations in this enzyme are known to affect metabolic regioselectivity (as discussed in Section 1.1.6 above). This could result in unexpected and toxic products of metabolism or the reduced metabolism of a prodrug into active compound.

1.2.5 Current methods for assessing CYP metabolism and interactions in drug development

Until quite recently, methods used in drug development seldom took the effects of CYP polymorphic variation into account, despite the possibility of it being a major contributing factor to the high attrition rate of novel drugs in the costly development process. In fact, prior to the early 2000s, studies often had little or no *in vitro* kinetic data on how the drugs were even metabolised, let alone the possible effects of genetic variation (Baldrick, 2008b). In one report in 2008, which studied the various methodologies used by over 30 pharmaceutical companies during the period 1997-2006, only 38% of kinetic studies used cytochrome P450 enzyme activity as a marker for enzyme induction, inhibition, drug interactions, or even isoform mapping (i.e. determining which CYP/s metabolise the drug) (Baldrick, 2008b).

In accordance with more recent regulatory guidance, pre-clinical investigations are now required to determine all of the following: which enzymes are involved in the drug's biotransformation, whether the enzymes involved in the biotransformation are influenced by the concomitant use of other drugs, whether the drug influences the expression or activity of enzymes that metabolise it or other drugs, and what toxicity is associated with the drug and its metabolites (Donato *et al.*, 2008). Several *in vitro* models have been developed to help answer these questions. These include a number of cell- and tissue-based systems, as well as microsomal fractions and purified recombinant enzymes. These models all have various advantages and drawbacks depending on the purpose of their use.

1.2.5.1 Cell and tissue-based systems

Numerous cell and tissue-based systems have been developed in attempts to mimic human metabolism. These include isolated liver slices, primary hepatocytes, and transformed cultured human hepatoma cell lines (Donato *et al.*, 2008;Gillam, 1998;Venkatakrishnan *et al.*, 2003).

Liver slices are difficult to obtain in consistent quantities and qualities and deteriorate rapidly. The widespread use of hepatocytes is also hampered by restricted accessibility to suitable liver samples, as hepatocyte cell cultures have limited viability (2-3 days) and therefore require fresh liver that has to be processed and cultured immediately upon harvesting from the donor. Liver slices and hepatocyte preparations exhibit phenotypic instability and high batch-to-batch functional variability as they are derived from different donors. This can complicate their use in routine testing. However, to take some account of population variation, one form of the hepatocyte assay is to use a mixture of hepatocyte preparations from a variety of donors - although the numbers of individual human livers used in such assays are never large enough to represent inter-individual variation with statistical relevance.

Primary cultures of human hepatocytes are still the best *in vitro* model for studying enzyme induction. Hepatocytes are cells with intact plasma membranes, metabolic pathways, protein expression systems, as well as physiological levels of cofactors and coenzymes. Potential inducers can simply be added to the culture medium and levels of mRNA and/or enzyme activity can then be measured. Human hepatocytes in primary culture respond well to enzyme inducers during the first few days while the cells are still viable. Hepatocytes are also good models to predict liver toxicity, although toxicity in other organs will of course not be predicted.

Hepatoma cell lines exhibit variable metabolic activity upon passaging or have a limited lifespan and must be freshly derived. Fully differentiated cells such as primary hepatocytes have to be used, since hepatoma cell lines grow inefficiently and often have low and partial CYP expression in the absence of specific known inducers. Hepatoma cell lines are also inferior to hepatocytes as a model for potential drug induction as they respond poorly to inducers.

Since there are many different CYP isoforms in human liver tissue (and other liver tissue-based systems), quantifying the relative contribution of each to a metabolic reaction requires form-specific probes; i.e. specific substrates, chemical inhibitors or inhibitory antibodies (Gillam, 1998).

However, because many CYP forms are very closely related most *in vitro* probes are selective but not specific, which complicates attempts at accurate metabolic profiling. Another disadvantage of these systems is that their CYP expression profile differs to that of extrahepatic tissues, which means that extrahepatic metabolism of a compound by other CYPs (e.g. in the target tissue) is unaccounted for, despite it possibly being critical in toxicity, efficacy and overall drug clearance.

1.2.5.2 Liver microsomal systems

Microsomal fractions of human liver are the most widely used source of human enzymes for *in vitro* metabolic studies (Donato *et al.*, 2008;Gillam, 1998). Microsomes are formed from the smooth ER during tissue homogenisation and are fractionated by differential ultracentrifugation. This system contains individual CYPs mixed together in proportions that reflect the proportions found in human liver. Major drawbacks of this model though are limited tissue availability, artefactual variability in CYP protein expression and function caused by the time between tissue harvesting and freezing, as well as variability introduced by the donors themselves. In addition, technical difficulties in purification, problems with the reconstitution of enzymes and the difficulty in separating the various closely related CYP forms makes it problematic to assess the role of any particular individual CYP in such preparations. Another major limitation of microsomes is that unlike hepatocyte preparations, they contain phase I enzymes but only some of the phase II enzymes and drug transport is not accounted for at all. Microsomes can also only be used for short incubation times (usually <1hr), and as a consequence poorly metabolised drugs as well as secondary metabolism are hardly detected. Screening of CYP induction can also not be carried out with microsomes, as they lack a full protein expression system.

1.2.5.3 Recombinant enzyme systems

Genetically engineered cells expressing human DMEs have become potent tools for investigating drug metabolism, drug-drug interactions and bioactivation (Donato *et al.*, 2008). A variety of expression systems such as bacteria, yeast, insect cells and mammalian cells have been used to stably express CYP enzymes. Recombinant enzymes show catalytic properties comparable to those of human liver microsomes whilst activity levels are similar to or higher than that of hepatocytes.

In recent years, the use of recombinant systems expressing a single CYP enzyme has become relatively common in drug discovery and development (Donato *et al.*, 2008;Gillam, 1998). The

obvious advantage is that the activity and metabolites of one specific human CYP can be studied in isolation. Unlike tissue-derived CYPs, recombinant CYPs can be regarded as a renewable resource and are much easier to obtain in sufficient quantities and purity for biophysical studies. Moreover, structure-function studies (such as those examining the effects of polymorphisms), can be carried out by site-directed mutagenesis, while extrahepatic or foetal forms can also be investigated as long as the cDNA sequence is known. Microsomal preparations (called 'supersomes') or purified enzymes can be produced with high batch-to-batch consistency and can be stored for long periods at -80°C without loss of viability.

However, single-CYP assays alone cannot accurately predict the degree of involvement of the given CYP in a particular reaction *in vivo*. Instead, this system offers an inclusion/exclusion screen for the role of each CYP in the metabolism of a drug, metabolites produced can easily be identified and attributed to a particular CYP whilst potential inhibitors (such as other drugs) can also be characterised. This system is therefore an excellent tool for confirming the metabolic pathways suggested by liver microsome or hepatocyte models (Lahoz *et al.*, 2008).

Different strategies have been proposed to mathematically reconstruct the relative contribution of each isoform to the metabolism of a given compound using a combination of data from the various *in vitro* models. Such strategies are likely to be used in the near future, as CYPs are increasingly used in *in vitro* drug metabolism studies. Although a complete understanding of the quantitative relationship between *in vitro* and *in vivo* findings on metabolism, transport and drug-drug interactions is still emerging, it is already clear that *in vitro* studies can frequently serve as a screening mechanism to rule out the importance of a metabolic or transport pathway and thus the potential for drug-drug interactions affecting this pathway, so that subsequent *in vivo* testing is unnecessary (Huang *et al.*, 2007).

Drug metabolism activity is usually detected by quantifying the metabolites using HPLC separation and GC/GC-MS analysis (Lahoz *et al.*, 2008;Urlacher *et al.*, 2004). Another advantage of the single-CYP system is that it is the most easily adapted for high-throughput (HTP) screening using this technology, due to the difficulties involved in the detection of multiple substrates and metabolites produced in HTP applications of other models. However, few laboratories are readily able to engage in such studies, as typical HTP CYP assays rely on the use of large quantities of

CYP preparations together with banks of automated LC-MS machines. The costs involved are therefore currently prohibitive.

1.3 CONCLUSIONS & OBJECTIVES

There is currently a clear need for a high-throughput platform that allows quick, cost-effective and accurate measurement of the metabolism of novel drugs by not only wild-type cytochrome P450s, but also by all the clinically relevant polymorphic forms. Developing such a platform would provide an important addition to the current repertoire of tools used by pharmaceutical companies for pre-clinical drug screening. Not only could it provide HTP information as to which CYP isoforms metabolise a given drug and how genetic variation affects this metabolism, it could also be used to screen potential inhibitors and drug-drug interactions. This will be valuable in improving the efficiency of the development pipeline by facilitating decisions as to which lead compounds to progress and which to discontinue. It would also provide pharmacogenetic data that could inform and guide PGx studies carried out in clinical trials.

Novel miniaturised assay formats could bring high-throughput CYP screening assays within reach of smaller companies and academic laboratories, as well as providing the pharmaceutical industry with a tool that is much more cost-effective, comprehensive, and adaptable than current HTP methods. However, miniaturising protein assays is not without its challenges. This is especially true when attempting to develop a cytochrome P450 protein array platform, as these enzymes are notoriously complex in their expression, interactions and kinetics. These are the aims and challenges addressed in this thesis.

CHAPTER TWO

Construct Cloning, Expression & Purification

2.1 INTRODUCTION

Prior to developing the P450 Biochip platform, standardised methods for generating adequate amounts of soluble and functional CYP and CPR enzymes had to be developed for later array fabrication. For the purposes of this thesis, the CYP3A4 isoform was chosen for all development work due to its notoriety as being the most complex in its requirements for reconstitution of enzyme activity - the hypothesis being that if a functional array was possible using CYP3A4 it would be very likely that the remaining isoforms (which are potentially less exacting in their requirements) should also succeed (Gillam *et al.*, 1993). It is also of great interest in terms of its pharmacogenetics, as comparatively little is understood about the impact of genetic variation on the kinetics of this isoform, especially in terms of its regioselectivity.

The heterologous expression host, cloning strategy, and expression and purification protocols had to be determined and various protein constructs tested for best yield and activity. Unlike the few previous attempts by other groups at cytochrome P450 function arrays, which utilised commercial baculosome CYP and CPR enzyme preparations (discussed in Section 5.1.1), our aim was to first set up an expression system that would be easily adaptable to high-throughput, low-cost protein production at a later stage - keeping in mind the requirements for protein immobilisation onto the array surface. This was a necessary consideration as the many possible future applications of this platform include the metabolic screening of libraries of new chemical entities and the parallel functional comparison of CYP genetic variants on drug metabolism, both of which involve large sample numbers.

The experimental objectives described in this chapter are therefore the following:

1) To evaluate heterologous expression hosts and cloning strategies used in previous CYP studies in the literature.
2) To create the various requisite CYP3A4 and CPR constructs by recombinant cloning.
3) To express and purify the recombinant CYP3A4 and CPR proteins.

2.2 RESULTS AND DISCUSSION

2.2.1 Choice of heterologous expression host and cloning strategy

2.2.1.1 Choice of heterologous expression host

Mammalian CYP was first successfully expressed in a heterologous host in 1991 in *E. coli* cells (Barnes *et al.*, 1991). Since then, several other hosts and expression strategies have been investigated and CYPs with catalytic properties comparable to those of human liver microsomes have been expressed in bacteria, yeast, insect cells and mammalian cells (Masimirembwa *et al.*, 1999;Yun *et al.*, 2006).

In most laboratories where protein expression is regularly carried out, whether for small- or large-scale studies, *E. coli* and baculovirus-mediated insect cell expression are the two main workhorses used (Hunt, 2005). In most cases, *E. coli* is tried first unless there are clear precedents in the literature to the contrary or where post-translational modifications are critical for protein function. If yields and solubility levels are deemed satisfactory for experimental purposes, it generally becomes the preferred expression host due to the significant cost and timesavings it offers. *E. coli* has also become the favoured workhorse for high-throughput protein expression in proteomics research in general due to its robustness, which allows automated expression and screening (Korf *et al.*, 2005). For the purposes of developing the P450 Biochip platform, we therefore opted to utilise *E. coli* as the expression host for both cytochrome P450 and reductase expression. In addition, it has previously been used extensively in the production and characterisation of the CYP enzymes (Yun *et al.*, 2006).

E. coli expression systems offer a range of advantages compared to other host systems, including lower cost, ease of use, high potential protein yields, and short protein production cycles (Yun *et al.*, 2006) (see Table 2.1 below). A large variety of mutant host strains and expression vectors are commercially available and cloning, protein engineering and mutagenesis are all relatively easy. Culture, protein expression and purification are also relatively simple and rapid and are more easily adaptable to large-scale applications. In addition, *E. coli* has no native cytochrome P450 or CPR genes and purification of recombinant CYP proteins is therefore simplified.

Factors	Expression system			
	Mammalian cells	**Baculovirus-infected insect cells**	**Yeast**	**Bacteria (*E. coli*)**
Technical demand for culture	High	Medium	Low	Low
Cost of culture	High	High	Low	Low
Length of culture period to harvesting	Weeks	Days to a week	Days to a week	Hours to days (usually 1-2 days)
Requirement for δ-ALA or hemin supplementation	Not necessary	Usually required	Not necessary	Usually required to increase yields
Protein yield	Low to moderate	High	Moderate	High
Presence of endogenous CYP and/or reductase system	CYPs, CPR and Cyt b_5 are present	No significant endogenous CYP, CPR and Cyt b_5	CYPs and CPR are present	Contains flavodoxin and NADPH-flavodoxin reductase
Requirement for CPR co-expression	Not necessary (levels might be rate limiting though)	Co-expression required	Endogenous CPR may couple poorly, co-expression might be required	Co-expression usually required (poor coupling to host reductase system)
Requirement for N-terminal modifications	Not necessary	Not necessary	Not necessary	Usually necessary
Occurrence of post-translational modifications	Yes	Yes	Yes	No
Ease of mutagenesis	Straightforward	More complicated	Straightforward	Straightforward
Ease of multi-parallel optimisation of expression conditions/ high-throughput adaptation	Difficult	Difficult (Extra optimisations required include time of infection and multiplicity of infection)	Moderately difficult	Relatively easy

Table 2.1: Comparison of the advantages and disadvantages of the various CYP heterologous expression systems. [Adapted from (Gillam, 1998;Hunt, 2005;Yun *et al.*, 2006).]

One major disadvantage of the *E. coli* system however, is that recombinant proteins seldom undergo post-translational modifications such as phosphorylation or glycosylation (Yun *et al.*, 2006). However, there are few known CYP post-translational modifications and, of those characterised, none appear to be critical for enzymatic activity (Aguiar *et al.*, 2005;Neve and Ingelman-Sundberg, 2008). One exception to the absence of post-translational modifications is that of heme incorporation, which still occurs in bacterial cells unless the recombinant proteins are incorrectly folded (Yun *et al.*, 2006).

In addition, disulphide bonds are incapable of forming in the strongly reducing environment of the bacterial cytosol, which can hamper the correct folding of certain proteins (Fujita and Kamataki, 2002). This problem can sometimes be overcome by targeted secretion into the more aerobic environment of the periplasmic space, where disulphide bond formation is likely to occur. However, the CYPs and cytochrome P450 reductase (CPR) are not cysteine-rich proteins, and do not appear to require disulphide bond formation for correct folding.

Having chosen to use an *E. coli* expression host, the various different expression strategies described in the literature for the successful *E. coli* expression of CYP enzymes were then investigated. These include the use of different host strains, different vector elements, co-expression of molecular chaperones, targeting to specific cellular or extracellular compartments, various N-terminal sequence modifications, and the use of fusion proteins (Yun *et al.*, 2006). Based on this information, we devised an expression strategy that is briefly outlined below.

2.2.1.2 Choice of protein targeting strategy

The majority of human CYPs expressed in *E. coli* have been designed to incorporate into the inner membrane *via* their N-terminal membrane anchor (Yun *et al.*, 2006). However, CYPs can also be expressed such that they localise in the cytosol or the periplasm, are targeted for secretion into the culture medium, or are targeted for aggregation into inclusion bodies.

Of the various protein targeting methods described (see Table 2.2 below), the method that has been the most extensively and successfully used for CYP expression is that of inner membrane targeting, with cytosolic expression the second-most (Yun *et al.*, 2006). For the purposes of HTP protein production and array fabrication however, cytosolic targeting appeared the most amenable. Compared to inner membrane targeting, targeting CYP proteins to localise in the cytoplasm has the benefits of simpler purification and simpler plasmid constructs. Cytosolic targeting is also the most routinely used strategy for recombinant protein production and therefore has better-defined problems (e.g. purification/proteolysis vs. translocation/secretion/refolding problems). In addition, one promising future approach for CYP array fabrication is *in vitro*, on-chip translation and concomitant immobilisation (He and Taussig, 2007;He and Wang, 2007). This approach is not explored in this thesis but would be more easily explored at a later stage using protein constructs designed for soluble expression. We therefore opted for cytosolic protein targeting.

Targeting strategy	Advantages	Disadvantages
Inner membrane	Full-length CYP and CPR proteins can be expressed this way (i.e. with N-terminal membrane anchors)	Various CYP isoform-specific sequence modifications of the N-terminus usually required to ensure adequate expression Over-expression toxic to *E. coli* cells Proteins subject to proteolysis Purification more complex (more protein types)
Cytoplasm	Easier to purify and more scalable to HTP than membrane-targeted forms High potential yields	Despite truncation of membrane anchor isoform-specific peripheral membrane interactions still occur, which require disruption by detergent and/or salts Proteins subject to proteolysis Purification more complex
Inclusion bodies	Proteins protected from proteases Purification is simple (fewer protein types) High potential yields	Protein recovery and refolding costly, laborious and not always effective
Periplasm	Proteolysis less extensive Purification is simple	Translocation across inner membrane not always efficient Cleavage of signal peptide not always efficient
Extracellular medium	Little proteolysis Purification is simple Protein folding often improved	Efficient secretion difficult to obtain Proteins are diluted, difficult to obtain adequate concentrations

Table 2.2: Comparison of the advantages and disadvantages of the various CYP protein targeting strategies (Makrides, 1996;Yun *et al.*, 2006).

2.2.1.3 Choice of N-terminal modifications

The main caveat to CYP expression in *E. coli* is that in general, modifications of the N-terminus are required to achieve adequate expression levels because native human CYPs are usually expressed at low levels, if at all, in *E. coli* (Gillam, 1998). This has been attributed partly to the fact that codon preferences differ between mammals and bacteria, and partly to the fact that proteins highly expressed in bacteria contain features at the 5′ ends of their genes (such as relatively high concentrations of AT base pairs) that may be absent in mammalian sequences.

The first successful attempts at CYP expression in *E. coli* therefore focused mainly on altering CYP N-terminal nucleotide sequences in order to facilitate expression. The first catalytically active mammalian CYP to be expressed, bovine CYP17A1 by Waterman's group in 1991, contained several such N-terminal sequence modifications (Barnes *et al.*, 1991). Specifically, the second codon was changed from TGG (tryptophan) to GCT (alanine), a preferred second codon for expression of the *E. coli lacZ* gene; codons 4 and 5 were changed to TTA (silent mutations), since

this region of *E. coli* mRNAs has been shown to be rich in adenosine and uridine nucleotides; and the last nucleotides of codons 6 and 7 were changed to adenosine and thymidine (silent mutations) respectively to reduce rare codon usage and prevent mRNA secondary structure formation, both of which inhibit translation. These modifications led to abundant CYP17A1 expression in *E. coli* compared to the unsuccessful attempts at producing this enzyme from the native sequence.

After the pioneering work by Barnes *et al.*, several other groups used this modified CYP17A1 N-terminal sequence to augment the expression of other mammalian CYPs, *via* sequence alignment and site-directed mutagenesis (Yun *et al.*, 2006). However, this N-terminal sequence has not been as successful for all of the isoforms tested and only a few of these chimeric constructs yielded substantial expression levels. It therefore appears that various N-terminal sequence modifications need to be empirically tested for each CYP isoform in order to obtain satisfactory to optimal yields.

In addition to N-terminal sequence modifications, several studies have also attempted to increase the expression and solubility of the CYPs by truncating the N-terminal membrane anchor domain, which otherwise targets these proteins for incorporation into the *E. coli* inner membrane with the active site facing the cytosol (Pernecky and Coon, 1996;Yun *et al.*, 2006). For certain isoforms this was shown to increase the yield of soluble enzyme, whereas other isoforms still associated strongly with the membrane. However, this interaction was shown to be a peripheral membrane association mediated by areas on the catalytic domain distinct from that of the N-terminal anchor. In addition, the strength and extent of these interactions were found to be isoform-specific. These peripheral membrane interactions can usually be disrupted during purification by high ionic strength buffers (e.g. 1 M NaCl, 0.5 M potassium phosphate) or the use of detergents (Guengerich *et al.*, 1997;Yun *et al.*, 2006).

The above N-terminal modifications were predicated on the hypothesis that the N-terminal hydrophobic domain (usually < 30 a.a.) of mammalian CYPs functions mainly as a membrane anchor and does not play a direct role in catalysis, substrate binding, or coupling with CPR or cytochrome b_5 (Yun *et al.*, 2006). This hypothesis has been borne out by the observation that these modifications do not appear to significantly alter CYP enzymatic properties relative to native constructs expressed in non-*E. coli* systems (see e.g. (Blake *et al.*, 1996;Sandhu *et al.*, 1993;Sandhu *et*

al., 1994;Yamazaki *et al.*, 1999)). However, it remains difficult to directly and accurately compare the catalytic activities and other biophysical properties of recombinant CYP enzymes to those purified from human tissue, due mainly to the difficulty in purifying the corresponding enzymes from tissue and ensuring that only one CYP isoform is present.

Based on our decision to use cytosolic protein targeting, N-terminal membrane anchor deletions rather than sequence modifications were chosen to facilitate *E. coli* expression. This also facilitated our requirement for a simple cloning strategy that could be quickly and easily extended in a HTP format to other CYP isoforms and their various genetic variants at a later stage.

2.2.1.4 Choice of fusion tags for purification and increased solubility

A variety of fusion tags have been described for the heterologous expression of proteins in *E. coli*. In addition to offering a convenient one-step generic purification strategy, several tags have also been shown to enhance the solubility of the protein partner (Korf *et al.*, 2005). Granted that truncation of the CYP membrane anchor in itself does not automatically confer adequate levels of soluble and functional expression for all CYP isoforms, N-terminal truncation in addition to fusion to a soluble protein tag was therefore chosen as a promising strategy. However, there is no single tag available that is optimal for the expression of proteins in general, nor is there an ideal strategy to express and purify a large number of different proteins. Four of the most popular tags utilised in *E. coli* are NusA (*E. coli* N-utilisation substance protein A), His (polyhistidine, often His_6), MBP (maltose binding protein), and GST (glutathione *S*-transferase). These tags are compared in Table 2.3 below.

GST, and especially MBP and NusA tags are all known to increase protein solubility. However, the size of these three tags can be prohibitive due to the possible occlusion of protein interactions or changes to the native conformation of the fusion partner (Hunt, 2005). Although the removal of large fusion proteins can be facilitated by the incorporation of protease cleavage recognition sites, there are a number of problems associated with their use such as incomplete or non-specific cleavage. Since we did not wish to include a large protein tag due to the possibility of it later interfering with CYP-CPR interactions on the array surface, none of these three tags were deemed suitable for our purposes.

Fusion tag	Size (kDa)	Use	Matrix/elution	Comments
His_6	0.84	Purification	Divalent metal resin/ imidazole	Smallest of the tags Tag or elution may affect protein properties More modest purification compared to other tags
MBP	40	Purification Enhanced solubility	Amylose resin/ Maltose	Matrix not compatible with reducing agents Large tag which may affect protein properties
GST	26	Purification Enhanced solubility	Glutathione resin/ Glutathione	GST dimerisation and glutathione elution may effect protein properties Slow binding kinetics to affinity resin Large tag which may affect protein properties
NusA	54.9	Increased expression Enhanced solubility	N/A	Must be used in conjunction with an affinity tag Large tag which may affect protein properties

Table 2.3: Comparison of the various fusion tags commonly used for heterologous expression. [Adapted from (Hunt, 2005).]

For purification purposes we therefore chose to utilise the His-tag. Unlike GST and MBP, His-tags seldom effect the folding of the fusion protein due to their small size (Hunt, 2005). The His-tag (usually 4-6 residues) is a popular CYP fusion tag due to the ease of purification it confers and its good reversible binding attributes, both of which allow for simple, non-denaturing purification with Ni^{2+}-affinity chromatography.

In addition to the His-tag, we chose to incorporate a tag derived from the biotin carboxyl carrier protein (BCCP) domain of the *E. coli* enzyme acetyl-coenzyme A carboxylase. The BCCP domain (called accB) of this heteromultimeric enzyme complex is 83 a.a. in size (Roberts *et al.*, 1999). BCCP folds autonomously into an all β–strand structure and is biotinylated post-translationally *in vivo* by biotin ligase on a single, specific lysine residue, Lys122. The BCCP-tag was specifically chosen because the high affinity biotin-streptavidin interaction (K_d = 10^{-15} M) could be exploited downstream to obtain specific and oriented immobilisation on the protein array surface (further discussed in Section 2.2.2.2 and Chapter Five), whereas the His-tag was not suitable for this purpose - the binding between Ni^{2+} and the His-tag being easily disrupted even under simple wash conditions (Lue *et al.*, 2007). The 17 kDa AccB protein itself is unlikely to compete for binding to the array surface, as it forms part of a heteromultimeric complex within which the biotinylated lysine residues are buried (Cronan and Waldrop, 2002).

The BCCP-tag was also chosen for its potential to facilitate the folding and soluble expression of its fusion partners (Samaddar *et al.*, 2005). There is still little knowledge as to the molecular mechanisms of the effects of the various fusion tags on the solubility of their fusion partners and rational design towards this end is only possible to a limited extent (Korf *et al.*, 2005). The most probable reason for the improved folding conferred by certain fusion tags such as MBP and BCCP is that the tag is a protein that is itself normally expressed by the host and therefore efficiently and rapidly reaches a native conformation as it emerges from the ribosome, thereby promoting the acquisition of the correct structure in downstream fusion partners (Hunt, 2005). It is also likely that fusion tags promote protein folding by recruiting *E. coli* chaperones to the nascent polypeptide during translation. These effects are of course more likely to apply to N-terminal BCCP-tagged proteins.

Besides its potential for facilitating protein folding and solubility, the BCCP-tag has also serendipitously proven useful as a protein folding marker for both N- and C-terminal BCCP-tagged proteins (Samaddar *et al.*, 2005). There are several examples in the literature where fusion tags, most notably GFP, have been used as protein folding markers with a very high correlation seen for example between the correct folding and consequent activity of the fusion partner and the fluorescence of the GFP-tag [see e.g. (Pédelacq *et al.*, 2002;Waldo, 2003)]. Similarly, the BCCP domain has to be fully folded in order to become biotinylated *in vivo*, as when the domain is linearised the lysine is no longer recognised by biotin ligase and remains unbiotinylated (Chapman-Smith and Cronan Jr, 1999). For those constructs with the BCCP-tag at the C-terminus, it is highly unlikely that where the fusion partner folds aberrantly that the downstream BCCP domain would fold correctly and become biotinylated. If somehow the correct folding of the BCCP domain still occurred, the other mechanism that could prevent its biotinylation is that the misfolded protein partners would be likely to form aggregates, thereby physically occluding the BCCP domain from the biotin ligase. The latter effect would of course be relevant for N-terminal tagged constructs as well.

The conclusion therefore is that if the fusion partner misfolds, the biotinylation of the BCCP-tag is unlikely to occur. Consequently, it is likely that only correctly folded proteins will later be immobilised onto the array surface *via* the streptavidin-biotin interaction, thereby facilitating the enrichment of functional CYP and CPR proteins on the array surface.

2.2.1.5 Use of molecular chaperones

The co-overproduction of molecular chaperones can greatly increase the soluble yields of aggregation-prone heterologously expressed proteins as they are known to facilitate correct protein folding (Makrides, 1996;Waldo, 2003). Bacterial and mammalian chaperone proteins have been found to increase the level of soluble expression of several human CYP enzymes by up to >3-fold (Yun *et al.*, 2006). The best characterised molecular chaperones in *E. coli* cytoplasm are the ATP-dependent DnaK-KnaJ-GrpE and GroEL-GroES systems (Makrides, 1996), while examples of chaperones found to be successful in CYP studies specifically are GroEL (*E. coli*) and Hsp40 and Hsp70 (human) (Yun *et al.*, 2006). Different chaperones show different levels of efficacy depending on the CYP isoform under investigation, so in order to obtain optimal expression a variety of chaperones would need to be tested for each isoform. For this reason, this approach was not attempted here. However, this is a promising tool to be kept in mind for further work, especially for those CYP isoforms that do not prove amenable to *E. coli* expression using the expression strategy chosen and tested in this thesis.

2.2.1.6 Choice of CYP and CPR reconstitution strategy

Bacterial cells contain endogenous electron transport systems capable of transferring electrons to human CYP, however, the level of activity thus provided is insufficient to support the full catalytic activities of the CYPs (Yun *et al.*, 2006). Therefore, it is necessary to add human CPR for the full reconstitution of the CYP monooxygenase system. Several strategies have been explored in the literature for optimal CYP-CPR reconstitution: simply adding purified CPR from tissue or recombinant sources to purified CYP; co-expression using a bicistronic vector (one promoter upstream of two ribosome binding sites each adjacent to a recombinant coding sequence) (Dong and Porter, 1996;Parikh *et al.*, 1997;Shimada *et al.*, 1998); co-expression using a dual promoter vector (one promoter for each gene on the same vector) (Blake *et al.*, 1996;Iwata *et al.*, 1998); co-expression using different plasmids for each protein in the same cell (Pritchard *et al.*, 1998); and co-expression of the two enzymes as a fusion protein with a joining linker peptide (Fisher *et al.*, 1992;Shet *et al.*, 1993).

For the purposes of this study, since we were testing various constructs for both CYP3A4 and CPR individually (described in the following experimental section), and since we were interested in testing various reconstitution ratios of CYP3A4 to CPR on the array platform, we focused on

using the simplest strategy of utilising separate culture systems for the CYP and the reductase and combining the two enzymes after purification.

2.2.2 Recombinant cloning of the CYP3A4 and CPR constructs

2.2.2.1 Choice of plasmid vector and host strain

The DH5α (Invitrogen) and JM109 (Invitrogen, Strategene) *E. coli* strains are the most popular strains in the expression of human CYPs, and are known for their increased insert stability and good plasmid yield and quality (Yun *et al.*, 2006). We therefore used DH5α for all expression work.

The pMD004 and pBJW102.2 plasmids were a gift from Procognia Ltd, Maidenhead, UK. These vectors were originally derived from the *E. coli* expression vector, pQE-80L (4.8 kb; Qiagen, USA). The original pQE-80L vector contains amongst other elements, a T5 promoter/lac operator element, a His_6-tag coding sequence and a β-lactamase coding sequence that confers ampicillin resistance. The BCCP domain (a.a. 74 - 156) of the *E. coli* AccB enzyme and a glycine-serine linker sequence (GGSGSG) were later cloned into the pQE-80L vector, in frame and either 3′ or 5′ to the His_6-tag, thereby creating the pMD004 vector for N-terminal tagged constructs and the pBJW102.2 vector for C-terminal tagged constructs respectively (see Fig. 2.1 below; personal communication Jonathan Blackburn, (Boutell *et al.*, 2004)).

Wild-type human CYP3A4 and CPR sequences were amplified by PCR from cDNA libraries and then cloned into both of these two vectors (for the pBJW102.2 vector the stop codons were removed from CYP3A4 and CPR protein sequences). pMD004 plasmids containing Δ34N-CYP3A4 and FLN-CPR and pBJW102.2 plasmids containing Δ24C-CYP3A4 and FLC-CPR were obtained from Procognia and used as the parent vectors for all further construct cloning work carried out in our laboratory. (Note: Δ34 represents for example a deletion construct that has had the first 34 N-terminal amino acid residues deleted, ‘FL’ signifies full-length protein while ‘N’ and ‘C’ signify N- and C-terminal tag placement respectively.)

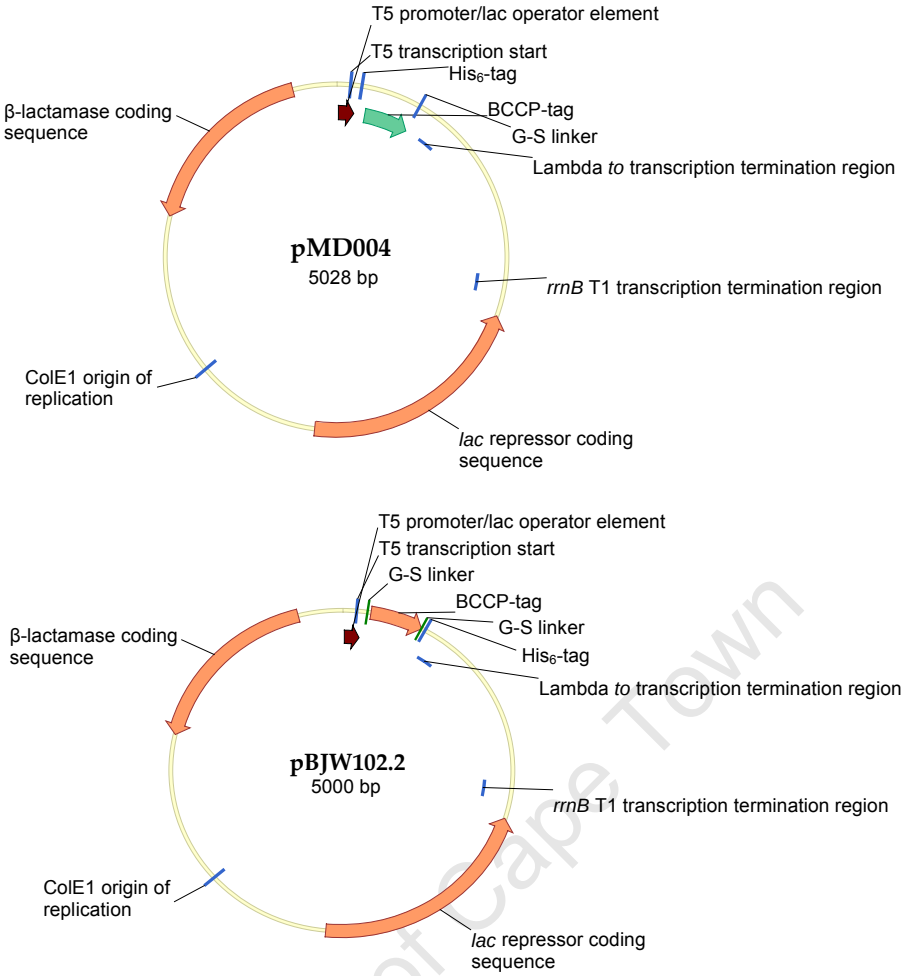


Figure 2.1: Plasmid maps of the pMD004 (*top*) and pBJW102.2 (*bottom*) plasmid vector backbones used for cloning and expressing all CYP3A4 and CPR constructs. pMD004 was used for creating all N-terminal BCCP-tagged constructs while pBJW102.2 was used for creating all C-terminal BCCP-tagged constructs.

The T5 promoter/lac operator element allows graded, high-level protein production that is tightly regulated to minimise metabolic burdens and toxic effects. Recombinant protein expression is often carried out with stronger promoters such as T7, lambda P1 and *atra*B (Hunt, 2005). However, the resulting high levels of mRNA can cause ribosome destruction and cell death. An additional limitation of strong promoter systems is that the target protein is often unable to reach a native conformation and either partially or completely segregates within inclusion bodies. For these reasons, the weaker *lac* promoter or hybrids thereof (as present in pQE-80L) are often preferred for CYP expression (Barnes, 1996). The *lac* promoter is derived from the *E. coli* lactose utilisation (*lac*) operon, and its *in vivo* basal activity can be effectively controlled by the *lac* repressor protein, *lac* I^q (Makrides, 1996). The *lac* promoter is also highly valuable for facilitating the graded expression of proteins, as a range of induction levels are possible using isopropyl-β-D-thiogalactopyranoside (IPTG). This helps to increase the levels of soluble protein by allowing

more time for protein folding, as well as improving yields of recombinant proteins that are toxic or disruptive to *E. coli* cells (such as membrane proteins).

2.2.2.2 *Choice of His$_6$-BCCP-tag placement*

Besides being useful as a folding marker and for high-affinity immobilisation, the BCCP-tag is also very valuable in that it allows the favourable orientation of both N- and C-terminal tagged proteins when immobilised on the array surface: the BCCP-tag binds to the streptavidin-coated surface *via* the biotinylated lysine residue, which is ~44 Å from both the N- and C-termini of BCCP, and these termini are themselves located on the protein surface within close proximity [see Fig. 2.2; (Athappilly and Hendrickson, 1995)]. This allows for the fusion of CYP or CPR proteins to either the N- or C-termini of the BCCP-tag while still ensuring that the proteins are presented away from the surface, thereby minimising the potential perturbation of protein folding and function that might otherwise arise from hydrophobic surface interactions. The tag also facilitates enzyme activity by conferring a limited yet critical degree of mobility to the fusion partner. The latter characteristic is particularly essential in the development of our array format, and will be discussed further in Chapter Five.

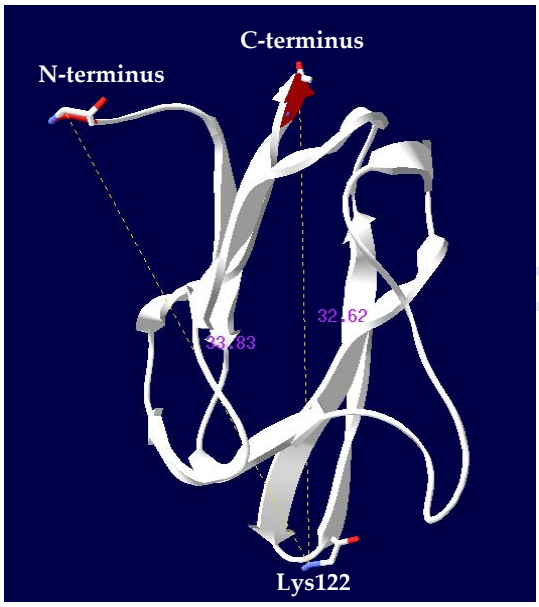


Figure 2.2: Structure of the BCCP-domain as viewed using Swiss PDBViewer (1bdo.pdb). Distances between Lys122 and the two termini are shown (measurements in Å). The Lys122-attached biotin molecule, when extended, adds an additional ~10 Å to the tag (not depicted here).

While His-tags can in principle be placed on either terminus of the protein, the optimal location depends on the folding and biochemical characteristics of the specific fusion protein partner involved (Hunt, 2005). The majority of studies expressing CYPs with this tag used labelling at the C-terminus, due to concerns about the interference the charged tag might pose to the membrane anchor domain if placed on the N-terminus. However, there have also been studies showing the successful replacement of the truncated N-terminal membrane anchor with the His-tag [e.g. (Hanna *et al.*, 2000;Kempf *et al.*, 1995)]. In some cases, the addition of the positively charged

residues of the His-tag appeared to increase the level of cytosolic expression relative to that of untagged truncated CYPs.

We therefore chose to clone both the His_6-tag and the BCCP-tag in tandem on the N- and C-terminals of each CYP and CPR construct to determine their optimal configuration in terms of expression and activity levels.

2.2.2.3 Choice of N-terminal deletions

CYP3A4:

As discussed in Section 1.1.4, microsomal CYPs have a characteristic N-terminal hydrophobic domain of 20-25 residues, which acts as a signal-anchor sequence for insertion into the ER membrane (Neve and Ingelman-Sundberg, 2008). The signal-anchor sequence is usually followed by a short stretch rich in basic residues, which are thought to provide a halt-transfer signal between the hydrophobic membrane anchor and the globular part of the protein. In order to create an N-terminal deletion construct of the full-length protein to remove the membrane anchor, several online software tools were used to map the likely membrane-spanning domain of CYP3A4 (see Materials and Methods). Based on the consensus sequence from these tools, a Δ24-CYP3A4 deletion was chosen (see Fig. 2.3 below). This deletion has also been used by other groups for CYP3A4 crystallisation studies (Williams *et al.*, 2004;Yano *et al.*, 2004).

MALIPDLAME TWLLLAVSLV LLYLYGTHSH GLFKKLGIPG PTPLPFLGN-----

Figure 2.3: N-terminal human CYP3A4 sequence depicting the hydrophobic membrane anchor (*yellow*) as calculated using various software tools. The Δ24-CYP3A4 N-terminal deletion (numbered from the first a.a. residue) is as illustrated (*boxed*).

Cytochrome P450 Reductase:

The cytochrome P450 reductase protein consists of two functional domains: a hydrophobic N-terminal membrane-binding domain (~6 kDa) and a hydrophilic C-terminal catalytic domain (~72 kDa). The catalytic domain can be further subdivided into four structural domains (from N- to C-terminus): the FMN-binding domain, the connecting domain, and the FAD- and NADPH-binding domains (Wang *et al.*, 1997). The linear sequence for the latter three domains is intertwined, and these domains are therefore not functionally separable. The membrane anchor region is preceded by a hydrophilic sequence and an acidic residue (Asp), and is followed by a highly charged

sequence that is positioned at the ER-cytosol interface and likely acts as a stop-transfer signal (Black and Coon, 1982). Based on software predictions from sequence analysis, we chose to create a Δ43-CPR N-terminal deletion construct of the full-length protein to remove the membrane anchor (see Fig. 2.4 below).

MGDSHVDTSSTVSEAVAEEVSLFSMTDMILFSLIVGLLTYWFLFRKKKEEV-----

Figure 2.4: N-terminal human CPR sequence depicting the hydrophobic membrane anchor (*yellow*) as calculated using various software tools. The Δ43-CPR N-terminal deletion (numbered from the first a.a. residue) is as illustrated (*boxed*).

The reductase was first isolated in the literature as a protease-treated form, which is a soluble monomeric protein of MW = 71,000 Da. This truncated form was found to be highly active in the reduction of cytochrome c as an artificial electron acceptor, but no activity was observed towards the physiological acceptors, the CYPs (Black and Coon, 1982). This result has been taken to mean that the hydrophobic domain of the reductase is essential for its biological function in transferring electrons to the CYPs. To the best of our knowledge, no activity towards the CYPs has yet been demonstrated using the truncated form of mammalian CPRs (Backes and Kelley, 2003). However, we were still interested in testing such a construct as we deemed it possible that the CPR membrane anchor is not actually essential for electron transfer in itself but rather provides a physical means of bringing the two enzymes within close enough proximity *via* the ER membrane for electron transfer to actually occur.

2.2.2.4 Cloning of the various CYP3A4 and CPR constructs

The pMD004 plasmids containing Δ34N-CYP3A4 and FLN-CPR and the pBJW102.2 plasmids containing Δ24C-CYP3A4 and FLC-CPR were used as the parent vectors for constructing FLN-CYP3A4, FLC-CYP3A4, Δ24N-CYP3A4, Δ43N-CPR, and Δ43C-CPR *via* inverse PCR (see Fig. 2.5 below for a flow diagram of the cloning process). The FLN-CYP3A4 and Δ24N-CYP3A4 constructs were created by the author, while the FLC-CYP3A4, Δ43N-CPR, and Δ43C-CPR constructs were created by a post-doctoral fellow in our laboratory, Dr Siddharth Sharma. In addition, the Δ24C-CYP3A4 plasmid was used to generate the pBJW102.2 backbone vector for further sub-cloning, while the pMD004 backbone was generated from a pMD004 plasmid containing the p53 gene (also obtained from Procognia). Membrane-targeted (i.e. full-length)

CYP3A4 and CPR constructs were generated in addition to the cytosol-targeted constructs in the interests of comparing the effect of the N-termini of these enzymes on expression and activity.

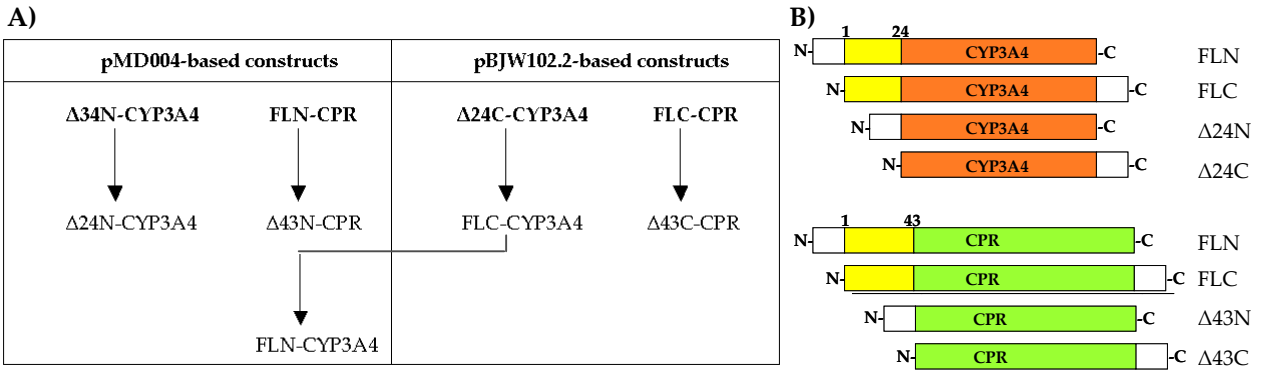


Figure 2.5: A) Flow diagram depicting the order in which the various CYP3A4 and CPR constructs were generated from the parent plasmids (shown in *bold*) obtained from Procognia. B) Diagram depicting the elements of the various CYP3A4 and CPR constructs. The His_6-BCCP-tags (*white*), membrane anchor domains (*yellow*) and catalytic domains of CYP3A4 (*orange*) and CPR (*green*) are depicted.

Depending on the construct being generated, either inverse (i.e. 'back-to-back') PCR primers or normal (i.e. 'face-to-face') primers were created. In general, inverse PCR primers were used when sequence additions or deletions had to be performed, while normal PCR primers were used when sub-cloning a sequence into a different vector (i.e. the creation of the FLN-CYP3A4 construct from FLC-CYP3A4). Primers were designed such that they contained a unique restriction site at either of their 5′ ends so that the amplified product could later be digested and cloned unidirectionally into the multiple cloning site (MCS) of the relevant plasmid vector.

The relevant parent plasmids were extracted from O/N cultures and used as PCR templates. The gel-extracted PCR products and the relevant backbone vector (i.e. either pMD004 or pBJW102.2) were digested with restriction enzymes and then ligated together using various insert:vector ratios. The ligation mixes were then used to transform competent DH5α cells. Potential recombinant colonies were screened using colony PCR (and in some cases restriction enzyme mapping) and those colonies that produced positive bands were sent for sequencing. The final nucleotide and amino acid sequences for each construct as determined by sequencing are listed in Appendix I. A flow diagram of the basic process is shown in Fig. 2.6 below.

The pMD004 and pBJW102.2 backbone vectors were also re-created using inverse PCR. In this case however the PCR products were first treated with T4 polynucleotide kinase to phosphorylate

the 5′ ends prior to self-ligation. In addition, the PCR reactions were subjected to DpnI digestion to remove the parent plasmid. These backbone vector sequences are also listed in Appendix I.

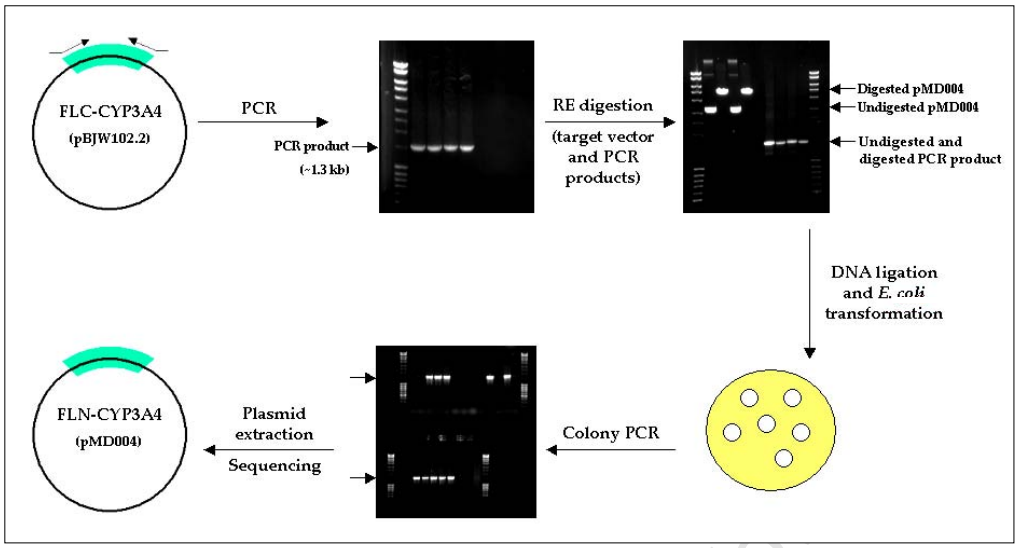


Figure 2.6: Flow diagram of the basic cloning procedure used to create the various CYP3A4 and CPR constructs. Here the generation of the FLN-CYP3A4 construct is used to illustrate the process.

2.2.3 CYP3A4 and CPR construct expression and purification

2.2.3.1 Determining the relative expression levels of the recombinant CYP3A4 and CPR constructs

CYP3A4:

Of the CYP3A4 constructs, only Δ24C-CYP3A4 produced purified protein detectable by Coomassie Blue and Ponceau-S staining at the expected molecular weight (Fig. 2.7A, C and D). Western blots performed on the same samples using anti-His antibody showed that all constructs except Δ24N-CYP3A4 were expressed in the soluble fraction, with Δ24C-CYP3A4 demonstrating the highest level of expression overall (Fig. 2.7 B and C). Based on these gels and later experiments (discussed in Section 2.2.3.2), the relative expression levels for the CYP3A4 constructs was determined to be Δ24N-CYP3A4 <<< FLN-CYP3A4 < FLC-CYP3A4 << Δ24C-CYP3A4.

No protease inhibitors were used during protein extraction and the relative stability of each construct in *E. coli* could therefore be determined. All of the CYP3A4 constructs exhibited some degree of proteolysis, with a greater degree generally visible for the purified proteins relative to the crude lysate extracts (Fig. 2.7 B, D). One proteolytic fragment that was particularly prominent, migrated at ~40 kDa and was visible on both the Coomassie Blue-stained gels and the Western

blots for all constructs, except for the Western blots of Δ24N-CYP3A4 (Fig. 2.7 A-E). It is possible that the 40 kDa band visible by Coomassie Blue-staining resulted mainly from an *E. coli* protein that co-migrated with the His-tagged proteolytic fragment detected on the Western blots – which would explain why a band at this molecular weight was visible for all constructs exhibiting soluble expression but not for Δ24N-CYP3A4 when using Anti-His antibody. Smaller proteolytic fragments were also observed at the dye front for all expressed proteins.

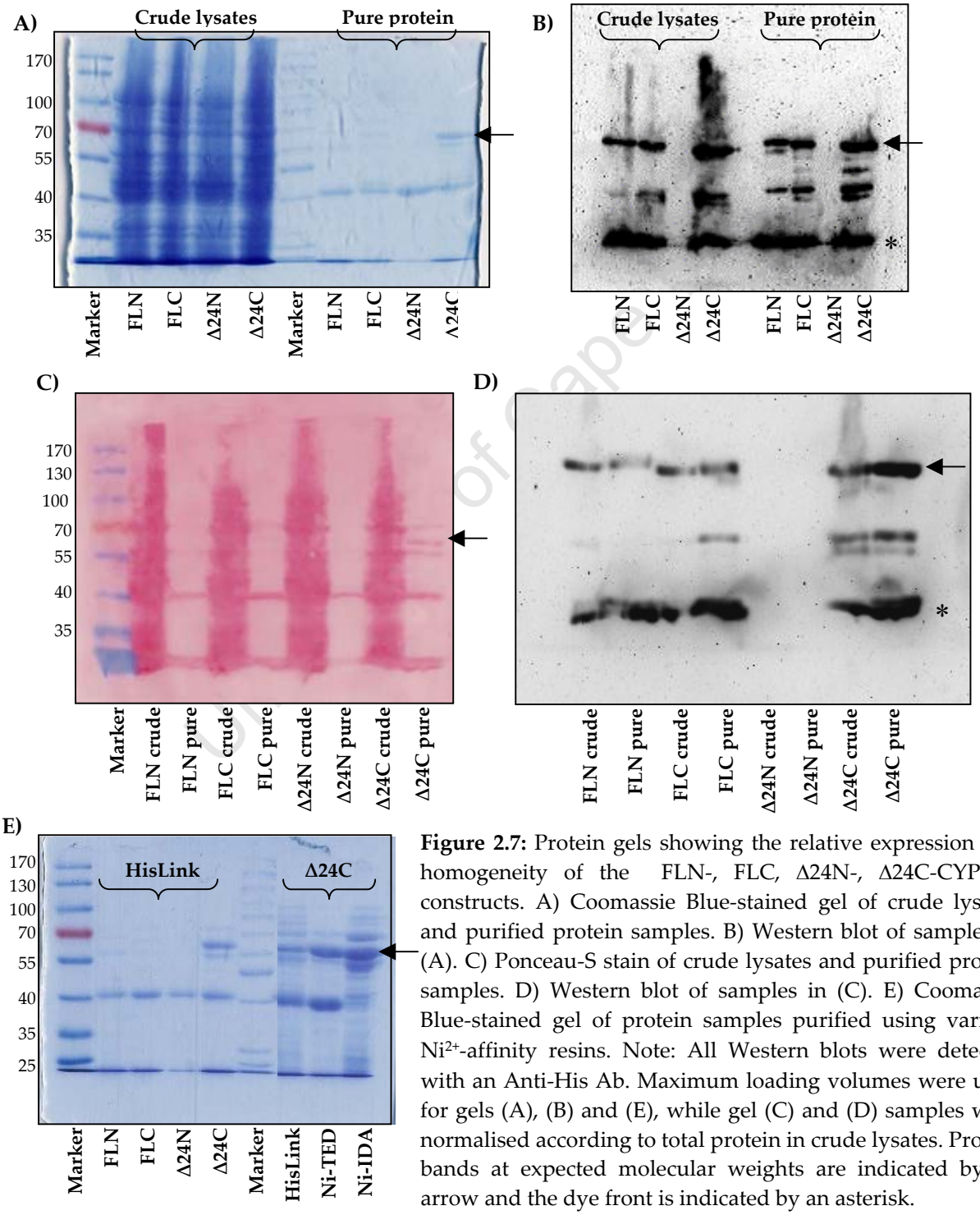


Figure 2.7: Protein gels showing the relative expression and homogeneity of the FLN-, FLC, Δ24N-, Δ24C-CYP3A4 constructs. A) Coomassie Blue-stained gel of crude lysates and purified protein samples. B) Western blot of samples in (A). C) Ponceau-S stain of crude lysates and purified protein samples. D) Western blot of samples in (C). E) Coomassie Blue-stained gel of protein samples purified using various Ni^{2+}-affinity resins. Note: All Western blots were detected with an Anti-His Ab. Maximum loading volumes were used for gels (A), (B) and (E), while gel (C) and (D) samples were normalised according to total protein in crude lysates. Protein bands at expected molecular weights are indicated by an arrow and the dye front is indicated by an asterisk.

Various affinity resins for His-tag protein purification were tested over the course of this thesis, including TALON, HisLink, Ni-TED and Ni-IDA (see Materials and Methods). The first is a cobalt-based resin while the rest are nickel-based. The nickel-based columns provided higher yields than the TALON resin (data not shown), with Ni-TED providing the best balance between yield and homogeneity (see Fig. 2.7 E).

Cytochrome P450 Reductase:

Of the CPR constructs, FLC-CPR, Δ43C-CPR and Δ43N-CPR were all well expressed in the soluble fraction, while FLN-CPR expression was not detected (Fig. 2.8 B and C). The Δ43-CPR constructs expressed the best with recombinant protein bands easily visible on the Ponceau-S stains. Based on these gels and later experiments the relative expression levels for the CPR constructs was determined to be FLN-CPR <<< FLC-CPR << Δ43C-CPR < Δ43N-CPR.

Much less proteolysis was visible for the CPR constructs compared to those of CYP3A4 (Fig. 2.8 B and C). In fact, it is possible that the main secondary band observed - which closely co-migrated with the expected purified protein bands - may be unbiotinylated CPR protein (biotin MW = 8.651 kDA; see Fig. 2.8C). This secondary band was most visible for the FLC- and Δ43C-CPR constructs on the Coomassie Blue-stained gels (Fig. 2.8B) but may have been obscured by high signal in the Western blot (and similarly for Δ43N-CPR on the Coomassie Blue gel).

The CPR constructs generally express better than the CYP3A4 constructs and it is possible that endogenous biotinylation was therefore limiting for the former. It is known that the overexpression of recombinant proteins in bacterial cells can result in an appreciable portion remaining un-biotinylated due to competition with the essential *E. coli* lipid synthetic protein AccB for the endogenous biotin ligase (Lue *et al.*, 2007;Sakanyan, 2005). (The BCCP-tag can also be biotinylated *in vitro* using the same enzyme but substances commonly present in enzyme preparations reduce biotinylation efficiency. We therefore opted for *in vivo* biotinylation facilitated by the addition of biotin to the culture upon induction.) Biotinylation of CYP3A4 and CPR constructs of interest was later explored using streptavidin-HRP (see Section 3.2.4).

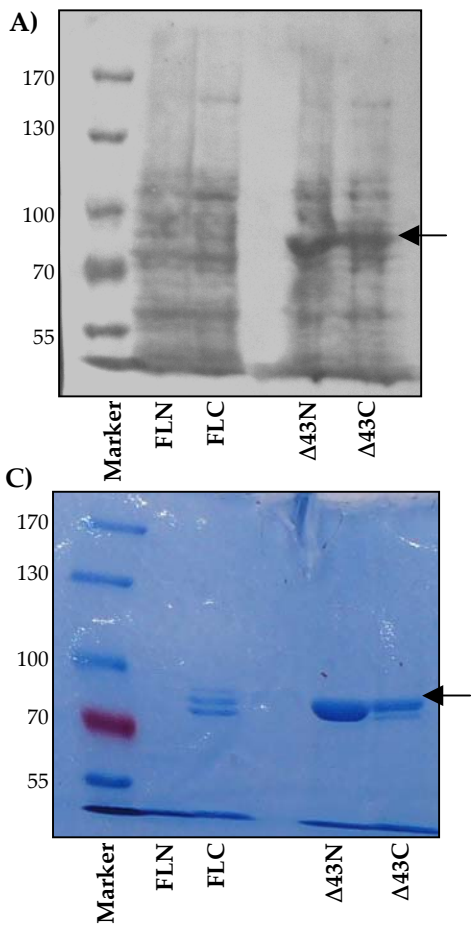


Figure 2.8: Various protein gels showing the relative expression and homogeneity of the FLN-, FLC, Δ43N-, Δ43C-CPR constructs. A) Ponceau-S stain of CPR crude lysates normalised according to total protein. B) Western blot of samples in (A) using an Anti-His Ab. C) Coomassie Blue-stained gel of CPR protein samples purified using HisLink resin (maximum loading volumes used).

The variable expression levels of the different CYP3A4 and CPR constructs was highly reproducible and several observations drawn from the literature may be helpful in rationalising these differences. The factors known to influence the successful expression of mammalian proteins in *E. coli* are diverse and complex and include subtle structural features of the nucleotide sequence, the stability and translational efficiency of the mRNA, the ease of protein folding, degradation of the protein by host cell proteases, major differences in codon usage between the foreign gene and native *E. coli*, and the potential toxicity of the protein to the host (Makrides, 1996). There are however some empirical 'rules' that can be followed to limit this unpredictability, many of which are based on observations as to optimal expression elements utilised by the most highly expressed *E. coli* proteins.

One of these factors is the potential of the nascent mRNA for forming secondary structures in the translation initiation region, thereby blocking the binding of the ribosomal complex and subsequent translation (Gillam, 1998;Makrides, 1996). Structures with a free energy (ΔG_f^0) of > -6 kcal/mol generally do not reduce translational efficiency, while structures of < -6 kcal/mol are

more stable and can cause significant decreases in expression (de Smit and van Duin, 1994). Both 5′ and 3′ regions of *E. coli* ribosome binding sequences (RBS) exhibit a bias towards high adenine content for this reason. Using the Genebee server (www.genebee.msu.su) and a cut-off value of -6 kcal/mol, the potential for secondary structure formation in the region of the RBS [-26 to + 27 with reference to the start codon (Gillam *et al.*, 1993)] was determined for each construct (see Table 2.4).

None of the N-terminal tagged constructs exhibited any potential for mRNA secondary structure. Of the C-terminal tagged constructs, no structures were predicted for FLC-CYP3A4, while one to two structures were predicted for the Δ24C-CYP3A4, FLC-CPR and Δ43C-CPR constructs (between -7.9 and -6.3 kcal/mol). These results suggested that mRNA secondary structure was not an important contributing factor to the variable expression levels observed as no correlation was found with the expression data discussed above.

Constructs	Potential for mRNA secondary structure in RBS[a]	Second codon	Nucleotide usage within first 8 codons	Minor codon usage within first 60 bp	3′ UTR[b]
FLN-CYP3A4	None predicted	Arginine	33% A, 50% A+T	AGA, GGA, AGT	37 bp
Δ24N-CYP3A4	None predicted	Arginine	33% A, 50% A+T	AGA, GGA, AGT	37 bp
FLN-CPR	None predicted	Arginine	33% A, 50% A+T	AGA, GGA, AGT	37 bp
Δ43N-CPR	None predicted	Arginine	33% A, 50% A+T	AGA, GGA, AGT	37 bp
FLC-CYP3A4	None predicted	Alanine	17% A, 38% A+T	None	8 bp
Δ24C-CYP3A4	- 7.5; -6.9	Alanine	25% A, 46% A+T	TCA, 3 x GGA	8 bp
FLC-CPR	- 7.9; - 7.8	Alanine	21% A, 33% A+T	GGA	8 bp
Δ43C-CPR	- 6.3	Alanine	38% A, 54% A+T	AGA, CCC	8 bp

Table 2.4: Comparison of the various CYP3A4 and CPR constructs in terms of various sequence elements known to play a role in efficient protein expression in *E. coli*. Notes: a) Values are in kcal/mol. Only structures of < -6 kcal/mol are listed. b) Number of nucleotides between the TAG stop codon and the start of the λ t_0 transcription termination region.

Another factor known to influence the recombinant expression of mammalian proteins in *E. coli* is the prevalence of the so-called 'minor' codons in the nucleotide sequence. There is a bias in the usage of synonymous codons for almost all degenerate codon families in *E. coli* (Makrides, 1996). The frequency of use for a particular codon usually reflects the abundance of the cognate tRNA, which means that heterologous sequences enriched with codons rarely used by *E. coli* will not be expressed as efficiently. Where the minor codons do appear in *E. coli* genes, there is a bias towards their presence within the first 25 codons, which supports the role of minor codons in regulating gene expression (Chen and Inouye, 1990). Where the distance between the initiation

codon and the minor codon increases past 50-60 codons, no destabilisation effect is seen and the ribosome no longer falls off.

The arginine codons AGA and AGG are the rarest codons found in *E. coli* genes, whereas these codons are common in eukaryotes (Chen and Inouye, 1990;Makrides, 1996). Other rare *E. coli* codons include CUA (Leu), UCA (Ser), AGU (Ser), AUA (Ile), ACA (Thr), GGA (Gly) and CCC (Pro). Comparing the nucleotide sequences of the various constructs we have expressed, almost all constructs have minor codons within the first 20 codons (see Table 2.4 above). For all the N-terminal tagged constructs there is also an arginine codon as the second codon (i.e. the second codon of the entire coding sequence). However, based on relative expression levels, there does not appear to be any correlation between the number and position of these minor codons and soluble expression yields. Other studies have also shown that there are no consistent and unambiguous rules for the effect of minor codons, and several eukaryotic proteins with high arginine content have been expressed very efficiently (see reviews referenced above). This is likely to be due to interplay with some of the other factors discussed here.

ATG is the most commonly utilised start codon in *E. coli* (91%), and is therefore typically incorporated into recombinant cDNAs to increase expression (Makrides, 1996). All of our constructs utilise this codon. Incorporating alanine as the second codon and the sequence TTAA embedded within the fourth and fifth codons has also been shown to increase recombinant protein expression (Barnes, 1996). Alanine is present in this position for all our C-terminal tagged constructs while arginine is present for all the N-terminal constructs. It is possible that this plays a role in the observation that the C-terminal tagged constructs generally express better than the N-terminal tagged constructs (although it is noted that Δ43N-CPR also expresses very well). None of our constructs contain the sequence TTAA however. Studies have also indicated that there is a distinct nucleotide preference of A>T>G>C within the sequence of the first 7 or 8 codons in *E. coli* proteins. Increasing the A+T richness in this region using silent mutations is one of the more consistently successful strategies for increasing recombinant protein expression. However, there is yet again no clear correlation between A+T richness and relative construct expression (see Table 2.4 above), despite other studies showing the clearest correlation for this factor with increased expression when compared to the potential for secondary structure formation, the use of Ala as the second codon and the effect of rare codons (Gillam, 1998).

The transcription terminator is located downstream of the stop codon and serves both as a signal to terminate transcription and as a protective element composed of stem-loop structures, which help protect the mRNA from exonucleolytic degradation and thereby extend mRNA half-life (Makrides, 1996). The inclusion of extraneous 3'-untranslated sequence between the stop codon and the terminator region can have a detrimental effect on the level of CYP expression in *E. coli* and should be avoided if possible. Comparing the nucleotide sequences of the N- and C-terminal tagged constructs, the latter have much less extraneous 3'-UTR sequence (see Table 2.4 above). This too might play a role in the higher relative expression in general of the C- versus N-terminal tagged constructs.

2.2.3.2 Determining the relative insoluble and soluble expression levels of the constructs

Based on the above results it appears more likely that the greatest correlation between construct elements and soluble expression exists between the presence or absence of the membrane anchor and the placement of the His-BCCP-tag. This conclusion is supported when comparing the relative soluble and insoluble expression levels of each construct described in the following experiment. Note that here the soluble fraction refers to proteins isolated from the cytoplasm, while the insoluble fraction refers to a combination of protein sequestered into inclusion bodies and proteins bound to or associated with *E. coli* membranes.

For all the CYP3A4 constructs tested, significant expression was observed in the insoluble fraction (Fig. 2.9 A and D). In fact, for Δ24N-CYP3A4, only insoluble expression was detectable, while for FLC-CYP3A4 the greater majority of protein was insoluble. As expected from the previous results, Δ24C-CYP3A4 exhibited the highest level of soluble expression amongst the CYP3A4 constructs. However, this construct also exhibited a proportionally higher level of expression in the insoluble fraction. This result is not entirely surprising - despite the absence of a membrane anchor for the CYP3A4 deletion constructs, the CYPs are known to have membrane association regions within the catalytic region that are distinct from the N-terminal hydrophobic domain. Of course, the insoluble expression observed here might also result from the formation of inclusion bodies as a consequence of high-level CYP protein production in the *E. coli* cytoplasm (Makrides, 1996).

In terms of CPR expression, a greater proportion of the full-length constructs were expressed in the insoluble fraction, while the deletion constructs exhibited similar or greater yields in the

soluble fraction as might be expected (see Fig. 2.9B). Once again it is possible that the insoluble protein was membrane-associated and/or from within inclusion bodies.

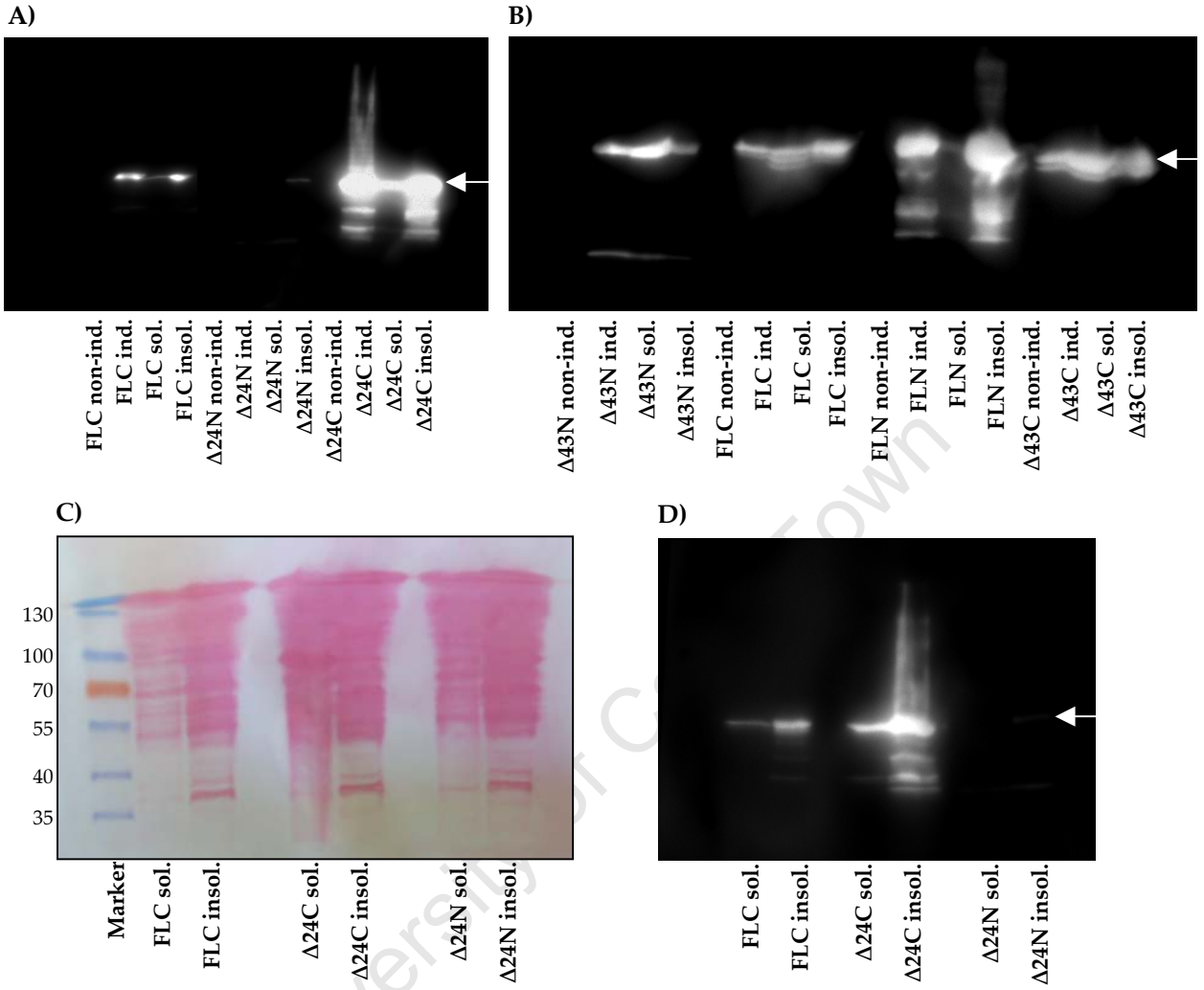


Figure 2.9: Various protein gels showing the relative expression of the CYP3A4 and CPR constructs in the soluble and insoluble protein fractions. The regulation of expression was also tested. A) Western blot of CYP3A4 protein fractions. B) Western blot of CPR protein fractions. C) Ponceau-S stain of CYP3A4 samples [repeat of (A) except with soluble and insoluble fractions only]. D) Western blot of samples in (C). Note: All Western blots were detected with an Anti-His Ab and samples were. normalised according to total protein in crude lysates. 'Non-ind.' samples were whole cell lysates prepared prior to induction with IPTG. 'Ind.' samples were whole cell lysates prepared post-induction. 'Sol.' and 'Insol.' samples represent the soluble and insoluble protein fractions respectively.

A non-induced control was also tested to determine whether expression was indeed tightly regulated using the T5 promoter/lac operon element and no leaky expression was observed (see Fig. 2.9B and D above).

From these results, it appears that both tag placement and N-terminal deletions have an effect on soluble expression levels: N-terminal deletions and C-terminal tag placements are generally more favourable than full-length constructs and N-terminal tag placement. The most highly expressed

CYP3A4 and CPR constructs are all deletion constructs and two of these are C-terminal tagged, i.e. Δ24C-CYP3A4 and Δ43C-CPR. This however does not adequately explain why Δ24N-CYP3A4 exhibits such low expression while Δ43N-CPR is so highly expressed – it is possible that tag placement is somehow more significant for CYP constructs compared to CPR constructs.

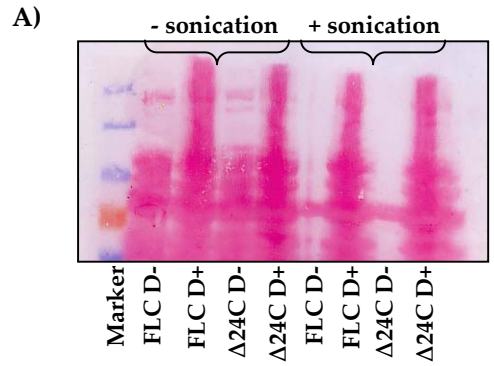
2.2.3.3 Determining the effect of high ionic strength, detergents and sonication on soluble protein yields

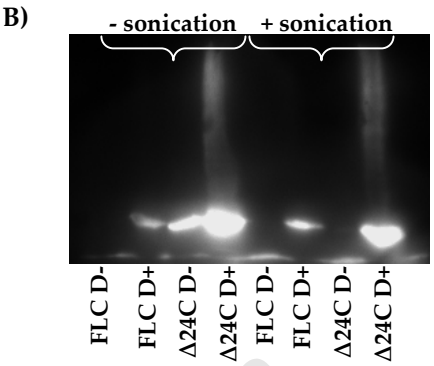
Due to the high level of expression of several of the construct proteins in the insoluble fraction, various protein extraction parameters were briefly investigated to determine whether the soluble yield could be improved. The use of detergents (CHAPS, Triton X-100 and IgePal; see Materials and Methods) had a significant effect on the soluble yield of all three constructs tested (FLC-CYP3A4, Δ24C-CYP3A4, and Δ43C-CPR; see Fig. 2.10). In the absence of detergents, sonication appeared to reduce the yield of total soluble protein, while in the presence of detergents, similar soluble protein yields were observed after sonication in comparison to extractions carried out with detergents in the absence of sonication (see Fig. 2.10 A and B). Sonication was therefore not included in further protein purifications. (Note therefore that since the experiment in Section 2.2.3.2 above was carried out using sonicated protein preparations, it is possible that the actual levels of insoluble protein may be less than that observed.there.)

When comparing the effects of detergents and high ionic strength on soluble protein yields, detergents were found to give a far greater increase in total soluble protein as well as of purified Δ24C-CYP3A4 protein (see Fig. 2.10 D). Samples containing both detergent and high salt concentrations generally became far too viscous to carry forwards for further purification and were therefore not included in this experiment.

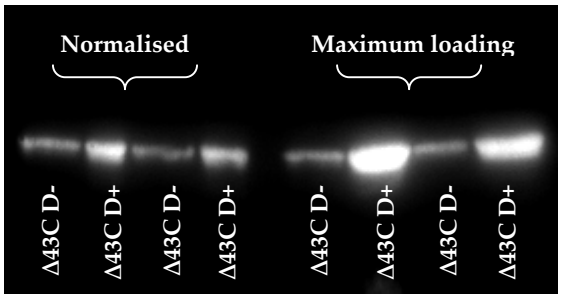
The effects of high salt concentrations and detergents on CYP purification have been studied previously. It is thought that detergents [usually non-ionic at 0.1 – 0.2% final (v/v)] disrupt peripheral membrane associations between the CYP proteins and the *E. coli* inner membrane due to the hydrophobic nature of these associations (Pernecky and Coon, 1996), similarly with CPR (Murataliev *et al.*, 2004). The use of high salt concentrations is thought to disrupt any electrostatic associations present between positively charged residues on the protein surface and the negatively charged phospholipid heads of the inner membrane. From the above experiments it appears that peripheral membrane interactions play a far greater role in reducing soluble protein

yields of our constructs than electrostatic interactions, and that this effect can be mitigated using detergents.





C)



D)

	Lysozyme only	Lysozyme + Salt	Lysozyme + Detergents
Yellowness of lysate	+	++	++++
Crude lysate [protein]	5.45 mg/ml	8.78 mg/ml	14.22 mg/ml
Pure [protein]	23.22 μg/ml	48.55 μg/ml	121.37 μg/ml

Figure 2.10: Determining the effect of detergents and sonication on CYP3A4 and CPR protein yields. A) Ponceau-S stain of FLC- and Δ24C-CYP3A4 crude lysates extracted with ('D+') and without ('D-') detergents and with and without sonication. B) Western blot of samples in (B). C) Western blot of Δ43C-CPR purified protein samples extracted with and without detergents. D) Table illustrating the effects of high salt concentrations and detergents on cell lysis and purified Δ24C-CYP3A4 protein yield. The colour of the lysate is influenced by the amount of heme-containing P450. Note: All Western blots were detected with an Anti-His Ab. Maximum loading volumes were used for gels (A) and (B), while for gel (C) both maximum loading and crude lysate-normalised volumes were used (see labels).

2.3 CONCLUSIONS

Using traditional cloning and protein expression methods, the generation of many possible combinations of constructs and their analysis in different expression systems is far too labour-intensive and time-consuming to be practical (Hunt, 2005). The traditional approach has therefore been to intuitively select a handful of constructs in the first instance, analyse, optimise, and repeat the process until the most appropriate conditions have been found. It is this latter basic approach that we have pursued here.

Based on cloning and expression methodologies chosen after careful literature reviews, we have successfully cloned and expressed various full-length and N-terminal deletion CYP3A4 and CPR constructs. The relative soluble and insoluble expression yields of the various constructs were compared and in general we found that constructs without the membrane anchor and with a C-terminal His-BCCP tag provided the highest soluble protein expression. Tag placement and N-terminal modifications appeared to have a greater influence on protein yields than nucleotide sequence elements. Despite the high level of apparently insoluble expression observed even for certain deletion constructs, the use of detergents was shown to significantly increase soluble protein yield probably through the disruption of peripheral membrane interactions.

While many protein expression and purification factors could be further optimised so as to further increase CYP3A4 and CPR soluble protein yield (e.g. IPTG concentrations, expression temperatures and purification conditions) [see e.g. ((Barnes, 1996;Hannig and Makrides, 1998;Makrides, 1996)], the expression levels obtained were deemed sufficient to continue with further biophysical characterisation, for determining which of the constructs were the most likely candidates for array construction. Ultimately activity was a more important criterion for our purposes than yield. These experiments are discussed in the following chapter.

2.4 MATERIALS AND METHODS

2.4.1 Recombinant cloning of the CYP3A4 and CPR constructs

2.4.1.1 Design of N-terminal deletion constructs

Various online software tools were used to delimit the likely membrane-spanning sequences at the N-termini of human CYP3A4 and CPR. These tools were Statistical Analysis of Protein Sequences (SAPS; http://www.ebi.ac.uk/Tools/saps/), SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/) and MEMSAT (http://bioinfadmin.cs.ucl.ac.uk/memsat/). The membrane anchor deletions for the Δ24-CYP3A4 and Δ43-CPR were then chosen based on the consensus sequence obtained.

2.4.1.2 Original parent vectors

Agar stab cultures containing Δ34N-CYP3A4 and FLN-CPR (both cloned into the pMD004 vector), as well as Δ24C-CYP3A4 (cloned into the pBJW102.2 vector) were a gift from Procognia Ltd, Maidenhead, UK. These vectors were used as the parent vectors for all further construct

cloning work. The agar stab cultures were used to inoculate 5 ml each of Luria-Bertani medium supplemented with 100 μg/ml ampicillin [LB(Amp)]. The cultures were incubated O/N at 37°C with shaking. Glycerol stocks of these (and later constructs) were then produced by mixing 0.5 ml glycerol together with 0.5 ml of each culture followed by storage at -80°C.

2.4.1.3 Primer design and synthesis

All primers were designed according to standard guidelines with regards to length, T_m, GC content and non-self or cross-complementarity (Ausubel *et al.*, 1999). Each primer set encoded an appropriate restriction site such that the amplified product could later be digested and cloned unidirectionally into the multiple cloning site (MCS) of the relevant plasmid vector (i.e. pMD004 or pBJW102.2). Inverse primers (i.e. 'back-to-back' primers) were used to create all constructs except for FLN-CYP3A4, for which normal primers were used. Software programs such as Vector NTI Advance 11.0 (Invitrogen, USA) and DNAMan 4.13 (Lynnon Corporation, Canada) were used to design and analyse potential primers according to these guidelines. Primers were synthesised using standard methods by the Synthetic DNA Laboratory, Molecular and Cell Biology Department, University of Cape Town. Primers used for constructs generated by the author are listed in Appendix I.

2.4.1.4 Plasmid DNA extraction and quantification

A freshly plated colony grown on LB agar was used to inoculate 5 ml LB(Amp) medium. The culture was incubated O/N at 37°C with shaking. Plasmid extraction from cultures was carried out using the Qiaprep Spin Miniprep kit (Qiagen, Germany), according to manufacturer's instructions. Plasmid DNA was quantified using UV spectroscopy at A_{260} with a Nanodrop spectrophotometer (Thermo Scientific, USA).

2.4.1.5 PCR amplifications

In general, PCR reactions consisted of 1 μM of each primer, 1-2 ng/μl plasmid DNA, 200 μM dNTP mix, 1.5 mM $Mg2^+$, 2.5 U of the high fidelity Pwo polymerase enzyme (Roche, Germany), 1 x Pwo PCR buffer and water in a 50 μl reaction volume. For some PCR reactions, titration of Mg^{2+} was necessary to obtain satisfactory yields. The PCR amplification cycles were carried out according to standard guidelines and relevant to the T_m of the primer pair used (Ausubel *et al.*, 1999).

For recombinant colony screening after transformations, crude colony PCRs were performed on bacterial cell suspensions using standard methods and incorporating the PCR procedure detailed above, with the exception that the Taq polymerase (Fermentas, Canada) was used (Ausubel *et al.*, 1999).

2.4.1.6 Restriction enzyme (RE) digestions

PCR products were subjected to RE digestion with the relevant enzymes such that the PCR products could be unidirectionally ligated into a similarly digested preparation of plasmid vector. Restriction enzyme mapping was also carried out for certain constructs during recombinant colony screening. Enzymes were obtained from New England Biolabs (USA) and were used according to standard protocols and manufacturer's instructions. In general, the RE digests consisted of ~500 ng of DNA, ≥ 2 U of each RE enzyme, 1 x RE buffer, water, and any recommended additives (e.g. BSA) up to a final volume of 20 µl. Digests were then left for 1 hr - O/N at the recommended temperature. Where RE conditions differed greatly, dual-enzyme restrictions were carried out sequentially to ensure satisfactory yields. REs used for constructs generated by the author are listed in Appendix I.

2.4.1.7 Other DNA manipulations

For those inverse PCR products that were self-ligated without prior RE digestion (i.e. during the generation of the pMD004 and pBJW102.2 backbone vectors), T4 kinase treatment was carried out to phosphorylate the 5' ends such that ligation could occur. In general, kinase reactions consisted of ≤ 300 pmol DNA, 1 x kinase buffer, 1 mM ATP, 10 U T4 polynucleotide kinase (New England Biolabs) and water in a 50 µl reaction volume. Reactions were incubated 30 min - O/N at 37°C.

To minimise the potential for a high background of parental plasmid transformation, PCR reactions were further treated by DpnI digestion, which cleaves methylated parental DNA. In general, these digestions consisted of ~500 ng DNA, 1 x reaction buffer, 10 U DpnI (New England Biolabs) and water in a 20 µl reaction volume. Reactions were incubated for 3 hrs at 37°C.

2.4.1.8 DNA ligations

In general, DNA ligations consisted of ~50 - 150 ng/µl DNA, 1 x ligation buffer, 20 U/µl T4 DNA Ligase (New England Biolabs) and water in a 20 µl reaction volume. Reactions were incubated O/N at RT or 16°C. For ligations of PCR product into linearised vector, a variety of PCR

product:vector ratios were performed for testing in later transformations. In general, 3:1, 1:1, 1:3 insert:vector molar ratios as well as a high (70:1) insert:vector ratio were used, and the latter consistently gave the best results. A negative control was also included for later transformation (linearised vector only).

2.4.1.9 Agarose gel electrophoresis and DNA clean-up

DNA products resulting from PCR amplification or processing steps such as RE digestion were subjected to agarose gel electrophoresis to determine yield. In general, 1% agarose containing 1 μg/ml ethidium bromide was used to cast the gel, while 1 x Tris-acetate buffer (TAE, pH 8) was used as the running buffer according to standard procedures (Ausubel *et al.*, 1999). Gels were then visualised using a standard UV lightbox. DNA ladders in the relevant MW range were included on each gel for size determination of the PCR/RE digestion products (Fermentas). When non-specific PCR products were present in addition to the expected PCR bands, the target PCR bands were excised from the gel and the DNA extracted using a QIAquick gel extraction kit (Qiagen) according to manufacturer's instructions. When only the target bands were present, the relevant PCR reactions were pooled and the DNA was subjected to a QIAquick PCR purification kit (Qiagen) to remove unincorporated primers and dNTPs. DNA yields were then quantified as described above.

2.4.1.10 Recombinant DNA transformations

Competent DH5α cells were prepared as follows: 5 ml LB was inoculated with a freshly plated colony of *E. coli* DH5α cells and incubated O/N at 37°C with shaking. This pre-culture was used to inoculate 100 ml LB media and incubated at 37°C with shaking until $OD_{550} \approx 0.5$. The culture was put on ice for 15 min and then centrifuged at 5000 rpm in a pre-chilled rotor at 4°C for 5 min. The cell pellet was resuspended in 40 ml TFBI solution (30 mM KOAc, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$ and 15% (v/v) glycerol, pH 5.8), put on ice for 15 min and then centrifuged as before. The cell pellet was then resuspended in 4 ml TFBII solution (10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl and 15% (v/v) glycerol, pH 6.5), put on ice for 15 min, and then divided into 100 μl aliquots for storage at -80°C.

One tube of 100 μl DH5α competent cells per transformation was thawed on ice for 5 min. 10 ng DNA from each ligation reaction was added to each tube, the tube was flicked gently to mix and incubated on ice for 30 min. The cells were heat-shocked at 42°C for 90 sec and then put back on

ice for 5 min. 900 µl LB media was added to each tube and the tubes were incubated at 37°C with shaking for 1 hr. 100 µl of each transformation mix (therefore a 1:10 final cell dilution), as well as a 100 µl of a 1:10 dilution of the transformation mix (therefore a 1:100 final cell dilution) was prepared in LB(Amp) media and plated onto LB(Amp) plates. The plates were incubated at 37°C O/N and colonies were then screened using colony PCR and/or RE mapping as described above.

2.4.1.11 DNA Sequencing

Potential recombinant clones identified by colony PCR screening were sent for sequence verification. All sequencing was carried out by the DNA Sequencing Service, Molecular and Cell Biology Department, University of Cape Town. For each sequencing reaction 5 µM of each primer and 500-1000 ng plasmid DNA were supplied. Sequence data was analysed using Chromas 2.01 (Technelyesium Pty Ltd., Australia) and aligned to Swiss-Prot entry P08684 for CYP3A4 constructs and P16435 for CPR constructs.

2.4.2 CYP3A4 and CPR construct expression and purification

2.4.2.1 CYP3A4 and CPR construct expression

For each culture a freshly plated colony was picked and used to inoculate a 5 ml LB(Amp) pre-culture. The pre-culture was incubated O/N at 37°C with shaking and then used to inoculate 500 ml Terrific Broth media supplemented with 100 µg/ml ampicillin, 1 mM thiamine and trace elements (per litre: 27 g $FeCl_3$, 2 g $ZnCl_2$, 2 g $CoCl_2$, 2 g Na_2MoO_4, 1.3 g $CaCl_2$, 1 g $CuCl_2$, 0.5 g H_3BO_3, 100 ml concentrated HCl; use at 2.5 ml/L media). The 500 ml culture was incubated at 37°C with shaking at 150 rpm until the culture OD_{600} was between 0.4-0.6. At this point CYP3A4 cultures were supplemented with 0.5 mM δ-aminolevulinic acid (δ-ALA) to facilitate heme synthesis and incubated at 30°C, 150 rpm for 30 min. CYP3A4 and CPR protein expression was induced by the addition of 1 mM IPTG and the culture was supplemented with 50 µM biotin to facilitate BCCP-tag biotinylation. Induced cultures were then incubated at 30°C, 150 rpm for 18 hrs.

2.4.2.2 Preparation of crude lysates

Expression cultures were harvested by centrifugation at 5000 xg for 15 min at 4°C. The cell pellets were resuspended in 250 ml each of 1 x phosphate-buffered saline (PBS; pH 7.4) to remove any unincorporated biotin. The cells were then pelleted as before. The cell pellets were resuspended in

5 ml lysis buffer (20 mM phosphate buffer, pH 7.4; 20% glycerol (v/v), 10 mM β-mercaptoethanol), with the addition of 1 mg/ml lysozyme, 500 U DNase I, 0.5 mM $CaCl_2$, 2.5 mM $MgCl_2$, using a 10 ml final volume for each culture (DNaseI sourced from New England Biolabs). The cell suspensions were incubated on ice for 30 min with gentle shaking. Following this, 0.15% IgePal CA-630, 0.5% CHAPS and 0.1% Triton X-100 (Sigma Aldrich, USA) were added and the cell suspension were incubated on ice as before. The cell suspensions were then centrifuged at 10,000 xg for 30 min at 4°C and the resulting supernatants were the crude soluble lysate preparations. 100 µl of each lysate was removed for storage at -20°C for later analysis.

2.4.2.3 Protein purification

Protein purification resin was packed into PD-10 columns (GE Healthcare, USA). Either HisLink (Promega, USA), TALON (Clontech, USA), Ni-NTA or Ni-TED resins (Macherey-Nagel, Germany) were used according to manufacturers' instructions with the provided buffers, with the exception that the HisLink wash buffer was supplemented to 75 mM imidazole, while the elution buffer was supplemented to 250 and 500 mM imidazole for sequential elutions of 2.5 ml each. Centriprep YM-10 columns (Millipore, Ireland) were then rinsed with 15 ml dH_20 for 15 min at 3000 xg. Each resin-purified sample was added to a Centriprep column and centrifuged at 3000 xg at 4°C until the protein samples were concentrated to < 700 µl (~35 min). Each protein concentrate was then added to a Zeba column (Pierce, USA) for buffer exchange into P450 Storage Buffer (100 mM potassium phosphate, pH 7.4; 20% glycerol (v/v), 0.2 mM EDTA, 1 mM DTT) according to manufacturer's instructions. Proteins were eluted in <1 ml each from the Zeba columns and divided into aliquots for -20°C storage.

2.4.2.4 Total protein quantification

Low expression levels made routine spectral quantification of proteins impractical. Total protein concentrations were therefore estimated using the Bio-Rad Protein Assay Dye based on the Bradford assay (Bio-Rad, USA) according to manufacturer's instructions. The standard assay was used for crude lysate samples while the microassay was used for purified samples. Bovine serum albumin (BSA; Sigma Aldrich) was used to create standard curves. Note that the concentrations of detergents used during cell lysis were within the recommended range for this assay.

2.4.2.5 Gel electrophoresis and Western blot transfer

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot transfers of protein were carried out according to standard methods (Ausubel *et al.*, 1999). In general, 7% polyacrylamide gels were used for CPR samples while 10% gels were used for CYP3A4 samples. Protein molecular weight markers used were the PageRuler pre-stained protein ladder and the PageRuler unstained protein ladder (both from Fermentas). SDS-PAGE gels were either stained using Coomassie Blue solution [0.25% (w/v) Coomassie Blue, 50% (v/v) methanol, 10% (v/v) acetic acid] O/N with gentle shaking and then destained >1 hr in destaining solution [45% (v/v) methanol and 10% (v/v) acetic acid], or were used for Western blot protein transfer onto Hybond-C extra membrane (GE Healthcare). Successful protein transfer, sample loading and protein marker positions were determined by staining the membranes with Ponceau-S solution (Sigma Aldrich) for < 5 min and destaining afterwards in dH_20. The membranes were blocked in 5% fat-free milk powder in 1 X TBS buffer (10 x TBS: 0.5 M Tris, 0.66 M NaCl, pH 7.6) for 1-2 hours and then incubated in a 1:2000 dilution of Anti-His HRP-conjugated antibody (#A7058; Sigma Aldrich) in 2% fat-free milk powder in 1 x TBS, O/N at RT. Membranes were then washed 3 x 5 min in 0.1% Tween-20 in 1 X TBS, overlaid with a luminol solution (Sigma Aldrich) and detected with either a chemiluminescence detector or X-ray films.

2.4.2.6 Protocol for determining relative expression of constructs in insoluble and soluble protein fractions

For each culture a freshly plated colony was picked and used to inoculate a 10 ml LB(Amp) pre-culture. The pre-culture was incubated O/N at 37°C with shaking and 2.5 ml was used to inoculate 50 ml Terrific Broth supplemented with ampicillin and trace elements. The 50 ml culture was incubated at 37°C with shaking at 150 rpm until the culture OD_{600} was between 0.4-0.6. At this point 1 ml of each culture was removed and centrifuged at 13,000 rpm for 2 min and the resulting cell pellet was resuspended in 50 µl 5 x sample application buffer (SAB) and stored at -20°C (non-induced control). CYP cultures were then supplemented with δ-ALA and incubated at 30°C, 150 rpm for 30 min, after which CYP and CPR protein expression was induced by the addition of 1 mM IPTG. Expression was carried out at 30°C, 150 rpm for 4-5 hrs.

After expression 1 ml of each culture was removed and centrifuged as before and the cell pellet was resuspended in 100 µl 5 x SAB and stored at -20°C (induced control). The remaining culture was centrifuged at 4100 rpm for 20 min at 4°C. The cell pellets were resuspended in 1 ml lysis buffer, with the addition of 0.2 mg/ml lysozyme, 20 U DNase I, 1 x DNase I buffer and 4 mM

$MgCl_2$ using a 2 ml final volume for each culture. The cell suspensions were incubated 4°C for 30 min with gentle shaking. The cell suspensions were then sonicated for 6 x 10 sec with 10 sec pauses while on ice. The cell lysates were centrifuged at 10,000 xg for 20 min at 4°C. The pellet was resuspended in 1 ml P450 Storage Buffer. 10 µl 5 x SAB was added to 10 µl of each supernatant (soluble fraction) and pellet suspension (insoluble fraction) for storage at -20°C. All fractions were later analysed by SDS-PAGE and Western blot.

2.4.2.7 Protocol for determining the effect of high ionic strength, detergents and sonication on soluble protein yields

The following procedure was used to prepare the protein samples illustrated in Fig. 2.10A and B. FLC- and Δ24C-CYP3A4 protein expression cultures were carried out as in Section 2.4.2.1. After expression the cultures were centrifuged at 5000 xg for 15 min at 4°C. The cell pellets were resuspended in 1 x PBS and harvested as before. The cell pellets were then lysed in lysis buffer supplemented with lysozyme, DNase I, $CaCl_2$, $MgCl_2$ as in Section 2.4.2.2. Each lysate was then divided into two equal volumes. To one half of each lysate, 0.15% IgePal, 0.5% CHAPS, 0.1% Triton X-100 and an equal final volume of Buffer A (1 M potassium phosphate, pH 7.4; 20% glycerol (v/v), 10 mM β-mercaptoethanol) was added (D+ samples), while nothing was added to the other half of each lysate (D- samples). All lysates were then incubated on ice for 30 min at 4°C with gentle shaking. At this point, 100 µl of each lysate was removed and centrifuged at 13,000 rpm for 2 min at 4°C and the resulting supernatants were stored at -20°C (un-sonicated samples). The rest of the lysate was then sonicated for 3 x 45 sec with 90 sec pauses while on ice. 100 µl of each sonicated lysate was removed and centrifuged as before and the resulting supernatants were stored at -20°C (sonicated samples). Prior to SDS-PAGE analysis, those samples that had not been treated with detergents and Buffer A were diluted 1:1 in P450 Storage Buffer.

The same procedure was carried out for the Δ43C-CPR protein samples illustrated in Fig. 2.10 C, with the following amendments. Subsequent to lysozyme treatment, each lysate was divided into two equal volumes. To one half of each lysate, 0.15% IgePal, 0.5% CHAPS, 0.1% Triton X-100 and an equal final volume of Buffer A was added (D+ samples) while only an equal final volume of Buffer A was added to the other half (D- samples). All lysates were then incubated on ice for 30 min at 4°C with gentle shaking. At this point, each lysate was centrifuged at 10,000 xg for 30 min at 4°C and the resulting supernatants stored at -20°C.

A similar procedure was carried out for the Δ24C-CYP3A4 protein samples described in the table inset in Fig. 2.10 D, with the following amendments. The culture was divided into four equal volumes. Subsequent to lysozyme treatment, one culture had detergents and an equal volume of Buffer A added ('lysozyme + detergents + salt'), the second culture only had detergents added ('lysozyme + detergents'), the third only had salt added ('lysozyme + salt), while nothing was added to the fourth culture ('lysozyme only'). All lysates were incubated on ice for 30 min at 4°C with gentle shaking and then centrifuged at 10,000 xg for 30 min at 4°C. 100 µl of each of the resulting supernatants were put in -20°C storage as crude lysate samples while the rest of the supernatants were purified as described in Section 2.4.2.3.

CHAPTER THREE

Biophysical Characterisation of CYP3A4 & CPR Constructs

3.1 INTRODUCTION

Once adequate levels of soluble protein expression had been achieved for the majority of the CYP3A4 and CPR constructs, these constructs were taken forward for further characterisation using various biophysical methods. This was carried out to determine which of the various constructs most closely resembled the native proteins in terms of conformation and catalytic activity. Once the optimal CYP3A4 and CPR constructs were determined it would then be possible to investigate a potential format for co-immobilising these partner enzymes in an array format such that they could form a functional complex for electron transfer. The experimental objectives of this chapter were therefore the following:

1) To compare the level of secondary structure exhibited by each of the constructs using circular dichroism methods.
2) To compare the catalytic activity of the constructs using known substrates.
3) Based on the above results, to determine those constructs with the most promise for protein array fabrication.
4) To determine the biotinylation status of the selected constructs.

3.2 RESULTS AND DISCUSSION

3.2.1 Circular dichroism investigation of CYP3A4 and CPR construct proteins

Circular dichroism spectroscopy (CD) was performed for all CYP3A4 and CPR constructs within the far-UV range, using proteins purified from the soluble fraction (see Fig. 3.1 and Fig. 3.2). As the CYP3A4 and CPR construct proteins all include protein tags and several of these contain N-terminal deletions, it was not possible to directly compare the CD spectras obtained here with those available in CD databases. However, several conclusions could be made based on a visual examination of the results. In general, the C-terminal tagged constructs of both the CYP3A4 and CPR proteins exhibited the most ordered secondary structure (predominantly α-helical in nature as expected from the crystal structures discussed earlier and from published CD spectra [see e.g. (Dodhia *et al.*, 2006;Kim *et al.*, 2003;Marohnic *et al.*, 2006)], while the N-terminal tagged proteins exhibited very little to no ordered structure (i.e. no defined peaks and signals approached zero). Once again however, Δ43N-CPR defies this general rule and shows similar levels of foldedness compared to Δ43C-CPR. Bovine serum albumin was included as a positive control for both experiments and yielded the expected curves (Nozaki *et al.*, 1974).

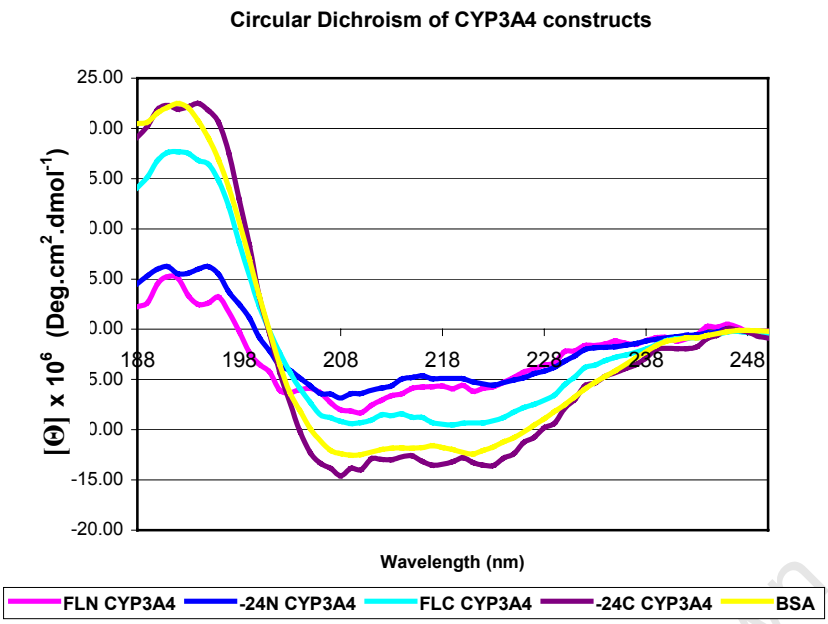


Figure 3.1: Circular dichroism spectroscopy of the various CYP3A4 constructs in the far-UV region illustrating the relative amount of ordered secondary structure present for each. Curves represent the average of 10 spectra.

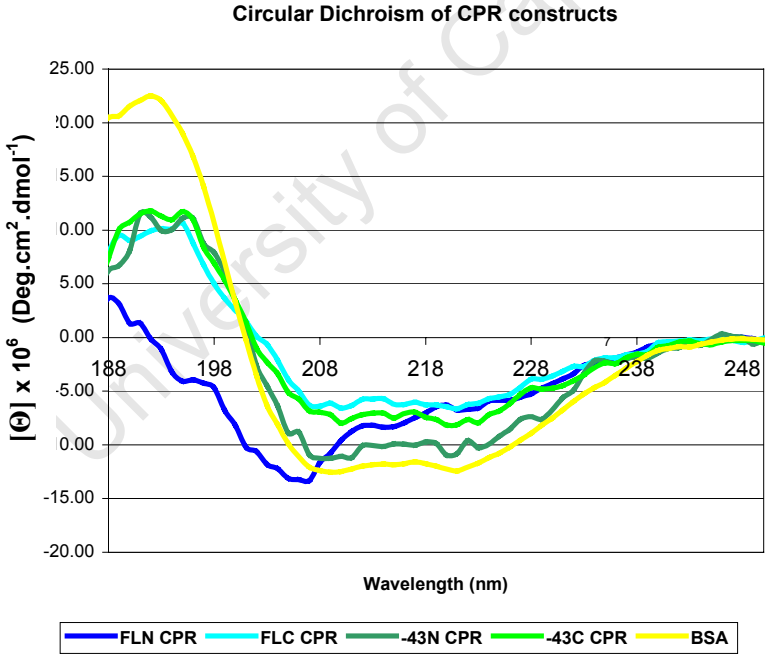


Figure 3.2: Circular dichroism spectroscopy of the various CPR constructs in the far-UV region illustrating the relative amount of ordered secondary structure present for each. Curves represent the average of 10 spectra.

These results correlate well to the relative insoluble and soluble expression levels of the constructs discussed in Chapter Two. Those constructs with little to no soluble expression exhibit the least ordered structure, suggesting that what little protein there is present in the soluble fraction is

denatured. It is therefore likely that these construct proteins are either not properly folded as nascent proteins or that their folded structure is unstable and therefore rapidly denatured. This implies that the insoluble fraction for these proteins consists of proteins mainly targeted towards inclusion bodies.

It is not immediately clear why the C-terminal tag would confer greater protein stability or favour correct protein folding compared to the N-terminal tag. However, such observations are typical for fusion tags in a protein-specific manner and are seldom easy to rationalise (Korf *et al.*, 2005). It is interesting that despite the C-terminal tag favouring the folding of the full-length constructs as well as of the deletion constructs, the FLC proteins are still targeted more towards the insoluble fraction (as discussed in Chapter Two). This suggests that the insoluble expression of the FLC proteins is not due to the aggregation of denatured proteins in inclusion bodies but rather due to insertion into or association with the *E. coli* inner membrane resulting from the hydrophobic nature of the N-terminal membrane anchor.

The following conclusions can therefore be drawn from the expression data from Chapter Two and the circular dichroism studies discussed here. FLN-CYP3A4, Δ24N-CYP3A4 and FLN-CPR are poorly expressed in the soluble fraction due to decreased stability or incorrect folding which targets them towards inclusion bodies. FLC-CYP3A4 and FLC-CPR are folded but the presence of the hydrophobic domain targets these proteins towards the inner membrane. Δ24C-CYP3A4, Δ43N-CPR and Δ43C-CPR are all folded and due to the absence of the hydrophobic domain yield the highest levels of soluble expression. These conclusions were further tested in the following section using functional assays.

3.2.2 Enzyme activity assays of CYP3A4 construct proteins

3.2.2.1 Enzyme activity assay of commercial baculosomal CYP3A4 protein as a positive control

Commercial baculosome preparations co-expressing human CYP3A4 and CPR were used as a positive control for testing the turnover of the Vivid Red substrate (both sourced from Invitrogen). A NADPH regeneration system was used as an electron source and enzyme turnover was assayed by monitoring the production of the fluorescent product, resorufin. Reactions were carried out according to the manufacturer's guidelines and similar kinetic data was obtained (see

Fig. 3.3 - 3.5 below). The reactions were linear over the time period used with an $R^2 > 0.95$ except for the 38.4 μM reaction (see Fig. 3.3).

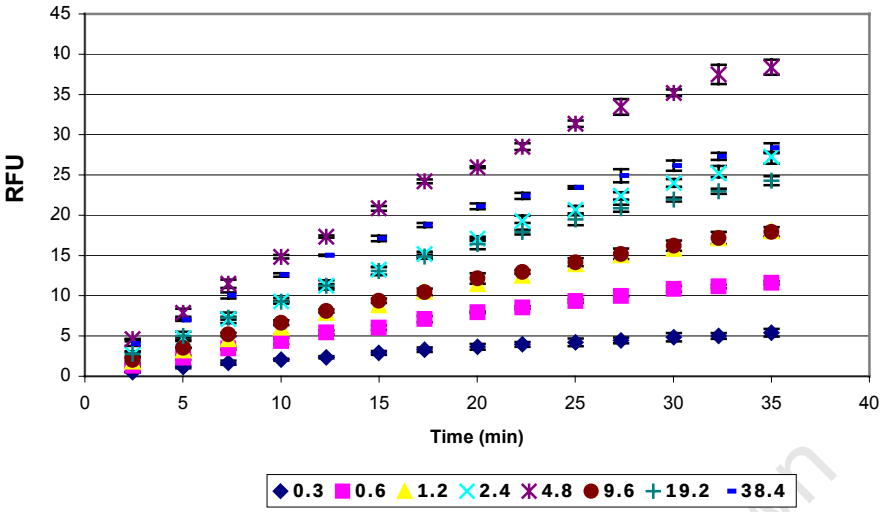


Figure 3.3: Graph of the increase in relative fluorescence units (RFU) over time resulting from baculosomal CYP3A4 metabolism of various Vivid Red concentrations (units in μM). Note: Values are an average of three replicates and error bars represent standard deviations.

The velocities obtained from linear regression of the averaged readings were then used to determine the kinetic constants of the reaction. Initial plots of the velocity vs. [substrate] data showed a non-hyperbolic kinetic profile (see Fig. 3.4). Above 4.8 μM substrate, a significant decrease in velocity was observed. Initially we surmised that this inhibition was due to solvent or substrate inhibition effects (the 9.6 μM reaction contains 0.48% acetonitrile, 19.2 μM contains 1% while the final 38.4 μM reaction contains 1.9%). However, acetonitrile is generally considered to only inhibit CYP activity at concentrations > 1% (Venkatakrishnan *et al.*, 2003) and an increase in activity was observed for the two highest substrate concentrations rather than a further decrease as might be expected for solvent inhibition. Solvent inhibition was therefore only likely to play a small role in the kinetic profile obtained here.

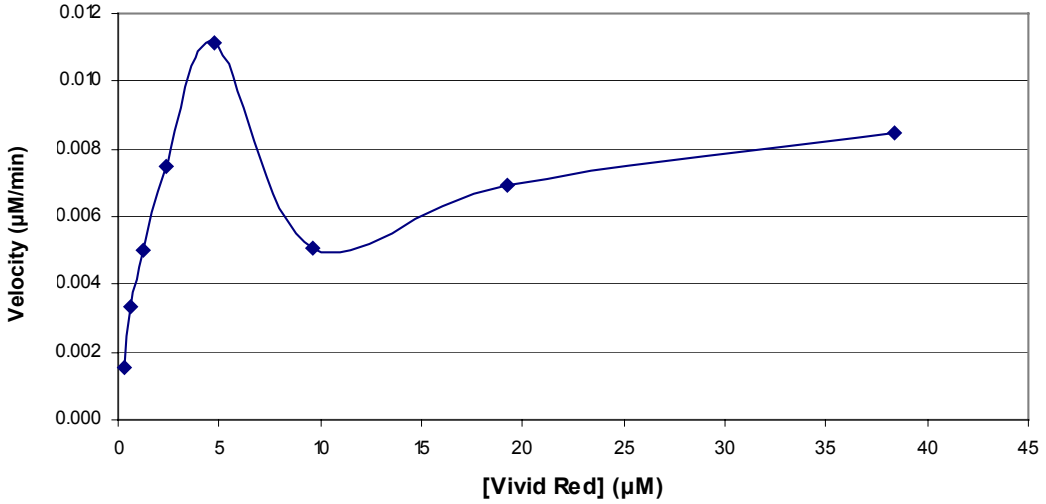


Figure 3.4: Baculosomal CYP3A4 turnover of Vivid Red displaying a non-hyperbolic kinetic profile.

The literature suggests a K_m of 3 µM for Vivid Red when using a baculosomal CYP3A4 preparation and our assay therefore included a substrate concentration range 10-fold above and below this value. However, our data for commercial baculosomal CYP3A4 only matched the literature K_m if we excluded data from the higher substrate concentrations that exhibited a decrease from the initial maximum (i.e. > 4.8 µM in this experiment; see Fig. 3.5). This anomaly was further investigated at a later stage using our recombinant CYP3A4 protein (see Chapter Six), however for the purposes of the rest of this chapter, only data obtained using ≤ 6 µM Vivid Red was used for analysis and typical Michaelis-Menten kinetics were used to model the data.

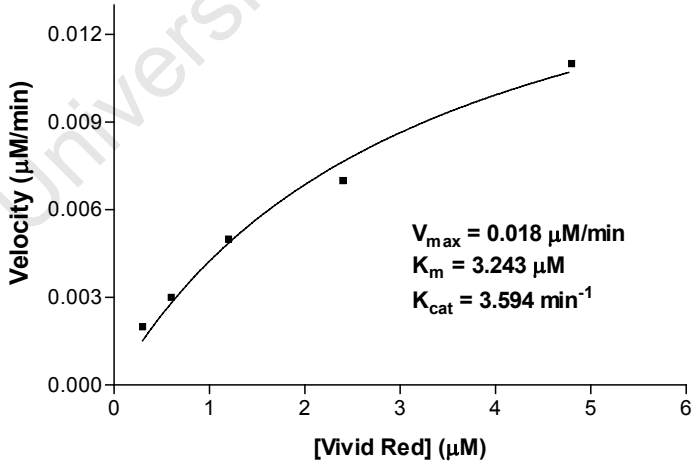


Figure 3.5: Baculosomal CYP3A4 turnover of Vivid Red using CPR and NADPH as an electron source. Note: Data points are based on velocities from Fig. 3.4 above.

3.2.2.2 Testing Vivid Red turnover using the Δ24C-CYP3A4 construct and cumene hydroperoxide

Once the basic Vivid Red assay protocol had been established using commercial baculosome CYP3A4 preparations, we moved our attention towards assaying the functional activity of the

various CYP3A4 construct proteins we had generated. However, rather than using electron transfer *via* one of the CPR construct proteins and the NADPH regeneration system in the first instance, we chose instead to use cumene hydroperoxide (CuOOH) as an electron source *via* the peroxide shunt pathway, so as not to complicate the analysis by relying on reconstitution of activity from two recombinantly engineered enzymes.

The use of hydrogen peroxide in the metabolism of substrates by CYPs has long been established (Chefson *et al.*, 2006;Miles *et al.*, 2000). However, the high concentrations of H_2O_2 typically used resulted in very rapid loss of enzyme activity due to the uncontrolled oxidative destruction of the heme group by the peroxide. By comparison, CuOOH is an organic peroxide that is capable of stimulating CYP activity at much lower concentrations and with lower rates of heme destruction. Since the level of CuOOH stimulation is known to vary according to the CYP isoform and the substrate used, we initially tested three widely-ranging CuOOH concentrations (0.05, 0.5 and 5 mM) to determine the lowest optimal concentration for Vivid Red metabolism.

For these initial experiments, we used the CYP3A4 construct most likely to be active based on its soluble expression levels and CD results, i.e. Δ24C-CYP3A4. The 5 mM CuOOH sample was observed to give both the lowest overall turnover and initial rate of reaction (see Fig. 3.6); the 0.05 mM sample gave the highest overall turnover and a median initial rate; while the 0.5 mM sample gave median turnover but was also observed to have the highest initial rate of reaction.

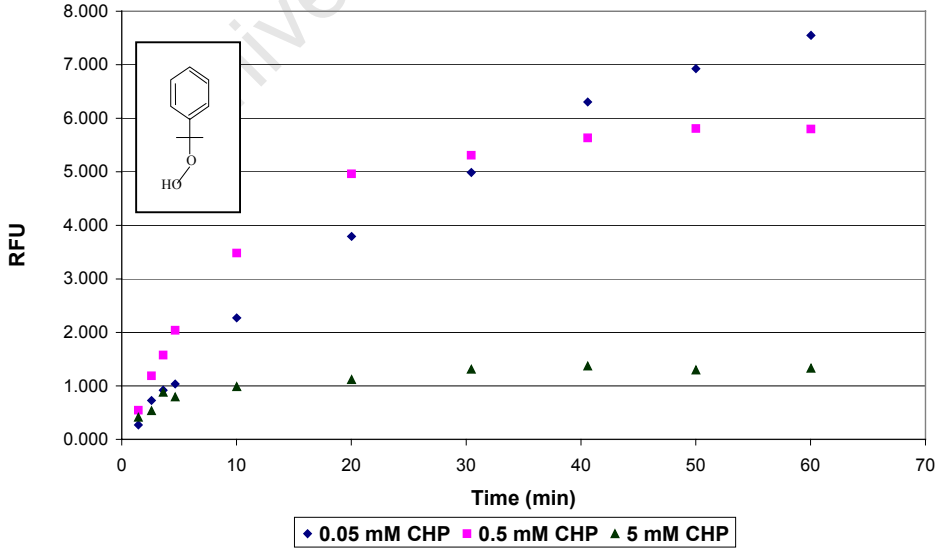


Figure 3.6: Graph of the increase in fluorescence over time resulting from Δ24C-CYP3A4 metabolism of Vivid Red at various CuOOH concentrations (0.05 – 5 mM). One series of data was performed for 2.5 μM and 25 μM Vivid Red with similar effect (2.5 μM results are shown here). The structure of CuOOH is shown (*inset*).

We therefore opted to study CuOOH concentrations over a narrower range (0.1 – 1 mM) to further refine the optimal concentration (see Fig. 3.7). The results showed that CuOOH concentrations between 0.5 – 1 mM yielded similar reaction rates and overall turnover, while the concentrations tested above and below this range yielded comparatively lower turnover. This confirmed that 0.5 mM CuOOH was optimal for further CYP3A4 construct testing using peroxide-driven metabolism (the lower CuOOH concentration was chosen due to the negative effects of peroxide as discussed previously).

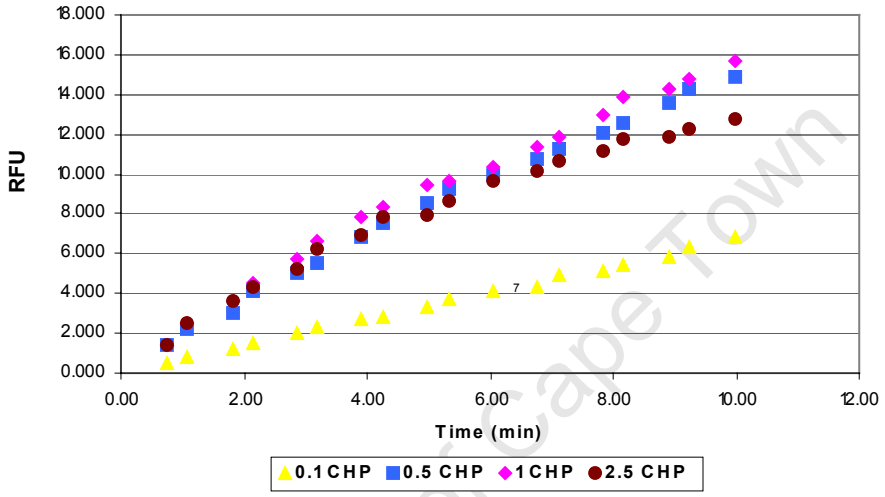


Figure 3.7: Graph of the increase in fluorescence over time resulting from Δ24C-CYP3A4 metabolism of Vivid Red at various CuOOH concentrations (0.1 – 2.5 mM). One series of data was performed for 4 μM Vivid Red for two different enzyme preparations with similar effect (results from the fresher enzyme preparation are shown here).

3.2.2.3 Enzyme activity assays of the CYP3A4 construct proteins

CuOOH-driven Vivid Red metabolism was used to compare the functional activity levels of the CYP3A4 construct proteins according to the assay conditions ascertained in the above experiments. Of the CYP3A4 constructs, the deletion construct Δ24C-CYP3A4 exhibited the greatest enzyme activity (in terms of both V_{max} and k_{cat}) and a K_m value closest to that of 3 μM reported in the literature (see Fig. 3.8; Table 3.1). The FLC-CPR construct exhibited around 3.5-fold less activity compared to Δ24C-CYP3A4, while the FLN-CYP3A4 and Δ24N-CYP3A4 constructs exhibited negligible activity. These results confirmed the conclusions drawn in Section 3.2.1 above.

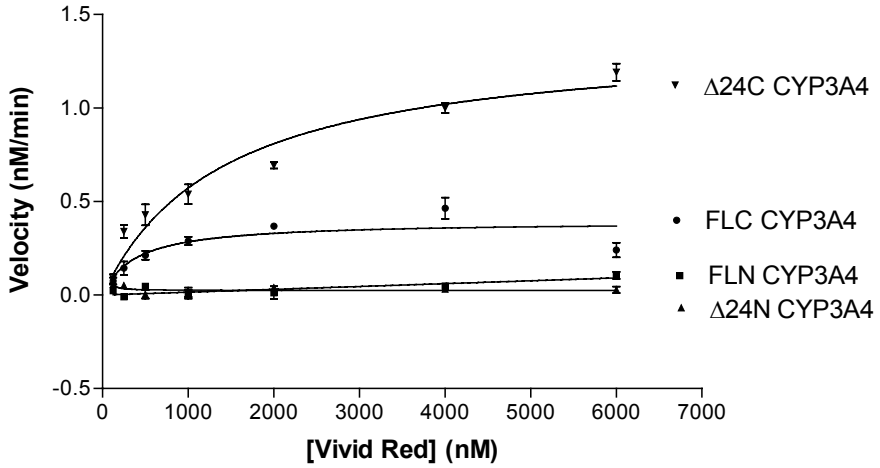


Figure 3.8: Enzyme assays of the CYP3A4 constructs using Vivid Red substrate and CuOOH as the electron donor. Note: Data points are an average of three replicates and error bars represent standard deviation.

Construct	V_{max} (nM/min)	K_m (μM)	k_{cat} (min^{-1})
FLN	n.d.	n.d.	n.d.
FLC	0.393 ± 0.04	0.383 ± 0.16	0.002
Δ24N	n.d.	n.d.	n.d.
Δ24C	1.380 ± 0.10	1.414 ± 0.27	0.004
Baculosomal CYP3A4*	17.97	3.243	3.594

Table 3.1: Kinetic constants generated from Vivid Red enzyme assays of the CYP3A4 construct proteins. Note: Each data point represents the average of three replicates ± standard deviation. Those reactions displaying negligible turnover are marked 'n.d.'. *Values for baculosomal CYP3A4 obtained from Fig. 3.5 above.

The turnover number of Δ24C-CYP3A4 of Vivid Red measured here was significantly lower than that exhibited by the commercial baculosome CYP3A4 preparation (see Table 3.1). However, the former reaction was carried out using CuOOH-driven electron transfer whereas the latter was carried out using the conventional CPR and NADPH regeneration system. The level of activity observed when testing the baculosomal CYP3A4 with CuOOH was ~18-fold lower than when utilising CPR (see Fig. 3.9), hence a partial explanation of our apparently low k_{cat} values may lie in the electron transfer system used and could be due to heme destruction by CuOOH.

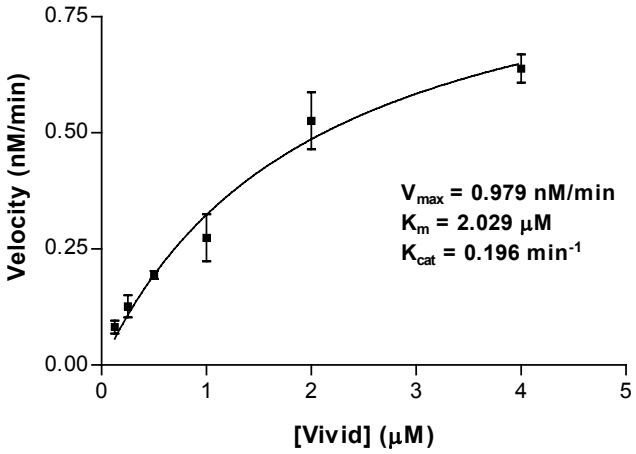


Figure 3.9: Baculosomal CYP3A4 turnover of Vivid Red using CuOOH as an electron donor. Note: Data points are an average of three replicates and error bars represent standard deviations.

Of course it was also possible that only a small fraction of the soluble CYP3A4 proteins expressed here were heme-bound. Despite Δ24C-CYP3A4 displaying the highest soluble expression, N-terminally truncated CYPs are generally known to be more unstable than full-length enzymes and therefore more easily perturbed. For example, truncated CYP2A4 and CYP2E1 were found to convert to P420 during purification (Larson *et al.*, 1991;Sueyoshi *et al.*, 1996). (The presence of a peak at 420 nm in the difference absorbance spectrum of a reduced, carbon-monoxide bound protein preparation is indicative of denatured protein and is termed 'P420'.) The presence of detergents, although often critical to obtain satisfactory yields of protein, is also known to destabilise the CYPs (Fisher *et al.*, 1992;Sandhu *et al.*, 1994).

Indeed, in our experience it was difficult to obtain distinctive CO-binding spectrums for the CYP3A4 constructs, a problem confounded by low soluble expression yields after purification of < 2 mg/L. The data we generated for Δ24C-CYP3A4 showed protein mostly in the P420 form (~7 nmol/mg), with a much lower P450 protein content of ~3 nmol/mg (see Fig. 3.10 below). Purified CYP proteins should ideally approach a specific content of ~20 nmol/mg (Guengerich *et al.*, 1996), which means that while approximately half of the total protein measured here contained heme, only a third of the same protein fraction was correctly folded so as to experience a spectral shift to 450 nm upon binding to carbon monoxide. These results suggested that our Δ24C-CYP3A4 protein preparations contained a significant proportion of denatured protein, as well as other non-heme bound protein (i.e. proteolytic fragments and possibly some impurities). This in turn implies an ~7-fold underestimation of the k_{cat} values obtained in the above experiments using our

recombinant enzyme, which would place the turnover number of Δ24C-CYP3A4.within one fold of magnitude of the commercial baculosomal CYP3A4. Flavin content determination of the various CPR constructs was not carried out, however further spectroscopic characterisation of these constructs would be informative and should be included in future work.

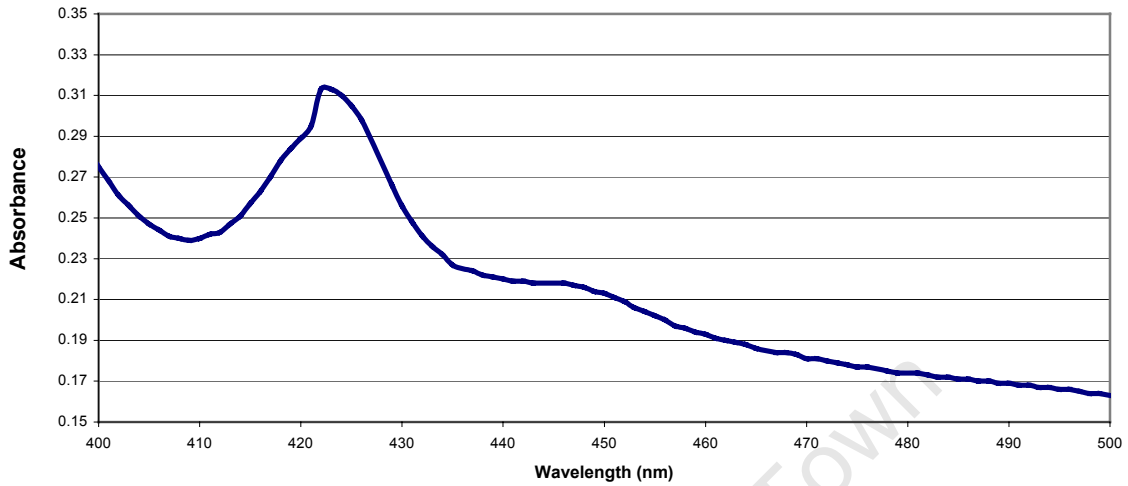


Figure 3.10: A typical Fe^{2+}.CO vs. Fe^{2+} carbon monoxide difference spectrum of purified Δ24C-CYP3A4 protein. $\Delta A_{450\text{-}490}$ = 0.044, representing a CYP heme protein content of 0.48 μM and a specific content of 3 $nmol.mg^{-1}$. $\Delta A_{420\text{-}490}$ = 0.120, representing a denatured heme-bound CYP protein content of 1.08 μM.

Several Δ24-CYP3A4 constructs have been previously reported in the literature. Gillam *et al*. expressed a human Δ24-CYP3A4 deletion construct (without tags) in *E. coli*, and detected no hemoprotein after purification from the membrane fraction (Gillam *et al.*, 1993). Δ24-CYP3A4 deletion constructs have also been used for CYP3A4 crystallisation studies. In the study by Williams *et al*., this construct was expressed with a His_4-tag was added to the C-terminus (Williams *et al.*, 2004). Although no expression or CO-binding data was given in this study, they found no change in the regiospecificity of the truncation mutant for testosterone, nifedipine and midazolam metabolism (no kinetic data was given however). In addition, they reported kinetics for dibenzylfluorescein comparable to commercial full-length enzyme. Yano *et al*. also successfully expressed a Δ23-CYP3A4 deletion construct for crystallisation purposes, however they also did not report expression, activity, or CO-binding data (Yano *et al.*, 2004).

There are several possibilities that could be investigated to increase the activity of Δ24C-CYP3A4. Besides the optimisation of the usual expression and purification parameters, the addition of a known inhibitory ligand during purification could also be tested as this has been shown to stabilise other N-terminally truncated CYPs. For example, the addition of 4-methylpyrazole stabilised truncated CYP2E1 protein by blocking the conversion of P450 to P420, thereby

increasing holoprotein yields (Larson *et al.*, 1991). α-Naphthoflavone was found to have a similar effect on CYP1A2 (Sandhu *et al.*, 1994). However, despite the apparently low activity observed here for the CYP3A4 constructs when using CuOOH in a 96-well plate fluorescence assay, we were satisfied that the activity levels were sufficient for further array construction due to the known sensitivity of fluorescence microarray scanners and the increased sensitivity intrinsic to the small volume reactions used to fabricate protein microarrays (see Chapter Five for further discussion).

3.2.3 Enzyme activity assays of CPR construct proteins

The substrate thiazolyl blue tetrazolium bromide (MTT) was used to assay and compare the functional activity levels of the CPR construct proteins according to the method of Yim *et al*. (Yim *et al.*, 2005). A continuous spectrophotometric assay at A_{610} was used to measure the increase in absorbance resulting from the formation of the blue product, formazon (reduced MTT). Of the CPR constructs, the deletion constructs Δ43N-CPR and Δ43C-CPR exhibited the highest turnover numbers (see Fig. 3.11 and Table 3.2). The FLC-CPR construct exhibited around two-fold less activity compared to the deletion constructs, while the FLN-CPR construct exhibited negligible activity. These results were again in accord with the conclusions drawn in Section 3.2.1 above.

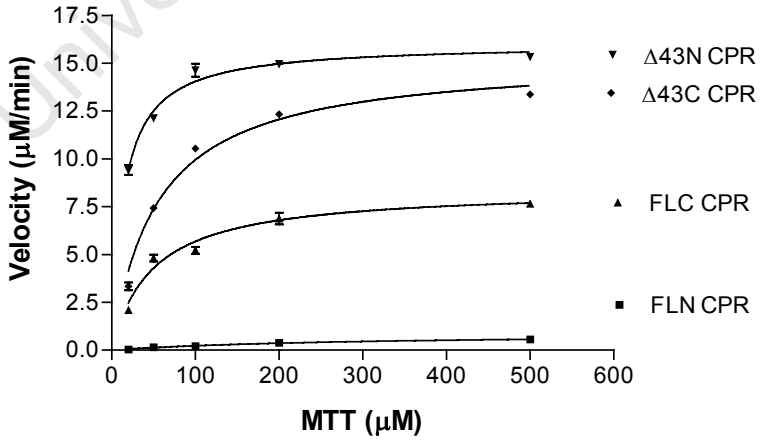


Figure 3.11: Enzyme assays of the CPR construct proteins using MTT reduction. Note: Each data point represents the average of three replicates and the error bars represent standard deviations.

The kinetic data followed classic Michaelis-Menten kinetics and the kinetic constants determined using this kinetic model are shown in Table 3.2 below. Yim *et al*. observed a K_m of 20 μM and a k_{cat} of 1910 min^{-1} for MTT using full-length rat recombinant enzyme purified from *E. coli* membranes

(Yim *et al.*, 2005). The two deletion constructs approach these values with Δ43N-CPR exhibiting a K_m of 14 μM and a k_{cat} of 801 min^{-1}. While the FLC construct was shown to be folded using CD it exhibited lower specific activity levels compared to the deletion constructs, probably due to a lower degree of purity obtained after affinity chromatography purification. The FLN construct exhibited the lowest level of activity as expected from its disordered structure demonstrated using CD.

Construct	V_{max} (μM/min)	K_m (μM)	k_{cat} (min^{-1})	Specific activity (U/mg)
FLN	0.899 ± 0.19	295 ± 120	45	0.52
FLC	8.439 ± 0.37	48.57 ± 7.05	422	4.80
Δ43N	16.020 ± 0.23	14.13 ± 1.16	801	9.68
Δ43C	15.320 ± 0.41	54.09 ± 4.86	766	9.20

Table 3.2: Kinetic constants generated from MTT-based enzyme assays of the CPR construct proteins. Note: Each data point represents the average of three replicates ± standard deviation.

Interestingly, the expression media used for the Yim *et al*. study was supplemented with 10 μg/ml riboflavin to facilitate the incorporation of the flavin cofactors. Riboflavin supplementation was not used in our expression protocol and it is therefore likely that its addition might further increase the specific activity of our purified CPR protein constructs. However the activity levels obtained here were satisfactory for our purposes.

3.2.4 Determining the biotinylation status of the CYP3A4 and CPR constructs

Based on the activity data for the CYP3A4 and CPR constructs, it was clear that the most promising candidates for array construction were Δ43N-CPR, Δ43C-CPR and Δ24C-CYP3A4. Of these, the two C-terminal tagged constructs were simplistically considered to be the mostly likely partner enzyme combination due to the similar location of their BCCP-tags. It was not immediately obvious however that these two constructs would be able to form a functional complex once immobilised *via* their C-terminals on the array surface. This was further investigated using *in silico* modelling as discussed in Chapter Five. However, prior to array construction it was necessary to ensure that the BCCP-tags of these constructs were indeed sufficiently biotinylated *in vivo* during expression in *E. coli*. Western blots of purified Δ24C-CYP3A4 and Δ43C-CPR using streptavidin conjugated to horseradish peroxidase (Strep-HRP) were carried out for this purpose.

The results showed that the Δ24C-CYP3A4 and Δ43C-CPR soluble protein fractions were both biotinylated and strong signals were present for both the crude lysate and pure protein fractions (see Fig. 3.12 E). No proteolytic fragments were observed for Δ43C-CPR, in correlation with the results obtained in Chapter Two when using Anti-His antibody. Δ24C-CYP3A4 results also correlated with previous data with a similar pattern of proteolytic fragments observed for the crude lysate and the purified protein as before.

These results also appear to confirm the idea suggested in Section 2.2.3 that the two closely migrating bands observed for Δ43C-CPR are biotinylated and non-biotinylated Δ43C-CPR protein (rather than biotinylated protein and a proteolytic fragment); only one band was visible on the Strep-HRP blot for Δ43C-CPR (Fig. 3.12 E), while two bands were clearly visible on Coomassie Blue-stained gels (Fig. 3.12 B) as well as on blots detected with Anti-His antibody (Fig. 3.12 D). In contrast, Δ24C-CYP3A4 only displays one band at the expected molecular weight on both the Anti-His blot (Fig. 3.12 C) and the Strep-HRP blot (Fig. 3.12 E) - although several secondary bands were visible on the Coomassie Blue-stained gel (Fig. 3.12 A). The explanation for this difference between the two constructs probably lies in the higher expression of the Δ43C-CPR protein relative to that of Δ24C-CYP3A4 leading to the incomplete biotinylation of the former, as suggested previously.

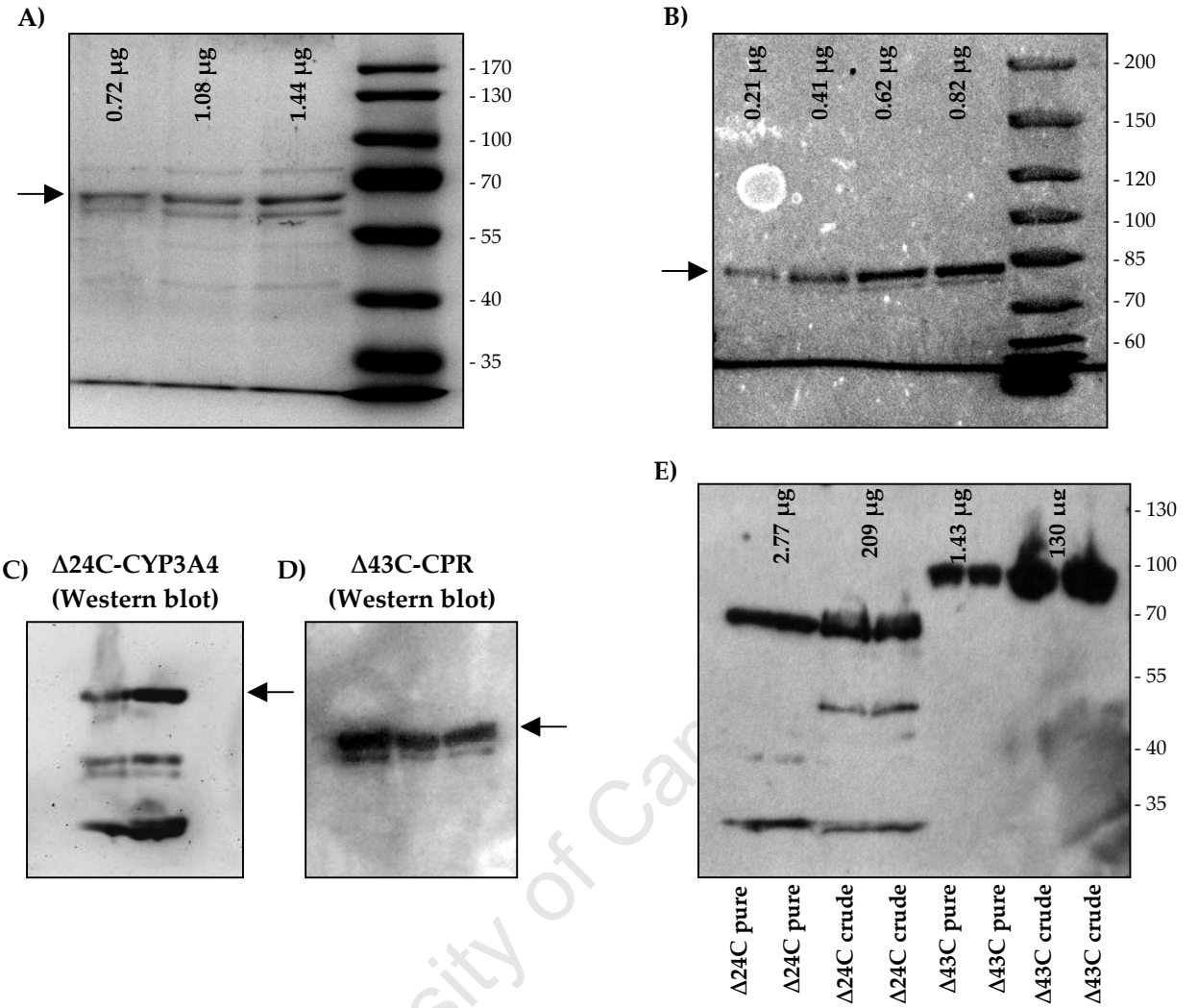


Figure 3.12: Protein gels of Δ24C-CYP3A4 and Δ43C-CPR protein using various detection methods. A) Coomassie Blue-stained gel of HisLink purified Δ24C-CYP3A4 protein. B) As for (A) except using Δ43C-CPR protein. C) Western blot of Δ24C-CYP3A4 protein using Anti-His Ab detection*. D) As for (C) except using Δ43C-CYP3A4 protein. E) Western blot of Δ24C-CYP3A4 and Δ43C-CPR protein using Strep-HRP detection. *Same gel described in Fig. 2.7 D.

3.3 CONCLUSIONS

C-terminal tag placement and the truncation of the N-terminal hydrophobic domains were shown to not only favour soluble protein expression (as shown in Chapter Two), but also protein folding and enzyme activity levels of both CYP3A4 and CPR construct proteins. The effect of tag placement was however more pronounced for the CYP3A4 constructs, the possible reasons for which were discussed above.

The combined expression, CD and activity data for the CYP3A4 and CPR constructs provided the necessary information for choosing those constructs that offered the most promise for future

array work. Both Δ43N-CPR and Δ43C-CPR offered sufficient soluble expression and activity levels, while only Δ24C-CYP3A4 was satisfactory out of the four CYP3A4 constructs. Due to the similar orientation of their BCCP-tags the two C-terminal tagged constructs were chosen for further investigation. Both of these constructs were found to be biotinylated *in vivo*, thereby ensuring that immobilisation onto the array surface would be possible.

3.4 MATERIALS AND METHODS

3.4.1 Circular dichroism investigation of CYP3A4 and CPR construct proteins

Proteins were expressed and purified (using HisLink resin) as described in Sections 2.4.2.1-3 of Materials and Methods, Chapter Two. The purified proteins were concentrated and buffer exchanged (using Centriprep and Zeba Spin columns respectively) into 20 mM potassium phosphate buffer (pH 7) that had been autoclaved and filter sterilised. A 0.1 mg/ml BSA solution was prepared in the same buffer. Circular dichroism spectra of each sample were measured using a Jasco J-810 spectropolarimeter (Jasco Co., Japan) in the far-UV range (188 - 250 nm) using a quartz cuvette with a 0.1 cm path length. Scans were carried out in continuous mode at 100 nm/min, a bandwidth of 1 nm and data increments of 0.1 nm. A baseline was set up using dH_20. Ten scans were performed for each sample including the potassium phosphate buffer and the ellipticity data were averaged to generate the final spectra. The average buffer reading at each wavelength was subtracted from that of the average sample readings. Readings were truncated at lower wavelengths (<189 nm) when the total voltage exceeded the recommended maximum of 600 V. Protein concentrations were estimated using the Bradford's assay and were used to normalise the ellipticity data. Final sample protein concentrations ranged between 0.5 - 3μM, however in some cases dilutions were performed in order to achieve scans below 190 nm. Data was reported in units of molar ellipticity (Θ), expressed in $deg.cm^2.dmol^{-1}$. These experiments were carried out without prior heme or flavin quantification.

3.4.2 Enzyme activity assays of CYP3A4 construct proteins

3.4.2.1 Enzyme activity assay of commercial baculosomal CYP3A4 protein as a positive control

The Vivid BOMR substrate (Vivid Red; #P2865) and the commercial baculosome preparations co-expressing human CYP3A4 and CPR (#P2377) were both obtained from Invitrogen. The resorufin sodium salt standard was obtained from Sigma Aldrich (#R3257), as was the NADPH

regeneration system. The Vivid Red substrate was dissolved in 100% acetonitrile forming a stock solution of 2 mM that was divided into aliquots and stored at -20°C. A fresh stock of the resorufin sodium salt standard was prepared on the day of each assay by dissolving the salt in 100% acetonitrile to form a 10 mM solution. A 29 nM – 2.5 μM resorufin dilution series was prepared using 100 mM potassium phosphate buffer (pH 7.4). Baculosomal CYP3A4 metabolism of Vivid Red was assayed using a fluorescence plate reader with a 10 nm slit width, λ_{ex} of 530 nm and λ_{em} of 585 nm. Readings were taken every 2.5 min over a 35-minute time period. Assay mixtures of 200 μl each contained final concentrations of 5 nM CYP3A4 enzyme, 0.3 – 38.4 μM Vivid Red, and an NADPH regeneration system (100 μM $NADP^+$, 3.3 mM glucose 6-phosphate and 0.4 U/ml glucose 6-phosphate dehydrogenase) in 100 mM potassium phosphate buffer (pH 7.4), according to manufacturer's recommendations (note that Invitrogen supplied no information as to the concentration of the CPR enzyme in the baculosomal enzyme preparation). 100 μl enzyme pre-mixes consisting of enzyme, NADPH regeneration system and buffer were first aliquoted into each well of a Nunc F96 Microwell plate (#436111, Nunc, Thermo Scientific) and pre-incubated at RT for 5 min. 100 μl of the substrate and $NADP^+$ pre-mixes were then added to each well to initiate the reactions and the plate was rapidly shaken on a plate vortex prior to starting the assay readings. Three replicates were carried out per sample and negative control (enzyme and substrate only). No appreciable increase in signal was observed for the negative controls relative to the overall activity.

Data was imported into GraphPad Prism 3.02 software (GraphPad Software Inc., USA). Averaged relative fluorescence unit (RFU) values at each time point were used to plot the reaction velocities for each substrate concentration using linear regression. These velocities were then converted from $RFU.min^{-1}$ to resorufin $(\mu M).min^{-1}$ using the standard curve generated from the resorufin dilution series. The velocity data was then analysed using non-linear regression and Michaelis-Menten kinetics.

3.4.2.2 Testing Vivid Red turnover using the Δ24C-CYP3A4 construct and cumene hydroperoxide

Δ24C-CYP3A4 protein was expressed and purified using HisLink resin as described in Sections 2.4.2.1-3 of Materials and Methods, Chapter Two. Cumene hydroperoxide was obtained from Sigma Aldrich and a fresh stock solution of 1 mM in 100 mM potassium phosphate buffer (pH 7.4) was prepared on the day of each assay. Δ24C-CYP3A4 metabolism of Vivid Red was assayed as described in Section 3.4.2.1 above with the following amendments. Readings were taken every

minute over a 5 min time period, thereafter a reading was taken every 10 min for a total assay time period of 1 hr to determine the linear reaction period. Assay mixtures of 200 µl each contained final concentrations of 250 nM Δ24C-CYP3A4 enzyme, 2.5 or 25 µM Vivid Red, and either 0.05, 0.5 or 5 mM final concentrations of CuOOH in 100 mM potassium phosphate buffer (pH 7.4). 100 µl pre-mixes consisting of enzyme and substrate only were first aliquoted into each plate well and pre-incubated at RT for 5 min. 100 µl of the 1 mM CuOOH stock solution was then added to each well to initiate the reactions and the plate was rapidly shaken on a plate vortex prior to starting the assay readings. Three replicates were carried out per sample and negative control (controls were enzyme and substrate only and substrate and CuOOH only). Negative controls were observed to yield a slight increase over time of ~1% of the total activity. A similar protocol was used to test CuOOH concentrations between 0.1 – 1 mM, except that 4 µM Vivid Red was used and data was generated for two different enzyme batches.

3.4.2.3 Enzyme activity assays of the CYP3A4 construct proteins

Turnover of Vivid Red by each CYP3A4 construct was assayed as described in Sections 3.4.2.1-2 above with the following amendments. Readings were taken every minute over a 20 min time period. Assay mixtures of 200 µl each contained final concentrations of 250 nM CYP3A4 enzyme, 0.125 - 6 µM Vivid Red, and 0.5 mM final concentrations of CuOOH in 100 mM potassium phosphate buffer (pH 7.4). A 3.13 – 100 nM resorufin dilution series was used to generate the standard curve. The CuOOH-based baculosomal assay was carried out similarly except with 5 nM of commercial baculosomal CYP3A4.

Carbon monoxide difference spectra of CYP3A4 were obtained following the methods of Omura and Sato (Omura and Sato, 1964b). Purified protein was diluted into 50 mM sodium phosphate buffer (pH 7.3) including 20% glycerol (v/v). Glycerol is often used in such assays to help stabilise the enzyme during the assay. Proteins were reduced by the addition of an equimolar concentration of sodium dithionite (Sigma Aldrich), using a 0.5 mM solution in the above buffer. The spectra were measured at 2 min intervals until no further reduction was observed (i.e. until the 420 nm peak remained constant). Carbon monoxide was then bubbled through the solutions for 30 sec and the resulting spectra were measured at 2 min intervals until no further increase in the 450 nm peak was observed. Each reduced protein sample spectrum was then subtracted from the relevant CO-bound reduced protein spectrum to obtain the final difference spectrum.

Denatured heme-bound CYP protein content was calculated using $\varepsilon_{420\text{-}490}$ = 111 $mM^{-1}.cm^{-1}$, while catalytically active heme-bound protein content was calculated using $\varepsilon_{450\text{-}490}$ = 91 $mM^{-1}.cm^{-1}$.

3.4.3 Enzyme activity assays of CPR construct proteins

Proteins were expressed and purified using HisLink resin as described in Sections 2.4.2.1-3 of Materials and Methods, Chapter Two. The substrate used was 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) obtained from Fluka (#88415, Sigma Aldrich). A fresh stock of 10 mM MTT was prepared in 100 mM potassium phosphate buffer (pH 7.6) on the day of each assay. CPR construct enzyme function was tested using a continuous spectrophotometric assay at A_{610} of MTT reduction over a one-minute time period according to the method of Yim *et al*. (Yim *et al*., 2005). Assay mixtures of 1 ml each contained 20 nM CPR enzyme, 0.1 – 500 μM MTT, and an NADPH regeneration system (0.5 mM NADP+, 5 mM glucose 6-phosphate and 0.5 U/ml glucose 6-phosphate dehydrogenase). Negative controls at each substrate concentration consisting of enzyme and substrate alone were included. Three replicates were carried out per sample. No enzyme activity was observed for the negative controls nor for the 0.1 – 5 μM MTT sample reactions and the latter results were therefore excluded from analysis. Data was imported into GraphPad Prism 3.02 software. Averaged absorbance values were used to plot the reaction velocities for each substrate concentration using linear regression. Velocities were then converted from $\Delta A_{610}.min^{-1}$ to formazon (μM).min^{-1} using an extinction coefficient of 11.3 $mM^{-1}.cm^{-1}$ for reduced MTT. The velocity data was then analysed using non-linear regression and Michaelis-Menten kinetics.

3.4.4 Determining the biotinylation status of the CYP3A4 and CPR constructs

Proteins were expressed and purified using HisLink resin as described in Sections 2.4.2.1-3 of Materials and Methods, Chapter Two. SDS-PAGE and Western blot transfer were carried out as described in Section 2.4.2.5. Western blot transfer for determining biotinylation used 0.5 μg/ml streptavidin-HRP conjugate (#14-30-00; KPL, USA).

CHAPTER FOUR

Overview of Protein Array Technology

4.1 INTRODUCTION

4.1.1 Overview

The completion of the human genome sequencing project has provided the impetus for unravelling one of the greatest challenges in the biological sciences – understanding the full complexity of cells and tissues in terms of their intra- and inter-cellular pathways (Wilson and Nock, 2003). Until quite recently this goal was approached at the individual gene or protein level. However, due to the complex interdependence of biological systems, such studies cannot provide a global view of what is occurring at the cellular level. Researchers are therefore constantly developing tools that enable the study of the interactions of many genes and proteins in parallel under different experimental conditions. Miniaturised and parallel assay systems are crucial to such large-scale and high-throughput biological analyses, as they are faster and more economical (Zhu and Snyder, 2003).

The development of DNA microarrays and cost-effective, high-throughput genome sequencing is addressing this need in genomic research. However, DNA microarrays cannot be used to answer all research questions, as transcript levels seldom correlate with levels of protein expression and/or activity (Kumble, 2003;Templin *et al.*, 2003). This is mostly due to differential stability of transcripts, alternative splicing, as well as a plethora of post-translational modifications such as phosphorylation, glycosylation, lipidation and cleavage. In contrast to the approximately 20,000 – 25,000 genes comprising the human genome, 10 - 100-fold more protein forms are thought to exist owing to the downstream processing of transcripts and the proteins they encode (Banks and Selby, 2003). Since proteins carry out most of the functions in the cell and are the targets for the vast majority of drugs, it is crucial that we complement knowledge of gene expression at the transcript level with information of expression and function at the protein level.

Bridging this gap is the relatively new field of proteomics (Takahashi *et al.*, 2003). Proteomics is defined as the study of the 'abundance, modification, activity, localisation and interaction of all proteins' in a cell, tissue or whole organism. Proteomics tools currently available for the global analysis of protein levels and activities include two-dimensional (2D) gel electrophoresis, yeast two-hybrid screening, chromatography and mass spectrometry (MS). These technologies traditionally require large sample volumes and are technically difficult to adapt to a high-throughput format (Kumble, 2003).

Within the last decade, protein microarray technology has emerged as a potentially powerful proteomics tool that enables high-throughput and parallel analysis of hundreds to thousands of proteins in a single experiment in a time- and cost-effective manner (Hartmann *et al.*, 2009). Protein microarrays are highly miniaturised assay systems where reactions take place in spots with diameters of less than 300 μm on a solid surface (typically measuring 75 mm x 25 mm), with densities of a >2000 spots/cm^2 possible (Wingren and Borrebaeck, 2007). Each spot can be regarded as a miniature reaction vessel with a volume in the picolitre to nanolitre range, and assay sensitivities as high as pM and fM levels are achievable. Protein arrays are scalable and automatible, and can rapidly provide quantitative, reproducible data of numerous parameters in a single experiment, in contrast to the few parameters allowed by traditional methods (Kumble, 2003;Zhu and Snyder, 2003). Considering that each spot is an individual assay, microarray platforms have the advantage of inherent self-consistency as all experiments can be carried out under essentially identical reaction conditions (Diaz-Mochón *et al.*, 2007). The fact that protein arrays utilise very low sample volumes (typically a few nanolitres), is another significant benefit, as protein samples are often very limited in supply and, unlike nucleic acids, cannot be amplified.

Protein microarrays are well-established research tools in basic and applied research, with several commercial products already available (Hartmann *et al.*, 2009;Templin *et al.*, 2003). Protein arrays have also shown great potential in areas such as clinical diagnostics and drug discovery and are predicted to become one of the most powerful tools in biological research as they can be used to analyse protein-protein, antibody-antigen, protein-lipid, protein-nucleic acid, protein-small molecule, protein-peptide, protein-oligosaccharide and enzyme-substrate interactions (Zhu and Snyder, 2003). In a single experiment, one can now for example determine all the substrates of a protein kinase, elucidate a cell-signalling network, or detect all the binding partners of a potential drug.

4.1.2 Current clinical applications

Protein arrays have already evolved into tools that deliver clinical data of high diagnostic and prognostic value, with a variety of protein array platforms already cleared for use in the US and the EU [see Table 4.1; (Hartmann *et al.*, 2009)]. This trend looks set to continue, especially since it is increasingly clear that while individual proteins typically do not make good biomarkers, multiplexed, multi-analyte assays of biosignatures may yet do so.

Commercial array manufacturers generally supply customers with ready-made arrays, while the customers carry out their own assays, data acquisition and analysis. Although multiplexed genetics tests currently dominate the diagnostics market, protein tests are rapidly catching up. The majority of current, FDA-approved protein array tests use antibodies and serum samples to diagnose allergies, autoimmune and infectious diseases. The two most popular platforms for these tests are the bead-based array platform from Luminex (USA, www.luminexcorp.com) and the Evidence automated biochip platform from Randox Laboratories (UK, www.randox.com), with several of their assays approved or currently being approved by the FDA.

Platform	Manufacturer	Description
Bioplex 2200 Analyser together with the AtheNA Multi-lyte ANA Test System	Bio-Rad Laboratories, Zeus Scientific	Bead-based suspension microarray for the semi-quantitative assessment of a panel of clinically relevant autoimmune IgG antibodies in human serum samples. This platform integrates the only FDA-approved algorithm for a protein array platform, which correlates patient antibody profiles with a database of autoimmune disease-associated profiles.
DynaChip Antibody Analysis System	Invitrogen	Automated chip-based system for the detection of human leukocyte antigen (HLA) antibodies for organ transplant donor-recipient matching
Evidence BioChip Array Technology	Randox Laboratories	Automated biochip system for multi-analyte testing. Applications already developed include cardiac, cytokine, drugs of abuse, tumour monitoring, and fertility assays.

Table 4.1: Some of the FDA-approved protein microarray platforms currently marketed and used for clinical diagnostics purposes. [Adapted from (Hartmann *et al.*, 2009).]

However, despite the huge diagnostic potential of protein arrays, they are currently far from being widely used for clinical research or drug discovery (Hartmann *et al.*, 2009). One factor contributing to this is the lack of researchers skilled in protein array production and analysis and the many technical challenges involved. Another factor is that protein array assays are mostly performed manually, but routine automation is required in order to enter the diagnostics market. Nevertheless, medical demand combined with overall cost reduction will likely become the driving force behind protein arrays gaining a substantial share of the *in vitro* diagnostics market, especially in cases where sample volume is limited.

4.1.3 History

Before the advent of protein arrays, proteins had been routinely immobilised onto a variety of surfaces in microtitre plate format for enzyme-linked immunosorbent assays (ELISA) for over three decades (Chen *et al.*, 2003). However, to apply ELISAs on a large scale requires large sample volumes and is hugely laborious without access to expensive robotic systems, putting this beyond the reach of most laboratories.

In the 1960s, the first miniaturised immunoassay was developed by Feinberg's group for detecting autoimmune diseases (Feinberg, 1961;Feinberg and Wheeler, 1963). Auto-antigens were immobilised onto a surface in microspots, and the surface was then incubated with serum samples from patients. They concluded that this format had several advantages over the more traditional methods, however it garnered little interest at the time.

In 1989, Roger Ekins published his ambient analyte theory on the feasibility of microspot-based immunoassays (Ekins, 1989;Templin *et al.*, 2002). According to his theory, miniaturisation is not only feasible - it actually leads to an increase in detection sensitivity. While the total signal for a spot will increase with increasing spot size (with a maximum signal reached when most targets are captured from the solution), the signal *density* increases with *decreasing* spot sizes, reaching a constant level when the captured molecule concentration is $<0.1/K$ (where K is the association constant for the analyte). Under these ambient analyte conditions, i.e. where the numbers of captured molecules are much lower than the number of molecules in the analyte solution, the target concentration in solution is only minimally altered by the amount of captured targets in the microspot. The signal from the microspot (which is a function of the fractional occupancy of the total number of capture sites available in the spot) then becomes volume-independent and can be used to directly calculate the concentration of the target in the analyte solution. This holds true even when target molecules display high affinity for the capture molecules on the surface and the analyte concentration is low. Higher sensitivity is achieved because the capture molecule-target complexes are concentrated into a small area in the microspot, resulting in a much higher signal intensity compared to a macrospot. This increased sensitivity allows the detection of concentrations down to the femtomolar range.

In an influential publication in 1991, Fodor *et al.* described the use of photolithographic techniques for fabricating the first high-density, spatially addressable peptide arrays (Fodor *et al.*,

1991). Photolithography is a technique used routinely in the semiconductor industry to manufacture computer chips. Using this approach, 1024 different peptides were synthesised *in situ* and then incubated with an antibody to determine specific antibody-interacting peptide sequences using standard immuno-detection methods.

Fodor and colleagues then extended this approach to the construction of the first prototypical DNA microarray in 1994, which consisted of oligonucleotides photolithographically synthesised on-chip as a tool for DNA sequence analysis (Pease *et al.*, 1994). This would later become the core technology of Affymetrix's GeneChip platform. In 1995, Brown and co-workers at Stanford University created the first DNA microarray using an automatic robotic spotter (Schena *et al.*, 1995). The next few years witnessed an exponential increase in the application of DNA microarray technology (Chen *et al.*, 2003;MacBeath, 2002). It is therefore of note that although arrays of peptides and other biopolymers preceded the DNA microarray, it is only in the last nine years that protein-, peptide- and small-molecule arrays have become popular tools - whereas DNA arrays have been in use for over 15 years and have been widely adopted. In fact, it is the field of genomics and the success of DNA microarray technology that has been the main driving force behind the development and adoption of protein arrays, with much of the technology simply being adapted for the purpose. It was only once the benefits of DNA microarray technology were well recognised and the equipment to manufacture these arrays became more accessible that attention shifted towards creating antibody arrays for proteomic research.

In 2000, MacBeath and Schreiber constructed one of the first protein arrays on a glass microscope slide using the robotic DNA microarray technology developed by the Brown group (MacBeath and Schreiber, 2000). They demonstrated that antibody-antigen interactions could be detected with similar sensitivity to that of conventional ELISA and Western Blot assays; protein-small molecule interactions in the micromolar range could be detected; and protein kinase activity could also be assayed. In the same year, Zhu and colleagues analysed the substrate specificities of 119 yeast protein kinases using 17 different substrates immobilised in nanowell microarrays (Zhu *et al.*, 2000). Not only were known kinase-substrate interactions identified but many novel activities were also revealed, including the ability of Ser-Thr protein kinases to phosphorylate tyrosine residues.

These seminal articles clearly demonstrated the vast potential of the technology, yet the adoption of protein arrays has still been much slower than that of DNA arrays. This is mainly due to the construction of protein arrays being technically more challenging and more complex in nature than that of nucleic acid arrays (Hartmann *et al.*, 2009). However, more recent developments arising from nanotechnology are now adding further momentum to the field (Wingren and Borrebaeck, 2007).

4.1.4 Challenges

When constructing an array of proteins rather than DNA, there are a variety of additional challenges that arise due to the more complex nature of proteins [see Table 4.2; (Hartmann *et al.*, 2009)]. DNA consists of combinations of only four different nucleic acids with the same hydrophilic backbone, while proteins consist of combinations of 20 different amino acids that vary in the size and nature of their side chains. DNA arrays consist of linear, single-stranded molecules that are quite robust, while the probes and/or analytes of protein arrays have three-dimensional folded structures that are sensitive to perturbations. Unlike the stable, one-to-one base pairing interaction typically observed for DNA binding, proteins bind to analytes through diverse interactions (including hydrogen bonding, electrostatic forces and hydrophobic Van der Waals interactions) that are easily effected by experimental conditions. While DNA polymers of similar length have similar physical and chemical properties, this is not true for proteins of similar size or shape.

Properties	DNA	Proteins
Structure	Uniform Hydrophilic acidic backbone Stable	Varies Hydrophobic and/or hydrophilic domains Generally unstable
Functional state	No loss of activity when denatured (can be stored dry) Activity not sensitive to perturbations	3D structure critical for activity, denaturation has to be avoided Activity sensitive to perturbations
Interaction sites	One-to-one interaction	Multiple interaction sites possible
Interaction affinity	High	Dependent on protein/s involved: very low to high
Interaction specificity	High	Dependent on protein/s involved: very low to high
Activity prediction	Easily predicted from primary nucleotide sequence	Many attempts but not possible yet with accuracy or confidence
Amplification	PCR	None

Table 4.2: Comparison of the properties of DNA and proteins important to their application in microarray fabrication. [Adapted from (Templin *et al.*, 2003).]

The tertiary, three-dimensional structure of proteins is critical for their activity. Some proteins rely not only on this tertiary structure but also on quaternary structure in order to carry out their functions (e.g. haemoglobin). This higher-order structure has to be maintained throughout the processes of printing, immobilisation, assaying and slide processing if protein arrays are to be anything but a collection of unfolded, immobilised polypeptides. This poses a major challenge, as proteins not only vary in shape and size, but also in the optimal conditions required to maintain their activity and stability. These conditions include factors such as pH, salt concentrations, detergents, and reducing agents. Many proteins also require various post-translational modifications such as cleavage, phosphorylation or glycosylation for activity.

An added difficulty in constructing protein microarrays is that unlike the easy amplification of DNA molecules by PCR, the direct amplification of proteins is not possible. This leads to the challenge of having to obtain suitable levels of protein from extraction, or, more commonly, through heterologous expression and subsequent purification. This can prove difficult, especially when looking at whole-proteome analysis, as proteins differ widely in the ease with which they can be expressed and purified. A further challenge is that high affinity and specificity capture molecules are easily predicted and generated for DNA analytes, however designing protein capture molecules is far from easy as such interactions are more complex and generally less well understood.

Laboratories using protein microarrays for their research therefore tend to spend a significant period of time on developing the technology and testing various arrayers, printing methodologies, surface chemistries, and other optimisations, before using these arrays for experimental purposes. These various aspects of protein array construction and application are discussed below.

4.2 PROTEIN ARRAY CONSTRUCTION

4.2.1 Array surface

The final performance of a protein array strongly depends on the surface material and the immobilisation strategy utilised, as proteins are prone to partial to complete denaturation upon immobilisation on array surfaces (Jonkheijm *et al.*, 2008). One of the main challenges therefore in constructing functional protein arrays lies in choosing an array surface chemistry that maintains

the native conformation and biological function of the proteins of interest upon their immobilisation. In addition, the chemical and physical properties of the surface will also influence the binding, orientation and density of target proteins, as well as the non-specific binding of target and non-target proteins. These factors in turn influence the signal-to-noise ratio produced (i.e. the sensitivity of the array and the limit of detection attainable by the chip). There is therefore increasing interest in developing different methods of surface activation and functionalisation to meet the needs of the various applications of protein arrays.

Proteins can either be arrayed on flat solid surfaces or on chips modified to carry pads, films, nanowells or microfluidic channels (Phizicky *et al.*, 2003). The selection of the surface and the immobilisation strategy depends on the nature of the capture and analyte molecules as well as the application for which the protein array will be used (Jonkheijm *et al.*, 2008). For example, gold is often used for biosensors with an electrochemical or surface plasmon resonance (SPR) read-out due to its excellent electrical conductivity and convenient functionalisation through thiol chemisorption. Glass and silicon are often used for optical sensors, due to the transparency of the former and the low intrinsic fluorescence of both. Ideally, surfaces must be uniformly and densely derivatised with stable, controllable surface properties, modifiable with a wide range of chemical functionalities and be as flat as possible. Surfaces should also have high protein-binding capacity, retain protein function after immobilisation, be easy to manufacture with minimum intra- and inter-batch variability, produce high signal-noise ratios, and have a reasonably long shelf-life (Kumble, 2003).

Below follows a brief description of the various surfaces used for constructing protein arrays, and the methods by which these are most often functionalised. The next section deals with the various immobilisation strategies used.

Glass:

The main method for functionalising glass slides makes use of the reactive silanol groups (Si-OH) on the glass surface (Jonkheijm *et al.*, 2008;Tao *et al.*, 2007). The silanol groups can be generated by pre-treating the glass with, for example, piranha solution (a mixture of H_2O_2 and H_2SO_4), or oxy plasma. The surface is then salinised to introduce a new functional group for subsequent immobilisation steps. A bifunctional silane cross-linker is commonly employed, which forms a self-assembled monolayer (SAM) with one functional group that reacts with the hydroxyl groups

on the glass surface, and another free group (such as an aldehyde, epoxy or N-hydroxysuccinimidyl ester (NHS) group) that can either react directly with primary amine groups of proteins or can be further chemically modified to reach maximum specificity. There are a large variety of these silane reagents commercially available. Saline reagents are deposited onto glass slides either by vapour deposition or by dipping into solution; both methods often result in an uneven coating giving rise to significant micro-heterogeneity across the slide surface and high intra- and inter-slide variability (Kumble, 2003).

Silicon:

Silicon has high electrical conductivity, high chemical resistance to solvents, good mechanical stability, and low intrinsic fluorescence (Jonkheijm *et al.*, 2008). However, the chemical functionalisation of silicon surfaces is complicated by the fact that silicon spontaneously oxidises in air to produce an amorphous silica layer. These surfaces therefore require pre-treatment, firstly to remove this oxide layer and secondly to activate the silicon surface for subsequent reaction with functional moieties. Alternatively, silicon can be oxidised with plasma (ionised gas), which produces silanols on the surface. Functionalisation with organosilanes can then be carried out as for glass slides.

Gold:

Gold has the advantage in that it is easily functionalised with self-assembling monolayers of thiols, disulfides and sulfides (Jonkheijm *et al.*, 2008;Zhu and Snyder, 2003). Bifunctional thio-alkanes are normally used, which have a SH-group that reacts with gold and another free group that reacts with capture molecules. These cross-linkers are dissolved in pure solvent and incubated on the gold surface upon which they slowly but spontaneously assemble to form a monolayer. The generation of SAMs strongly depends on the crystalline morphology of the underlying surface metal. Au(III) yields SAMs with the highest density and degree of regularity and is therefore most widely used. Gold films are usually applied onto polished glass, silicon or freshly cleaved mica. The thermal deposition of gold onto silicon wafers is the most feasible method, with an intermediate layer of chromium or titanium used to promote the adhesion of gold to the silicon.

The chemistry of the gold-thiol interface is well known and much easier to control than the organosilane chemistry used for glass and silicon surfaces. However, there are drawbacks to

using thiolates, including their mobility on solid surfaces (which limits the lifetime of such chips), as well as their susceptibility to photooxidation. The main advantage of using gold surfaces is that SPR and MS can then be used as detection methods to monitor reactions in real time or to identify captured molecules respectively.

Polymer surfaces:

These are a valuable alternative to inorganic surface materials, as they are inexpensive to produce and are becoming available for implementation for standard microarray use, e.g. in microfluidic chips which are usually manufactured with polymeric material such as poly(dimethylsiloxane) (PDMS), poly(methyl methacrylate) and polycarbonate (Jonkheijm *et al.*, 2008). For protein immobilisation, surface modification of polymer surfaces is required, as they have no suitable functional groups in native form. For example, PDMS can be treated by plasma oxidation, which then allows functionalisation with organosilanes. Binkert *et al*. recently achieved microwell densities of 1000/cm^2 using PDMS to construct picolitre-volume chambers (Binkert *et al.*, 2009). Another alternative includes polystyrene 96-well plates that are routinely used for conventional ELISA assays (Kumble, 2003). Large array spots are printed into each well and up to 96 different assays can then be carried out in this 'array of arrays' format. This is useful when a small number of different proteins or antibodies are being used to analyse large numbers of samples.

Chips with 3D matrices:

In contrast to the two-dimensional surface materials mentioned above, materials such as membranes and hydrogels have a porous matrix that allows molecules to diffuse in 3D space (Jonkheijm *et al.*, 2008;Zhu and Snyder, 2003). These matrices therefore show a higher capacity for protein binding and provide a more 'natural' aqueous environment (70-95% water) than that offered by flat surfaces, thus helping to prevent the denaturation of proteins. An alternative to directly functionalising glass slides is therefore to coat the slides with a polymeric hydrogel or a membrane.

Polymeric hydrogels are hydrophilic matrices with up to 100-fold more capacity for immobilisation than flat surfaces (Jonkheijm *et al.*, 2008). These gels generate stable surface chemistries as they are capable of covalent attachment to solid surfaces (e.g. agarose and acrylamide can be photopolymerised onto a surface functionalised with acryl groups). The polymer is subsequently activated with hydrazine or ethylenediamine to generate amine-reactive

groups on the surface for protein immobilisation. Other examples of polymeric gel surfaces include polysaccharides such as chitosan or dextran. Hydrogels can also be created from aldehyde- or lysine-functionalised poly(ethylene glycol) (PEG) spotted onto aldehyde-coated glass slides. Sol-gel matrices (proteins encapsulated within an inorganic silicate matrix formed by polymeric SiO_2 networks) are yet another 3D matrix and are often preferred for their optical transparency, chemical inertness and robustness (Gill and Ballesteros, 2000;Rupcich *et al.*, 2003). Their porosity can also be controlled relatively precisely. In a recent study by Tanaka *et al*. a novel method of co-polymerising the protein and the polymer reagent on the array surface was used to create a protein-based hydrogel that yielded superior sensitivity compared to older methods (Tanaka *et al.*, 2009).

A simpler method to make a 3D array surface is to simply place a membrane on top of a solid support, such as a glass slide (Jonkheijm *et al.*, 2008;Tao *et al.*, 2007). Membrane materials most commonly used are nitrocellulose and nylon, with the latter offering better physical strength, binding capacity and stable protein attachment. Nylon interacts with proteins through positive and negative electrostatic interactions, while nitrocellulose is thought to bind proteins through both hydrophobic and electrostatic interactions. Proteins can also be immobilised onto nylon using photocrosslinking. Casting these membranes onto glass surfaces enhances their stability and improves their spot resolution from the millimetre to the micrometer range, thereby making them suitable for array applications. Such array surfaces are commercially available.

In addition to the sophisticated processes required to create hydrogel matrices, one of the major disadvantages of these 3D array surfaces is that it is more difficult to exchange buffers and recover trapped molecules from them compared to 2D array surfaces (Jonkheijm *et al.*, 2008;Tao *et al.*, 2007). Other potential challenges are high background signals, cross-contamination by spot running, and mass transport effects (despite their increased protein capacity, 3D surfaces offer more resistance towards the diffusion of solutions and this can result in decreased signal when compared to a 2D surface). Sol-gels in particular exhibit variability in matrix properties over time, poor adhesion onto solid surfaces and the final structure often cracks due to gel shrinkage (Lebert *et al.*, 2008).

In addition to 3D surfaces generated from polymers, surfaces can also be engineered to contain nanowells or microfluidic channels (Phizicky *et al.*, 2003;Zhu and Snyder, 2003). Nanowell arrays

consist of wells with diameters of < 1 μm made of materials such as moulded PDMS or etched glass. Microfluidic chips consist of channels engineered from glass or silicon with micrometer - nanometer widths in which fluid flow is precisely controlled. Due to the open structure of nanowells and microfluidic channels, different components and buffers can be added sequentially and captured molecules can be recovered easily. In addition, these surface-modified chips reduce evaporation and denaturation during drying, increase protein-binding capacity, and prevent cross-contamination due to physical barriers separating reactions. A disadvantage of this technology however, is that specialized equipment is needed to engineer these surfaces and to load samples onto them.

4.2.2 Protein immobilisation strategies

Depending on the array application, the choice of immobilisation method can be critical (Cha *et al.*, 2005;Jonkheijm *et al.*, 2008;Zhu and Snyder, 2003). Many current methods result in the random attachment of proteins to the slide, which can result in some fraction of the proteins losing activity either through unfolding or through steric blockage of the active site. Although antibody arrays have been shown to retain sufficient binding when using random immobilisation, in most other applications, such random attachment leads to the loss of protein function.

In practice, the amount of signal detected will depend on the amount of capture molecules bound to the surface, which in turn depends on the surface density of the functionalised groups. If this number is below the critical threshold of detection, no signal will be observed. Oriented immobilisation increases the number of proteins free to interact with analyte thereby enabling detection with a lower number of capture molecules, while using a 3D support surface such as a hydrogel further increases this potential for interaction. Surfaces that allow for homogenous surface orientation without effecting protein conformation are therefore highly sought after.

Another important consideration when choosing an immobilisation strategy is whether or not to incorporate a spacer. The direct attachment of proteins onto a surface without a spacer group can impose steric constraints on the protein's reactivity or interaction capabilities. Moreover, multiple contacts with the surface can induce denaturation or partial denaturation and thereby decrease protein activity. Introducing a spacer between the protein and the surface can minimise these effects. Spacers can be of any desired length, rigidity or flexibility, with hydrophilic, hydrophobic,

charged or neutral characteristics. Surfaces functionalised with a spacer can then be further functionalised with any desired reactive group. Some common spacer molecules include PEG, gluteraldehyde, divinylsulfone, disuccinimidyl carbonate, and *N,N'*-carbonyldiimidazole.

Immobilisation strategies are categorised into the following methods: non-directional covalent and non-covalent (i.e. resulting in a random orientation), and directional covalent and non-covalent (i.e. resulting in a uniform orientation).

Random non-covalent:

This type of immobilisation occurs *via* adsorption utilising ionic, electrostatic, hydrophobic and polar interactions (Jonkheijm *et al.*, 2008). Which of these intermolecular forces dominate will depend on the proteins and surfaces used. Examples of such surfaces include nitrocellulose (hydrophobic and electrostatic interactions) and PVDF (hydrophobic interactions). The main advantage of this method is that no surface modification is needed prior to immobilisation. Disadvantages are limited adsorption capacity, potential dissociation of weakly bound proteins, loss of protein activity (for reasons mentioned above), as well as high non-specific binding.

Random covalent:

Proteins generally offer many functional groups suitable for immobilisation, mainly in their amino acid side chains, which can be used to covalently attach proteins to surfaces through a range of chemical reactions (Jonkheijm *et al.*, 2008). Various chemically modified surfaces are commercially available for these purposes, some designed to resist non-specific adsorption. Examples are surfaces coated with NHS ester, aldehyde, isothiocyanate, and epoxide (all of which react with the amine groups on proteins); or amine (which reacts with carboxylic acid groups). It is important to note that protein attachment can occur simultaneously through many residues thereby restricting the degree of conformational freedom of the protein and thus possibly its activity.

Of these methods, the most widely used is that of covalently attaching proteins to amine-reactive surfaces using the amine group of lysine side chains, with aldehyde-derivatised glass slides being extensively used for this purpose. However, the high abundance of these groups (>10%) can lead to loss of activity. Surface-bound NHS-activated carboxylic acids are also commonly used for coupling with protein amine groups, thereby forming stable peptide bonds.

A new, emerging immobilisation method is that of using direct photochemical attachment of proteins onto solid supports using photolithography technology [see (Shin *et al.*, 2003) for example]. After light activation, the photoreagents attached to the chip surface (e.g. arylazide or diazirine) undergo a distinct chemical transformation that leads to the formation of covalent bonds between photo-generated intermediates and proteins on the surface. This method is currently subject to limitations such as reaction conditions that possibly degrade or denature proteins. The advantages are that proteins can be printed without prior chemical modification or surface functionalisation, and protein immobilisation in specific patterns is possible at high resolution.

Uniform non-covalent:

Ideally, attachment of proteins should occur in a non-random, oriented manner. This can be accomplished using specialised surfaces such as nickel-coated glass for capturing hexahistidine-tagged proteins (Jonkheijm *et al.*, 2008;Kumble, 2003). These affinity tags are fused at either the N- or C-terminus of the protein, such that the protein's activity is preserved and all proteins orient uniformly away from the surface.

There are many popular fusion protein tags available such as GST, MBP, FLAG, and His_6, which were originally developed for affinity purification purposes. These allow specific, directed and reversible attachment to solid surfaces coated with the relevant affinity reagents. Reversibility is sometimes attractive as such chips can potentially be recycled: for example, His_6-tagged proteins binding on a Ni-NTA (Nickel-nitrilotriacetate) chip can be reversed with EDTA or imidazole. The low affinity of fusion tags to their affinity partners can however lead to the unwanted dissociation of immobilised proteins during the time frame of downstream experiments.

The specific binding of biotin to avidin and streptavidin is another strategy adopted from affinity chromatography. Streptavidin comprises four identical subunits each of which binds one biotin molecule (Gonzalez *et al.*, 1997). The high binding affinity ($K_d = 10^{-15}$M) observed in the formation of this complex is nearly irreversible, almost comparable in strength to the covalent bond. The biotin-streptavidin bond is formed very rapidly and is unaffected by factors such as pH, temperature, organic solvents, enzymatic proteolysis or other denaturing agents. Streptavidin is usually preferred over avidin since glycoamino acids occurring on avidin can potentially cause unwanted non-specific adsorption (Jonkheijm *et al.*, 2008).

The biotinylation of proteins can be carried out with standard bioconjugation techniques using chemically activated biotin derivatives (e.g. biotin-NHS which is amine-reactive, pentylamine biotin which is carboxylic acid reactive, and biotin BMCC that is thio-reactive) (Jonkheijm *et al.*, 2008). However, to avoid random biotinylation and the subsequent random orientation and potential inactivation of proteins, site-specific labelling of proteins using biotin ligase strategies (Chapman-Smith and Cronan, 1999), or tag-free intein-based methods (Lesaicherre *et al.*, 2002), have been developed.

Other recent advances include the development of heterodimeric leucine zipper pairs (Zhang *et al.*, 2005), as well as oligonucleotide-directed immobilisation (Wacker and Niemeyer, 2004). These take advantage of the excellent specificity of the Watson-Crick base-pairing of two complementary single-stranded nucleic acids, and self-assembling protein arrays can be generated where the probe proteins in bulk solution are each sent to their relevant 'address' on the array surface by unique 'zip-code' tags. This provides unique-site selectivity, exceptional stability and specificity, and uses well-established DNA chip production technology. However, it is still a demanding task to incorporate oligonucleotides into large proteins. Current methods include using chemically modified oligos capable of reacting to lysine or cysteine amino acid residues, but site-specific methods have also been developed.

Uniform covalent:

Uniform covalent immobilisation methods involve the derivatisation of a protein with a unique chemical group at a defined position, which can later react chemoselectively with a complementary group on a solid support (Jonkheijm *et al.*, 2008). In some cases oriented protein immobilisation can be achieved if the protein has a single accessible reactive amino acid. For example, cysteine has a relative abundance of < 1% and is the only naturally occurring amino acid with a thiol group in its side chain. This thiol group is sufficiently reactive at neutral pH to react with chemical functionalities such as maleimide-modified surfaces to form a stable thioether bond. Such cysteine residues should of course be exposed to a solvent-accessible region of the protein. Cysteine residues can also be introduced into proteins through site-specific mutagenesis.

One of the most efficient ways to chemoselectively immobilise proteins is through native chemical ligation (NCL) (Jonkheijm *et al.*, 2008). For example, proteins can be produced with terminal thioesters that can then be immobilised onto cysteine-modified glass slides, or terminal

cysteine-containing proteins can be spotted onto thioester-modified slides. Another example is where a C-intein fragment is covalently immobilised onto a glass surface, while the N-intein fragment is fused to the C-terminus of the protein to be immobilised. When the C- and N-intein fragments interact on the slide surface they form an active intein domain, which self-splices the intein from the protein while simultaneously binding the protein to the surface; the intein is then released into solution [see e.g. (Lesaicherre *et al.*, 2002)].

There are other, less commonly used methods for uniform covalent immobilisation of proteins onto surfaces. One of these utilises enzymatically-reactive capture surfaces: the target protein is fused to an enzyme, which reacts irreversibly with groups on the array surface thereby indirectly immobilising the target protein to the surface. For example, a slide coated with phosphonate groups reacts irreversibly with the serine residue in the active site of the enzyme cutinase, which can be fused to a target protein (Hodneland *et al.*, 2002). The Staudinger Ligation method (which requires an azide group and a phosphine-containing ester or thioester) (Watzke *et al.*, 2006), and immobilisation by cycloaddition ('click-chemistry') reactions, which require azide groups and terminal alkynes (Sun *et al.*, 2006), are further examples.

4.2.3 Minimising non-specific adsorption

Non-specific adsorption of macromolecules to surfaces presents a major obstacle in the development of protein arrays (Hartmann *et al.*, 2009). Unlike DNA, which has a uniformly negative charge, proteins tend to adsorb non-specifically to a variety of surfaces. This can give rise to high background signals and therefore low signal-to-noise ratios, thereby decreasing the sensitivity of the array.

This can be reduced in part by the careful choice of surface material and printing and assay buffers. Naturally-occurring polymeric surfaces such as elastin, sarcosine, agarose, cellulose and polysaccharides, or synthetic polymeric surfaces such as fluorocarbon polymers or poly(ethylene glycol), are often used for this reason. PEG is thought to suppress non-specific binding due to its intrinsic inertness, hydrophilic nature, mobility and chain density (Lu *et al.*, 2000), as well as by acting as a spacer between the slide and the immobilised proteins (Lesaicherre *et al.*, 2003). The latter also ensures the prevention of surface-induced denaturation. One very promising approach is using oligo(ethylene glycol) derivatives (Chapman *et al.*, 2000). Unlike PEG, these derivatives

are thought to resist protein adsorption due to the presence of a structured and tightly bound water layer on the polymer surface. These polymers are however susceptible to autooxidation. Surface phospholipids have also proven useful, which is believed to be due to their strong hydration capacity.

To help minimise non-specific protein adsorption, array surfaces are often treated with a blocking agent. Although many elaborate methods have proven effective, the old methods of blocking reactive sites with proteins such as bovine serum albumin, skimmed milk powder or other reagents, as well as the presence of surfactants such as Tween-20 and SDS, are often indispensable to the suppression of non-specific adsorption (Jonkheijm *et al.*, 2008). A confounding factor is that strategies effective in eliminating non-specific adsorption often do so at a price: the adsorption of blocking molecules passivates the sites for non-specific adsorption but might also obscure immobilised ligands thereby interfering with desired interactions (Houseman and Mrksich, 2002). On some surfaces, these blocking agents can actually lead to an increase in background, while the use of detergents can disrupt certain protein-substrate interactions or promote the denaturation of immobilised proteins. The final choice of blocking agent will therefore greatly depend on the surface functionalisation and immobilisation methods used, and what works for one application might not necessarily prove optimal for another application.

4.2.4 Sample spotting

Typically, protein arrays are prepared by immobilisation onto a chemically derivatised glass slide using either a contact spotter or a noncontact microarrayer, both of which are routinely used in constructing DNA microarrays (Jonkheijm *et al.*, 2008;Zhu and Snyder, 2003). Robotic contact printing machinery can be used to create protein microarrays with >30,000 spots per slide. These microarray printers transfer nanolitre sample volumes on the outside of solid pins or inside split- or ring-shaped pins directly onto the surface of the slide at defined *xy* coordinates. Solid pins need to be dipped into the sample between each spotting, while split pins do not need to be re-loaded with sample as frequently (they hold a greater volume of sample by drawing it up into the pin through capillary action). Split pins can therefore multi-dispense the same sample faster, however solid pins are arguably better as they are cheaper, more robust, and not prone to clogging.

However, contact printers have more difficulty aligning their pins to pre-fabricated structures, hence the reason for the development of non-contact printing robots. Non-contact robots employ ink-jet technology or electrospray deposition technology. These robots are not restricted to specific surface structures and are suited for more complex assays; however they are slower when printing multiple samples. Current contact and non-contact printers deposit volumes of <100 nl, producing feature sizes between 30-300 μm with array densities of >2000/cm^2 possible, depending on the technology used (Tao *et al.*, 2007;Wingren and Borrebaeck, 2007).

The conditions required for effective deposition of protein solutions differs significantly to that for DNA solutions. Protein solutions are often less concentrated but have a much higher viscosity and a tendency to adhere strongly to a variety of surfaces. Some problems that might therefore arise during printing are sample agglomeration on solid pins with subsequent sample loss and possible carryover, and capillary clogging and air bubble formation in inkjet and split-pin technologies (Barron *et al.*, 2005). Solid pins are less sensitive to variation in sample viscosity, however this can be a major concern when using split pins - and to a lesser degree, when using ink-jet printers and electrospray deposition. It is therefore necessary to routinely check the printing mechanism and to ensure that equal sample loading is occurring. As with DNA microarrays, careful preparation of samples and slides prior to printing is required, and printing conditions such as humidity, temperature and dust levels need to be monitored and controlled throughout the print run (Schweitzer *et al.*, 2003).

Newer, nanotech methods such as dip-pen nanolithography (DPN) have recently been developed that produce feature sizes in the nanometer range, allowing much higher feature densities than that possible with conventional arrayers (< 1 x 10^6/mm^2) (Wingren and Borrebaeck, 2007). However, most of these nanoarray methods are not yet capable of being multiplexed to the same extent as those used to produce microarrays, with <1000 unique probes possible for microarrays but <10 normal for nanoarrays.

4.2.5 Detection

Detection methods used in protein array analysis can be grouped into label-dependent and label-free technologies (Templin *et al.*, 2003;Yu *et al.*, 2006). Fluorescence, chemiluminescence, colourimetric and radioactivity are all label-dependent methods. Of these fluorescence is the most

popular, due to its sensitivity, stability, high resolution, and the availability of fluorescent scanners and commonly used fluorescent dyes, such as the Cy3 and Cy5 dyes used routinely for DNA microarrays. The various different fluorophores available also allow multiplexing in for example, differential protein expression profiling. Radioactivity is used less frequently as long exposure times and safety precautions are necessary, and although chemiluminescence is highly sensitive it can lead to low resolution through signal bleeding and limited dynamic range. In addition, microarrays utilising radioactivity or chemiluminescence rapidly lose signal whereas those using fluorescence can be archived, as fluorophores can undergo numerous excitation-emission cycles.

Label-dependent technologies are not always sensitive enough to detect low-abundance proteins (Kumble, 2003) and labelling proteins can affect activity and prevent epitope recognition by antibodies (Yu *et al.*, 2006). Test compounds and assay reagents can also interfere through auto-fluorescence. The other disadvantages are that one is restricted to the available lasers for detection and that labelling efficiencies can vary widely between different proteins and molecules, which makes accurate quantification difficult. In such cases label-free detection methods can be used (Wingren and Borrebaeck, 2007). These include surface plasmon resonance imaging, mass spectrometry, electrochemical impedance spectroscopy and atomic force microscopy. Proof-of-concept studies have also been performed for newer label-free methods, such as scanning Kelvin nanoprobe, nanoscale interferometry and microcantilevers, while Maven Biotechnologies' internal reflection ellipsometry-based system has been validated with protein array densities up to 2500 spots/cm^2, detecting molecules as small as 150 Da (www.mavenbiotech.com). There are also biosensor array platforms that produce electrochemical readouts of enzymatic reactions, e.g. GRAVI-Chip from DiagnoSwiss (www.diagnoswiss.com). Assay sensitivities in the picomolar range are already achievable for nanowire-based biosensors.

Several label-free technologies allow the real-time detection of reactions and interactions and thereby allow kinetic analyses to be performed (Hartmann *et al.*, 2009). In contrast to label-dependent detection where both unbound and bound label is detected, label-free methods can also distinguish between the two forms. However, the majority of label-free methods are currently low-throughput and in certain cases less sensitive than fluorescence methods (i.e. their signal-to-noise ratios are not yet optimal) and still have to improve in reliability and breadth of application before being used routinely in multiplexed assays. Nonetheless, several studies have

already successfully made use of this technology. For example, the interactions between an *E. coli* membrane protein receptor (FhuA) and viral ligands were investigated on protein arrays using resonating microcantilevers, a system that was found to be capable of measuring interactions at subpicomolar concentrations (Braun *et al.*, 2009).

4.2.6 Data acquisition and analysis

When using fluorescence-based detection, the binding of the probe to the array can be detected with either a scanning confocal laser or a charge coupled device (CCD) camera (Kumble, 2003). Once the array has been scanned, specialised software is used to correct for background noise and to normalise the data, with various options that are user-defined. The software used is often the same as that developed for DNA arrays, as the technical issues of spot detection for protein arrays are fundamentally similar (Hartmann *et al.*, 2009). Several public domain and commercial software packages are available to analyse microarray data.

There are several controls that should be included on each chip to assist with background subtraction, normalisation and calibration (Kumble, 2003). Negative control spots are used to subtract intrinsic background signal from the slide surface as well as the buffers, proteins and other analytes used. Depending on the application, negative control spots are also used to control for non-specific binding to the slide surface. If no barcodes or numbering is present on the slides, marker spots are also required for orientation. Normalisation, which accounts for factors such as biological variation as well as variation in sample quality and labelling, sample loading onto the surface, as well variation in the slide surface, also has to be performed. Decisions have to be made as to how failed spots or subarrays will be treated during analysis, how they effect the rest of the data, and whether or not such data can be used (Hartmann *et al.*, 2009)

While developing and optimising a protein array platform, it is necessary to determine and then minimise sources of technical variation produced by the specific set of microarray equipment used. These include the linearity and limits of detection of the method of detection, the variability in sample loading between replicate spots, as well as the variability between replicate subarrays and slides due to surface and batch variations. The temperature, humidity and printing buffers used during the print run need to be optimised to prevent spot running or evaporation. Processing of the slides, such as the blocking, washing and drying steps also need to be optimised

such that background is minimised while protein stability is maintained. Optimising these conditions will significantly improve the quality of the data obtained.

It is critical that protein microarray experiments and the data analysis thereof be standardised, and journals and authors increasingly follow set guidelines to facilitate data reproduction (Kricka and Master, 2008). The large number of diverse protocols for DNA array construction led to the requirement for a consensus set of minimal experimental information (the 'Minimal Information About a Microarray Experiment', or 'MIAME') that has to be available along with published results (Brazma *et al.*, 2001). A similar format for protein arrays (MIAPE) has started to emerge as part of the larger proteomics field (Taylor *et al.*, 2007), however most of the work to date has focused on protein expression data derived from gel electrophoresis and mass spectrometric analysis. The FDA has also set up a MicroArray Quality-Control (MAQC) project to evaluate the reliability and reproducibility of DNA arrays, and to investigate inter-platform and inter-laboratory concordance of results (FDA, 2009). While a similar platform is not yet well established for the protein array field, several groups (such as the Human Proteome Organisation's Proteomics Standards Initiative, www.hupo.org) have started to address this issue.

Large, multiplexed assays generate huge amounts of data. As multiplexing becomes more commonplace it is likely that effort will be put towards developing pattern recognition software to replace or complement current methods that assess each single reaction spot (Hartmann *et al.*, 2009). The BioPlex 2200 ANA system (Bio-Rad, USA) is an example of a platform that provides software utilising a pattern-recognition algorithm for protein arrays.

4.3 APPLICATIONS

Protein array assay formats are often classed as either forward-phase or reverse-phase assays (Hartmann *et al.*, 2009). The former involves the use of high numbers of purified capture molecules arrayed onto the surface, while the analyte applied across the whole array is the same complex sample (e.g. investigating which proteins are present in one patient's serum sample using various antibodies as capture molecules). This enables the analysis of several parameters of a single sample. The latter format involves the arraying of different complex samples, while the analyte applied to the whole array consists of a single, purified moiety (e.g. investigating the presence and levels of a specific protein in different patients' serum samples using a single

antibody as an analyte). This method is used to analyse a single parameter of several different samples. Reverse-phase arrays enable the detection of differential levels of protein in diseased versus healthy samples, as well as the identification of disease-specific biomarkers and cell signalling networks. Tissue arrays are a special form of reverse-phase arrays, which consist of tissue slices immobilised to the array surface. These arrays form a miniaturised equivalent to traditional immunohistochemistry assays and enable analysis of large numbers of tissue samples with minimum reagent consumption. Forward-phase arrays are the most commonly used format however, and are often used for protein function studies or for protein expression analysis.

Another frequently used grouping categorises protein arrays as either analytical or functional (Phizicky *et al.*, 2003). Analytical microarrays are used for detecting the presence and abundance of specific target molecules in a complex mixture, while functional microarrays are used to assay a wide range of biochemical activities such as protein-protein, protein-lipid, protein-nucleic acid, antibody-antigen, protein-carbohydrate, receptor-ligand and enzyme-substrate interactions. Target molecules can range from purified proteins to ligands present in cell lysates and tissue samples. Below follows a brief description of these formats and the challenges they present, as well as a few examples of the analytical and functional applications of protein microarrays used to date.

4.3.1 Analytical protein microarrays

On analytical protein arrays, affinity reagents such as antibodies, antigens, aptamers, affibodies, and nucleic acids are arrayed at high density (Hartmann *et al.*, 2009;Phizicky *et al.*, 2003). The arrays are incubated with complex samples such as cell lysates or body fluids, the proteins from these mixtures are captured by the affinity reagents and the captured proteins are then detected and quantified. Similar to the procedure in DNA microarray experiments, protein samples from two biological states can be labelled with two different fluorescent dyes, mixed, and then applied to the array. Red and green spots signify higher levels of specific proteins in one state compared to the other, while yellow spots show no significant difference.

Captured proteins are often detected by means of a labelled antibody that recognises another epitope on the protein targets. This method is called a 'sandwich assay' and specificity can be greatly enhanced by using matched antibody pairs in this format. Antibody arrays are the most

common form of analytical protein arrays, as they are relatively robust and extensive antibody libraries are readily available (Takahashi *et al.*, 2003;Zhu and Snyder, 2003). This has made antibody arrays the starting point for the adoption of protein microarray technology, as these proteins were already broadly used and well known from immunoassays.

The disadvantage of analytical protein arrays is that they require prior knowledge of the proteins of interest, in order to produce affinity reagents that will recognise them (Hartmann *et al.*, 2009;MacBeath, 2002). The greatest challenge facing these arrays however, is the generation of affinity reagents with sufficient specificity and sensitivity. The invention and commercial production of monoclonal (as opposed to polyclonal) antibodies was a major step forward for these types of arrays, as it meant that highly specific capture molecules could be generated for array use, and several *in vitro* methods of developing high quality affinity molecules have also been developed since then. Despite this, the generation and validation of any type of affinity reagent is still a time-consuming and costly process. Several global efforts have therefore been initiated to produce large numbers of diverse, highly specific and sensitive affinity reagents to address the needs of proteome-scale arrays (e.g. Human Protein Atlas project, HUPO Antibody Initiative, ProteomeBinder and Protein Capture Tools) (Hober and Uhlén, 2008).

Analytical arrays are mainly used for monitoring protein expression profiling, biomarker identification, cell surface profiling, and clinical diagnostics (Phizicky *et al.*, 2003). In one of the largest analytical array studies carried out**,** Sreekumar and colleagues spotted 146 distinct antibodies onto glass slides to monitor protein expression levels in LoVo colon carcinoma cells (Sreekumar *et al.*, 2001). Their results revealed radiation-induced up-regulation of several proteins, including p53 and tumour necrosis factor-related ligand, as well as several down-regulated proteins. In a study by Hiller *et al*. (Hiller *et al.*, 2002), 94 common allergen molecules were arrayed on glass slides and used to miniaturize the allergy test. The allergen molecules were a range of proteins, peptides and small molecules of common allergen sources. A minimal amount of the patient's serum was applied to the allergen array and allergic reactivity was then profiled. Results obtained using this array have been confirmed using traditional skin tests. This platform is now commercially available from VBC-Genomics (www.vbc-genomics.at). In a report by Hsu and colleagues, hierarchical clustering of the expression profiles of 11 soluble receptors from samples obtained from 36 critically ill patients enabled them to divide these patients into sepsis and trauma groups, thereby better informing their subsequent treatment (Hsu *et al.*, 2009).

Proteomic profiles of arrays of normal versus cancerous ovarian tissue were able to completely discriminate between the two tissue states in a study by Petricoin *et al.* (Petricoin *et al.*, 2002). This distinct change in the pattern of protein expression would never have been identified if individual protein expression levels had been measured. This method produced a much higher positive predictive value of 94% compared to the low 20% value obtained with the most commonly used biomarker for ovarian cancer, cancer antigen 125. In a similar study, the same group demonstrated that the activation of the signalling molecule Akt is a key event in the progression of prostate cancer (Paweletz *et al.*, 2001).

4.3.2 Functional protein microarrays

Functional protein microarrays are constructed by immobilising purified proteins at high density and then investigating their functional interactions with various analytes applied to the array surface (Hartmann *et al.*, 2009). The ultimate form of a functional protein array is one that contains all the proteins encoded by an organism's genome, i.e. a proteome array. Functional arrays containing subsets of the proteome, such as proteins with similar biological activities (e.g. protein kinases), proteins expressed in the same tissue or sub-cellular compartment, or proteins associated with a particular disease state, are also very useful (Schweitzer *et al.*, 2003).

The greatest obstacle facing the use of functional protein arrays, is the production of the large diversity of purified proteins that form the array elements (MacBeath, 2002;Schweitzer *et al.*, 2003). Current protein expression systems have several disadvantages: cDNA libraries are not normalised, they fail to express low abundance proteins and proteins are often not expressed for the full-length of their coding sequence. Proteins that are expressed are often not active as they are misfolded, or lack the post-translational and/or reversible covalent modifications required for their activity. In addition, protein diversity resulting from alternative mRNA splicing is usually not well represented. With whole-genome sequence data now available, there are projects underway to create cDNA libraries that are normalised and contain full-length open reading frames. These libraries coupled with high-throughput protein expression and purification systems, would motivate the widespread application of protein microarrays in functional proteomics.

Due to the many technical challenges of cell-based protein expression, cell-free systems are gaining ground as an attractive alternative that can produce large numbers of diverse proteins on-chip (He and Taussig, 2007;He and Wang, 2007). Cell-free systems consist of arrays of PCR templates immobilised onto the slide surface in discrete positions. All the necessary enzymes, amino acids, nucleotides, salts, etc., needed for *in vitro* protein transcription and translation are then added (these are often supplied in a crude lysate extracted from a source of choice, e.g. *E. coli*, human cells, or *Xenopus* oocytes), thereby initiating protein expression. This circumvents the need for cloning and the problem of heterologous expression. The system is open to manipulation, thereby allowing direct control of the conditions of protein folding (such as temperature or pH), as well as the insertion of PTMs or labels. The expressed protein is kept immobilised to the surface by pre-treating the surface with a protein-capturing reagent, such that newly synthesised proteins are captured as soon as they are released from the ribosomes. A specific antibody or an N- or C-terminal affinity tag can be used for this purpose. Alternatively, by deleting the stop codon from the PCR template, the nascent protein will remain coupled to the encoding mRNA, as the translating ribosome will be unable to dissociate from the 3' –end of the mRNA strand. Expression levels comparable and sometimes exceeding that of cell-based systems have been reported.

Functional protein arrays are used in basic research such as analysing protein binding partners, protein function and post-translational modifications, but are also particularly useful for more complex problems such as building cell signalling networks (Phizicky *et al.*, 2003). The pioneering work of the MacBeath and Zhu groups described earlier (Section 4.1.1) produced some of the first examples of functional protein arrays. In another seminal study in 2001, Zhu *et al.* built on their earlier achievements by developing the first ever proteome microarray (Zhu *et al.*, 2001). They cloned ~5800 of the 6200 ORFs present in the yeast genome into a yeast expression vector which expressed the proteins as GST-His_6 fusions. The proteins were then affinity-purified utilising the GST tags, and immobilised onto a nickel-coated glass slide *via* the hexaHis tag. The proteome microarray was subsequently probed with Cy-3-labelled calmodulin and various phosphatidylinositides (PIs). Calmodulin is a calcium-binding protein that is of great interest as it regulates many signalling pathways, while PIs are important secondary messenger molecules that regulate diverse cellular processes. In addition to identifying several known interactions, 33 novel calmodulin-binding partners were discovered. Sequence comparisons revealed a new consensus-binding site that was distinct from the previously discovered calmodulin-binding motif. 150 lipid-

binding proteins were also identified, 52 of which were novel. These achievements would not have been possible without the comprehensive screening enabled by the proteome microarray, and several of these interactions had been missed by earlier large-scale yeast two-hybrid and/or affinity purification-MS studies.

4.4 FUTURE CHALLENGES FOR PROTEIN ARRAY TECHNOLOGY

Protein microarray technology has enormous potential in basic research, diagnostics and drug discovery. In the early days of DNA microarray development, they were expensive, had few known applications and their utility had yet to be proven. Now however, DNA microarray technology is used extensively in the field of functional genomics. Protein microarray technology is currently undergoing the same maturation process and array-based technologies will continue to find an ever wider application within the field of proteomics.

In order to meet the ideal of developing protein arrays for proteome-wide investigations, nanotechnology-based capabilities have to be developed to allow for the fabrication of highly dense arrays (>50 000 spots/mm^2 with features of <1 μm), such that arraying a complex proteome with hundreds of thousands of proteins can be feasible (Wingren and Borrebaeck, 2007). There are some nanotech printing technologies already in use, such as microcontact printing, dip-pen nanolithography, and nanodispensing, however these are still suffering technical difficulties and have yet to address the need for multiplexed arrays where multiple different proteins are printed at individual locations on the array surface. Recent progress has been made by other groups who have used a DNA surface assembly protocol to prepare extraordinarily precise nanoscale patterns of biomolecules [e.g. (Park *et al.*, 2005)]. Another issue is that the dynamic detection range of current, commonly-used imaging technology is limited to 3 to 4 orders of magnitude, while complex protein samples contain protein concentrations of 6 to 10 orders of magnitude (Kumble, 2003).

The high costs and the expertise required to produce protein arrays remain limiting for most laboratories wanting to generate their own arrays, while the costs of purchasing commercial arrays are typically even more prohibitive (Tao *et al.*, 2007). Besides the material and equipment costs of array fabrication, additional cost and time constraints are often imposed by the need to generate and validate target and capture molecules. This is not only a problem for antibody-based

arrays, but also for protein function arrays. The traditional method of protein preparation consisting of recombinant cloning, heterologous expression and purification remains the gold standard; as although there are several promising cell-free alternatives, these are not yet feasible for large-scale studies. As mentioned previously, there are several global efforts underway to produce large numbers of highly specific and sensitive affinity reagents to address the needs of proteome-scale arrays, but these remain a long way from completion (Hober and Uhlén, 2008).

Finally, most protein array results are only semi-quantitative and as yet there are no widely accepted experimental and data analysis protocols for standardisation. This problem is more complex than that originally presented by DNA microarray analysis, given the wide variety of possible applications of protein arrays and their correspondingly varied data analysis requirements.

Most of the challenges above are exacerbated by large, multiplexed studies using hundreds to thousands of different proteins. In contrast, smaller studies are more focussed and already achievable, and large quantities of high quality data can be produced. The utility of such arrays has been shown in a wide range of applications and there is little doubt that this field will continue to progress at a rapid pace. This is demonstrated in the following two chapters of this thesis, which cover the development and optimisation of a novel cytochrome P450 protein array platform within our laboratory.

CHAPTER FIVE

Development of a Cytochrome P450 Protein Array Platform

5.1 INTRODUCTION

5.1.1 Current CYP protein array technology

Protein arrays are also still only sparsely integrated into drug discovery research (Hartmann *et al.*, 2009) and as yet few examples exist of array technology being used to investigate the cytochrome P450s. Besides the challenges generally posed when miniaturising assays in the protein array format (as discussed in Chapter Four), the cytochrome P450 system poses several additional significant hurdles.

The main obstacle facing researchers attempting further CYP assay miniaturisation from well-plates to protein arrays, is the immobilisation of these unstable membrane proteins and their reductase partners in such a way that enzyme activity is maintained and electron transfer can occur. These two proteins do not form a stable complex [K_d = 0.05 μM (Miwa *et al.*, 1979), K_m for electron transfer from the reductase = 0.2 μM (Voznesensky and Schenkman, 1992a)], yet it is critical that when they do associate that they be in the correct relative orientation for electron transfer to occur. Neither of these enzymes are particularly stable in solution, nor are they easily expressed in a fully folded, functional form in heterologous systems (Yun *et al.*, 2006). In terms of array-based assays, this poses the challenge of maintaining functionality that might already have been compromised to some degree during protein and array production.

The earliest approaches to the miniaturisation of CYP assays involved the immobilisation of CYP enzymes onto graphite, glassy-carbon or gold electrodes**,** e.g. (Fantuzzi *et al.*, 2004;Hvastkovs *et al.*, 2007), or in sol-gel droplets in a 96-well microtitre plate in an array of arrays format prior to assay (Sakai-Kato *et al.*, 2005). However, these approaches only work on a limited scale and are not amenable to high-throughput screening.

Lee *et al*. carried out the first significant attempt at CYP protein arrays in 2005 (Lee *et al.*, 2005). This group encapsulated commercial baculosome CYP and CPR enzyme preparations together with a NADPH regeneration system (glucose 6-phosphate, glucose 6-phosphate dehydrogenase and $NADP^+$) within a sol-gel matrix in an array format and used the array for basic toxicity screening. To elaborate, the sol-gel enzyme preparation was spotted onto a slide in 30 nl spots and once the spots had set for 24 hrs at 4°C, 60 nl of substrate was printed directly on top. The slide was then immediately 'stamped' onto a monolayer of MCF7 breast cancer cells and after 6 hrs the cells were screened for cytotoxicity of the CYP metabolites.

For the future applications of this thesis however, the sol-gel encapsulation method is limited due to the long preparation time (24 hrs required for gelation and aging), which would hamper HTP array fabrication and could result in decreased enzyme activity of the unstable CYPs. In fact, the authors noted a two-thirds decrease in enzyme activity compared to that in solution. They concluded that the remaining activity was sufficient for the purposes of their more qualitative assay, however this drop in activity could of course prove detrimental to attempts at quantitative kinetic profiling. This 3D matrix-based format is also less amenable to multiplexing and platform extension to other detection methods due to the problems associated with metabolite retrieval and solution exchange (see Section 4.1.3).

In the most quantitative study yet, published in 2009, Sukumaran *et al.* produced a CYP microarray by encapsulating commercial baculosome CYP and CPR enzyme preparations and a NADPH regeneration system within an alginate matrix and printing this onto functionalised slides (Sukumaran *et al.*, 2009). Up to 1134 discrete reactions in volumes of 80 nl were printed and these arrays were then used for fluorescence-based kinetic and inhibitory assays. Reactions were detected with a novel, custom-built wide-field CCD camera capable of imaging a 20 x 20 mm area. Six 20 x 20 mm images across the 25 x 75 mm slide were therefore imaged consecutively and then used to create a composite image of the entire slide, with a total readout time of 18 sec per composite image. It is unclear how long the encapsulation (gelation and aging) process took for the fabrication of these arrays, although the authors do note that they observed up to 60% of solution-phase enzyme activity on the array surface.

Due to the high spot volumes imposed by the microarrayer used in this study, this array format was unable to obtain truly high-throughput levels expected from a protein array. This format also makes use of a 3D matrix, which, as discussed previously, poses several possible limitations to format flexibility. In essence, this study is therefore simply an extension of the work carried out by Lee *et al.*, described above. One excellent achievement of this study however, was the development of a fluorescence-imaging system capable of much faster readout times compared to current microarray fluorescence-imaging technology. Most current detection methods are not suited to the rapid, real-time array-based imaging required for enzyme kinetics. This is therefore likely to pave the way for future protein array detection technology.

Studies of CYP and CPR immobilisation in applications that are not electrode- or array-based are also limited. What interest there was in developing CYP immobilisation techniques was mainly stimulated by the discovery that immobilisation often has a stabilising effect on the enzyme, which increases its utility for assays and storage compared to enzyme in solution (see references in this paragraph). Most of these studies also make use of membrane preparations (others use purified enzymes reconstituted with lipids), with encapsulation in matrices such as agarose, alginate, or polyacrylamide (Azari and Wiseman, 1980;Hara *et al.*, 1999;Ibrahim *et al.*, 1986;Yawetz *et al.*, 1984), or cross linking to activated Sepharose or Sephadex for example (Ibrahim *et al.*, 1986;Lehman *et al.*, 1981). Other immobilisation methods include random covalent binding to cyanogen bromide-activated sepharose (Baess *et al.*, 1975;Jänig *et al.*, 1979) and random adsorption to ion exchange matrices (Taylor *et al.*, 2000). Using these various methods, equivalent or lower activities compared to solution-phase assays were observed.

It is precisely because of the problem of maintaining electron transfer between the partner enzymes that immobilisation and array-based work to date has focused on using matrix-encapsulated membrane (i.e. baculosomal/microsomal) preparations of cytochrome P450 and the reductase. Despite the membrane preparations being encapsulated on the array surface the partner enzymes are still able to diffuse laterally towards each other *via* their membrane anchor domains - much like they do in solution-based assays, which utilise similar membrane preparations or purified enzymes reconstituted with various lipids. The array-based assays described above are therefore essentially miniaturisations of conventional solution-phase assays. There has therefore long been much room for technological improvements towards fully realising the goal of high-throughput CYP protein function microarrays.

5.1.2 Developing a novel cytochrome P450 platform

This chapter describes the successful development and optimisation of a novel cytochrome P450 array format that allows functional interaction between CYP and its reductase partner on a 2D surface, using membrane-free, recombinant enzymes within sub-nanolitre volumes. The fabrication of CYP arrays using direct, lipid-free and non-random immobilisation of these partner enzymes was therefore successfully achieved for the first time.

In fabricating our arrays, we chose commercial hydrogel-coated glass slides functionalised with N-hydroxysuccinimidyl ester (NHS) groups (Schott Nexterion H slides). The hydrogel on these slides is composed of long chain PEG covalently attached to the glass slide (Nath *et al.*, 2008;Schott, 2006), which offers good resistance to non-specific adsorption (see Section 4.1.3). The slides are laser cut, which ensures that the surface is flat with minimal slide thickness deviation. This is critical, as deviations in surface height introduce loading variation across a slide during printing, thereby reducing the quality of the data. In a recent study by Guilleaume *et al.*, the Nexterion H slides were compared to seven other protein array slide surfaces and found to be one of the two best commercial surfaces with the highest signal-to-noise ratios and the best performance in antibody and protein detection (the other top performer was the Perkin Elmer HydroGel slides) (Guilleaume *et al.*, 2005).

These slides were coated with streptavidin in our laboratory to facilitate directed and highly specific binding of our recombinant biotin-labelled CYP and CPR enzymes. As discussed in the preceding chapter, directed immobilisation is critical for maintaining optimal enzyme function. Except for the bound biotinylated tag on the C-termini of the enzymes, each immobilised protein remains away from the array surface due to the flexible BCCP-protein linker and the repulsive nature of the PEG environment. As a result, there should be minimal disturbance of the native conformation of the proteins due to hydrophobic surface interactions (Cha *et al.*, 2005). Despite being non-covalent, the biotin-streptavidin interaction is almost irreversible (K_d = 10^{-15} M) and is therefore not susceptible to dissociation through washing and block steps or perturbations in pH, salts, etc. (Gonzalez *et al.*, 1997). In addition, the binding interaction between these two proteins is very rapid, thereby omitting the need for a lengthy incubation step once the proteins have been printed onto the array surface (MacBeath and Schreiber, 2000;Zhu *et al.*, 2000). The high specificity of this interaction is also extremely useful, as this allows for one-step binding and purification of biotinylated proteins from crude lysates (Boutell *et al.*, 2004). This is a significant benefit in terms of time, cost and enzyme integrity. Levels of enzyme activity are generally higher in the crude extract compared to downstream purified fractions allowing for more robust kinetic assays.

For printing and detection of our arrays conventional DNA microarray technology was used: a contact printer using solid pins and a fluorescence laser scanner. Besides being amenable to the technology available to us, fluorescence detection was chosen as it is a well-established method for detecting CYP activity and suffers few of the drawbacks faced by other methods (see Section

4.1.3). However, as mentioned earlier in the introduction, current microarray fluorescence-imaging technology does not include a real-time fluorescence monitoring capacity (a scan at each enzyme assay time point has to be prompted manually), while the readout times required to scan the entire slide are too long to capture the linear rates of certain enzyme reactions. For the purposes of this thesis however, the available fluorescence microarray scanner proved both convenient and sufficient for the P450 Biochip development and optimisation described here.

.

The experimental objectives of the development and optimisation process were therefore the following:

1) To determine the detection sensitivity required to measure CYP3A4 activity on-chip.
2) To perform hand-spotted CYP3A4 enzyme arrays based on cumene hydroperoxide to confirm that turnover was indeed detectable in the array format.
3) To build an *in silico* model to explore the choice of CYP3A4 and reductase constructs for array construction.
4) To prepare and test home-made streptavidin-coated array surfaces.
5) To produce and optimise the first robotically printed Δ24C-CYP3A4 and Δ43C-CPR enzyme arrays.
6) To determine the optimal PMT for scanning CYP arrays for quantitative kinetics.

5.2 RESULTS AND DISCUSSION

5.2.1 Determining the detection sensitivity required to measure CYP3A4 activity on-chip

Based on the levels of solution-phase Vivid Red turnover by the Δ24C-CYP3A4 construct using the cumene hydroperoxide assay format (see Section 3.2.2.3), calculations were performed to determine how much resorufin would be produced on an array surface in assay volumes of ≤ 1 µl (see Materials and Methods). Using this information, a range of resorufin concentrations were used to check the sensitivity of the microarray scanner and thereby determine the feasibility of detecting Vivid Red turnover in microvolumes on the slide surface.

The Tecan scanner used for our experiments contains two standard lasers, λ_{ex} = 532 nm ('green') and λ_{ex} = 635 nm ('red'), which are optimised for the use of dyes such as Cy3 and Cy5 respectively (these are fluorescent dyes widely used in DNA microarray applications). The λ_{ex} of resorufin is

within 530 nm, which makes the 532 nm laser suitable for excitation. The Tecan scanner is supplied with a range of emission filters specific to each laser fitted, and resorufin's λ_{em} of 585 nm is within the range of that of the conventional 'green' dyes (~ 560 – 590 nm).

The sensitivity of the laser scanner can be adjusted using different PMT (photo multiplier tube) settings. The PMT amplifies the number of electrons generated from each photon that reaches the photo cathode, thereby increasing the signal levels. This signal amplification can be up to 6 orders of magnitude, depending on the PMT voltage used. The Tecan PMT voltage is adjusted using the 'PMT Gain' setting, which can be set anywhere between 70-255. These gain settings are linearly proportional to the 330-1200 V range.

The chosen resorufin concentration range of 10 nM – 5.12 μM was not detectable at 100 PMT while at 200 PMT all the spots were saturated (see Fig. 5.1). Visible, non-saturated signal across the concentration range was observed at 145 PMT. This clearly demonstrated that nanomolar (and indeed, lower) resorufin concentrations are easily detectable with the laser scanner using PMT settings well within the recommended range. Very little signal was detectable in the red channel, as expected.

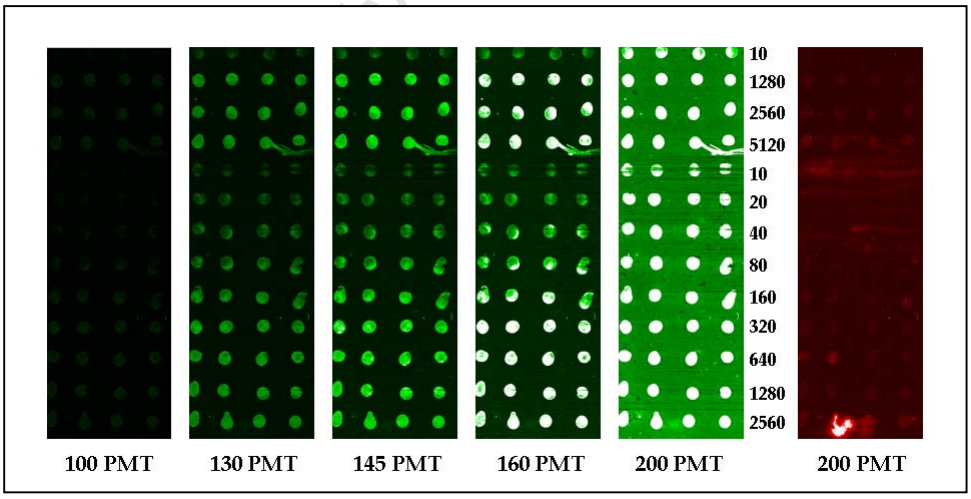


Figure 5.1: Scanned images of a slide hand-spotted with a resorufin dilution series (10 nM-5.12 μM) at various PMT Gain settings. The first five images were acquired in the green channel, while the last image was acquired in the red channel. These images indicated that nanomolar concentrations of resorufin are easily detectable in the array format.

5.2.2 First hand-spotted CYP3A4 enzyme arrays using cumene hydroperoxide

Before commencing with robotically printed protein arrays, several crude hand-spotted enzyme arrays were prepared to confirm that Δ24C-CYP3A4 enzyme turnover of Vivid Red on the slide surface would be detectable. The cumene hydroperoxide assay format was used for this due to its simplicity and its independence of the reductase. Two different assay formats were compared to determine optimal assay conditions and thereby infer the practical feasibility of the cumene hydroperoxide-based assay for father protein array work. Initial testing of the effect of glycerol on enzyme turnover on the array surface was also carried out.

The first crude hand-spotted enzyme arrays confirmed that Δ24C-CYP3A4 enzyme turnover was readily detectable in microvolumes on the slide surface. Increasing signal was observed for 250 nM enzyme with increasing Vivid Red concentrations (except for 2 μM Vivid, possibly due to a dilution error, see Fig. 5.2). Although no negative controls were included in these arrays, it is clear that turnover is occurring due to the marked differences in signal between the two assay formats in reactions with the same concentrations of enzyme and substrate. The second slide format consisting of enzyme and substrate pre-mix spots overlaid with CuOOH solution yielded up to ~3-fold more signal than the first slide format, which consisted of enzyme spots overlaid with CuOOH and substrate pre-mix. This was consistent with earlier observations from solution-phase assays carried out in 96-well plates (data not shown) and suggests the deactivation of the enzyme during incubation with CuOOH alone.

The implication of this finding was that the CuOOH-based assay was not ideally suited for further protein array work. Preferably and for simplicity, the slide surface should first be spotted with enzyme and then washed to remove unbound/non-biotinylated proteins before the pre-mixed substrate and cofactors are printed down in overlaying spots. Since this experiment showed that overlaying enzyme spots with CuOOH and substrate pre-mix yields inferior signal and since washing the slide would remove most of the substrate pre-incubated with the enzyme, the conclusion was drawn that future efforts at building an optimal CYP array would require the focus to move towards the traditional and more complex assay system consisting of CYP3A4 and cytochrome P450 reductase, with NADPH as the electron donor.

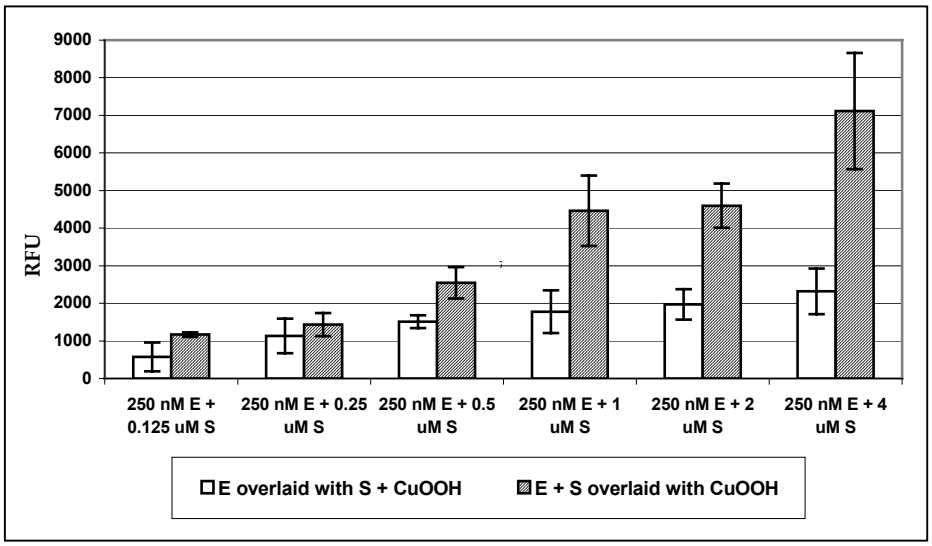


Figure 5.2: Cumene hydroperoxide-based Δ24C-CYP3A4 turnover of Vivid Red in a hand-spotted enzyme array. Two different assay formats were investigated: enzyme spots overlaid with CuOOH and substrate (*white bars*), and enzyme plus substrate spots overlaid with CuOOH (*hatched bars*). The second assay format was found to give superior turnover. Note: E = enzyme, S = substrate, RFU = relative fluorescence units. Error bars represent standard deviation of three replicates.

In the second experiment, two hand-spotted enzyme arrays were prepared using the assay format found to be optimal in the first experiment. Several negative controls were included in these arrays to confirm that signal generated in the earlier hand-spotted arrays was due to enzyme turnover rather than background fluorescence of the substrate. The results conclusively showed that CYP3A4 enzyme turnover was indeed occurring (see Fig. 5.3).

Since the reaction times in the earlier hand-spotted arrays were observed to be limited due to rapid evaporation, glycerol was included in the reaction mixtures for these arrays to slow evaporation. This is a common additive in protein arrays construction for this purpose, as well as for stabilising proteins (Nath *et al.*, 2008). The addition of glycerol was observed to visibly reduce the evaporation rate on the CYP arrays, while 7.5% glycerol appeared to yield ~1.5-fold higher signal than 10% glycerol (although the large error bars exclude any final conclusion). Glycerol is known to inhibit cytochrome P450 turnover, although the exact mechanism/s of this inhibition are still unclear (the possible implications of glycerol inhibition on CYP kinetic profiling are discussed in Section 6.2.4). However, as glycerol was found to be necessary for the reactions to proceed on the array surface, it was used for all further array development.

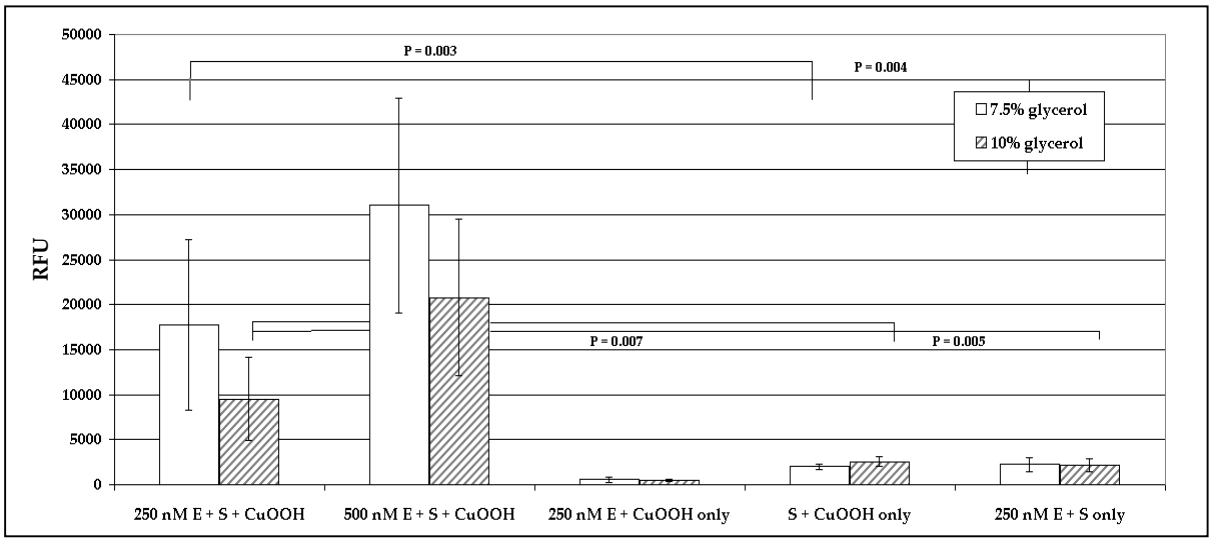


Figure 5.3: Cumene hydroperoxide-based Δ24C-CYP3A4 turnover of Vivid Red on a hand-spotted enzyme array. The effect of two different glycerol concentrations was investigated and several negative controls were included. Despite the large error introduced by hand-spotting, there are significant increases in signal in the reactions versus the negative controls. Note: Error bars represent standard deviation of three replicates, P-values calculated from type 2, two-tailed t-test.

5.2.3 Choice of CYP3A4 and reductase constructs for fabricating the array

From the separate solution-phase enzyme assays carried out for each of the CYP3A4 and CPR constructs (see Sections 3.2.2 and 3.2.3), it was clear that the Δ24C-CYP3A4 and the Δ43N-CPR and Δ43C-CPR constructs yielded the highest enzyme activities in solution. For each of these constructs, the N-terminal deletions consisted of the entire membrane-anchoring fragment. It therefore seemed likely that these constructs would be able to re-form productive associations on the array surface because their cytosolic domains remained intact. Replacement of the N-terminal membrane anchors of CYP3A4 and CPR with an N-terminal BCCP-tag should intuitively be compatible with preserving the correct relative geometric arrangement of these two enzymes when immobilised *via* the tag onto an array surface. It was less immediately obvious though that this would be the case when the BCCP-tag was positioned at the C-terminal ends of CYP3A4 and CPR.

However, since the choice of CYP3A4 constructs was constrained functionally to the C-terminal tagged CYP3A4 construct, the Δ43C-CPR construct was chosen as the most likely partner for

Δ24C-CYP3A4. In order to explore the possible geometric impact of this choice, various protein modelling tools were used.

5.2.3.1 Superimposition of CYP3A4 and CPR onto the BM3 bacterial template

As mentioned previously, the comparison of various crystal structures of eukaryotic and prokaryotic CYPs have shown that even where global sequence identity is low, the overall fold topography is highly conserved (Otyepka *et al.*, 2007). Of the bacterial CYPs, the well-known CYP BM3 bears the most structural similarity to mammalian CYPs (Sevrioukova *et al.*, 1999b). BM3 is a 119 kDa soluble protein comprising of a typical CYP heme domain and a typical FMN/FAD-containing reductase domain joined together in a single polypeptide. Due to the high structural similarity observed (including in the FMN domain (Zhao *et al.*, 1999)), BM3 provides an excellent model system for studying mammalian CYP-CPR interactions and electron transfer mechanisms.

The crystal structure of the BM3 heme- and FMN-containing domains was determined by Sevrioukova *et al*. in 1999, which depicted an interaction complex between these two domains at a 2.03 Å resolution (Sevrioukova *et al.*, 1999b). The FAD-binding domain was omitted to simplify their studies of heme-FMN interaction and electron transfer. Using this structure (1BVY), we set out to build a model for the interaction between the human CYP3A4 and CPR enzymes, as well as to explore whether this interaction can still occur in principle after immobilisation to the array surface *via* the C-terminal tags.

There are several CYP3A4 structures available, most of which are complexed to one or more substrates. As mentioned previously, CYP3A4 is the most conformationally flexible of the CYPs. Different ligand-bound structures can vary greatly in terms of the size of the active site and in the positions of the ordered C^{α} atoms (Ekroos and Sjörgren, 2006). Substrate-bound CYP structures are referred to as 'closed' conformations, while ligand-free structures are termed 'open'. Depending on the substrate, minimal to large conformational changes can occur between open and closed CYP conformations, mainly due to movements of the F and G helices and the F/G loop (and in some cases to a lesser extent in the I and B' helices and the β4 region) (Li and Poulos, 2004). To simplify the structural superimposition of CYP3A4 onto the heme domain of BM3, a ligand-free structure (1TQN) was therefore used (Yano *et al.*, 2004).

As there were no complete human CPR structures available containing both the FMN and FAD domains, a full-length structure of a rat (*Rattus norvegicus*) CPR was used instead. Specifically, an 'open' conformation structure (3ES9) was used that depicts the FMN domain facing outwards, away from the FAD domain interaction interface - as would presumably occur during electron transfer between the FMN domain of CPR and the heme of the CYP (Hamdane *et al.*, 2009). The rat CPR is known to be very similar in both sequence and FMN domain structure to human CPR (Zhao *et al.*, 1999). This was confirmed using sequence and structure alignment tools for these two proteins, which were found to share 92% sequence identity and a structural fit with 0.42 Å RMSD. It was therefore tenable to substitute the one structure for the other.

Using the Swiss-PDBViewer software, human CYP3A4 was aligned and fitted onto the heme domain of BM3 with a resulting RMSD of 1.50 Å. Similarly, the FMN domain of rat CPR was fitted onto the FMN domain of BM3 with a resulting RMSD of 1.13 Å (see Fig. 5.4 A). Both of these RMSD values indicate a satisfactory fit. To further evaluate these structural alignments, the parallel stacking of both the heme and FMN cofactors was confirmed, with < 2 Å distance between their corresponding atoms. In addition, the highly conserved I-helices of the heme domains were found to be completely superimposed (see Fig. 5.4 B). The proximal surface of the heme domain on CYP3A4 was observed to face towards the FMN domain of CPR, which conforms to previous findings in the literature (Sevrioukova *et al.*, 1999a;Zhao *et al.*, 1999). In addition, by using the software to map the electrostatic potential of the proteins' surfaces, the docking interface was shown to correspond to a mostly basic interface for CYP3A4 (depicted in blue in Fig. 5.4 C) and a mostly acidic interface for CPR (depicted in red in Fig. 5.4 D). This in agreement with previous studies of the CYP3A4-CPR interaction interface (see Section 5.2.3.2 below).

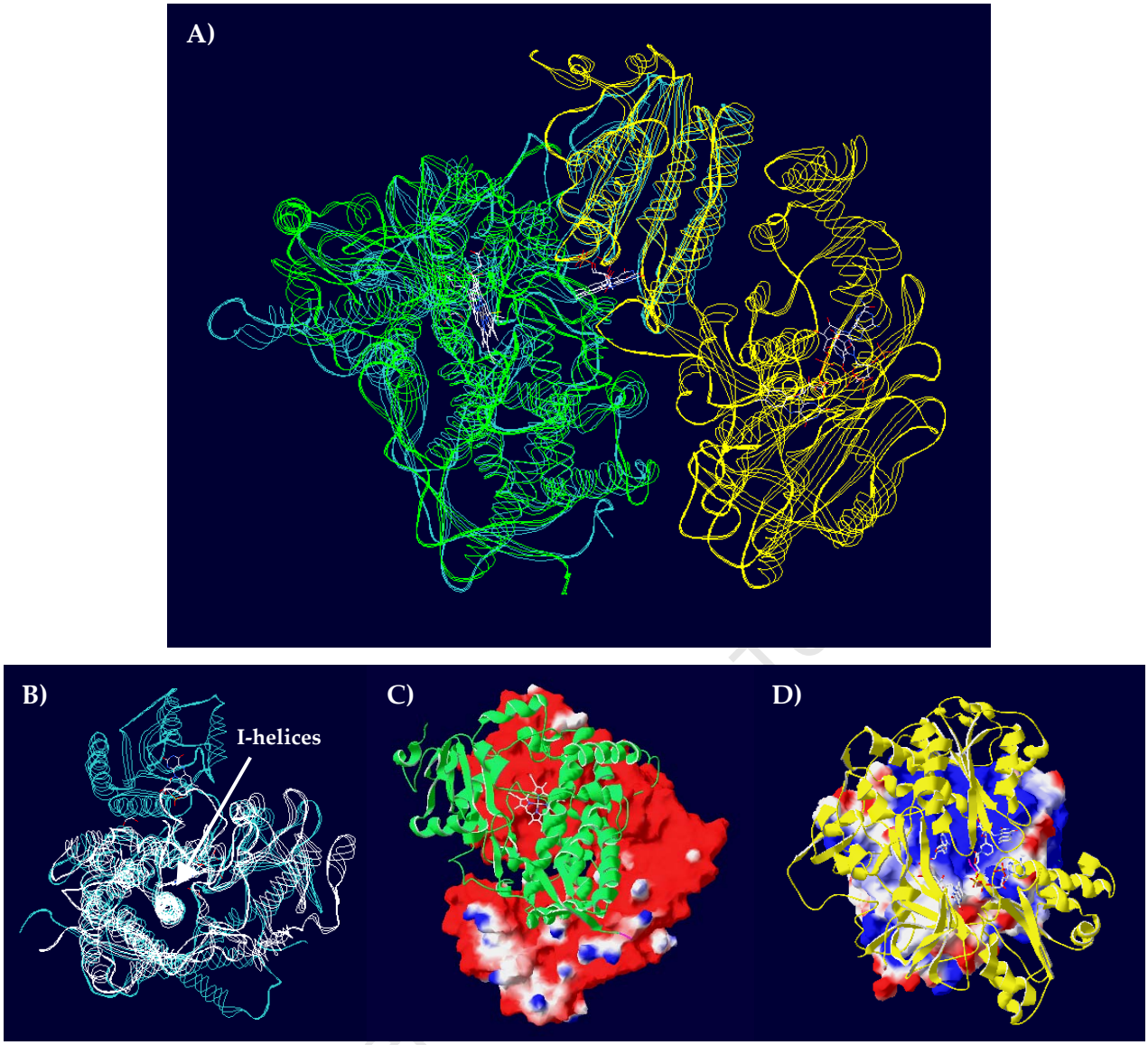


Fig. 5.4: A) Superimposition of CYP3A4 (*green*) and CPR (*yellow*) onto the bacterial BM3 protein (*blue*). Note the close parallel stacking of the heme and FMN cofactors. B) Image depicting the excellent superimposition of the I-helices of the CYP3A4 (*white*) and BM3 (*blue*) proteins. C) Electrostatic potential map of CPR depicting CYP3A4 (*green*) docked onto its surface. D) Electrostatic potential map of CYP3A4 depicting CPR (*yellow*) docked onto its surface.

5.2.3.2 Further validation of the CYP3A4-CPR interaction model using charge cluster pairings

The CYP3A4 and CPR superimposition onto BM3 was then evaluated according to potential electrostatic interactions allowed by the model. There have been several studies illustrating the critical importance of electrostatic interactions in the functional association of CYPs and the reductase (see references in Table 5.2 below). Complementary charge pairing between acidic residues on the CPR surface and basic residues on the CYP surface is thought to be the predominant force in the interaction of these proteins.

A previous study carried out by our group (not described in detail here), describes an extensive literature survey of all the residues involved in electrostatic interactions between various CYP isoforms and either cytochrome b_5 or the reductase (Zawaira *et al.*, 2010). These residues were then 'mapped' onto rabbit CYP2B4 (*Oryctolagus cuniculus*) and the FMN domain of rat CPR using a combination of sequence and structure alignments, for the purpose of validating a superimposition of these two proteins onto BM3 and using the resulting complex to investigate potential charge cluster pairings at the interaction interface. Using similar methods, the same procedure was carried out in this thesis for the purpose of developing a CYP3A4-CPR interaction model. Drawing on the published study, CYP3A4 residues homologous to experimentally defined residues described in the literature were identified using sequence and structure alignments between human CYP3A4 and rabbit CYP2B4 (2BDM). A similar procedure was used to identify experimentally defined residues on the human reductase homologous to those on the rat reductase.

Using the Swiss-PDBViewer software, acidic (Glu and Asp) and basic (Arg and Lys) residues within 4, 6, 8 and 10 Å distance between the CYP3A4 and CPR interface of our model were delimited (see Table 5.1 and Fig. 5.5). These residues were then visually examined for putative charge clusters which were then compared to known charge clusters described in the literature. All of the known charge clusters were identified, while several putative, previously undescribed clusters were also found (see Table 5.2). This suggested that our CYP3A4-CPR interaction model was satisfactory for further modelling of the likely interaction and orientation of these partner enzymes while immobilised. Note that while a distance of 4 Å is often used as the default cut-off distance within which salt-bridge interactions between residues occur (e.g. in the web server, ESBRI, which evaluates salt-bridges in protein structures – see http://bioinformatica.isa.cnr.it/ESBRI/), this restriction is more valid when using structures that have been processed to allow the re-orientation of sidechains for optimal complex formation. Given that this was not carried out during the construction of this model, it is still satisfactory to note that most of the charge clusters identified contain at least some residues within <6 Å of the protein interaction interface (these residues are highlighted in bold in Table 5.2).

Residues within 4 Å		Residues within 4-6 Å		Residues within 6-8 Å		Residues within 8-10 Å	
CYP3A4	CPR	CYP3A4	CPR	CYP3A4	CPR	CYP3A4	CPR
Lys91	**Glu93**	**Arg130**	**Glu92**	**Arg128**	**Glu142**	Arg260	**Asp101**
Lys96	**Asp113**	**Lys141**	**Asp144**	Lys288	**Asp151**	**Arg375**	Asp154
Lys127	**Glu115**	**Lys342**	**Asp209**	**Lys422**	**Asp208**	**Lys378**	**Glu179**
Lys143	**Glu116**	**Lys421**	**Glu214**	**Arg440**			**Asp207**
Lys266	**Asp147**	**Lys424**		**Arg458**			**Asp215**
Arg268	**Glu213**	**Arg446**					
	Glu246						
	Asp277						
	Asp352						
	Glu353						
	Glu354						

Table 5.1: Delimitation of acidic residues (CPR) and basic residues (CYP3A4) within various distances of the CYP-CPR interaction interface. Residues previously described in the literature are highlighted in bold (see references in Table 5.2).

CYP3A4 residues	CPR residues	Literature references for cluster pairings
Lys91, Lys96 (Shimizu *et al.*, 1991)	Glu142, **Asp144**, **Asp147** (Zhao *et al.*, 1999) Asp151 (N/A)	(85R)-(142E-144D-147D) (Zawaira *et al.*, 2010)
Lys127, Arg128, **Arg130** (Bridges *et al.*, 1998;Davydov *et al.*, 1992;Omata *et al.*, 2000)	**Glu92**, **Glu93** (Karyakin *et al.*, 2007;Zawaira *et al.*, 2010)	(133R)-(92E-93E) (Karyakin *et al.*, 2007;Zawaira *et al.*, 2010)
Lys141 (Bridges *et al.*, 1998;Davydov *et al.*, 1992)	**Glu93** (see above) Asp101 (Zawaira *et al.*, 2010)	(133R)-(101D) (Zawaira *et al.*, 2010) (133R)-(93E) (see above)
Lys141, **Lys143** (Bridges *et al.*, 1998;Davydov *et al.*, 1992) **Arg446** (Mayuzumi *et al.*, 1993;Shimizu *et al.*, 1991)	**Glu213**, **Glu214**, Asp215 (Shen and Kasper, 1995)	(443R)-(213E) (Zawaira *et al.*, 2010)
Arg260, **Lys266**, **Arg268** (Cvrk and Strobel, 2001)	**Glu246**, **Asp352**, **Glu353**, **Glu354** N/A	N/A
Lys288 (N/A)	Asp101, **Glu93** (see above)	N/A
Lys342, Arg458 (Bridges *et al.*, 1998;Mayuzumi *et al.*, 1993;Shimizu *et al.*, 1991;Zawaira *et al.*, 2010)	**Asp277** (N/A)	N/A
Lys421, Lys422, **Lys424**, **Arg446** (Mayuzumi *et al.*, 1993;Shimizu *et al.*, 1991)	Asp207, Asp208, **Asp209** (Shen and Kasper, 1995;Zhao *et al.*, 1999)	(421K-422R)-(207D-208D-209D) (Zawaira *et al.*, 2010) (443R)-(208D) (Zawaira *et al.*, 2010)
Arg440 (Mayuzumi *et al.*, 1993;Shimizu *et al.*, 1991)	**Asp113**, **Glu115**, **Glu116** (Zawaira *et al.*, 2010)	(433K-434R)-(113D-115E-116E) (Zawaira *et al.*, 2010)
Lys422 (see above)	Glu179 (Zawaira *et al.*, 2010)	(421K)-(179E) (Zawaira *et al.*, 2010)

Table 5.2: Putative charge cluster pairings for CYP3A4 and CPR interaction based on the *in silico* model. Residues within 6 Å of the interface are shown in bold. References for homologous residues and cluster pairings previously described in the literature are provided in the relevant columns [residue notation for cluster pairings as for rabbit CYP2B4 and rat CPR described in (Zawaira *et al.*, 2010)].

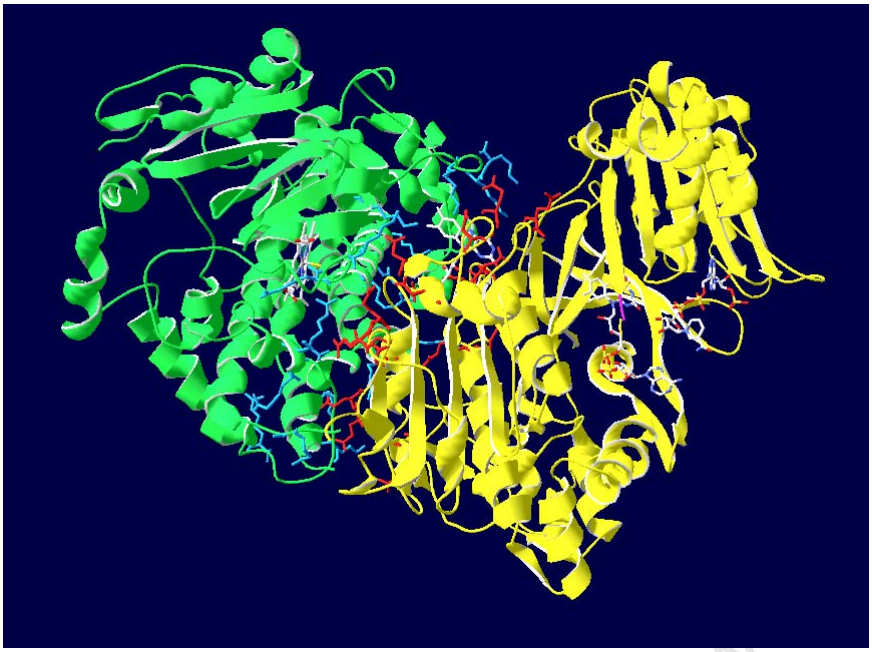


Figure 5.5: Basic CYP3A4 residues (*blue*) and acidic CPR residues (*red*) within 10 Å of the interaction interface of the *in silico* model.

Several potential 'reverse' charge clusters were also identified, i.e. where acidic residues present at the CYP3A4 interface appeared to cluster with basic residues present on the CPR interface. However, these are not described here as no experimental evidence exists in the literature (to our knowledge), which confirms their importance in CYP-CPR interaction.

5.2.3.3 Modelling the final CYP3A4-CPR interaction complex on the array surface

Once the initial CYP3A4-CPR model had been validated, the model was further modified by the addition of the BCCP immobilisation tag to the C-terminals of both proteins. This tag consists of (from the N- to the C-terminus), a glycine-serine linker (GGSGSG), the BCCP domain, a His_6 sequence and a final G-S linker. A known structure of the BCCP domain was used (3BDO), while the G-S linkers and the His_6-tag were initially built onto this structure as loops using the Swiss-PDBViewer software. This software was then used to detect and model the potential secondary structure of these added sequences. The final model of the BCCP tag was then merged with that of the CPY3A4-CPR interaction model and added to the C-termini of both proteins to create the final model of their potential interaction on a two-dimensional surface (see Fig. 5.6).

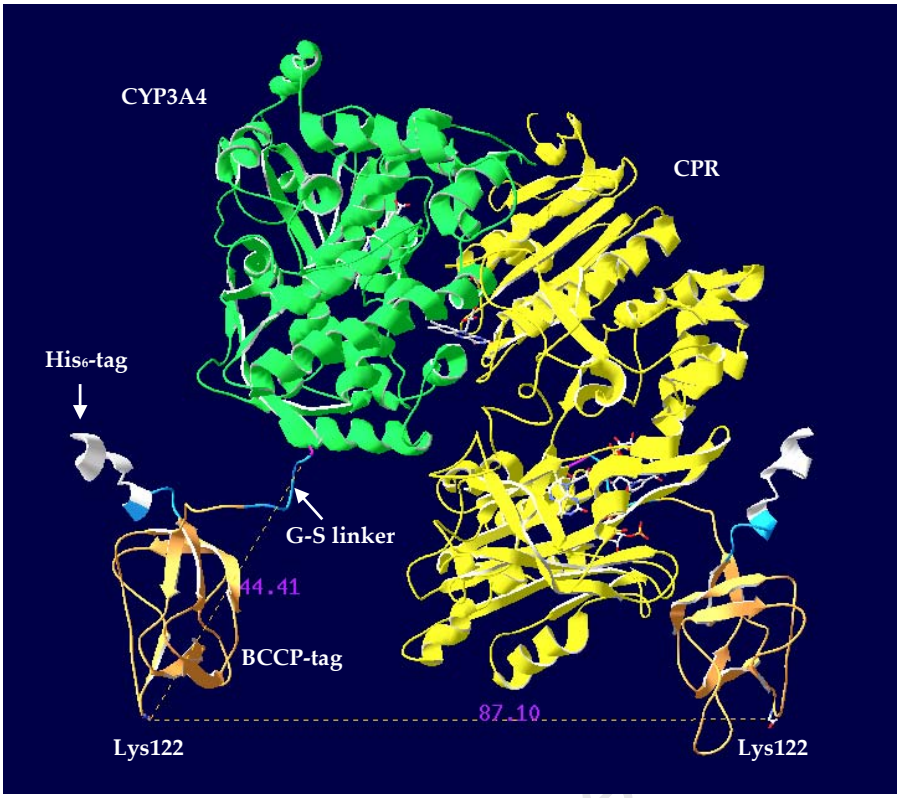


Fig 5.6: Final *in silico* model of CYP3A4 and CPR interaction while 'immobilised' to a horizontal plane *via* the C-terminal BCCP-tags. The BCCP-tag provides a < 50 Å stalk between the surface and the C-termini of both CYP3A4 and CPR (*C-terminal residues shown in purple*), while the distance between adjacent BCCP-tags on the surface is ~87 Å. The BCCP-tag attaches to the surface *via* a single biotinylated lysine residue (Lys122).

As can be seen from Fig. 5.6, the BCCP-tag provides a < 50 Å 'stalk' (depending on the actual loop conformation), which both immobilises the fusion partners indirectly to the surface and directs them away from the surface. *Via* the flexible G-S linker, the BCCP-tag also provides both enzymes with limited, yet presumably beneficial degrees of freedom to move in three dimensions (the streptavidin proteins are themselves also bound to long, flexible polyethylene glycol chains). Thus, despite being much 'bulkier' than CYP3A4, this model suggests that the reductase is able to form a productive complex with the CYP while still remaining away from the array surface. In addition, the C-terminus of the CPR is tucked inwards, towards the centre of the FMN/FAD domain, which provides space for the G-S linker to attach to the reductase while still maintaining the correct relative geometry of the two partner enzymes.

It is also interesting to consider that the success of our approach might also depend to some degree on the fortuitous domain arrangement within the reductase. The FMN domain lies N-terminal to both the FAD and NADPH domains and is connected to the latter domains *via* an ~30

a.a. long linker that is highly flexible (Wang *et al.*, 1997). This flexible 'hinge' - together with a flexible area within the FAD domain (residues 325 – 450) - allows the movement of the FMN domain from an interface with the FAD domain (from which it accepts electrons), outwards towards an interface with the CYP heme domain (to which it then transfers these electrons). For the C-terminally immobilised reductase, the smaller N-terminal FMN domain (~18.5 kDa) is therefore free to carry out this movement without steric constraint, whereas for an N-terminally immobilised reductase, the much bulkier FAD and NADPH domains (~46.5 kDa) would have to perform this movement, as the mobility of the FMN domain would be constrained by the BCCP-tag.

Streptavidin can bind up to four biotin molecules simultaneously, with an average distance between biotin molecules of ~20Å (1SWR.pdb, see www.pdb.org). The above model therefore suggests that the interacting CYP3A4 and CPR enzymes would probably be bound to separate streptavidin molecules on the array surface, as the computed distance between the adjacent enzymes was ~87Å (see Fig. 5.6 above). Binding to separate streptavidin proteins would further increase the relative mobility CYP3A4 and CPR partner enzymes, thereby increasing the probability of the formation of a functional complex.

Whilst it is recognised that this model of the putative array-bound CYP3A4-CPR complex contains many caveats, it does suggest that both the flexibility and mobility provided by the C-terminal BCCP-tags, together with the inherent structure of the reductase, should permit the productive association of the partner enzymes after immobilisation *via* their C-termini. The C-terminal tagged CYP and CPR constructs were therefore chosen for further array construction.

5.2.4 Preparation and testing of home-made streptavidin-coated array surfaces

For all previous array experiments commercial streptavidin-coated slides were used. For further array work 'home-made' streptavidin-coated slides were required so that fresh slides could be prepared as needed, using standardised protocols. Array surfaces coated in streptavidin were therefore prepared and tested. The specificity of binding to these surfaces was investigated using biotinylated BSA and non-biotinylated BSA, while the homogeneity of the coating was tested and compared between batches.

5.2.4.1 Binding specificity

After incubation of the home-made streptavidin-coated slides with Cy3-BSA and Cy3-biotin-BSA, only Cy3-biotin-BSA was observed binding to the array surface (see Fig. 5.7 A). This demonstrates that binding of proteins to these slides is mediated *via* the streptavidin-biotin interaction, with little to no non-specific binding, as expected. Some fluorescence was observed on the Cy3-BSA slides from wash artefacts where buffer dried onto the surface prior to being spun dry. Wash conditions were later optimised to avoid this occurrence.

Initially slides were coated by hand using a pipette and a glass microscope slide as a cover slip. The streptavidin-biotin interaction is known to be formed rapidly (Lesaicherre *et al.*, 2003) and this is clearly demonstrated here by the visible rings that delineate where solution was spotted down prior to placing a glass slide on top for solution spreading. For a more homogenous coating, subsequent batches were prepared using glass microarray Lifterslips (see Fig. 5.7 B). The Lifterslips contain thin supports down their length on each side, are the same width as the array slides and cover all except the bottom 1.5 cm. The solution is aliquoted along the width of the slide along the top and the bottom of the Lifterslip. The solution is then drawn quickly and evenly across the slide surface through capillary action.

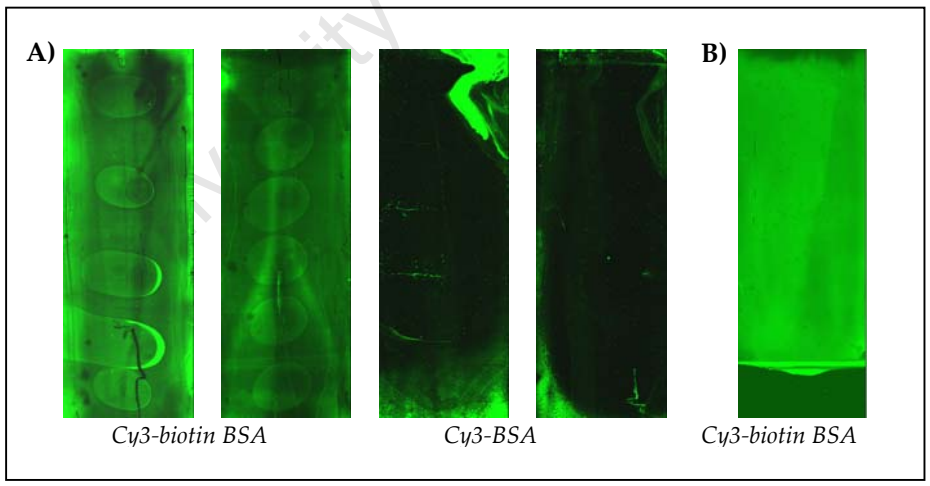


Figure 5.7: A) Streptavidin-coated slides incubated with Cy3-biotin-BSA and Cy3-BSA to test binding specificity. Only the biotinylated BSA bound to the surface, demonstrating that surface binding is mediated by the streptavidin-biotin interaction. These slides were prepared using a pipette and a glass slide as a cover slip. B) A streptavidin-coated slide prepared using a Nunc Lifterslip. This method ensured a more homogenous coating of the surface.

5.2.4.2 Binding homogeneity

Across batches, binding homogeneity within the print region varied between 15-24% (coefficient of variation, CV). For an area of this size (22 x 58 mm), CVs of ≤ 25% are deemed suitable by commercial standards (personal communication, Jonathan Blackburn), and these batches were therefore used for further experimental work. These findings compare favourably to the 19-33% (dependent on fluorophore) intraslide CV seen for the Nexterion H slides in the study by Guilleaume *et al.* (Guilleaume *et al.*, 2005), which used CVs of individual printed spots averaged across the whole slide. For the smaller area (7 x 7 mm) determination however, the results varied widely between various points within the print region. CVs of ≤10% are generally considered to be very good (personal communication, Jonathan Blackburn) and this was achieved for some batches but not all. An unknown level of variation is of course introduced into each lot of slides during commercial coating with the amine-reactive surface. However, the overall binding homogeneity was deemed sufficient for our purposes.

5.2.5 First robotically printed Δ24C-CYP3A4 and Δ43C-CPR enzyme arrays

Using information gained from earlier arrays, the first successful robotically printed Δ24C-CYP3A4 and Δ43C-CPR enzyme arrays were produced. In this experiment various assay parameters were tested. These included the number of stamps per spot, final glycerol concentrations in the reaction mixtures and the molar ratio of Δ24C-CYP3A4:Δ43C-CPR.

5.2.5.1 Effects of printing and reaction buffers

The choice of enzyme printing buffer proved critical for maintaining protein function and interaction throughout printing and incubation on the slide surface. Prior to the fabrication of robotically printed arrays, crude Δ24C-CYP3A4 and Δ43C-CPR hand-spotted arrays were produced by incubating the purified enzymes on the slide surface in the reaction buffer (100 mM potassium phosphate pH 7.4, plus 7.5% glycerol (v/v) to reduce evaporation), rather than in P450 Storage Buffer. Proteins immobilised on these arrays showed little to no turnover (data not shown). Turnover was only observed when enzyme was spotted down in P450 Storage Buffer and this was therefore used as the printing buffer for all subsequent printed arrays.

These results appear contradictory to those obtained in Section 5.2.2, where Δ24C-CYP3A4 spotted down in reaction buffer (no glycerol) yielded substrate turnover using the CuOOH-based

assay format. This is possibly due to these CuOOH-based slides not undergoing a wash step between the first and second rounds of spotting - as was the case for the reductase-based slides described here. It is likely that in the absence of glycerol in the spotting buffer, the proteins undergo denaturation while the slides are dried down.

Initial array-based assays using a final concentration of 7.5% glycerol in the reaction buffer for the second round of printing showed no turnover, probably due to the rapid evaporation of spots (data not shown). This was not observed in the hand-spotted arrays using the same glycerol concentration, possibly due to the smaller surface area/volume ratio in those assays [see Section 5.2.2; similar effect noted in (Jackman *et al.*, 1998)]. For this reason, 15 and 20% final glycerol concentrations were tested on these robotically printed arrays. As observed in the hand-spotted arrays however, increasing glycerol concentrations led to increasing inhibition of turnover (see Fig. 5.8-9 below). For both stamp routines, 15% glycerol was found to yield ~1.3-fold higher turnover compared to 20% glycerol and in both cases the spots did not visibly dry out but remained fluid. 15% glycerol was therefore used for all subsequent arrays.

Schott recommends using 5% glycerol for printing on the Nexterion H surfaces (Schott, 2006), however this is for slides that are processed shortly after printing (i.e. washed/blocked, etc.) and then scanned, rather than assayed for a reaction taking place within the microdroplets. The latter requires that a liquid environment be maintained to allow diffusion of the substrate and the metabolites and to facilitate CYP-CPR interactions. Our finding of 15% glycerol being optimal for maintaining longer-term droplet fluidity for enzyme arrays corresponds to that found in a different study by Doong *et al*. (Doong and Shih, 2006).

5.2.5.2 Effect of the intermediate wash step between print rounds

The choice of the wash buffer used for washing the slides between the first and the second print rounds was also found to be critical. Initial experiments in which arrays were washed with PBST (1 x PBS pH 7.4 with 0.1% Tween 20), a detergent that is used for many protein array applications because it assists in preventing non-specific binding (Jonkheijm *et al.*, 2008;Schott, 2006), resulted in little to no observable turnover (data not shown). It is possible that this is due to either the buffer not being optimal for maintaining CYP protein folding during the wash step or the active promotion of unfolding of these enzymes by the detergent (Katakam *et al.*, 1995;Kreilgaard *et al.*, 1999).

In addition, when using PBST as the wash buffer the subsequent spots printed down during the second round tended to run into each other (this was particularly noticeable for hand-spotted arrays) and exhibited poor spot morphology. This is likely due to decreased surface tension between the buffer and the slide surface resulting from the presence of the detergent (water: 75.64 dyn/cm, Tween 20: 38.9 dyn/cm). Therefore, for this and further array experiments, only P450 Slide Wash Buffer (P450 Storage Buffer with 0/5% glycerol and no DTT) was used for the wash step. For this experiment, all glycerol was excluded from the Slide Wash Buffer due to concerns about high background; however, all further experiments included 5% glycerol (see Section 6.2.1.6).

High background was noted for the upper half of the slide despite the wash step (see Fig. 5.9 below). This was likely due to unbound enzyme washing off and then rebinding elsewhere on the streptavidin-coated surface. In subsequent arrays, a biotin-based blocking step was introduced to counter this problem (see Section 6.2.1.6).

5.2.5.3 Effect of the number of stamps per spot

For both glycerol concentrations, the arrays printed with three stamps per spot for both print rounds gave superior turnover (≥ 2-fold more) compared to those printed with one stamp per spot for both print rounds (see Fig. 5.8-9 below). The simplistic interpretation here is that a single stamp does not provide sufficient protein to saturate the surface. Since a certain critical density of enzymes on the slide surface must be present to ensure that electron transfer can occur between the reductase and the CYP enzymes, the higher loading per spot provided by increasing the number of stamps possibly helps to increase the number of these associations.

This conclusion appears confounded however, when comparing the turnovers of the 0.25:0.25 μM and 3.09:3.09 μM CYP3A4:CPR enzyme concentrations. Both of these concentrations show ≥ 2-fold increase in signal when three stamps per spot are used and yet they show similar relative turnover levels for both the one- and three-stamp print routines (see Fig. 5.8; note: used a 100 μm feature-finder across all spots for analysis). However, the increase in protein loading within the spot was also associated with a visible increase in the diameter of the spots by up to ~30 μm for the brightest spots (i.e. a change from 90 to 120 μm for the 8 μM substrate reactions). This increase in spot diameter resulted in an increase in the surface area available for enzyme binding by up to

two-fold. Since a constant feature-finder diameter of 100 μm was used for analysis across all spots however, this increase in spot diameter can only partially explain the increase in signal measured. Another possible explanation for the increased signal observed for the three-stamp routine, is that by increasing the number of stamps per spot the surface area to volume ratio of the spots decreased compared to those formed by one stamp. This would maintain the aqueous environment of the proteins immobilised below the spot for longer periods during the print run, by further delaying the drying down of spots through evaporation. This of course would be beneficial for maintaining protein stability and activity on the array surface prior to assay.

The above results suggest that 250:250 nM is enough protein to adequately saturate the surface and that by increasing the number of stamps per spot, the signal not only increases linearly with the concomitant increase in surface area (due to the linear increase in the number of molecules of enzyme bound), it also increases in density within the spot as well. While the origin of the first effect appears obvious, the explanation for the latter effect is not completely clear.

5.2.5.4 Effect of the CYP3A4:CPR ratio

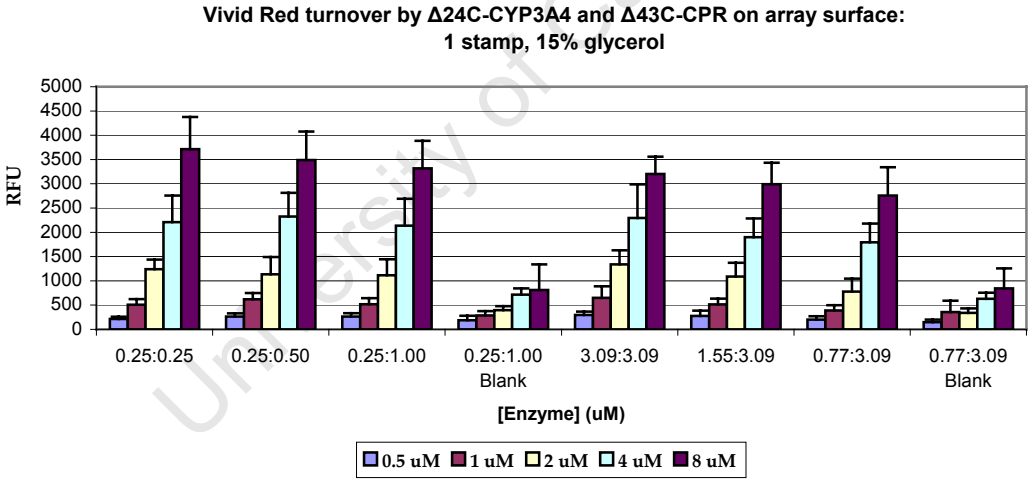
The biological ratios of CYP to reductase in liver and microsomal preparations range from 10:1-20:1 CYP:CPR (Backes and Kelley, 2003), while artificial, reconstituted systems often use greater than 1:1 CPR:CYP ratios (Miwa *et al.*, 1979). The influence of this ratio on the rate of substrate metabolism is reflected in the higher turnover numbers obtained with artificial, reconstituted systems; the higher ratio of the reductase increases the probability that when a substrate molecule enters the CYP active site, a reductase enzyme will be close enough (by chance diffusion within the membrane) for reduction to occur. However, in the case of the P450 Biochips developed in this thesis, both the CYP and the reductase enzymes are bound to the surface and are not free to diffuse. The only freedom of movement allowed is that prescribed by the length and flexibility of the BCCP-tag and that of the underlying PEG surface. It therefore seems reasonable to suppose that lower ratios of reductase to CYP might be required.

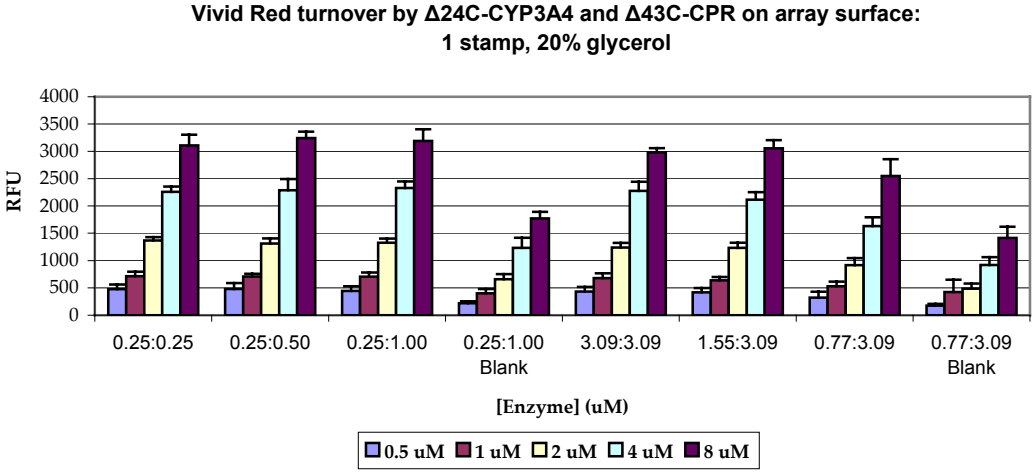
This hypothesis is supported when comparing the turnovers of the various CYP3A4:CPR ratios (1:1, 1:2, and 1:4) on the biochip. No increase in turnover is observed when increasing the amount of reductase relative to the CYP (see Fig. 5.8 below). In fact, as the reductase increases, in most reactions a slight decrease in turnover is apparent. This suggests that equimolar amounts of the

CYP and the reductase provided the optimal ratio for association on the array surface and that the frequency of productive CYP-CPR association is not rate limiting here.

1:1 monomeric stoichiometry is in fact the ratio at which a functional association is known to occur, with a K_d of ~ 50 nM (Miwa *et al.*, 1979). This stoichiometry is also reflected in the structure of the most enzymatically active CYP known, BM3, a bacterial CYP that consists of one heme domain and one reductase domain joined in a single polypeptide (Sevrioukova *et al.*, 1999b). Of course we cannot be certain that the absolute amounts of functional CYP protein and CPR protein are equivalent on the array surface as total protein contents were used in these assays rather than total P450 and flavin contents. Δ

The slight decrease observed upon increasing the amount of reductase protein relative to CYP is possibly the result of a subsequent decrease in the number of sites available for the CYP protein to bind within the spot, thereby decreasing the total amount of CYP available for association with the reductase.





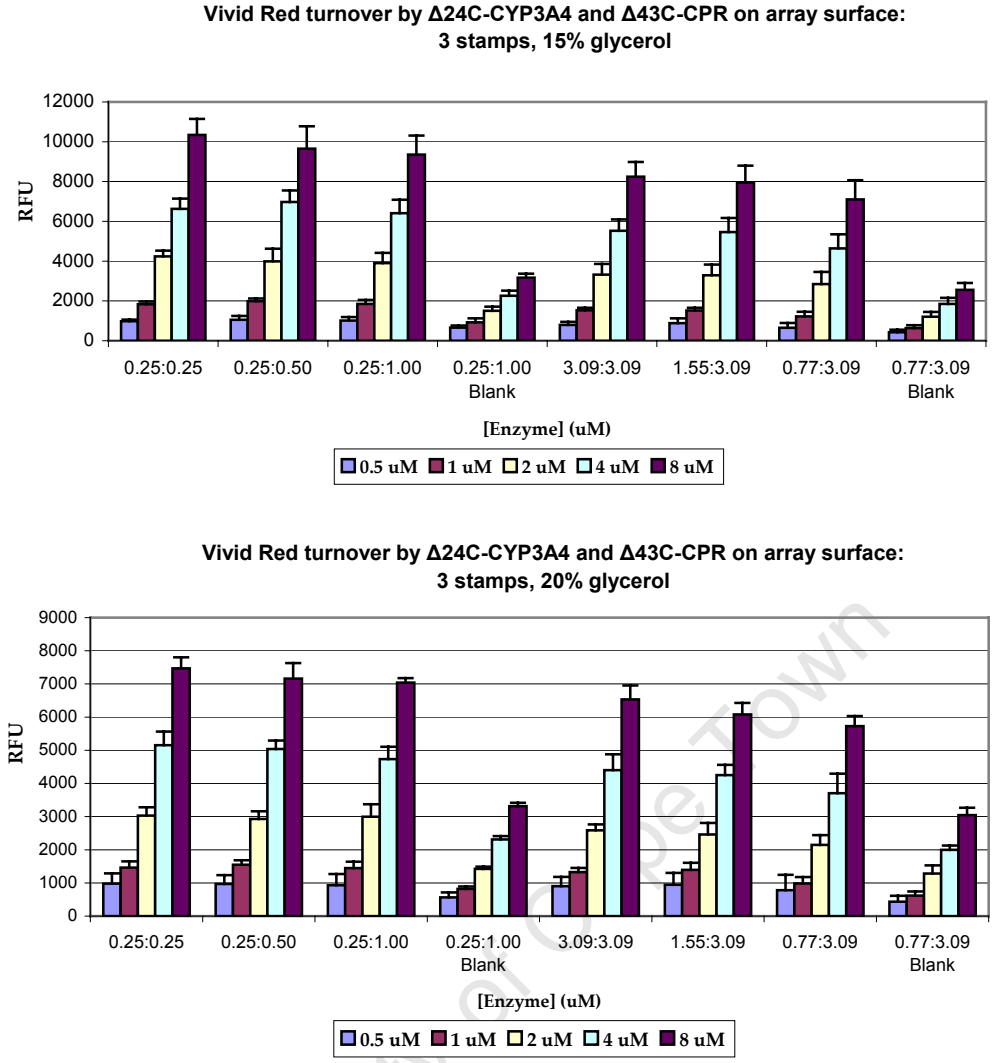


Figure 5.8: Optimisation of Vivid Red metabolism on a streptavidin-coated array surface by Δ24C-CYP3A4 and Δ43C-CPR. Various print parameters (1 stamp per spot vs. 3 stamps, 15% glycerol vs. 20%) and various assay parameters (different CYP:CPR ratios, different enzyme concentrations) were tested as indicated. Note: Blank = negative controls (enzyme and substrate only). Error bars represent standard deviations for 27 replicates for reactions and 9 replicates for negative controls.

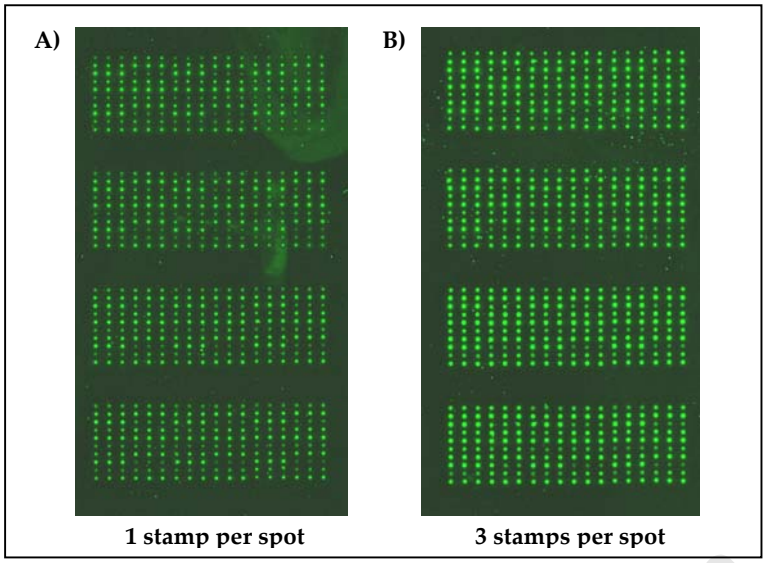


Figure 5.9: Scanned fluorescence image of the Δ24C-CYP3A4 and Δ43C-CPR enzyme arrays showing Vivid Red turnover. The three stamps per spot print routine (B) yielded visibly higher turnover than the one stamp per spot routine (A).

5.2.5.5 Choice of cofactor

NADPH is roughly four times more expensive than the oxidised form, so traditional solution-phase assays typically use the latter together with the cofactor regenerating system (i.e. glucose 6-phosphate and glucose 6-phosphate dehydrogenase). However, due to the much lower volumes required when carrying out our array-based assays (working on the assumption that a spot's volume is constrained by its visible diameter of 150 μm and that it forms a full hemisphere, the volume of each reaction is calculated at only 0.884 nl), working directly with NADPH not only becomes a cost-effective option, but also reduces the assay complexity. It is also likely that using saturating amounts of NADPH (greater than twice the substrate concentration) increases reaction efficiency compared to the regenerating system, as reducing equivalents are readily available and CYP-CPR associations on the array surface are not potentially disrupted by the presence of another enzyme (glucose 6-phosphate dehydrogenase MW = 128 kDa).

5.2.6 Determination of the optimal PMT for scanning CYP enzyme arrays

Before any quantitative CYP kinetic assays could be carried out, the linear range of the Tecan fluorescence microarray scanner had to be determined as well as the optimal PMT setting for capturing data in the signal range relevant to such assays. It is critical that the optimal PMT be determined as otherwise signal measurements could fall prey to bias by the non-linearity of the

scanner. All expected signal levels on an array should fall within the dynamic range of the scanner and should be obtained using an optimal PMT setting that does not produce spots with saturated pixels or signal within range of the background noise.

5.2.6.1 Optimal PMT

The resorufin range of 1 nM to 500 nM was determined to be the most relevant to future robotically printed assays using calculations similar to that in Section 5.2.1. This range was therefore used to determine which PMT setting was optimal. Linear regression plots were generated for each PMT setting and used to determine whether the above range of resorufin concentrations was contained within the linear range of detection of that particular setting. PMT settings of 200 and above gave poor correlations (R^2 values) as the signal levels in the 500 nM spots began to saturate at these settings. PMT settings of 150 and below also gave poor correlations as signal levels for the lower resorufin concentrations (e.g. 10 nM and below) were dropping within range of the background signal (SNR < 3, where SNR is the signal to noise ratio, which is equal to the background corrected signal divided by the standard deviation of the background). A PMT of 175 was therefore determined to be optimal as it gave excellent correlation over the 1 - 500 nM range (see Fig. 5.10 below).

5.2.6.2 Dynamic range

The manufacturers of the Tecan scanner state that it has a dynamic range over five orders of magnitude. Although signal was observed across the 10,000-fold dilution series, the quality of the data (i.e. SNR > 3, saturation < 5%) was only satisfactory over three orders of magnitude (see Fig. 5.10 A). However, this was deemed sufficient for kinetic assays, which usually use substrate concentrations within this range.

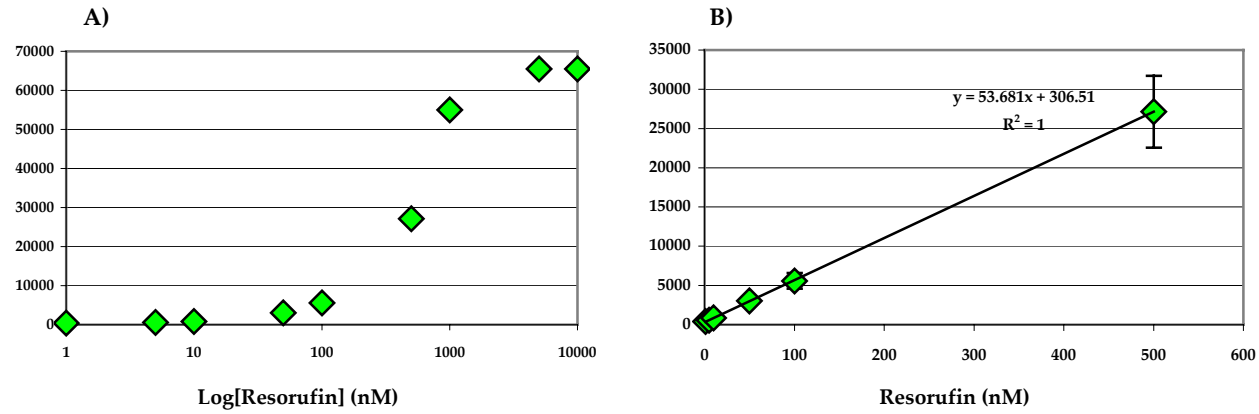


Figure 5.10: Optimisation of the PMT used for CYP enzyme arrays. A) Log plot of signal obtained at 175 PMT for a 10,000-fold resorufin dilution series. B) A resorufin standard curve generated at 175 PMT plotted for concentrations within the linear range.

5.2.6.3 Sample carryover

Printing for these arrays was carried out by field and not by sample, so all samples of Field 1 were printed before the same samples were printed in a replicate Field 2 further down the slide (see Fig. 5.11). Carryover of sample was observed in the buffer-only spots of the second field of subarrays, as the signal levels of these spots were higher than those in the first field. This signal was also seen to decrease with each subsequent spot within the triplicate set, further confirming it as sample carryover. The pins are washed between each unique sample visit, however this wash step was insufficient to prevent carryover from the last sample printed in the first field (10 μM resorufin) to the first sample printed in the second field (the buffer-only spots). In order to mitigate this effect, future arrays contained six buffer spots between each new field, and substrate concentrations were printed in order of increasing concentration. In addition, potential carryover between spots containing different substrate concentrations was checked during data analysis by observing the signal patterns obtained and the levels of signal variation.

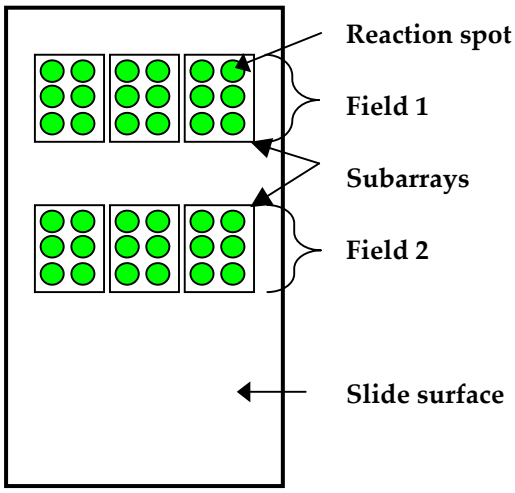


Figure 5.11: Simplistic representation of a typical array print layout. The printing area is divided into fields, each of which consists of an identical number of subarrays (three are depicted here). Each subarray within a field is printed by a different pin. For this experiment, three pins were used to print both fields.

It is of course possible to increase the number and duration of the pin washes, however the washes take up the majority of the print run time and therefore already significantly effect the amount of time from the start of the print run to the point of scanning. This is especially important considering that the enzyme might unfold over time (as the print run and scanning currently occur at RT) and that for kinetic assays it is necessary to capture the data while enzyme turnover of the substrate is still linear with time. It was therefore concluded that the final printing parameters used were good enough for further work.

5.3 CONCLUSIONS

This chapter has described the successful development and optimisation of a novel CYP array format that allows functional interaction between cytochrome P450 and its reductase partner *via* direct, non-random immobilisation of membrane-free, purified enzymes on a 2D surface.

Prior to initiating array experiments, various surface chemistries and immobilisation methods were considered through reviewing the literature [including studies carried out previously by our group's principal investigator (Boutell *et al.*, 2004)] and the method most likely to work for our application was chosen. Several parameters and protocols throughout the array fabrication process were tested, such as various buffers, printing parameters, enzyme concentrations, partner enzyme ratios, slide preparation protocols, etc. The various CYP and CPR constructs were also

examined to decide which pair was the most plausible theoretically. These choices were validated by successful catalytic turnover in this novel format.

This data demonstrates for the first time catalytic turnover after the direct immobilisation of cytochrome P450 and its reductase partner onto an array surface, using only their cytosolic domains and without the use of lipids. In contrast to all other current assay formats and despite the challenges posed in achieving electron transfer between the two enzymes on a 2D array surface, this has been successfully realised here for mammalian CYP enzymes in a direct and optimal one-to-one functional relationship through the design of this platform. In addition, the method for fabricating these arrays is much faster than that allowed by the membrane encapsulation methods used thus far in other studies – it only takes as long as a print run, which can be less than half an hour when a full pinhead is used. The number of possible reactions per array is also much higher. With a working surface area per slide of 58 x 22 mm, our array format is capable of at least 4200 reactions per slide when producing 150 μm features with 350 μm centre to centre spacing, with a resulting feature density of ~ 329 spots/cm^2. Both of these attributes are significant improvements in terms of high-throughput capacity. This format also provides greater ease of multiplexing and platform extension to other detection methods (e.g. mass spectrometry) due to the greater facility with which metabolites can be retrieved and solutions exchanged.

These initial experiments have shown that CYP enzyme assay miniaturisation is feasible in a protein microarray format and that it has several advantages over traditional solution-phase assays when considering its use as a tool for high-throughput compound screening. Reagent and protein consumption are obviously much lower (5 μl of sample would yield over 1800 reactions using our 3-stamp print regime), which becomes an important consideration when screening thousands of compounds (carrying out a CYP inhibition assay for example can require over 100 reactions, including triplicates, per compound). More importantly, with the basic format in hand, we were able to generate large numbers of high quality data points over several different parameters within a single experiment. Significant time- and cost-savings are therefore implicit in this novel assay format. In addition, since various different parameters or assays can be carried out simultaneously under the same conditions, potentially complicating factors that often effect experiments carried out sequentially are eliminated.

5.4 MATERIALS AND METHODS

5.4.1 Determining the detection sensitivity required to measure CYP3A4 activity on-chip

5.4.1.1 Calculations

In the solution-phase cumene-hydroperoxide assay, a k_{cat} of ≈ 0.005 min^{-1} was obtained for Vivid Red using Δ24C-CYP3A4. Hence, during a 20 min reaction using 500 nM enzyme, ~50 nM resorufin will be produced at substrate concentrations twice that of K_m (~2.8 μM Vivid Red). The slide scanner must therefore be able to detect < 5 x 10^{-14} moles resorufin in a 1 μl volume.

5.4.1.2 Preparation of slide hand-spotted with resorufin dilution series

A 10 mM resorufin solution was prepared by dissolving the resorufin sodium salt (# R3257, Sigma Aldrich) in 100% acetonitrile. From this, three 10 μM dilutions were prepared in 100 mM potassium phosphate buffer (pH 7.4) and used to check the actual concentration with a spectrophotometer at A_{571} using an extinction coefficient of 54 x 10^3 $M^{-1}.cm^{-1}$. Using the measured concentration, a dilution series of 10 nM – 5.12 μM was prepared in phosphate buffer. Replicate rows of 1 μl spots were deposited with a pipette onto a poly-lysine coated slide (#406/1078/00, BDH Laboratory Supplies, UK). The slide was then scanned using the Tecan LS Series Laser Scanner (Tecan Group Ltd., Switzerland) in the 'green' channel (laser λ_{ex} = 532 nm) at various PMT settings.

5.4.2 First hand-spotted CYP3A4 enzyme arrays using cumene hydroperoxide

5.4.2.1 Preparation of enzyme array slides

For the first experiment, two commercially prepared streptavidin-coated slides were used (Streptavidin-OptArray, Accelr8 Technology Corporation, US). The slides were pre-equilibrated for 1 hr at RT before opening. The first slide was hand-spotted with 1 μl spots of enzyme pre-mixes containing purified CYP enzyme and 100 mM potassium phosphate buffer (pH 7.4) only, while the second slide was hand-spotted with 1 μl spots of enzyme pre-mixes containing both CYP enzyme and Vivid Red substrate in buffer (see Table 2.3 below). Immediately afterwards, 1 μl spots of the second round of pre-mixes were overlaid onto the first spots on both slides. On the first slide, these second round pre-mixes consisted of Vivid Red and 1 mM CuOOH in buffer, while for the second slide the second round pre-mixes consisted of 1 mM CuOOH only in buffer (the final concentration of CuOOH in all reactions was therefore ~0.5 mM). The slides were incubated at RT for 30 min and then scanned at 140 PMT.

For the second experiment, two hand-spotted Δ24C-CYP3A4 enzyme arrays were similarly prepared using the second assay format mentioned above. The effect of two different final concentrations of glycerol in the reactions was tested and various negative controls were included. The first slide contained spots with a final glycerol concentration of 7.5% while the second contained spots with a 10% final concentration. These slides were scanned at 125 PMT due to signal saturation occurring at higher settings.

Experiment and slide no.	First round printing pre-mix	Second round printing pre-mix
Experiment 1, Slide 1	Enzyme only	Substrate and CuOOH
Experiment 1, Slide 2	Enzyme and substrate	CuOOH only
Experiment 2, Slides 1 and 2	Enzyme and substrate	CuOOH only

Table 5.3: Outline of the pre-mixes used for the first and second print rounds of the hand-spotted arrays.

5.4.2.2 Data analysis

Data from scanned slides was captured using GenePix Pro 6.0 microarray analysis software (Axon Instruments Inc., US). The median fluorescence of each spot was captured using 600 μm diameter circular feature finders. The median background was measured within 3 feature diameters from the centre of each feature, excluding the first 6 pixels surrounding the feature to avoid high signal from any irregular spot morphology. The median background was then subtracted from the median fluorescence for each feature.

5.4.3 Choice of CYP3A4 and reductase constructs for fabricating the array

5.4.3.1 Superimposition of CYP3A4 and CPR onto BM3 bacterial template

A ligand-free CYP3A4 structure (1TQN) and a rat CPR structure in the open conformation (3ES9) were superimposed consecutively onto the BM3 structure (1BVY) using the Iterative Magic Fit function (carbon alpha atoms only) of the Swiss-PDBViewer software [www.expasy.org/spdbv; (Guex and Peitsch, 1997)]. After fitting, the Improve Fit function was used. The fit was evaluated by the final RMSD values obtained and by the parallel and close stacking of the heme and FMN cofactors, as well as by the complete superimposition of the I-helix of the heme domain. After superimposition the BM3 layer was discarded leaving the CYP3A4/CPR superimposition in place. Electrostatic potential was mapped onto the protein surfaces using the Compute Electrostatic Potential tool of SPDBV.

5.4.3.2 Further validation of the CYP3A4-CPR interaction model using charge cluster pairings

Swiss-PDBViewer was used to delimit all residues within certain user-defined cut-off distances (4, 6, 8 and 10 Å) between the CYP3A4-CPR interaction interface. These residue lists were refined to include only acidic and basic residues, which were then modelled onto the protein interface. Identification of putative charged cluster pairings was carried out by visual inspection of the protein interface and using the distance measurement tools provided by the software.

For homologous residue identification, pairwise sequence alignments of CYP3A4 and rabbit CYP2B4, as well as human and rat CPR were carried out using the ClustalW program (Thompson *et al.*, 1994). Structure alignments of the same proteins were also carried out using the Swiss-PDBViewer Iterative Magic Fit function.

5.4.3.3 Modelling the final CYP3A4-CPR interaction complex on the array surface

A known BCCP structure (3BDO) was modified by the addition of a missing alanine residue at the N-terminus to compare with the full sequence used in our constructs. A G-S linker was then built onto the N-terminus of this structure by the sequential addition of each residue. The Build Loop function was then used to join these added residues to the BCCP backbone. Similarly, a G-S linker and a His_6 sequence were built onto the C-terminus of the BCCP structure. Any extraneous sequence arising from the plasmid-encoded residues prior to the stop codon were also included. The Detect Secondary Structure function was then used to detect potential α-helical, β-sheet and random coil structure within all the added sequences (the whole structure was used for this computation). This modelled BCCP-tag structure was then merged with the CYP3A4-CPR interaction model such that the BCCP-tags could be attached to the C-terminus of both enzymes while maintaining an approximately 90° angle contact with an arbitrary horizontal plane. (Note: see Appendix I for the final protein sequences for Δ24C-CYP3A4 and Δ43C-CPR used for the model.)

5.4.4 Preparation and testing of home-made streptavidin-coated array surfaces

5.4.4.1 Preparation of streptavidin solutions

A vial of streptavidin lyophilisate (#S0677, Sigma Aldrich) was equilibrated to RT. The lyophilisate was then dissolved in water to ≤ 10 mg/ml and the reconstitution was allowed to proceed for 30 min. The solution was divided into 1 mg aliquots and stored at –20°C. 1 mg/ml

working solutions were prepared in 1 x Spotting Buffer (final concentrations of 100 mM KCl, 0.01% Triton X-100, 25 mM HEPES pH 8.5) as needed. Care had to be taken that the Spotting Buffer had a pH > 8.3, as this was required for the protein amine groups to remain mostly unprotonated and therefore reactive towards amine-reactive groups on the slide surface.

5.4.4.2 Preparation of streptavidin-coated slides

The sealed package containing Nexterion H slides (#1070936, Schott, Germany) was removed from –20°C storage and equilibrated to RT for 1 hour to minimise condensation. A fresh 1 mg/ml streptavidin solution was prepared and spread across the slide surface. [Schott recommends 0.1-1 mg/ml probe protein concentration to ensure sufficient loading (Schott, 2006).] For the first batch this was carried out by hand using a pipette and a glass microscope slide. Subsequent batches were prepared using glass microarray Lifterslips (#231650, Nunc, USA). The slides were then incubated for 1 hr at RT in a humidified chamber. After incubation, the coverslips were removed and the slides were incubated on a platform shaker at RT for 1 hr in 50 mM ethanolamine in 50 mM potassium phosphate buffer (pH 8.0), to quench any remaining amine-reactive groups on the surface. The slides were then washed 3 x 5 min in P450 Slide Wash Buffer (20 mM potassium phosphate pH 7.4, 0.2 mM EDTA*) and 1 x 5 min in dH_2O with shaking. Each slide was then placed in a 50 ml Falcon tube and centrifuged at 1000 rpm for 5 min at 20°C to spin dry. Coated slides were placed into slide boxes which were sealed in Ziploc bags and stored at -20°C. (*Note: the P450 Slide Wash Buffer was later amended to include 5% glycerol, see Section 6.2.1.6.)

5.4.4.3 Preparation and reactivity testing of biotin-NHS ester solution

The biotin-NHS ester lyophilisate (#14405, Sigma Aldrich) was fully equilibrated to RT prior to dissolving it in DMSO to form a 35 mg/ml solution and aliquots were stored at -20°C. The ester is prone to hydrolysis, therefore, prior to use a reactivity assay was carried out to ensure that enough intact biotin-NHS remained. The NHS group absorbs strongly at 260-280 nm when the biotin group has been released by hydrolysis. The reactivity of the ester can therefore be determined by comparing A_{280} of the ester before and after base hydrolysis by 100 mM NaOH using the appropriate blank solutions for each (Miron and Wilchek, 1982). An increase of ~1 OD_{280} unit was observed after base hydrolysis of the biotin-NHS ester, demonstrating that a sufficient proportion of the compound consisted of the intact biotin-NHS ester rather than its hydrolysis products. The ester was therefore suitable for labelling proteins with biotin.

5.4.4.4 Preparation of Cy3-BSA and Cy3-biotin-BSA

500 µl of 10 mg/ml BSA in PBS pH 7.4 was aliquoted into two eppendorf tubes. 0.5 mg of Biotin-NHS ester was added to one of the tubes at a 1 mg/ml final concentration. The tubes were incubated for 4 hrs at RT on a platform shaker to allow the biotin-NHS to react with amine groups on the BSA in the one aliquot. A total of 2 mg each of the biotin-BSA and unbiotinylated BSA from each of these tubes (still in the original buffer and concentration) were added to two separate foil-wrapped vials of Cy3 dye powder (#PA23001, GE Healthcare). The vials were vortexed to ensure the dye was completely dissolved and the vials were then incubated for 1 hr at RT with vortexing every 10 mins. After incubation, unbound Cy3 dye was removed by applying the samples to SigmaSpin Post-Reaction Purification Columns (#S5059, Sigma Aldrich) as per manufacturers instructions for Cy-labelled DNA. Aliquots of Cy3-BSA and Cy3-biotin-BSA were stored at -20°C.

To ensure that adequate Cy3-labelling had been obtained, spectrophotometry assays were carried out to determine the dye-to-protein (D/P) ratio, which, according to the manufacturers, should optimally be ≥ 2. This is particularly important in the case of the Cy3-biotin-BSA, which was initially reacted with biotin prior to dye labelling, as both labels will react with amine groups on the protein surface. This ratio can be calculated by measuring the protein's absorbance at A_{280} and the dye's absorbance at A_{552} (8% of the dye's A_{552} reading is then subtracted from the A_{280} reading to correct for the dye's absorbance at A_{280}). However, the A_{280} measurement requires large quantities of the protein sample unless microvolume cuvettes are available. Therefore, a Bradford standard assay was carried out at A_{595} instead to determine the BSA concentration. No absorbance of the dye at A_{595} was observed and no correction was therefore required. Using a MW of 66.43 kDa for BSA and an extinction coefficient of 0.15 $\mu M^{-1}.cm^{-1}$ for Cy3, the concentrations of the BSA and the dye were determined and the D/P ratio was calculated (molar dye concentration/molar BSA concentration). A D/P of 8.05 was observed for Cy3-BSA, while a D/P of 5.63 was found for Cy3-biotin-BSA. As expected, the D/P was lower for the biotinylated BSA. Despite this, the labelling efficiency for both samples was more than adequate for further use.

5.4.4.5 Testing of binding specificity to streptavidin-coated slides

The slides were removed from -20°C storage and equilibrated to RT for 1 hr. Cy3-BSA and Cy3-biotin BSA were diluted 1:1250 (6 ng/ml) in 1 x Spotting Buffer. Two slides each were incubated in Cy3-BSA and Cy3-biotin-BSA for 30 min at RT on a platform shaker. The slides were then washed

3 x 5 min in PBST (0.1% Tween-20 in 1 x PBS pH 7.4) and 1 x 5 min in dH_2O with shaking. Slides were spun dry at 1000 rpm for 5 min and scanned at 120 PMT at 10 μm resolution.

5.4.4.6 Testing of coat homogeneity across slide surface

After each fresh batch of streptavidin-coated slides was prepared, a batch quality control test was performed to determine whether the homogeneity across the slide surface was adequate. One to two slides from each batch (depending on batch size) were incubated with Cy3-biotin-BSA and then washed, spun dry and scanned (as described for the binding specificity test above). Using GenePix Pro measuring tools, a rectangular region of 22 x 58 mm (the active area of the slide) was drawn. The average and median pixel fluorescence and the standard deviation across this area were then calculated. The same measurements were also performed for a rectangular region of 7 x 7 mm to determine homogeneity across a smaller area.

5.4.5 First robotically printed Δ24C-CYP3A4 and Δ43C-CPR enzyme arrays

5.4.5.1 Preparation of assay pre-mixes

For the first print round enzyme pre-mixes were prepared. These consisted of the following different CYP3A4:CPR ratios prepared in P450 Storage Buffer: 0.25:0.25 μM, 0.25:0.5 μM, 0.25:1 μM, 3.09:3.09 μM, 1.55:3.09 μM, and 0.77:3.09 μM. Purified proteins prepared as described in Sections 2.4.2.1-3 were used. For the second print round, two sets of substrate and cofactor pre-mixes were prepared in 100 mM potassium phosphate buffer (pH7.4) with either 15% or 20% final glycerol concentrations. A range of 0.5-8 μM Vivid Red substrate was used together with 1 mM NADPH. NADPH is unstable in solution and was therefore prepared fresh every two weeks at 30 mM in 100 mM potassium phosphate (pH7.4) and stored at 4°C (Wu *et al.*, 1986). Negative control samples consisting of enzyme and substrate but excluding the NADPH were prepared for each substrate concentration. The pre-mixes were then aliquoted into a 384-well Genetix source plate (#X7022, Genetix Ltd., UK) for printing. These microtitre plates are especially designed for microarray use as their conical-shaped wells allow the maximum utilisation of sample (as little as 5 μl sample per well can be used). The pattern of sample aliquoting into the source plate is determined by the desired pattern of samples on the slide and the number of pins used in the pinhead.

5.4.5.2 Printing parameters

The slides were printed with a Qarray² System robotic microarrayer (Genetix Ltd.) using 300 μm, solid, flat-tipped stainless steel Telechem pins. The print run was carried out at ambient temperature (~25°C) and humidity (~40%). A sample inking time of 500 ms was used. Both one and three stamps per spot were tested. The pins were washed between unique samples for 60 sec in water, 10 sec in ethanol, and dried for 11 sec. Prior to starting a print run, a datum pointing test was carried out using an old slide and buffer only to ensure that the pins were correctly adjusted to the height of the slides used (for example, when using the streptavidin-coated slides and four pins only, the Z-axis of the pinhead had to be routinely adjusted to -1250 μm). In earlier attempts, incorrect adjustment resulted in uneven protein loading across the slide surface due to slight surface height variations. This was evidenced by high signal variation between spots printed in different sub-arrays and different fields across the slide, while intra-sub-array variation was much lower.

5.4.5.3 Printing procedure

One streptavidin-coated slide was removed from storage and equilibrated to RT prior to use. On the upper third of the slide all spots were printed with only 1 sample stamp per spot, while on the middle third of the slide, 3 stamps per spot were used. In both sections four fields were printed, two each respectively for 15% and 20% final glycerol concentrations.

Two overlaying print runs were carried out. The first print run spotted down the enzyme pre-mixes. After this print run completed, the enzymes were incubated on the slide surface for 30 min at RT. The slides were then washed 3 x 5 min in P450 Slide Wash Buffer (no glycerol) and spun dry. The second print run overlaid the first spot positions with the substrate and cofactor pre-mixes. Obviously this required that the slide be placed in exactly the same position on the array bed for both print runs; this was facilitated by the slide placeholders present on the print bed. The reactions were allowed to proceed for 1 hr at RT after which the slide was scanned at 150 PMT.

5.4.5.4 Data analysis

As for Section 5.4.2.2, except that the diameter of the feature sizes were adjusted to 100 μm.

5.4.6 Determination of the optimal PMT for scanning CYP enzyme arrays

A 0.1 nM – 10 μM dilution series of the resorufin standard was prepared in 100 mM potassium phosphate (pH 7.4) with 15% final glycerol in a manner similar to that described in Section 5.4.1.2. The resorufin dilution series was then printed onto a streptavidin-coated slide under the same conditions as for the previous arrays, however 3 stamps per spot were used for all samples. A set of triplicate buffer spots was printed at the start of each subarray as a blank and to check for sample carryover by the pins. The slides were then scanned at a range of PMTs between 70 and 255. Data capture and analysis were carried out as before, except that only spots with ≥ 90% of pixels with signal greater than the median background plus two standard deviations were included in data analysis to remove any potential background interference.

CHAPTER SIX

Using the P450 Biochip to generate initial kinetic data

6.1 INTRODUCTION

While the Michaelis-Menten kinetic model adequately describes a substantial number of *in vitro* CYP kinetic data sets and has been used for many years to quantify compound clearance and inhibition, it has become increasingly recognised that a large number of substrates do not conform to this hyperbolic kinetic model (Houston and Galetin, 2005;Tracy and Hummel, 2004). In the past, these 'aberrant' kinetic profiles might have been dismissed as being a result of experimental error, or simply forced onto the Michaelis-Menten model resulting in a poor 'fit' of the data. However, we now know that unlike many other enzyme systems, the cytochrome P450s routinely display atypical kinetic behaviour. These observations have also led to the realisation that simple substrate substitution and inhibitor reciprocity in CYP assays are not possible - contrary to what was often assumed before. Most reproducible cases of atypical kinetics observed so far involve the CYP3A4 isoform, however other CYPs have also been implicated and include the CYPs 1A2, 2B4, 2B6, 2C8, 2C9 and 3A5 (Atkins, 2004;Korzekwa *et al.*, 1998). While we still lack a full understanding of the mechanistics of all atypical kinetic phenomena, at least partial explanations are becoming available for certain cases.

These atypical phenomena cannot be readily described using the Michaelis-Menten single site model and three types of approaches have therefore been adopted (Houston and Galetin, 2005). First is the 'naïve' approach of using the Michaelis-Menten model regardless of the kinetic behaviour observed, either by forcing the data onto a hyperbolic curve or by truncating data points. Second is the 'empirical' approach, which, for example, employs the Hill equation to model autoactivation, or uncompetitive inhibition equations to model substrate inhibition. This approach is useful for the preliminary analysis of data but provides no mechanistic information about ligand interactions. The final method is the 'mechanistic' approach, which involves the use of multisite kinetic models. These multisite models are derived using the same rapid equilibrium/steady state assumptions as the Michaelis-Menten model and describe atypical phenomena in greater detail and with better accuracy than the first two approaches. There are many studies demonstrating that accurate estimation of substrate turnover can be achieved using these models, supporting the underlying hypothesis that two or even three substrate binding sites are present on the CYP enzymes. Although more complex than the first two approaches, the multisite approach is more comprehensive as simple multisite models specific for a given reaction can be readily generated from a basic generic model [see (Houston and Galetin, 2005) for an

overview]. The caveat however, is that such models require a large number of data points to ensure confident fits.

6.1.1 Basis of atypical CYP kinetics

The basis of the atypical kinetics displayed by the CYP enzymes has been interpreted primarily in the context of two binding domains within the active site of the enzyme, although other hypotheses also exist such as the presence of two binding sites plus a distinct effector binding site; multiple and kinetically distinct enzyme conformations; or a combination of multiple binding sites accompanied by conformational changes (Wienkers and Heath, 2005).

The observation that a number of CYP substrates fail to conform to Michaelis-Menten kinetics when assayed alone, nor to the expected mutually competitive inhibition type of interaction when combined, first indicated the potential existence of several binding sites within a large active site (Houston and Galetin, 2005). Early crystal structures of CYP3A4 supported this idea of an active site cavity large enough to allow the simultaneous presence of at least two molecules [PDB file 1TQN; (Yano *et al.*, 2004)], while later structures gave a direct demonstration thereof, e.g. the CYP3A4 structure showing two ketoconazole molecules bound within the active site [PDB file 2VOM; (Ekroos and Sjörgren, 2006)], and NMR studies of CYP2C9 showing that the presence of the effector molecule causes the substrate to be shifted closer to the heme iron (Hummel *et al.*, 2003).

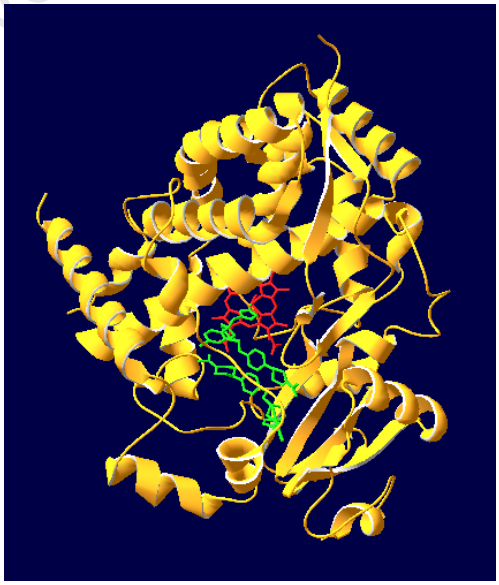


Figure 6.1: Crystal structure of CYP3A4 complexed to two ketoconazole molecules (*green*). The heme group of the active site is also shown (*red*). [2VOM.pdb rendered with Swiss-PDBViewer (Ekroos and Sjörgren, 2006).]

In addition to this and other biophysical evidence (e.g. stoichiometry experiments), site-directed mutagenesis studies have also supported the multisite hypothesis and have indicated that in the case of CYP3A4, these sites, although separate, are closely linked within the active site cavity of the enzyme (Domanski *et al.*, 2001;He *et al.*, 2003). Analysis has shown that each CYP3A4 substrate sub-group (based on structural similarities) not only has a preferential binding site but also shares a mutual binding site within the active site. It is likely, as described by the Shou *et al*. model, that these preferred and mutual binding sites differ in terms of steric hindrance, hydrophobic interactions and electron characteristics (Shou *et al.*, 1999). All of this results in an orientation difference such that SE ≠ ES (where S is substrate and E is enzyme), with each site displaying different binding affinities and rates of catalytic turnover. The exact binding conformations (i.e. which molecule binds to which site) appears to depend on the substrates involved and their relative concentration and affinity for the enzyme, while the interactions between these sites might arise through steric hindrance, the induction of conformational changes, or an alteration in the heme electron environment.

An additional basis of the atypical kinetics displayed by these enzymes appears to lie in their large degree of conformational flexibility. Crystal structures showing CYP enzymes adopting more than one ligand-bound conformation support the hypothesis of the presence of several kinetically distinct conformers [see (Ekroos and Sjörgren, 2006;Wang *et al.*, 1999) for discussion]. This hypothesis is also supported by other biophysical methods, including presteady-state kinetics and flash photolysis results with CO binding. Based on currently available data, it thus appears that both of the above mechanisms (multisite binding and multiple conformers) might play a role in the atypical kinetics of the CYPs and may even proceed simultaneously.

6.1.2 Overview of the various kinetic profiles observed for the CYPs

In addition to typical Michaelis-Menten kinetics, the various atypical kinetics routinely displayed by the CYPs include auto- and heteroactivation; partial, cooperative and substrate-dependent inhibition; substrate inhibition; and concentration-dependent heterotropic responses (activation/inhibition by an effector molecule) (Houston and Galetin, 2005). These non-hyperbolic kinetic profiles are classed as positive homotropy (substrate self-activation), negative homotropy (substrate self-inhibition) or heterotropy (activation and/or inhibition by an effector molecule).

Fig. 6.2 below depicts some of the kinetic profiles observed together with their characteristic Eadie-Hofstee plots.

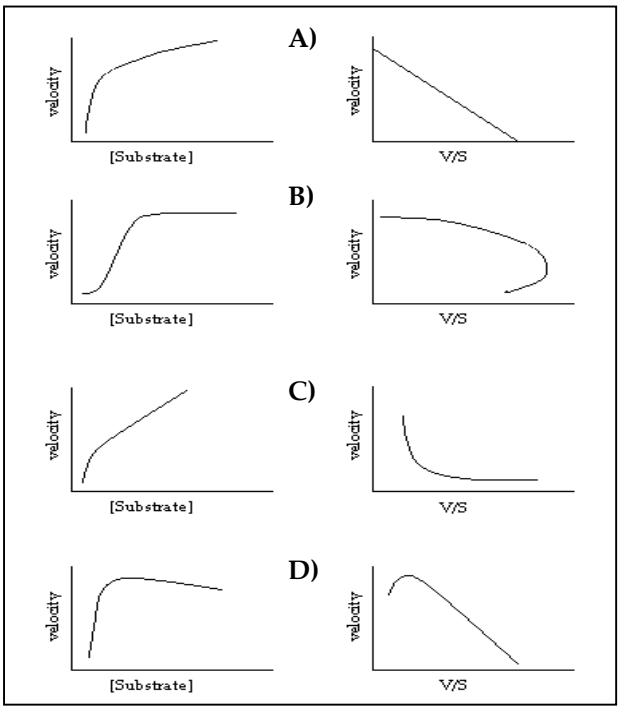


Figure 6.2: Representative graphs of various homotropic kinetic profiles [adapted from (Houston and Galetin, 2005)]. Normal velocity vs. substrate concentration plots (*left panel*) and their respective Eadie-Hofstee plots (*right panel*) are shown for Michaelis-Menten/hyperbolic kinetics (A), sigmoidal kinetics (B), biphasic kinetics (C) and substrate inhibition kinetics (D).

Both homotropic and heterotropic effects represent alterations in either the binding affinity or rate of product formation after the binding of the second molecule (whether it be of the same substrate as in homotropy, or of a different substrate as in heterotropy) (Houston and Galetin, 2005). This effect could be due to changes in access to oxidation by the heme group, or to changes in the enzyme conformation or the electron environment of the heme group (Wienkers and Heath, 2005). As a net effect, CYP activity is either increased or decreased in a substrate/effector concentration-dependent manner.

6.1.2.1 Hyperbolic kinetics

Hyperbolic (Michaelis-Menten) kinetics is the most common form of kinetics observed in general enzyme kinetics studies and this model assumes a single substrate binding site (Tracy and Hummel, 2004). By using non-linear regression to fit velocity data to the Michaelis-Menten equation, estimates of the apparent kinetic parameters V_{max} and K_m of the binding site can be determined. Conventionally, only a few data points (~6) are used to describe this curve. The cytochrome P450s often display hyperbolic kinetics, however many such cases reported in the literature might actually be incorrect due to the dismissal of atypical kinetic features.

6.1.2.2 Positive homotropic kinetics

Sigmoidal kinetics is an example of cooperative positive homotropy, where the presence of increasing concentrations of a substrate results in increased enzyme activity in a manner that describes a sigmoidal curve (Atkins, 2004;Tracy, 2003;Tracy and Hummel, 2004). In most cases it

is not yet clear whether sigmoidal kinetics results from two substrate molecules binding within the active site, or from the binding of the second molecule to a separate, allosteric site. In either case, the SES complex exhibits either a lower K_m or a higher V_{max} than the ES/SE complexes. This type of kinetic profile has probably been under-reported to date, due to a lack of analytical sensitivity required to characterise the initial part of the curve. An example of a reaction exhibiting this profile is the metabolism of dapsone by CYP2C9 (Hutzler *et al.*, 2001;Korzekwa *et al.*, 1998).

Forcing a hyperbolic curve through sigmoidal data can result in the over- or underestimation of drug clearance depending on the concentrations used. Multisite models of sigmoidal kinetics provide K_{m1}, V_{max1}, and K_{m2}, V_{max2} for the first and second substrate-binding sites respectively, rather than a single K_m and V_{max} as provided by the Michaelis-Menten model (Korzekwa *et al.*, 1998). Other, more complex models also allow for the estimation of the dissociation constants between the various enzyme-substrate complexes (Shou *et al.*, 1999). Around 15-20 data points are necessary to adequately describe these curves.

Biphasic kinetics is another example of positive homotropy (Tracy, 2003;Tracy and Hummel, 2004). This profile has two distinct phases: at low substrate concentrations, the kinetic profile is curved as with hyperbolic kinetics; however, at high substrate concentrations, the velocity of the reaction continues to increase as opposed to becoming asymptotic. This is often observed when more than one enzyme metabolises the substrate; however, when using a purified source of one enzyme, this occurs when there are multiple binding regions within the active site each with different affinities or substrate turnover rates. Thus, the single enzyme acts as though there were multiple enzymes participating in the reaction; one exhibiting a low K_m/low V_{max} (resulting in the pseudo-hyperbolic part of the curve), and another a high K_m/high V_{max} (resulting in the linear part of the cure). An example of a reaction exhibiting this profile is the metabolism of naproxen by CYP2C9 (Hutzler *et al.*, 2001).

Forcing a hyperbolic curve through biphasic data results in the overestimation of both K_m and V_{max}. Multisite models of biphasic kinetics provide the K_{m1} and V_{max1} exhibited at lower substrate concentrations by the low K_m/low V_{max} binding site, while a CL_{int} (intrinsic clearance) value is used to describe the linear portion of the plot exhibited at higher substrate concentrations and is the ratio of V_{max2}/K_{m2} (Korzekwa *et al.*, 1998). Because this upper portion is linear, it is difficult to

estimate the actual V_{max} and K_m of the second binding site as saturation is not achieved. Thus, CL_{int} in essence represents the part of the curve formed by a combination of substrate metabolism by both binding sites (Tracy and Hummel, 2004). Around 10-15 data points are necessary to adequately describe these curves.

6.1.2.3 Negative homotropic kinetics

Substrate inhibition kinetics is an example of negative homotropy (Tracy and Hummel, 2004;Venkatakrishnan *et al.*, 2003). As the substrate concentration is increased, the reaction rate starts to decrease from an achieved maximum. As with the above kinetic profiles, substrate inhibition is thought to be due to two or more substrate molecules binding to the enzyme active site simultaneously, but in this case, binding to one of the sites results in a decrease in substrate turnover rates in the other site. The SES complex of enzymes exhibiting this profile therefore has a lower V_{max} than the ES/SE complexes. An example of a reaction exhibiting this profile is the metabolism of piroxicam by CYP2C9 (Tracy and Hummel, 2004).

Forcing a hyperbolic curve through substrate inhibition data results in the underestimation of both K_m and V_{max}. Multisite models of substrate inhibition provide the K_m and V_{max} for the initial binding site and a K_i value representing the dissociation constant of substrate binding to the inhibitory site (Korzekwa *et al.*, 1998;Lin *et al.*, 2001). Since K_i must be larger than K_m for the reaction to be observable, the data at high substrate concentrations is not normally clinically relevant. The use of the CL_{int} parameter therefore remains valid, with the V_{max}/K_m ratio describing substrate turnover at non-inhibitory substrate concentrations. However, it is still necessary to model the *in vitro* kinetics correctly in order for an accurate CL_{int} estimation to be obtained. Other, more complex substrate inhibition models are also available that can provide further kinetic parameters such as the binding site interaction factors, α and β. Around 10-35 data points are necessary to adequately describe substrate inhibition curves, depending on the model used and the depth of information required.

6.1.2.4 Heterotropic kinetics

A heterotropic kinetic profile occurs when the presence of one substrate increases (positive heterotropy) or decreases (negative heterotropy) the metabolism of a different substrate, and occurs in contrast to the expected, simple mutual inhibition that results from competition for a single binding site (Atkins, 2004;Houston and Galetin, 2005). Instead, atypical heterotropic

interactions occur when the one substrate - termed the 'effector' - either activates or inhibits the metabolism of the other substrate, and where this effect is not reciprocal between the two substrates. Other, even more complex cases of heterotropic kinetics have been observed, such as an effector activating substrate metabolism at low concentrations but inhibiting it at higher concentrations; more than two different molecules binding simultaneously; or the presence of regioselective- or oxidative pathway-differential effects (i.e. the differential production of multiple products from the same substrate or the differential degree of coupling along the various steps of the catalytic cycle). Examples of substrate pairs displaying heterotropic effects include aflatoxin B1-α-naphthoflavone, testosterone-diazepam and testosterone-7-benzyloxyquinoline interactions.

Modelling heterotropic phenomena can obviously become rather complicated due to the increased number of different substrate, effector, and enzyme complexes that exist, and these models result in complex, 3-dimensional kinetic profiles [see (Korzekwa *et al.*, 1998)]. Due to the complexity of these models with their high number of parameters, an even larger number of data points are necessary to accurately estimate heterotropic kinetic parameters, and great care must be taken to minimise experimental error that may hinder data fitting (Galetin *et al.*, 2002;Tracy and Hummel, 2004).

6.1.2.5 Partial and substrate-dependent inhibition

Partial and substrate-dependent inhibition refer to atypical inhibition effects by a compound that is not itself a substrate. Partial inhibition refers to incomplete inhibition of substrate metabolism, even in the presence of saturating concentrations of the inhibitor (Tracy, 2003). This effect is possibly due to the binding of the inhibitor to an allosteric site with a resultant conformational change. However, it could also be due to the binding of the inhibitor to the active site in a distinct or overlapping position compared to the substrate binding site, such that either steric hindrance or a change in the heme electron environment occurs. This phenomenon is common to CYP-metabolised substrates and can also be modelled using multisite models (Korzekwa *et al.*, 1998). In the case of substrate-dependent inhibition, the observed K_i of an inhibitor for a specific isoform varies significantly depending on the substrate against which it is tested (Tracy, 2003). This has profound implications for the prediction and study of drug interactions, since the drug interaction potential of a compound is traditionally extrapolated from a very limited range of substrates – in some cases, only a single, 'prototypical' substrate has been used. This of course

also holds true when testing the inhibitory effects of interacting compounds that are both substrates (as discussed for heterotropic effects).

6.1.3 Using the P450 biochip platform to generate quantitative kinetic data

As discussed in the previous chapter, the P450 biochip platform was successfully developed and optimised such that satisfactory enzyme activity was achieved. However, the nature of these experiments was more qualitative in nature and it was therefore necessary to further validate the platform by generating full kinetic data sets using a commonly used CYP3A4 substrate and inhibitor. In addition, crude *E. coli* cell lysates containing Δ24C-CYP3A4 and Δ43C-CPR were tested for the first time in the array format, so as to demonstrate the feasibility and advantages of this approach for further work. The experimental objectives discussed in this chapter were therefore the following:

1) To determine initial array-based kinetics using purified Δ24C-CYP3A4 and Δ43C-CPR enzymes and a common CYP3A4 substrate, Vivid Red.
2) To determine array-based inhibition kinetics using the conventional CYP3A4 inhibitor, ketoconazole.
3) To explore the use of crude *E. coli* cell lysates containing Δ24C-CYP3A4 and Δ43C-CPR in a basic array format.

The chapter is concluded with a discussion of the potential effect of the use of glycerol in the array assays on the kinetic profiles observed.

6.2 RESULTS AND DISCUSSION

6.2.1 The cytochrome P450 kinetic array

Once the array platform had been optimised it was possible to move forward and generate kinetic data to further prove the utility of this format. For the initial kinetic array, CPR-supported CYP3A4 turnover (using 250:250 nM enzymes) of a series of Vivid Red substrate concentrations (0.5 - 43.25 μM) was assayed on the array surface and used to calculate kinetic parameters. These parameters were then compared to those obtained from the earlier cumene hydroperoxide-based

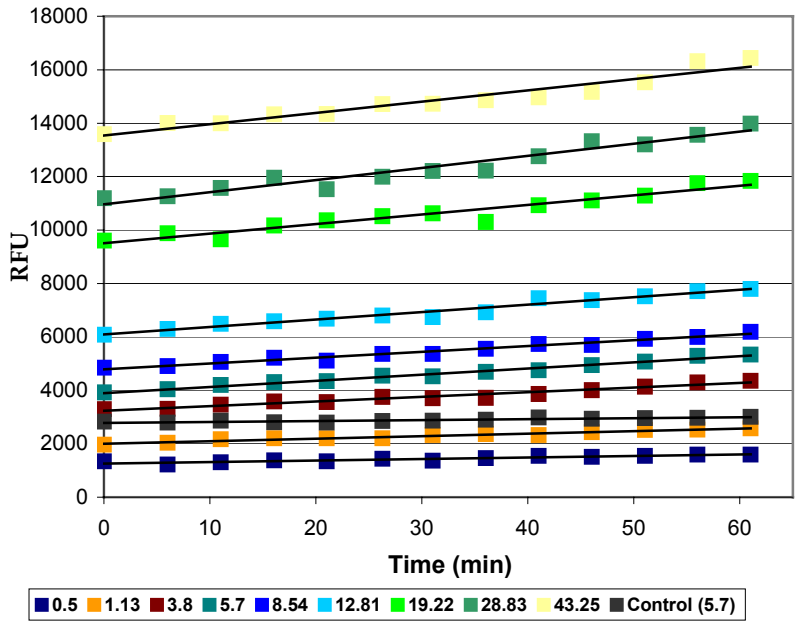
solution-phase assays and assays using the commercial baculosome preparation, as well as to that described in the literature.

6.2.1.1 Reaction linearity and reproducibility

All reactions were observed to have linear rates of at least one hour, from the start of the second round of printing (i.e. of the substrate premixes) and throughout the hour of scanning (see Fig. 6.3). This differs to the linear reaction times observed for the two solution-phase assays: the cumene hydroperoxide and commercial baculosome-based assays had linear reaction times of < 20 min, the latter in line with that reported in the literature (Trubetskoy *et al.*, 2005). The lower linear reaction times for the solution-phase assays may be the result of enzyme inactivation to some degree as the reaction time is too short to assume substrate depletion.

The longer reaction times observable for the array format is therefore likely to be due to an increased stability of the partner enzymes conferred through immobilisation - a common trend observed for the CYPs (see Section 5.1.1). This extended reaction period proved useful in the array assays as it negated the longer time periods required by the conventional microarray scanner to scan a full slide as compared to a fluorometric 96-well plate reader (see Section 5.1.2). This linearity of reaction satisfies one of the requirements for assuming steady-state kinetics, i.e. that the product increases linearly with respect to time (FDA, 2006b;Tracy, 2003). In addition, all reactions exhibited less than 15% substrate depletion.

The median CV values across all replicates and time points for each substrate concentration were also found to be satisfactory with < 15% CVs for all except the lowest substrate concentration, demonstrating the high level of reproducibility possible for enzyme reactions across an array surface (see inset, Fig. 6.3). Throughout analysis only features with ≥ 90% of pixels with signal greater than the median background plus two standard deviations were included in order to remove any potential background interference. This yielded spots with SNR values > 17, which is also satisfactory [a SNR > 10 is considered acceptable performance for array assays (Leifert *et al.*, 2005)].



Reaction (μM Vivid Red)	Median CVs (%) across time points
0.5	18.2
0.75	9.1
1.13	14.2
1.89	11.2
2.53	10.1
3.8	9.0
5.7	7.3
8.54	9.8
12.81	10.8
19.22	9.1
28.83	14.0
43.25	5.9
Control (5.7)	6.7

Figure 6.3: Velocity (RFU vs. time) graphs for the turnover of Vivid Red (0.5 – 43.25 μM) on the P450 biochip surface. All reaction rates remained linear over the 1 hr assay period. The table (*right*) lists the median CV values for each of the reactions across all 13 time points. Note: Each data point represents the average RFU of a maximum of 18 replicates. SEM error bars and certain reactions omitted from the graph for clarity. Data does not extrapolate to zero due to intrinsic fluorescence of the substrate.

The stability of the resorufin standard curve over the 1 hr assay period was also investigated. A standard curve was generated from data from both the first and the last time points and was found to be sufficiently similar – the slope of the standard curve generated from data from the last time point was 98.7% of that from the first time point (data not shown). The curve generated from the last time point was therefore used in this and subsequent array assays.

6.2.1.2 Generating a kinetic profile of Vivid Red metabolism by CYP3A4

Prior to deciding which kinetic model to fit the above velocity data to, an Eadie-Hofstee plot was generated to determine, if possible, which type of kinetic profile the data described. In this case, the data yielded an Eadie-Hofstee plot suggesting biphasic kinetics (see Fig. 6.4D). The data was then imported into GraphPad Prism for non-linear regression curve fits to both the Michaelis-Menten and biphasic models, the latter using Korzekwa's multisite equation (Korzekwa *et al.*, 1998) [see Materials and Methods for details on curve generation and use of error bars]. The goodness of fit of these individual curves was then assessed using various statistical parameters

calculated by the software, while the comparative fit of the two models was calculated using an F test, which reports a P value to guide appropriate model selection (see Table 6.1 below).

Three subsets of the kinetic data were used: the full data set comprising all 12 data points, a 'right-censored' set with only the first 11 data points (i.e. excluding the highest substrate concentration data point) and a further right-censored set with only the first 10 data points. These three variations were used to determine whether the 28.83 μM and 43.25 μM samples might effect the data unduly due to the final acetonitrile solvent concentrations in these samples being > 1% and 2% respectively. Based on the comparative fit parameter described above and on the goodness of fit statistics for each individual graph, the biphasic kinetic model fitted all three data sets the best.

However, when comparing the fit between the Michaelis-Menten and biphasic models, the Prism software penalises the more complex biphasic model due to its higher number of parameters. For this reason, the comparative fit resulted in the biphasic model fitting only the second data set (11 data points) with a highly significant P value (0.0068) compared to the Michaelis-Menten model. Despite this, all three data sets yielded similar kinetic parameters for both models, with the similarity being strongest between the first and second data sets for the Michaelis-Menten model, and the second and third data sets for the biphasic model (see Fig. 6.4A-C).

Based on these results and that of the Eadie-Hofstee plot, it seems reasonable to conclude that the biphasic kinetic model is the most suitable for profiling the turnover of Vivid Red by CYP3A4, and that the exclusion of the final data point (i.e. the sample containing 2.2% solvent) is warranted as it is likely that at this concentration solvent inhibition is occurring (see further discussion in Section 6.2.1.4 below), whereas its inclusion skews the data to produce a false asymptote as favoured by the fit to the Michaelis-Menten model. In contrast, excluding the sample containing 1.4% solvent favours the comparative fit to the Michaelis-Menten model due to the lower number of data points remaining to define the linear part of the biphasic curve, while including it greatly improves the fit of the data to the biphasic curve.

This reiterates the point made in the Introduction, that larger (and well spread) numbers of data points are necessary to provide statistically significant fits to atypical kinetic profiles and that in their absence, many a data set may be incorrectly modelled by using the Michaelis-Menten

equation by default. In the case of the Vivid Red substrate however, this does not result in a very large disparity in results. Further confirmation of the biphasic nature of our data was generated by modelling the data to the Hill equation, as an $n < 1$ value was obtained for all three data sets [data not shown, see (Korzekwa *et al.*, 1998)].

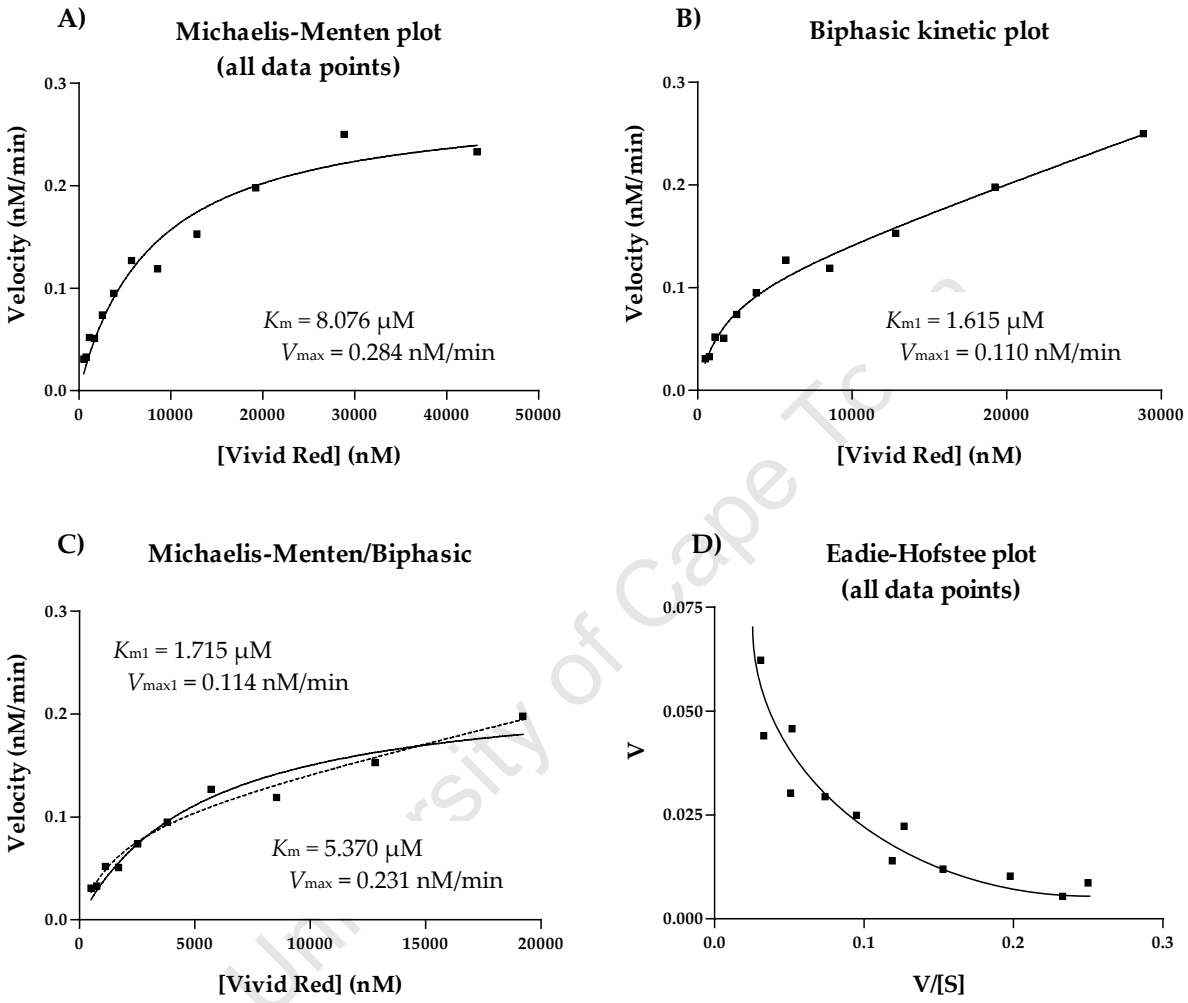


Figure 6.4: Graphs fitted to the array-generated kinetic data set for the turnover of Vivid Red using non-linear regression. A) Michaelis-Menten fit to the full data set (12 data points). B) Biphasic (multisite) fit to right-censored data (11 data points). C) Michaelis-Menten (*solid*) and biphasic (*dotted*) fit to right-censored data (10 data points). D) Eadie-Hofstee plot demonstrating the non-Michaelis-Menten behaviour of the data set (12 data points)..

All 12 data points		11 data points		10 data points	
Michaelis-Menten		*Michaelis-Menten*		*Michaelis-Menten*	
Best-fit values		Best-fit values		Best-fit values	
*V*max	0.2837	*V*max	0.2962	*V*max	0.2306
*K*m	8076	*K*m	8824	*K*m	5370
Goodness of Fit		Goodness of Fit		Goodness of Fit	
R^2	0.9575	R^2	0.9468	R^2	0.9533
Abs. sum of squares	0.002742	Abs. Sum of squares	0.002664	Abs. sum of squares	0.001295
Sy.x	0.01656	Sy.x	0.0172	Sy.x	0.01272
Biphasic		*Biphasic*		*Biphasic*	
Best-fit values		Best-fit values		Best-fit values	
V_{max1}	0.2062	V_{max1}	0.1104	V_{max1}	0.1144
V_{max2}	9.04E-07	V_{max2}	7.59E-06	V_{max2}	2.25E-05
K_{m2}	0.5506	K_{m2}	1.435	K_{m2}	4.463
K_{m1}	4683	K_{m1}	1615	K_{m1}	1715
Goodness of Fit		Goodness of Fit		Goodness of Fit	
R^2	0.9621	R^2	0.9872	R^2	0.9771
Abs. sum of squares	0.002445	Abs. Sum of squares	0.0006397	Abs. sum of squares	0.0006352
Sy.x	0.01748	Sy.x	0.00956	Sy.x	0.01029
Comparison of Fits		*Comparison of Fits*		*Comparison of Fits*	
P value	0.632	P value	0.0068	P value	0.1181
Best Fit Equation (penalised)	Michaelis-Menten	Best Fit Equation (penalised)	Biphasic	Best Fit Equation (penalised)	Michaelis-Menten

Table 6.1: Prism results table of non-linear regression analysis of Vivid Red kinetic data, showing the comparative fits of the Michaelis-Menten and biphasic kinetic models using censored and uncensored data sets. Note: V_{max} and K_m units are nM.min^{-1} and nM respectively. Full results shown in Appendix II.

6.2.1.3 Comparison of array-generated Vivid Red kinetic profile to that reported in the literature and to previous baculosome- and CuOOH-based assays

Much of the literature (including that produced by the substrate manufacturer) either report or assume typical Michaelis-Menten kinetic profiles and parameters for the turnover of Vivid Red by CYP3A4, with a resulting K_m of ~3 μM [and a V_{max} of ~25 nM.min^{-1} and k_{cat} of 5 min^{-1} where reported; see (Trubetskoy *et al.*, 2005) for example]. An exception to this is a study published by Cohen *et al*. in 2003 that observed an atypical kinetic profile corresponding to that found in our array- and baculosome-based experiments within the substrate range used (Cohen *et al.*, 2003). The Cohen study measured Vivid Red turnover using commercial insect microsomal CYP preparations across a substrate range of 0.1 - 8 μM. Based on the profile they observed within this narrow substrate range they concluded that at concentrations above 1.5 μM partial substrate inhibition was occurring, while at concentrations below 1.5 μM typical kinetics were followed with a K_m of 0.5 μM. They therefore suggested that this substrate should only be used for IC_{50} inhibitor determinations at 0.5 μM.

When comparing the results of the Cohen study to that observed in our baculosome- and array-based assays, it appears that despite the data modelling well to a biphasic curve it may in fact be deconvoluted further into at least three separate binding events that are each dependent on substrate concentration. The first phase appears to occur at Vivid Red concentrations between ~0.1 - 1.5 μM, the second at ~1.5 - 6 μM, and the third at ~6 - 30 μM (this latter part of the profile was not assayed in the Cohen study). Indeed, upon generating a Hanes-Woolf plot of our data, this is exactly the pattern that emerges (see Fig. 6.5 below).

Looking at the kinetic parameters generated from this plot by linear regression, the last two binding events have a K_m of 7.785 μM and V_{max} of 0.297 nM/min for the 1.5 - 6 μM phase and a K_m of 25.983 μM and V_{max} of 0.472 nM/min for the 6 - 30 μM phase. This graph suggests that after the first achieved maximum reaction rate, Vivid Red metabolism undergoes inhibition followed by activation (this correlates with the data in Fig. 3.4). The possibility of the simultaneous binding of three substrate molecules within the CYP3A4 active site has often been discussed in the literature, with some biophysical evidence already available (Kenworthy *et al.*, 1999;Koley *et al.*, 1995;Yano *et al.*, 2004).

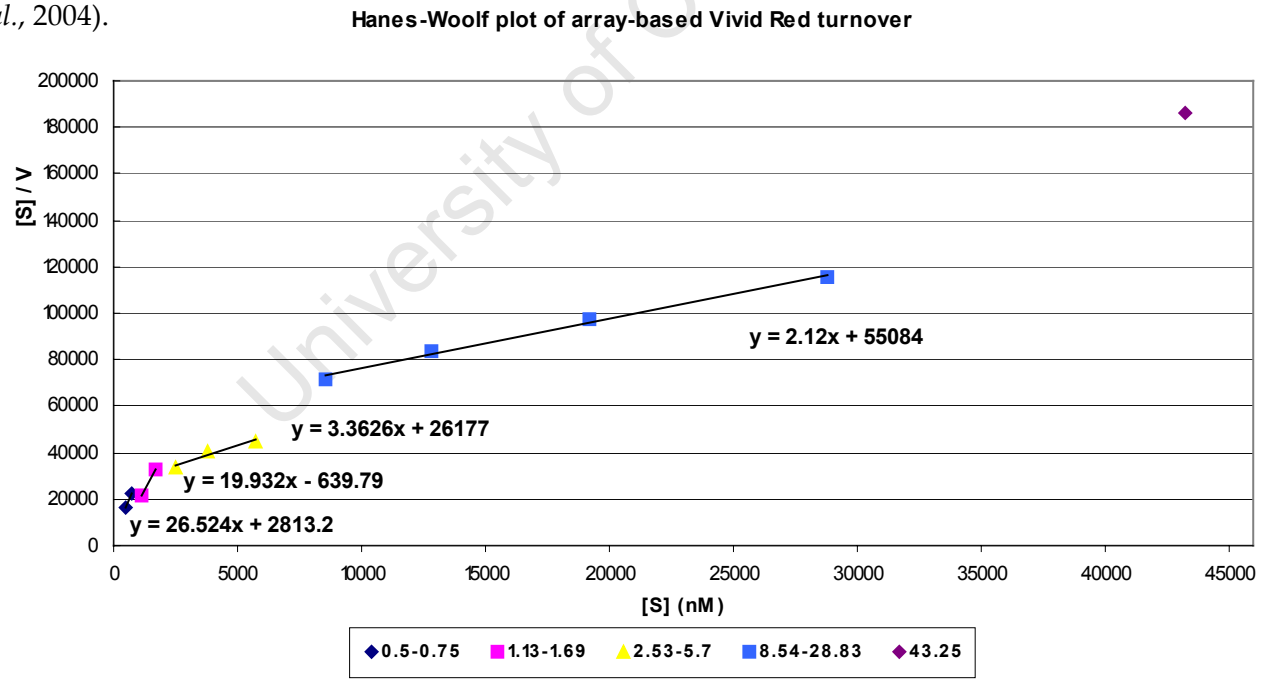


Figure 6.5: Hanes-Woolf plot fitted to the array-generated kinetic data for CYP3A4 Vivid Red turnover. Each trendline potentially represents a separate binding event with its own kinetic parameters.

Vivid Red is known to have two sites of potential metabolism, both of which produce the fluorescent product resorufin. It is therefore possible that CYP3A4 exhibits a preference for metabolism at the one or the other site depending on the concentration of Vivid Red present, and

that this regioselectivity is in turn dependent on the orientation of the substrate molecules within the active site. This may go some way towards explaining the biphasic nature of Vivid Red metabolism described here.

As discussed previously, biphasic/non-monophasic kinetic data can result as an artefact of the presence of more than one enzyme within the assay mixture, as well as from the presence of multiple substrate binding sites on a single enzyme. However, the V/[S] profile formed by the full data set (see Fig. 6.3A) conforms to that observed for the solution-phase commercial baculosome assay carried out previously (see Section 3.2.2.1). For this reason, and despite the moderate homogeneity of our recombinant enzyme extracts (specifically for CYP3A4), it appears unlikely that the atypical profile we observed was due to fortuitous turnover by *E. coli* proteins that might have co-purified and then non-specifically bound to the array surface – a combination of events that are unlikely within themselves.

Atypical kinetics can also result from the available substrate concentration being lower than expected due to, for example, high non-specific binding of the substrate to the microsomal fraction in microsome-based assays (Venkatakrishnan *et al.*, 2003). Concerns might therefore be raised that non-specific substrate adsorption to the enzymes at the array surface or even absorption into the hydrogel matrix might be reducing the available substrate concentration. However, this once again appears unlikely when considering that similar atypical profiles were obtained for Vivid Red in our baculosome assay using the recommended low microsomal concentrations (5 nM), as well as in the baculosome assay carried out by Cohen *et al*. (Cohen *et al.*, 2003).

Due to initial concerns of substrate or solvent inhibition based on our baculosome assay results and the relative lack of literature describing atypical kinetics for this substrate, our earlier CuOOH-based assays (Section 3.2.2) were carried out within a narrower substrate range (0.125 - 6 μM) compared to the above array-based assay. For this reason (and in part due to the difference in electron donors used), it is not possible to directly compare the kinetic parameters obtained by the two assay formats. Despite this, the K_m value of 1.414 μM obtained for the Δ24C-CYP3A4 in the CuOOH-based assay is similar to that obtained in this experiment using either model, suggesting that the immobilisation of CYP3A4 and the use of 15% glycerol did not affect substrate

affinity. The V_{max} for the CuOOH-based assays was ~1.38 nM/min, while for the array-based assay the turnover levels appeared to decrease by ~10-fold.

This decrease in V_{max} was likely due to a number of factors, including the longer set-up times to prepare the array assays compared to the smaller plate-based assays, the use of 15% glycerol in the array assay, and the requirement for a functional CYP3A4-CPR complex when using immobilised CPR as the electron donor rather than CuOOH. In addition, despite immobilisation generally increasing the stability of the CYPs, it is also known to decrease their catalytic rates (Lee *et al.*, 2005;Sukumaran *et al.*, 2009). It is important to note here that V_{max} is not a fundamental parameter unlike k_{cat}, as the former is dependent on enzyme concentration. However, we cannot directly compare the k_{cat} values of the solution-phase and array-based assays due to the difficulties inherent in quantifying the number of enzyme molecules immobilised within a microarray spot.

The fact that the array-based turnover described here was still sufficient despite all of these factors and the inherently low activity of these recombinant enzymes in solution is very promising for the further development of this platform.

6.2.1.4 Potential effect of solvent on atypical kinetics

Due to the majority of the CYP substrates being lipophilic, substrates are often dissolved in organic solvents such as methanol, acetonitrile and polyethylene glycol, and many of these are known to inhibit or activate substrate turnover or even alter kinetic profiles (Venkatakrishnan *et al.*, 2003). Other factors, including pH, ionic strength, specific salts, and the presence of cations, can also markedly affect substrate turnover within a relatively narrow range. In terms of achieving maximal turnover, different assay conditions are therefore optimal for different CYPs - and in some cases, even for different substrates metabolised by the same CYP isoform (although of course the goal should rather be to simulate *in vivo* conditions as closely as possible rather than to achieve optimal turnover). These factors, in addition to atypical kinetics inherent to the CYP metabolism of many substrates, increases assay complexity and further complicates data analysis and interpretation – as well as data comparison between studies carried out by different laboratories. All of these issues have to therefore be taken into account when carrying out kinetic assays.

In terms of the above array-based assay, the substrate range of 0.5 – 19.22 μM Vivid Red used maintained the acetonitrile solvent below the recommended 1% maximum (v/v) at which little to no solvent inhibition of CYP3A4 occurs (Venkatakrishnan *et al.*, 2003). However, this could not be avoided for the 28.83 μM and 43.25 μM reactions, which exceeded this at 1.4 and 2.2% respectively (kinetic analyses should preferably contain a substrate range between $10^{-1}.K_m$ and $10^{1}.K_m$, as was used here). Ideally, a substrate series containing the maximum solvent percentage required (e.g. 2.2% for the above assay) constant across all reactions should also be included so as to exclude any confounding solvent concentration effects. This is especially important when the solvent effect on the enzyme is unknown. However, the atypical kinetic profile obtained here for Vivid Red is very unlikely to be a result of this due to the low and narrow concentration range of acetonitrile used. In fact, exclusion of the highest acetonitrile concentration samples favoured the fit of the data to the atypical model, as discussed above.

6.2.1.5 Negative control results

The negative control used in the array assay was similar in concept to that used in the solution-phase assays: 250:250 nM Δ24C-CYP3A4:Δ43C-CPR, 5.7 μM Vivid Red and no NADPH. A slight increase in signal over time was observed for this control (see Fig. 6.3). This increase was lower than that of the 0.5 μM reaction and ~15% that of the 5.7 μM reaction. In contrast, the negative controls used in the CuOOH-based assays showed an increase of only around 1-2% of the total activity. An NADPH-only control showed no signal above background and was therefore not found to contribute significantly to the overall signal (data not shown). Such controls are rarely discussed in detail in the literature, so it is not possible to draw any conclusions at this point without further work.

The fact that the negative control exhibited a much higher percentage of the total activity in the array-based assays suggested that it would be prudent to include a negative control at each substrate concentration in future assays. From this it might be determined whether the percentage activity is concentration-dependent or may be extrapolated across the concentration range. For the purpose of this assay, no contribution of the negative control was therefore included in the analysis. A further, similar array-based experiment carried out using a negative control at each substrate concentration, demonstrated that although the percentage contribution varied randomly between 13-33% (median of 27%), it did not significantly effect the estimation of the kinetic parameters compared to that described above (data not shown). This effect was further

confirmed when carrying out the inhibitory assay described in the following section (where the median control contribution was found to be 15% of total activity).

6.2.1.6 Addition of slide blocking step and requirement of glycerol in slide buffers

For all experiments described in this Chapter a biotin-blocking step was introduced to bind to free streptavidin on the array surface, thereby minimising any potential background signal from excess CYP3A4 and CPR enzyme washing out of the spots and binding elsewhere (see Section 5.2.5.2). Initial array experiments that included this blocking step yielded little to no enzyme turnover (data not shown). However, upon addition of 5% glycerol to both the slide washing and blocking buffers, enzyme activity was reinstated. Comparing this result to the successful turnover observed for the arrays in Chapter 5 that had no blocking step and no glycerol in the wash buffer, it appears that the combination of a wash and a block step was too harsh for the immobilised enzymes in the absence of glycerol.

Nath *et al*. recently showed that enzymes captured on slide surfaces (the same Nexterion H slides utilised in our arrays were tested) undergo a significant loss of activity when washed in phosphate buffer only, as is routinely done for antibody arrays (Nath *et al.*, 2008). However, the addition of 5% glycerol to the wash buffer was shown to help retain activity during washing and spin-drying. The presence of glycerol in the wash buffer stabilises the proteins, and during spin-drying, the glycerol probably functions to prevent the complete dehydration of the surface. As the authors of that study note, most previous studies on factors influencing protein array performance have focused on those important for antibody array production. The optimisation of slide washing and handling to ensure that protein structure is maintained is not critical for such arrays, in contrast to enzyme arrays where this is essential. This of course reflects the difference in sensitivity to perturbation of catalytic sites versus protein-protein interaction sites and antigenic sites, and this was reiterated by our findings.

6.2.2 Testing the CYP array with the prototypical CYP3A4 inhibitor, ketoconazole

6.2.2.1 Overview of ketoconazole-based CYP3A4 inhibition

Once the utility of the P450 Biochip had been demonstrated using the kinetic analysis of Vivid Red turnover, a full inhibitory array was carried out using the CYP3A4 inhibitor, ketoconazole, to further explore the potential uses of this platform. Ketoconazole is considered to be both a strong

in vivo and *in vitro* inhibitor of CYP3A4 activity and is therefore recommended as a 'prototypical' CYP3A4 inhibitor control (Donato *et al.*, 2004;FDA, 2006b). It generally exhibits a K_i of between 0.0037 – 0.18 μM and IC_{50} values of < 3 μM depending on substrate and assay conditions.

Mechanisms of CYP inhibition can be broadly divided into three categories: reversible, quasi-irreversible and irreversible inhibition (Lin and Lu, 1998). Reversible interactions are the most common type of inhibition and mostly arise as a result of direct competition for binding at the active site and probably only involves the first step of the catalytic cycle. Members of the CYP family are particularly prone to this type of inhibition due to their broad substrate specificity. Inhibitors that act during and subsequent to the oxygen-binding step are generally irreversible or quasi-irreversible inhibitors and require at least one catalytic cycle to take effect and are a result of the formation of inhibitory reactive metabolites.

The effect of an inhibitor on drug clearance/substrate metabolism will be quite different depending on whether the inhibition is reversible or irreversible (Lin and Lu, 1998). The most important distinction between the two is the time-dependent loss of activity associated with irreversible inhibition (i.e. observed IC_{50} values decrease over time as the enzyme becomes inactivated), whereas for reversible inhibitors this does not occur (IC_{50} values remain constant over time as enzyme function is unaltered). Kinetically, reversible inhibition can be further sub-classified as competitive, non-competitive, uncompetitive or mixed inhibition. In competitive inhibition, the binding of inhibitor to the active site prevents the binding of the substrate; in non-competitive inhibition, the inhibitor binds to a different site on the enzyme and the enzyme-substrate-inhibitor (ESI) complex is non-productive; in uncompetitive inhibition, the inhibitor does not bind to free enzyme but only to the enzyme-substrate complex and the consequent ESI complex is non-productive. Mixed inhibition exhibits characteristics of more than one of the above inhibition types.

Many of the potent reversible CYP inhibitors are nitrogen-containing compounds such as imidazoles, pyridines and quinolines (Lin and Lu, 1998). The potency of these inhibitors is determined both by their lipophilicity and by the strength of the bond between their nitrogen lone pairs and the CYP heme iron. Ketoconazole is a potent reversible CYP inhibitor as it contains an imidazole group that binds strongly to the heme iron and it has a high lipophilicity (log P = 3.7), allowing it to undergo strong hydrophobic interactions with the CYP enzyme. Despite the large

number of studies investigating CYP inhibition by ketoconazole there still appears to be a lack of clarity as to the exact mechanisms of its inhibition. Various studies refer to ketoconazole in turn as a competitive (Sekiguchi *et al.*, 2009), non-competitive (Gibbs *et al.*, 1999) or mixed-type inhibitor (Bourrie *et al.*, 1996). However, due to its potency and prevalence of use in published CYP assays, ketoconazole was chosen as the test compound for investigating array-based inhibition of Vivid Red turnover.

6.2.2.2 Testing the inhibitory effect of ketoconazole on array-based CYP3A4 Vivid Red turnover

The inhibitor array was carried out in a similar manner to the kinetic array but with the addition of 0.1 nM - 1 mM ketoconazole. Experiments to quantify inhibitory potency usually determine the parameters IC_{50} (the concentration of the inhibitor at which the substrate turnover is reduced by 50%) or K_i (the concentration of the inhibitor at which half the enzyme is inhibitor-bound) (FDA, 2006b). For IC_{50} evaluations, only one substrate concentration is used, which should be $\leq K_m$ (ideally it should be $< K_m$ to more closely relate the inhibitor's IC_{50} to its K_i), together with a relevant range of inhibitor concentrations. For K_i determinations, both substrate and inhibitor concentrations should be varied to cover ranges above and below the drug's K_m and the inhibitor's likely K_i. K_i is particularly useful for expressing the potency of an inhibitor because, unlike the IC_{50} value, it is independent of substrate concentration (Wienkers and Heath, 2005). In our array-based inhibitor assay, both parameters were included for analysis.

Competitive inhibitors are known to raise the apparent K_m without affecting the V_{max}, while non-competitive inhibitors have no effect on K_m but decrease V_{max}. To determine whether ketoconazole inhibition of array-based Vivid Red turnover conformed to either inhibition profile, various data transformations were used. The apparent K_m values for the substrate series at each inhibitor concentration were determined using non-linear regression and these were then plotted against the inhibitor concentrations used – a linear relationship is expected for competitive inhibition. To investigate the effect of the inhibitor on substrate turnover, the percentage total turnover for each substrate concentration was plotted against the inhibitor concentrations (i.e. in the same manner used to determine IC_{50} values).

Based on these analyses it was clear that ketoconazole inhibition of Vivid Red turnover by immobilised CYP3A4 enzyme did not conform to either expected inhibitory profile. The plots of the apparent K_m vs. inhibitor concentration yielded linear relationships with low correlation

values, particularly for the K_m values generated using the Michaelis-Menten model (see Fig. 6.6 and Appendix II). Based on the linear regression results generated using K_m values derived from the biphasic model, the K_m (y-intercept) is 1.77 µM while the K_i (negative x-intercept) is 0.663 mM. While the K_m value is similar to that obtained in the initial kinetic array described above, the K_i value is clearly much larger than expected. In fact, as illustrated by the data represented in Fig. 6.6, at inhibitor concentrations between 0.1 nM – 10 µM, activation of Vivid Red turnover to greater than 100% of that observed in the absence of ketoconazole occurred across all substrate concentrations, while significant inhibition was only observed in the presence of 100 µM ketoconazole. These results were later confirmed by a repetition of this experiment using a different batch of recombinant enzyme as well as a fresh lot of ketoconazole (data not shown).

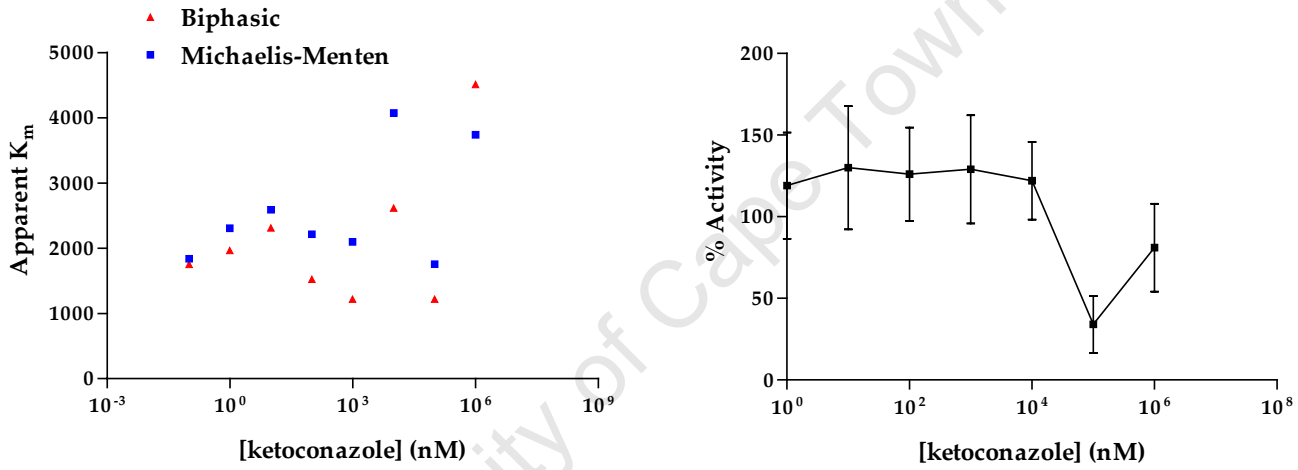


Figure 6.6: Graphs depicting the change in apparent K_m across the inhibitor concentration series (*left*) and the change in median activity across all substrate concentrations at each inhibitor concentration (*right*). Note: The bars illustrate the spread of data across substrate concentrations and are not standard errors of measurement. Inhibitor concentrations for both graphs are depicted in $\log_{10}$ scale.

The obvious question therefore, is why this prototypical and potent CYP3A4 inhibitor is exhibiting such a low inhibitory effect towards array-based Vivid Red turnover. The answer may lie in the fact that ketoconazole is known to interact with ferric CYP at the sixth axial position, which induces a type II difference spectrum (Lin and Lu, 1998).

Both the coordination of a strong ligand to the pentacoordinated iron (which has cysteine only as a ligand), or the displacement of a weaker ligand (either water or substrate) from the hexacoordinated state by a stronger ligand, can give rise to a type II binding spectrum. As discussed in Section 6.2.4 below, glycerol is also known to induce a type II binding spectrum. The

main difference between the solution-phase assays and the array-based assays is the presence of 15% glycerol in the latter. It is possible that since both ketoconazole and glycerol potentially cause inhibition in a similar manner (i.e. by shifting the heme towards the low-spin state which is unfavourable for substrate oxidation), the presence of both of these compounds in the inhibitor array is resulting in the apparent decrease in the inhibitory effect of ketoconazole due to the vastly greater concentration of glycerol already present (~2 M).

Note that the potential inhibition displayed by ketoconazole is not competitive in terms of competing for metabolism as a substrate, but rather it is potentially competitive in that it binds to the active site and may therefore block access by the substrate. However, since it is also known to sequester the enzyme in a low-spin state, it should also display non-competitive characteristics. It is therefore arguable that this inhibitor should actually exert mixed type inhibition consisting of elements of both competitive and non-competitive inhibition.

Another possibility that might be put forward to explain the lack of inhibitory effect is that the inhibitor may have been 'mopped up' by inactive enzyme on the surface, thereby reducing the available inhibitor concentration. However, despite the relatively high CYP3A4 concentration used for our assays, 250 nM protein is still well below the recommended microsomal protein concentrations of less than 1 mg/ml (Ansede and Thakker, 2004;FDA, 2006b). Indeed, if this was the problem this effect would surely have been apparent for substrate turnover as well, which is unlikely as we obtained a K_m value within range of that expected. Another possibility is that ketoconazole is binding to the array surface itself, although this is difficult to test.

Any solvents used to dissolve the inhibitor should be present at ≤ 1% (v/v) and preferably at < 0.1% where possible (FDA, 2006b). Several papers describing the ketoconazole inhibition of Vivid Red turnover use DMSO as a solvent [DMSO is a universal organic solvent widely used to deliver drugs in cell culture-based biological assays (Venkatakrishnan *et al.*, 2003)], however DMSO is known to decrease CYP3A4 turnover by up to 50% at concentrations as low as 1% (Trubetskoy *et al.*, 2005); in fact it is a potent inhibitor of multiple CYP isoforms at concentrations as low as 0.1-0.2 % (Venkatakrishnan *et al.*, 2003). This renders it unsuitable for use in *in vitro* drug metabolism studies. For the array-based inhibitor assay described here, methanol was used instead as a solvent, since a recent study demonstrated that it imposes negligible inhibition of CYP3A4 turnover of Vivid Red at 1% (Hao *et al.*, 2008). Indeed, the authors of this study found an IC_{50}

value of 0.94 μM for ketoconazole on CYP3A4 Vivid Red metabolism (using 3 μM Vivid Red and 5 nM commercial baculosome preparations). The fact that the IC_{50} value they found was significantly higher than that generally reported in the literature for this substrate-inhibitor combination, suggests that part of the inhibitory effect attributed to ketoconazole might in fact be due to the choice of solvent used and the lack of adequate controls to compensate for this effect.

In either case, it is unlikely that the diminished inhibitory effect we observed for ketoconazole was in some way due to solvent inhibition, as the methanol concentration only exceeded the recommended 1% for the highest inhibitor sample (1 mM ketoconazole). An increase in activity for this sample relative to that observed for the 100 μM inhibitor sample (see Fig. 6.6) is possibly therefore due to solvent activation by the presence of a high (10%) methanol concentration.

These results suggest that for the further development of this array-based format, it may prove necessary to experiment with other alternatives to glycerol for reducing spot evaporation (some examples include arraying at lower temperatures and increased humidity, or the testing of other anti-evaporation agents not observed to induce spectral changes), such that the inhibitory effect is restored. This would of course also be beneficial in increasing array-based substrate turnover rates. However, these results also imply that caution needs to be taken in trying to correlate inhibitory (or indeed, substrate) kinetic data to that already provided in the literature.

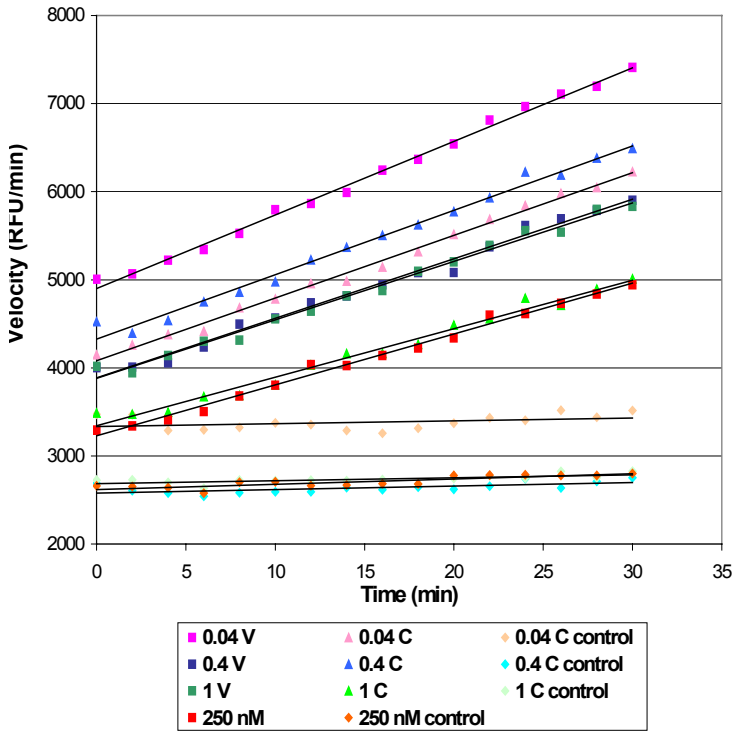
6.2.3 Testing crude *E. coli* lysates containing recombinant CYP3A4 and CPR in array format

Crude *E. coli* cell lysates containing Δ24C-CYP3A4 and Δ43C-CPR were tested for the first time in a basic array format to determine the potential of using lysates instead of purified recombinant enzymes for the preparation of the P450 Biochips. The enzyme pre-mixes contained pooled Δ24C-CYP3A4 and Δ43C-CPR crude lysates at three different final concentrations of 0.04, 0.4 and 1 mg/ml total protein using either equal volumes of each crude lysate or equivalent concentrations of either (e.g. 0.5 mg/ml each for the 1 mg/ml total). 250:250 nM purified Δ24C-CYP3A4 and Δ43C-CPR were included as a positive control and only one substrate concentration, 4 μM, was tested across the array. Negative controls at each protein concentration were also present and consisted of enzyme and substrate only.

The highest turnover was observed for the 0.04 mg/ml total protein crude lysate reactions (see Fig. 6.7), which reiterates the conclusion drawn in Section 5.2.5.3 that low protein concentrations are sufficient to saturate the array surface and to ensure the successful interaction between the immobilised partner enzymes. The negative controls across all samples yielded between 4 – 12% total signal, with the 0.04 mg/ml and 250 nM negative controls yielding the lowest and highest signals respectively.

In comparing the turnover levels obtained when using purified enzymes and when using crude lysates, an increase in turnover was observed for the latter; a result that was likely due to reduced enzyme handling prior to assay. This is possibly why the purified protein also exhibited a much higher (>10-fold) turnover at the substrate concentration tested, due to the greatly reduced set-up time needed for this basic assay compared to the full kinetic arrays described above (this brings the activity level in line with that observed for the CuOOH-based assays in Chapter 3). These results are very promising for further array development, as it suggests that using crude lysate preparations together with robotic liquid handling stations could yield improved assay sensitivity due to increased enzyme turnover levels. The use of freshly prepared crude lysates in itself would also significantly decrease the pre-array preparation time and cost.

Further work investigating a wider range of crude lysate concentrations optimised according to the relative percentage expression of the partner enzymes within *E. coli*, as well as assays using freshly prepared crude lysates, are all steps likely to further increase the turnover levels of the immobilised CYPs such that they more closely approach that observed for microsomal preparations in solution-phase assays. Another improvement would be the use of mechanised pre-array preparation tools including temperature-controlled liquid handling stations, such that the preparation time is further decreased and assay steps are carried out at lower temperatures – both important factors in helping to maintain enzyme activity prior to assay (all of the array work described in this thesis was carried out using manual pipetting at room temperature into 384-well plates). The influence of the latter factor was evident in the above simple assay, which had reduced sample numbers and therefore reduced pre-array preparation time; a significant increase in activity was observed for the 250 nM purified protein sample compared to the previous, larger kinetic and inhibitor arrays described earlier in this chapter. (Note that the same protein batch was used for both the above assay and the kinetic assay thereby excluding the possibility of enzyme batch-to-batch variation.)



Sample	Velocity (nM/min)	R^2	Minus controls
0.04 V	4.19	1.00	4.03
0.04 C	3.57	0.99	3.41
0.04 C control	0.16	0.13	-
0.4 V	3.39	0.99	3.19
0.4 C	3.67	0.99	3.47
0.4 C control	0.20	0.42	-
1 V	3.33	0.99	3.16
1 C	2.76	0.98	2.59
1 C control	0.17	0.46	-
250 nM	2.90	0.99	2.60
250 nM control	0.30	0.73	-

Figure 6.7: Graph depicting the turnover of Vivid Red by CYP3A4 and CPR crude lysates in a simple array format (protein concentration in mg/ml unless otherwise stated). Various lysate total protein concentrations were tested. A purified protein positive control and several negative controls were also included. Note: data points represent an average of 27 replicates. Error bars omitted for clarity.

6.2.4 Deconvoluting the potential effect of glycerol on CYP enzyme kinetics

Vivid Red is a commercial substrate that has been used in many CYP studies and the vast majority of these studies have reported typical Michaelis-Menten kinetics. Since we have repeatedly observed atypical enzyme kinetics for the same substrate, concerns might reasonably be raised regarding the possible role that the relatively high glycerol concentrations used in our array-based kinetic assays might play in inducing these atypical kinetic profiles. The presence of glycerol may also play a role in reducing the array-based inhibitory effect of the potent CYP3A4 inhibitor, ketoconazole. The following section therefore discusses the various hypotheses around the mechanisms of glycerol inhibition of CYP activity and whether or not it is likely to play a role.

6.2.4.1 Overview of the use of glycerol in protein biochemistry and protein array applications

Glycerol markedly decreases the rate of thermal inactivation of CYPs to the P420 form, probably due to its effect on ordering the solvent and making the unfolding of proteins more energetically unfavourable than in aqueous solutions alone (Metelitza *et al.*, 1982;Yang and Cederbaum, 1997).

Glycerol is also widely used to reduce denaturation during the solubilisation, purification, storage and analysis of CYPs (Headlam and Tuckey, 2002).

Despite the fact that glycerol has a slightly lower surface tension than water, glycerol exhibits much lower volatility and is therefore often used to reduce the rate of evaporation of microarray spots - the addition of glycerol lowers the vapour pressure of aqueous solutions, an effect that becomes particularly significant when working with nano- to picolitre volumes [see Table 6.2; (Diaz-Mochón *et al.*, 2007)]. When fabricating microarrays in a humidified environment, the presence of glycerol in the spots can also result in the re-absorption of water from the air, thereby further reducing the rate of evaporation (Wong and Diamond, 2008). Another benefit in using protein printing buffers with a high viscosity (such as glycerol provides), is that these buffers produce more compact spots resulting in more reliable data (Preininger *et al.*, 2005).

Property	Water	Glycerol
Mass (g/mol)	18	92
Vapour pressure (mmHg)	18	0.00016
Surface tension (J/m^2)	0.073	0.063
Viscosity (kg/m · s)	0.001	1.5
Density (g/ml)	0.99	1.3
Dielectric constant (ε)	78	43

Table 6.2: Physical properties of water and glycerol at 20°C (Ebbing and Gammon, 1999).

In addition to the beneficial effects of glycerol in stabilising proteins and reducing evaporation, glycerol also serves as a useful solvent for substrates in drug discovery in that its hydrophobicity helps to minimise compound precipitation (Gosalia and Diamond, 2003). It is also a water mimic, capable of forming multiple hydrogen bonds and providing the necessary lubrication for enzyme catalysis. However, enzyme reaction rates are generally reduced in glycerol, thus necessitating longer assay times or higher enzyme concentrations.

6.2.4.2 Enzyme inhibition effect

There have been many hypotheses put forward as to the exact mechanism of enzyme inhibition by glycerol, both for enzymes in general as well as specifically for the CYP system. In general terms, this inhibition is thought to be partially due to the high viscosity of glycerol, which may slow down the diffusion of substrate, cofactors and products as well as any necessary

conformational movements of the enzymes (Gosalia and Diamond, 2003). This is however only likely to be an issue for highly active enzymes that operate at close to the diffusion limit (k_{cat}/K_m ~10^8 $M^{-1}s^{-1}$). In addition, the hydrophilic aspect of organic solvents such as glycerol may reduce the water partition coefficient, stabilise hydrophobic substrates, raise the activation barrier and thereby potentially slow the reaction down. Binding equilibria might also be modulated by glycerol *via* hydrophobic or osmolarity effects, which in some instances may reduce assay sensitivity and yet enhance assay specificity by reducing non-specific interactions.

In terms of CYP catalysis, several potential avenues of glycerol inhibition have been investigated. One of the earliest hypotheses is that glycerol may act as a competitive substrate for binding in the active site. Subsequent studies have however shown that this is unlikely (Headlam and Tuckey, 2002;Voznesensky and Schenkman, 1992b). Levels of glycerol inhibition were found to be independent of substrate concentration and the affinity of substrates were observed to be independent of the glycerol concentrations. In addition, the K_d for CO binding was not altered by the presence of glycerol, whereas the presence of known substrates in the active site has been shown to markedly influence the strength of CO binding. Furthermore, no distinct substrate-like binding spectrum was observed when 2 M glycerol was added to microsomal systems.

It was also suggested that the increase in viscosity upon the addition of glycerol might slow down the necessary conformational changes required for electron transfer between the FAD and FMN domains of CPR. It might also decrease the frequency of CYP and CPR collision and thereby decrease the rate of electron transfer between the FMN and heme domains. However, the addition of glycerol to 20% (w/v) only increases the viscosity of water from 1.005 mPa · s to 1.76 mPa · s (Cheng, 2008), whereas the viscosity of rat liver microsomal membrane is two orders of magnitude higher than that of water (Feinstein *et al.*, 1975). The addition of glycerol would therefore be expected to only have minimal influence on the translational movement of these proteins.

Glycerol has been found to have minimal effects on lipid hydrocarbon dynamics and the rotational mobility of CYPs in membranes (Headlam and Tuckey, 2002), thereby ruling out a possible detrimental effect on membrane structure. In addition, both vesicle-associated and aqueous CYPs were shown to be inhibited to the same degree by glycerol.

Another potential mechanism of inhibition lies in the ability of glycerol to shift the P450 heme towards the low-spin state. In its resting state, the heme group exists as a mixture of low-spin iron with a water molecule bound, and high-spin iron with cysteine as the only ligand (Meunier *et al.*, 2004). The absolute absorption maximum of CYPs in the low-spin state is ~416-419 nm, while that of the high-spin state peaks at ~390-416 nm (Danielson, 2002). Substrate binding is therefore monitored by measuring the difference spectrum between unbound and substrate-bound enzyme. The binding of substrate in the active site displaces the axial water molecule (as well as other water molecules within the binding pocket), which induces the spin equilibria to shift towards the high-spin form. This produces what is termed a Type I difference spectrum, characterised by an absorbance maximum at ~385-390 and a trough at ~420 nm. Some ligands induce a Type II difference spectrum, which is characterised by an absorbance maximum at ~425-435 nm and a trough at ~390-405 nm. The presence of a Type II spectrum indicates that the ligand coordinates to the iron more strongly than the water molecule it displaced. Glycerol induces an absolute absorption maximum at approximately 417 nm, while its difference spectrum yields a Type II spectrum (Headlam and Tuckey, 2002), indicating that glycerol induces a shift of the CYP heme to the low-spin state and that it coordinates more strongly to the heme iron as the sixth ligand than water does.

This shift in the spin equilibria would result in a higher proportion of CYP enzyme being in the 'resting' state rather than in the high-spin state necessary for electron transfer. For substrates that exhibit a Type I spectrum, it is possible that even after substrate binding the presence of glycerol in the active site will maintain the low-spin state due to its stronger coordination to the heme iron. The glycerol molecule would therefore be displaced from the active site with greater difficulty, thereby slowing down catalysis. Following this line of argument, it would be expected that substrates exhibiting a Type II spectrum would be less susceptible to glycerol inhibition.

Indeed, the degree of glycerol inhibition was found to be substrate-dependent (Headlam and Tuckey, 2002). For example, the $K_{1/2}$ for P450scc inhibition by glycerol was found to be 22% in the presence of cholesterol (a Type I substrate), and 32% in the presence of 20α-hydroxycholesterol (a Type II substrate). Some substrates are therefore better than others at displacing water or glycerol from coordination to the iron, an idea that is corroborated by the fact that certain substrates induce a greater spin state transition than others, both in and without the presence of glycerol. However, the degree of glycerol inhibition was not observed to bear a direct correlation to

substrate K_d; hence this 'disruptive' property can not be simply extrapolated from substrate affinities and requires direct experimental measurement.

Another mechanism whereby glycerol might result in CYP inhibition was suggested to be *via* its osmotic effect in dehydrating the enzyme through the reduction of the macromolecular volume of the protein upon transition to a low-spin state, with the resultant forced dissociation of essential water molecules from the active site (Headlam and Tuckey, 2002). However, as mentioned above, it is the binding of substrate and the transition to the high-spin state that is accompanied by the displacement of water molecules from the active site – in fact, their release produces an entropy change which is thought to drive the substrate binding process in the first place (Meunier *et al.*, 2004). In addition, electron transfer cannot occur from the low-spin state in any case, since it is the high-spin state that causes a redox potential change and triggers electron transfer. Upon completion of the catalytic cycle, it is the re-entry of water molecules into the binding pocket that displaces the metabolic product and causes its release from the enzyme.

Glycerol has been shown to affect the affinity between the CYPs and the reductase, which may result in a decreased rate of electron transfer between the two enzymes. For example, through measuring actual electron transfer, the apparent K_m of rabbit CYP2B4 for CPR was observed to be 9 times higher in 25% glycerol than in 5% glycerol, although V_{max} was unaffected (Voznesensky and Schenkman, 1992b). However, through measuring spin state transition, the K_d of P450scc (a mitochondrial CYP) for adrenodoxin at 20% glycerol was found to decrease to 63% of that without glycerol (Headlam and Tuckey, 2002).

The first study suggests weaker binding of CYP and CPR in the presence of glycerol, while the second study suggests stronger binding. A possible explanation for these contradictory results lies in the following: glycerol has a lower dielectric constant than water and its addition to aqueous solutions therefore reduces the overall dielectric constant and conductivity thereof (Headlam and Tuckey, 2002;Voznesensky and Schenkman, 1992b). The force of electrostatic interactions relates inversely to the dielectric constant of the medium, so the addition of glycerol should increase the strength of complementary charge pairing between the charged residues on the CYP and reductase surfaces. This would result in stronger protein-protein interactions. It is not immediately obvious why an increased strength of electrostatic interaction would result in a decreased affinity (as observed by the first study), as it is now known that the productive

association of these partner enzymes is governed by such interactions (see Section 5.2.3.2). One possibility is that this increase leads to a higher probability of the formation of non-productive charge-pairing complexes, and perhaps the slower dissociation of both productive and non-productive charged complexes. (In the case of slower dissociation of productive complexes, the rate of the first electron transfer might increase, while that of the second electron transfer will decrease, as the CPR FMN domain has to dissociate from the CYP in order to be re-reduced by the FAD domain. This would therefore slow down overall electron transfer rates to CYP.) Both of these effects could result in overall decreased electron transfer rates and therefore apparent decreases in affinity as measured by K_m. Total interactions between CYP and CPR would increase however, as measured by the K_d for the spin transition.

Besides glycerol's potential reduction of the dielectric constant, yet another possibility of how glycerol might effect CYP-CPR electrostatic interactions lies in its ability to stabilise proteins by ordering the solvent molecules around them (Voznesensky and Schenkman, 1992b). The ordering of solvent molecules has the effect of internalising hydrophobic residues from the protein surface, thereby focusing surface charged residues towards the medium. This too could enhance non-productive charge-pairing and slow dissociation of charged complexes.

The lack of glycerol inhibition of transfer rates to uncharged electron acceptors such as 1,4-benzoquinone by CPR (Voznesensky and Schenkman, 1992b), appears to confirm the above hypothesis of the importance of the effect of glycerol inhibition *via* electrostatic interactions. In addition, glycerol was shown to reduce the rates of reduction of other charged electron acceptors such as cytochrome c and, as for CYP, this reduction was directly correlated with a decrease in the conductivity of the medium (Voznesensky and Schenkman, 1992b).

It is also possible that glycerol might induce changes in the structure of enzymes and thereby affect their activity. Yun *et al*. reported a marked conformational change of CYP1A2 from 47 to 64% α-helical content upon the addition of 10-15% glycerol, while the β-sheet content decreased from 17 to 10% at 15% glycerol (Yun *et al.*, 1996). Gutierrez *et al*. found that high glycerol concentrations (> 40%) induced changes in the optical spectra of both oxidised and reduced CPR samples, which was indicative of changes in the average flavin environments (Gutierrez *et al.*, 2002). Increasing the glycerol concentration was shown to progressively diminish the observed rate of internal reductase electron transfer: at 75% (w/v) glycerol, electron transfer was almost

completely inhibited. However, Voznesensky *et al*. did not find any difference evident in the vicinity of the flavin prosthetic groups in the fluorescence-emission spectrum of the reductase in 20% glycerol compared to that without glycerol (Voznesensky and Schenkman, 1992b).

6.2.4.3 Conclusions

Of the above hypotheses regarding the mechanism of CYP inhibition by glycerol, it appears unlikely that glycerol's effect lies in increased solution viscosity, altered membrane environments, or in competitive binding to the active site. While altered CYP structure might play a role at lower (<20%) glycerol concentrations, it is not yet clear whether these secondary structure changes translate into a detrimental alteration of the overall 3D fold and catalytic activity of the CYP enzyme. From current knowledge, it appears therefore that increasing glycerol concentrations increase the frequency of non-productive CYP-CPR associations, slow down CYP-CPR dissociations, and decrease reaction rates due to a shift towards the low-spin state. In addition, we know that the level of glycerol inhibition of CYP activity is substrate-dependent and sensitivity towards inhibition is not directly correlated to the substrate's affinity for CYP [in general however, low glycerol (<4%) causes less than 5% inhibition of CYP activity] (Headlam and Tuckey, 2002). Indeed, the fact that inhibition is substrate-dependent seems to point towards an inhibitory effect that is mediated through the active site, rather than towards a universal effect on CYP-CPR interaction and reaction environment.

For these reasons, it seems unlikely that glycerol is responsible for inducing the non-Michaelis-Menten kinetics we observed in our array experiments using 15% final glycerol. In addition, these atypical kinetics were also observed in earlier solution phase experiments that contained minimal glycerol (< 5%). It appears that the glycerol inhibition effect at this percentage simply results in a general decrease in the rate of reaction observed (i.e. in V_{max} as observed here), rather than in any changes in K_m or the kinetic profile (neither were disrupted here). Corroborating results were found in a study by Inouye *et al*. that observed biphasic kinetic behaviour of CYP1A1 turnover of 7-ethoxycoumarin using a recombinant yeast microsome system (Inouye *et al.*, 2000). They found that the addition of 10 and 20% glycerol to the enzyme reaction had no significant effects on the kinetic profile, while the addition of 30% glycerol significantly increased the K_{m2} and decreased the V_{max2}. Despite this, the inhibitory effect of glycerol at 30% was small compared to that of the solvent, methanol.

However, it does appear that glycerol was playing a critical role in inducing the atypical inhibition profile for ketoconazole on the array surface. This effect is potentially a result of ketoconazole and glycerol both causing inhibition by binding as Type II ligands to the CYP3A4 active site. The high glycerol concentrations required for the array-based assay therefore obscured the inhibitory effect of ketoconazole thereby significantly decreasing its apparent inhibition of Vivid Red turnover.

6.3 CONCLUSIONS

We used the conventional fluorescent CYP3A4 substrate, Vivid Red, to validate the P450 Biochip platform by generating a full kinetic data set. The array-generated data yielded a K_m for Vivid Red similar to that reported in the literature, illustrating that the array conditions and immobilisation of CYP3A4 did not affect substrate affinity. The maximal velocities observed were however much lower than those generated by the solution-phase assays. This was most likely due to the inhibitory presence of glycerol, which was used to minimise evaporation throughout array fabrication and analysis.

We observed atypical kinetics for CYP3A4 metabolism of the Vivid Red substrate, suggesting a biphasic kinetic profile and the presence of at least two substrate binding sites. These results were in contrast to the majority of the typical Michaelis-Menten kinetic profiles reported in the literature for this substrate, but were in correlation with data generated by our solution-phase assays for both recombinant and commercial baculosomal enzyme. The CYP3A4 isoform used in this study is known to exhibit the most ligand promiscuity and conformational flexibility of all the CYP isoforms (Williams *et al.*, 2004), which would suggest that is also the most 'susceptible' towards exhibiting atypical kinetics. The literature contains numerous examples of Michaelis-Menten curves being forced through data that clearly show atypical kinetic features, as discussed earlier in this chapter. In some of these cases, the insufficient number of data points rules out any meaningful selection of an alternative model (Galetin *et al.*, 2003). Performing assays that require high sample numbers (such as cases where atypical kinetics occur) therefore becomes much more feasible using this format, as we have found here.

Once we had obtained kinetic parameters for Vivid Red turnover on the array surface, we tested the platform using the prototypical inhibitor, ketoconazole. Ketoconazole is considered to be a

potent CYP3A4 inhibitor, however the data we generated showed very little inhibition of Vivid Red turnover on the array surface. It is possible that inhibition by ketoconazole was obscured by the presence of high glycerol concentrations in the array reactions, as both of these compounds are known to bind as Type II ligands. These results appear to suggest that ketoconazole does indeed act as a mixed-type inhibitor.

Crude lysates expressing Δ24C-CYP3A4 and Δ43C-CPR were also tested on the array surface to investigate the use of crude preparations rather than purified proteins. Relatively low total protein concentrations were found to be sufficient for obtaining Vivid Red turnover, and turnover at the substrate concentration used was higher than that for the purified proteins. In addition, the reduction of array preparation time for this experiment due to a lower sample number resulted in a significant increase in turnover by the purified enzyme, suggesting that the introduction of robotic liquid handling systems and the optimisation of array printing conditions to favour the folding and functionality of the immobilised enzymes are both promising avenues for future work.

6.4 MATERIALS AND METHODS

6.4.1 The cytochrome P450 kinetic array

6.4.1.1 Preparation of pre-mixes for the first and second print rounds

For the first print round, a 250 nM:250 nM Δ24C-CYP3A4 :Δ43C-CPR solution and a 0.05 mg/ml Cy3-biotin-BSA solution were each prepared in P450 Storage Buffer and aliquoted into the relevant wells in the microtitre plate. For the second print round, 12 substrate pre-mixes (0.5 – 43.25 μM Vivid, 1 mM NADPH, in 100 mM potassium phosphate with 15% final glycerol), two negative control pre-mixes (enzyme plus 5.7 μM Vivid, no NADPH; and NADPH only), and a resorufin dilution series (1 – 500 nM) were prepared. The negative controls and resorufin series were also prepared in 100 mM potassium phosphate with 15% final glycerol.

6.4.1.2 Printing procedure

After running a buffer only datum pointing test, the first print round (enzyme and BSA pre-mixes) was printed onto a home-made streptavidin-coated slide pre-equilibrated to RT, using the optimised print parameters previously established. After the print run had completed the slide was incubated on the print bed for 15 min at RT before removing and blocking 1 x 15 min in

chilled P450 Slide Blocking Buffer (20 mM potassium phosphate pH 7.4, 0.2 mM EDTA, 5% glycerol, 100 μM biotin, 0.1 mg/ml BSA) on ice with shaking, and then washing 3 x 5 min in chilled P450 Slide Wash Buffer (including 5% glycerol) and 1 x 5 min in chilled dH_20, also on ice with shaking. All block and wash steps were performed in a sterile falcon tube. The slide was then spun dry and the second print run (substrate pre-mixes and resorufin series) was then carried out. As soon as the print run completed, the slide was removed and scanned every 5 min for 1 hr at 175 PMT at 20 μm resolution.

6.4.1.3 Data analysis

To capture the signal from each spot a 150 μm diameter feature finder was used, while the median background was measured within 3 feature diameters from the centre of each feature, excluding the first 6 pixels. The median background signal was subtracted from the median fluorescence signal for each spot. To ensure that any background present did not compromise the quality of the kinetic data, only features containing ≥ 90% pixels with signal greater than background plus two standard deviations and ≤ 5% saturation were used for further data analysis. For the generation of the resorufin standard curve only features with a SNR ≥ 5 were used, while no censoring was possible for the negative controls except for that of obvious outliers due to visible background artefacts.

All the signals (median minus median background) for each reaction at each time point were averaged (maximum 18 replicates per substrate concentration). These averaged values were used to plot the reaction velocities for each substrate concentration in Microsoft Excel using linear regression. The velocities were then converted from $RFU.min^{-1}$ to resorufin $(nM).min^{-1}$ using a standard curve generated from the resorufin dilution series (using background- and buffer-subtracted data from the last time point). These values were then imported into GraphPad Prism and used to calculate kinetic parameters using non-linear regression. Note that sample signals were averaged at each time point and then used to generate a rate curve, rather than generating individual rate curves for each sample and then averaging the rates - hence the lack of error bars in the V/[S] graphs in Fig. 6.3 (median CV values for the averaged sample points were instead reported as in Fig. 6.2). The two approaches are mathematically equivalent, however the former greatly decreases array data analysis time.

Prior to deciding which kinetic model to fit the data to, an Eadie-Hofstee plot was produced (velocity vs. velocity/[substrate]). The relevant kinetic model was then chosen depending on the curve described by this plot, in this case the biphasic multisite model was used. The goodness of fit of these individual curves was then assessed using various statistical parameters (R^2, sum of squares of the residuals, number of runs observed, standard error of parameter estimates, etc.). The fit of atypical kinetic models was then compared to the fit to a traditional Michaelis-Menten model using an F test, which reports a P value for the comparison of fit.

The following equation was used to model biphasic kinetics and was manually inputted into Prism (Korzekwa *et al.*, 1998): $v = \frac{(V_{max1} \times [S]) + (CL_{int} \times [S]^2)}{K_{m1} + [S]}$

6.4.2 Testing the CYP array with the prototypical CYP3A4 inhibitor, ketoconazole

6.4.2.1 Preparation of pre-mixes for the first and second print rounds

As for Section 6.4.1.1 above, with the following exceptions. A full negative control series was printed, i.e. enzyme plus substrate and no NADPH, one at each substrate concentration tested. A 10 mM ketoconazole stock solution was prepared in 100% methanol and used to prepare a concentration range between 0.1 nM – 1 mM at each Vivid Red substrate concentration (i.e. the inhibitor was added to the substrate pre-mixes used for the second print round).

6.4.2.2 Printing procedure

As for Section 6.4.1.2 above.

6.4.2.3 Data analysis

As for Section 6.4.1.3 above, with the following additions. The apparent K_m for the substrate series at each inhibitor concentration was determined using non-linear regression and comparative fits to both the Michaelis-Menten and the biphasic kinetic model. To determine whether the inhibitor was conforming to standard competitive or non-competitive inhibition and to determine the K_i, these apparent K_m values were then plotted against the inhibitor concentrations. The inhibitor parameters were then determined and compared for both kinetic models in Prism using linear regression. In order to determine the IC_{50} values of the inhibitor at each substrate concentration, the percentage total turnover (expressed as the percentage of the reaction velocity observed in the

absence of inhibitor) was plotted against the inhibitor concentration series. Due to the similar pattern observed across all substrate concentrations, a graph was generated of the median percentage total activity across all substrate concentrations at each inhibitor concentration to illustrate this (bars were included to demonstrate the spread of the data for each substrate series).

6.4.3 Testing crude *E. coli* lysates containing recombinant CYP3A4 and CPR in array format

6.4.3.1 Preparation of pre-mixes for the first and second print rounds

As for Section 6.4.1.1 above, with the following exceptions. For the first print round, the enzyme pre-mixes contained pooled Δ24C-CYP3A4 and Δ43C-CPR crude lysates in three different series. The first series contained three different final concentrations of 0.04, 0.4 and 1 mg/ml total protein using equal volumes of each crude lysate. The second series contained three different final concentrations of 0.04, 0.4 and 1 mg/ml total using equal concentrations of each crude lysate (e.g. 0.5 mg/ml each of the Δ24C-CYP3A4 and Δ43C-CPR crude lysates for the 1 mg/ml total). The third series consisted of 250:250 nM purified Δ24C-CYP3A4 and Δ43C-CPR only. Only one substrate concentration, 4 μM, was tested. A negative control for each protein concentration was also included, which consisted of enzyme plus Vivid Red and no NADPH.

6.4.3.2 Printing procedure and data analysis

As for Sections 6.4.1.2 and 6.4.1.3 above with the following exceptions. Data was acquired every 2 min for 30 min. Only a basic velocity vs. substrate concentration curve was plotted for both the reaction and the negative controls.

CHAPTER SEVEN

Conclusions and Future Work

7.1 OVERVIEW

There is currently a clear niche for a low-cost, high-throughput assay platform that allows quick, cost-effective and accurate measurement of the metabolism of novel drugs by the cytochrome P450s, as well as enabling drug-drug interactions to be assessed at an early stage. Such a platform would provide an important addition to the current repertoire of tools used by pharmaceutical companies for pre-clinical drug screening as well as opening up such screening to smaller organisations and academic laboratories. Our aim therefore was to develop and optimise a novel protein array format that allowed quantitative and truly high-throughput measurement of cytochrome P450 turnover. Creating the P450 Biochip platform was certainly not without its challenges, and the results we obtained reiterated these enzymes' notoriety for complexity in expression, interactions and kinetics.

We successfully cloned and expressed various full-length and N-terminal deletion CYP3A4 and CPR constructs in *E. coli*. We opted to focus on the cytosolic fraction only for protein purification, as this was in line with our interests in developing a HTP platform. The relative soluble and insoluble expression of the various constructs were compared and in general we found that constructs without the N-terminal membrane anchor and with a C-terminal His-BCCP tag provided the highest soluble protein expression. Despite the high level of apparently insoluble expression observed even for certain deletion constructs, the use of detergents was shown to significantly increase soluble protein yield, probably through disruption of peripheral membrane interactions. C-terminal tag placement and the truncation of the N-terminal hydrophobic domains were shown to not only favour soluble protein expression, but also protein folding and enzyme activity levels of both CYP3A4 and CPR construct proteins.

Two of the CPR constructs, Δ43N-CPR and Δ43C-CPR, yielded solution-phase turnover numbers within the range reported in the literature. However, using solution-phase CuOOH-based assays, the most active CYP3A4 construct (Δ24C-CYP3A4), was observed to have a much lower turnover number compared to commercial baculosomal preparations. We have demonstrated that this was at least in part due to a low fraction of the purified recombinant enzyme being in the catalytically active P450 form, which correlated with results observed for similar truncation constructs described elsewhere in the literature. While a number of protein expression and purification factors could be further optimised so as to further increase the CYP3A4 holoprotein yield (e.g. IPTG concentrations, expression temperatures and purification conditions), the expression and

activity levels obtained were deemed sufficient to carry forward for array development. Due to the similar orientation of their BCCP-tags, Δ24C-CYP3A4 and Δ43C-CPR were chosen for further investigation using *in silico* protein modelling. Based on the results of this model, it appeared likely that these partner enzymes would be able to re-form a productive association on the array surface after immobilisation *via* their C-terminal BCCP-tags.

Prior to initiating array experiments, various surface chemistries and immobilisation methods were considered through a literature review and the method most likely to work for our application was chosen. We opted to use directed immobilisation *via* a streptavidin-biotin interaction mediated by the biotinylated BCCP-tags present on our constructs, and we prepared and tested home-made streptavidin slides for this purpose. Prior to using the array-immobilised enzymes for generating kinetic data, several parameters and protocols throughout the array fabrication process were tested. These included printing, slide blocking and wash buffers; printing parameters; CYP3A4 and CPR enzyme concentrations and molar ratios; slide preparation protocols, etc. These choices were then validated by the observation of successful catalytic turnover of the substrate on the array surface.

The next step in validating the P450 Biochip platform was the generation of kinetic data for a conventional CYP3A4 substrate, Vivid Red. The array-generated kinetic data yielded a K_m in line with that reported in the literature, illustrating that the array conditions and immobilisation of the CYP3A4 had not effected the enzyme's affinity for the substrate. The maximal velocities observed were however much lower than that observed in the solution-phase assays and this was thought to be due to the inhibitory presence of glycerol, which we had found to be necessary to maintain microdroplet fluidity on the array surface. Based on the results we obtained and data available in the literature, we concluded that the likely effect that glycerol was having was related to its ability to bind as the sixth axial heme ligand, thereby sequestering the enzyme in the catalytically unproductive low-spin Fe^{3+} form. It was also likely that glycerol was detrimental to the interaction of CYP3A4 and CPR by promoting non-specific interactions between these two enzymes and by slowing down their dissociation. The presence of glycerol therefore led to a decrease in reaction rates but had no effect on K_m or the kinetic profile.

We observed atypical kinetics for the Vivid Red substrate, suggesting two to three possible substrate binding sites, in contrast to most of the results reported in the literature but in

correlation with data generated by our solution-phase assays for both recombinant and commercial baculosomal enzyme. It is possible that many data sets on cytochrome P450 metabolism in the literature have previously been incorrectly modelled or truncated to obtain a fit to the Michaelis-Menten model, thereby potentially leading to erroneous reporting of the kinetic parameters when atypical kinetics are truly at play. This aspect of CYP kinetics was therefore thoroughly reviewed in this thesis.

Once we had obtained kinetic parameters for Vivid Red turnover on the array surface, we moved our attention towards testing the P450 array with a prototypical inhibitor, ketoconazole. Ketoconazole is considered to be a potent CYP3A4 inhibitor and is often used to investigate substrate metabolism by this enzyme. However, the data we generated showed very little inhibition of Vivid Red turnover by ketoconazole on the array surface. A review of the literature revealed that the likely mechanism of inhibition by ketoconazole is - similarly to glycerol - through its ability to bind as a sixth axial heme ligand. It is therefore likely that the atypical inhibition data we observed was due to the presence of high glycerol concentrations obscuring the inhibitory effect of ketoconazole in the array-based assays. Our results did however open up many questions for future work, since although ketoconazole is considered to be a prototypical inhibitor, there is still confusion as to whether it acts as a competitive, non-competitive or mixed type inhibitor and the reported K_i values for this compound differ over several orders of magnitude. Our results lead us to suggest that ketoconazole acts as a mixed type inhibitor: non-competitive inhibition is offered through its sequestration of the heme in the low-spin form, while competitive inhibition is offered through its binding in the active site thereby physically competing with substrate molecules for space within the binding pocket (particularly at higher ketoconazole concentrations where a second molecule of ketoconazole is known to bind in the CYP3A4 active site).

In summary, we have described the successful development of a novel cytochrome P450 array format that allows the functional interaction between membrane-free CYP and CPR enzymes on a 2D surface within sub-nanolitre volumes. In contrast to all other previous assay formats and despite the challenges posed in achieving electron transfer between the two enzymes *via* a transient protein-protein interaction on a 2D array surface, this has been successfully and optimally realised here for recombinant human CYP and CPR enzymes in a direct one-to-one functional relationship through the design of this platform. The method for fabricating these

arrays is much faster than that allowed by the membrane encapsulation methods used in previous studies and the number of possible reactions per array is also much higher. With a working surface area per slide of 58 x 22 mm, our array format is capable of at least 4200 reactions per slide, with a resulting feature density of ~ 329 spots/cm^2. Both of these attributes are significant improvements in terms of high-throughput capacity for the screening of cytochrome P450 metabolism. This format also provides greater ease of multiplexing and platform extension to other detection methods (e.g. mass spectrometry) due to the greater facility with which metabolites can be retrieved and solutions exchanged. Reagent and protein consumption are much lower (5 µl of enzyme yields over 1800 reactions), which becomes an important consideration when screening thousands of compounds and measuring the kinetic or inhibitory constants for each. More importantly, with the basic format in hand, we were able to generate large numbers of high quality data points over several different parameters within a single experiment. Significant time- and cost-savings are therefore implicit in this novel assay format, thereby enabling the HTP study of the effects of hundreds of clinically relevant polymorphisms on drug metabolism and propensity towards ADRs. The fact that this format is based on the use of recombinant proteins makes it even more adaptable to this type of application compared to previous formats.

7.2 FUTURE WORK

For the further development of the P450 Biochip platform described here, it will be necessary to investigate various factors important in the production of soluble CYP holoprotein in order to obtain turnover levels closer to that obtained for native enzymes. In addition, the various CPR constructs should be more fully characterised in terms of flavin content and its correlation to the protein yields and levels of catalytic activity obtained. It does not appear as though the array format in itself is detrimental to enzyme activity and any improvements obtained in holoprotein yield are very likely to translate into increased turnover on the array platform. The various expression and purification parameters that could possibly be tested have been discussed in this thesis.

In terms of the array format, the most pressing avenues for future work is in finding alternatives to glycerol for maintaining droplet fluidity, the reduction of array preparation time by the introduction of robotic liquid handling systems, and the optimisation of array printing conditions

to favour the folding and functionality of the immobilised enzymes. Further *in silico* modelling studies would also be useful to ascertain whether different linkers, linker lengths and/or deletion constructs would impart further flexibility and therefore potentially increased functional interaction between the immobilised CYP and CPR enzymes.

Other important routes of enquiry to further increase the versatility and HTP capacity of this platform are the use of on-chip protein translation, the extension of this platform to other detection methods such as MS or SPR, and the use of alternative pre-fabricated array structures such as nanowell arrays.

7.3 CONCLUSIONS

This work has shown that CYP enzyme assay miniaturisation is feasible in a genuine, quantitative protein microarray format and that it has several advantages over traditional solution-phase assays when considering its use as a tool for high-throughput compound screening by the pharmaceutical industry and others. It is possible that the lessons learnt during the development of this platform might later be extended to other membrane protein families as well. Despite the many protein array studies in recent years, studies describing membrane protein arrays and their use for ligand screening or substrate profiling remain limited. The first such study was only published in 2002, which described the fabrication and application of arrays of G protein-coupled protein receptors (Fang *et al.*, 2002). As the authors of this paper noted, the lack of microarray methods for membrane proteins remains a major limitation, particularly considering that membrane-bound proteins currently represent 50% of drug targets (Drews, 2000). The platform we have developed here might therefore not only provide a novel HTP screen for investigating cytochrome P450 drug metabolism but may also offer new ideas for utilising other membrane proteins in a microarray format. To the best of our knowledge, our work also represents one of the first demonstrations that enzymatic turnover reactions can be measured quantitatively in a true microarray format, potentially opening the door to similar studies on other classes of proteins.

APPENDIX

I. VECTOR AND CONSTRUCT SEQUENCES

Below follows the DNA sequences of both the pMD004 and pBJW102.2 plasmid vectors used for cloning the N- and C-terminal tagged constructs of CYP3A4 and CPR respectively. The protein sequence of each construct is also provided. Various sequence elements of importance are highlighted as follows:

Bold - T5 transcription start (N-terminus) and transcription termination region end (C-terminus). The first residues of each are highlighted.

Boxed - Start and stop codons

Pink - His_6-tag sequence

Blue - BCCP-tag sequence

Orange - GS-linker sequence

Underlined – first residue coding for CYP3A4/CPR protein sequence

1.1 Vector sequences

1.1.1 pMD004 plasmid vector backbone sequence

```
1       CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA
61      ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA
121     GGATCTCACC ATCACCATCA CCATGGGATC CCAGCAGCAG CGGAAATCAG TGGTCACATC
181     GTACGTTCCC CGATGGTTGG TACTTTCTAC CGCACCCCAA GCCCGGACGC AAAAGCGTTC
241     ATCGAAGTGG GTCAGAAAGT CAACGTGGGC GATACCCTGT GCATCGTTGA AGCCATGAAA
301     ATGATGAACC AGATCGAAGC GGACAAATCC GGTACCGTGA AAGCAATTCT GGTCGAAAGT
361     GGACAACCGG TAGAATTTGA CGAGCCGCTG GTCGTCATCG AGGGTGGCAG CGGTTCTGGC
421     CCCGCGGCCG CTCGAGCATG CATCTAGGGG TCGACCTGCA GCCAAGCTTA ATTAGCTGAG
481     CTTGGACTCC TGTTGATAGA TCCAGTAATG ACCTCAGAAC TCCATCTGGA TTTGTTCAGA
541     ACGCTCGGTT GCCGCCGGGC GTTTTTTATT GGTGAGAATC CAAGCTAGCT TGGCGAGATT
601     TTCAGGAGCT AAGGAAGCTA AAATGGAGAA AAAAATCACT GGATATACCA CCGTTGATAT
661     ATCCCAATGG CATCGTAAAG AACATTTTGA GGCATTTCAG TCAGTTGCTC AATGTACCTA
721     TAACCAGACC GTTCAGCTGG ATATTACGGC CTTTTTAAAG ACCGTAAAGA AAAATAAGCA
781     CAAGTTTTAT CCGGCCTTTA TTCACATTCT TGCCCGCCTG ATGAATGCTC ATCCGGAATT
841     TCGTATGGCA ATGAAAGACG GTGAGCTGGT GATATGGGAT AGTGTTCACC CTTGTTACAC
901     CGTTTTCCAT GAGCAAACTG AAACGTTTTC ATCGCTCTGG AGTGAATACC ACGACGATTT
961     CCGGCAGTTT CTACACATAT ATTCGCAAGA TGTGGCGTGT TACGGTGAAA ACCTGGCCTA
1021    TTTCCCTAAA GGGTTTATTG AGAATATGTT TTTCGTCTCA GCCAATCCCT GGGTGAGTTT
1081    CACCAGTTTT GATTTAAACG TGGCCAATAT GGACAACTTC TTCGCCCCCG TTTTCACCAT
1141    GGGCAAATAT TATACGCAAG GCGACAAGGT GCTGATGCCG CTGGCGATTC AGGTTCATCA
1201    TGCCGTTTGT GATGGCTTCC ATGTCGGCAG AATGCTTAAT GAATTACAAC AGTACTGCGA
1261    TGAGTGGCAG GGCGGGGCGT AATTTTTTTA AGGCAGTTAT TGGTGCCCTT AAACGCCTGG
1321    GGTAATGACT CTCTAGCTTG AGGCATCAAA TAAAACGAAA GGCTCAGTCG AAAGACTGGG
1381    CCTTTCGTTT TATCTGTTGT TTGTCGGTGA ACGCTCTCCT GAGTAGGACA AATCCGCCCT
1441    CTAGATTACG TGCAGTCGAT GATAAGCTGT CAAACATGAG AATTGTGCCT AATGAGTGAG
1501    CTAACTTACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG
1561    CCAGCTGCAT TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCCA
1621    GGGTGGTTTT TCTTTTCACC AGTGAGACGG GCAACAGCTG ATTGCCCTTC ACCGCCTGGC
1681    CCTGAGAGAG TTGCAGCAAG CGGTCCACGC TGGTTTGCCC CAGCAGGCGA AAATCCTGTT
1741    TGATGGTGGT TAACGGCGGG ATATAACATG AGCTGTCTTC GGTATCGTCG TATCCCACTA
```

```
1801   CCGAGATATC CGCACCAACG CGCAGCCCGG ACTCGGTAAT GGCGCGCATT GCGCCCAGCG
1861   CCATCTGATC GTTGGCAACC AGCATCGCAG TGGGAACGAT GCCCTCATTC AGCATTTGCA
1921   TGGTTTGTTG AAAACCGGAC ATGGCACTCC AGTCGCCTTC CCGTTCCGCT ATCGGCTGAA
1981   TTTGATTGCG AGTGAGATAT TTATGCCAGC CAGCCAGACG CAGACGCGCC GAGACAGAAC
2041   TTAATGGGCC CGCTAACAGC GCGATTTGCT GGTGACCCAA TGCGACCAGA TGCTCCACGC
2101   CCAGTCGCGT ACCGTCTTCA TGGGAGAAAA TAATACTGTT GATGGGTGTC TGGTCAGAGA
2161   CATCAAGAAA TAACGCCGGA ACATTAGTGC AGGCAGCTTC CACAGCAATG GCATCCTGGT
2221   CATCCAGCGG ATAGTTAATG ATCAGCCCAC TGACGCGTTG CGCGAGAAGA TTGTGCACCG
2281   CCGCTTTACA GGCTTCGACG CCGCTTCGTT CTACCATCGA CACCACCACG CTGGCACCCA
2341   GTTGATCGGC GCGAGATTTA ATCGCCGCGA CAATTTGCGA CGGCGCGTGC AGGGCCAGAC
2401   TGGAGGTGGC AACGCCAATC AGCAACGACT GTTTGCCCGC CAGTTGTTGT GCCACGCGGT
2461   TGGGAATGTA ATTCAGCTCC GCCATCGCCG CTTCCACTTT TTCCCGCGTT TTCGCAGAAA
2521   CGTGGCTGGC CTGGTTCACC ACGCGGGAAA CGGTCTGATA AGAGACACCG GCATACTCTG
2581   CGACATCGTA TAACGTTACT GGTTTCACAT TCACCACCCT GAATTGACTC TCTTCCGGGC
2641   GCTATCATGC CATACCGCGA AAGGTTTTGC ACCATTCGAT GGTGTCGGAA TTTCGGGCAG
2701   CGTTGGGTCC TGGCCACGGG TGCGCATGAT CTAGAGCTGC CTCGCGCGTT TCGGTGATGA
2761   CGGTGAAAAC CTCTGACACA TGCAGCTCCC GGAGACGGTC ACAGCTTGTC TGTAAGCGGA
2821   TGCCGGGAGC AGACAAGCCC GTCAGGGCGC GTCAGCGGGT GTTGGCGGGT GTCGGGGCGC
2881   AGCCATGACC CAGTCACGTA GCGATAGCGG AGTGTATACT GGCTTAACTA TGCGGCATCA
2941   GAGCAGATTG TACTGAGAGT GCACCATATG CGGTGTGAAA TACCGCACAG ATGCGTAAGG
3001   AGAAAATACC GCATCAGGCG CTCTTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTCGGTC
3061   GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA
3121   TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT
3181   AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC CCCCTGACGA GCATCACAAA
3241   AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT
3301   CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACCTG
3361   TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA GCTCACGCTG TAGGTATCTC
3421   AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC
3481   GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA
3541   TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG CGAGGTATGT AGGCGGTGCT
3601   ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC
3661   TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA
3721   CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC AGCAGATTAC GCGCAGAAAA
3781   AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA
3841   AACTCACGTT AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT
3901   TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT ATGAGTAAAC TTGGTCTGAC
3961   AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC
4021   ATAGTTGCCT GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC
4081   CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG CTCCAGATTT ATCAGCAATA
4141   AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC
4201   CAGTCTATTA ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC
4261   AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG TATGGCTTCA
4321   TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA
4381   GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA
4441   CTCATGGTTA TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT
4501   TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG GCGACCGAGT
4561   TGCTCTTGCC CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG
4621   CTCATCATTG GAAAACGTTC TTCGGGGCGA AAACTCTCAA GGATCTTACC GCTGTTGAGA
4681   TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC
4741   AGCGTTTCTG GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG
4801   ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTTCAAT ATTATTGAAG CATTTATCAG
4861   GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG
4921   GTTCCGCGCA CATTTCCCCG AAAAGTGCCA CCTGACGTCT AAGAAACCAT TATTATCATG
4981   ACATTAACCT ATAAAAATAG GCGTATCACG AGGCCCTTTC GTCTTCAC
```

1.1.2 pBJW102.2 plasmid vector backbone sequence

```
1      CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA
61     ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGGCA
121    CTTACGGCCG GGGGTGGCAG CGGTTCTGGC GCAGCAGCGG AAATCAGTGG TCACATCGTA
181    CGTTCCCCGA TGGTTGGTAC TTTCTACCGC ACCCCAAGCC CGGACGCAAA AGCGTTCATC
241    GAAGTGGGTC AGAAAGTCAA CGTGGGCGAT ACCCTGTGCA TCGTTGAAGC CATGAAAATG
```

```
301    ATGAACCAGA TCGAAGCGGA CAAATCCGGT ACCGTGAAAG CAATTCTGGT CGAAAGTGGA
361    CAACCGGTAG AATTTGACGA GCCGCTGGTC GTCATCGAGG GTGGCAGCGG TTCTGGCCAC
421    CATCACCATC ACCATAAGCT TAATTAGCTG AGCTTGGACT CCTGTTGATA GATCCAGTAA
481    TGACCTCAGA ACTCCATCTG GATTTGTTCA GAACGCTCGG TTGCCGCCGG GCGTTTTTTA
541    TTGGTGAGAA TCCAAGCTAG CTTGGCGAGA TTTTCAGGAG CTAAGGAAGC TAAAATGGAG
601    AAAAAAATCA CTGGATATAC CACCGTTGAT ATATCCCAAT GGCATCGTAA AGAACATTTT
661    GAGGCATTTC AGTCAGTTGC TCAATGTACC TATAACCAGA CCGTTCAGCT GGATATTACG
721    GCCTTTTTAA AGACCGTAAA GAAAAATAAG CACAAGTTTT ATCCGGCCTT TATTCACATT
781    CTTGCCCGCC TGATGAATGC TCATCCGGAA TTTCGTATGG CAATGAAAGA CGGTGAGCTG
841    GTGATATGGG ATAGTGTTCA CCCTTGTTAC ACCGTTTTCC ATGAGCAAAC TGAAACGTTT
901    TCATCGCTCT GGAGTGAATA CCACGACGAT TTCCGGCAGT TTCTACACAT ATATTCGCAA
961    GATGTGGCGT GTTACGGTGA AAACCTGGCC TATTTCCCTA AAGGGTTTAT TGAGAATATG
1021   TTTTTCGTCT CAGCCAATCC CTGGGTGAGT TTCACCAGTT TTGATTTAAA CGTGGCCAAT
1081   ATGGACAACT TCTTCGCCCC CGTTTTCACC ATGGGCAAAT ATTATACGCA AGGCGACAAG
1141   GTGCTGATGC CGCTGGCGAT TCAGGTTCAT CATGCCGTTT GTGATGGCTT CCATGTCGGC
1201   AGAATGCTTA ATGAATTACA ACAGTACTGC GATGAGTGGC AGGGCGGGGC GTAATTTTTT
1261   TAAGGCAGTT ATTGGTGCCC TTAAACGCCT GGGGTAATGA CTCTCTAGCT TGAGGCATCA
1321   AATAAAACGA AAGGCTCAGT CGAAAGACTG GGCCTTTCGT TTTATCTGTT GTTTGTCGGT
1381   GAACGCTCTC CTGAGTAGGA CAAATCCGCC CTCTAGATTA CGTGCAGTCG ATGATAAGCT
1441   GTCAAACATG AGAATTGTGC CTAATGAGTG AGCTAACTTA CATTAATTGC GTTGCGCTCA
1501   CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC
1561   GCGGGGAGAG GCGGTTTGCG TATTGGGCGC CAGGGTGGTT TTTCTTTTCA CCAGTGAGAC
1621   GGGCAACAGC TGATTGCCCT TCACCGCCTG GCCCTGAGAG AGTTGCAGCA AGCGGTCCAC
1681   GCTGGTTTGC CCCAGCAGGC GAAAATCCTG TTTGATGGTG GTTAACGGCG GGATATAACA
1741   TGAGCTGTCT TCGGTATCGT CGTATCCCAC TACCGAGATA TCCGCACCAA CGCGCAGCCC
1801   GGACTCGGTA ATGGCGCGCA TTGCGCCCAG CGCCATCTGA TCGTTGGCAA CCAGCATCGC
1861   AGTGGGAACG ATGCCCTCAT TCAGCATTTG CATGGTTTGT TGAAAACCGG ACATGGCACT
1921   CCAGTCGCCT TCCCGTTCCG CTATCGGCTG AATTTGATTG CGAGTGAGAT ATTTATGCCA
1981   GCCAGCCAGA CGCAGACGCG CCGAGACAGA ACTTAATGGG CCCGCTAACA GCGCGATTTG
2041   CTGGTGACCC AATGCGACCA GATGCTCCAC GCCCAGTCGC GTACCGTCTT CATGGGAGAA
2101   AATAATACTG TTGATGGGTG TCTGGTCAGA GACATCAAGA AATAACGCCG GAACATTAGT
2161   GCAGGCAGCT TCCACAGCAA TGGCATCCTG GTCATCCAGC GGATAGTTAA TGATCAGCCC
2221   ACTGACGCGT TGCGCGAGAA GATTGTGCAC CGCCGCTTTA CAGGCTTCGA CGCCGCTTCG
2281   TTCTACCATC GACACCACCA CGCTGGCACC CAGTTGATCG GCGCGAGATT TAATCGCCGC
2341   GACAATTTGC GACGGCGCGT GCAGGGCCAG ACTGGAGGTG GCAACGCCAA TCAGCAACGA
2401   CTGTTTGCCC GCCAGTTGTT GTGCCACGCG GTTGGGAATG TAATTCAGCT CCGCCATCGC
2461   CGCTTCCACT TTTTCCCGCG TTTTCGCAGA AACGTGGCTG GCCTGGTTCA CCACGCGGGA
2521   AACGGTCTGA TAAGAGACAC CGGCATACTC TGCGACATCG TATAACGTTA CTGGTTTCAC
2581   ATTCACCACC CTGAATTGAC TCTCTTCCGG GCGCTATCAT GCCATACCGC GAAAGGTTTT
2641   GCACCATTCG ATGGTGTCGG AATTTCGGGC AGCGTTGGGT CCTGGCCACG GGTGCGCATG
2701   ATCTAGAGCT GCCTCGCGCG TTTCGGTGAT GACGGTGAAA ACCTCTGACA CATGCAGCTC
2761   CCGGAGACGG TCACAGCTTG TCTGTAAGCG GATGCCGGGA GCAGACAAGC CCGTCAGGGC
2821   GCGTCAGCGG GTGTTGGCGG GTGTCGGGGC GCAGCCATGA CCCAGTCACG TAGCGATAGC
2881   GGAGTGTATA CTGGCTTAAC TATGCGGCAT CAGAGCAGAT TGTACTGAGA GTGCACCATA
2941   TGCGGTGTGA AATACCGCAC AGATGCGTAA GGAGAAAATA CCGCATCAGG CGCTCTTCCG
3001   CTTCCTCGCT CACTGACTCG CTGCGCTCGG TCGTTCGGCT GCGGCGAGCG GTATCAGCTC
3061   ACTCAAAGGC GGTAATACGG TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT
3121   GAGCAAAAGG CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
3181   ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG AGGTGGCGAA
3241   ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG AAGCTCCCTC GTGCGCTCTC
3301   CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG
3361   CGCTTTCTCA TAGCTCACGC TGTAGGTATC TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC
3421   TGGGCTGTGT GCACGAACCC CCCGTTCAGC CCGACCGCTG CGCCTTATCC GGTAACTATC
3481   GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT GGCAGCAGCC ACTGGTAACA
3541   GGATTAGCAG AGCGAGGTAT GTAGGCGGTG CTACAGAGTT CTTGAAGTGG TGGCCTAACT
3601   ACGGCTACAC TAGAAGGACA GTATTTGGTA TCTGCGCTCT GCTGAAGCCA GTTACCTTCG
3661   GAAAAAGAGT TGGTAGCTCT TGATCCGGCA AACAAACCAC CGCTGGTAGC GGTGGTTTTT
3721   TTGTTTGCAA GCAGCAGATT ACGCGCAGAA AAAAAGGATC TCAAGAAGAT CCTTTGATCT
3781   TTTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG TTAAGGGATT TTGGTCATGA
3841   GATTATCAAA AAGGATCTTC ACCTAGATCC TTTTAAATTA AAAATGAAGT TTTAAATCAA
3901   TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTACCA ATGCTTAATC AGTGAGGCAC
3961   CTATCTCAGC GATCTGTCTA TTTCGTTCAT CCATAGTTGC CTGACTCCCC GTCGTGTAGA
4021   TAACTACGAT ACGGGAGGGC TTACCATCTG GCCCCAGTGC TGCAATGATA CCGCGAGACC
4081   CACGCTCACC GGCTCCAGAT TTATCAGCAA TAAACCAGCC AGCCGGAAGG GCCGAGCGCA
```

```
4141   GAAGTGGTCC TGCAACTTTA TCCGCCTCCA TCCAGTCTAT TAATTGTTGC CGGGAAGCTA
4201   GAGTAAGTAG TTCGCCAGTT AATAGTTTGC GCAACGTTGT TGCCATTGCT ACAGGCATCG
4261   TGGTGTCACG CTCGTCGTTT GGTATGGCTT CATTCAGCTC CGGTTCCCAA CGATCAAGGC
4321   GAGTTACATG ATCCCCCATG TTGTGCAAAA AAGCGGTTAG CTCCTTCGGT CCTCCGATCG
4381   TTGTCAGAAG TAAGTTGGCC GCAGTGTTAT CACTCATGGT TATGGCAGCA CTGCATAATT
4441   CTCTTACTGT CATGCCATCC GTAAGATGCT TTTCTGTGAC TGGTGAGTAC TCAACCAAGT
4501   CATTCTGAGA ATAGTGTATG CGGCGACCGA GTTGCTCTTG CCCGGCGTCA ATACGGGATA
4561   ATACCGCGCC ACATAGCAGA ACTTTAAAAG TGCTCATCAT TGGAAAACGT TCTTCGGGGC
4621   GAAAACTCTC AAGGATCTTA CCGCTGTTGA GATCCAGTTC GATGTAACCC ACTCGTGCAC
4681   CCAACTGATC TTCAGCATCT TTTACTTTCA CCAGCGTTTC TGGGTGAGCA AAAACAGGAA
4741   GGCAAAATGC CGCAAAAAAG GGAATAAGGG CGACACGGAA ATGTTGAATA CTCATACTCT
4801   TCCTTTTTCA ATATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC GGATACATAT
4861   TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC CGAAAAGTGC
4921   CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAT AGGCGTATCA
4981   CGAGGCCCTT TCGTCTTCAC
```

1.2 N-terminal tagged CYP3A4 and CPR construct sequences (pMD004 vector)

1.2.1.1 FLN-CYP3A4 5′ flanking UTR and coding sequences

```
1      ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA
61     GGATCTCACC ATCACCATCA CCATGGGATC CCAGCAGCAG CGGAAATCAG TGGTCACATC
121    GTACGTTCCC CGATGGTTGG TACTTTCTAC CGCACCCCAA GCCCGGACGC AAAAGCGTTC
181    ATCGAAGTGG GTCAGAAAGT CAACGTGGGC GATACCCTGT GCATCGTTGA AGCCATGAAA
241    ATGATGAACC AGATCGAAGC GGACAAATCC GGTACCGTGA AAGCAATTCT GGTCGAAAGT
301    GGACAACCGG TAGAATTTGA CGAGCCGCTG GTCGTCATCG AGGGTGGCAG CGGTTCTGGC
361    CCCGCGGCCG CTATGGCTCT CATCCCAGAC TTGGCCATGG AAACCTGGCT TCTCCTGGCT
421    GTCAGCCTGG TGCTCCTCTA TCTATATGGA ACCCATTCAC ATGGACTTTT TAAGAAGCTT
481    GGAATTCCAG GGCCCACACC TCTGCCTTTT TTGGGAAATA TTTTGTCCTA CCATAAGGGC
541    TTTTGTATGT TTGACATGGA ATGTCATAAA AAGTATGGAA AAGTGTGGGG CTTTTATGAT
601    GGTCAACAGC CTGTGCTGGC TATCACAGAT CCTGACATGA TCAAAACAGT GCTAGTGAAA
661    GAATGTTATT CTGTCTTCAC AAACCGGAGG CCTTTTGGTC CAGTGGGATT TATGAAAAGT
721    GCCATCTCTA TAGCTGAGGA TGAAGAATGG AAGAGATTAC GATCATTGCT GTCTCCAACC
781    TTCACCAGTG GAAAACTCAA GGAGATGGTC CCTATCATTG CCCAGTATGG AGATGTGTTG
841    GTGAGAAATC TGAGGCGGGA AGCAGAGACA GGCAAGCCTG TCACCTTGAA AGACATCTTT
901    GGGGCCTACA GCATGGATGT GATCACTAGC ACATCATTTG GAGTGAACAT CGACTCTCTC
961    AACAATCCAC AAGACCCCTT TGTGGAAAAC ACCAAGAAGC TTTTAAGATT TGATTTTTTG
1021   GATCCATTCT TTCTCTCAAT AACAGTCTTT CCATTCCTCA TCCCAATTCT TGAAGTATTA
1081   AATATCTGTG TGTTTCCAAG AGAAGTTACA AATTTTTTAA GAAAATCTGT AAAAAGGATG
1141   AAAGAAAGTC GCCTCGAAGA TACACAAAAG CACCGAGTGG ATTTCCTTCA GCTGATGATT
1201   GACTCTCAGA ATTCAAAAGA AACTGAGTCC CACAAAGCTC TGTCCGATCT GGAGCTCGTG
1261   GCCCAATCAA TTATCTTTAT TTTTGCTGGC TATGAAACCA CGAGCAGTGT TCTCTCCTTC
1321   ATTATGTATG AACTGGCCAC TCACCCTGAT GTCCAGCAGA AACTGCAGGA GGAAATTGAT
1381   GCAGTTTTAC CCAATAAGGC ACCACCCACC TATGATACTG TGCTACAGAT GGAGTATCTT
1441   GACATGGTGG TGAATGAAAC GCTCAGATTA TTCCCAATTG CTATGAGACT TGAGAGGGTC
1501   TGCAAAAAAG ATGTTGAGAT CAATGGGATG TTCATTCCCA AAGGGGTGGT GGTGATGATT
1561   CCAAGCTATG CTCTTCACCG TGACCCAAAG TACTGGACAG AGCCTGAGAA GTTCCTCCCT
1621   GAAAGATTCA GCAAGAAGAA CAAGGACAAC ATAGATCCTT ACATATACAC ACCCTTTGGA
1681   AGTGGACCCA GAAACTGCAT TGGCATGAGG TTTGCTCTCA TGAACATGAA ACTTGCTCTA
1741   ATCAGAGTCC TTCAGAACTT CTCCTTCAAA CCTTGTAAAG AAACACAGAT CCCCCTGAAA
1801   TTAAGCTTAG GAGGACTTCT TCAACCAGAA AAACCCGTTG TTCTAAAGGT TGAGTCAAGG
1861   GATGGCACCG TAAGTGGAGC CTGA
```

1.2.1.2 FLN-CYP3A4 protein sequence

```
Total amino acid number: 609, MW=68381
1          MRGSHHHHHH GIPAAAEISG HIVRSPMVGT FYRTPSPDAK AFIEVGQKVN VGDTLCIVEA
61         MKMMNQIEAD KSGTVKAILV ESGQPVEFDE PLVVIEGGSG SGPAAAMALI PDLAMETWLL
121        LAVSLVLLYL YGTHSHGLFK KLGIPGPTPL PFLGNILSYH KGFCMFDMEC HKKYGKVWGF
```

```
181        YDGQQPVLAI TDPDMIKTVL VKECYSVFTN RRPFGPVGFM KSAISIAEDE EWKRLRSLLS
241        PTFTSGKLKE MVPIIAQYGD VLVRNLRREA ETGKPVTLKD IFGAYSMDVI TSTSFGVNID
301        SLNNPQDPFV ENTKKLLRFD FLDPFFLSIT VFPFLIPILE VLNICVFPRE VTNFLRKSIK
361        RMKESRLEDT QKHRVDFLQL MIDSQNSKET ESHKALSDLE LVAQSIIFIF AGYETTSSVL
421        SFIMYELATH PDVQQKLQEE IDAVLPNKAP PTYDTVLQME YLDMVVNETL RLFPIAMRLE
481        RVCKKDVEIN GMFIPKGVVV MIPSYALHRD PKYWTEPEKF LPERFSKKNK DNIDPYIYTP
541        FGSGPRNCIG MRFALMNMKL ALIRVLQNFS FKPCKETQIP LKLSLGGLLQ PEKPVVLKVE
601        SRDGTVSGA*
```

1.2.2.1 Δ24N-CYP3A4 5′ flanking UTR and coding sequences

```
1       ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA
61      GGATCTCACC ATCACCATCA CCATGGGATC CCAGCAGCAG CGGAAATCAG TGGTCACATC
121     GTACGTTCCC CGATGGTTGG TACTTTCTAC CGCACCCCAA GCCCGGACGC AAAAGCGTTC
181     ATCGAAGTGG GTCAGAAAGT CAACGTGGGC GATACCCTGT GCATCGTTGA AGCCATGAAA
241     ATGATGAACC AGATCGAAGC GGACAAATCC GGTACCGTGA AAGCAATTCT GGTCGAAAGT
301     GGACAACCGG TAGAATTTGA CGAGCCGCTG GTCGTCATCG AGGGTGGCAG CGGTTCTGGC
361     CCCGGGATGG CTTATGGAAC CCATTCACAT GGACTTTTTA AGAAGCTTGG AATTCCAGGG
421     CCCACACCTC TGCCTTTTTT GGGAAATATT TTGTCCTACC ATAAGGGCTT TTGTATGTTT
481     GACATGGAAT GTCATAAAAA GTATGGAAAA GTGTGGGGCT TTTATGATGG TCAACAGCCT
541     GTGCTGGCTA TCACAGATCC TGACATGATC AAAACAGTGC TAGTGAAAGA ATGTTATTCT
601     GTCTTCACAA ACCGGAGGCC TTTTGGTCCA GTGGGATTTA TGAAAAGTGC CATCTCTATA
661     GCTGAGGATG AAGAATGGAA GAGATTACGA TCATTGCTGT CTCCAACCTT CACCAGTGGA
721     AAACTCAAGG AGATGGTCCC TATCATTGCC CAGTATGGAG ATGTGTTGGT GAGAAATCTG
781     AGGCGGGAAG CAGAGACAGG CAAGCCTGTC ACCTTGAAAG ACATCTTTGG GGCCTACAGC
841     ATGGATGTGA TCACTAGCAC ATCATTTGGA GTGAACATCG ACTCTCTCAA CAATCCACAA
901     GACCCCTTTG TGGAAAACAC CAAGAAGCTT TTAAGATTTG ATTTTTTGGA TCCATTCTTT
961     CTCTCAATAA CAGTCTTTCC ATTCCTCATC CCAATTCTTG AAGTATTAAA TATCTGTGTG
1021    TTTCCAAGAG AAGTTACAAA TTTTTTAAGA AAATCTGTAA AAAGGATGAA AGAAAGTCGC
1081    CTCGAAGATA CACAAAAGCA CCGAGTGGAT TTCCTTCAGC TGATGATTGA CTCTCAGAAT
1141    TCAAAAGAAA CTGAGTCCCA CAAAGCTCTG TCCGATCTGG AGCTCGTGGC CCAATCAATT
1201    ATCTTTATTT TTGCTGGCTA TGAAACCACG AGCAGTGTTC TCTCCTTCAT TATGTATGAA
1261    CTGGCCACTC ACCCTGATGT CCAGCAGAAA CTGCAGGAGG AAATTGATGC AGTTTTACCC
1321    AATAAGGCAC CACCCACCTA TGATACTGTG CTACAGATGG AGTATCTTGA CATGGTGGTG
1381    AATGAAACGC TCAGATTATT CCCAATTGCT ATGAGACTTG AGAGGGTCTG CAAAAAAGAT
1441    GTTGAGATCA ATGGGATGTT CATTCCCAAA GGGGTGGTGG TGATGATTCC AAGCTATGCT
1501    CTTCACCGTG ACCCAAAGTA CTGGACAGAG CCTGAGAAGT TCCTCCCTGA AAGATTCAGC
1561    AAGAAGAACA AGGACAACAT AGATCCTTAC ATATACACAC CCTTTGGAAG TGGACCCAGA
1621    AACTGCATTG GCATGAGGTT TGCTCTCATG AACATGAAAC TTGCTCTAAT CAGAGTCCTT
1681    CAGAACTTCT CCTTCAAACC TTGTAAAGAA ACACAGATCC CCCTGAAATT AAGCTTAGGA
1741    GGACTTCTTC AACCAGAAAA ACCCGTTGTT CTAAAGGTTG AGTCAAGGGA TGGCACCGTA
1801    AGTGGAGCCT GA
```

1.2.2.2 Δ24N-CYP3A4 protein sequence

```
Total amino acid number: 585, MW=65746
1          MRGSHHHHHH GIPAAAEISG HIVRSPMVGT FYRTPSPDAK AFIEVGQKVN VGDTLCIVEA
61         MKMMNQIEAD KSGTVKAILV ESGQPVEFDE PLVVIEGGSG SGPGMAYGTH SHGLFKKLGI
121        PGPTPLPFLG NILSYHKGFC MFDMECHKKY GKVWGFYDGQ QPVLAITDPD MIKTVLVKEC
181        YSVFTNRRPF GPVGFMKSAI SIAEDEEWKR LRSLLSPTFT SGKLKEMVPI IAQYGDVLVR
241        NLRREAETGK PVTLKDIFGA YSMDVITSTS FGVNIDSLNN PQDPFVENTK KLLRFDFLDP
301        FFLSITVFPF LIPILEVLNI CVFPREVTNF LRKSIKRMKE SRLEDTQKHR VDFLQLMIDS
361        QNSKETESHK ALSDLELVAQ SIIFIFAGYE TTSSVLSFIM YELATHPDVQ QKLQEEIDAV
421        LPNKAPPTYD TVLQMEYLDM VVNETLRLFP IAMRLERVCK KDVEINGMFI PKGVVVMIPS
481        YALHRDPKYW TEPEKFLPER FSKKNKDNID PYIYTPFGSG PRNCIGMRFA LMNMKLALIR
541        VLQNFSFKPC KETQIPLKLS LGGLLQPEKP VVLKVESRDG TVSGA*
```

1.2.3.1 FLN-CPR 5′ UTR flanking and coding sequences

```
1       ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA
61      GGATCTCACC ATCACCATCA CCATGGGATC CCAGCAGCAG CGGAAATCAG TGGTCACATC
121     GTACGTTCCC CGATGGTTGG TACTTTCTAC CGCACCCCAA GCCCGGACGC AAAAGCGTTC
181     ATCGAAGTGG GTCAGAAAGT CAACGTGGGC GATACCCTGT GCATCGTTGA AGCCATGAAA
241     ATGATGAACC AGATCGAAGC GGACAAATCC GGTACCGTGA AAGCAATTCT GGTCGAAAGT
301     GGACAACCGG TAGAATTTGA CGAGCCGCTG GTCGTCATCG AGGGTGGCAG CGGTTCTGGC
361     CCCGGAGACT CCCACGTGGA CACCAGCTCC ACCGTGTCCG AGGCGGTGGC CGAAGAAGTA
421     TCTCTTTTCA GCATGACGGA CATGATTCTG TTTTCGCTCA TCGTGGGTCT CCTAACCTAC
481     TGGTTCCTCT TCAGAAAGAA AAAAGAAGAA GTCCCCGAGT TCACCAAAAT TCAGACATTG
541     ACCTCCTCTG TCAGAGAGAG CAGCTTTGTG GAAAAGATGA AGAAAACGGG GAGGAACATC
601     ATCGTGTTCT ACGGCTCCCA GACGGGGACT GCAGAGGAGT TTGCCAACCG CCTGTCCAAG
661     GACGCCCACC GCTACGGGAT GCGAGGCATG TCAGCGGACC CTGAGGAGTA TGACCTGGCC
721     GACCTGAGCA GCCTGCCAGA GATCGACAAC GCCCTGGTGG TTTTCTGCAT GGCCACCTAC
781     GGTGAGGGAG ACCCCACCGA CAATGCCCAG GACTTCTACG ACTGGCTGCA GGAGACAGAC
841     GTGGATCTCT CTGGGGTCAA GTTCGCGGTG TTTGGTCTTG GGAACAAGAC CTACGAGCAC
901     TTCAATGCCA TGGGCAAGTA CGTGGACAAG CGGCTGGAGC AGCTCGGCGC CCAGCGCATC
961     TTTGAGCTGG GGTTGGGCGA CGACGATGGG AACTTGGAGG AGGACTTCAT CACCTGGCGA
1021    GAGCAGTTCT GGCCGGCCGT GTGTGAACAC TTTGGGGTGG AAGCCACTGG CGAGGAGTCC
1081    AGCATTCGCC AGTACGAGCT TGTGGTCCAC ACCGACATAG ATGCGGCCAA GGTGTACATG
1141    GGGGAGATGG GCCGGCTGAA GAGCTACGAG AACCAGAAGC CCCCCTTTGA TGCCAAGAAT
1201    CCGTTCCTGG CTGCAGTCAC CACCAACCGG AAGCTGAACC AGGGAACCGA GCGCCACCTC
1261    ATGCACCTGG AATTGGACAT CTCGGACTCC AAAATCAGGT ATGAATCTGG GGACCACGTG
1321    GCTGTGTACC CAGCCAACGA CTCTGCTCTC GTCAACCAGC TGGGCAAAAT CCTGGGTGCC
1381    GACCTGGACG TCGTCATGTC CCTGAACAAC CTGGATGAGG AGTCCAACAA GAAGCACCCA
1441    TTCCCGTGCC CTACGTCCTA CCGCACGGCC CTCACCTACT ACCTGGACAT CACCAACCCG
1501    CCGCGTACCA ACGTGCTGTA CGAGCTGGCG CAGTACGCCT CGGAGCCCTC GGAGCAGGAG
1561    CTGCTGCGCA AGATGGCCTC CTCCTCCGGC GAGGGCAAGG AGCTGTACCT GAGCTGGGTG
1621    GTGGAGGCCC GGAGGCACAT CCTGGCCATC CTGCAGGACT GCCCGTCCCT GCGGCCCCCC
1681    ATCGACCACC TGTGTGAGCT GCTGCCGCGC CTGCAGGCCC GCTACTACTC CATCGCCTCA
1741    TCCTCCAAGG TCCACCCCAA CTCTGTGCAC ATCTGTGCGG TGGTTGTGGA GTACGAGACC
1801    AAGGCCGGCC GCATCAACAA GGGCGTGGCC ACCAACTGGC TGCGGGCCAA GGAGCCTGCC
1861    GGGGAGAACG GCGGCCGTGC GCTGGTGCCC ATGTTCGTGC GCAAGTCCCA GTTCCGCCTG
1921    CCCTTCAAGG CCACCACGCC TGTCATCATG GTGGGCCCCG GCACCGGGGT GGCACCCTTC
1981    ATAGGCTTCA TCCAGGAGCG GGCCTGGCTG CGACAGCAGG GCAAGGAGGT GGGGGAGACG
2041    CTGCTGTACT ACGGCTGCCG CCGCTCGGAT GAGGACTACC TGTACCGGGA GGAGCTGGCG
2101    CAGTTCCACA GGGACGGTGC GCTCACCCAG CTCAACGTGG CCTTCTCCCG GGAGCAGTCC
2161    CACAAGGTCT ACGTCCAGCA CCTGCTAAAG CAAGACCGAG AGCACCTGTG GAAGTTGATC
2221    GAAGGCGGTG CCCACATCTA CGTCTGTGGG GATGCACGGA ACATGGCCAG GGATGTGCAG
2281    AACACCTTCT ACGACATCGT GGCTGAGCTC GGGGCCATGG AGCACGCGCA GGCGGTGGAC
2341    TACATCAAGA AACTGATGAC CAAGGGCCGC TACTCCCTGG ACGTGTGGAG CTAG
```

1.2.3.2 FLN-CPR protein sequence

```
Total amino acid number: 779, MW=87350
1         MRGSHHHHHH GIPAAAEISG HIVRSPMVGT FYRTPSPDAK AFIEVGQKVN VGDTLCIVEA
61        MKMMNQIEAD KSGTVKAILV ESGQPVEFDE PLVVIEGGSG SGPGDSHVDT SSTVSEAVAE
121       EVSLFSMTDM ILFSLIVGLL TYWFLFRKKK EEVPEFTKIQ TLTSSVRESS FVEKMKKTGR
181       NIIVFYGSQT GTAEEFANRL SKDAHRYGMR GMSADPEEYD LADLSSLPEI DNALVVFCMA
241       TYGEGDPTDN AQDFYDWLQE TDVDLSGVKF AVFGLGNKTY EHFNAMGKYV DKRLEQLGAQ
301       RIFELGLGDD DGNLEEDFIT WREQFWPALC EHFGVEATGE ESSIRQYELV VHTDIDAAKV
361       YMGEMGRLKS YENQKPPFDA KNPFLAAVTT NRKLNQGTER HLMHLELDIS DSKIRYESGD
421       HVAVYPANDS ALVNQLGKIL GADLDVVMSL NNLDEESNKK HPFPCPTSYR TALTYYLDIT
481       NPPRTNVLYE LAQYASEPSE QELLRKMASS SGEGKELYLS WVVEARRHIL AILQDCPSLR
541       PPIDHLCELL PRLQARYYSI ASSSKVHPNS VHICAVVVEY ETKAGRINKG VATNWLRAKE
601       PAGENGGRAL VPMFVRKSQF RLPFKATTPV IMVGPGTGVA PFIGFIQERA WLRQQGKEVG
661       ETLLYYGCRR SDEDYLYREE LAQFHRDGAL TQLNVAFSRE QSHKVYVQHL LKQDREHLWK
721       LIEGGAHIYV CGDARNMARD VQNTFYDIVA ELGAMEHAQA VDYIKKLMTK GRYSLDVWS*
```

1.2.4.1 Δ43N-CPR 5′ UTR flanking and coding sequences

```
1       ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA
61      GGATCTCACC ATCACCATCA CCATGGGATC CCAGCAGCAG CGGAAATCAG TGGTCACATC
121     GTACGTTCCC CGATGGTTGG TACTTTCTAC CGCACCCCAA GCCCGGACGC AAAAGCGTTC
181     ATCGAAGTGG GTCAGAAAGT CAACGTGGGC GATACCCTGT GCATCGTTGA AGCCATGAAA
241     ATGATGAACC AGATCGAAGC GGACAAATCC GGTACCGTGA AAGCAATTCT GGTCGAAAGT
301     GGACAACCGG TAGAATTTGA CGAGCCGCTG GTCGTCATCG AGGGTGGCAG CGGTTCTGGC
361     CCCTTCAGAA AGAAAAAAGA AGAAGTCCCC GAGTTCACCA AAATTCAGAC ATTGACCTCC
421     TCTGTCAGAG AGAGCAGCTT TGTGGAAAAG ATGAAGAAAA CGGGGAGGAA CATCATCGTG
481     TTCTACGGCT CCCAGACGGG GACTGCAGAG GAGTTTGCCA ACCGCCTGTC CAAGGACGCC
541     CACCGCTACG GGATGCGAGG CATGTCAGCG GACCCTGAGG AGTATGACCT GGCCGACCTG
601     AGCAGCCTGC CAGAGATCGA CAACGCCCTG GTGGTTTTCT GCATGGCCAC CTACGGTGAG
661     GGAGACCCCA CCGACAATGC CCAGGACTTC TACGACTGGC TGCAGGAGAC AGACGTGGAT
721     CTCTCTGGGG TCAAGTTCGC GGTGTTTGGT CTTGGGAACA AGACCTACGA GCACTTCAAT
781     GCCATGGGCA AGTACGTGGA CAAGCGGCTG GAGCAGCTCG GCGCCCAGCG CATCTTTGAG
841     CTGGGGTTGG GCGACGACGA TGGGAACTTG GAGGAGGACT TCATCACCTG GCGAGAGCAG
901     TTCTGGCCGG CCGTGTGTGA ACACTTTGGG GTGGAAGCCA CTGGCGAGGA GTCCAGCATT
961     CGCCAGTACG AGCTTGTGGT CCACACCGAC ATAGATGCGG CCAAGGTGTA CATGGGGGAG
1021    ATGGGCCGGC TGAAGAGCTA CGAGAACCAG AAGCCCCCCT TTGATGCCAA GAATCCGTTC
1081    CTGGCTGCAG TCACCACCAA CCGGAAGCTG AACCAGGGAA CCGAGCGCCA CCTCATGCAC
1141    CTGGAATTGG ACATCTCGGA CTCCAAAATC AGGTATGAAT CTGGGGACCA CGTGGCTGTG
1201    TACCCAGCCA ACGACTCTGC TCTCGTCAAC CAGCTGGGCA AAATCCTGGG TGCCGACCTG
1261    GACGTCGTCA TGTCCCTGAA CAACCTGGAT GAGGAGTCCA ACAAGAAGCA CCCATTCCCG
1321    TGCCCTACGT CCTACCGCAC GGCCCTCACC TACTACCTGG ACATCACCAA CCCGCCGCGT
1381    ACCAACGTGC TGTACGAGCT GGCGCAGTAC GCCTCGGAGC CCTCGGAGCA GGAGCTGCTG
1441    CGCAAGATGG CCTCCTCCTC CGGCGAGGGC AAGGAGCTGT ACCTGAGCTG GGTGGTGGAG
1501    GCCCGGAGGC ACATCCTGGC CATCCTGCAG GACTGCCCGT CCCTGCGGCC CCCCATCGAC
1561    CACCTGTGTG AGCTGCTGCC GCGCCTGCAG GCCCGCTACT ACTCCATCGC CTCATCCTCC
1621    AAGGTCCACC CCAACTCTGT GCACATCTGT GCGGTGGTTG TGGAGTACGA GACCAAGGCC
1681    GGCCGCATCA ACAAGGGCGT GGCCACCAAC TGGCTGCGGG CCAAGGAGCC TGCCGGGGAG
1741    AACGGCGGCC GTGCGCTGGT GCCCATGTTC GTGCGCAAGT CCCAGTTCCG CCTGCCCTTC
1801    AAGGCCACCA CGCCTGTCAT CATGGTGGGC CCCGGCACCG GGGTGGCACC CTTCATAGGC
1861    TTCATCCAGG AGCGGGCCTG GCTGCGACAG CAGGGCAAGG AGGTGGGGGA GACGCTGCTG
1921    TACTACGGCT GCCGCCGCTC GGATGAGGAC TACCTGTACC GGGAGGAGCT GGCGCAGTTC
1981    CACAGGGACG GTGCGCTCAC CCAGCTCAAC GTGGCCTTCT CCCGGGAGCA GTCCCACAAG
2041    GTCTACGTCC AGCACCTGCT AAAGCAAGAC CGAGAGCACC TGTGGAAGTT GATCGAAGGC
2101    GGTGCCCACA TCTACGTCTG TGGGGATGCA CGGAACATGG CCAGGGATGT GCAGAACACC
2161    TTCTACGACA TCGTGGCTGA GCTCGGGGCC ATGGAGCACG CGCAGGCGGT GGACTACATC
2221    AAGAAACTGA TGACCAAGGG CCGCTACTCC CTGGACGTGT GGAGCTAG
```

1.2.4.2 Δ43N-CPR protein sequence

```
Total amino acid number: 737, MW=82761
1         MRGSHHHHHH GIPAAAEISG HIVRSPMVGT FYRTPSPDAK AFIEVGQKVN VGDTLCIVEA
61        MKMMNQIEAD KSGTVKAILV ESGQPVEFDE PLVVIEGGSG SGPFRKKKEE VPEFTKIQTL
121       TSSVRESSFV EKMKKTGRNI IVFYGSQTGT AEEFANRLSK DAHRYGMRGM SADPEEYDLA
181       DLSSLPEIDN ALVVFCMATY GEGDPTDNAQ DFYDWLQETD VDLSGVKFAV FGLGNKTYEH
241       FNAMGKYVDK RLEQLGAQRI FELGLGDDDG NLEEDFITWR EQFWPALCEH FGVEATGEES
301       SIRQYELVVH TDIDAAKVYM GEMGRLKSYE NQKPPFDAKN PFLAAVTTNR KLNQGTERHL
361       MHLELDISDS KIRYESGDHV AVYPANDSAL VNQLGKILGA DLDVVMSLNN LDEESNKKHP
421       FPCPTSYRTA LTYYLDITNP PRTNVLYELA QYASEPSEQE LLRKMASSSG EGKELYLSWV
481       VEARRHILAI LQDCPSLRPP IDHLCELLPR LQARYYSIAS SSKVHPNSVH ICAVVVEYET
541       KAGRINKGVA TNWLRAKEPA GENGGRALVP MFVRKSQFRL PFKATTPVIM VGPGTGVAPF
601       IGFIQERAWL RQQGKEVGET LLYYGCRRSD EDYLYREELA QFHRDGALTQ LNVAFSREQS
661       HKVYVQHLLK QDREHLWKLI EGGAHIYVCG DARNMARDVQ NTFYDIVAEL GAMEHAQAVD
721       YIKKLMTKGR YSLDVWS*
```

1.3 C-terminal tagged CYP3A4 and CPR construct sequences (pBJW102.2 vector)

1.3.1.1 FLC-CYP3A4 5′ UTR flanking and coding sequences

```
1       ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGGCA
61      CTTACGGCCG CTCTCATCCC AGACTTGGCC ATGGAAACCT GGCTTCTCCT GGCTGTCAGC
121     CTGGTGCTCC TCTATCTATA TGGAACCCAT TCACATGGAC TTTTTAAGAA GCTTGGAATT
181     CCAGGGCCCA CACCTCTGCC TTTTTTGGGA AATATTTTGT CCTACCATAA GGGCTTTTGT
241     ATGTTTGACA TGGAATGTCA TAAAAAGTAT GGAAAAGTGT GGGGCTTTTA TGATGGTCAA
301     CAGCCTGTGC TGGCTATCAC AGATCCTGAC ATGATCAAAA CAGTGCTAGT GAAAGAATGT
361     TATTCTGTCT TCACAAACCG GAGGCCTTTT GGTCCAGTGG GATTTATGAA AAGTGCCATC
421     TCTATAGCTG AGGATGAAGA ATGGAAGAGA TTACGATCAT TGCTGTCTCC AACCTTCACC
481     AGTGGAAAAC TCAAGGAGAT GGTCCCTATC ATTGCCCAGT ATGGAGATGT GTTGGTGAGA
541     AATCTGAGGC GGGAAGCAGA GACAGGCAAG CCTGTCACCT TGAAAGACAT CTTTGGGGCC
601     TACAGCATGG ATGTGATCAC TAGCACATCA TTTGGAGTGA ACATCGACTC TCTCAACAAT
661     CCACAAGACC CCTTTGTGGA AAACACCAAG AAGCTTTTAA GATTTGATTT TTTGGATCCA
721     TTCTTTCTCT CAATAACAGT CTTTCCATTC CTCATCCCAA TTCTTGAAGT ATTAAATATC
781     TGTGTGTTTC CAAGAGAAGT TACAAATTTT TTAAGAAAAT CTGTAAAAAG GATGAAAGAA
841     AGTCGCCTCG AAGATACACA AAAGCACCGA GTGGATTTCC TTCAGCTGAT GATTGACTCT
901     CAGAATTCAA AAGAAACTGA GTCCCACAAA GCTCTGTCCG ATCTGGAGCT CGTGGCCCAA
961     TCAATTATCT TTATTTTTGC TGGCTATGAA ACCACGAGCA GTGTTCTCTC CTTCATTATG
1021    TATGAACTGG CCACTCACCC TGATGTCCAG CAGAAACTGC AGGAGGAAAT TGATGCAGTT
1081    TTACCCAATA AGGCACCACC CACCTATGAT ACTGTGCTAC AGATGGAGTA TCTTGACATG
1141    GTGGTGAATG AAACGCTCAG ATTATTCCCA ATTGCTATGA GACTTGAGAG GGTCTGCAAA
1201    AAAGATGTTG AGATCAATGG GATGTTCATT CCCAAAGGGG TGGTGGTGAT GATTCCAAGC
1261    TATGCTCTTC ACCGTGACCC AAAGTACTGG ACAGAGCCTG AGAAGTTCCT CCCTGAAAGA
1321    TTCAGCAAGA AGAACAAGGA CAACATAGAT CCTTACATAT ACACACCCTT TGGAAGTGGA
1381    CCCAGAAACT GCATTGGCAT GAGGTTTGCT CTCATGAACA TGAAACTTGC TCTAATCAGA
1441    GTCCTTCAGA ACTTCTCCTT CAAACCTTGT AAAGAAACAC AGATCCCCCT GAAATTAAGC
1501    TTAGGAGGAC TTCTTCAACC AGAAAAACCC GTTGTTCTAA AGGTTGAGTC AAGGGATGGC
1561    ACCGTAAGTG GAGCCGGGGG TGGCAGCGGT TCTGGCGCAG CAGCGGAAAT CAGTGGTCAC
1621    ATCGTACGTT CCCCGATGGT TGGTACTTTC TACCGCACCC CAAGCCCGGA CGCAAAAGCG
1681    TTCATCGAAG TGGGTCAGAA AGTCAACGTG GGCGATACCC TGTGCATCGT TGAAGCCATG
1741    AAAATGATGA ACCAGATCGA AGCGGACAAA TCCGGTACCG TGAAAGCAAT TCTGGTCGAA
1801    AGTGGACAAC CGGTAGAATT TGACGAGCCG CTGGTCGTCA TCGAGGGTGG CAGCGGTTCT
1861    GGCCACCATC ACCATCACCA TAAGCTTAAT TAG
```

1.3.1.2 FLC-CYP3A4 protein sequence

```
Total amino acid number: 612, MW=68573
1         MALTAALIPD LAMETWLLLA VSLVLLYLYG THSHGLFKKL GIPGPTPLPF LGNILSYHKG
61        FCMFDMECHK KYGKVWGFYD GQQPVLAITD PDMIKTVLVK ECYSVFTNRR PFGPVGFMKS
121       AISIAEDEEW KRLRSLLSPT FTSGKLKEMV PIIAQYGDVL VRNLRREAET GKPVTLKDIF
181       GAYSMDVITS TSFGVNIDSL NNPQDPFVEN TKKLLRFDFL DPFFLSITVF PFLIPILEVL
241       NICVFPREVT NFLRKSIKRM KESRLEDTQK HRVDFLQLMI DSQNSKETES HKALSDLELV
301       AQSIIFIFAG YETTSSVLSF IMYELATHPD VQQKLQEEID AVLPNKAPPT YDTVLQMEYL
361       DMVVNETLRL FPIAMRLERV CKKDVEINGM FIPKGVVVMI PSYALHRDPK YWTEPEKFLP
421       ERFSKKNKDN IDPYIYTPFG SGPRNCIGMR FALMNMKLAL IRVLQNFSFK PCKETQIPLK
481       LSLGGLLQPE KPVVLKVESR DGTVSGAGGG SGSGAAAEIS GHIVRSPMVG TFYRTPSPDA
541       KAFIEVGQKV NVGDTLCIVE AMKMMNQIEA DKSGTVKAIL VESGQPVEFD EPLVVIEGGS
601       GSGHHHHHHK LN*
```

1.3.2.1 Δ24C-CYP3A4 5′ UTR flanking and coding sequences

```
1       ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGGCA
61      CTTACGGCCT ATGGAACCCA TTCACATGGA CTTTTTAAGA AGCTTGGAAT TCCAGGGCCC
121     ACACCTCTGC CTTTTTTGGG AAATATTTTG TCCTACCATA AGGGCTTTTG TATGTTTGAC
181     ATGGAATGTC ATAAAAAGTA TGGAAAAGTG TGGGGCTTTT ATGATGGTCA ACAGCCTGTG
```

```
241     CTGGCTATCA CAGATCCTGA CATGATCAAA ACAGTGCTAG TGAAAGAATG TTATTCTGTC
301     TTCACAAACC GGAGGCCTTT TGGTCCAGTG GGATTTATGA AAAGTGCCAT CTCTATAGCT
361     GAGGATGAAG AATGGAAGAG ATTACGATCA TTGCTGTCTC CAACCTTCAC CAGTGGAAAA
421     CTCAAGGAGA TGGTCCCTAT CATTGCCCAG TATGGAGATG TGTTGGTGAG AAATCTGAGG
481     CGGGAAGCAG AGACAGGCAA GCCTGTCACC TTGAAAGACA TCTTTGGGGC CTACAGCATG
541     GATGTGATCA CTAGCACATC ATTTGGAGTG AACATCGACT CTCTCAACAA TCCACAAGAC
601     CCCTTTGTGG AAAACACCAA GAAGCTTTTA AGATTTGATT TTTTGGATCC ATTCTTTCTC
661     TCAATAACAG TCTTTCCATT CCTCATCCCA ATTCTTGAAG TATTAAATAT CTGTGTGTTT
721     CCAAGAGAAG TTACAAATTT TTTAAGAAAA TCTGTAAAAA GGATGAAAGA AAGTCGCCTC
781     GAAGATACAC AAAAGCACCG AGTGGATTTC CTTCAGCTGA TGATTGACTC TCAGAATTCA
841     AAAGAAACTG AGTCCCACAA AGCTCTGTCC GATCTGGAGC TCGTGGCCCA ATCAATTATC
901     TTTATTTTTG CTGGCTATGA AACCACGAGC AGTGTTCTCT CCTTCATTAT GTATGAACTG
961     GCCACTCACC CTGATGTCCA GCAGAAACTG CAGGAGGAAA TTGATGCAGT TTTACCCAAT
1021    AAGGCACCAC CCACCTATGA TACTGTGCTA CAGATGGAGT ATCTTGACAT GGTGGTGAAT
1081    GAAACGCTCA GATTATTCCC AATTGCTATG AGACTTGAGA GGGTCTGCAA AAAAGATGTT
1141    GAGATCAATG GGATGTTCAT TCCCAAAGGG GTGGTGGTGA TGATTCCAAG CTATGCTCTT
1201    CACCGTGACC CAAAGTACTG GACAGAGCCT GAGAAGTTCC TCCCTGAAAG ATTCAGCAAG
1261    AAGAACAAGG ACAACATAGA TCCTTACATA TACACACCCT TTGGAAGTGG ACCCAGAAAC
1321    TGCATTGGCA TGAGGTTTGC TCTCATGAAC ATGAAACTTG CTCTAATCAG AGTCCTTCAG
1381    AACTTCTCCT TCAAACCTTG TAAAGAAACA CAGATCCCCC TGAAATTAAG CTTAGGAGGA
1441    CTTCTTCAAC CAGAAAAACC CGTTGTTCTA AAGGTTGAGT CAAGGGATGG CACCGTAAGT
1501    GGAGCCGGGG GTGGCAGCGG TTCTGGCGCA GCAGCGGAAA TCAGTGGTCA CATCGTACGT
1561    TCCCCGATGG TTGGTACTTT CTACCGCACC CCAAGCCCGG ACGCAAAAGC GTTCATCGAA
1621    GTGGGTCAGA AAGTCAACGT GGGCGATACC CTGTGCATCG TTGAAGCCAT GAAAATGATG
1681    AACCAGATCG AAGCGGACAA ATCCGGTACC GTGAAAGCAA TTCTGGTCGA AAGTGGACAA
1741    CCGGTAGAAT TTGACGAGCC GCTGGTCGTC ATCGAGGGTG GCAGCGGTTC TGGCCACCAT
1801    CACCATCACC ATAAGCTTAA TTAG
```

1.3.2.2 Δ24C-CYP3A4 protein sequence

```
Total amino acid number: 589, MW=66023
1         MALTAYGTHS HGLFKKLGIP GPTPLPFLGN ILSYHKGFCM FDMECHKKYG KVWGFYDGQQ
61        PVLAITDPDM IKTVLVKECY SVFTNRRPFG PVGFMKSAIS IAEDEEWKRL RSLLSPTFTS
121       GKLKEMVPII AQYGDVLVRN LRREAETGKP VTLKDIFGAY SMDVITSTSF GVNIDSLNNP
181       QDPFVENTKK LLRFDFLDPF FLSITVFPFL IPILEVLNIC VFPREVTNFL RKSIKRMKES
241       RLEDTQKHRV DFLQLMIDSQ NSKETESHKA LSDLELVAQS IIFIFAGYET TSSVLSFIMY
301       ELATHPDVQQ KLQEEIDAVL PNKAPPTYDT VLQMEYLDMV VNETLRLFPI AMRLERVCKK
361       DVEINGMFIP KGVVVMIPSY ALHRDPKYWT EPEKFLPERF SKKNKDNIDP YIYTPFGSGP
421       RNCIGMRFAL MNMKLALIRV LQNFSFKPCK ETQIPLKLSL GGLLQPEKPV VLKVESRDGT
481       VSGAGGGSGS GAAAEISGHI VRSPMVGTFY RTPSPDAKAF IEVGQKVNVG DTLCIVEAMK
541       MMNQIEADKS GTVKAILVES GQPVEFDEPL VVIEGGSGSG HHHHHHKLN*
```

1.3.3.1 FLC-CPR 5′ UTR flanking and coding sequences

```
1       ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGGCA
61      CTTACGGCCG GAGACTCCCA CGTGGACACC AGCTCCACCG TGTCCGAGGC GGTGGCCGAA
121     GAAGTATCTC TTTTCAGCAT GACGGACATG ATTCTGTTTT CGCTCATCGT GGGTCTCCTA
181     ACCTACTGGT TCCTCTTCAG AAAGAAAAAA GAAGAAGTCC CCGAGTTCAC CAAAATTCAG
241     ACATTGACCT CCTCTGTCAG AGAGAGCAGC TTTGTGGAAA AGATGAAGAA AACGGGGAGG
301     AACATCATCG TGTTCTACGG CTCCCAGACG GGGACTGCAG AGGAGTTTGC CAACCGCCTG
361     TCCAAGGACG CCCACCGCTA CGGGATGCGA GGCATGTCAG CGGACCCTGA GGAGTATGAC
421     CTGGCCGACC TGAGCAGCCT GCCAGAGATC GACAACGCCC TGGTGGTTTT CTGCATGGCC
481     ACCTACGGTG AGGGAGACCC CACCGACAAT GCCCAGGACT TCTACGACTG GCTGCAGGAG
541     ACAGACGTGG ATCTCTCTGG GGTCAAGTTC GCGGTGTTTG GTCTTGGGAA CAAGACCTAC
601     GAGCACTTCA ATGCCATGGG CAAGTACGTG GACAAGCGGC TGGAGCAGCT CGGCGCCCAG
661     CGCATCTTTG AGCTGGGGTT GGGCGACGAC GATGGGAACT TGGAGGAGGA CTTCATCACC
721     TGGCGAGAGC AGTTCTGGCC GGCCGTGTGT GAACACTTTG GGGTGGAAGC CACTGGCGAG
781     GAGTCCAGCA TTCGCCAGTA CGAGCTTGTG GTCCACACCG ACATAGATGC GGCCAAGGTG
841     TACATGGGGG AGATGGGCCG GCTGAAGAGC TACGAGAACC AGAAGCCCCC CTTTGATGCC
```

```
901     AAGAATCCGT TCCTGGCTGC AGTCACCACC AACCGGAAGC TGAACCAGGG AACCGAGCGC
961     CACCTCATGC ACCTGGAATT GGACATCTCG GACTCCAAAA TCAGGTATGA ATCTGGGGAC
1021    CACGTGGCTG TGTACCCAGC CAACGACTCT GCTCTCGTCA ACCAGCTGGG CAAAATCCTG
1081    GGTGCCGACC TGGACGTCGT CATGTCCCTG AACAACCTGG ATGAGGAGTC CAACAAGAAG
1141    CACCCATTCC CGTGCCCTAC GTCCTACCGC ACGGCCCTCA CCTACTACCT GGACATCACC
1201    AACCCGCCGC GTACCAACGT GCTGTACGAG CTGGCGCAGT ACGCCTCGGA GCCCTCGGAG
1261    CAGGAGCTGC TGCGCAAGAT GGCCTCCTCC TCCGGCGAGG GCAAGGAGCT GTACCTGAGC
1321    TGGGTGGTGG AGGCCCGGAG GCACATCCTG GCCATCCTGC AGGACTGCCC GTCCCTGCGG
1381    CCCCCCATCG ACCACCTGTG TGAGCTGCTG CCGCGCCTGC AGGCCCGCTA CTACTCCATC
1441    GCCTCATCCT CCAAGGTCCA CCCCAACTCT GTGCACATCT GTGCGGTGGT TGTGGAGTAC
1501    GAGACCAAGG CCGGCCGCAT CAACAAGGGC GTGGCCACCA ACTGGCTGCG GGCCAAGGAG
1561    CCTGCCGGGG AGAACGGCGG CCGTGCGCTG GTGCCCATGT TCGTGCGCAA GTCCCAGTTC
1621    CGCCTGCCCT TCAAGGCCAC CACGCCTGTC ATCATGGTGG GCCCCGGCAC CGGGGTGGCA
1681    CCCTTCATAG GCTTCATCCA GGAGCGGGCC TGGCTGCGAC AGCAGGGCAA GGAGGTGGGG
1741    GAGACGCTGC TGTACTACGG CTGCCGCCGC TCGGATGAGG ACTACCTGTA CCGGGAGGAG
1801    CTGGCGCAGT TCCACAGGGA CGGTGCGCTC ACCCAGCTCA ACGTGGCCTT CTCCCGGGAG
1861    CAGTCCCACA AGGTCTACGT CCAGCACCTG CTAAAGCAAG ACCGAGAGCA CCTGTGGAAG
1921    TTGATCGAAG GCGGTGCCCA CATCTACGTC TGTGGGGATG CACGGAACAT GGCCAGGGAT
1981    GTGCAGAACA CCTTCTACGA CATCGTGGCT GAGCTCGGGG CCATGGAGCA CGCGCAGGCG
2041    GTGGACTACA TCAAGAAACT GATGACCAAG GGCCGCTACT CCCTGGACGT GTGGAGCGGG
2101    GGTGGCAGCG GTTCTGGCGC AGCAGCGGAA ATCAGTGGTC ACATCGTACG TTCCCCGATG
2161    GTTGGTACTT TCTACCGCAC CCCAAGCCCG GACGCAAAAG CGTTCATCGA AGTGGGTCAG
2221    AAAGTCAACG TGGGCGATAC CCTGTGCATC GTTGAAGCCA TGAAAATGAT GAACCAGATC
2281    GAAGCGGACA AATCCGGTAC CGTGAAAGCA ATTCTGGTCG AAAGTGGACA ACCGGTAGAA
2341    TTTGACGAGC CGCTGGTCGT CATCGAGGGT GGCAGCGGTT CTGGCCACCA TCACCATCAC
2401    CATAAGCTTA ATTAG
```

1.3.3.2 FLC-CPR protein sequence

```
Total amino acid number: 786, MW=87856

1          MALTAGDSHV DTSSTVSEAV AEEVSLFSMT DMILFSLIVG LLTYWFLFRK KKEEVPEFTK
61         IQTLTSSVRE SSFVEKMKKT GRNIIVFYGS QTGTAEEFAN RLSKDAHRYG MRGMSADPEE
121        YDLADLSSLP EIDNALVVFC MATYGEGDPT DNAQDFYDWL QETDVDLSGV KFAVFGLGNK
181        TYEHFNAMGK YVDKRLEQLG AQRIFELGLG DDDGNLEEDF ITWREQFWPA LCEHFGVEAT
241        GEESSIRQYE LVVHTDIDAA KVYMGEMGRL KSYENQKPPF DAKNPFLAAV TTNRKLNQGT
301        ERHLMHLELD ISDSKIRYES GDHVAVYPAN DSALVNQLGK ILGADLDVVM SLNNLDEESN
361        KKHPFPCPTS YRTALTYYLD ITNPPRTNVL YELAQYASEP SEQELLRKMA SSSGEGKELY
421        LSWVVEARRH ILAILQDCPS LRPPIDHLCE LLPRLQARYY SIASSSKVHP NSVHICAVVV
481        EYETKAGRIN KGVATNWLRA KEPAGENGGR ALVPMFVRKS QFRLPFKATT PVIMVGPGTG
541        VAPFIGFIQE RAWLRQQGKE VGETLLYYGC RRSDEDYLYR EELAQFHRDG ALTQLNVAFS
601        REQSHKVYVQ HLLKQDREHL WKLIEGGAHI YVCGDARNMA RDVQNTFYDI VAELGAMEHA
661        QAVDYIKKLM TKGRYSLDVW SGGGSGSGAA AEISGHIVRS PMVGTFYRTP SPDAKAFIEV
721        GQKVNVGDTL CIVEAMKMMN QIEADKSGTV KAILVESGQP VEFDEPLVVI EGGSGSGHHH
781        HHHKLN*
```

1.3.4.1 Δ43C-CPR 5′ UTR flanking and coding sequences

```
1       ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGGCA
61      CTTACGGCCT TCAGAAAGAA AAAAGAAGAA GTCCCCGAGT TCACCAAAAT TCAGACATTG
121     ACCTCCTCTG TCAGAGAGAG CAGCTTTGTG GAAAAGATGA AGAAAACGGG GAGGAACATC
181     ATCGTGTTCT ACGGCTCCCA GACGGGGACT GCAGAGGAGT TTGCCAACCG CCTGTCCAAG
241     GACGCCCACC GCTACGGGAT GCGAGGCATG TCAGCGGACC CTGAGGAGTA TGACCTGGCC
301     GACCTGAGCA GCCTGCCAGA GATCGACAAC GCCCTGGTGG TTTTCTGCAT GGCCACCTAC
361     GGTGAGGGAG ACCCCACCGA CAATGCCCAG GACTTCTACG ACTGGCTGCA GGAGACAGAC
421     GTGGATCTCT CTGGGGTCAA GTTCGCGGTG TTTGGTCTTG GGAACAAGAC CTACGAGCAC
481     TTCAATGCCA TGGGCAAGTA CGTGGACAAG CGGCTGGAGC AGCTCGGCGC CCAGCGCATC
541     TTTGAGCTGG GGTTGGGCGA CGACGATGGG AACTTGGAGG AGGACTTCAT CACCTGGCGA
601     GAGCAGTTCT GGCCGGCCGT GTGTGAACAC TTTGGGGTGG AAGCCACTGG CGAGGAGTCC
```

```
661     AGCATTCGCC AGTACGAGCT TGTGGTCCAC ACCGACATAG ATGCGGCCAA GGTGTACATG
721     GGGGAGATGG GCCGGCTGAA GAGCTACGAG AACCAGAAGC CCCCCTTTGA TGCCAAGAAT
781     CCGTTCCTGG CTGCAGTCAC CACCAACCGG AAGCTGAACC AGGGAACCGA GCGCCACCTC
841     ATGCACCTGG AATTGGACAT CTCGGACTCC AAAATCAGGT ATGAATCTGG GGACCACGTG
901     GCTGTGTACC CAGCCAACGA CTCTGCTCTC GTCAACCAGC TGGGCAAAAT CCTGGGTGCC
961     GACCTGGACG TCGTCATGTC CCTGAACAAC CTGGATGAGG AGTCCAACAA GAAGCACCCA
1021    TTCCCGTGCC CTACGTCCTA CCGCACGGCC CTCACCTACT ACCTGGACAT CACCAACCCG
1081    CCGCGTACCA ACGTGCTGTA CGAGCTGGCG CAGTACGCCT CGGAGCCCTC GGAGCAGGAG
1141    CTGCTGCGCA AGATGGCCTC CTCCTCCGGC GAGGGCAAGG AGCTGTACCT GAGCTGGGTG
1201    GTGGAGGCCC GGAGGCACAT CCTGGCCATC CTGCAGGACT GCCCGTCCCT GCGGCCCCCC
1261    ATCGACCACC TGTGTGAGCT GCTGCCGCGC CTGCAGGCCC GCTACTACTC CATCGCCTCA
1321    TCCTCCAAGG TCCACCCCAA CTCTGTGCAC ATCTGTGCGG TGGTTGTGGA GTACGAGACC
1381    AAGGCCGGCC GCATCAACAA GGGCGTGGCC ACCAACTGGC TGCGGGCCAA GGAGCCTGCC
1441    GGGGAGAACG GCGGCCGTGC GCTGGTGCCC ATGTTCGTGC GCAAGTCCCA GTTCCGCCTG
1501    CCCTTCAAGG CCACCACGCC TGTCATCATG GTGGGCCCCG GCACCGGGGT GGCACCCTTC
1561    ATAGGCTTCA TCCAGGAGCG GGCCTGGCTG CGACAGCAGG GCAAGGAGGT GGGGGAGACG
1621    CTGCTGTACT ACGGCTGCCG CCGCTCGGAT GAGGACTACC TGTACCGGGA GGAGCTGGCG
1681    CAGTTCCACA GGGACGGTGC GCTCACCCAG CTCAACGTGG CCTTCTCCCG GGAGCAGTCC
1741    CACAAGGTCT ACGTCCAGCA CCTGCTAAAG CAAGACCGAG AGCACCTGTG GAAGTTGATC
1801    GAAGGCGGTG CCCACATCTA CGTCTGTGGG GATGCACGGA ACATGGCCAG GGATGTGCAG
1861    AACACCTTCT ACGACATCGT GGCTGAGCTC GGGGCCATGG AGCACGCGCA GGCGGTGGAC
1921    TACATCAAGA AACTGATGAC CAAGGGCCGC TACTCCCTGG ACGTGTGGAG CGGGGGTGGC
1981    AGCGGTTCTG GCGCAGCAGC GGAAATCAGT GGTCACATCG TACGTTCCCC GATGGTTGGT
2041    ACTTTCTACC GCACCCCAAG CCCGGACGCA AAAGCGTTCA TCGAAGTGGG TCAGAAAGTC
2101    AACGTGGGCG ATACCCTGTG CATCGTTGAA GCCATGAAAA TGATGAACCA GATCGAAGCG
2161    GACAAATCCG GTACCGTGAA AGCAATTCTG GTCGAAAGTG GACAACCGGT AGAATTTGAC
2221    GAGCCGCTGG TCGTCATCGA GGGTGGCAGC GGTTCTGGCC ACCATCACCA TCACCATAAG
2281    CTTAATTAG
```

1.3.4.2 Δ43C-CPR protein sequence

```
Total amino acid number: 744, MW=83267
1          MALTAFRKKK EEVPEFTKIQ TLTSSVRESS FVEKMKKTGR NIIVFYGSQT GTAEEFANRL
61         SKDAHRYGMR GMSADPEEYD LADLSSLPEI DNALVVFCMA TYGEGDPTDN AQDFYDWLQE
121        TDVDLSGVKF AVFGLGNKTY EHFNAMGKYV DKRLEQLGAQ RIFELGLGDD DGNLEEDFIT
181        WREQFWPALC EHFGVEATGE ESSIRQYELV VHTDIDAAKV YMGEMGRLKS YENQKPPFDA
241        KNPFLAAVTT NRKLNQGTER HLMHLELDIS DSKIRYESGD HVAVYPANDS ALVNQLGKIL
301        GADLDVVMSL NNLDEESNKK HPFPCPTSYR TALTYYLDIT NPPRTNVLYE LAQYASEPSE
361        QELLRKMASS SGEGKELYLS WVVEARRHIL AILQDCPSLR PPIDHLCELL PRLQARYYSI
421        ASSSKVHPNS VHICAVVVEY ETKAGRINKG VATNWLRAKE PAGENGGRAL VPMFVRKSQF
481        RLPFKATTPV IMVGPGTGVA PFIGFIQERA WLRQQGKEVG ETLLYYGCRR SDEDYLYREE
541        LAQFHRDGAL TQLNVAFSRE QSHKVYVQHL LKQDREHLWK LIEGGAHIYV CGDARNMARD
601        VQNTFYDIVA ELGAMEHAQA VDYIKKLMTK GRYSLDVWSG GGSGSGAAAE ISGHIVRSPM
661        VGTFYRTPSP DAKAFIEVGQ KVNVGDTLCI VEAMKMMNQI EADKSGTVKA ILVESGQPVE
721        FDEPLVVIEG GSGSGHHHHH HKLN*
```

1.4 Primer sequences

The primer sets used to create the FLN-CYP3A4 and Δ24N-CYP3A4 constructs were as follows.

1.4.1 FLN-CYP3A4

Normal PCR primers were used to amplify the full length CYP3A4 off the FLC-CYP3A4 plasmid for sub-cloning into the pMD004 vector. Sticky end ligation was used (restriction enzyme sites underlined).

Forward: 5′-GTGTCATGCGGCCGCTATGGCTCTCATCCCAGAC-3′ (NotI RE site –**check** RE seqs for all.)

Reverse: 5′-GCAACGTGTCGACTCAGGCTCCACTTACGGTGCC-3′ (SalI RE site)

1.4.2 Δ24N-CYP3A4

Inverse PCR primers were used to create the Δ24N-CYP3A4 construct from the Δ34N-CYP3A4 plasmid for sub-cloning into the pMD004 vector. Blunt end ligation was used.

Forward: 5′-CCCGGGATGGCTTATGGAACCCATTCACATGGACTTTTTAAGAAGCTTGGA ATTCC-3′ (SmaI RE site introduced)

Reverse: 5′-GCCAGAACCGCTGCCACCCTCGATGACGACC-3′

Note: Two non-synonymous mutations were present in the CYP3A4 sequences (V175I and V253I) while one such mutation was present in the CPR sequences (V226L). These mutations were already present in the parent plasmids obtained from Procognia and were confirmed during construct sequencing in our laboratory (personal communication, Jonathan Blackburn). None of these mutations are located near the enzyme active site nor in the SRSs of CYP3A4 and are therefore unlikely to affect protein function. The mutations are reflected in the above protein sequences.

II. KINETIC DATA

2.1 Vivid Red kinetic data

Below is the Prism results table of the non-linear regression analysis of data sets discussed in Section 6.2.1.2, showing the comparison between the fits of the Michaelis-Menten and biphasic kinetic models using censored and uncensored data sets. Note: V_{max} and K_m units are nM.min^{-1} and nM respectively.

All 12 data points		**11 data points**		**10 data points**	
Michaelis-Menten		*Michaelis-Menten*		*Michaelis-Menten*	
Best-fit values		Best-fit values		Best-fit values	
*V*max	0.2837	*V*max	0.2962	*V*max	0.2306
*K*m	8076	*K*m	8824	*K*m	5370
Std. Error		Std. Error		Std. Error	
*V*max	0.02077	*V*max	0.03155	*V*max	0.02175
*K*m	1598	*K*m	2189	*K*m	1218
95% Confidence Intervals		95% Confidence Intervals		95% Confidence Intervals	
*V*max	0.2374 to 0.3300	*V*max	0.2248 to 0.3675	*V*max	0.1804 to 0.2808
*K*m	4516 to 11636	*K*m	3872 to 13776	*K*m	2560 to 8179
Goodness of Fit		Goodness of Fit		Goodness of Fit	
Degrees of Freedom	10	Degrees of Freedom	9	Degrees of Freedom	8

R^2	0.9575	R^2	0.9468	R^2	0.9533
Abs. Sum of squares	0.002742	Abs. Sum of squares	0.002664	Abs. sum of squares	0.001295
Sy.x	0.01656	Sy.x	0.0172	Sy.x	0.01272
Runs test		Runs test		Runs test	
Expected	5	Expected	4	Expected	5
Observed	4	Observed	3	Observed	7
P value (runs test)	0.1091	P value (runs test)	0.06667	P value (runs test)	0.881
Deviation from Model	Not Significant	Deviation from Model	Not Significant	Deviation from Model	Not Significant
Biphasic		*Biphasic*		*Biphasic*	
Best-fit values		Best-fit values		Best-fit values	
V_{max1}	0.2062	V_{max1}	0.1104	V_{max1}	0.1144
V_{max2}	9.04E-07	V_{max2}	7.59E-06	V_{max2}	2.25E-05
K_{m2}	0.5506	K_{m2}	1.435	K_{m2}	4.463
K_{m1}	4683	K_{m1}	1615	K_{m1}	1715
Std. Error		Std. Error		Std. Error	
V_{max1}	0.05175	V_{max1}	0.01976	V_{max1}	0.02983
K_{m1}	2203	K_{m1}	639.9	K_{m1}	879.8
95% Confidence Intervals		95% Confidence Intervals		95% Confidence Intervals	
V_{max1}	0.08692 to 0.3256	V_{max1}	0.06369 to 0.1571	V_{max1}	0.04141 to 0.1874
K_{m1}	-398.0 to 9763	K_{m1}	101.7 to 3129	K_{m1}	-437.9 to 3868
Goodness of Fit		Goodness of Fit		Goodness of Fit	
Degrees of Freedom	8	Degrees of Freedom	7	Degrees of Freedom	6
R^2	0.9621	R^2	0.9872	R^2	0.9771
Abs. Sum of squares	0.002445	Abs. Sum of squares	0.0006397	Abs. sum of squares	0.0006352
Sy.x	0.01748	Sy.x	0.00956	Sy.x	0.01029
Runs test		Runs test		Runs test	
Expected	6	Expected	5	Expected	5
Observed	8	Observed	7	Observed	7
P value (runs test)	0.8535	P value (runs test)	0.7381	P value (runs test)	0.8333
Deviation from Model	Not Significant	Deviation from Model	Not Significant	Deviation from Model	Not Significant
Comparison of Fits		*Comparison of Fits*		*Comparison of Fits*	
P value	0.632	P value	0.0068	P value	0.1181
Best Fit Equation (penalised)	Michaelis-Menten	Best Fit Equation (penalised)	Biphasic	Best Fit Equation (penalised)	Michaelis-Menten

2.2 Ketoconazole kinetic data

Below is the Prism results table of the linear regression analysis of apparent K_m vs. [ketoconazole] for the array-based inhibition assay of Vivid Red turnover as discussed in Section 6.2.2, showing the comparitive fits of the Michaelis-Menten and biphasic kinetic models.

	Michaelis-Menten	Biphasic
Variables		
Slope	0.00128 ± 0.000865	0.00267 ± 0.000630
Y-intercept	2400 ± 307	1770 ± 224
X-intercept	-1880000	-663000
1/slope	782	375
95% Confidence Intervals		
Slope	-0.000838 to 0.00340	0.00113 to 0.00421

Y-intercept	1650 to 3150	1220 to 2320
Goodness of Fit		
r^2	0.267	0.749
Sy.x	800	583
Is slope significantly non-zero?		
F	2.19	17.9
DFn, DFd	1.00, 6.00	1.00, 6.00
P value	0.1897	0.0055
Deviation from zero?	Not Significant	Significant
Runs test		
Points above line	3	4
Points below line	5	4
Number of runs	6	6
P value (runs test)	0.9286	0.8857
Significantly nonlinear?	Not Significant	Not Significant

REFERENCES

Aguiar, M, Masse, R, and Gibbs, BF. Regulation of cytochrome P450 by posttranslational modification. Drug Metabolism Reviews 37, 379-404. 2005.

Alvares, AP, Schilling, G, Levin, W, and Kuntzman, R. Studies on the induction of CO-binding pigments in liver microsomes by phenobarbital and 3-methylcholanthrene. Biochemical and Biophysical Research Communications 29[4], 521-526. 1967.

Ansede, JH and Thakker, DR. High-throughput screening for stability and inhibitory activity of compounds toward cytochrome P450-mediated metabolism. Journal of pharmaceutical sciences 93[2], 239-255. 2004.

Athappilly, FK and Hendrickson, WA. Structure of the biotinyl domain of acetyl-coenzyme A carboxylase determined by MAD phasing. Structure 3[12], 1407-1419. 1995.

Atkins, WM. Implications of the allosteric kinetics of Cytochrome P450s. Drug Discovery Today 9[11], 478-484. 2004.

Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J and Struhl K (1999) *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., USA.

Axelrod, J. The enzymatic demethylation of ephedrine. Journal of Pharmacology and Experimental Therapeutics 114[4], 430-438. 1955.

Azari, MR and Wiseman, A. Solubilisation of Cytochrome P450 in a high yield from *Saccharomyces cerevisiae* microsomal membrane: stabilisation effect. Biochemical Society Transactions 8, 713-714. 1980.

Backes, WL and Kelley, RW. Organisation of multiple cytochrome P450s with NADPH-cytochrome P450 reductase in membranes. Pharmacology & Therapeutics 98, 221-233. 2003.

Baess, D, Jänig, GR, and Ruckpaul, K. Interaction of the components of the Cytochrome P450 monooxygenase system from liver microsomes. I: Immobilisation of the solubilised and partially purified protein components. Acta Biologica et Medica Germanica 34[11-12], 1745-1754. 1975.

Baldrick, P. Safety evaluation to support first-in-man investigations II: toxicology studies. Regulatory Toxicology and Pharmacology 51, 237-243. 2008a.

Baldrick, P. Safety evaluation to support first-in-man investigations I: kinetic and safety pharmacology studies. Regulatory Toxicology and Pharmacology 51, 230-236. 2008b.

Banks, R and Selby, P. Clinical proteomics - insights into pathologies and benefits for patients. Lancet 362, 415-416. 2003.

Barnes, HJ. Maximising expression of eukaryotic Cytochrome P450s in *Escherichia coli*. Methods in Enzymology 272, 3-14. 1996.

Barnes, HJ, Arlotto, MP, and Waterman, MR. Expression and enzymatic activity of recombinant cytochrome P450 17α-hydroxylase in *Escherichia coli*. Proceedings of the National Academy of Science 88, 5597-5601. 1991.

Barron, JA, Young, HD, Dlott, DD, Darfler, MM, Krizman, DB, and Ringeisen, BR. Printing of protein microarrays via a capillary-free fluid jetting mechanism. Proteomics 5, 4138-4144. 2005.

Benet, LZ, Cummins, CL, and Wu, CY. Unmasking the dynamic interplay between efflux transporters and metabolic enzymes. International Journal of Pharmaceutics 277, 3-9. 2004.

Bhathena, A and Spear, BB. Pharmacogenetics: improving drug and dose selection. Current Opinion in Pharmacology 8, 639-646. 2008.

Binkert, A, Studer, P, and Vörös, J. A microwell array platform for picoliter membrane protein assays. Small 5[9], 1070-1077. 2009.

Black, SD and Coon, MJ. Structural features of liver microsomal NADPH-cytochrome P450 reductase. The Journal of Biological Chemistry 257[10], 5929-5938. 1982.

Blake, JAR, Pritchard, M, Ding, S, Smith, GCM, Burchell, B, Wolf, CR, and Friedberg, T. Coexpression of a human P450 (CYP3A4) and P450 reductase generates a highly functional monooxygenase system in *Escherichia coli*. FEBS Letters 397, 210-214. 1996.

Blake, JF. Identification and evaluation of molecular properties related to preclinical optimisation and clinical fate. Medicinal Chemistry 1, 649-655. 2005.

Bollinger, JM and Krebs, C. Stalking intermediates in oxygen activation by iron enzymes: motivation and method. Journal of Inorganic Biochemistry 103, 586-605. 2006.

Bourrie M, Meunier V, Berger Y and Fabre G. Cytochrome P450 Isoform Inhibitors As a Tool for the Investigation of Metabolic Reactions Catalyzed by Human Liver Microsomes. Journal of Pharmacology and Experimental Therapeutics 277, 321-332. 1996.

Boutell, JM, Hart, DJ, Godber, BLJ, Kozlowski, RZ, and Blackburn, JM. Functional protein microarrays for parallel characterisation of p53 mutants. Proteomics 4[7], 1950-1958. 2004.

Braun, T, Ghatkesar, MK, Backmann, N, Grange, W, Boulanger, P, Letellier, L, Lang, HP, Bietsch, A, Gerber, C, and Hegner, M. Quantitative time-resolved measurement of membrane protein-ligand interactions using microcantilever array sensors. Nature Nanotechnology 4[3], 179-185. 2009.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball C and Causton H. Minimum Information About a Microarray Experiment (MIAME): Toward Standards for Microarray Data. Nature Genetics 29, 365-371. 2001.

Bridges, A, Gruenke, L, Chang, Y, Vakser, IA, Loew, G, and Waskell, L. Identification of the binding site on cytochrome P450 2B4 for cytochrome b_5 and cytochrome P450 reductase. The Journal of Biological Chemistry 273[27], 17036-17049. 1998.

Brockmöller, J and Tzvetkov, VM. Pharmacogenetics: data, concepts and tools to improve drug discovery and drug treatment. European Journal of Clinical Pharmacology 64, 133-157. 2008.

Brodie, BB, Axelrod, J, Cooper, JR, Gaudette, L, La Du, BN, Mitoma, C, and Udenfriend, S. Detoxification of drugs and other foreign compounds in liver microsomes. Science 121, 603-604. 1955.

Bromberg, Y and Rost, B. Correlating protein function and stability through the analysis of single amino acid substitutions. BMC Bioinformatics 10[Suppl8], S8. 2009.

Carson, PE, Alving, AS, Flanagan, CL, and Ickes, CE. Enzymatic deficiency in primaquine-sensitive erythrocytes. Science 124[3220], 484-485. 1956.

Caskey, CT. The drug development crisis: efficiency and safety. Annual Review of Medicine 58[1], 16. 2007.

Cha, T, Guo, A, and Zhu, X. Enzymatic activity on a chip: the critical role of protein orientation. Proteomics 5, 416-419. 2005.

Chapman RG, Ostuni E, Takayama S, Holmlin R E, Yan L and Whitesides G M. Surveying for Surfaces That Resist the Adsorption of Proteins. Journal of the American Chemical Society 122, 8303-8304. 2000.

Chapman-Smith A and Cronan Jr J. Molecular Biology of Biotin Attachment to Proteins. Journal of Nutrition 129, 477S. 1999.

Chapman-Smith A and Cronan J E. In Vivo Enzymatic Protein Biotinylation. Biomolecular Engineering 16, 119-125. 1999.

Chefson, A, Zhao, J, and Auclair, K. Replacement of natural cofactors by selected hydrogen peroxide donors or organic peroxides results in improved activity for CYP3A4 and CYP2D6. ChemBioChem 7, 916-919. 2006.

Chen, GT and Inouye, M. Supression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. Nucleic Acids Research 18[6], 1465-1473. 1990.

Chen, GYJ, Uttamchandani, M, Lue, RYP, Lesaicherre, M, and Yao, SQ. Array-based technologies and their applications in proteomics. Current Topics in Medicinal Chemistry 3, 705-724. 2003.

Cheng, S. Formula for the viscosity of a glycerol-water mixture. Industrial Engineering & Chemical Research 47, 3285-3288. 2008.

Christians, U. Transport proteins and intestinal metabolism: P-glycoprotein and cytochrome P4503A. Therapeutic Drug Monitoring 26[2], 104-106. 2004.

Cockburn I (2006) Is the pharmaceutical industry in a productivity crisis?, in *NBER/Innovation Policy and Economy* (National Bureau of Economic Research ed) pp 1-32, MIT Press, Cambridge, Massachusetts.

Cohen, LH, Remley, MJ, Raung, D, and Vaz, ADN. *In vitro* drug interactions of cytochrome P450: an evaluation of fluorogenic to conventional substrates. Drug Metabolism and Disposition 31[8], 1005-1015. 2003.

Correia, MA, Sadeghi, S, and Mundo-Paredes, E. Cytochrome P450 ubiquitination: branding for the proteolytic slaughter? Annual Reviews in Pharmacology & Toxicology 45, 439-464. 2005.

Cronan J and Waldrop G. Multi-Subunit Acetyl-CoA Carboxylases. Progress in Lipid Research 41, 407-435. 2002.

Cvrk, T and Strobel, HW. Role of Lys271 and Lys279 residues in the interaction of cytochrome P4501A1 with NADPH-cytochrome P450 reductase. Archives of Biochemistry and Biophysics 385[2], 290-300. 2001.

Danielson, PB. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. Current Drug Metabolism 3, 561-597. 2002.

Davydov, DR, Darovsky, BV, Dedinsky, IR, Kanaeva, IP, Bachmanova, GI, Blinov, VM, and Archakov, AI. Cytochrome c (Fe2+) as a competitive inhbitor of NADPH-dependent reduction of cytochrome P450 LM2: locating protein-protein interaction sites in microsomal electron carriers. Archives of Biochemistry and Biophysics 297[2], 304-313. 1992.

de Smit, M and van Duin, J. Control of translation by mRNA secondary structure in *Escherichia coli*. Journal of Molecular Biology 244, 144-150. 1994.

de Visser, SP. Density functional theory (DFT) and combined quantum mechanical/molecular mechanics (QM/MM) studies on the oxygen activation step in nitric oxide synthase enzymes. Biochemical Society Transactions 37[2], 373-377. 2009.

Dervieux, T, Meshkin, B, and Neri, B. Pharmacogenetic testing: proofs of principle and pharmacoeconomic implications. Mutation Research 573, 180-194. 2005.

Diaz-Mochón, JJ, Tourniaire, G, and Bradley, M. Microarray platforms for enzymatic and cell-based assays. Chemical Society Reviews 36, 449-457. 2007.

DiMasi, JA, Hansen, RW, and Grabowski, HG. The price of innovation: new estimates of drug development costs. Journal of Health Economics 22, 151-185. 2003.

Dodhia VR, Fantuzzi A and Gilardi G. Engineering Human Cytochrome P450 Enzymes into Catalytically Self-Sufficient Chimeras Using Molecular Lego. Journal of Biological Inorganic Chemistry 11, 903-916. 2006.

Domanski, TL, He, Y, Khan, KK, Roussel, F, Wang, Q, and Halpert, JR. Phenylalanine and tryptophan scanning mutagenesis of CYP3A4 substrate recognition site residues and effect on substrate oxidation and cooperativity. Biochemistry 40, 10150-10160. 2001.

Donato, MT and Castell, JV. Strategies and molecular probes to investigate the role of cytochrome P540 in drug metabolism. Clinical Pharmacokinetics 42[2], 153-178. 2003.

Donato, MT, Jiménez, N, Castell, JV, and Gómez-Lechón, MJ. Florescence-based assays for screening nine cytochrome P450 (P450) activities in intact cells expressing individual human P450 enzymes. Drug Metabolism and Disposition 32[7], 699-706. 2004.

Donato, MT, Lahoz, A, Castell, JV, and Gómez-Lechón, MJ. Cell lines: a tool for *in vitro* drug metabolism studies. Current Drug Metabolism 9, 1-11. 2008.

Dong, J and Porter, TD. Coespression of mammalian cytochrome P450 and reductase in *Escherichia coli*. Archives of Biochemistry and Biophysics 327[2], 254-259. 1996.

Doong, R and Shih, H. Glutamate optical biosensor based on the immobilisation of glutamate dehydrogenase in titanium dioxide sol-gel matrix. Biosensors and Bioelectronics 22, 185-191. 2006.

Drews J. Drug Discovery: a Historical Perspective. Science 287, 1960-1964. 2000.

Ebbing D and Gammon S (1999) *General Chemistry*. Houghton Mifflin Company, New York.

Ekins, RP. Multi-analyte immunoassay. Journal of Pharmaceutical & Biomedical Analysis 7, 155-168. 1989.

Ekroos, M and Sjörgren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. Proceedings of the National Academy of Science 103[37], 13682-13687. 2006.

Estabrook, RW. A passion for P450s (remembrances of the early history of research on Cytochrome P450). Drug Metabolism and Disposition 31[12], 1461-1473. 2003.

Fang, Y, Frutos, AG, and Lahiri, J. Membrane protein microarrays. Journal of the American Chemical Society 124[11], 2394-2395. 2002.

Fantuzzi, A, Fairhead, M, and Gillardi, G. Direct electrochemistry of immobilised human cytochrome P450 2E1. Journal of the American Chemistry Society 126, 5040-5041. 2004.

FDA. Challenge and opportunity on the critical path to new medical products. Food and Drug Administration March. 2006a. http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm113411.pdf

FDA. Guidance for industry: drug interaction studies - study design, data analysis, and implications for dosing and labeling. Food and Drug Administration . 2006b. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM072101.pdf

FDA. Microarray Quality Control (MAQC). Food and Drug Administration . 2009. http://www.fda.gov/ScienceResearch/BioinformaticsTools/MicroarrayQualityControlProject/default.htm

Feinberg, JG. A microspot test for antigen and antibodies. Nature 192, 985-986. 1961.

Feinberg, JG and Wheeler, AW. Detection of auto-immune antibody and tissue antigens by the microspot technique. British Medical Journal 16, 282-284. 1963.

Feinstein, MB, Fernandez, SM, and Sha'afi, RI. Fluidity of natural membranes and phosphatidylserine and ganglioside dispersions. Effects of local anaesthetics, cholesterol and protein. Biochimica et Biophysica Acta 413, 354-370. 1975.

Fisher, CW, Shet, MS, Caudle, DL, Martin-Wixtrom, CA, and Estabrook, RW. High-level expression in *Escherichia coli* of enzymatically active fusion proteins containing the domains of mammalian cytochromes P450 and NADPH-P450 reductase flavoprotein. Proceedings of the National Academy of Science 89, 10817-10821. 1992.

Fletcher, AP. Spontaneous adverse drug reaction reporting vs. event monitoring: a comparison. Journal of the Royal Society of Medicine 84, 341-344. 1991.

Fodor S, Read J, Pirrung M, Stryer L, Lu A and Solas D. Light-Directed, Spatially Addressable Parallel Chemical Synthesis. Science 251, 767-773. 1991.

Fujita, K and Kamataki, T. Genetically engineered bacterial cells co-expressing human cytochrome P450 with NADPH-cytochrome P450 reductase: prediction of metabolism and toxicity of drugs in humans. Drug Metabolism and Pharmacokinetics 17[1], 1-22. 2002.

Galetin, A, Clarke, SE, and Houston, JB. Quinidine and haloperidol as modifiers of CYP3A4 activity: multisite kinetic model approach. Drug Metabolism and Disposition 30[12], 1512-1522. 2002.

Galetin, A, Clarke, SE, and Houston, JB. Multisite kinetic analysis of interactions between prototypical CYP3A4 subgroup substrates: midazolam, testosterone, and nifedipene. Drug Metabolism and Disposition 31[9], 1108-1116. 2003.

Galonic, DP, Barr, EW, Walsh, CT, Bollinger, JM, and Krebs, C. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. Nature Chemical Biology 3, 113-116. 2007.

Gibbs MA, Thummel K E, Shen D D and Kunze K L. Inhibition of Cytochrome P-450 3A (CYP3A) in Human Intestinal and Liver Microsomes: Comparison of K_i Values and Impact of CYP3A5 Expression. Drug Metabolism and Disposition 27, 180-187. 1999.

Gilbert, J, Henske, P, and Singh, A. Rebuilding big pharma's business model. Business Med Rep 21[10], ? 2003.

Gill I and Ballesteros A. Bioencapsulation Within Synthetic Polymers (Part 1): Sol-Gel Encapsulated Biologicals. TRENDS in Biotechnology 18, 282-296. 2000.

Gillam, EMJ. Human Cytochrome P450 enzymes expressed in bacteria: reagents to probe molecular interactions in toxicology. Clinical and Experimental Pharmacology and Physiology 25, 877-886. 1998.

Gillam, EMJ, Baba, T, Kim, B, Ohmori, S, and Guengerich, FP. Expression of modified human cytochrome P450 3A4 in *Escherichia coli* and the purification and reconstitution of the enzyme. Archives of Biochemistry and Biophysics 305[1], 123-131. 1993.

Glurich, I, Burmester, JK, and Caldwell, MD. Understanding the pharmacogenetic approach to warfarin dosing. Heart Failure Reviews 15[3], 239-248. 2008.

Gonzalez, FJ and Nebert, DW. Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. TRENDS in Genetics 6[6], 182-186. 1990.

Gonzalez M, Bagatolli L A, Echabe I, Arrondo J L R, Argarana C E, Cantor C R and Fidelio G D. Interaction of Biotin With Streptavidin: Thermostability and Conformational Changes Upon Binding. Journal of Biological Chemistry 272, 11288-11294. 1997.

Gosalia, DN and Diamond, SL. Printing chemical libraries on microarrays for fluid phase nanolitre reactions. Proceedings of the National Academy of Science 100[15], 8721-8726. 2003.

Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. The Journal of Biological Chemistry 267[1], 83-90. 1992.

Graham, SE and Peterson, JA. How similar are P450s and what can their differences teach us. Archives of Biochemistry and Biophysics 369, 24-29. 1999.

Guengerich, FP. Cytochrome P450: what have we learned and what are the future issues? Drug Metabolism Reviews 36[2], 159-197. 2004.

Guengerich, FP. Cytochromes P450, drugs and diseases. Molecular Interventions 3[4], 194-204. 2003.

Guengerich, FP. A malleable catalyst dominates the metabolism of drugs. Proceedings of the National Academy of Science 103[37], 13565-13566. 2006.

Guengerich, FP, Martin, MV, Guo, Z, and Chun, Y. Purification of functional recombinant P450s from bacteria. Methods in Enzymology 272, 35-44. 1996.

Guengerich, FP, Parikh, A, Johnson, EF, Richardson, TH, Von Wachenfeldt, C, Cosme, J, Jung, F, Strassburg, CP, Manns, MP, Tukey, RH, Pritchard, M, Fournel-Gigleux, S, and Burchell, B. Heterlogous expression of human drug-metabolizing enzymes. Drug Metabolism and Disposition 25[11], 1234-1241. 1997.

Guex, N and Peitsch, MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714-2723. 1997.

Guilleaume, B, Bruneß, A, Schmidt, C, Klimek, F, Moldenhauer, G, Huber, W, Arlt, D, Korf, U, Wiemann, S, and Poustka, A. Systematic comparison of surface coatings for protein microarrays. Proteomics 5, 4705-4712. 2005.

Gutierrez, A, Paine, M, Wolf, CR, Scrutton, NS, and Roberts, GCK. Relaxation kinetics of cytochrome P450 reductase: internal electron transfer is limited by conformational change and regulated by coenzyme binding. Biochemistry 41, 4626-4637. 2002.

Güzey, C and Spigset, O. Genotyping as a tool to predict adverse drug reactions. Current Topics in Medicinal Chemistry 4, 1409-1419. 2004.

Hamby, L, Weeks, WB, and Malikowski, C. Complications of warfarin therapy: causes, costs, and the role of the anticoagulation clinic. Effective Clinical Practice 3, 179-184. 2000.

Hamdane, D, Xia, C, Im, SC, Zhang, H, Kim, JJ, and Waskell, L. Structure and function of an NADPH-cytochrome P450 oxidoreductase in an open conformation capable of reducing cytochrome P450. J Biol Chem 284, 11374-11384. 2009.

Hanna I, Dawling S and Roodi N. Cytochrome P 1B1 (CYP1B1) Pharmacogenetics: Association of Polymorphisms With Functional Differences in Estrogen Hydroxylation Activity. Cancer Research 60, 3440-3444. 2000.

Hannig, G and Makrides, SC. Strategies for optimizing heterologous protein expression in *Escherichia coli*. Tibtech 16, 54-60. 1998.

Hao, M, Zhao, Y, Chen, P, Huang, H, Liu, H, Jiang, H, Zhang, R, and Wang, H. Structure-activity relationship and substrate-dependent phenomena in effects of ginsenosides on activities of drug-metabolising P450 enzymes. PLoS ONE 3[7], e2697-1-e2697-10. 2008.

Hara, M, Iazvovskaia, S, Ohkawa, H, Asada, Y, and Miyake, J. Immobilisation of P450 monooxygenases and chloroplast for use in light-driven bioreactors. Journal of Bioscience and Bioengineering 87[6], 793-797. 1999.

Hartmann, M, Roeraade, J, Stoll, D, Templin, MF, and Joos, TO. Protein microarrays for diagnostic assays. Analytcal & Bioanalytical Chemistry 393, 1407-1416. 2009.

He, M and Taussig, MJ. Rapid discovery of protein interactions by cell-free protein technologies. Biochemical Society Transactions 35[5], 962-965. 2007.

He, M and Wang, M. Arraying proteins by cell-free synthesis. Biomolecular Engineering 24, 375-380. 2007.

He, YA, Roussel, F, and Halpert, JR. Analysis of homotropic and heterotropic cooperativity of diazepam oxidation by CYP3A4 using site-directed mutagenesis and kinetic modeling. Archives of Biochemistry and Biophysics 409, 92-101. 2003.

Headlam, MJ and Tuckey, RC. The effect of glycerol on cytochrome P450scc (CYP11A1) spin state, activity, and hydration. Archives of Biochemistry and Biophysics 407, 95-102. 2002.

Hernandez-Boussard, T, Whirl-Carrillo, M, Hebert, JM, Gong, L, Owen, R, Gong, M, Gor, W, Liu, F, Truong, C, Whaley, R, Woon, M, Zhou, T, Altman, RB, and Klein, TE. The pharmacogenetics and pharmacogenomics database: accentuating the knowledge. Nucleic Acids Research 36[Database Issue], D913-D918. 2008.

Hiller, R, Laffer, S, Harwanegg, C, Huber, M, Schmidt, WM, and et al. Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. FASEB Journal 16, 414-416. 2002.

Hirota, T, Takane, H, Higuchi, S, and Ieri, I. Epigenetic regulation of genes encoding drug metabolising enzymes and transporters; DNA methylation and other mechanisms. Current Drug Metabolism 9, 34-38. 2008.

Hlavica, P. Models and mechanisms of O-O bond activation by cytochrome P450. European Journal of Biochemistry 271, 4335-4360. 2004.

Hober, S and Uhlén, M. Human protein atlas and the use of microarray technologies. Current Opinion in Biotechnology 19, 30-35. 2008.

Hodges, C. Regulating risk or advancing therapies? Regulation and sustainability of medicines in a cash-limited economy. European Business Law Review 19[2], 365-386. 2008.

Hodneland CD, Lee Y S, Min D H and Mrksich M. Selective Immobilization of Proteins to Self-Assembled Monolayers Presenting Active Site-Directed Capture Ligands. Proceedings of the National Academy of Sciences 99, 5048-5052. 2002.

Hoffart, LM, Barr, EW, Guyer, RB, Bollinger, JM, and Krebs, C. Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. Proceedings of the National Academy of Science 103, 14738-14743. 2006.

Hou, T, Wang, J, Zhang, W, Wang, W, and Xu, X. Recent advances in computational prediction of drug absorption and permeability in drug discovery. Current Medicinal Chemistry 13, 2653-2667. 2006.

Houseman, BT and Mrksich, M. Towards quantitative assays with peptide chips: a surface engineering approach. TRENDS in Biotechnology 20[7], 279-281. 2002.

Houston, JB and Galetin, A. Modelling atypical CYP3A4 kinetics: principles and pragmatism. Archives of Biochemistry and Biophysics 433, 351-360. 2005.

Hsu, HY, Wittemann, S, Schneider, EM, Weiss, M, and Joos, TO. Suspension microarrays for the identification of the response patterns in hyperinflammatory diseases. Medical Engineering & Physics 30[8], 976-983. 2009.

Huang, S, Temple, R, Throckmorton, DC, and Lesko, LJ. Drug interaction studies: study design, data analysis, and implications for dosing and labeling. Clinical Pharmacology & Therapeutics 81[2], 298-304. 2007.

Hummel, MA, Gannett, PM, and Tracy, TS. Evaluation of proton to heme distances of flurbiprofen and dapsone within the active site of CYP2C9. 12th North American ISSX Meeting. 201. 2003.

Hunt, I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. Protein Expression and Purification 40, 1-22. 2005.

Hutzler JM, Hauer M J and Tracy T S. Dapsone Activation of CYP2C9-Mediated Metabolism: Evidence for Activation of Multiple Substrates and a Two-Site Model. Drug Metabolism and Disposition 29, 1029-1034. 2001.

Hvastkovs, EG, So, M, Krishnan, S, Bajrami, B, Tarun, M, Jansson, I, Schenkman, JB, and Rusling, JF. Electrochemiluminescent arrays for Cytochrome P450-activated genotoxicity screening: DNA damage from benzo[*a*]pyrene metabolites. Analytical Chemistry 79[5], 1897-1906. 2007.

Ibrahim, M, Decolin, M, Bett, AM, Dellarcherie, E, and Siest, G. Immobilisation of pig liver microsomes: stability of cytochrome P450-dependent monooxygenase activities. Applied Biochemistry & Biotechnology 12[3], 199-213. 1986.

Ingelman-Sundberg, M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. TRENDS in Pharmacological Sciences 25[4], 193-200. 2004.

Ingelman-Sundberg, M. The human genome project and novel aspects of cytochrome P450 research. Toxicology and Applied Pharmacology 207, S52-S56. 2005.

Ingelman-Sundberg, M. Pharmacogenomic biomarkers for prediction of severe adverse drug reactions. New England Journal of Medicine 358[6], 637-639. 2008.

Inouye, K, Mizokawa, T, Saito, A, Tonomura, B, and Ohkawa, H. Biphasic kinetic behaviour of rat cytochrome P450 1A1-dependent monooxygenation in recombinant yeast microsomes. Biochimica et Biophysica Acta 1481, 265-272. 2000.

Ioannides, C and Lewis, DFV. Cytochromes P450 in the bioactivation of chemicals. Current Topics in Medicinal Chemistry 4, 1767-1788. 2004.

Iwata, H, Fujita, K, Kushida, H, Suzuki, A, Konno, Y, Nakamura, K, Fujino, A, and Kamataki, T. High catalytic activity of human cytochrome P450 co-expressed with human NADPH-cytochrome P450 reductase in *Escherichia coli*. Biochemical Pharmacology 55, 1315-1325. 1998.

Jackman, RJ, Duffy, DC, Ostuni, E, Willmore, ND, and Whitesides, GM. Fabricating large arrays of microwells with arbitrary dimensions and filling them using discontinuous dewetting. Analytical Chemistry 70, 2280-2287. 1998.

Jänig, GR, Pfell, D, and Ruckpaul, K. Enzymatic activities of matrix-bound components of the liver microsomal Cytochrome P450 system. Acta Biologica et Medica Germanica 38[2-3], 409-422. 1979.

Johnson E, Kronbach T and Hsu M. Analysis of the Catalytic Specificity of Cytochrome P450 Enzymes Through Site-Directed Mutagenesis. The FASEB Journal 6, 700-705. 1992.

Jonkheijm, P, Weinrich, D, Schroder, H, Niemeyer, CM, and Waldmann, H. Chemical strategies for generating protein biochips . Angewandte Chemie 47[50], 9618-9647. 2008.

Kalow, W, Tang, BK, and Endrenyi, L. Hypothesis comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. Pharmacogenetics 8, 283-289. 1998.

Kang, L, Chung, BG, Langer, R, and Khademhosseini, A. Microfluidics for drug discovery and development: from target selection to product lifecycle management. Drug Discovery Today 12[1-2], 1-13. 2008.

Karyakin, A, Motiejunas, D, Wade, RC, and Jung, C. FTIR studies of the redox partner interaction in cytochrome P450: the Pdx-P450cam couple. Biochimica et Biophysica Acta 1770, 420-431. 2007.

Katakam M, Bell L N and Banga A K. Effect of Surfactants on the Physical Stability of Recombinant Human Growth Hormone. Journal of pharmaceutical sciences 84, 713-716. 1995.

Kempf, AC, Zanger, UM, and Meyer, UA. Truncated human P450 2D6: expression in *Escherichia coli*, Ni^{2+}-chelate affinity purification, and characterisation of solubility and aggregation. Archives of Biochemistry and Biophysics 321[2], 277-288. 1995.

Kenworthy K, Bloomer J, Clarke S and Houston J. CYP3A4 Drug Interactions: Correlation of Ten *in Vitro* Probe Substrates. British Journal of Clinical Pharmacology 48, 716-727. 1999.

Kim KH, Ahn T and Yun C H. Membrane Properties Induced by Anionic Phospholipids and Phosphatidylethanolamine Are Critical for the Membrane Binding and Catalytic Activity of Human Cytochrome P450 3A4. Biochemistry 42, 15377-15387. 2003.

Klingenberg, M. Pigments of rat liver microsomes. Archives of Biochemistry and Biophysics 75[2], 376-386. 1958.

Koch, WH. Technology platforms for pharmacogenomic diagnostic assays. Nature Reviews Drug Discovery 3, 749-761. 2004.

Kolata, G. Merck and Vioxx: the overview; A widely used arthritis drug is withdrawn. New York Times 1 October. 2004. New York, The New York Times Company.

Koley, AP, Buters, JTM, Robinson, RC, Markowitz, A, and Friedman, FK. CO binding kinetics of human cytochrome P450 3A4. The Journal of Biological Chemistry 270[10], 5014-5018. 1995.

Korf, U, Kohl, T, van der Zandt, H, Zahn, R, Schleeger, S, Ueberle, B, Wandschneider, S, Bechtel, S, Schnölzer, M, Ottleberr, H, Wiemann, S, and Poustka, A. Large-scale protein expression for proteome research. Proteomics 5, 3571-3580. 2005.

Korzekwa, KR, Krishnamachary, N, Shou, M, Ogai, A, Parise, RA, Rettie, AE, Gonzalez, FJ, and Tracy, TS. Evaluation of atypical cytochrome P450 kinetics with two-substrate models: evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites. Biochemistry 37, 4137-4147. 1998.

Kramer, JA, Sagartz, JE, and Morris, DL. The application of discovery toxicology and pathology towards the desing of safer pharmaceutical lead candidates. Nature Reviews Drug Discovery 6, 636-649. 2007.

Kreilgaard L, Frokjaer S, Flink J M, Randolph T W and Carpenter J F. Effects of Additives on the Stability of *Humicola Lanuginosa* Lipase During Freeze-Drying and Storage in the Dried Solid. Journal of pharmaceutical sciences 88, 281-290. 1999.

Kricka, LJ and Master, SR. Validation and quality control of protein microarray-based analytical methods. Molecular Biotechnology 38[19], 31. 2008.

Kronbach T and Johnson E. An Inhibitory Monoclonal Antibody Binds in Close Proximity to a Determinant for Substrate Binding in Cytochrome P450IIC5. Journal of Biological Chemistry 266, 6215-6220. 1991.

Kumble, KD. Protein microarrays: new tools for pharmaceutical development. Analytcal & Bioanalytical Chemistry 377, 812-819. 2003.

Kusano, K, Kagawa, N, Sakaguchi, M, Omura, T, and Waterman, MR. Importance of a proline-rich sequence in the amino-terminal region for correct folding of mitochondrial and soluble microbial P450s. Journal of Biochemistry (Tokyo) 129, 271-277. 2001a.

Kusano, K, Sakaguchi, M, Kagawa, N, Waterman, MR, and Omura, T. Microsomal P450s use specific proline-rich sequences for efficient folding, but not for maintenance of the folded structure. Journal of Biochemistry (Tokyo) 129, 259-269. 2001b.

Lahoz, A, Donato, MT, Castell, JV, and Gómez-Lechón, MJ. Strategies to *in vitro* assessment of major human CYP enzyme activities using liquid chromatography tandem mass spectrometry. Current Drug Metabolism 9, 12-19. 2008.

Lam, KS. New aspects of natural products in drug discovery. TRENDS in Microbiology 15[6], 279-289. 2007.

Lamba, JK, Lin, YS, Schuetz, EG, and Thummel, KE. Genetic contribution to variable human CYP3A-mediated metabolism. Advanced Drug Delivery Reviews 54, 1271-1294. 2002.

Larson, JR, Coon, MJ, and Porter, TD. Purification and properties of the shortened form of cytochrome P-450 2E1: deletion of the NH_2-terminal membrane-insertion signal peptide does not alter the catalytic activities. Proceedings of the National Academy of Science 88, 9141-9145. 1991.

Lazarou, J, Pomeranz, BH, and Corey, PN. Incidence of adverse drug reactions in hospitalised patients: meta-analysis of prospective studies. JAMA 279, 1200-1205. 1998.

Lebert, JM, Forsberg, EM, and Brennan, JD. Solid-phase assays for small molecule screening using sol-gel entrapped proteins. Biochemistry & Cell Biology 86, 100-110. 2008.

Lee, M, Park, CB, Dordick, JS, and Clark, DS. Metabolizing enzyme toxicology assay chip (MetaChip) for high-throughput microscale toxicity analyses. Proceedings of the National Academy of Science 102[4], 983-987. 2005.

Lehman, JP, Ferrin, L, Fenselau, C, and Yost, GS. Simultaneous immobilisation of Cytochrome P450 and glucuronyltransferase for synthesis of drug metabolites. Drug Metabolism and Disposition 9[1], 15-18. 1981.

Leifert, WR, Aloia, AL, Bucco, O, Glatz, RV, and McMurchie, EJ. G-protein coupled receptors in drug discovery: nanosizing using cell-free technologies and molecular biology approaches. Journal of Biomolecular Screening 10[8], 765-779. 2005.

Lesaicherre, M, Uttamchandani, M, Chen, GYJ, and Yao, SQ. Developing site-specific immobilization strategies of peptides in a microarray. Bioorganic & Medicinal Chemistry Letters 12, 2079-2083. 2003.

Lesaicherre ML, Lue R Y P, Chen G Y J, Zhu Q and Yao S Q. Intein-Mediated Biotinylation of Proteins and Its Application in a Protein Microarray. Journal of the American Chemical Society 124, 8768-8769. 2002.

Lesko, LJ and Woodcock, J. Translation of pharmacogenomics and pharmacogenetics: a regulatory perspective. Nature Reviews Drug Discovery 3, 763-769. 2004.

Lewis, DFV. On the recognition of mammalian microsomal Cytochrome P450 substrates and their characteristics. Biochemical Pharmacology 60, 293-306. 2000.

Li, H and Poulos, TL. Crystallisation of cytochromes P450 and substrate-enzyme interactions. Current Topics in Medicinal Chemistry 4, 1789-1802. 2004.

Lichtenburg, FR. The impact of new drug launches on longevity: evidence from longitudinal, disease-level data from 52 countries, 1982-2001. International Journal of Health Care Finance & Economics 1, 47-73. 2005.

Lin, JH and Lu, AYH. Inhibition and induction of cytochrome P450 and the clinical implications. Clinical Pharmacokinetics 35[5], 361-390. 1998.

Lin, Y, Lu, P, Tang, C, Mei, Q, Sandig, G, Rodrigues, AD, Rushmore, TH, and Shou, M. Substrate inhibition kinetics for cytochrome P450-catalyzed reactions. Drug Metabolism and Disposition 29[4], 368-374. 2001.

Lindberg, RL and Negishi, M. Alteration of mouse cytochrome P450coh substrate specificity by mutation of a single amino acid residue. Nature 339[6226], 632-634. 1989.

Lindner, MD. Clinical attrition due to biased preclinical assessments of potential efficacy. Pharmacology & Therapeutics 115, 148-175. 2007.

Lu HB, Campbell C T and Castner D G. Attachment of Functionalized Poly(Ethylene Glycol) Films to Gold Surfaces. Langmuir 16, 1711-1718. 2000.

Lue R, Chen G, Zhu Q, Lesaicherre M and Yao S (2007) Site-specific immobilization of biotinylated proteins for protein microarray analysis, in *Methods in Molecular Biology* (Fung E ed) pp 85-100, Humana Press Inc., Totowa, NJ.

MacBeath, G. Protein microarrays and proteomics. Nature Genetics 32[Supplement 2], 526-532. 2002.

MacBeath, G and Schreiber, S. Printing proteins as microarrays for high-throughput function determination. Science 289, 1760-1763. 2000.

Makrides, SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiological Reviews 60[3], 512-538. 1996.

Makris, TM, von Koenig, K, Schlichting, I, and Sligar, SG. The status of high-valent metal oxo complexes in the P450 cytochromes. Journal of Inorganic Biochemistry 100, 507-518. 2006.

Marohnic CC, Panda S P, Martásek P and Masters B S. Diminished FAD Binding in the Y459H and V492E Antley-Bixler Syndrome Mutants of Human Cytochrome P450 Reductase. Journal of Biological Chemistry 281, 35975-35982. 2006.

Masimirembwa, CM, Otter, C, Berg, M, Jönsson, M, Leidvik, B, Jonsson, E, Johansson, T, Bäckman, A, Edlund, A, and Andersson, TB. Heterlogous expression and kinetic characterization of human cytochromes P-450: validation of a pharmaceutical tool for drug metabolism research. Drug Metabolism and Disposition 27[10], 1117-1122. 1999.

Mayne, JT, Ku, WW, and Kennedy, SP. Informed toxicity assessment in drug discovery: systems-based toxicology. Current Opinion in Drug Discovery 9, 75-83. 2006.

Mayuzumi, H, Sambongi, C, Hiroya, K, Shimizu, T, Tateishi, T, and Hatano, M. Effect of mutations of ionic amino acids of cytochrome P450 1A2 on catalytic activities toward 7-ethoxycoumarin and methanol. Biochemistry 32, 5622-5628. 1993.

McLeod, HL. Pharmacogenetic analysis of clinically relevant genetic polymorphisms. Clinical Infectious Diseases 41, S449-S452. 2003.

McWilliam, A, Lutter, R, and Nardinelli, C. Health care savings from personalisng medicine using genetic testing: the case of warfarin. American Enterprise Institute-Brookings Joint Center, Working Paper , 6-23. 2006.

Merck & Co., Inc. Merck announces voluntary worldwide withdrawal of VioxxR. Merck & Co., Inc. 2004. http://www.merck.com/newsroom/vioxx/pdf/vioxx_press_release_final.pdf

Metelitza, DI, Eryomin, AN, and Usanov, SA. Quantitative characteristics of the stabilising effect of glycerol on cytochrome P450 in rabbit liver microsomes. Acta Biologica et Medica Germanica 41[1], 17-21. 1982.

Meunier, B, de Visser, SP, and Shaik, S. Mechanism of oxidation reactions catalysed by cytochrome P450 enzymes. Chemical Reviews 104, 3947-3980. 2004.

Miles, CS, Ost, TWB, Noble, MA, Munro, AW, and Chapman, SK. Protein engineering of cytochrome P-450. Biochimica et Biophysica Acta 1543, 383-407. 2000.

Mills, E, Cooper, C, Seely, D, and Kanfer, I. African herbal medicines in the treatment of HIV: *Hypoxis* and *Sutherlandia*. An overview of evidence and pharmacology. Nutrition Journal 4[1], 19. 2005.

Miron T and Wilchek M. A Spectrophotometric Assay for Soluble and Immobilized N-Hydroxysuccinimide Esters. Analytical Biochemistry 126, 433-435. 1982.

Miwa GT, West S B, Huang M T and Lu A Y. Studies on the Association of Cytochrome P450 and NADPH-Cytochrome c Reductase During Catalysis in a Reconstituted Hydroxylating System. Journal of Biological Chemistry 254, 5695-5700. 1979.

Miyazaki M, Nakamura K, Fujita Y, Guengerich F, Horiuchi R and Yamamoto K. Defective Activity of Recombinant Cytochromes P450 3A4.2 and 3A4.16 in Oxidation of Midazolam, Nifedipine, and Testosterone. Drug Metabolism and Disposition 36, 2287. 2008.

Murataliev, MB, Feyereisen, R, and Walker, AH. Electron transfer by diflavin reductases. Biochimica et Biophysica Acta 1698, 1-26. 2004.

Nam, W. High-valent iron(IV)-oxo complexes of heme and non-heme ligands in oxygenation reactions. Accounts of Chemical Research 40[7], 522-531. 2007.

Nath, N, Hurst, R, Hook, B, Meisenheimer, P, Zhao, KQ, Nassif, N, Bulleit, RF, and Storts, DR. Improving protein array performance: focus on washing and storage conditions. Journal of Proteome Research 7[10], 4475-4482. 2008.

Nelson DR, Kamataki T, Waxman D, Guengerich F, Estabrook R, Feyereisen R, Gonzalez F, Coon M, Gunsalus I and Gotoh O. The P450 Superfamily: Update on New Sequences, Gene Mapping, Accession Numbers, Early Trivial Names of Enzymes, and Nomenclature. DNA and Cell Biology 12, 1-51. 1993.

Neve, EPA and Ingelman-Sundberg, M. Intracellular transport and localisation of microsomal cytochrome P450. Analytcal & Bioanalytical Chemistry 392, 1075-1084. 2008.

Ning G, Ouyang H, Wang S, Chen X, Xu B, Yang J, Zhang H, Zhang M and Xia G. 3', 5'-Cyclic Adenosine Monophosphate Response Element Binding Protein Up-Regulated Cytochrome P450 Lanosterol 14 {Alpha}-Demethylase Expression Involved in Follicle-Stimulating Hormone-Induced Mouse Oocyte Maturation. Molecular Endocrinology 22, 1682-1694. 2008.

Nozaki Y, Reynolds J and Tanford C. The Interaction of a Cationic Detergent With Bovine Serum Albumin and Other Proteins. Journal of Biological Chemistry 249, 4452-4459. 1974.

Olson, H, Betton, G, Robinson, D, Thomas, K, Monro, A, Kolaja, G, Lilly, P, Sanders, J, Sipes, G, Bracken, W, Dorato, M, Van Deun, K, Smith, P, Berger, B, and Heller, A. Concordance of the toxicity of pharmaceuticals in humans and animals. Regulatory Toxicology and Pharmacology 32, 56-67. 2000.

Omata, Y, Dai, R, Smith, SV, Robinson, RC, and Friedman, FK. Synthetic peptide mimics of a predicted topographical interaction surface: the cytochrome P450 2B1 recognition domain for NADPH-cytochrome P450 reductase. Journal of Protein Chemistry 19[1], 23-31. 2000.

Omura, T and Sato, R. The carbon monoxide-binding pigment of liver microsomes: I. Evidence for its hemoprotein nature. The Journal of Biological Chemistry 239[7], 2370-2378. 1964a.

Omura, T and Sato, R. The carbon monoxide-binding pigment of liver microsomes: II. Solubilization, purification and properties. The Journal of Biological Chemistry 239[7], 2379-2385. 1964b.

Otyepka, M, Skopalik, J, Anzenbacherová, E, and Anzenbacher, P. What common structural features and variations of mammalian P450s are known to date? Biochimica et Biophysica Acta 1770, 376-389. 2007.

Parikh, A, Gillam, EMJ, and Guengerich, FP. Drug metabolism by *Escherichia coli* expressing human Cytochrome P450. Nature Biotechnology 15, 784-788. 1997.

Park SH, Yin P, Liu Y, Reif J H, LaBean T H and Yan H. Programmable DNA Self-Assemblies for Nanoscale Organization of Ligands and Proteins. Nano Letters 5, 729-734. 2005.

Patel, J, Buddha, B, Dey, S, Pal, D, and Mitra, AK. *In vitro* interaction of the HIV protease inhibitor Ritonavir with herbal constituents: changes in P-gp and CYP3A4 activity. American Journal of Therapeutics 11, 262-277. 2004.

Pavek, P and Dvorak, Z. Xenobiotic-induced transcriptional regulation of xenobiotic metabolising enzymes of the Cytochrome P450 superfamily in human extrahepatic tissues. Current Drug Metabolism 9, 129-143. 2008.

Paweletz CP, Charboneau L, Bichsel V E, Simone N L, Chen T, Gillespie J W, Emmert-Buck M R, Roth M J, Petricoin III E F and Liotta L A. Reverse Phase Protein Microarrays Which Capture Disease Progression Show Activation of Pro-Survival Pathways at the Cancer Invasion Front. gene 20, 1981-1989. 2001.

Pease A, Solas D, Sullivan E, Cronin M, Holmes C and Fodor S. Light-Generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis. Proceedings of the National Academy of Sciences 91, 5022-5026. 1994.

Peck, RW. Driving earlier clinical attrition: if you want to find the needle, burn down the haystack. Considerations for biomarker development. Drug Discovery Today 12, 289-294. 2007.

Pédelacq J, Piltch E, Liong E, Berendzen J, Kim C, Rho B, Park M, Terwilliger T and Waldo G. Engineering Soluble Proteins for Structural Genomics. Nature Biotechnology 20, 927-932. 2002.

Pernecky, SJ and Coon, MJ. N-terminal modifications that alter P450 membrane targeting and function. Methods in Enzymology 272, 25-34. 1996.

Petricoin EF, Ardekani A M, Hitt B A, Levine P J, Fusaro V A, Steinberg S M, Mills G B, Simone C, Fishman D A and Kohn E C. Use of Proteomic Patterns in Serum to Identify Ovarian Cancer. The Lancet 359, 572-577. 2002.

Pharmaceutical Research and Manufacturers of America. Pharmaceutical Industry Profile 2009. PhRMA April. 2009. Washington, DC. http://www.phrma.org/files/attachments/PhRMA%202009%20Profile%20FINAL.pdf

Phillips, KA, Veenstra, DL, Oren, E, Lee, JK, and Sadee, W. Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. JAMA 286, 2270-2279. 2001.

Phizicky, E, Bastiaens, PIH, Zhu, H, Snyder, M, and Fields, S. Protein analysis on a proteomic scale. Nature 422, 208-215. 2003.

Pirmohamed, M and Park, BK. Cytochrome P450 enzyme polymorphisms and adverse drug reactions. Toxicology 192[1], 23-32. 2003.

Poulos, TL. Intermediates in P450 catalysis. Philosophical Transactions of The Royal Society 363, 793-806. 2005.

Poulos, TL, Finzel, BC, and Gunsalus, IC. The 2.6A structure of *Pseudomonas putida* Cytochrome P450. Journal of Biological Chemistry 260[30], 16122-16130. 1985.

Preininger, C, Sauer, U, Dayteg, J, and Pichler, R. Optimising processing parameters for signal enhancement of oligonucleotide and protein arrays on ARChip Epoxy. Bioelectrochemistry 67[2], 155-162. 2005.

Pritchard, MP, Glancey, MJ, Blake, JAR, Gilham, DE, Burchell, B, Wolf, RC*CR, and Friedberg, T. Functional co-expression of CYP2D6 and human NADPH-cytochrome P450 reductase in *Escherichia coli*. Pharmacogenetics 8, 33-42. 1998.

Pylypenko, O and Schlichting, I. Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s. Annual Reviews in Biochemistry 73, 991-1018. 2004.

Roberts E, Shu N, Howard M, Broadhurst R, Chapman-Smith A, Wallace J, Morris T, Cronan Jr J and Perham R. Solution Structures of Apo and Holo Biotinyl Domains From Acetyl Coenzyme A Carboxylase of *Escherichia Coli* Determined by Triple-Resonance Nuclear Magnetic Resonance Spectroscopy. Biochemistry 38, 5045-5053. 1999.

Roots, I, Gerloff, T, Meisel, C, Kirchheiner, J, Goldammer, M, Kaiser, R, Laschinski, G, Brockmöller, J, Cascorbi, I, Kleeberg, U, and Hilderbrandt, AG. Pharmacogenetics-based new therapeutic concepts. Drug Metabolism Reviews 36[3 & 4], 617-638. 2004.

Rupcich N, Goldstein A and Brennan J D. Optimisation of Sol Gel Formulations and Surface Treatments for the Development of Pin-Printed Protein Microarrays. Chemistry of Materials 15, 1803-1811. 2003.

Sadée, W and Dai, Z. Pharmacogenetics/genomics and personalised medicine. Human Molecular Genetics 14[2], R207-R214. 2005.

Sakai-Kato, K, Kato, M, Homma, H, Toyo'oko, T, and Utsunomiya-Tate, N. Creation of a P450 array toward high-throughput analysis. Analytical Chemistry 77[21], 7080-7083. 2005.

Sakanyan, V. High-throughput and multiplexed protein array technology: protein-DNA and protein-protein interactions. Journal of Chromatography B 815, 7-95. 2005.

Samaddar, M, Blackburn, JM, Hart, DJ, and Dyson, MR. Protein tag comprising a biotinylation domain and method for increasing solubility and determining folding state. 10/502,581[US 2005/0221308 A1]. 2005. US, Google Patents.

Sandhu, P, Baba, T, and Guengerich, FP. Expression of modified cytochrome P450 2C10 (2C9) in *Escherichia coli*, purification, and reconstitution of catalytic activity. Archives of Biochemistry and Biophysics 306[2], 443-450. 1993.

Sandhu, P, Guo, Z, Baba, T, and Martin, MV. Expression of modified human cytochrome P450 1A2 in Escherichia coli: Stabilisation, purification, spectral characterisation, and catalytic activities of the enzyme. Archives of Biochemistry and Biophysics 309[1], 168-177. 1994.

Schena M, Shalon D, Davis R W and Brown P O. Quantitative Monitoring of Gene Expression Patterns With a Complementary DNA Microarray. Science 270, 467-470. 1995.

Schott. Nexterion Slide H Protein application. [1.0]. 2006. Schott, Germany. www.schott.com/nexterion

Schroth, W, Antoniadou, L, Fritz, P, Schwab, M, Muerdter, T, Zanger, UM, Simon, W, Eichelbaum, M, and Brauch, H. Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. Journal of Clnical Oncology 25[33], 5187-5193. 2007.

Schweitzer, B, Predki, P, and Snyder, M. Microarrays to characterise protein interactions on a whole-proteome scale. Proteomics 3, 2190-2199. 2003.

Sekiguchi N, Higashida A, Kato M, Nabuchi Y, Mitsui T, Takanashi K, Aso Y and Ishigai M. Prediction of Drug-Drug Interactions Based on Time-Dependent Inhibition From High Throughput Screening of Cytochrome P450 3A4 Inhibition. Drug Metabolism and Pharmacokinetics 24, 500-510. 2009.

Sevrioukova, IF, Hazzard, JT, Tollin, G, and Poulos, TL. The FMN to heme electron transfer in Cytochrome P450BM-3. The Journal of Biological Chemistry 274[51], 36097-36106. 1999a.

Sevrioukova, IF, Li, H, Zhang, H, Peterson, JA, and Poulos, TL. Structure of a cytochrome P450-redox partner electron-transfer complex. Proceedings of the National Academy of Science 96, 1863-1868. 1999b.

Shen, AL and Kasper, CB. Role of acidic residues in the interaction of NADPH-cytochrome P450 oxidoreductase with cytochrome P450 and cytochrome c. Journal of Biological Chemistry 270, 27475-27480. 1995.

Shet, MS, Fisher, CW, Holmans, PL, and Estabrook, RW. Human cytochrome P450 3A4: Enzymatic properties of a purified recombinant fusion protein containing NADPH-P450 reductase. Proceedings of the National Academy of Science 90, 11748-11752. 1993.

Shimada, T, Wunsch, RM, Hanna, IH, Sutter, TR, Guengerich, FP, and Gillam, EMJ. Recombinant human cytochrome P450 1B1 expression in *Escherichia coli*. Archives of Biochemistry and Biophysics 357[1], 111-120. 1998.

Shimizu, T, Tateishi, T, Hatano, M, and Fujii-Kuriyama, Y. Probing the role of lysines and arginines in the catalytic function of cytochrome P450d by site-directed mutagenesis. Interaction with NADPH-cytochrome P450 reductase. Journal of Biological Chemistry 266, 3372-3375. 1991.

Shin DS, Lee K N, Jang K H, Kim J K, Chung W J, Kim Y K and Lee Y S. Protein Patterning by Maskless Photolithography on Hydrophilic Polymer-Grafted Surface. Biosensors and Bioelectronics 19, 485-494. 2003.

Shou, M, Mei, Q, Ettore, MW, Baillie, DR, and Rushmore, TH. Sigmoidal kinetic model for two cooperative substrate-binding sites in a cytochrome P450 3A4 active site: an example of the metabolism of diazepam and its derivatives. Biochemal Journal 340[3], 845-853. 1999.

Sollana, JA, Kirsch, JM, Bala, MV, Chambers, MG, and Harpole, LH. The economics of drug discovery and the ultimate valuation of pharmacotherapies in the marketplace. Clinical Pharmacology & Therapeutics 84[2], 263-266. 2008.

Sreekumar, A, Nyati, MK, Varambally, S, Barrette, TR, Ghosh, D, Lawrence, TS, and Chinnaiyan, AM. Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. Cancer Res. 61, 7585-7593. 2001.

Sueyoshi, T, Park, LJ, Moore, R, Juvonen, RO, and Negishi, M. Molecular engineering of microsomal P450 2a-4 to a stable, water-soluble enzyme. Archives of Biochemistry and Biophysics 322[1], 265-271. 1996.

Sukumaran, SM, Potsaid, B, Lee, M, Clark, DS, and Dordick, JS. Development of a fluorescence-based, ultra high-throughput screening platform for nanoliter-scale Cytochrome P450 microarrays. Journal of Biomolecular Screening 14[6], 668-678. 2009.

Sun XL, Stabler C L, Cazalis C S and Chaikof E L. Carbohydrate and Protein Immobilization Onto Solid Surfaces by Sequential Diels Alder and Azide Alkyne Cycloadditions. Bioconjugate Chemistry 17, 52-57. 2006.

Takahashi, N, Kaji, H, Yanagida, M, Hayano, T, and Isobe, T. Proteomics: Advanced technology for the analysis of cellular function. Journal of Nutrition 133, 2090S-2096S. 2003.

Tanaka, H, Hanasaki, M, Isojima, T, Takeuchi, H, Shiroya, T, and Kawaguchi, H. Enhancement of sensitivity of SPR protein microarray using a novel 3D protein immobilization. Colloids and Surfaces B: Biointerfaces 70[2], 259-265. 2009.

Tao, A, Chen, C, and Zhu, H. Applications of protein microarray technology. Combinatorial Chemistry & High Throughput Screening 10, 706-718. 2007.

Taylor CF, Paton N W, Lilley K S, Binz P A, Julian R K, Jones A R, Zhu W, Apweiler R, Aebersold R and Deutsch E W. The Minimum Information About a Proteomics Experiment (MIAPE). Nature Biotechnology 25, 887-893. 2007.

Taylor, M, Lamb, DC, Cannel, RJ, Dawson, MJ, and Kelly, SL. Cofactor recycling with immobilised heterologous Cytochrome P450 105D1 (CYP105D1). Biochemical and Biophysical Research Communications 279[2], 708-711. 2000.

Templin M, Stoll D, Schrenk M, Traub P, Vöhringer C and Joos T. Protein Microarray Technology. Drug Discovery Today 7, 815-822. 2002.

Templin, MF, Stoll, D, Schwenk, JM, Pötz, O, Kramer, S, and Joos, TO. Protein microarrays: Promising tools for proteomic research. Proteomics 3, 2155-2166. 2003.

Thomas, FJ, McLeod, HL, and Watters, JW. Pharmacogenomics: the influence of genomic variation on drug response. Current Topics in Medicinal Chemistry 4, 1397-1407. 2004.

Thompson, JD, Higgins, DG, and Gibson, TH. CLUSTAL W: improving the senstivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673-4680. 1994.

Tracy, TS. Atypical enzyme kinetics: their effect on *in vitro-in vivo* pharmacokinetic predictions and drug interactions. Current Drug Metabolism 4, 341-346. 2003.

Tracy, TS and Hummel, MA. Modeling kinetic data from *in vitro* drug metabolism enzyme experiments. Drug Metabolism Reviews 36[2], 231-242. 2004.

Trubetskoy, OV, Gibson, JR, and Marks, BD. Highly miniaturised 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. 2005.

Ukelis U, Kramer P, Olejniczak K and Mueller S. Replacement of *in Vivo* Acute Oral Toxicity Studies by *in Vitro* Cytotoxicity Methods: Opportunities, Limits and Regulatory Status. Regulatory Toxicology and Pharmacology 51, 108-118. 2008.

Urlacher, VB, Lutz-Wahl, S, and Schmid, RD. Microbial P450 enzymes in biotechnology. Applied Microbial Biotechnology 64, 317-325. 2004.

Venkatakrishnan, K, von Moltke, LL, Obach, RS, and Greenblatt, DJ. Drug metabolism and drug interactions: application and clinical value of *in vitro* models. Current Drug Metabolism 4, 423-459. 2003.

Vizirianakis, IS. Clinical translation of genotyping and haplotyping data. Clinical Pharmacokinetics 46[10], 807-824. 2007.

Voznesensky, AI and Schenkman, JB. Inhibition of cytochrome P450 reductase by polyols has an electrostatic nature. European Journal of Biochemistry 210, 741-746. 1992b.

Voznesensky, AI and Schenkman, JB. The Cytochrome P450 2B4-NADPH Cytochrome P450 Reductase electron transfer complex is not formed by charge-pairing. The Journal of Biological Chemistry 267[21], 14669-14676. 1992a.

Wacker, R and Niemeyer, CM. DDI-microFIA - a readily configurable microarray-fluorescence immunoassay based on DNA-directed immobilisation of proteins. ChemBioChem 5, 453-459. 2004.

Waldo G. Genetic Screens and Directed Evolution for Protein Solubility. Current Opinion in Chemical Biology 7, 33-38. 2003.

Wang H, Dick R, Yin H, Licad-Coles E, Kroetz D, Szklarz G, Harlow G, Halpert J and Correia M. Structure-Function Relationships of Human Liver Cytochromes P450 3A: Aflatoxin B1 Metabolism As a Probe. Biochemistry 37, 12536-12545. 1998.

Wang, M, Roberts, DL, Paschke, R, Shea, TM, Masters, BSS, and Kim, JP. Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes. Proceedings of the National Academy of Science 94, 8411-8416. 1997.

Wang, RW, Newton, DJ, Liu, N, Atkins, WM, and Lu, AYH. Human cytochrome P-450: *in vitro* drug-drug interaction patterns are substrate-dependent. Drug Metabolism and Disposition 28[3], 360-366. 1999.

Watzke A, K÷hn M, Gutierrez-Rodriguez M, Wacker R, Schr÷der H, Breinbauer R, Kuhlmann J, Alexandrov K, Niemeyer C M and Goody R S. Site-Selective Protein Immobilization by Staudinger Ligation. Angewandte Chemie 45, 1408-1412. 2006.

Werck-Reichhart, D and Feyereisen, R. Cytochromes P450: a success story. Genome Biology 1[6], 3003.1-3003.9. 2000.

Wienkers, LC and Heath, TG. Predicting *in vivo* drug interactions from *in vitro* drug discovery data. Nature Reviews Drug Discovery 4, 825-833. 2005.

Wilkinson, GR. Drug metabolism and variability among patients in drug response. The New England Journal of Medicine 352[21], 2211-2221. 2005.

Williams, PA, Cosme, J, Sridhar, V, Johnson, EF, and McRee, DE. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. Molecular Cell 5[1], 121-131. 2000.

Williams, PA, Cosme, J, Vinkovic, DM, Ward, A, Angove, HC, Day, PJ, Vonrhein, C, Tickle, IJ, and Jhoti, H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. Science 305, 683-686. 2004.

Wilson, DS and Nock, S. Recent developments in protein microarray technology. Angewandte Chemie 42[5], 494-500. 2003.

Wingren, C and Borrebaeck, CAK. Progress in miniaturization of protein arrays - a step closer to high-density nanoarrays. Drug Discovery Today 12[19/20], 813-819. 2007.

Wishart, DS. Improving early drug discovery through ADME modelling: an overview. Drugs in R&D 8[6], 349-362. 2007.

Wong, EY and Diamond, SL. Enzyme microarrays assembled by acoustic dispensing technology. Analytical Biochemistry 381, 101-106. 2008.

Woodcock, J. Making medical products better, faster, and cheaper. Food and Drug Administration . 2007. http://www.fda.gov/consumer/updates/criticalpath091007.html

Wu, JT, Wu, LH, and Knight, JA. Stability of NADPH: effect of various factors on the kinetics of degradation. Clinical Chemistry 32[2], 314-319. 1986.

Yamazaki, H, Nakajima, M, Nakamura, M, Asahi, S, Shimada, N, Gillam, EMJ, Guengerich, FP, Shimada, T, and Yokoi, T. Enhancement of cytochrome P-450 3A4 catalytic activities by cytochrome b_5 in bacterial membranes. Drug Metabolism and Disposition 27[9], 999-1004. 1999.

Yang, M and Cederbaum, AI. Glycerol increases content and activity of human cytochrome P450 2E1 in a transduced HepG2 cell line by protein stabilisation. Alcoholism: Clinical & Experimental Research 21[2], 340-347. 1997.

Yano, JK, Wester, MR, Schoch, GA, Griffin, KJ, Stout, CD, and Johnson, EF. 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. 2004.

Yawetz, A, Perry, AS, Freeman, A, and Katchalski-Katzir, E. Monooxygenase activity of rat liver microsomes immobilised by entrapment in a crosslinked prepolymerised polyacrylamide hydrazide. Biochimica et Biophysica Acta 798[2], 204-209. 1984.

Yim, S, Yun, C, Ahn, T, Jung, H, and Pan, J. 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. 2005.

Yong, WP and Innocenti, F. Translation of pharmacogenetic knowledge into cancer therapeutics. Clinical Advances in Hematology & Oncology 5[9], 698-706. 2007.

Yu, X, Xu, D, and Cheng, Q. Label-free detection methods for protein microarrays. Proteomics 6[20], 5493-5503. 2006.

Yun, C, Song, M, Ahn, T, and Kim, H. Conformational change of cytochrome P450 1A2 induced by sodium chloride. The Journal of Biological Chemistry 271[49], 31312-31316. 1996.

Yun, C, Yim, S, Kim, B, and Ahn, E. Functional expression of human Cytochrome P450 enzymes in *Escherichia coli*. Current Drug Metabolism 7, 411-429. 2006.

Zangar, RC, Davydov, DR, and Verma, S. Mechanisms that regulate production of reactive oxygen species by cytochrome P450. Toxicology and Applied Pharmacology 199, 316-331. 2004.

Zanger, UM, Turpeinen, M, Klein, K, and Schwab, M. Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. Analytcal & Bioanalytical Chemistry 392, 1093-1108. 2008.

Zawaira, A, Gallotta, M, Beeton-Kempen, N, Coulson, L, Marais, P, Kuttel, M, and Blackburn, JM. Exhaustive computational search of ionic-charge clusters that mediate interactions between mammalian cytochrome P450 (CYP) and P450-oxidoreductase (POR) proteins. Computational Biology and Chemistry 34[1], 42-52. 2010.

Zhang K, Diehl M R and Tirrell D A. Artificial Polypeptide Scaffold for Protein Immobilization. Journal of the American Chemical Society 127, 10136-10137. 2005.

Zhao, Q, Modi, S, Smith, G, Paine, M, McDonagh, PD, Wolf, CR, Tew, D, Lian, LY, Roberts, GC, and Driessen, HP. Crystal structure of the FMN-binding domain of the human cytochrome P450 reductase at 1.93 A resolution . Protein Science 8, 298-306. 1999.

Zhu, H, Bilgin, M, Bangham, R, Hall, D, Casamayor, A, Bertone, P, Lan, N, Jansen, R, Bidlingmaier, S, Houfek, T, Mitchell, T, Miller, P, Dean, RA, Gerstein, M, and Snyder, M. Global analysis of protein activities using proteome chips. Science 293, 2101-2105. 2001.

Zhu, H, Klemic, JF, Chang, S, Bertone, P, Casamayor, A, Klemic, KG, Smith, D, Gerstein, M, Reed, MA, and Snyder, M. Analysis of yeast protein kinases using protein chips. Nature Genetics 26, 283-289. 2000.

Zhu, H and Snyder, M. Protein chip technology. Current Opinion in Chemical Biology 7, 55-63. 2003.