FUNCTIONAL ANNOTATION OF ORPHAN HUMAN P450 ENZYMES: HETEROLOGOUS EXPRESSION AND SUBSTRATE SEARCHES BY METABOLOMIC APPROACHES

By

Yi Xiao

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

December 2014

Nashville, Tennessee

Approved:	Date:
F. Peter Guengerich, Ph.D.	10 September 2014
Richard Armstrong, Ph.D.	10/ Sept / 2014
Alan Brash, Ph.D.	Sept 10 2014
David Hachey, Ph.D.	10 - Sep - 2014
Kevin Schey, Ph.D.	10/Sept/2014

FUNCTIONAL ANNOTATION OF ORPHAN HUMAN P450 ENZYMES: HETEROLOGOUS EXPRESSION AND SUBSTRATE SEARCHES BY METABOLOMIC APPROACHES

By

Yi Xiao

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

December 2014

Nashville, Tennessee

Approved by:

Professor F. Peter Guengerich

Professor Richard Armstrong

Professor Alan Brash

Professor David Hachey

Professor Kevin Schey

To my dearest Mom and Dad, and my beloved wife Wenyue, for their unconditional love and support

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my mentor and Ph.D. advisor, Dr. F. Peter Guengerich, for his enormous guidance and support, and for securing the research funding (National Institutes of Health Grants R37 CA090426 and P30 ES000267). He is not only a role model as a great scientist but also committed to mentoring and teaching students. He was always there to encourage and lead me through the twists and turns of scientific journey and also gave me the time and freedom to explore, learn, and grow. He is also very supportive for my overall professional development. Working with him is a great privilege and I am forever indebted to him.

I would also like to thank the members of my thesis committee, Dr. Richard Armstrong, Dr. Alan Brash, Dr. David Hachey, and Dr. Kevin Schey, for their guidance on my thesis projects and advice on my career development. The project would not have proceeded so well without their support. Dr. David Hachey helped tremendously on state-of-the-art MS instrumentations and data interpretation. Dr. Alan Brash gave invaluable help on characterizing fatty acid analogs by liquid chromatography. Dr. Richard Armstrong and Dr. Kevin Schey always offered valuable input and thoughtful suggestions. I was so fortunate to have the privilege of meeting with the amazing panel of experts every six months.

The laboratory is a very supportive community. Kathy Trisler, Ellen Rochelle, Dr. Martha Martin, Lindsay Folkmann, and Leslie Nagy made

everything run smoothly. Dr. Wade Calcutt, Dr. Raku Shinkyo, Dr. Goutam Chowdhury, and Dr. Qian Cheng are great sources of advice, training, and friendship. I also owe thanks to all the brilliant and kind people that I have worked with on my thesis projects for their generous assistance on instrumentations and data analysis. I am thankful for my friends I met in the Interdisciplinary Graduate Program and the community of international students/scholars. They provided me with support, empathy, and fun, which greatly expanded my life in Nashville and prevented me from tunnel vision. I am eternally thankful to them and wish them the best in their future endeavors. I am also thankful for Vanderbilt International Scholar Program for providing full stipend support for two years and the director, Dr. Kathy Gould, for her support and guidance.

Last but certainly not least, I would like to thank my Mom and Dad for their enduring love and support and Wenyue, my beloved wife, for her companionship through this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APCI	atmospheric pressure chemical ionization
AQ4	1,4-*bis*{[2-(dimethylamino)ethyl]amino}-5,8-dihydroxy-anthracene-9,10-dione
AQ4M	1-{[2-(dimethylamino-*N*-oxide)ethyl]amino}-4-{[2-(dimethylamino)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione
AQ4N	1,4-*bis*{[2-(dimethylamino-*N*-oxide)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione
BSTFA	*N,O*-*bis*-(trimethylsilyl)-trifluoroacetamide
DLPC	L-α-dilauoryl-*sn*-glycero-3-phosphocholine
ESI	electrospray ionization
GC-MS	gas chromatography–mass spectrometry
HPLC	high-performance liquid chromatography
IPTG	isopropyl-β-D-thiogalactoside
IRES	internal ribosome entry site
HRMS	high resolution mass spectrometry
LC-MS	liquid chromatography–mass spectrometry
LPA	lysophosphatidic acid
LPC	lysophosphatidylcholine
LPE	lysophosphatidylethanolamine
LPG	lysophosphatidylglycerol

LPI	lysophosphatidylinositol
LPS	lysophosphatidylserine
MOI	multiplicities of infection
MS	mass spectrometry
NMR	nuclear magnetic resonance
P450	cytochrome P450
PC	phosphatidylcholine
pFBB	pentafluorobenzyl bromide
PPARα	peroxisome proliferator-activated receptor alpha
TMCS	trimethylchlorosilane
TMS	trimethylsilyl
TMSI	trimethylsilylimidazole
UPLC	ultra-performance liquid chromatography
RhPV	*Rhopalosiphum padi* virus

CHAPTER I

INTRODUCTION

Cytochrome P450 enzymes

Cytochrome P450 (P450) enzymes are versatile biocatalysts with broad substrate specificities. The heme protoporphyrin IX prosthetic group, linked with the axial cysteine residue, gives P450 enzymes their unique chemistry and spectroscopic properties. Research on P450 enzymes can be traced back as early as in 1964, when Omura and Sato named the enzyme "cytochrome P450" for "pigment 450" due to the maximal absorbance at 450 nm following reduction and CO binding (*1, 2*). Absorbance at 450 nm of the Fe^{2+}·CO complex has been used to determine the concentration of the active enzyme (*3, 4*).

It is now well accepted that P450 enzymes are a superfamily of proteins with similar molecular weight (~ 50 kiloDaltons). To date, more than 20,000 P450 genes have been identified across all kingdoms of life with sequence information stored in the P450 database (http://drnelson.uthsc.edu/CytochromeP450.html). With the launching of genome sequencing projects towards numerous other organisms, sequence information for more and more P450 genes will become available.

In order to give a unique name for each P450 enzyme, a nomenclature system was adopted: P450 enzymes with > 40% amino acid sequence identity

are categorized in the same family and are designated with a number (P450 1, P450 2, *etc.*); subfamilies with > 55% amino acid sequence identity are designated with a letter following the number (P450 1A, P450 1B, *etc.*); each enzyme in the subfamily is named with another number (P450 1A1, P450 1A2, *etc.*).

P450 enzymes are versatile biocatalysts. With numerous members found in nature, P450 enzymes are well known for their involvement in diverse and important physiological functions. Substrates for P450 enzymes include a plethora of compounds involved in both primary and secondary metabolism (e.g., hormones, cholesterol, fatty acids, terpenoids, antibiotics, pesticides, herbicides, environmental toxins). In addition, P450 enzymes can catalyze a variety of types of reactions.

The most commonly seen reactions are monooxygenation reactions, with stoichiometry shown below (RH is a substrate and ROH is the monooxygenated product):

$$RH + O_2 + NADPH + H^+ \rightarrow ROH + H_2O + NADP^+$$

Some P450 enzymes also catalyze reduction (*5*), carbon-carbon cleavage (*6*), dealkylation (*7*), ring expansion (*8*), desaturation (*9*), and other reactions based on the same high-valent iron chemistry (Figure 1) (*10*).

To appreciate their physiological significance and utilize their catalytic activities for our benefit, studies on P450 enzymes have been continuing in many industries for decades, including drug discovery, insect control, and crop science.

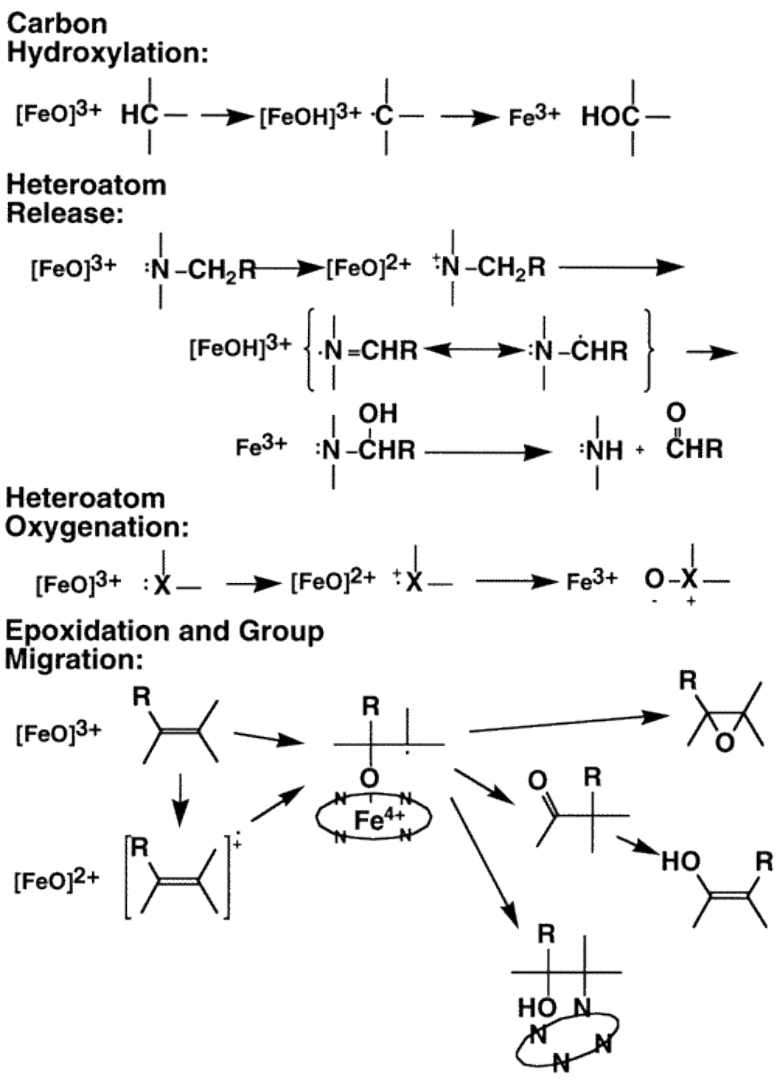


Figure 1. Common reactions based on the high-valent iron chemistry catalyzed by P450 enzymes.
From (*11*).

Human cytochrome P450 enzymes

Human P450 enzymes are membrane-bound proteins either targeted to endoplasmic reticulum or to mitochondria. Redox partners and lipids are needed to mimic membrane environment so that catalytic activities of P450 enzymes can be reproduced *in vitro* for functional assays (*12*): microsomal P450 enzymes can be reconstituted with NADPH-P450 reductase (*12*-*14*); mitochondrial P450 enzymes can be reconstituted with adrenodoxin reductase and adrenodoxin (*15*).

It is well established that human P450 enzymes play crucial roles in many important biological processes. Initial interest in the field was fostered in endocrinology (*16*), chemical carcinogenesis (*17*), and drug metabolism (*18*). Substrates for human P450 enzymes include a variety of compounds, including cholesterol, vitamins, steroids, fatty acids, eicosanoids, environmental toxicants, and drugs (*19, 20*).

Studies on human P450 enzymes have been continuing for decades and knowledge on their catalytic properties has been expanding. To date, it is well accepted that human P450 enzymes are involved in the metabolism of ~ 75% of drugs currently on the market (*21*-*23*) (Figure 2). The majority (99%) of human P450 enzyme-mediated drug metabolism is carried out by only a few enzymes (i.e., human P450 3A4, 3A5, 2C9, 2C19, 2D6, 1A1, 1A2, 2B6, and 2E1) (*23*). Although current knowledge suggests that only < 1% of drug metabolism is contributed by other human P450 enzymes, it is increasingly appreciated that those less-studied P450 enzymes can play important roles as well. For example,

human P450 2J2 metabolizes pomalidomide, terfenadine, and ebastine (*24, 25*), and recombinant human P450 2J2 has been made commercially available for pharmaceutical companies to study P450 2J2-mediated drug metabolism. Another good example is the activation of several duocarmycin analogues by human P450 2W1 (*26*). Human P450 2W1 and 2S1 are also involved in the metabolism of several fluorinated 2-aryl-benzothiazole antitumor molecules (*27*). These less-studied human P450 enzymes have also been found to be involved in vitamin metabolism (*28*), tumor prognosis (*29-31*), and a recessive degenerative eye disease (*32*). Thus, the need to investigate all human P450 enzymes is warranted.

Human Genome Project identified 57 human P450 enzymes, which have been classified by their substrate specificities (Table 1). It has to be noted that one P450 enzyme can have substrates that fit into two different categories. For example, human P450 27A1 oxidizes both cholesterol (an endogenous steroid) (*33*) and vitamin D_3 (*34*); human P450 2W1, 2S1, and 2J2 are all involved in drug (xenobiotics) metabolism (*24, 26, 27*); human P450 1A2, 2C8, and 2C9 also oxidize fatty acids (*35*).

Among the 57 human P450 enzymes, 13 of them are considered ‘orphans’ because knowledge towards their physiological significances is limited. Much effort has been devoted to functionally annotate these orphan human P450 enzymes in recent years. For example, heterologous expression systems of a few orphan human P450 enzymes have been established (*36-42*); expression profiles of most orphan human P450 enzymes have been characterized (*41, 43-*

51); a few catalytic activities have been reported (*5, 20, 28, 36, 39-42*). The investigation of catalytic activities of orphan human P450 2S1, 2W1, and 4X1 is the focus of Chapter II and III.

Functional studies of human P450 enzymes have been greatly benefited from the development of heterologous protein expression systems (i.e., *Escherichia coli*, baculovirus, yeast, and mammalian cell expression systems). Historically human P450 enzymes were isolated from human tissue samples (*52*). One major disadvantage of this method is the basal expression of other P450 enzymes in the same tissue and the technical difficulties in enzyme purification without cross-contamination from these enzymes. Other problems include the difficulties in obtaining human tissue samples and in purifying enzymes without losing much catalytic activity. Thanks to the development of heterologous protein expression systems, now it is a common practice to express recombinant P450 enzymes in many laboratories.

The first P450 enzyme ($P450_{cam}$) was crystallized in 1985 and the structure was solved two years afterwards (*53, 54*). The crystal structure of the first mammalian microsomal P450 enzyme was solved in 2000 (*55*) (Figure 3). Since then, the number of Protein Data Bank deposits of human P450 enzymes has been increasing, but none of them was generated with orphan human P450 enzymes, partially because of the difficulties in expressing P450 enzymes in good yield. Although tools have being developed for decades to facilitate heterologous protein expression, it can still be the bottleneck for research projects, especially if large amounts of protein are needed. Heterologous

expression of several human P450 enzymes for functional studies is the focus of Chapter IV.

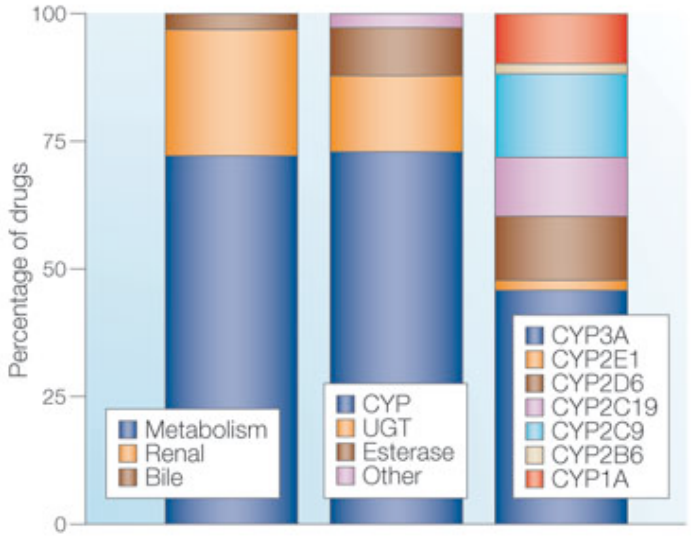


Figure 2. Routes of elimination of the top 200 most prescribed drugs in 2002.
Human P450-mediated metabolism represents the listed clearance mechanism for ~ 73% of the top 200 drugs. For the P450 (CYP)-mediated clearance mechanisms, the majority (46%) are carried out by members of the CYP3A family; followed by 28% by members of the CYP2C family; 12% for CYP2D6; 9% for members of the CYP1A family; and 2% for both CYP2B6 and CYP2E1. From (*22*).

Sterols	Xenobiotics	Fatty Acids	Eicosanoids	Vitamins	Unknown
1B1	1A1	2J2	4F2	2R1	2A7
7A1	1A2	4A11	4F3	24A1	2S1
7B1	2A6	4B1	4F8	26A1	2U1
8B1	2A13	4F12	5A1	26B1	2W1
11A1	2B6		8A1	26C1	3A43
11B1	2C8			27B1	4A22
11B2	2C9				4F11
17A1	2C18				4F22
19A1	2C19				4V2
21A2	2D6				4X1
27A1	2E1				4Z1
39A1	2F1				20A1
46A1	3A4				27C1
51A1	3A5				
	3A7				

Table 1. Classification of human P450 enzymes based on their substrate specificities.
From (*20*).

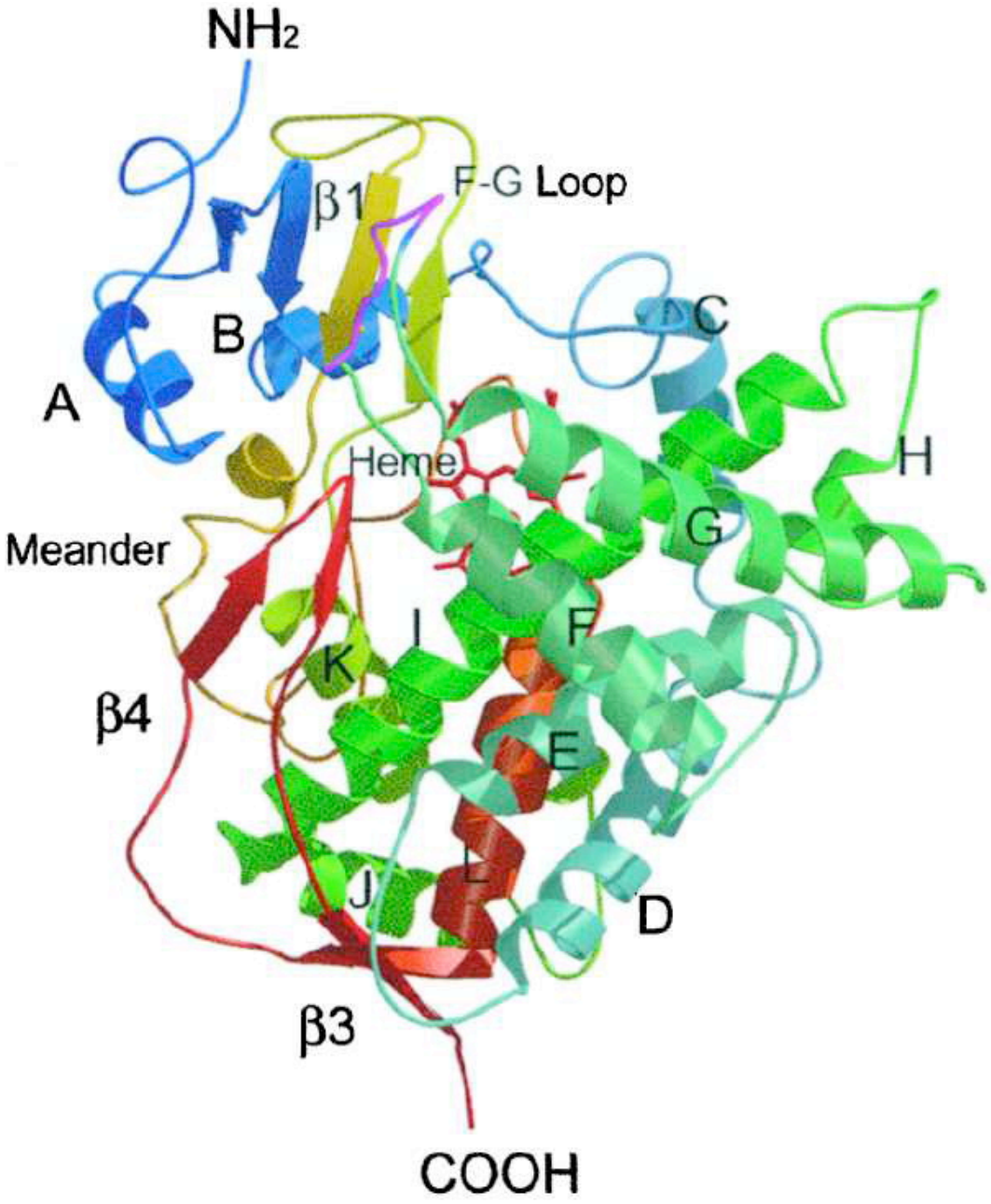


Figure 3. The topology of crystallized P450 2C5/3LVdH.
From (*55*).

Increasing need for functional annotation of gene products

The classical biochemistry paradigm of the 20th century involves several events that happen in a chronological order: 1) identification of an *in vivo* phenomena of physiological significance; 2) development of *in vitro* assays to purify the key proteins involved; 3) information about the protein (or an antibody) can be used to clone a cDNA, and its nucleotide sequence can be determined; 4) the gene product is finally annotated with physiological significance (Figure 4) (*56*).

Due to the advent of genome sequencing technologies, today we have sequence information of all the genes in numerous organisms including humans. Open reading frames and splicing variants can be predicted and confirmed easily by experiments. However, methods for rapidly annotating gene products with physiological significances are still being developed, and functional annotation of gene products is considered one of the grand challenges in modern biomedical research. Even in an organism as well studied as *E. coli*, functions of only 54% of the gene products have been experimentally determined (*57*). Not surprisingly, the prevalence of gene products with unknown function is well beyond 50% in higher plants and animals (*58*). Therefore, strategies are urgently needed for functional annotation of gene products in a backward order: from individual coding sequences to their physiological significances *in vivo* (Figure 4) without any preconceived biological phenomena.

Strategies for functional annotation of gene products often include the followings: sequence comparison with other genes (in DNA levels and in protein levels); characterization of expression profiles (information at different developmental stages may help); the establishment of heterologous protein expression and purification methods; reproducing physiological significances with functional assays *in vitro* and *in vivo*. Although the establishment of heterologous expression methods can still be very problematic (see Chapter IV), the more daunting challenge is choosing the 'right' functional assays to investigate possible protein functions (see Chapter III). Among the numerous functional assays that can be chosen from, some are as easy as an *in vitro* incubation that takes only a few hours, some are as daunting and time consuming as developing genetically modified animal models followed by further experiments.

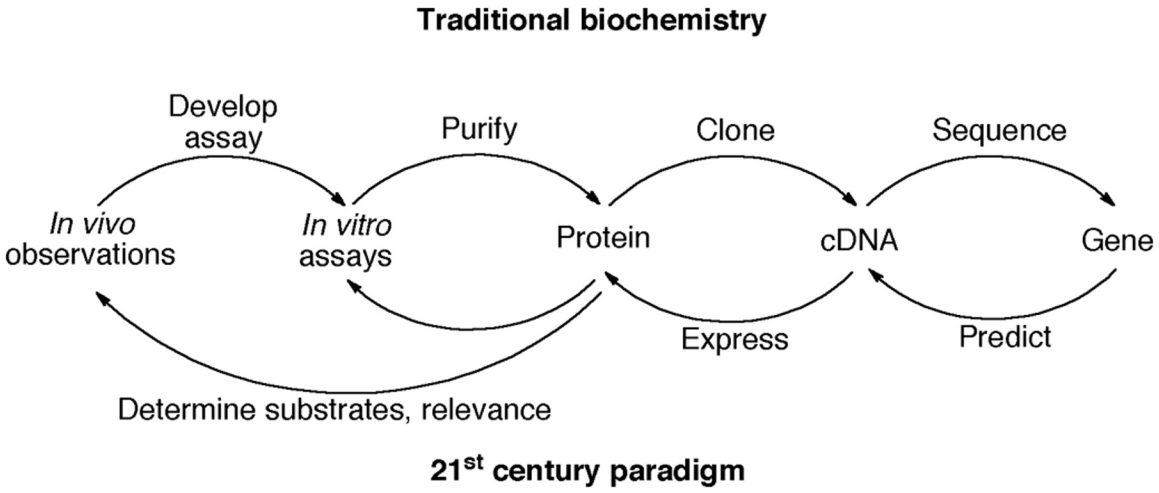


Figure 4. Classical and 21st century approaches in functional annotation of gene products.
From (*56*).

Functional annotation of gene products and applications in studying human P450 enzymes

Targeted strategies

Sequence similarities

In some cases, insight into protein function can be hypothesized from sequence similarities. Given the fact that human P450 3A4 and 3A5 share 85% amino acid identity, it is not surprising that their substrate specificities overlap (*59*). Another example is human P450 7A1, whose catalytic activity was deduced from its animal orthologue (rat P450 7A1) (*60*).

The advantage of this approach is its potential to annotate gene function in the most quick and efficient manner. However, the major problem is that proteins with high sequence similarities can still have completely different functions. For example, although human P450 2C8, 2C9, 2C18, and 2C19 share sequence similarities between 77% and 93%, their substrate specificities differ tremendously (*61*). Enzymes in the enolase and vicinal oxygen chelate (VOC) families (*62*) also catalyze a variety of enzymatic reactions.

Structure insight and *in silico* docking

Another appealing approach is protein crystallization and computational modeling. However, protein-substrate interaction is a dynamic process and structure obtained by crystallization can only provide information of a snapshot. Since it is well accepted that P450 enzymes usually undergo major

conformational changes upon ligand binding (*63*), predicting substrate specificities from protein crystallization and *in silico* docking can be very difficult. In addition, P450 enzymes can have large active sites, which requires tremendous computational power to predict potential substrates. For some P450 enzymes, it can also be very challenging to obtain large amounts of purified protein (*40*). Thus, successful example of functional annotation of P450 enzymes by this approach is limited.

Trial and error by functional assays

Potential function of a protein can be hypothesized and tested with functional assays. Commonly used rationales include expression profiles, regulatory mechanisms, and analysis of phenotypes observed from animals/patients without the functional protein.

This strategy can also annotate protein function in a very quick and efficient manner. Successful examples include the identification of fatty acids as substrates for many orphan P450 enzymes, although with low catalytic efficiencies (*36, 39, 40, 42, 64*). However, there could be endless hypothesized functions waiting to be tested, and it could be even more laborious if an individual method has to be developed for each hypothesis. In terms of fatty acid oxidations by orphan human P450 enzymes, although there are scenarios where fatty acid oxygenation products do have biological activities (*65*), it is not clear that any of these fatty acid oxidations catalyzed by orphan human P450 enzymes are physiologically important. It is hard to believe that fatty acid oxidation is the only

physiological significance for these orphan human P450 enzymes (*39, 40, 42*), and it is very likely that there are other substrates waiting to be identified.

Screen with compound libraries

Compound libraries can be used to enable high-throughput capability.

A library of procarcinogens can be incubated with P450 enzymes to screen for P450-mediated procarcinogen activation and DNA damage (*Umu* assay) (Figure 5) (*36, 64, 66, 67*). Several substrates of human P450 2W1 have been identified with this approach, with low turnover rates (*36*).

Another method is to test a few representatives from each chemical class of potential substrates (e.g., fatty acids, steroid hormones, cholesterol, vitamins). Substrates of each class can be mixed together and used as a substrate pool for enzymatic incubations. Reaction products can be analyzed by a liquid chromatography-mass spectrometry (LC-MS) method that detects every compound in the substrate pool. If any catalytic activity can be detected towards one of these compounds (e.g., cholesterol), then the pool (of cholesterol) can be expanded with more cholesterol analogs to find the ones with better catalytic efficiencies. However this approach has not been successful to date.

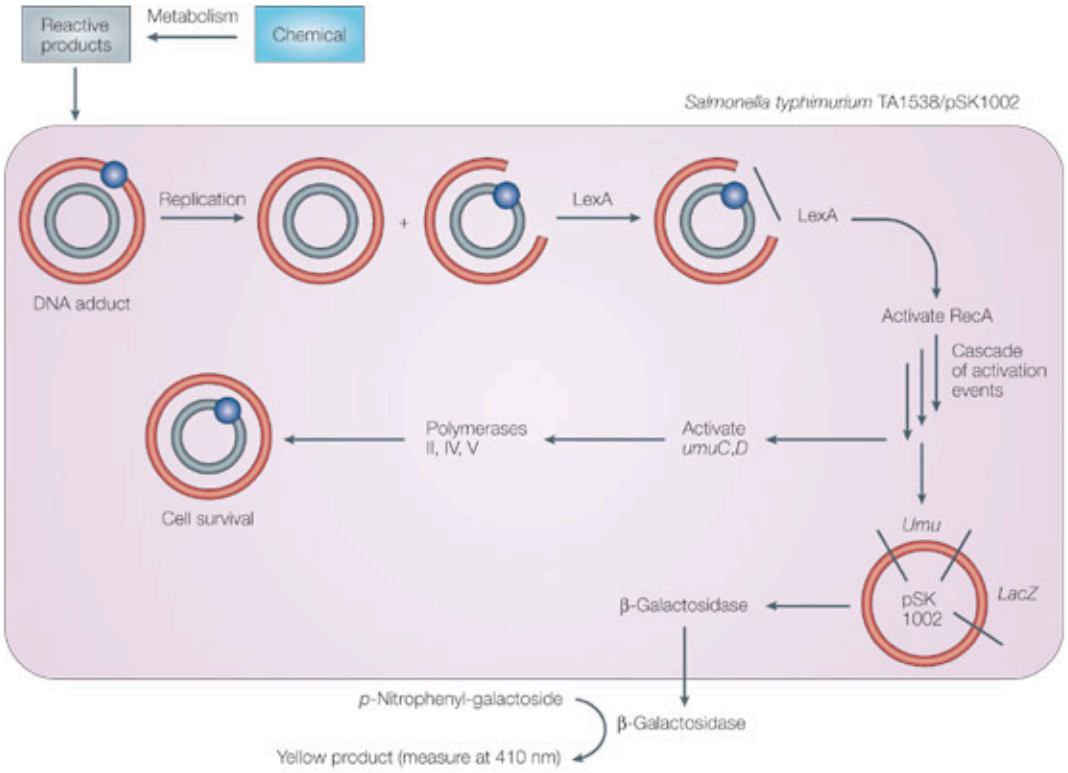


Figure 5. Identification of P450-mediated procarcinogen activation and DNA damage with *umu* assay.
Chemicals are incubated with Salmonella typhimurium cells (containing the plasmid pSK1002) and a metabolic system. The metabolic system can be either liver microsomes or a recombinant cytochrome P450 system expressed within the cells. The formation of DNA adducts would block the replication of the modified strand and generate regions of single-stranded DNA, which would activate the binding of protease LexA followed by the activation of more than 30 other genes. *umu*C, one of the activated gene, encodes a translesion polymerase that the bacteria use to bypass DNA adducts and replicate. A reporter plasmid (pSK1002) contains the regulatory region of the umu gene followed by the coding sequence of *Lac*Z. Thus, the generation of DNA adducts in Salmonella typhimurium cells would activate the expression of β-galactosidase and the production of β-galactosidase, which can be quantified using a colorimetric assay. From (*67*).

Untargeted strategies

Untargeted approaches have the potential to identify substrate(s) in a less biased way. The general hypothesis is that substrate(s) of a P450 enzyme can be identified in the tissue/organ where the enzyme is expressed. This is a common hypothesis seen in many metabolomic studies (*68-71*). Caution must be taken in that this hypothesis may not work if the enzyme of interest works only on xenobiotics.

Binding assays

One approach is to look for ligands that bind to the protein of interest. Ligand-bound protein can be immune-precipitated or immobilized so that the ligand can be enriched and characterized afterwards. This approach is commonly used in laboratories that investigate signal transduction pathways and protein-protein interactions (*72*). One successful example in identifying small molecule ligands is the discovery of linoleic acid as an endogenous ligand of the hepatocyte nuclear factor (HNF) 4α (*73*). Another good example is the identification of interaction between unsaturated fatty acids and the nuclear receptor Nur77 (*74*).

The major disadvantage of this approach is that ligands could disassociate from the targeted protein during protein purification. Thus, only ligands with very high affinities can be captured. Another disadvantage of this approach is the fact that the binding of a ligand to a protein does not necessarily warrant any

physiological significance. Thus, this approach has not been used widely for functional annotation of P450 enzymes.

Metabolomic approaches for functional annotation of orphan enzymes

The metabolome represents a chemically diverse mixture of physiologically relevant compounds, including those that have not been reported in literature or not commercially available. Recent technological advances in global metabolomic profiling enabled simultaneous screening of thousands of metabolites from a variety of chemical classes, making them useful tools for functional annotation of orphan enzymes (*68-71, 75*).

Metabolomic studies can be conducted both *in vivo* and *in vitro*, and require state-of-the-art analytical instruments to analyze complicated biological matrices (if working with genetically modified model organisms or human samples) or incubation mixture (if working with cell lines, subcellular fractions, and reconstituted enzymes) in order to pinpoint any interesting features that could shed light on enzymatic activities. Commonly used instruments include LC-MS, gas chromatography–mass spectrometry (GC-MS), and nuclear magnetic resonance (NMR). Further characterization of these identified features also require the use of sophisticated methodologies, such as high resolution mass spectrometry (HRMS), preparative HPLC, and NMR. LC-MS based metabolomic approaches that can be used for functional annotation of orphan enzymes are summarized in Figure 6.

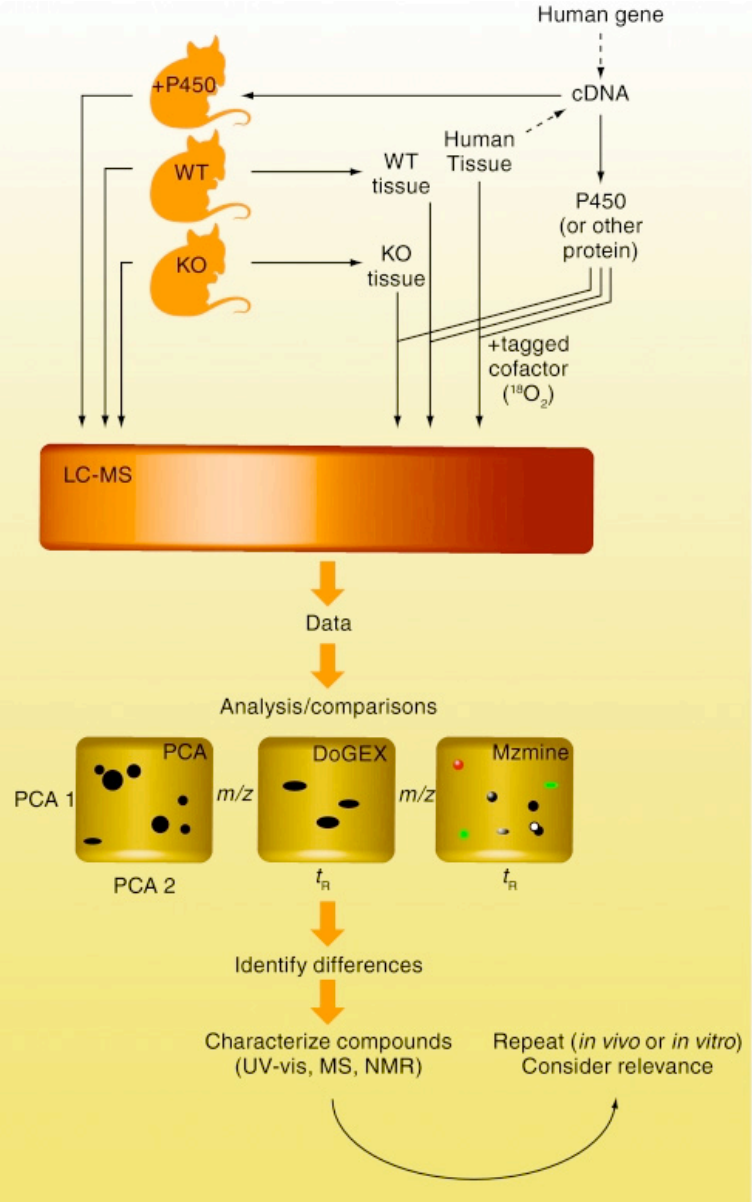


Figure 6. Untargeted LC-MS based strategies for functional annotation of orphan enzymes.
One strategy utilizes genetically modified model organisms, with the orthologue of the protein knocked-out or over-expressed. The metabolome from body fluids (or tissues) can be analyzed by LC-MS. As an alternative, tissue extracts can be incubated with and without the protein or the co-factor. A $^{18}O_2$ labeling strategy can also be used (Figure 7) to identify monooxygenation reactions. Specialized software systems can be used to identify doublets or features that are consumed (putative substrates) or generated (putative products) in a time- and enzyme-dependent manner. From (*56*).

Study design

Based on the hypothesis that the substrate(s) of the enzyme can be identified in the tissue/organ where the enzyme is expressed, the metabolome from the tissue/organ or from cultured cells with a modified enzyme expression level can be compared with that from control samples with an unchanged enzyme expression level. Alternatively, tissue extracts can be incubated with and without enzyme or co-factor and be scrutinized for any differences. Many specialized software systems can be used to identify difference(s) between sets of data (*76*-*81*). This approach has the potential to identify different classes of reactions (e.g., reduction, dealkylation, ring expansion).

Although tissue samples from humans are not always easy to obtain, genetically modified model organisms can be used as an alternative. Thanks to the development of modern molecular biology techniques, it is now a common practice to knock out or over-express a gene in many model organisms (e.g., mouse, zebrafish, fruit fly). Any phenotype resulting from the modified gene expression level could shed light on the physiological function of the gene. The phenotype could be as dramatic as embryonic lethality (which may require tissue-specific genetic modification for further investigation), or as subtle as tiny changes in the abundance of only a few chemicals. Characterization of such subtle changes often require the help of state-of-the-art analytical instruments and require careful study design to exclude false positives and reduce variances. Successful examples include the studies of fatty acid amide hydrolase (FAAH) (*71*), organic anion transporter (Oat1) (*70*), and mammalian

glycerophosphodiesterase (GDE) (*69*): the metabolome from knockout mice was compared with that from wild-type mice. Chemicals that accumulate in the knockout mice were later confirmed as substrates.

Besides the technical requirement in generating genetically modified model organisms and the time spent in producing them, one major problem with studying genetically modified model organism is that the orthologue may not always have the same function as the gene in humans, or there could be no orthologue in the model organism at all. Another concern is the possibility that other proteins may show compensatory activities in the model organism (there are 103 mouse P450 genes and 89 rat P450 genes, in contrast with 57 P450 genes in humans). Thus any phenotype due to the functional loss/gain of the gene could be masked.

Cultured mammalian cell lines can also be modified to increase or decrease the expression level of a gene of interest, transiently or permanently. Controls would be cells transfected with a blank expression vector. For example, the integral membrane enzyme α/β-hydrolase domain–containing 3 (ABHD3) was transiently over-expressed in HEK293T cells and the accumulated phospholipids were confirmed to be the substrates (*68*). One of the advantages of studying cultured cells is the lower cost compared with generating genetically modified model organisms. It is also much easier and takes less time to modify the expression level in cultured cells. A ‘rescue’ experiment can be used to provide solid evidence about physiological relevance, if the phenotype of the permanent knockout cells can be rescued by transient over-expression of the

gene. In addition, cell lines from many different human tissues are available for experiment. Therefore, issues with orthologues (i.e., no corresponding orthologue in model organisms, orthologues in model organisms have different functions) are no longer a concern. If the protein of interest is expressed in human liver, cell lines originally from human liver (i.e., HepG2) should be used due to their physiological relevance. Disadvantages of using cultured cells include low protein expression levels and compensatory effect. Also, the extent of substrate accumulation or product consumption in cultured cells may not be as much as it would be in genetically modified model organisms (depends on the level of protein expression and abundance of substrate in the system). Another disadvantage is that most phenotypes cannot be observed by naked eye and have to be characterized by complicated assays.

Modified cell lines (protein of interest over-expressed or knocked out) can also be incubated with tissue extracts. The metabolome from modified cell lines can be compared with that from ordinary cell lines.

Isolated recombinant enzymes (isolated membrane/microsomal fractions; purified enzymes) can also be incubated with tissue extracts for untargeted metabolomic studies (*14, 64*). Controls are incubations without the enzyme or without co-factor. This approach provides reduced level of complexity in background and less burden to analytical methodologies, so that the chance of identifying a subtle phenotypic change may increase. Compensatory effect is less of a concern, depends on the purity of the isolated protein. One disadvantage of using purified enzymes is the need to establish heterologous

protein expression systems. If a high protein expression level of a microsomal P450 enzyme cannot be achieved for protein purification, co-expression of the enzyme and NADPH-P450 reductase with a bi-cistronic expression vector could be used for functional assays (*82*): coding sequences of both the enzyme and its redox partner are connected with an internal ribosome entry site (IRES) so that both sequences will be transcribed on a single mRNA and translated individually; since both P450 enzyme and NADPH-P450 reductase are membrane-targeted proteins, isolated membrane fractions can be used to reproduce catalytic activities. If a high protein expression level of a mitochondrial P450 enzyme cannot be achieved for protein purification, co-expression of the enzyme, adrenodoxin, and adrenodoxin reductase can also be achieved with a tri-cistronic expression vector, in which the three coding sequences are connected with two IRES sequences (*83*). Since adrenodoxin is a soluble protein and is not membrane-targeted, functional studies have to be conducted with whole cells rather than with isolated membrane fractions (if expression is achieved with an *E. coli* expression system) (*83*). For tri-cistronic expressions in cells with mitochondria subcellular structure (i.e., mammalian cells, insect cells), isolated mitochondria supplemented with extra adrenodoxin can be used for functional assays (*84*).

An approach was recently developed in Guengerich lab to identify monooxygenation reactions after incubations (*85*). This approach is based on the fact that most P450-mediated reactions involve the incorporation of an oxygen atom into the substrate, such that the product is 16 amu heavier. Thus, for

incubations performed with equal amount of $^{16}O_2$ and $^{18}O_2$ gas, the monooxygenated products would show up with M/M+2 doublet patterns in the mass spectra, which can be identified by a specialized computer software system (*14, 64, 85*). Technically, the mixture of the reconstituted P450 system and tissue extracts (corresponding to the localization of the P450 enzyme) are added into the bottom of Thunberg tube while NADPH is added into the neck (Figure 7). Air is removed by several cycles of vacuum and purging with argon. Then the samples (under vacuum) can be charged with either $^{16}O_2$ gas or $^{18}O_2$ gas, followed by tipping the tube to add NADPH into the incubation mixture to start the reaction (figure 7). Both incubation mixtures (one with $^{16}O_2$ gas and another with $^{18}O_2$ gas) are combined before extraction, so that doublet patterns are expected for monooxygenated products. Controls are incubations with only $^{16}O_2$ gas or without NADPH. This methodology identified a group of fatty acids as substrates of human P450 1A2, 2C8, 2C9, and 4F11 (*14, 64*). In principle, this approach can be used with sulfotransferases, epoxide hydrolases, and any other enzymes that catalyze reactions incorporating a co-factor that can be modified with a stable isotope.

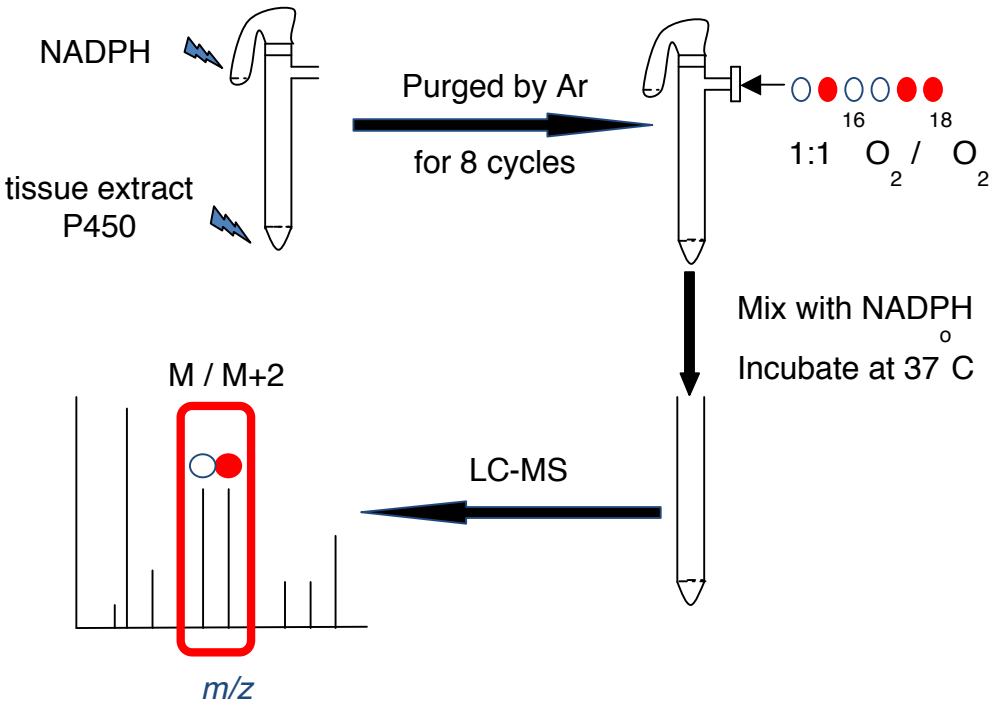


Figure 7. A strategy to identify monooxygenation reactions using $^{18}O_2$ gas.

Analytical methodologies for metabolomic studies

The success of metabolomic studies very much depends on the sophisticated analytical instruments that are used to identify small changes in complex mixtures. One important contributing factor is sensitivity, as high sensitivity ensures the detection of chemicals at very low concentrations. NMR (*86*), LC-MS (*14, 68-70, 85*), and GC-MS (*87*) have all been used in metabolomic studies (Figure 8). Direct infusion coupled with mass spectrometry (MS) analysis is not commonly used due to the lack of chromatography separation.

Ultra performance liquid chromatography (UPLC) coupled with MS analysis provides many benefits over NMR and GC-MS: accurate mass of the parent and fragment ions can be used to determine elemental composition; ultra-resolving power of UPLC provides the capability to analyze complex tissue extracts; if needed, LC methods can be scaled up for metabolites isolation followed by NMR analysis. Low-molecular-weight compounds, particularly those relatively nonpolar ones that might be the substrates of P450 enzymes, can usually be resolved well with a reversed-phase octadecylsilane (C_{18})-based column and a water-to-acetonitrile gradient. Other columns are also available (e.g., HILIC columns can be used to resolve polar compounds (*88*)). Chromatography in metabolomic studies needs to be very reproducible. Achieving symmetrical peak shapes and baseline separations for all compounds in the metabolome are not feasible to accomplish, and are not crucial in most cases. In order to cover as many compounds as possible, multiple runs with

positive- and negative-ion electrospray ionization (ESI) and atmospheric-pressure chemical ionization (APCI) are desired. Molecules bearing a positive charge (e.g., amines) ionize best with positive-ion ESI (with HCO_2H in mobile phase to help ionization); molecules bearing a negative charge (e.g., carboxylic acids) ionize best with negative-ion ESI (with $NH_4CH_3CO_2$ in mobile phase to help ionization); neutral compounds (e.g., steroids) may ionize with APCI. Although there are other ionization modes available (e.g., APPI, MALDI), they are not commonly used in metabolomic studies.

While compounds that ionize well in MS will be easily picked up during data collection and analysis, those that do not ionize well would be invisible. In order to cover as many compound classes as possible, a derivatization method with dansyl chloride has been developed to improve the sensitivity towards unactivated alcohols (*89*), which are common products of many P450 reactions. Dansylated compounds have excellent ionization properties in the positive-ion ESI mode due to the positive charge on the amine moiety (*89*). This method is particularly effective for uncharged molecules such as sterols and can improve sensitivities to up to several thousand fold (*89*). Dansylated compounds can be confirmed by MS-based fragmentation, which removes the dansyl group and produces characteristic daughter ions (*89*). Due to the increased hydrophobicity of dansylated compounds, tert-butyl methyl ether can be added to mobile phase to adjust retention time.

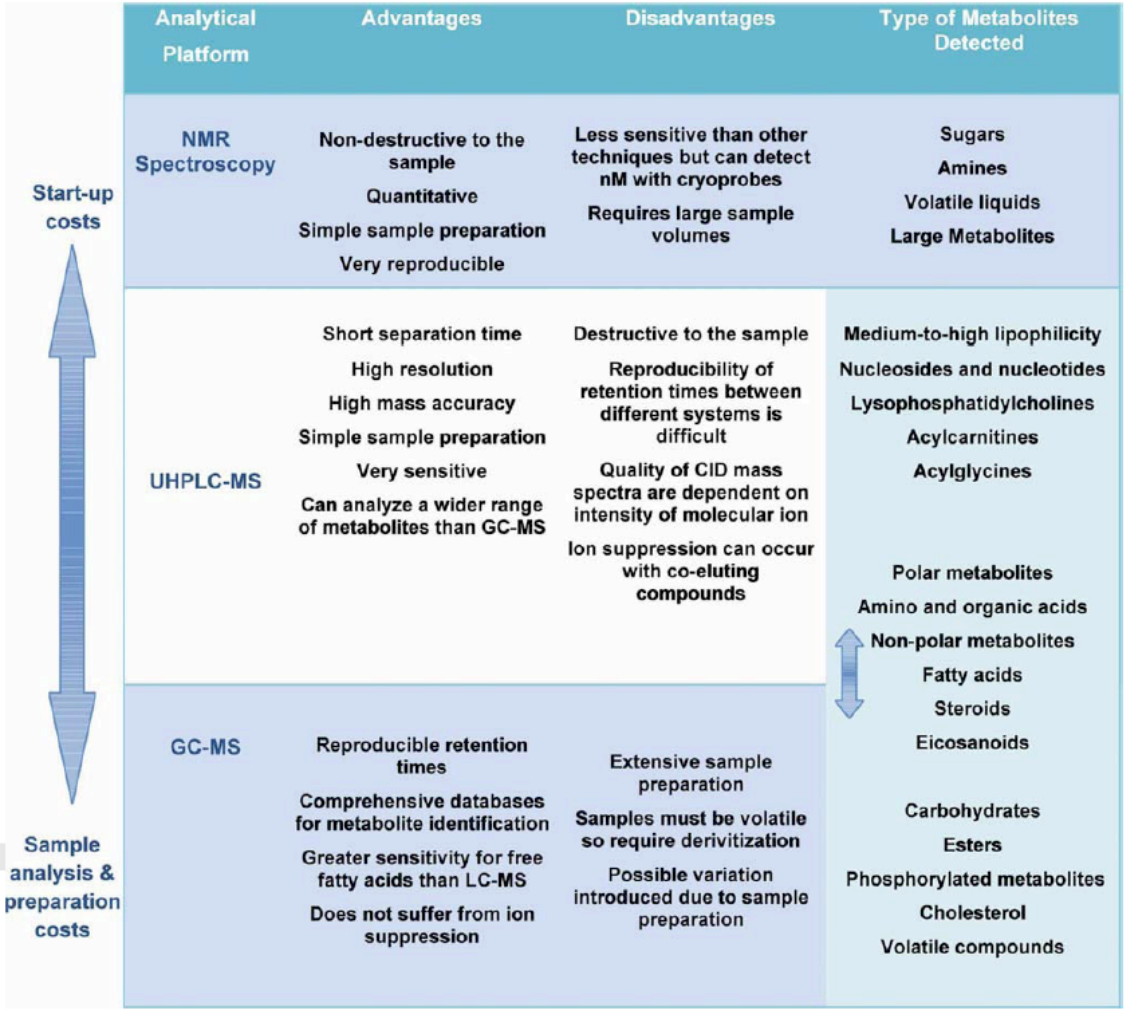


Figure 8. Advantages and disadvantages of different analytical methodologies for metabolomic studies.
From (*90*).

Data analysis software systems for metabolomic studies

LC-MS data collected in full scan mode are stored in the format of a matrix: signal intensities are stored as entries; retention times and *m/z* values can be inferred from the i values and j values. For example, a LC-MS data file from a 90 second gradient scanning from *m/z* 200 to 400 can be stored in a matrix [$a_{90,200}$]: 90 different i values represent 90 seconds; 200 different j values represent 200 different *m/z* values. The entry $a_{60,100}$ represents the signal intensity of the ion with a retention time of 60 seconds and a *m/z* value of 300.

LC-MS data can be visualized as a heat map with signal intensities represented with color (Figure 9). Alternatively, LC-MS data can be visualized in three dimensions, in which signal intensities are indicated with values in the z-axis (Figure 10).

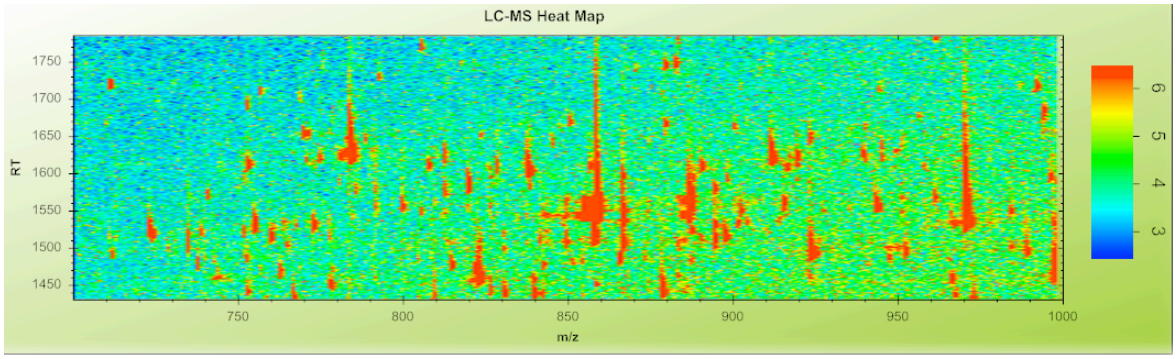


Figure 9. LC-MS data visualized in heat map by spectrolyzer
Data from Medicwave, Sweden.

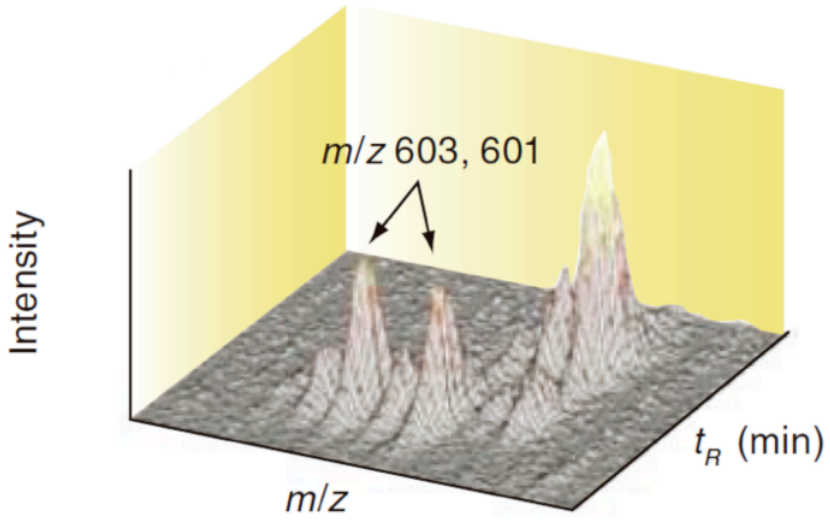


Figure 10. LC-MS data visualized in three dimensions.
From (*56*).

The goal of metabolomic studies for functional annotation is to identify a few changes in abundance among thousands of chemical features in several sets of LC-MS data. A single LC-MS data file from a 30 min gradient collected in full scan profile mode could be as big as 900 Mb, thus it is not realistic for manual inspections. Data analysis software systems with solid mathematical foundations have to be applied to facilitate this process. A number of software systems can be used to analyze and compare sets of LC-MS data for metabolomic studies (e.g., enzymatic incubations with and without co-factor; samples from wild-type versus knockout model organisms), including MZmine (*78, 79*), MetAlign (*80*), MAVEN (*81, 91*), MetaboAnalyst (*92*), and XCMS (*76*), as well as those sold by mass spectrometry venders, such as MarkerLynx (Waters), Mass Profiler Pro (Agilent), and Metabolic Profiler (Bruker).

A common practice in data analysis with LC-MS data is to deconvolute (e.g., noise removal, peak detection) each data file before further analysis. Deconvolution keeps the highest intensity of the peak feature and removes unnecessary information (e.g., background noise and signal intensities on peak shoulders), thus makes it much easier for further analysis (Figure 11). To compare sets of LC-MS data, retention times have to be aligned across different data files. Principal Component Analysis (PCA) and Partial Least Squares (PLS) analysis are commonly used mathematical procedures for data comparison after retention time alignment (*77*).

Signal normalization is another common practice used in data comparison. Several methods can be used: 1) peak intensities can be normalized by the peak with the highest signal intensity, assuming the abundance of the coressponding chemical are constant cross all samples. A major concern is that the assumption above may not be valid. Another concern is that any error in sample prepration (lost of recovery) will significantly affect downstream data analysis and leads to increased false-positives; 2) peak intensities can be normalized by total ion intensities. This method is good only if small changes are occured across samples, thus total ion intensities are similar across all samples. Errors in sample prepration is no longer a major concern with this data normalization method.

Although it is a common practice to normalize data before data comparison, it should be recognized that no data normalization is needed if sample preparation methods are highly reproducible and well designed.

A

		retention time (min)				
		t	t+0.02	t+0.04	...	t+0.2
	M	13225558	14971650	14124689		4872419
	M+0.067	14275327	21428922	18174298		4908518
	M+0.13	14715285	16374091	15119488		4818509
	...					
m/z	M+1	299557	569943	408393		664936
	...					
	M+2	1234126	3402372	1820382		910316
	...					
	M+3	577824	665613	643904		645588

B

		retention time (min)				
		t	t+0.02	t+0.04	...	t+0.2
	M	0	0	0		0
	M+0.067	0	21428922	0		0
	M+0.13	0	0	0		0
	...					
m/z	M+1	0	0	0		0
	...					
	M+2	0	3402372	0		0
	...					
	M+3	0	0	0		0

Figure 11. LC-MS data visualized in matrices: before (A) and after (B) data deconvolution.
Each entry in the matrices represents the signal intensity of the ion with the corresponding retention time (indicated by column number) and *m/z* value (indicated by row number). Data after deconvolution only have information of the highest intensity of each peak feature, and unnecessary information is removed (e.g., background noise and signal intensities on peak shoulders).

A software system, DoGEX ("Discovery of General Endo- and Xenobiotics"), was developed to identify doublet patterns after incubations with $^{16}O_2/^{18}O_2$ gas (*85*) (Figure 12). In principle, this software system can be used to identify any isotopic patterns in LC-MS data (*85*).

All data analysis approaches inevitably bring many false-positives, which have to be excluded by manual inspections and appropriate control experiments. Parameters (e.g., signal intensity threshold) carelessly set-up could give tremendous amounts of false-positives and shall be carefully adjusted. Thus, it is always a good practice to optimize these parameters at the beginning of any metabolomic studies. Many P450-mediated fatty acid oxidation reactions (*14*) can be used to train software systems to pick up enzymatic activities with low turnover rates.

Choice of instrumentation

Instruments from Thermo, Waters, Agilent, and Bruker all can be used for metabolomic studies. Data files from different instruments all can be transformed and visualized as shown in Figure 9-11.

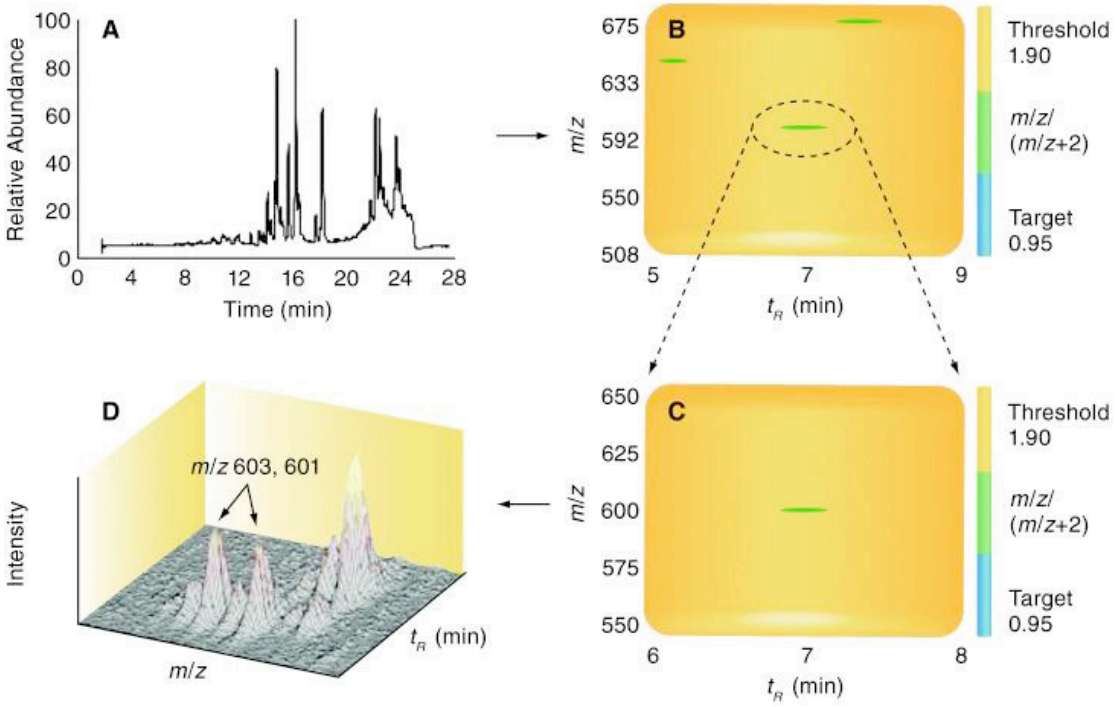


Figure 12. DoGEX is used in the analysis of substrates of P450 7A1 with human liver extracts.

A reconstituted system with P450 7A1 was incubated with a human liver extract with $^{16}O_2/^{18}O_2$ gas. LC-MS analysis of the product extract (derivatized with succinic anhydride) yielded the total ion current trace (A). Data analysis with DoGEX software generated the figure shown in B, where a green band indicates a potential doublet with M and M+2 intensity ratio near a targeted value, 0.95. The same region is expanded in C and shown in three dimensions in D, with a *m/z* 603/601 doublet identified. From (*56*).

Heterologous expression systems for human P450 enzymes

E. coli expression systems

Native human P450 coding sequences are typically expressed at low yields in *E. coli* expression systems. However, years of investigations have revealed several useful modifications and have made *E. coli* expression systems very popular for the expression of human P450 enzymes (*33, 37, 64, 93-102*). They are the cheapest and the least labor-intensive systems, and have the potential to produce active human P450 enzymes in large amounts. Thus, *E. coli* expression systems are usually tested for every p450 enzyme. The expression vector pCW(+ori), with two *tac* promoters arranged in tandem, has been tested for the expression of most human P450 enzymes (*33, 37, 64, 93-102*). Expression is normally induced by the addition of isopropyl-β-D-thiogalactoside (IPTG) to a bacterial culture with an OD_{600} between 0.5 and 1.3. Bi-cistronic and tri-cistronic expression systems can be established to co-express redox partners and P450 enzyme of interest, so that whole cell or membrane fractions can be directly used for functional assays (*82, 83*). DH5α has been the most used *E. coli* strain for heterologous expression of human P450 enzymes, but other strains have also been successful (e.g., JM109 (*97*), BL21 (*64*), TOPP3 (*98*), and TOP10 (*103*)). Trace elements ($FeCl_3$, $ZnCl_2$, $CoCl_2$, Na_2MoO_4, $CaCl_2$, $CuCl_2$, H_3BO_3) (*104*), δ-aminolevulinic acid, and thiamine are also commonly added to the bacterial culture to improve P450 expression levels (*33, 37, 64, 93-102*). The coding sequence is usually constructed with a histidine tag to facilitate protein

purification with a Ni-NTA column (*36, 64, 99*).

Besides expression temperature (23-30 °C) and expression time (20-48 h), several contributing factors have been summarized in published review articles (*93, 105*), including codon-optimization, N-terminal modification, and chaperone co-expression.

Codon-optimization

As elongation proceeds, *E. coli* ribosomes may not readily recognize certain codon sequences commonly seen in coding sequences from eukaryotic cells. Expressions with optimized codon sequences that favor *E. coli* tRNA usage have the potential to tremendously increase protein expression levels (*36, 99, 106*). Many public available tools can be used to design codon-optimized coding sequences, such as DNA 2.0 technology (*106*) and IDT (*107*). It is also believed that codons can be optimized to increase mRNA stabilities.

N-terminal modification

Mammalian P450 enzymes are membrane-bound proteins, and the subcellular location (mitochondria or endoplasmic reticulum) is largely determined by the hydrophobic amino acids at the N-terminal of the protein (*108*). Replacing these amino acids with coding sequences that are more friendly to *E. coli* expression systems has been commonly used to increase P450 expression levels (*64, 93-95, 97, 98, 100, 103, 109-112*). Commonly seen N-terminal modifications include amino acids MALLLAVFL (*64, 93, 97, 109*)

and MAKKTSS (*113*), although other modifications have been used as well (*64, 93-95, 97, 98, 100, 103, 109-112*). These N-terminal modifications also favor the tRNA usage of *E. coli* and might increase protein expression levels with mechanisms similar to codon-optimization.

Chaperone co-expression

The ability of the host system to fold P450 enzymes and incorporate heme correctly also affects P450 expression levels. Chaperones aid in protein folding by providing a hydrophilic environment (*114*). Thus, co-transformation and subsequent induction of chaperones is frequently used to increase expression levels of P450 enzymes (*36, 37, 99*). Although other chaperone systems are also available for co-expression, GroES/EL is the most commonly used one to increase P450 expression levels.

Baculovirus expression systems

Baculovirus expression systems are known for their capabilities to express proteins with high molecular weights and with post-translational modifications. Formation of inclusion bodies is usually not a problem for baculovirus expression systems. The Bac-to-Bac system (Invitrogen, CA) has become more and more popular for baculovirus expression: native coding sequences can be integrated into the baculovirus genome by site-directed transposition, and the isolated genome can be transfected into insect cells to produce high-titer virus stock. Bulk

cultured insect cells (in suspension) can be infected with high-titer baculovirus at different multiplicities of infection (MOI) (ratio of competent virus to cells) to optimize protein expression yields (Figure 13). Supplemented with heme precursors (e.g., δ-aminolevulinic acid), baculovirus expression systems can also express mammalian P450 enzymes in large amounts (*105*) and are oftentimes the last resort to obtain functional protein, after failure in other expression systems.

However, the disadvantages of baculovirus expression systems also made them the last approaches to be tested for protein expressions. One major problem is the high cost, both in time and consumables (i.e., expression medium, fetal bovine serum, T-flasks). A dedicated incubator (28 °C) is also required for culturing insect cells. If the expression vector needs to be constructed, and cryopreserved insect cells need to be revived, it often takes at least 1 month to be ready for the first expression trial even for highly experienced personnel. Accurately measuring the titer of the virus stock is another concern: plaque assay, which co-culture insect cells with diluted virus stocks in low melting point agarose, is the gold standard but takes 4-7 days to complete and is technically challenging (i.e., insect cells will be damaged if the agarose is too hot; agarose will solidify too soon if it is too cold); the qPCR based method usually overestimate the titer because it measures the amount of total genome DNA (virus with no infectivity will not be differentiated from fully-competent ones); immunoassays can give non-reproducible data (personal experience and communication with others). As a result, many highly experienced personnel

choose to perform expression trials with many different volumes of the virus stock, rather than spending time to measure the titer. To avoid the problems of setting up baculovirus expression systems from scratch, many institutes offer services to express proteins in insect cells for a charge.

Similar to *E. coli* expression systems, bi-cistronic or tri-cistronic systems can be established in baculovirus expression systems to co-express P450 enzymes and their redox partners (*115-117*).

While codon-optimization helped increase protein expression levels in *E. coli* expression systems (*36, 99, 106*), it is still not clear if it can benefit baculovirus expression systems.

Yeast expression systems

Compared with *E. coli*, yeast has endoplasmic reticulum and in general gives less trouble with the formation of inclusion bodies. The disadvantages of yeast expression systems include low P450 expression levels and the requirement of special instruments to effectively break the cell wall. Nevertheless, yeast expression systems are still used in several laboratories for heterologous expression of recombinant P450 enzymes (*40*).

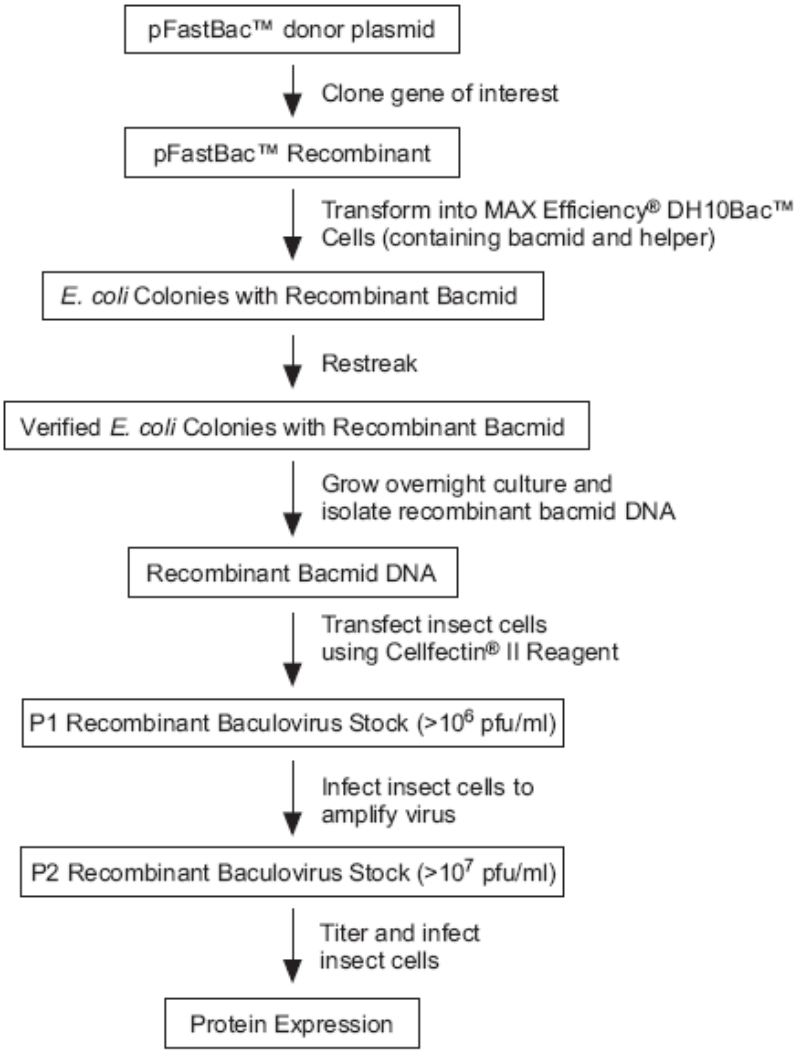


Figure 13. Flowchart for heterologous protein expression in insect cells with Bac-to-Bac system.

Mammalian cell expression systems

Mammalian cells offer the closest physiological conditions to express mammalian proteins compared with all other heterologous protein expression systems. However, the basal expression of other P450 enzymes could be a problem. The high cost and low P450 expression levels also made mammalian cell expression systems less popular.

Orphan human P450 enzymes covered in this thesis work

P450 2W1

Human P450 2W1 is preferentially expressed in colorectal cancer tissues (*118, 119*), and is regulated by gene methylation and reverse membrane orientation (*119, 120*). mRNA expression of human P450 2W1 has also been detected in colon, ileum, and testes in mice (*121*), but intensive searches in corresponding human tissues have not been reported. Increased levels of P450 2W1 expression positively correlate with the degree of tumor malignancy, and with decreased 10-year survival in colon cancer patients (*30, 122*).

Heterologous expression has been achieved in an *E. coli* expression system (*36*). Several P450 2W1-catalyzed reactions have been identified, including *N*-demethylation of benzphetamine (*36*), reduction of the drug candidate 1,4-*bis*{[2-(dimethylamino-*N*-oxide)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione (AQ4N) (*5*), oxidation of indole and its derivatives (*36, 119, 123*), oxidation of the free fatty acid (FFA) arachidonic

acid (at a very low turnover rate) (*36, 118*), and activation of a variety of chemical carcinogens to genotoxic forms (*36*).

P450 2S1

The highest mRNA expression of human P450 2S1 has been detected in lung, followed by small intestine, spleen, and colon (*47, 48, 124*). Human P450 2S1 protein has been detected in liver (*46*). P450 2S1 expression can be induced by 2,3,7,8-tetrachlorodibenzodioxin, indicating possible regulation by aryl hydrocarbon receptor (*49, 125*).

Heterologous expression has been achieved in an *E. coli* expression system (*36*). The catalytic specificity of P450 2S1 has been controversial (*56*). Interestingly, a reduction reaction has been reported with AQ4N as the substrate (*5*).

P450 2U1

The strongest mRNA expression of human P450 2U1 has been detected in brain and thymus (*39, 43*). Moderate mRNA levels have also been detected in heart, kidney, liver, lung, testis, and leucocytes (*39*).

Heterologous expression of human P450 2U1 has been achieved in a baculovirus expression system (*39*) and an *E. coli* expression system (*37*). Hydroxylations of fatty acids, both *ω*- and *ω*-1, have been reported (*39*).

P450 4V2

Human P450 4V2 protein can be detected in retina and corneum (*32, 51*). Mutations in the CYP4V2 gene (chromosome 4q35.1) are associated with Bietti's Crystalline Dystrophy (BCD), an autosomal recessive chorioretinal dystrophy featured with progressive night blindness (*51*).

Human P450 4V2 has been expressed and purified from a baculovirus expression system and the *ω*-hydroxylations of medium-chain fatty acids have been reported (*42*).

P450 4X1

The highest mRNA expression level has been detected in trachea and aorta, followed by breast, ovary, uterus, liver, and kidney (*50*). Expression has also been detected in brain, especially in the cerebellum, amygdala, and basal ganglia (*41*). Expression of human P450 4X1 is regulated by the peroxisome proliferator-activated receptor (PPAR) α (*50*).

Heterologous expression has been achieved in an *E. coli* expression system (*41*). Reported substrates include anandamide and arachidonic acid, both with very slow turnover rates (*41*).

P450 4Z1

High mRNA expression has been reported in breast, prostate, and ovarian tumors (*31, 44, 126-128*). Expression in normal breast tissue has also been

reported (50). Expression of human P450 4Z1 is also regulated by PPARα receptor (50).

Heterologous expression has been achieved in a yeast expression system (*40*). The only reported reactions so far are fatty acid oxidations (*40*).

CHAPTER II

HUMAN P450 2S1 IS REDUCED BY NADPH-P450 REDUCTASE

Introduction

Human P450 2S1 is one of the orphan human P450 enzymes with unknown physiological function (*35, 56*). Although human P450 2S1 has been expressed in *E. coli*, purified, and studied, its substrate specificity is still controversial (*35, 36*). To explain the limited catalytic activity of human P450 2S1, Bui and Hankinson (*129-131*) attempted to measure the formation of Fe^{2+}·CO complex under aerobic conditions and concluded that human P450 2S1 is not capable of interacting with NADPH-P450 reductase or accepting electrons from it. In contrast, Nishida *et al.* (*5*) reported the observation of a Fe^{2+}·CO complex anaerobically. In that study, human P450 2S1 was reported to reduce the pro-drug AQ4N to its mono *N*-oxide intermediate AQ4M [1-{[2-(dimethylamino-*N*-oxide)ethyl]amino}-4-{[2-(dimethylamino)ethyl]amino}-5,8-dihydroxyanthracene-9,10-dione] and finally to AQ4 [1,4-*bis*{[2-(dimethylamino)ethyl]amino}-5,8-dihydroxy-anthracene-9,10-dione]. A net 2-electron transfer mechanism was proposed (*5*). In this study, we reproduced the anaerobic reduction of AQ4N, measured NADPH oxidation rates, and measured the reduction kinetics using a stopped-flow apparatus. Our results confirm that human P450 2S1 can efficiently accept electrons from NADPH- P450 reductase.

Materials and methods

Chemical and reagents

Deferoxamine, mitoxantrone, protocatechuate, and protocatechuate dioxygenase were purchased from Sigma-Aldrich (St. Louis, MO). AQ4N, AQ4M, and AQN were gifts from Dr. Klaus Pors (University of Bradford, UK). Human P450 2S1 (*36*) and rat NADPH-P450 reductase (*132*) were expressed in *E. coli* and purified as previously reported.

Anaerobic reduction

Anaerobic reduction experiments were performed using an OLIS RSM-1000 stopped-flow instrument (On-Line Instrument Systems, Bogart, GA). Samples (in glass tonometers) were deaerated using an argon/vacuum manifold as described previously (*133*), utilizing a protocatechuate/protocatechuate dioxygenase oxygen-scrubbing system (*134*). Human P450 2S1 (2 μM) was pre-incubated with NADPH-P450 reductase (4 μM), 160 μM L-α-dilauoryl-*sn*-glycero-3-phosphocholine (DLPC), 100 mM potassium phosphate buffer (pH 7.4), and AQ4N (200 μM), and reduced upon the addition of NADPH (150 μM) from a second syringe. Rates were analyzed using the OLIS software and in GraphPad Prism (GraphPad, San Diego, CA).

Enzyme activity determination of human P450 2S1 towards AQ4N

Reactions were carried out in duplicate in 1.5-ml Eppendorf tubes at 37 °C. Anaerobic incubations were carried out under a nitrogen atmosphere inside of a glovebox (Labconco Protector® Controlled Atmosphere). Enzyme reaction mixtures typically contained 0.1 μM human P450 2S1, 0.2 μM NADPH-P450 reductase, 160 μM DLPC, 1 mM desferoxamine, 100 mM potassium phosphate buffer (pH 7.4), and 200 μM AQ4N (*5*). After pre-incubation for 5 min, the reactions were started by the addition of an NADPH-generating system including 100 μl of 100 mM glucose 6-phosphate, 50 μl of 10 mM $NADP^+$, and 2 μl of a 1 mg ml^{-1} solution of yeast glucose 6-phosphate dehydrogenase (*3*), and aliquots were terminated by the addition of three volumes of CH_3OH (with 5 μM mitoxantrone as an internal standard) at times of 0, 1, 2, 3, 4, 5, and 6 min. Samples were prepared and analyzed by high-performance liquid chromatography (HPLC) as described (*135*) with a Thermo Hypersil GOLD column (150 mm × 2.1 mm I.D.) employing isocratic elution with 50 mM NH_4HCO_2 buffer (pH 3.6):CH_3CN (89:11, v/v). Formation of AQ4M was used to determine enzyme activity.

Determination of NADPH consumption rates

NADPH oxidation rates were determined using 0.05 μM human P450 2S1, 0.10 μM NADPH-P450 reductase, 160 μM DLPC, 1 mM desferoxamine, and 100 mM potassium phosphate buffer (pH 7.4). For the determination of NADPH

consumption rates under anaerobic conditions, reconstituted enzyme mixtures were deaerated using an argon/vacuum manifold as described previously (*133*). Reconstituted enzymes were pre-incubated for 5 min at 37 °C in the presence or absence of AQ4N (200 μM). Reactions were initiated with the addition of NADPH to a final concentration of 150 μM and A_{340} was monitored (Cary 14/OLIS instrument). Rates were calculated using the value $\Delta\epsilon_{340}$ = 6.22 mM^{-1} cm^{-1}. Experiments were conducted in duplicate.

Results and discussion

The reduction of AQ4N to AQ4M is proposed to be a net 2-electron transfer mechanism (*5*). The hypoxic activation of AQ4N (to AQ4M) by human P450 2S1 was reproduced, measured to be 18 min^{-1} under anaerobic conditions and <0.5 min^{-1} under aerobic conditions.

The steady-state rates of oxidation of NADPH by human P450 2S1 in the presence of NADPH-P450 reductase were also measured (Table 2). The higher oxidation rate under aerobic conditions confirms the acceptance of electrons by human P450 2S1 from NADPH-P450 reductase. Because the rate of AQ4N catalysis under aerobic conditions is very low, the increased NADPH consumption in the presence of AQ4N presumably comes from electron transfer to form partially reduced oxygen products instead of being transferred to AQ4N.

The rate of product formation was measured to be 18 min^{-1}. The anaerobic NADPH consumption rate without air in the presence of AQ4N was 28

min^{-1} and 13 min^{-1} in the absence of AQ4N, suggesting roughly quantitative coupling efficiency (i.e., all additional electrons are transferred to substrate AQ4N).

The reduction step can be studied by monitoring the formation of the Fe^{2+}·CO complex near 450 nm in the absence of O_2, which will compete with CO and lead to P450 re-oxidation (*136, 137*). Our kinetic reduction data, measured in a stopped-flow apparatus (Figure 14), clearly showed a rapid increase in A_{450}, indicating the acceptance of electron by P450 2S1 from NADPH-P450 reductase.

		v (min^{-1})
Plus air	Minus AQ4N	23.3 ± 2.6
	Plus AQ4N	45.7 ± 4.7
Minus air	Minus AQ4N	12.7 ± 2.9
	Plus AQ4N	28.2 ± 0.8

Table 2. NADPH oxidation rates.

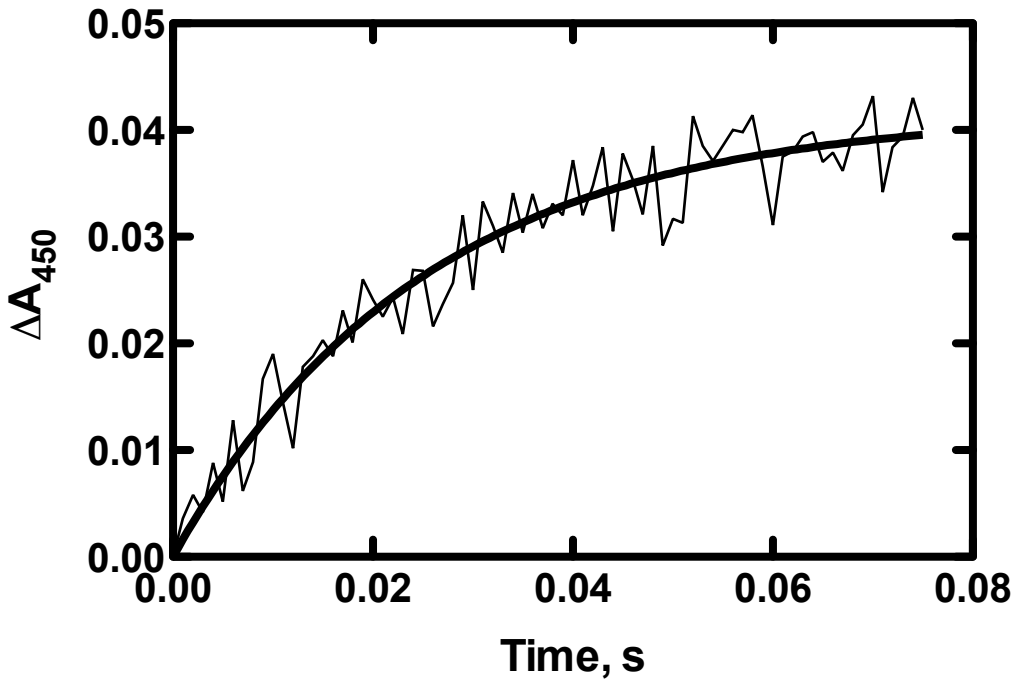


Figure 14. Reduction kinetics of P450 2S1 (with AQ4N).
Stopped-flow absorbance trace of the reduction of P450 2S1 (2 μM) by NADPH-P450 reductase (4 μM) in the presence of DLPC (120 μM) and AQ4N (200 μM) upon the addition of NADPH (150 μM). The rate was 40 s^{-1} (measured by averaging results from sixteen independent shots).

Bui and Hankinson reported that human P450 2S1 could not be reduced by NADPH-P450 reductase and that several catalytic activities of P450 2S1 could be observed if reactions were supported by oxygen surrogates (i.e., hydroperoxides) (*129-131*). In contrast, Nishida *et al.* reported the observation of a Fe^{2+}·CO complex produced by human P450 2S1 and AQ4N (rate not measured) (*5*). A serious caveat in the work of Bui and Hankinson (*129*) is that the reduction work was done aerobically, under which conditions ferrous P450s are reoxidized (*138*). Nishida *et al.* (*5*) also emphasized that the conclusion of Bui *et al.* (*129-131*) that human P450 2S1 normally uses lipid peroxides for catalytic function is invalid, in that many P450s can react with lipid hydroperoxides through a shunt pathway that generates lipid alkoxy and peroxy radicals. These radicals can enter co-oxidation reactions outside of P450 active sites (*139, 140*). In another publication, Bui *et al.* (*131*) reported the isomerization activity of P450 2S1 in a NADPH-independent manner. However, such isomerization activity has been seen with several other P450 enzymes and its physiological role still to be validated (*141-143*).

In conclusion, our results are in agreement with previous findings (*5*) that human P450 2S1 can accept electrons from NADPH-P450 reductase. A physiological role of human P450 2S1, if one exists, remains to be revealed.

CHAPTER III

FUNCTIONAL ANNOTATION OF ORPHAN HUMAN P450 2W1, 2S1, AND 4X1 BY METABOLOMIC APPROACHES

Introduction

Although the genomic sequences of human and numerous other organisms have been established, the functions of less than one-half of the proteins have been annotated, even in *E. coli*. Thus, an important and challenging task in modern biochemistry is the annotation of protein functions, including the establishment of the catalytic activities of novel enzymes with unknown substrates (*144, 145*). P450 enzymes play important roles in the metabolism of a large number of compounds, including sterols, fatty acids, eicosanoids, vitamins, and xenobiotics (*63*). It has been estimated that P450 reactions are involved in ~75% of the enzymatic transformations of small molecule drugs (*21, 23*). There are 57 human P450 genes identified in the human genome, and about one fourth of them can be termed “orphans” because of their unknown physiological or other functions (*20, 21*).

LC-MS is one of the most widely used analytical methods for metabolomic analysis and has proved to be a powerful approach in substrate searches (*68, 71, 75, 76*). The most commonly used method is to compare incubation mixture with and without the functional enzyme or co-factors to identify any changes.

Accumulated chemical features after incubation are potential products (*68-71*). Recently we have developed a general strategy for the identification of endogenous substrates of human P450s in tissue extracts using LC-MS assays and the program DoGEX (*14, 85*). The approach is based on the fact that the majority of P450-mediated reactions involve the incorporation of an oxygen atom into the substrate (i.e., the product is 16 amu heavier than the substrate). Incubation of a 1:1 mixture of $^{18}O_2$ and $^{16}O_2$ gas with tissue extracts generates products as M/M+2 doublets in the MS spectra, which can be identified by the program DoGEX (*14, 85*). This strategy has been validated (*14*) and used to identify endogenous substrates for P450 4F11 (*64*).

The aim of this study was to identify endogenous substrates for orphan human P450 2W1, 2S1, and 4X1. Purified enzyme was used to conduct untargeted substrate searches with tissue extracts (i.e., site of P450 expression) using either XCMS analysis (*69-71*) or an isotopic labeling approach (*14, 64, 68-71*). A series of lysophospholipids and FFAs were identified as novel substrates for P450 2W1, and the isomer- and enantiomer-selectivity of P450 2W1-catalyzed lysophospholipid oxidations have been characterized. The identities of the oxidation products were defined, and steady-state kinetics of the P450 reactions was determined. No endogenous substrate has been identified for human P450 2S1 or 4X1.

Experimental procedures

Materials and reagents

Human P450s 1A2 (*97*), 2A6 (*146*), 2C8 (*147, 148*), 2D6 (*149*), 2E1 (*102*), 2S1 (*36*), 2W1 (*36*), 3A4 (*101*), 7A1 (*150*), 4X1 (*41*), and rat NADPH-P450 reductase (*132*) were expressed in *E. coli* and purified as previously described. Human P450 2C19 was expressed (*151*) and purified using the same protocol as that for human P450 2C9 (*96*). All phospholipids were purchased from Avanti Polar Lipids (Alabaster, AL). The BSTFA:TMCS:TMSI:pyridine mixture (3:2:3:10, v/v/v/v) was purchased from Regis Technologies (Morton Grove, IL). Preparative TLC was conducted on pre-coated 2 mm silica gel GF-254 plates (Analtech, Newark, DE). All other reagents and solvents were obtained from general commercial suppliers.

Tissue extracts

Malignant human colorectal cancer samples were obtained from the Translational Pathology Shared Resource, Vanderbilt University School of Medicine. Human liver and lung samples (from organ donors) were obtained from Tennessee Donor Services. Bovine brains were purchased from Pel-Freez (Rogers, AR). If not otherwise stated, extracts were prepared from pooled samples using Folch reagent ($CHCl_3$:MeOH, 2:1, v/v) as previously described

(*14*). Ethanolic tissue extracts were prepared by homogenizing 20 volumes of EtOH with tissues from different brain substructures or pooled human lung/liver samples. Tissue homogenate was stirred for 20 min at 23 °C, filtered through paper, and dried under a stream of N_2.

Enzymatic incubations

In vitro P450 incubations were performed in 1.0 ml of 100 mM potassium phosphate buffer (pH 7.4) each containing purified human P450 enzyme (1.0 μM), NADPH-P450 reductase (2.0 μM), DLPC (150 μM), an NADPH-generating system (*3*), and an aliquot of an ethanolic solution of tissue extracts (1%, v/v). For comparison of enzymatic incubation with and without co-factor (i.e., NADPH-generating system) by XCMS, incubations were performed in duplicate at 37 °C for 30 min. For incubations without the NADPH-generating system, the system was added after quenching. Enzymatic incubation mixtures (with tissue extracts made with Folch reagent) were quenched with 2 ml of CH_2Cl_2 and centrifuged at 3×10^3 *g*. Lower organic layer was dried under a stream of N_2. Enzymatic incubation mixtures (with tissue extracts made with EtOH) were quenched with MeOH and centrifuged at 3×10^3 *g*. Supernatant was dried under a stream of N_2. All samples were redissolved in 100 μl of MeOH, and centrifuged again at 3×10^3 *g* followed by LC-MS data collection and data analysis by XCMS.

For $^{16}O_2/^{18}O_2$ isotopic labeling experiments, reactions were performed using the method described previously (*14*), except that 100% $^{16}O_2$ and 97% $^{18}O_2$

gas were used in two individual Thunberg tubes. The enzyme reactions were initiated by the addition of an NADPH-generating system (*3*). After incubation at 37 °C for 60 min, the contents of the two Thunberg tubes were combined (equal volumes) and quenched with CH_2Cl_2 (for incubations with tissue extracts made with Folch reagent) or MeOH (for incubations with tissue extracts made with EtOH) and centrifuged at 3×10^3 *g*. Samples was dried under a stream of N_2, redissolved in 100 µl of MeOH, and centrifuged again at 3×10^3 *g* for LC-MS data collection. The oxidation products were identified as M and M+2 doublets (*14*) with a newly developed approach, which was based on the software MZmine2 (*79*) and an in-house made Matlab program.

Dansylation was performed as previously described (*89*) with slight modifications. Briefly, samples were dissolved in 200 µl of CH_2Cl_2 containing 2 mg of dansyl chloride, 0.5 mg of 4,4-dimethylaminopyridine, and 2 µl of triethylamine and heated at 65 °C for 1 h. Each sample was dried under a stream of N_2 and then dissolved in 100 µl of CH_3CN before LC-MS analysis.

LC-MS based metabolomics and data analysis

LC separation was performed with a Waters Acquity UPLC system (Waters, Milford, MA) with an Acquity BEH octadecylsilane (C_{18}) UPLC column (1.7 µm, 1.0 mm × 100 mm) at 50°C. Samples (10 µl) were injected onto the UPLC column, and components were eluted with a linear gradient increasing from 95% (v/v) mobile phase A (10 mM $NH_4CH_3CO_2$ in a 5:95 (v/v) CH_3CN/H_2O

mixture) to 100% mobile phase B (10 mM $NH_4CH_3CO_2$ in a 95:5 (v/v) CH_3CN/H_2O mixture) over 20 min, and held at 100% mobile phase B for 5 min at a flow rate of 0.15 ml min^{-1}. Data was collected with a ThermoFinnigan LTQ mass spectrometer (ThermoFisher, Watham, MA) equipped with an ESI source or APCI source scanning from *m/z* 80 to 800 (profile mode for doublet search; centroid mode for XCMS analysis), using the same instrument parameters as previously described (*14*). A ThermoFinnigan Orbitrap mass spectrometer (ThermoFisher, Watham, MA) was used for the collection of HRMS data.

Characterization of oxidation products by human P450 2W1

Characterization of oxidation products was performed by GC-MS after preparing the corresponding TMS ethers. Oxidation products of FFAs were obtained by incubating each FFA (100 μM) in 1.0 ml of reaction mixture containing phosphate buffer, P450 2W1, NADPH-P450 reductase, DLPC, and an NADPH-generating system (see above). Each sample was extracted with 2.0 ml of CH_2Cl_2 and dried under a stream of N_2. Epoxides were converted to dihydrodiols after incubation with H_2O (adjusted to pH 2) at 23 °C for 10 min and extracted again with CH_2Cl_2. Oxidation products of lysophosphatidylcholines (LPCs) were obtained by incubating each LPC (100 μM) with 1.0 ml of the reaction mixture (see above). The reactions were quenched with 2.0 ml of CH_3OH containing butylated hydroxytoluene (0.005%, w/v) and 1.0 ml of aqueous KOH (15%, w/v). The mixtures were then mixed with a vortex device,

purged with Ar, and incubated at 37 °C for 30 min to hydrolyze the oxidized LPCs and release the oxidized fatty acids (*152*). The mixtures were acidified to pH 2 with HCl, and the oxidized fatty acids were extracted into CH_2Cl_2. TMS derivatization was performed with 20 µl of silylation reagent (BSTFA:TMCS:TMSI:pyridine, 3:2:3:10, v/v/v/v) at 60 °C for 30 min. The resulting TMS derivatives were analyzed with a Finnigan Trace GC DSQ mass spectrometer (Thermo Fisher, Watham, MA) using the electron impact negative ion mode(*14*). Samples (4 µL) were injected onto a HP-5 capillary column (30 m × 0.25 mm × 0.25 µm, Agilent Technologies, Santa Clara, CA). The temperature program started at 150 °C (held for 1 min), and then increased to 300 at 8 °C min^{-1}, and held at 300 °C for 6 min more.

Kinetic analysis of P450 2W1-mediated oxidations

Substrate concentrations ranging from 0 to 200 µM were used for steady-state kinetic studies. Reactions were run in duplicate at 37 °C for 15 min. Oxidation products of FFAs and LPCs were extracted as described above. The products were derivatized with 20 µl of 10% (v/v) *N*,*N*-diisopropylethylamine in CH_3CN and 40 µl of 10% (v/v) pFBB in CH_3CN at 37°C for 20 min (*152*). pFBB-derivatized samples were dried under a N_2 stream and then derivatized with 20 µl of BSTFA and 7 µl of dry dimethylformamide at 37°C for 20 min (*152*) and analyzed by GC-MS in the chemical ionization mode. FFA C17:0 and C19:0 standards were used to prepare calibration curves for kinetic analysis of FFA

C16:0 and C18:1 oxidation; 17:0 LPC and C19:0 LPC were used to prepare calibration curve for kinetic analysis of 16:0 LPC oxidation. Epoxy-18:1 LPC and epoxy-16:1 PC, purified and quantified by a phosphorus assay (*153*), were used to prepare calibration curves for kinetic analysis of 18:1 LPC oxidation. Epoxy-18:1 LPC was chemically synthesized by incubating 18:1 LPC with an excess amount of *m*-chloroperoxybenzoic acid. The reaction mixture was streaked on a preparative fluorescent TLC plate, developed with CH_3OH:$CHCl_3$ (1:1 v/v), and visualized by UV light. The lower band was scraped and extracted by the same solvent, taken to dryness using a rotary evaporator, and dissolved in C_2H_5OH containing 1% diisopropylethylamine (v/v). LC-MS analysis confirmed that all 18:1 LPC was converted into epoxy-18:1 LPC. Epoxy-16:1 PC was synthesized and quantified with the same method (see above). For rate comparisons of different lysophospholipids, P450 2W1 was incubated with 100 μM 18:1 LPC, 18:1 lysophosphatidylinositol (LPI), 18:1 lysophosphatidylserine (LPS), 18:1 lysophosphatidylglycerol (LPG), 18:1 lysophosphatidylethanolamine (LPE), or 18:1 lysophosphatidic acid (LPA) in triplicate, and the rates were determined as described above.

Purification of sn-1 and sn-2 LPC

HPLC was used to separate the two isomers. *sn*-1 and *sn*-2 LPCs were monitored at 196 nm and baseline separation was achieved with a Phenomenex prodigy ODS HPLC column (5 μm, 2.0 mm × 150 mm) (*154*). An isocratic

solution of 1:1 (v/v) CH_3CN/H_2O (pH adjusted to 5 with NH_4CO_2H) was used to resolve the two isomers at 40 °C, at a flow rate of 0.5 ml min^{-1}. The collected *sn*-1 and *sn*-2 LPC fractions were frozen and concentrated by lyophilization prior to enzymatic reaction.

Chiral analysis

Optically pure (9*S*,10*R*)- and (9*R*,10*S*)-epoxystearic acids were produced by hydrogenating pure (9*S*,10*R*)-epoxy-12*Z*-octadecenoic acid and (9*R*,10*S*)-epoxy-12*Z*-octadecenoic acid (*155*) with Pd powder under a H_2 stream for 3 min (*14*). The enantiomers of 9,10-epoxystearic acid were separated by normal phase HPLC with a Waters Alliance 2695 HPLC pump (Waters, Milford, MA) and a Chiralpak AD column (5 μm, 4.6 mm × 25 cm). An isocratic solvent of a 100:2:0.05 (v/v/v) hexanes/CH_3OH/CH_3CO_2H mixture was used to resolve the enantiomers at a flow rate of 1 ml min^{-1} at room temperature. The retention times of (9*S*,10*R*)- and (9*R*,10*S*)-epoxystearic acids were determined to be 16.9 min and 18.7 min, respectively. Epoxide generated from FFA C18:1 was extracted with CH_2Cl_2 after enzymatic reaction. Epoxide generated from 18:1 LPC was subjected to hydrolysis as described above, and five volumes of 1 M potassium phosphate buffer (pH 7.4) was added to neutralize the pH before extraction with CH_2Cl_2. After dried under a N_2 stream, the epoxide was analyzed as described above and detected with the ESI negative ion mode.

Other methods

UV-visible spectra were recorded using an Aminco DW-2a/OLIS spectrophotometer. P450 concentrations were estimated spectrally as previously described (*3*).

Results

New approach for doublet searches

A new approach for doublet searches was developed based on the MZmine2 software and an in-house made Matlab program. It has improved performance regarding both precision and recall compared with the previously described DoGEX program (*85*), although the results cannot be displayed as heat maps. LC-MS raw data files were imported to MZmine2. After chromatogram construction and deconvolution (*79*), the data is displayed as a peak list, which includes the *m/z* value, retention time, duration, and peak height of individual peaks. The peak list is copied to an Excel file and exported as a CSV (Comma Separated Values) file. A Matlab program was developed to extract peak information from the CSV file and search for doublet patterns.

A data file generated from incubation of human P450 2C8 with human liver extracts, an NADPH generating system (*3*), and $^{16}O_2/^{18}O_2$ gas—for which number of doublets was already known due to previously reported P450 reactions (*14*)—was used to set up parameters (e.g., peak intensity threshold) so that as many doublets can be identified as possible. Two peaks would be

identified as a doublet if they fit the following criteria: mass difference of 1.9-2.2 Dalton, retention time difference <1.5 s, height ratio of 0.67-1.5, peak height >1000 (arbitrary units), and duration >1 s. To compare the performance of the new approach with the previously reported DoGEX program, two data files (P450 2C8 incubated with liver extracts and $^{16}O_2/^{18}O_2$ gas, and P450 7A1 incubated with liver extracts and $^{16}O_2/^{18}O_2$ gas, followed by dansylation (*89*)) were analyzed by both approaches and their true-positive rate (TPR) and positive predictive value (PPV) were calculated as follows (*156*):

$$TPR = \frac{TP}{TP + FN}$$

$$PPV = \frac{TP}{TP + FP}$$

Where TP (true-positive) is the percentage of hits identified by the software and confirmed in Qual Browser (Thermo Fisher, Watham, MA), FP (false-positive) is the percentage of hits that does not share the features of doublets, and FN (false-negative) is the number of hits that were not identified by the software but yet they are true doublets generated from P450 enzyme - catalyzed oxidation reactions. TPR is called “recall” and PPV is called “precision”.

Unlike DoGEX, which uses profile data during the entire data processing procedures, the new approach uses deconvoluted data before searching for doublets, and therefore many false-positives (from intensities of peak shoulders)

can be avoided. The new software performs better in both precision and recall (Table 3).

		2C8 data	7A1 data
New approach	Precision	11.4 %	25 %
	Recall	100 %	100 %
DoGEX	Precision	1.1 %	0.8 %
	Recall	75 %	100 %

Table 3. Comparison of doublet search approaches.

Searches for P450 2W1 substrates in malignant human colorectal cancer extracts

Purified P450 2W1 was incubated with malignant human colorectal cancer extracts, an NADPH generating system (*3*), and $^{16}O_2/^{18}O_2$ gas mixtures. In principle, all doublets (*m/z* M/M+2) in LC-MS data result from the addition of an oxygen atom to endogenous substrates, with the general concept described previously (*85*). To profile as many metabolites as possible, samples were analyzed with both ESI and APCI sources in both the positive and negative ionization modes.

The doublets *m/z* 538/540 (Figure 15) in the ESI positive ion mode and *m/z* 269/271, 271/273, 295/297, 297/299, and 319/321 in the ESI negative ion mode were identified (Table 4). Product candidates were found only in the samples incubated with P450 2W1, NADPH-P450 reductase, and NADPH-generating system but not in the samples absent any of these. All doublets were further confirmed to be oxidation products by comparison with the incubations done only with $^{16}O_2$ gas, in which the *m/z* M+2 peaks were absent.

The *m/z* values of the respective substrates can be deduced from the *m/z* values of the products by subtracting 16 amu (oxygen). Therefore, the molecular masses of the putative substrates were calculated to be 521, 254, 256, 280, 282, and 304. MS fragmentation of the *m/z* 538 ion in the ESI positive ion mode produced a daughter ion of *m/z* 184 (Figure 15), indicative of a phosphocholine group (*157*). The LIPIDMAPS database (http://www.lipidmaps.org) suggested that 18:1 LPC (*m/z* 521) was a likely substrate. MS fragmentation analysis of the products detected in the ESI negative ion mode and the search of LIPIDMAPS database suggested that FFAs C16:0, C16:1, C18:1, C18:2, and C20:4 were likely substrates. To confirm the identities of the substrates, 18:1 LPC and five FFAs were incubated with P450 2W1 and NADPH, and the extracted products were analyzed by LC-MS/MS. All of the product peaks formed in the incubations with the authentic compounds yielded exactly the same peaks identified by the new software system. These results identified 18:1 LPC and the C16:0, C16:1, C18:1, C18:2, and C20:4 FFAs as substrates for P450 2W1 in malignant human colorectal cancer tissue.

ID	mass (*m/z*)	t_R (min)	intensity	ID	mass (*m/z*)	t_R (min)	intensity
Identified in the ESI negative ion mode							
6099	269.17	9.76	3554	1748	271.25	9.78	4001
6091	269.17	8.36	4387	19	271.17	8.37	3216
4236	269.17	10.01	2612	20	271.17	9.99	2744
6098	269.17	9.60	9022	1747	271.17	9.61	6466
3437	269.25	10.59	4499	5748	271.17	10.58	3020
1740	271.25	9.04	20985	7446	273.25	9.05	14853
1739	271.17	8.88	1873	4228	273.17	8.88	1328
10838	295.25	10.44	6876	2795	297.25	10.46	7904
2793	297.25	10.20	18480	6803	299.25	10.20	13914
7957	297.25	11.15	19947	9905	299.25	11.15	17247
7959	297.25	11.32	113883	9907	299.25	11.32	103711
203	319.17	10.54	1037	3414	321.25	10.55	1329
Identified in the ESI positive ion mode							
12861	538.33	10.12	7169	16842	540.33	10.12	6923
5691	538.33	9.36	8220	21497	540.25	9.35	6262

Table 4. Doublets identified by Matlab program and confirmed in Qual Browser.

XCMS analysis with P450 2W1 incubating mixtures with and without NADPH-generating system (*3*) identified the accumulation of oxidation products from C16:0, C16:1, C18:1, C18:2, and C20:4 FFAs, but not the accumulation of the oxidation product of 18:1 LPC. A signal decrease in C16:0, C16:1, C18:1, C18:2, and C20:4 FFAs as well as in 18:1 LPC was not observed by XCMS nor by manual inspections with LC-MS data in Qual Browser.

Dansylation followed by enzymatic incubation (for XCMS and for doublet searches) did not identify extra substrate.

Searches for human P450 2S1 substrates in human lung extracts

Both XCMS and doublet searches were conducted to identify endogenous substrate for human P450 2S1 in human lung extracts. Dansylation was also used to improve the sensitivity towards unactivated alcohols. However, no candidate was identified as a potential substrate or product.

Searches for human P450 4X1 substrates in bovine brain extracts, human liver extracts, and human lung extracts

Both Folch reagent and EtOH were used to make tissue extracts to expand compound diversities. Extracts were made with bovine brain samples, human liver samples, and human lung samples. In order to preserve the compound diversity in ethanolic extracts, enzymatic incubation mixtures with ethanolic tissue extracts were quenched with MeOH for LC-MS analysis (protein was removed by centrifugation).

Both XCMS and doublet searches were conducted to identify endogenous substrate for human P450 4X1. Dansylation was also used to improve the sensitivity towards unactivated alcohols. However, no candidate was identified as a potential substrate or product.

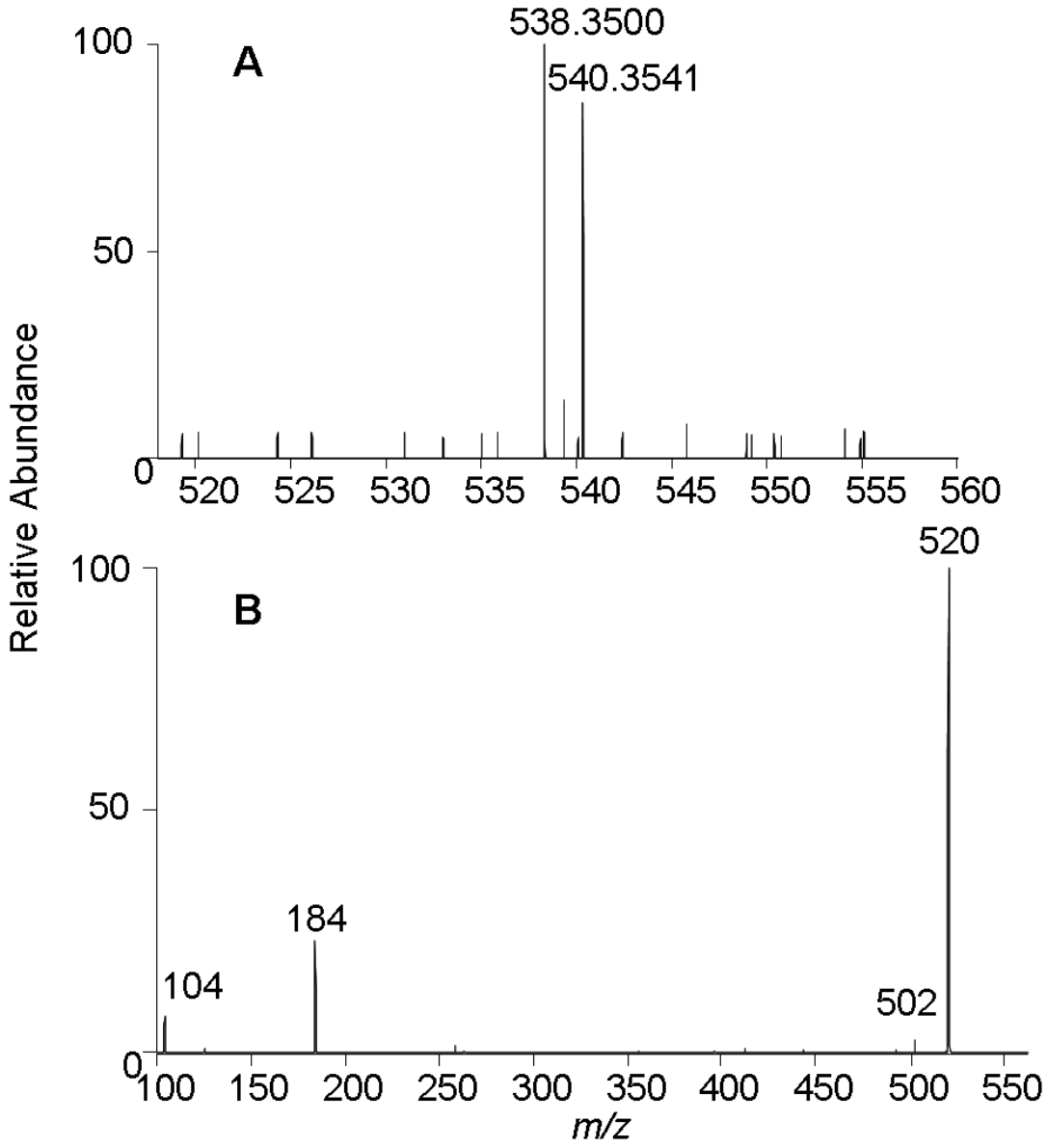


Figure 15. LC-MS/MS analysis of the *m/z* 538/540 doublet.
(A) HRMS of *m/z* 538/540 produced from the incubation of P450 2W1, NADPH, colon cancer extracts, and $^{18}O_2/^{16}O_2$ gas (1:1, v/v). (B) MS/MS fragmentation of the 18:1 LPC oxidation product *m/z* 538 in the ESI positive ion mode.

Specificity of 18:1 LPC as a substrate for other human P450 enzymes

The specificity of 18:1 LPC as a substrate for different human P450 enzymes was investigated. 18:1 LPC was incubated with purified human P450s 1A2, 2A6, 2C8, 2C19, 2D6, 2E1, 2S1, 2W1, 3A4, and 7A1, and the extracted products were analyzed by LC-MS (Figure 16). Although many other human P450 enzymes share similar catalytic efficiencies toward fatty acids (*14, 64*), they showed little activity with 18:1 LPC. We conclude that 18:1 LPC is a substrate relatively specific for P450 2W1.

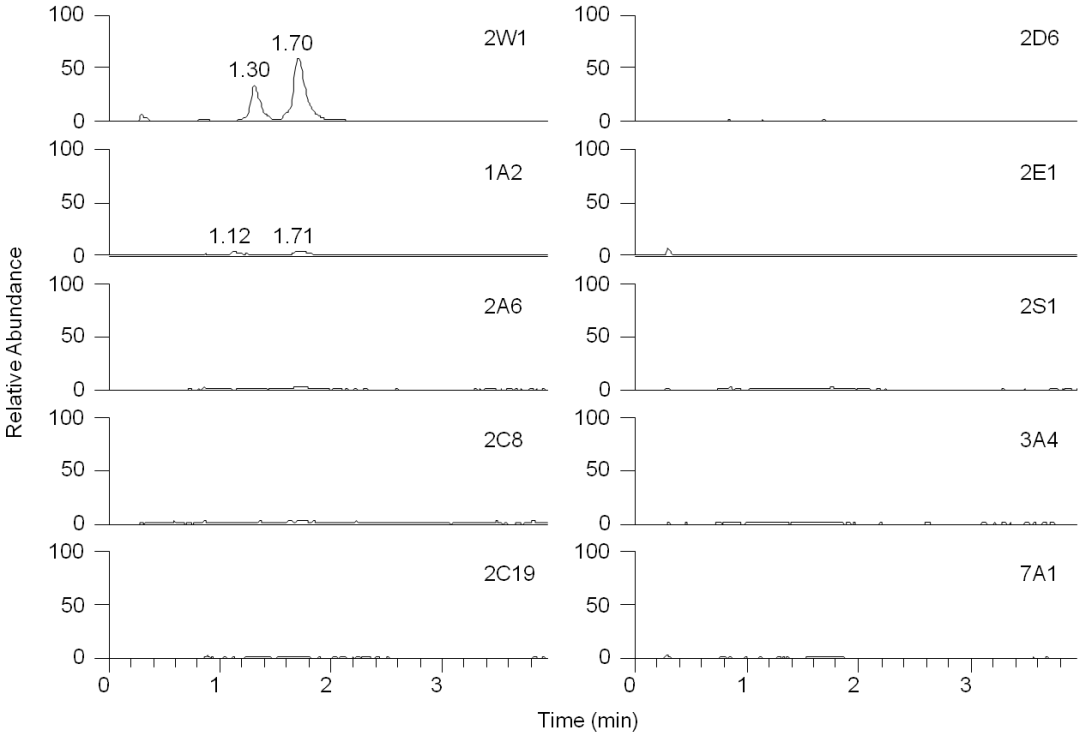


Figure 16. Specificity of 18:1 LPC as a substrate for human P450 enzymes. Selected ion chromatograms of 18:1 LPC oxidation products (*m/z* 538) after incubating 18:1 LPC with purified P450 enzymes. Retention times (t_R) are indicated on the chromatograms.

Characterization of oxidation products

GC-MS assays of TMS derivatives were used to characterize the oxidation products of P450 2W1 reactions. For FFAs, oxidation products were extracted with CH_2Cl_2 and their identities were established after silylation. For LPCs, base hydrolysis was performed to release the oxidized fatty acids prior to silylation. Multiple doublets in the LC-MS data shared the same *m/z*, suggesting that each substrate may have multiple products (Figure 16). As summarized in Table 5 and Figure 17, P450 2W1 catalyzed both hydroxylation and epoxidation in the middle of fatty acid chains.

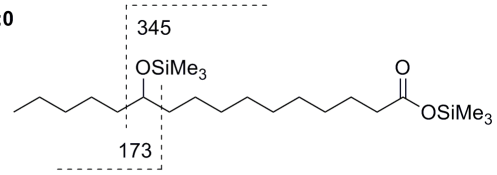
Steady-state kinetic analysis of P450 2W1 reactions

Kinetic studies were performed for 16:0 LPC and 18:1 LPC, as well as for the FFAs C16:0 and C18:1. The kinetic parameters k_{cat} and K_m were estimated based on Michaelis-Menten plots and nonlinear regression analysis (Table 6). The catalytic efficiencies (k_{cat}/K_m) of P450 2W1-catalyzed fatty acid oxidations were comparable with those catalyzed by other human P450 enzymes (*14, 64*). The catalytic efficiency of 18:1 LPC oxidation was ~6-fold greater than that of the FFA C18:1.

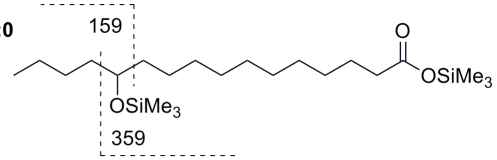
Substrate		[$M^{+\bullet}$]	Major fragment ions (*m/z*)	Product
16:0[1, 2]	Hydroxylation	416	145, 159, 173, 345, 359, 373	11-OH, 12-OH, 13-OH
18:1[1, 2]	Hydroxylation	442	201, 241, 303, 343	8-OH, 11-OH
18:1[1, 2]	Epoxidation	532	215, 317	9, 10-epoxide
16:0 LPC[1]	Hydroxylation	416	215, 229, 243, 275, 289, 303	6-OH, 7-OH, 8-OH
18:0 LPC[1]	Hydroxylation	444	229, 243, 257, 289, 303, 317	7-OH, 8-OH, 9-OH
18:1 LPC[1]	Hydroxylation	442	201, 241, 303, 343	8-OH, 11-OH
18:1 LPC[1]	Epoxidation	532	215, 317	9, 10-epoxide
20:0 LPC[1]	Hydroxylation	472	243, 257, 271, 303, 317, 331	8-OH, 9-OH, 10-OH

Table 5. **EI GC-MS analysis of TMS derivatives of the oxidation products of fatty acids and LPCs by human P450 2W1.**
[1] See Figure 17 for fragmentation patterns.
[2] FFA

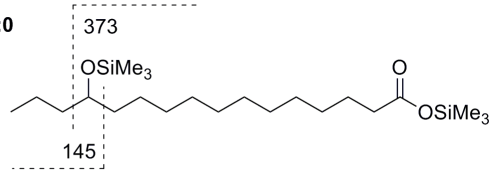
A, 11-OH C16:0



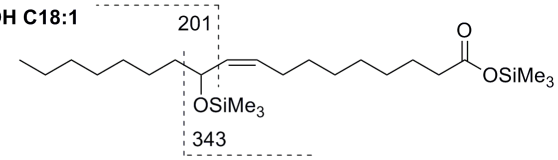
B, 12-OH C16:0



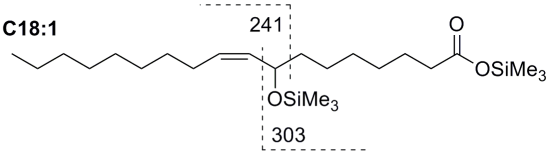
C, 13-OH C16:0



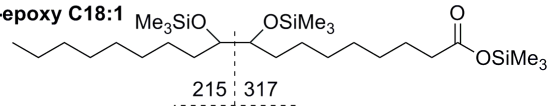
D, 11-OH C18:1



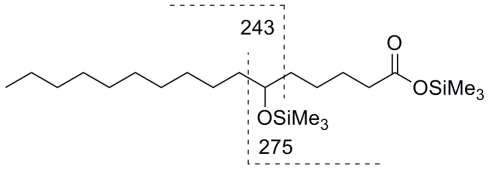
E, 8-OH C18:1



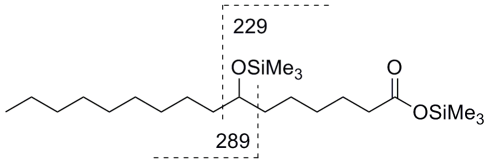
F, 9, 10-epoxy C18:1



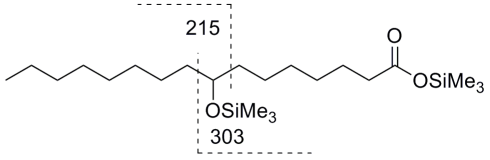
G, 6-OH C16:0



H, 7-OH C16:0



I, 8-OH C16:0



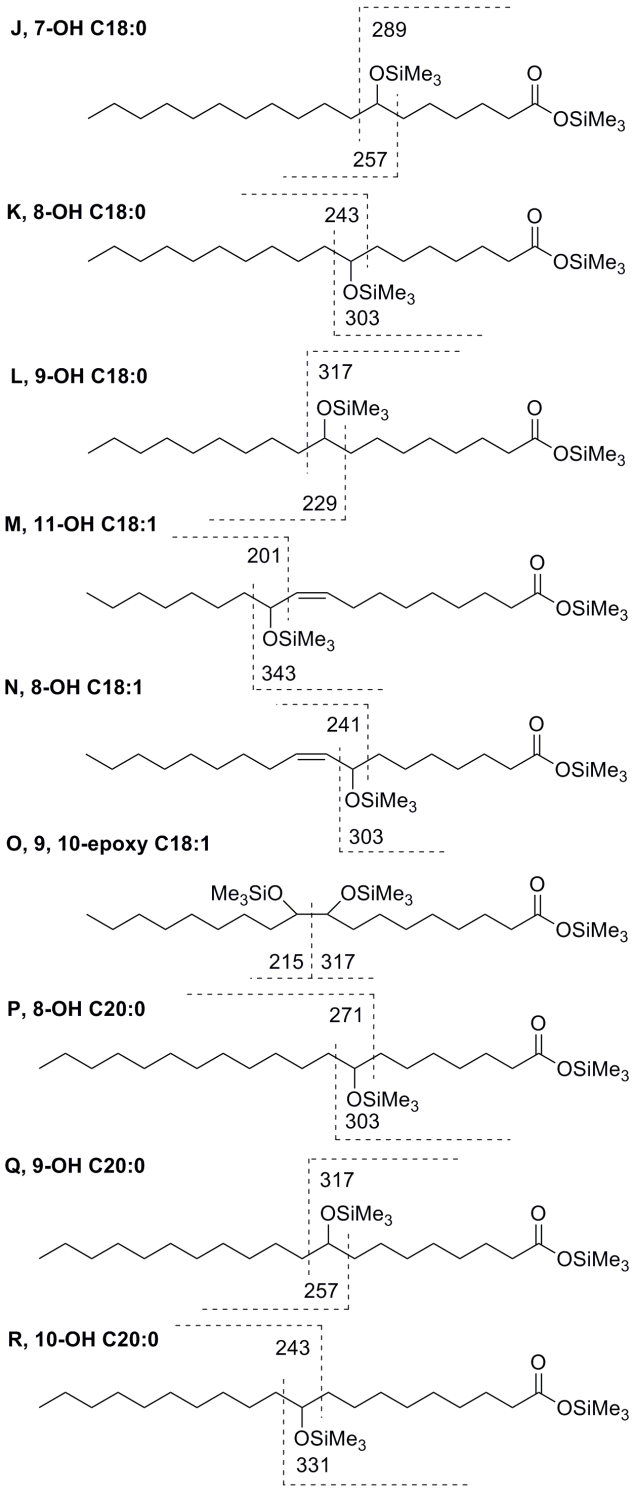


Figure 17. EI GC-MS fragmentation patterns of TMS-derivatized oxidized fatty acid moieties.

Oxidized fatty acid moieties from enzymatic incubations with P450 2W1 with (A-C) C16:0, (D-F) C18:1, (G-I) C16:0; after incubation and base hydrolysis with (J-L) 18:0 LPC, (M-O) 18:1 LPC, and (P-R) 20:0 LPC.

Substrate		k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} mM^{-1})
16:0[1]	Hydroxylation	0.77 ± 0.06	83 ± 17	9.3 ± 0.7
18:1[1]	Hydroxylation	0.63 ± 0.09	101 ± 31	6.2 ± 0.9
18:1[1]	Epoxidation	0.044 ± 0.003	95 ± 13	0.46 ± 0.03
16:0 LPC	Hydroxylation	0.012 ± 0.002	14 ± 5	0.9 ± 0.1
18:1 LPC	Hydroxylation	1.36 ± 0.16	38 ± 12	36 ± 4
18:1 LPC	Epoxidation	0.21 ± 0.02	34 ± 11	6.2 ± 0.6

Table 6. Steady-state kinetics of P450 2W1-catalyzed oxidations
[1] FFA

Isomer specificity of lysophospholipid oxidation

Commercially available LPCs are mixtures of both *sn*-1 and *sn*-2 isomers due to acyl migration, which occurs even at neutral pH (*158*). The inter-conversion can be attenuated at pH 4–5 (*158*), and baseline separation could be achieved using a pH 5 solvent in HPLC (Figure 18). The identities of recovered isomers were confirmed by MS/MS fragmentation patterns, due to the different ratios of fragmentation ions at *m*/*z* 184 and 104 (*157*): the *sn*-1 isomer produces (Figure 19, D, E) more *m/z* 104 daughter ion than the *sn*-2 isomer (Figure 15, B). A mock incubation of pure *sn*-1 and *sn*-2 isomers in a P450 reaction mixture was performed for 10 min, and little inter-conversion between the two isomers was observed. To determine the isomer specificity of lysophospholipid oxidation,

equivalent amounts of *sn*-1 and *sn*-2 isomers were incubated with P450 2W1 and NADPH for 10 min. The reactions were quenched with three volumes of CH_3CN, and the pH was decreased to 5 with one-half volume of 1 M NH_4CO_2H buffer (pH 5). Oxidation of 18:1 LPC was observed only in the incubation with *sn*-1 isomer (Figure 19). The *sn*-1 isomeric nature of the oxidation product, confirmed by MS/MS fragmentation (Figure 19), also indicated that *sn*-1 18:1 LPC is the preferred substrate. We conclude that the *sn*-1 isomers of lysophospholipids are preferred substrates.

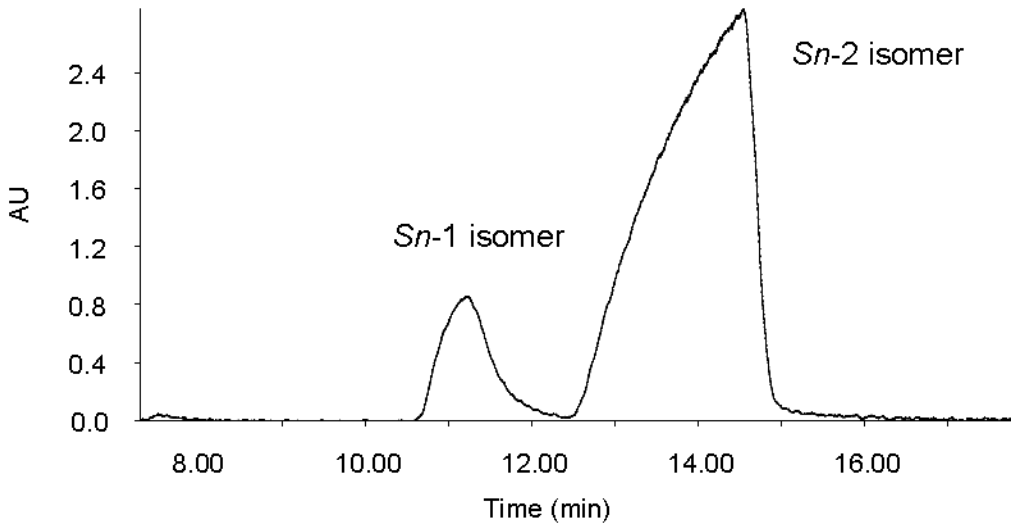


Figure 18. Baseline separation of *sn*-1 and *sn*-2 isomer of 18:1 LPC.

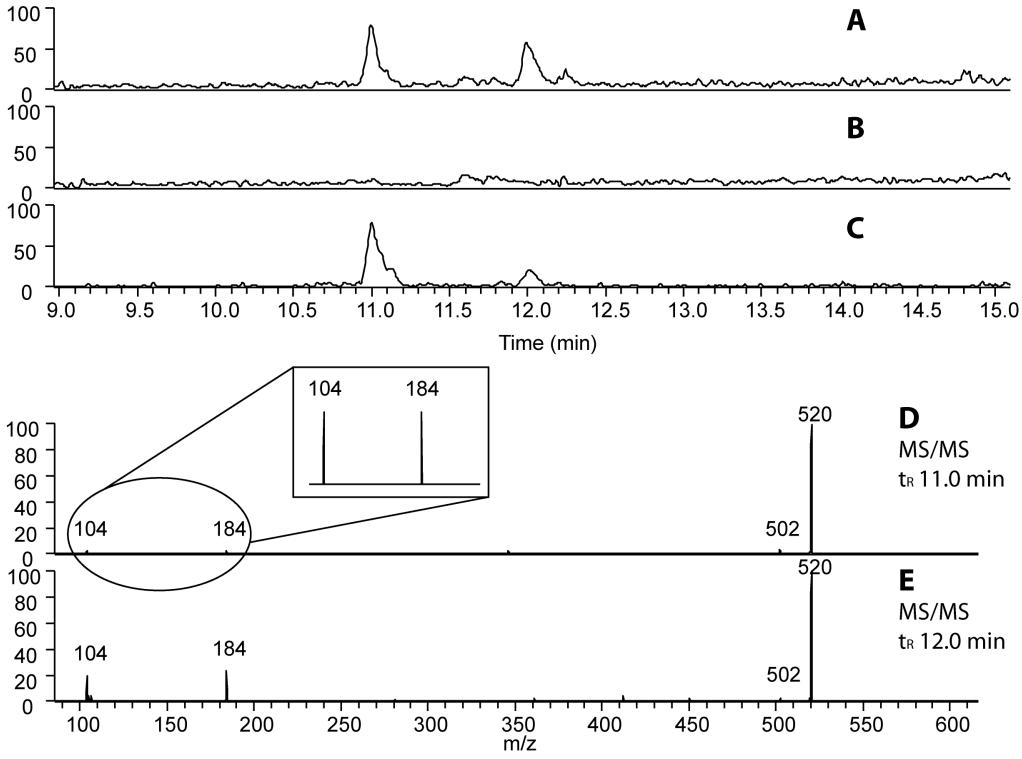


Figure 19. Isomer specificity of lysophospholipid oxidation.
Selected ion chromatograms of 18:1 LPC oxidation products (*m/z* 538) after incubating P450 2W1 with (A) *sn*-1 18:1 LPC, (B) *sn*-2 18:1 LPC, or (C) an equilibrated mixture of *sn*-1 and *sn*-2 18:1 LPC (containing 90% *sn*-1 18:1 LPC and 10% *sn*-2 18:1 LPC). Fragmentation of (D) the peak at t_R 11 min and (E) the peak at t_R 12 min.

Substrate specificity for other phospholipids

Diacylphospholipids, including 16:0 PC, 16:0-18:1 PC, and 16:0-20:4 PC, were incubated with the P450 2W1 reaction mixture and analyzed using the same conditions as for LPCs; no oxidation products were detected. Other classes of lysophospholipids, including 18:1 LPI, 18:1 LPS, 18:1 LPG, 18:1 LPE, and 18:1 LPA, were also confirmed to be substrates for P450 2W1, with similar rates of oxidation (Figure 20).

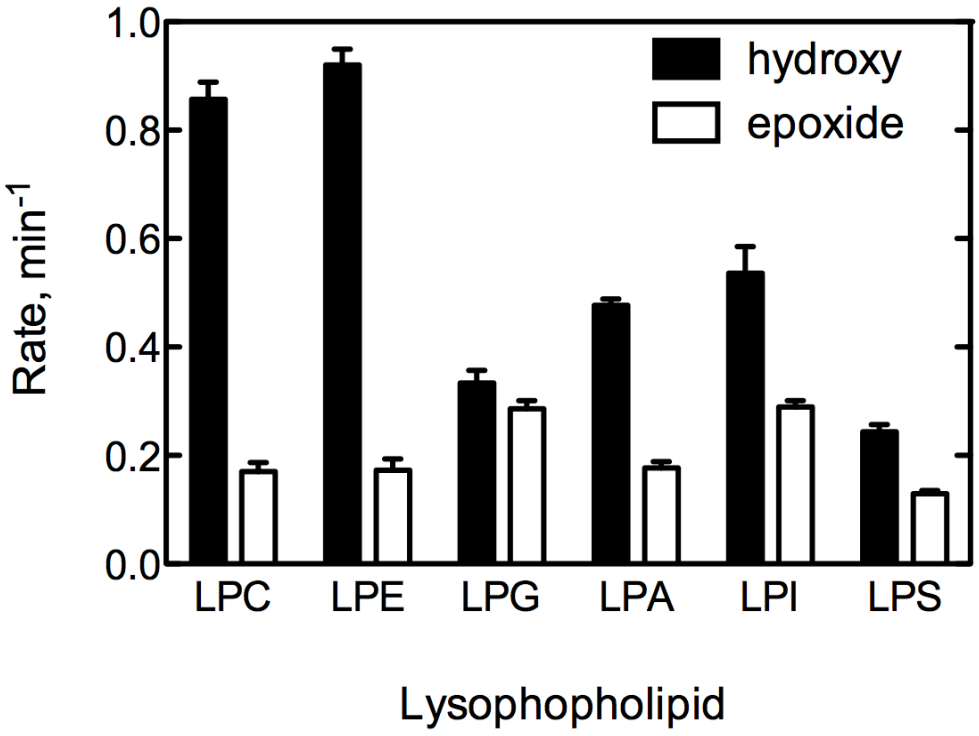


Figure 20. Rates of P450 2W1-catalyzed oxidation with different 18:1 lysophospholipids.

Chiral analysis

P450 2W1-catalyzed epoxidation was investigated with chiral HPLC. The epoxide generated from FFA C18:1 was a mixture of (9*S*,10*R*) and (9*R*,10*S*) in the molar ratio of 1:10 (Figure 21). The epoxide generated from 18:1 LPC was extracted after hydrolysis, and chiral analysis showed that epoxy-18:1 LPC was also a mixture of (9*S*,10*R*) and (9*R*,10*S*) but in the ratio of 1:3 (Figure 21).

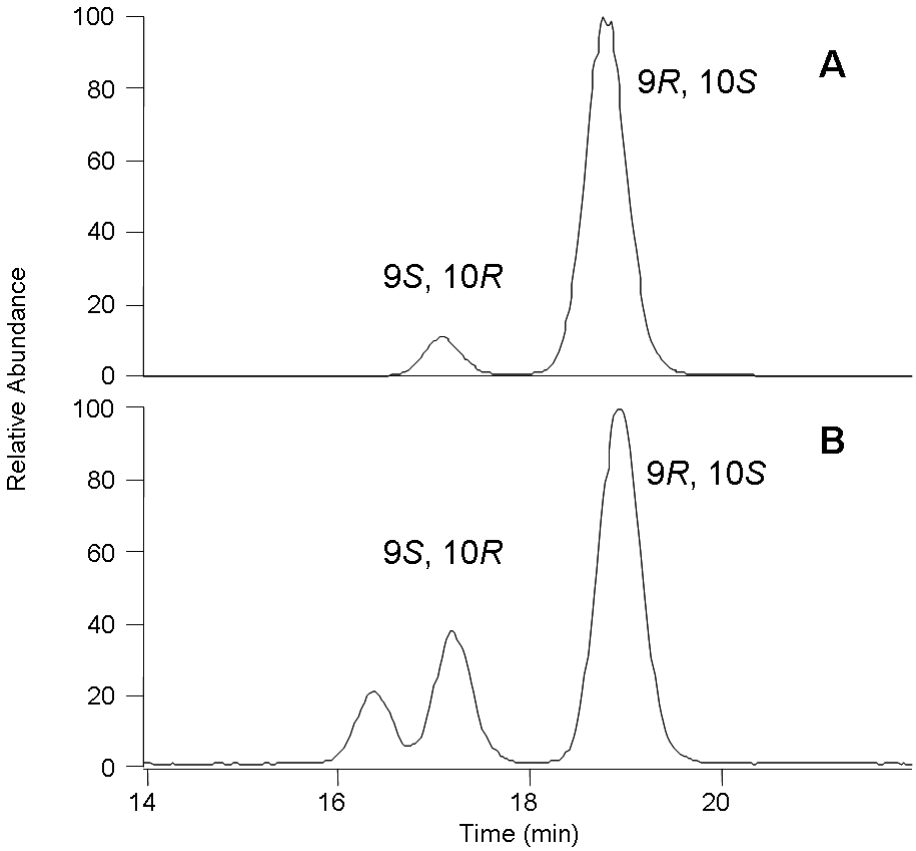


Figure 21. Chiral analysis of P450 2W1-catalyzed epoxidation of FFA C18:1 and 18:1 LPC.
Selected ion chromatograms of P450 2W1-catalyzed epoxidations (*m/z* 297 > 171) of (A) FFA C18:1 (B) 18:1 LPC. The identity of the peak at t_R 16.5 min is unknown.

Discussion

To elucidate the substrate specificities and possible physiological functions of human P450 2W1, 2S1, and 4X1, untargeted substrate searches were performed using XCMS (*76*) and an LC-MS-based metabolomic and isotopic-labeling approach (*14, 68-71, 85*). In addition to FFAs, 18:1 LPC was identified as a substrate for P450 2W1 (Table 4). Other human P450 enzymes tested showed little activity toward 18:1 LPC (Figure 16). In addition to LPCs, a series of other lysophospholipids, including 18:1 LPI, 18:1 LPS, 18:1 LPG, 18:1 LPE, and 18:1 LPA (but not diacylphospholipids), were identified as substrates for P450 2W1 (Figure 20). *sn*-1 18:1 LPC was the preferred substrate over the *sn*-2 isomer (Figure 19), and we conclude that the *sn*-1 isomers of lysophospholipids are the preferred substrates. For the 18:1 epoxidation product, chiral analysis showed enantio-selectivity for formation of (9*R*,10*S*) over (9*S*,10*R*). The position specificities (Figure 21) and the kinetics (Table 6) of P450 2W1-catalyzed oxygenation of lysophospholipids (16:0 LPC and 18:1 LPC) and FFAs (C16:0 and C18:1) were also determined. The epoxidation and hydroxylation for 18:1 LPC were considerably more efficient than for the FFA C18:1. The reaction with an unsaturated acyl LPC (18:1 LPC) was also considerably more efficient than with a saturated acyl LPC (16:0 LPC). No endogenous substrate was identified for human P450 2S1 or 4X1.

Different extraction methods could enrich different classes of compounds. Most substrate and product of P450 enzymes are hydrophobic and can be

extracted by CH_2Cl_2 and Folch reagent. To increase the compound diversity in the tissue extracts, EtOH can be used to include more hydrophilic compounds. However, caution must be taken in that more compounds in tissue extracts could be a burden for analytical instruments, and signals from unrelated compounds could potentially mask the signals from substrates or products of the enzyme. To date, no successful example has been reported for a P450 enzyme for which an additional substrate has been identified with an ethanolic tissue extract.

FFAs are common substrates for a number of human and other P450 enzymes (*36, 39, 42, 64, 159*). Among the five fatty acids identified in our study, only C20:4 has been reported as a substrate for P450 2W1 previously (36). This is the first report that the FFAs C16:0, C16:1, C18:1, and C18:2 are substrates for P450 2W1. P450 2W1 catalyzes both hydroxylation and epoxidation at the middle of fatty acid chains (Table 5). Although some of oxidation products of FFAs have been shown to have interesting physiological functions in vivo (*160*), the fact that many P450 enzymes catalyze FFA oxidations at very slow rates raises doubts about the physiological importance of many of the products (e.g., ω-1, ω-2).

To my knowledge, this is the first report that lysophospholipids are substrates for any P450 enzyme. The most abundant LPCs in plasma are 16:0 LPC, 18:0 LPC, 18:1 LPC, 18:2 LPC, and 20:4 LPC (*161*). All of these, except 18:2 LPC and 20:4 LPC, were confirmed to be substrates for P450 2W1. 18:2 LPC and 20:4 LPC are also likely to be substrates for P450 2W1, but these were not commercially available and, therefore, were not tested. The results of these

kinetic studies suggest that unsaturated acyl LPCs are more efficiently oxidized by P450 2W1 than are saturated acyl LPCs.

Although commercial LPCs are composed of ~90% *sn*-1 and ~10% *sn*-2 isomers, the percentage of each isomer in plasma has been reported to be ~50% (*162*). In plasma, 90% of the unsaturated acyl LPCs were *sn*-1 isomers (*162*), which can be more efficiently oxidized by P450 2W1 compared with saturated acyl LPC.

Lysophospholipids are lipid mediators involved in a vast variety of biological functions (*163*). In particular, LPCs are endogenous proinflammatory lipids that stimulate chemo taxis of T lymphocytes (*164*) and macrophages (*165*). Decreased concentrations of LPCs have been identified in the plasma of colorectal cancer and lung cancer patients (*157, 166*). Thus, decreased LPC levels may be an important contributing factor for tumor development. LPAs are also potent lipid mediators that lead to a plethora of biological actions, including cell proliferation, survival, motility, and invasion, which are critically required for tumor initiation and progression (*167, 168*). 18:1 LPA, one of the substrates of P450 2W1, has been reported to enhance the metastatic potential of human colon cancer cells and to protect them from apoptosis (*169-171*). One aspect of LPA action is its role as a ligand for several cell surface G-protein coupled receptors [e.g., LPA1, LPA2, LPA3, LPA4/GPG23, and LPA5/GPR92 (*172*)]; it is not known whether these receptors are isomer-selective or enantio-selective. LPA enhances cell proliferation by activation of the transcription factor Krüppel-like fractor 5 (KLF5) (*170, 173*). Considering the presence of hydroxyl- and

epoxy-lysophospholipids/phospholipids in vivo (*174, 175*), it is possible that P450 2W1-catalyzed lysophospholipid oxidations are involved in inflammation and tumor development by producing ligands to these receptors and modulate downstream signaling pathways, although further information is not available.

The balance of concentrations of oxidized lysophospholipids is very delicate. They can be synthesized from at least three pathways: P450 2W1-catalyzed lysophospholipid oxidation, deacylation from oxidized phospholipids (*175*), and esterification of oxidized fatty acids to the glycerol derivatives. At the meantime, they can be degraded by several lysophospholipase-based pathways. It is possible that the upregulation of P450 2W1 expression in colorectal cancer tissues disturbs the balance of oxidized lysophospholipids and leads to pathological consequences, but further conjecture about the role of P450 2W1 in cancer is speculative.

In conclusion, we have identified P450 oxidation reactions that selectively occur on lysophospholipid fatty acid chains (but not on diacylphospholipids). They were found to be selectively catalyzed by an orphan human P450, P450 2W1, that had not been clearly shown to have definitive catalytic activities with physiological substrates previously. The physiological functions of these oxidized lysophospholipids, if any, remain to be established. We have also introduced new software for the analysis of isotopic compositions of compounds in MS, which can be used in other metabolomic studies. No endogenous substrate was identified for human P450 2S1 or 4X1.

CHAPTER IV

HETEROLOGOUS EXPRESSION OF HUMAN P450 ENZYMES

Introduction

Historically, human P450 enzymes were isolated from human tissue samples (*52*). However, due to the difficulties in obtaining human samples and in purifying specific P450 enzyme without cross-contamination from others, the establishment of heterologous expression systems has been used as a popular alternative to obtain active P450 enzymes for functional studies.

Among the orphan human P450 enzymes covered in this thesis work, human P450 2S1 and 2W1 have been expressed and purified in *E. coli* expression systems with good yields (*36*). Human P450 4V2 was expressed and purified from insect cells, but the expression level was low (~ 50 nM) (*42*). Human P450 2U1 and 4Z1 have been expressed either in *E. coli* (*37*), insect cells (*39*), or yeast (*40*). However, the expression levels were not satisfactory.

This chapter describes the establishment of heterologous protein expression systems to obtain active human P450 enzymes for future functional studies. Heterologous expressions of P450 46A1 and orphan P450 2U1, 4V2, and 4Z1 have been tested in *E. coli* expression systems. Codon-optimization tremendously increased the expression levels of P450 46A1 and P450 2U1. Baculovirus expression systems were tested for P450 4V2 and 4Z1 after the

failure in *E. coli* expression systems. Expressions of P450 4V2 and 4Z1 in insect cells were achieved with low yields. Co-expressions of P450 4V2/4Z1 with NADPH-P450 reductase were achieved in insect cells, and the isolated membrane fractions can be used for future functional studies.

Experimental procedures

Materials and reagents

Codon-optimized coding sequences for heterologous expressions in *E. coli* systems (i.e., P450 2U1, 4V2, 4Z1, and 46A1) were designed with DNA 2.0 technology (*106*) or IDT (*107*), and synthesized either in-house (*36, 38*) or by Genewiz, NJ) (*36*). Native coding sequences for heterologous expressions in insect cells were obtained from DF/HCC DNA Resource Core (Harvard, Boston). The baculovirus IRES sequence was generously provided by Dr. Tzong-Yan Wu (National Tsing Hua University, Hsinchu, Taiwan, ROC). All other reagents, solvents, and competent cells were obtained from general commercial suppliers.

Heterologous expressions in E. coli *cells*

Expression vectors were constructed with codon-optimized coding sequences and a His_6-tag at C-terminal. Expression trials were conducted as previously described (*36*) at 23 °C or 28 °C with different *E. coli* strains (i.e., DH5α, BL21 DE3, BL21 DE3 pLyss, TOP10, XL1-Blue, XL10-Blue, and C41) with and without the co-expression of different chaperones (i.e., pG-KJE8,

pGro7, pKJE7, pTf16, pACYC). P450 spectra were measured 24, 48, and 72 h after induction.

Heterologous expressions in insect cells

Expression vectors (pFastbac and pFastbac-dual) were constructed with native coding sequences and a His_6-tag at C-terminal. Insect cells (i.e., sf9 and sf21 cells) were maintained in spinner flasks in Sf-900 III SFM media containing 5% fetal bovine serum (v/v). The expression culture was supplemented with 0.3 mM δ-aminolevulinic acid and 0.2 mM ferric citrate 24 h post infection with the recombinant virus. Cells were harvested 48 and 72 h post infection, followed by the isolation of microsomal fraction and collection of P450 spectra with the method described (*42*).

Other methods

Immunoblotting was conducted with the method described (*83*). UV-visible spectra were recorded using an Aminco DW-2a/OLIS spectrophotometer. P450 concentrations were estimated spectrally as previously described (*153*). Methyl viologen (final concentration 1.2 μM) was used in the measurement of human P450 4V2 and 4Z1 expressed in insect cells.

Results

Previous expression trials of P450 4Z1 in E. coli *expression systems*

Dr. Zhongliu Wu and Dr. G. Salamanca-Pinzón tested several N-terminal modifications with a codon-optimized 4Z1 sequence but no satisfactory expression level could be obtained (data not shown) (Table 7).

Constructs	*N-terminal Amino Acid Sequences*
4Z1 native	*MEPSWLQELMAHPFLLLILLCMSLLLFQVIRLYQRRRWMIRALHLFPAPPAH*
Wu #1	*MALLLAVFLQELMAHPFLLLILLCMSLLLFQVIRLYQRRRWMIRALHLFPAPPAH*
Wu #2	*MALLLAVFLMAHPFLLLILLCMSLLLFQVIRLYQRRRWMIRALHLFPAPPAH*
Wu #3	*MAHPFLLLILLCMSLLLFQVIRLYQRRRWMIRALHLFPAPPAH*
Wu #4	*MALLLAVFLRLYQRRRWMIRALHLFPAPPAH*
Wu #5	*MAKKTSSKGKLYQRRRWMIRALHLFPAPPAH*
Wu #6	*MALLLFQVIRLYQRRRWMIRALHLFPAPPAH*
Wu #7	*MARQSFGRGKLPAPPAH*
Wu #8	*MAKKTSSKGKLPAPPAH*
Salamanca #2	*MARQVHSSWNLHPFLLLILLCMSLLLFQVIRLYQRRRWMIRALHLFPAPPAH*
Salamanca #3	*MAKKTSSKGKLHPFLLLILLCMSLLLFQVIRLYQRRRWMIRALHLFPAPPAH*
Salamanca #4	*MAKKTSSKGRLQVIRLYQRRRWMIRALHLFPAPPAH*
Salamanca #5	*MAQVIRLYQRRRWMIRALHLFPAPPAH*
Salamanca #6	*MAKKTSSKGKLPAPPAH*

Table 7. Previously tested N-terminal modification for 4Z1 expression.
Eight were tested by Dr. Zhongliu Wu (Wu # 1-8); five were tested by Dr. G. Salamanca-Pinzón (Salamanca # 2-6). Underlined letters are different N-terminal modifications.

Expression trials of P450 4Z1 with new N-terminal modifications

New N-terminal modifications (Table 8) were designed with the codon-optimized 4Z1 sequence. Expression trials were conducted with different *E. coli* strains (DH5α, DH5α F'IQ, BL21 DE3, BL21 DE3 pLyss, TOP10, XL1-Blue, XL10-Blue, and C41) and co-expression of different chaperones (pG-KJE8, pGro7, pKJE7, pTf16, pACYC) at 23 °C or 28 °C after induction. Reproducible P450 spectra (Figure 22) that indicate an expression level of 300 nM with crude cell culture could be obtained with construct # 4 in DH5α cells 48 h after induction with the co-expression of GroES/EL. However, no P450 spectrum could be observed in the membrane or cytosol fractions (data not shown).

Constructs	N-terminal Amino Acid Sequences
4Z1 native	MEPSWLQELMAHPFLLLILLCMSLLLFQVIRLYQRRRWMIRALHLFPAPPAH
#1	*MALLLAVFL*PAPPAH
#2	*MALLLAVFL*MSLLLFQVIRLYQRRRWMIRALHLFPAPPAH
#3	*MAKKTSSKGKL*RLYQRRRWMIRALHLFPAPPAH
#4	*MA*RLYQRRRWMIRALHLFPAPPAH
#5	*MA*RRRWMIRALHLFPAPPAH
#6	*MAKKT*PAPPAH

Table 8. New N-terminal modifications for heterologous expression of P450 4Z1 in *E. coli*.
Underlined letters are different N-terminal modifications.

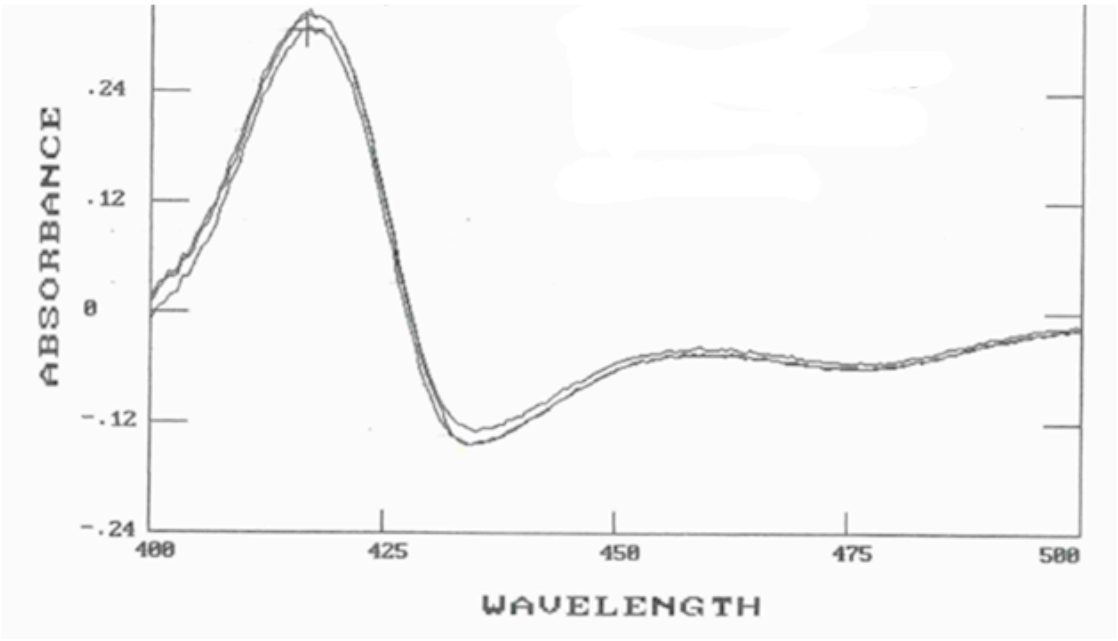


Figure 22. P450 spectrum of expressed human P450 4Z1 from *E. coli* cells. With construct # 4 in DH5α cells 48 h after induction with the co-expression of GroES/EL. The calculated concentration of P450 4Z1 was 300 nM with crude cell culture.

Troubleshooting P450 4Z1 expression in E. coli *expression systems*

Quantitative immunoblotting was performed to quantify the expression of P450 4Z1 protein in the membrane fraction. Although the immunoblotting result indicates a high concentration of P450 4Z1 (~ 5 μM) in the membrane fraction (Figure 23), no P450 spectrum can be observed (data not shown). Thus, two possibilities were speculated: 1) P450 4Z1 polypeptide was translated and was detected by the anti-His_6-tag antibody, but no active protein was generated (possibly due to misfolding or difficulty in heme incorporation); 2) active P450

4Z1 was generated but the catalytic activity was lost quickly during membrane isolation. Since it is technically challenging to address either possibility, a baculovirus expression system was tested to obtain functional P450 4Z1.

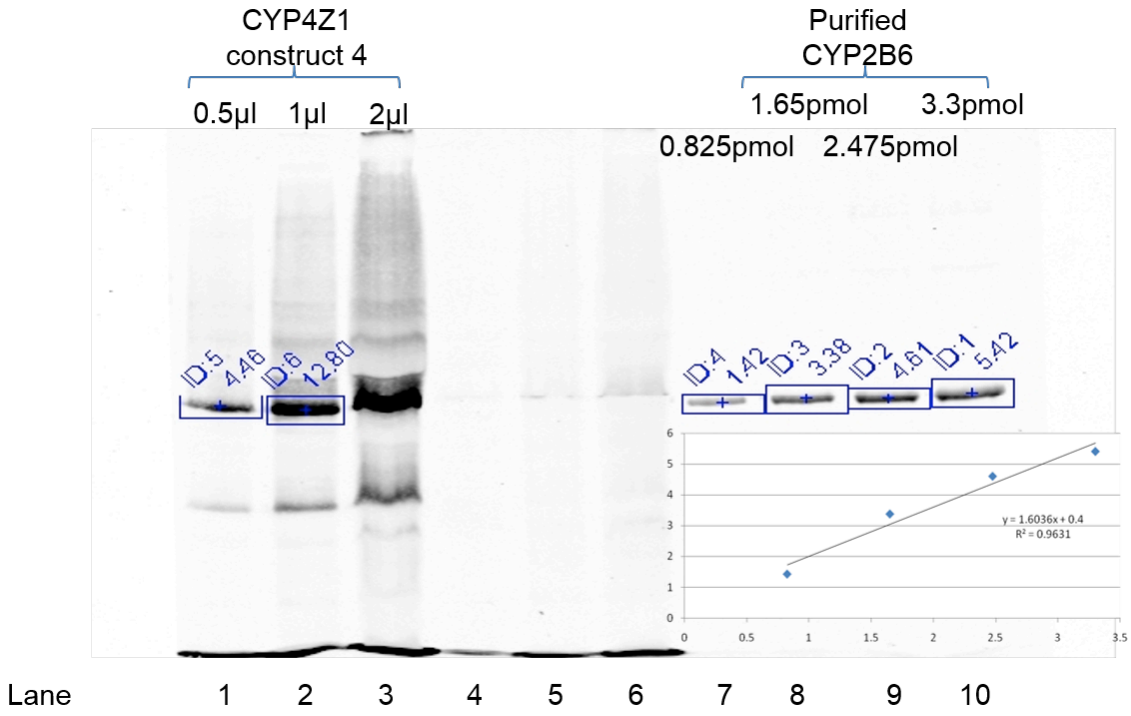


Figure 23. Quantitative immunoblotting to detect P450 4Z1 expression.
Lanes 1, 2, and 3 were loaded with 0.5, 1, and 2 µl of P450 4Z1-expressed membrane fraction. Lanes 7, 8, 9, and 10 were loaded with 0.825 pmol, 1.65 pmol, 2.475 pmol, and 3.3 pmol of a purified P450 2B6 protein. Protein was detected with anti-His_6-tag antibody and LI-COR system. The standard curve was built based on the signal intensity of purified P450 2B6 protein (shown in the bottom right corner).

Expression of P450 4Z1 and 4V2 in insect cells

Mono-cistronic expression vectors for P450 4V2 (previous expression trials failed in *E. coli* expression systems, data not shown) (sequence in Figure 24) and 4Z1 (sequence in Figure 25) have been constructed with native coding sequences. Due to technical difficulties in generating high-titer virus stock in-house, the isolated baculovirus genome was shipped to the Wistar Institute (Philadelphia, PA) to generate high-titer bacoluvirus stock for expression trials. No significant difference in expression levels was observed with different insect cell stains (sf9 versus sf21) (data not shown). The best expression levels (24 nM for P450 4V2; 8 nM for P450 4Z1) were obtained 48 h after infection with a MOI of 5 (Figure 26).

GAATTCATGGCGGGGCTCTGGCTGGGGCTCGTGTGGCAGAAGCTGCTGCT
GTGGGGCGCGGCGAGTGCCCTTTCCCTGGCCGGCGCCAGTCTGGTCCTGA
GCCTGCTGCAGAGGGTGGCGAGCTACGCGCGGAAATGGCAGCAGATGCG
GCCCATCCCCACGGTGGCCCGCGCCTACCCACTGGTGGGCCACGCGCTGC
TGATGAAGCCGGACGGGCGAGAATTTTTTCAGCAGATCATTGAGTACACAG
AGGAATACCGCCACATGCCGCTGCTGAAGCTCTGGGTCGGGCCAGTGCCC
ATGGTGGCCCTTTATAATGCAGAAAATGTGGAGGTAATTTTAACTAGTTCAA
AGCAAATTGACAAATCCTCTATGTACAAGTTTTTAGAACCATGGCTTGGCCT
AGGACTTCTTACAAGTACTGGAAACAAATGGCGCTCCAGGAGAAAGATGTT
AACACCCACTTTCCATTTTACCATTCTGGAAGATTTCTTAGATATCATGAATG
AACAAGCAAATATATTGGTTAAGAAACTTGAAAAACACATTAACCAAGAAGC
ATTTAACTGCTTTTTTTACATCACTCTTTGTGCCTTAGATATCATCTGTGAAAC
AGCTATGGGGAAGAATATTGGTGCTCAAAGTAATGATGATTCCGAGTATGTC
CGTGCAGTTTATAGAATGAGTGAGATGATATTTCGAAGAATAAAGATGCCCT
GGCTTTGGCTTGATCTCTGGTACCTTATGTTTAAAGAAGGATGGGAACACAA
AAAGAGCCTTAAGATCCTACATACTTTTACCAACAGTGTCATCGCGGAACGG
GCCAATGAAATGAACGCCAATGAAGACTGTAGAGGTGATGGCAGGGGCTCT
GCCCCCTCCAAAAATAAACGCAGGGCCTTTCTTGACTTGCTTTTAAGTGTGA
CTGATGACGAAGGGAACAGGCTAAGTCATGAAGATATTCGAGAAGAAGTTG
ACACCTTCATGTTTGAGGGGCACGATACAACTGCAGCTGCAATAAACTGGT
CCTTATACCTGTTGGGTTCTAACCCAGAAGTCCAGAAAAAAGTGGATCATGA
ATTGGATGACGTGTTTGGGAAGTCTGACCGTCCCGCTACAGTAGAAGACCT
GAAGAAACTTCGGTATCTGGAATGTGTTATTAAGGAGACCCTTCGCCTTTTT
CCTTCTGTTCCTTTATTTGCCCGTAGTGTTAGTGAAGATTGTGAAGTGGCAG
GTTACAGAGTTCTAAAAGGCACTGAAGCCGTCATCATTCCCTATGCATTGCA
CAGAGATCCGAGATACTTCCCCAACCCCGAGGAGTTCCAGCCTGAGCGGTT
CTTCCCCGAGAATGCACAAGGGCGCCATCCATATGCCTACGTGCCCTTCTC
TGCTGGCCCCAGGAACTGTATAGGTCAAAAGTTTGCTGTGATGGAAGAAAA
GACCATTCTTTCGTGCATCCTGAGGCACTTTTGGATAGAATCCAACCAGAAA
AGAGAAGAGCTTGGTCTAGAAGGACAGTTGATTCTTCGTCCAAGTAATGGC
ATCTGGATCAAGTTGAAGAGGAGAAATGCAGATGAACGCCATCATCATCATC
ATCATTAA**GTCGAC**

Figure 24. Sequence information of P450 4V2 in pFastbac.
Restriction sites for cloning purpose are shown in **Bold**.

GCGCGCATGGAACCTTCTTGGCTGCAGGAGCTGATGGCACACCCGTTCTTA
CTGCTGATCCTGTTATGTATGAGCCTGCTGTTATTCCAGGTCATTCGTTTATA
TCAGCGCCGTCGCTGGATGATTCGTGCGCTGCATTTATTCCCAGCCCCACC
GGCGCACTGGTTCTACGGCCATAAGGAATTCTACCCGGTGAAGGAGTTTGA
GGTCTATCACAAACTGATGGAGAAATACCCATGTGCGGTTCCGTTATGGGT
CGGTCCATTCACGATGTTCTTCAGCGTTCACGACCCAGATTACGCCAAAATT
CTGTTAAAGCGTCAGGACCCGAAATCTGCGGTGTCTCACAAGATTCTGGAA
AGCTGGGTTGGCCGTGGCCTGGTCACCCTGGACGGTTCTAAATGGAAAAAA
CATCGCCAAATTGTGAAGCCGGGCTTCAACATCAGCATTCTGAAAATCTTCA
TCACCATGATGAGCGAGAGCGTCCGCATGATGCTGAACAAGTGGGAGGAG
CATATTGCGCAGAACAGCCGCCTGGAGCTGTTCCAACACGTCAGCCTGATG
ACGTTAGATAGCATTATGAAATGCGCCTTTAGCCACCAGGGCAGCATCCAA
CTGGACAGCACCTTAGACAGCTACTTAAAGGCCGTGTTTAATCTGAGCAAG
ATTAGCAACCAACGTATGAACAACTTCCTGCACCACAACGACTTAGTGTTCA
AGTTCTCTAGCCAAGGCCAGATTTTTAGCAAGTTTAACCAGGAACTGCATCA
ATTCACCGAAAAGGTCATCCAAGATCGCAAGGAAAGCCTGAAGGACAAGCT
GAAACAGGACACGACGCAAAAGCGCCGCTGGGATTTCCTGGACATTCTGCT
GTCTGCCAAGAGCGAAAACACGAAGGATTTCTCTGAGGCAGATCTCCAGGC
CGAAGTGAAAACCTTCATGTTTGCCGGTCATGATACCACGTCTAGCGCAATT
TCTTGGATCCTGTATTGTCTGGCGAAATACCCGGAACACCAACAGCGTTGC
CGTGATGAAATTCGCGAGTTACTGGGTGACGGCAGCTCTATTACGTGGGAG
CACCTGTCTCAGATGCCATACACGACCATGTGTATTAAGGAATGCCTGCGC
CTGTATGCGCCGGTGGTCAACATCAGCCGCCTGCTGGACAAACCGATCACC
TTTCCAGACGGTCGCAGCCTGCCAGCGGGCATCACCGTCTTTATCAACATT
TGGGCACTGCATCATAACCCATACTTTTGGGAGGACCCGCAAGTGTTTAAC
CCGCTGCGCTTCAGCCGTGAAAATAGCGAGAAGATCCACCCGTATGCGTTC
ATCCCATTTAGCGCAGGTTTACGCAACTGCATTGGCCAGCACTTTGCGATCA
TTGAATGCAAAGTTGCAGTGGCGCTGACCCTGCTGCGTTTCAAGTTAGCGC
CAGATCATAGCCGTCCACCGCAGCCGGTTCGTCAAGTTGTGTTAAAGAGCA
AGAACGGTATCCACGTGTTTGCGAAAAAGGTCTGCCATCACCACCATCATC
ATTAA**GTCGAC**

Figure 25. Sequence information of P450 4Z1 in pFastbac.
Restriction sites for cloning purpose are shown in **Bold**.

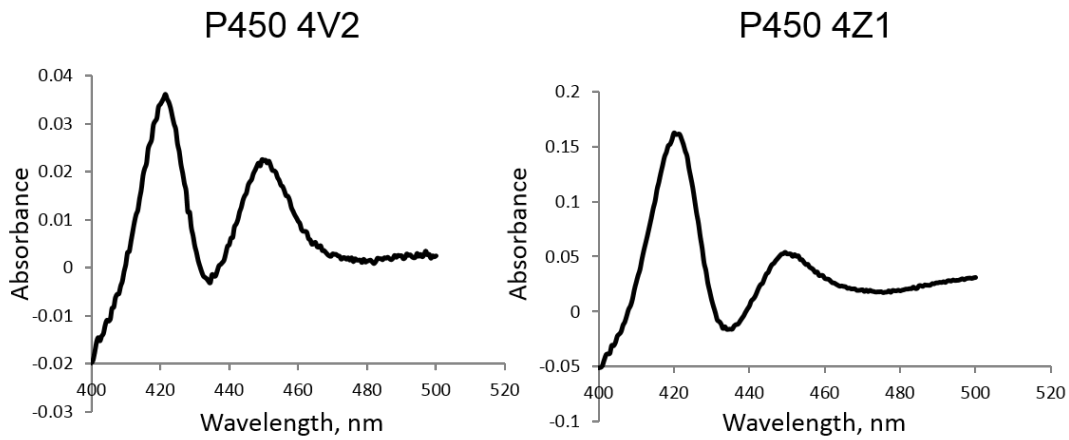


Figure 26. P450 spectra with P450 4V2 and 4Z1 from insect cells.
Methyl viologen was added at a final concentration of 1.2 μM.

It would be difficult to obtain enough purified protein with the mono-cistronic expression systems described above. Therefore, systems that co-express P450 enzyme and NADPH-P450 reductase were tested. If succeeded, the isolated membrane fractions can be used for functional assays.

Several systems can be used to co-express P450 enzyme and NADPH-P450 reductase: 1) infect insect cells with two strains of virus, one expressing a P450 enzyme and the other one expressing NADPH-P450 reductase. However, this method is known to cause a huge reduction in expression levels and thus is rarely used (*115*); 2) infect insect cells with a single virus strain with a dual promoters system, one promoter expressing NADPH-P450 reductase (sequence in Figure 27) and the another one expressing P450 4V2 (sequence in Figure 24) or 4Z1 (sequence in Figure 25); 3) infect insect cells with a bi-cistronic virus strain, in which one single promoter drives the expression of both the P450

enzyme and NADPH-P450 reductase. The coding sequences of two proteins are connected with an IRES so that both proteins can be transcribed and translated (*117*). The IRES from *E. coli* (lg10) and from insect cell virus *Rhopalosiphum padi* (RhPV) were both tested for the co-expression systems (constructed sequences in Figures 28-31). No significant difference in expression levels was observed with different insect cell stains (sf9 versus sf21) (data not shown). For both the dual promoter and bi-cistronic systems, the expression levels of NADPH-P450 reductase were ~ 8 nM. The best expression levels of P450 enzymes (Table 9) were obtained 48 h after infection with a MOI of 5.

In retrospect there is a missense mutation in the bi-cistronic expression vector 4V2-RhPV-IRES-OR (Figure 28).

CTCGAGATGGGAGACTCCCACGTGGACACCAGCTCCACCGTGTCCGAGGC
GGTGGCCGAAGAAGTATCTCTTTTCAGCATGACGGACATGATTCTGTTTTCG
CTCATCGTGGGTCTCCTAACCTACTGGTTCCTCTTCAGAAAGAAAAAAGAAG
AAGTCCCCGAGTTCACCAAAATTCAGACATTGACCTCCTCTGTCAGAGAGA
GCAGCTTTGTGGAAAAGATGAAGAAAACGGGGAGGAACATCATCGTGTTCT
ACGGCTCCCAGACGGGGACTGCAGAGGAGTTTGCCAACCGCCTGTCCAAG
GACGCCCACCGCTACGGGATGCGAGGCATGTCAGCGGACCCTGAGGAGTA
TGACCTGGCCGACCTGAGCAGCCTGCCAGAGATCGACAACGCCCTGGTGG
TTTTCTGCATGGCCACCTACGGTGAGGGAGACCCCACCGACAATGCCCAG
GACTTCTACGACTGGCTGCAGGAGACAGACGTGGATCTCTCTGGGGTCAAG
TTCGCGGTGTTTGGTCTTGGGAACAAGACCTACGAGCACTTCAATGCCATG
GGCAAGTACGTGGACAAGCGGCTGGAGCAGCTCGGCGCCCAGCGCATCTT
TGAGCTGGGGTTGGGCGACGACGATGGGAACTTGGAGGAGGACTTCATCA
CCTGGCGAGAGCAGTTCTGGCCGGCCGTGTGTGAACACTTTGGGGTGGAA
GCCACTGGCGAGGAGTCCAGCATTCGCCAGTACGAGCTTGTGGTCCACAC
CGACATAGATGCGGCCAAGGTGTACATGGGGGAGATGGGCCGGCTGAAGA
GCTACGAGAACCAGAAGCCCCCCTTTGATGCCAAGAATCCGTTCCTGGCTG
CAGTCACCACCAACCGGAAGCTGAACCAGGGAACCGAGCGCCACCTCATG
CACCTGGAATTGGACATCTCGGACTCCAAAATCAGGTATGAATCTGGGGAC
CACGTGGCTGTGTACCCAGCCAACGACTCTGCTCTCGTCAACCAGCTGGGC
AAAATCCTGGGTGCCGACCTGGACGTCGTCATGTCCCTGAACAACCTGGAT
GAGGAGTCCAACAAGAAGCACCCATTCCCGTGCCCTACGTCCTACCGCACG
GCCCTCACCTACTACCTGGACATCACCAACCCGCCGCGTACCAACGTGCTG
TACGAGCTGGCGCAGTACGCCTCGGAGCCCTCGGAGCAGGAGCTGCTGCG
CAAGATGGCCTCCTCCTCCGGCGAGGGCAAGGAGCTGTACCTGAGCTGGG
TGGTGGAGGCCCGGAGGCACATCCTGGCCATCCTGCAGGACTGCCCGTCC
CTGCGGCCCCCCATCGACCACCTGTGTGAGCTGCTGCCGCGCCTGCAGGC
CCGCTACTACTCCATCGCCTCATCCTCCAAGGTCCACCCCAACTCTGTGCA
CATCTGTGCGGTGGTTGTGGAGTACGAGACCAAGGCCGGCCGCATCAACA
AGGGCGTGGCCACCAACTGGCTGCGGGCCAAGGAGCCTGCCGGGGAGAA
CGGCGGCCGTGCGCTGGTGCCCATGTTCGTGCGCAAGTCCCAGTTCCGCC
TGCCCTTCAAGGCCACCACGCCTGTCATCATGGTGGGCCCCGGCACCGGG
GTGGCACCCTTCATAGGCTTCATCCAGGAGCGGGCCTGGCTGCGACAGCA
GGGCAAGGAGGTGGGGGAGACGCTGCTGTACTACGGCTGCCGCCGCTCG
GATGAGGACTACCTGTACCGGGAGGAGCTGGCGCAGTTCCACAGGGACGG
TGCGCTCACCCAGCTCAACGTGGCCTTCTCCCGGGAGCAGTCCCACAAGG
TCTACGTCCAGCACCTGCTAAAGCAAGACCGAGAGCACCTGTGGAAGTTGA
TCGAAGGCGGTGCCCACATCTACGTCTGTGGGGATGCACGGAACATGGCC
AGGGATGTGCAGAACACCTTCTACGACATCGTGGCTGAGCTCGGGGCCAT
GGAGCACGCGCAGGCGGTGGACTACATCAAGAAACTGATGACCAAGGGCC
GCTACTCCCTGGACGTGTGGAGCTAG**GCTAGC**

Figure 27. Sequence information of NADPH-P450 reductase in pFastbac-dual.
Restriction sites for cloning purpose are shown in **Bold**.

GAATTCATGGCGGGGCTCTGGCTGGGGCTCGTGTGGCAGAAGCTGCTGCT
GTGGGGCGCGGCGAGTGCCCTTTCCCTGGCCGGCGCCAGTCTGGTCCTGA
GCCTGCTGCAGAGGGTGGCGAGCTACGCGCGGAAATGGCAGCAGATGCG
GCCCATCCCCACGGTGGCCCGCGCCTACCCACTGGTGGGCCACGCGCTGC
TGATGAAGCCGGACGGGCGAGAATTTTTTCAGCAGATCATTGAGTACACAG
AGGAATACCGCCACATGCCGCTGCTGAAGCTCTGGGTCGGGCCAGTGCCC
ATGGTGGCCCTTTATAATGCAGAAAATGTGGAGGTAATTTTAACTAGTTCAA
AGCAAATTGACAAATCCTCTATGTACAAGTTTTTAGAACCATGGCTTGGCCT
AGGACTTCTTACAAGTACTGGAAACAAATGGCGCTCCAGGAGAAAGATGTT
AACACCCACTTTCCATTTTACCATTCTGAAAGATTTCTTAGATATCATGAATG
AACAAGCAAATATATTGGTTAAGAAACTTGAAAAACACATTAACCAAGAAGC
ATTTAACTGCTTTTTTTACATCACTCTTTGTGCCTTAGATATCATCTGTGAAAC
AGCTATGGGGAAGAATATTGGTGCTCAAAGTAATGATGATTCCGAGTATGTC
CGTGCAGTTTATAGAATGAGTGAGATGATATTTCGAAGAATAAAGATGCCCT
GGCTTTGGCTTGATCTCTGGTACCTTATGTTTAAAGAAGGATGGGAACACAA
AAAGAGCCTTAAGATCCTACATACTTTTACCAACAGTGTCATCGCGGAACGG
GCCAATGAAATGAACGCCAATGAAGACTGTAGAGGTGATGGCAGGGGCTCT
GCCCCCTCCAAAAATAAACGCAGGGCCTTTCTTGACTTGCTTTTAAGTGTGA
CTGATGACGAAGGGAACAGGCTAAGTCATGAAGATATTCGAGAAGAAGTTG
ACACCTTCATGTTTGAGGGGCACGATACAACTGCAGCTGCAATAAACTGGT
CCTTATACCTGTTGGGTTCTAACCCAGAAGTCCAGAAAAAAGTGGATCATGA
ATTGGATGACGTGTTTGGGAAGTCTGACCGTCCCGCTACAGTAGAAGACCT
GAAGAAACTTCGGTATCTGGAATGTGTTATTAAGGAGACCCTTCGCCTTTTT
CCTTCTGTTCCTTTATTTGCCCGTAGTGTTAGTGAAGATTGTGAAGTGGCAG
GTTACAGAGTTCTAAAAGGCACTGAAGCCGTCATCATTCCCTATGCATTGCA
CAGAGATCCGAGATACTTCCCCAACCCCGAGGAGTTCCAGCCTGAGCGGTT
CTTCCCCGAGAATGCACAAGGGCGCCATCCATATGCCTACGTGCCCTTCTC
TGCTGGCCCCAGGAACTGTATAGGTCAAAAGTTTGCTGTGATGGAAGAAAA
GACCATTCTTTCGTGCATCCTGAGGCACTTTTGGATAGAATCCAACCAGAAA
AGAGAAGAGCTTGGTCTAGAAGGACAGTTGATTCTTCGTCCAAGTAATGGC
ATCTGGATCAAGTTGAAGAGGAGAAATGCAGATGAACGCCATCATCATCATC
ATCATTAAGAATTCGATAAAAGAACCTATAATCCCTTCGCACACCGCGTCAC
ACCGCGCTATATGCTGCTCATTAGGAATTACGGCTCCTTTTTTGTGGATACA
ATCTCTTGTATACGATATACTTATTGTTAATTTCATTGACCTTTACGCAATCCT
GCGTAAATGCTGGTATAGGGTGTACTTCGGATTTCCGAGCCTATATTGGTTT
TGAAAGGACCTTTAAGTCCCTACTATACTACATTGTACTAGCGTAGGCCACG
TAGGCCCGTAAGATATTATAACTATTTTATTATATTTTATTCACCCCCCACATT
AATCCCAGTTAAAGCTTTATAACTATAAGTAAGCCGTGCCGAAACGTTAATC
GGTCGCTAGTTGCGTAACAACTGTTAGTTTAATTTTCCAAAATTTATTTTTCA
CAATTTTTAGTTAAGATTTTAGCTTGCCTTAAGCAGTCTTTATATCTTCTGTAT
ATTATTTTAAAGTTTATAGGAGCAAAGTTCGCTTTACTCGCAATAGCTATTTT
ATTTATTTTAGGAATATTATCACCTCGTAATTATTTAATTATAACATTAGCTTTA
TCTATTTATACCCGGGATTGGATCCACCGGTCGCCACCATGGGAGACTCCC
ACGTGGACACCAGCTCCACCGTGTCCGAGGCGGTGGCCGAAGAAGTATCT
CTTTTCAGCATGACGGACATGATTCTGTTTTCGCTCATCGTGGGTCTCCTAA
CCTACTGGTTCCTCTTCAGAAAGAAAAAAGAAGAAGTCCCCGAGTTCACCAA

AATTCAGACATTGACCTCCTCTGTCAGAGAGAGCAGCTTTGTGGAAAAGATG
AAGAAAACGGGGAGGAACATCATCGTGTTCTACGGCTCCCAGACGGGGAC
TGCAGAGGAGTTTGCCAACCGCCTGTCCAAGGACGCCCACCGCTACGGGA
TGCGAGGCATGTCAGCGGACCCTGAGGAGTATGACCTGGCCGACCTGAGC
AGCCTGCCAGAGATCGACAACGCCCTGGTGGTTTTCTGCATGGCCACCTAC
GGTGAGGGAGACCCCACCGACAATGCCCAGGACTTCTACGACTGGCTGCA
GGAGACAGACGTGGATCTCTCTGGGGTCAAGTTCGCGGTGTTTGGTCTTGG
GAACAAGACCTACGAGCACTTCAATGCCATGGGCAAGTACGTGGACAAGCG
GCTGGAGCAGCTCGGCGCCCAGCGCATCTTTGAGCTGGGGTTGGGCGACG
ACGATGGGAACTTGGAGGAGGACTTCATCACCTGGCGAGAGCAGTTCTGG
CCGGCCGTGTGTGAACACTTTGGGGTGGAAGCCACTGGCGAGGAGTCCAG
CATTCGCCAGTACGAGCTTGTGGTCCACACCGACATAGATGCGGCCAAGGT
GTACATGGGGGAGATGGGCCGGCTGAAGAGCTACGAGAACCAGAAGCCCC
CCTTTGATGCCAAGAATCCGTTCCTGGCTGCAGTCACCACCAACCGGAAGC
TGAACCAGGGAACCGAGCGCCACCTCATGCACCTGGAATTGGACATCTCG
GACTCCAAAATCAGGTATGAATCTGGGGACCACGTGGCTGTGTACCCAGCC
AACGACTCTGCTCTCGTCAACCAGCTGGGCAAAATCCTGGGTGCCGACCTG
GACGTCGTCATGTCCCTGAACAACCTGGATGAGGAGTCCAACAAGAAGCAC
CCATTCCCGTGCCCTACGTCCTACCGCACGGCCCTCACCTACTACCTGGAC
ATCACCAACCCGCCGCGTACCAACGTGCTGTACGAGCTGGCGCAGTACGC
CTCGGAGCCCTCGGAGCAGGAGCTGCTGCGCAAGATGGCCTCCTCCTCCG
GCGAGGGCAAGGAGCTGTACCTGAGCTGGGTGGTGGAGGCCCGGAGGCA
CATCCTGGCCATCCTGCAGGACTGCCCGTCCCTGCGGCCCCCCATCGACC
ACCTGTGTGAGCTGCTGCCGCGCCTGCAGGCCCGCTACTACTCCATCGCCT
CATCCTCCAAGGTCCACCCCAACTCTGTGCACATCTGTGCGGTGGTTGTGG
AGTACGAGACCAAGGCCGGCCGCATCAACAAGGGCGTGGCCACCAACTGG
CTGCGGGCCAAGGAGCCTGCCGGGGAGAACGGCGGCCGTGCGCTGGTGC
CCATGTTCGTGCGCAAGTCCCAGTTCCGCCTGCCCTTCAAGGCCACCACGC
CTGTCATCATGGTGGGCCCCGGCACCGGGGTGGCACCCTTCATAGGCTTC
ATCCAGGAGCGGGCCTGGCTGCGACAGCAGGGCAAGGAGGTGGGGGAGA
CGCTGCTGTACTACGGCTGCCGCCGCTCGGATGAGGACTACCTGTACCGG
GAGGAGCTGGCGCAGTTCCACAGGGACGGTGCGCTCACCCAGCTCAACGT
GGCCTTCTCCCGGGAGCAGTCCCACAAGGTCTACGTCCAGCACCTGCTAAA
GCAAGACCGAGAGCACCTGTGGAAGTTGATCGAAGGCGGTGCCCACATCT
ACGTCTGTGGGGATGCACGGAACATGGCCAGGGATGTGCAGAACACCTTC
TACGACATCGTGGCTGAGCTCGGGGCCATGGAGCACGCGCAGGCGGTGG
ACTACATCAAGAAACTGATGACCAAGGGCCGCTACTCCCTGGACGTGTGGA
GCTAG**GTCGAC**

Figure 28. Sequence information of 4V2-RhPV-IRES-OR in pFastbac.
Restriction sites for cloning purpose are shown in **Bold**. The missense mutation was labeled in red. **A** should be replaced by G for correct 4V2 protein sequence.

GAATTCATGGCGGGGCTCTGGCTGGGGCTCGTGTGGCAGAAGCTGCTGCT
GTGGGGCGCGGCGAGTGCCCTTTCCCTGGCCGGCGCCAGTCTGGTCCTGA
GCCTGCTGCAGAGGGTGGCGAGCTACGCGCGGAAATGGCAGCAGATGCG
GCCCATCCCCACGGTGGCCCGCGCCTACCCACTGGTGGGCCACGCGCTGC
TGATGAAGCCGGACGGGCGAGAATTTTTTCAGCAGATCATTGAGTACACAG
AGGAATACCGCCACATGCCGCTGCTGAAGCTCTGGGTCGGGCCAGTGCCC
ATGGTGGCCCTTTATAATGCAGAAAATGTGGAGGTAATTTTAACTAGTTCAA
AGCAAATTGACAAATCCTCTATGTACAAGTTTTTAGAACCATGGCTTGGCCT
AGGACTTCTTACAAGTACTGGAAACAAATGGCGCTCCAGGAGAAAGATGTT
AACACCCACTTTCCATTTTACCATTCTGGAAGATTTCTTAGATATCATGAATG
AACAAGCAAATATATTGGTTAAGAAACTTGAAAAACACATTAACCAAGAAGC
ATTTAACTGCTTTTTTTACATCACTCTTTGTGCCTTAGATATCATCTGTGAAAC
AGCTATGGGGAAGAATATTGGTGCTCAAAGTAATGATGATTCCGAGTATGTC
CGTGCAGTTTATAGAATGAGTGAGATGATATTTCGAAGAATAAAGATGCCCT
GGCTTTGGCTTGATCTCTGGTACCTTATGTTTAAAGAAGGATGGGAACACAA
AAAGAGCCTTAAGATCCTACATACTTTTACCAACAGTGTCATCGCGGAACGG
GCCAATGAAATGAACGCCAATGAAGACTGTAGAGGTGATGGCAGGGGCTCT
GCCCCCTCCAAAAATAAACGCAGGGCCTTTCTTGACTTGCTTTTAAGTGTGA
CTGATGACGAAGGGAACAGGCTAAGTCATGAAGATATTCGAGAAGAAGTTG
ACACCTTCATGTTTGAGGGGCACGATACAACTGCAGCTGCAATAAACTGGT
CCTTATACCTGTTGGGTTCTAACCCAGAAGTCCAGAAAAAAGTGGATCATGA
ATTGGATGACGTGTTTGGGAAGTCTGACCGTCCCGCTACAGTAGAAGACCT
GAAGAAACTTCGGTATCTGGAATGTGTTATTAAGGAGACCCTTCGCCTTTTT
CCTTCTGTTCCTTTATTTGCCCGTAGTGTTAGTGAAGATTGTGAAGTGGCAG
GTTACAGAGTTCTAAAAGGCACTGAAGCCGTCATCATTCCCTATGCATTGCA
CAGAGATCCGAGATACTTCCCCAACCCCGAGGAGTTCCAGCCTGAGCGGTT
CTTCCCCGAGAATGCACAAGGGCGCCATCCATATGCCTACGTGCCCTTCTC
TGCTGGCCCCAGGAACTGTATAGGTCAAAAGTTTGCTGTGATGGAAGAAAA
GACCATTCTTTCGTGCATCCTGAGGCACTTTTGGATAGAATCCAACCAGAAA
AGAGAAGAGCTTGGTCTAGAAGGACAGTTGATTCTTCGTCCAAGTAATGGC
ATCTGGATCAAGTTGAAGAGGAGAAATGCAGATGAACGCCATCATCATCATC
ATCATTAAGCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATAAATATG
GCTGACTCCCACGTGGACACCAGCTCCACCGTGTCCGAGGCGGTGGCCGA
AGAAGTATCTCTTTTCAGCATGACGGACATGATTCTGTTTTCGCTCATCGTG
GGTCTCCTAACCTACTGGTTCCTCTTCAGAAAGAAAAAAGAAGAAGTCCCCG
AGTTCACCAAAATTCAGACATTGACCTCCTCTGTCAGAGAGAGCAGCTTTGT
GGAAAAGATGAAGAAAACGGGGAGGAACATCATCGTGTTCTACGGCTCCCA
GACGGGGACTGCAGAGGAGTTTGCCAACCGCCTGTCCAAGGACGCCCACC
GCTACGGGATGCGAGGCATGTCAGCGGACCCTGAGGAGTATGACCTGGCC
GACCTGAGCAGCCTGCCAGAGATCGACAACGCCCTGGTGGTTTTCTGCATG
GCCACCTACGGTGAGGGAGACCCCACCGACAATGCCCAGGACTTCTACGA
CTGGCTGCAGGAGACAGACGTGGATCTCTCTGGGGTCAAGTTCGCGGTGTT
TGGTCTTGGGAACAAGACCTACGAGCACTTCAATGCCATGGGCAAGTACGT
GGACAAGCGGCTGGAGCAGCTCGGCGCCCAGCGCATCTTTGAGCTGGGGT
TGGGCGACGACGATGGGAACTTGGAGGAGGACTTCATCACCTGGCGAGAG
CAGTTCTGGCCGGCCGTGTGTGAACACTTTGGGGTGGAAGCCACTGGCGA

GGAGTCCAGCATTCGCCAGTACGAGCTTGTGGTCCACACCGACATAGATGC
GGCCAAGGTGTACATGGGGGAGATGGGCCGGCTGAAGAGCTACGAGAACC
AGAAGCCCCCCTTTGATGCCAAGAATCCGTTCCTGGCTGCAGTCACCACCA
ACCGGAAGCTGAACCAGGGAACCGAGCGCCACCTCATGCACCTGGAATTG
GACATCTCGGACTCCAAAATCAGGTATGAATCTGGGGACCACGTGGCTGTG
TACCCAGCCAACGACTCTGCTCTCGTCAACCAGCTGGGCAAAATCCTGGGT
GCCGACCTGGACGTCGTCATGTCCCTGAACAACCTGGATGAGGAGTCCAAC
AAGAAGCACCCATTCCCGTGCCCTACGTCCTACCGCACGGCCCTCACCTAC
TACCTGGACATCACCAACCCGCCGCGTACCAACGTGCTGTACGAGCTGGC
GCAGTACGCCTCGGAGCCCTCGGAGCAGGAGCTGCTGCGCAAGATGGCCT
CCTCCTCCGGCGAGGGCAAGGAGCTGTACCTGAGCTGGGTGGTGGAGGC
CCGGAGGCACATCCTGGCCATCCTGCAGGACTGCCCGTCCCTGCGGCCCC
CCATCGACCACCTGTGTGAGCTGCTGCCGCGCCTGCAGGCCCGCTACTAC
TCCATCGCCTCATCCTCCAAGGTCCACCCCAACTCTGTGCACATCTGTGCG
GTGGTTGTGGAGTACGAGACCAAGGCCGGCCGCATCAACAAGGGCGTGGC
CACCAACTGGCTGCGGGCCAAGGAGCCTGCCGGGGAGAACGGCGGCCGT
GCGCTGGTGCCCATGTTCGTGCGCAAGTCCCAGTTCCGCCTGCCCTTCAAG
GCCACCACGCCTGTCATCATGGTGGGCCCCGGCACCGGGGTGGCACCCTT
CATAGGCTTCATCCAGGAGCGGGCCTGGCTGCGACAGCAGGGCAAGGAGG
TGGGGGAGACGCTGCTGTACTACGGCTGCCGCCGCTCGGATGAGGACTAC
CTGTACCGGGAGGAGCTGGCGCAGTTCCACAGGGACGGTGCGCTCACCCA
GCTCAACGTGGCCTTCTCCCGGGAGCAGTCCCACAAGGTCTACGTCCAGC
ACCTGCTAAAGCAAGACCGAGAGCACCTGTGGAAGTTGATCGAAGGCGGT
GCCCACATCTACGTCTGTGGGGATGCACGGAACATGGCCAGGGATGTGCA
GAACACCTTCTACGACATCGTGGCTGAGCTCGGGGCCATGGAGCACGCGC
AGGCGGTGGACTACATCAAGAAACTGATGACCAAGGGCCGCTACTCCCTG
GACGTGTGGAGCTAGAAGCTT

Figure 29. Sequence information of 4V2-lg10-OR in pFastbac.
Restriction sites for cloning purpose are shown in **Bold**.

GCGCGCATGGAACCTTCTTGGCTGCAGGAGCTGATGGCACACCCGTTCTTA
CTGCTGATCCTGTTATGTATGAGCCTGCTGTTATTCCAGGTCATTCGTTTATA
TCAGCGCCGTCGCTGGATGATTCGTGCGCTGCATTTATTCCCAGCCCCACC
GGCGCACTGGTTCTACGGCCATAAGGAATTCTACCCGGTGAAGGAGTTTGA
GGTCTATCACAAACTGATGGAGAAATACCCATGTGCGGTTCCGTTATGGGT
CGGTCCATTCACGATGTTCTTCAGCGTTCACGACCCAGATTACGCCAAAATT
CTGTTAAAGCGTCAGGACCCGAAATCTGCGGTGTCTCACAAGATTCTGGAA
AGCTGGGTTGGCCGTGGCCTGGTCACCCTGGACGGTTCTAAATGGAAAAAA
CATCGCCAAATTGTGAAGCCGGGCTTCAACATCAGCATTCTGAAAATCTTCA
TCACCATGATGAGCGAGAGCGTCCGCATGATGCTGAACAAGTGGGAGGAG
CATATTGCGCAGAACAGCCGCCTGGAGCTGTTCCAACACGTCAGCCTGATG
ACGTTAGATAGCATTATGAAATGCGCCTTTAGCCACCAGGGCAGCATCCAA
CTGGACAGCACCTTAGACAGCTACTTAAAGGCCGTGTTTAATCTGAGCAAG
ATTAGCAACCAACGTATGAACAACTTCCTGCACCACAACGACTTAGTGTTCA
AGTTCTCTAGCCAAGGCCAGATTTTTAGCAAGTTTAACCAGGAACTGCATCA
ATTCACCGAAAAGGTCATCCAAGATCGCAAGGAAAGCCTGAAGGACAAGCT
GAAACAGGACACGACGCAAAAGCGCCGCTGGGATTTCCTGGACATTCTGCT
GTCTGCCAAGAGCGAAAACACGAAGGATTTCTCTGAGGCAGATCTCCAGGC
CGAAGTGAAAACCTTCATGTTTGCCGGTCATGATACCACGTCTAGCGCAATT
TCTTGGATCCTGTATTGTCTGGCGAAATACCCGGAACACCAACAGCGTTGC
CGTGATGAAATTCGCGAGTTACTGGGTGACGGCAGCTCTATTACGTGGGAG
CACCTGTCTCAGATGCCATACACGACCATGTGTATTAAGGAATGCCTGCGC
CTGTATGCGCCGGTGGTCAACATCAGCCGCCTGCTGGACAAACCGATCACC
TTTCCAGACGGTCGCAGCCTGCCAGCGGGCATCACCGTCTTTATCAACATT
TGGGCACTGCATCATAACCCATACTTTTGGGAGGACCCGCAAGTGTTTAAC
CCGCTGCGCTTCAGCCGTGAAAATAGCGAGAAGATCCACCCGTATGCGTTC
ATCCCATTTAGCGCAGGTTTACGCAACTGCATTGGCCAGCACTTTGCGATCA
TTGAATGCAAAGTTGCAGTGGCGCTGACCCTGCTGCGTTTCAAGTTAGCGC
CAGATCATAGCCGTCCACCGCAGCCGGTTCGTCAAGTTGTGTTAAAGAGCA
AGAACGGTATCCACGTGTTTGCGAAAAAGGTCTGCCATCACCACCATCATC
ATTAACAATTCGATAAAAGAACCTATAATCCCTTCGCACACCGCGTCACACC
GCGCTATATGCTGCTCATTAGGAATTACGGCTCCTTTTTTGTGGATACAATC
TCTTGTATACGATATACTTATTGTTAATTTCATTGACCTTTACGCAATCCTGC
GTAAATGCTGGTATAGGGTGTACTTCGGATTTCCGAGCCTATATTGGTTTTG
AAAGGACCTTTAAGTCCCTACTATACTACATTGTACTAGCGTAGGCCACGTA
GGCCCGTAAGATATTATAACTATTTTATTATATTTTATTCACCCCCCACATTAA
TCCCAGTTAAAGCTTTATAACTATAAGTAAGCCGTGCCGAAACGTTAATCGG
TCGCTAGTTGCGTAACAACTGTTAGTTTAATTTTCCAAAATTTATTTTTCACAA
TTTTTAGTTAAGATTTTAGCTTGCCTTAAGCAGTCTTTATATCTTCTGTATATT
ATTTTAAAGTTTATAGGAGCAAAGTTCGCTTTACTCGCAATAGCTATTTTATTT
ATTTTAGGAATATTATCACCTCGTAATTATTTAATTATAACATTAGCTTTATCT
ATTTATACCCGGGATTGGATCCACCGGTCGCCACCATGGGAGACTCCCACG
TGGACACCAGCTCCACCGTGTCCGAGGCGGTGGCCGAAGAAGTATCTCTTT
TCAGCATGACGGACATGATTCTGTTTTCGCTCATCGTGGGTCTCCTAACCTA
CTGGTTCCTCTTCAGAAAGAAAAAAGAAGAAGTCCCCGAGTTCACCAAAATT
CAGACATTGACCTCCTCTGTCAGAGAGAGCAGCTTTGTGGAAAAGATGAAG

AAAACGGGGAGGAACATCATCGTGTTCTACGGCTCCCAGACGGGGACTGC
AGAGGAGTTTGCCAACCGCCTGTCCAAGGACGCCCACCGCTACGGGATGC
GAGGCATGTCAGCGGACCCTGAGGAGTATGACCTGGCCGACCTGAGCAGC
CTGCCAGAGATCGACAACGCCCTGGTGGTTTTCTGCATGGCCACCTACGGT
GAGGGAGACCCCACCGACAATGCCCAGGACTTCTACGACTGGCTGCAGGA
GACAGACGTGGATCTCTCTGGGGTCAAGTTCGCGGTGTTTGGTCTTGGGAA
CAAGACCTACGAGCACTTCAATGCCATGGGCAAGTACGTGGACAAGCGGCT
GGAGCAGCTCGGCGCCCAGCGCATCTTTGAGCTGGGGTTGGGCGACGAC
GATGGGAACTTGGAGGAGGACTTCATCACCTGGCGAGAGCAGTTCTGGCC
GGCCGTGTGTGAACACTTTGGGGTGGAAGCCACTGGCGAGGAGTCCAGCA
TTCGCCAGTACGAGCTTGTGGTCCACACCGACATAGATGCGGCCAAGGTGT
ACATGGGGGAGATGGGCCGGCTGAAGAGCTACGAGAACCAGAAGCCCCCC
TTTGATGCCAAGAATCCGTTCCTGGCTGCAGTCACCACCAACCGGAAGCTG
AACCAGGGAACCGAGCGCCACCTCATGCACCTGGAATTGGACATCTCGGA
CTCCAAAATCAGGTATGAATCTGGGGACCACGTGGCTGTGTACCCAGCCAA
CGACTCTGCTCTCGTCAACCAGCTGGGCAAAATCCTGGGTGCCGACCTGGA
CGTCGTCATGTCCCTGAACAACCTGGATGAGGAGTCCAACAAGAAGCACCC
ATTCCCGTGCCCTACGTCCTACCGCACGGCCCTCACCTACTACCTGGACAT
CACCAACCCGCCGCGTACCAACGTGCTGTACGAGCTGGCGCAGTACGCCT
CGGAGCCCTCGGAGCAGGAGCTGCTGCGCAAGATGGCCTCCTCCTCCGGC
GAGGGCAAGGAGCTGTACCTGAGCTGGGTGGTGGAGGCCCGGAGGCACA
TCCTGGCCATCCTGCAGGACTGCCCGTCCCTGCGGCCCCCCATCGACCAC
CTGTGTGAGCTGCTGCCGCGCCTGCAGGCCCGCTACTACTCCATCGCCTCA
TCCTCCAAGGTCCACCCCAACTCTGTGCACATCTGTGCGGTGGTTGTGGAG
TACGAGACCAAGGCCGGCCGCATCAACAAGGGCGTGGCCACCAACTGGCT
GCGGGCCAAGGAGCCTGCCGGGGAGAACGGCGGCCGTGCGCTGGTGCCC
ATGTTCGTGCGCAAGTCCCAGTTCCGCCTGCCCTTCAAGGCCACCACGCCT
GTCATCATGGTGGGCCCCGGCACCGGGGTGGCACCCTTCATAGGCTTCAT
CCAGGAGCGGGCCTGGCTGCGACAGCAGGGCAAGGAGGTGGGGGAGACG
CTGCTGTACTACGGCTGCCGCCGCTCGGATGAGGACTACCTGTACCGGGA
GGAGCTGGCGCAGTTCCACAGGGACGGTGCGCTCACCCAGCTCAACGTGG
CCTTCTCCCGGGAGCAGTCCCACAAGGTCTACGTCCAGCACCTGCTAAAGC
AAGACCGAGAGCACCTGTGGAAGTTGATCGAAGGCGGTGCCCACATCTAC
GTCTGTGGGGATGCACGGAACATGGCCAGGGATGTGCAGAACACCTTCTA
CGACATCGTGGCTGAGCTCGGGGCCATGGAGCACGCGCAGGCGGTGGAC
TACATCAAGAAACTGATGACCAAGGGCCGCTACTCCCTGGACGTGTGGAGC
TAG**GTCGAC**

Figure 30. Sequence information of 4Z1-RhPV-IRES-OR in pFastbac.
Restriction sites for cloning purpose are shown in **Bold**.

GCGCGCATGGAACCTTCTTGGCTGCAGGAGCTGATGGCACACCCGTTCTTA
CTGCTGATCCTGTTATGTATGAGCCTGCTGTTATTCCAGGTCATTCGTTTATA
TCAGCGCCGTCGCTGGATGATTCGTGCGCTGCATTTATTCCCAGCCCCACC
GGCGCACTGGTTCTACGGCCATAAGGAATTCTACCCGGTGAAGGAGTTTGA
GGTCTATCACAAACTGATGGAGAAATACCCATGTGCGGTTCCGTTATGGGT
CGGTCCATTCACGATGTTCTTCAGCGTTCACGACCCAGATTACGCCAAAATT
CTGTTAAAGCGTCAGGACCCGAAATCTGCGGTGTCTCACAAGATTCTGGAA
AGCTGGGTTGGCCGTGGCCTGGTCACCCTGGACGGTTCTAAATGGAAAAAA
CATCGCCAAATTGTGAAGCCGGGCTTCAACATCAGCATTCTGAAAATCTTCA
TCACCATGATGAGCGAGAGCGTCCGCATGATGCTGAACAAGTGGGAGGAG
CATATTGCGCAGAACAGCCGCCTGGAGCTGTTCCAACACGTCAGCCTGATG
ACGTTAGATAGCATTATGAAATGCGCCTTTAGCCACCAGGGCAGCATCCAA
CTGGACAGCACCTTAGACAGCTACTTAAAGGCCGTGTTTAATCTGAGCAAG
ATTAGCAACCAACGTATGAACAACTTCCTGCACCACAACGACTTAGTGTTCA
AGTTCTCTAGCCAAGGCCAGATTTTTAGCAAGTTTAACCAGGAACTGCATCA
ATTCACCGAAAAGGTCATCCAAGATCGCAAGGAAAGCCTGAAGGACAAGCT
GAAACAGGACACGACGCAAAAGCGCCGCTGGGATTTCCTGGACATTCTGCT
GTCTGCCAAGAGCGAAAACACGAAGGATTTCTCTGAGGCAGATCTCCAGGC
CGAAGTGAAAACCTTCATGTTTGCCGGTCATGATACCACGTCTAGCGCAATT
TCTTGGATCCTGTATTGTCTGGCGAAATACCCGGAACACCAACAGCGTTGC
CGTGATGAAATTCGCGAGTTACTGGGTGACGGCAGCTCTATTACGTGGGAG
CACCTGTCTCAGATGCCATACACGACCATGTGTATTAAGGAATGCCTGCGC
CTGTATGCGCCGGTGGTCAACATCAGCCGCCTGCTGGACAAACCGATCACC
TTTCCAGACGGTCGCAGCCTGCCAGCGGGCATCACCGTCTTTATCAACATT
TGGGCACTGCATCATAACCCATACTTTTGGGAGGACCCGCAAGTGTTTAAC
CCGCTGCGCTTCAGCCGTGAAAATAGCGAGAAGATCCACCCGTATGCGTTC
ATCCCATTTAGCGCAGGTTTACGCAACTGCATTGGCCAGCACTTTGCGATCA
TTGAATGCAAAGTTGCAGTGGCGCTGACCCTGCTGCGTTTCAAGTTAGCGC
CAGATCATAGCCGTCCACCGCAGCCGGTTCGTCAAGTTGTGTTAAAGAGCA
AGAACGGTATCCACGTGTTTGCGAAAAAGGTCTGCCATCACCACCATCATC
ATTAAGCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATAAATATGGCT
GACTCCCACGTGGACACCAGCTCCACCGTGTCCGAGGCGGTGGCCGAAGA
AGTATCTCTTTTCAGCATGACGGACATGATTCTGTTTTCGCTCATCGTGGGT
CTCCTAACCTACTGGTTCCTCTTCAGAAAGAAAAAAGAAGAAGTCCCCGAGT
TCACCAAAATTCAGACATTGACCTCCTCTGTCAGAGAGAGCAGCTTTGTGGA
AAAGATGAAGAAAACGGGGAGGAACATCATCGTGTTCTACGGCTCCCAGAC
GGGGACTGCAGAGGAGTTTGCCAACCGCCTGTCCAAGGACGCCCACCGCT
ACGGGATGCGAGGCATGTCAGCGGACCCTGAGGAGTATGACCTGGCCGAC
CTGAGCAGCCTGCCAGAGATCGACAACGCCCTGGTGGTTTTCTGCATGGCC
ACCTACGGTGAGGGAGACCCCACCGACAATGCCCAGGACTTCTACGACTG
GCTGCAGGAGACAGACGTGGATCTCTCTGGGGTCAAGTTCGCGGTGTTTG
GTCTTGGGAACAAGACCTACGAGCACTTCAATGCCATGGGCAAGTACGTGG
ACAAGCGGCTGGAGCAGCTCGGCGCCCAGCGCATCTTTGAGCTGGGGTTG
GGCGACGACGATGGGAACTTGGAGGAGGACTTCATCACCTGGCGAGAGCA
GTTCTGGCCGGCCGTGTGTGAACACTTTGGGGTGGAAGCCACTGGCGAGG
AGTCCAGCATTCGCCAGTACGAGCTTGTGGTCCACACCGACATAGATGCGG

CCAAGGTGTACATGGGGGAGATGGGCCGGCTGAAGAGCTACGAGAACCAG
AAGCCCCCCTTTGATGCCAAGAATCCGTTCCTGGCTGCAGTCACCACCAAC
CGGAAGCTGAACCAGGGAACCGAGCGCCACCTCATGCACCTGGAATTGGA
CATCTCGGACTCCAAAATCAGGTATGAATCTGGGGACCACGTGGCTGTGTA
CCCAGCCAACGACTCTGCTCTCGTCAACCAGCTGGGCAAAATCCTGGGTGC
CGACCTGGACGTCGTCATGTCCCTGAACAACCTGGATGAGGAGTCCAACAA
GAAGCACCCATTCCCGTGCCCTACGTCCTACCGCACGGCCCTCACCTACTA
CCTGGACATCACCAACCCGCCGCGTACCAACGTGCTGTACGAGCTGGCGC
AGTACGCCTCGGAGCCCTCGGAGCAGGAGCTGCTGCGCAAGATGGCCTCC
TCCTCCGGCGAGGGCAAGGAGCTGTACCTGAGCTGGGTGGTGGAGGCCC
GGAGGCACATCCTGGCCATCCTGCAGGACTGCCCGTCCCTGCGGCCCCCC
ATCGACCACCTGTGTGAGCTGCTGCCGCGCCTGCAGGCCCGCTACTACTC
CATCGCCTCATCCTCCAAGGTCCACCCCAACTCTGTGCACATCTGTGCGGT
GGTTGTGGAGTACGAGACCAAGGCCGGCCGCATCAACAAGGGCGTGGCCA
CCAACTGGCTGCGGGCCAAGGAGCCTGCCGGGGAGAACGGCGGCCGTGC
GCTGGTGCCCATGTTCGTGCGCAAGTCCCAGTTCCGCCTGCCCTTCAAGGC
CACCACGCCTGTCATCATGGTGGGCCCCGGCACCGGGGTGGCACCCTTCA
TAGGCTTCATCCAGGAGCGGGCCTGGCTGCGACAGCAGGGCAAGGAGGTG
GGGGAGACGCTGCTGTACTACGGCTGCCGCCGCTCGGATGAGGACTACCT
GTACCGGGAGGAGCTGGCGCAGTTCCACAGGGACGGTGCGCTCACCCAG
CTCAACGTGGCCTTCTCCCGGGAGCAGTCCCACAAGGTCTACGTCCAGCAC
CTGCTAAAGCAAGACCGAGAGCACCTGTGGAAGTTGATCGAAGGCGGTGC
CCACATCTACGTCTGTGGGGATGCACGGAACATGGCCAGGGATGTGCAGA
ACACCTTCTACGACATCGTGGCTGAGCTCGGGGCCATGGAGCACGCGCAG
GCGGTGGACTACATCAAGAAACTGATGACCAAGGGCCGCTACTCCCTGGAC
GTGTGGAGCTAG**AAGCTT**

Figure 31. Sequence information of 4Z1-Ig10-OR in pFastbac.
Restriction sites for cloning purpose are shown in **Bold**.

	P450 4V2 (nM)	P450 4Z1 (nM)
mono-cistronic system	24	8
dual promoter system	3.6	1.8
bi-cistronic system (lg10 IRES)	6	4.8
bi-cistronic system (RhPV IRES)	5.4	6

Table 9. Expression levels of P450 4V2 and 4Z1 in insect cells.

Expression of human P450 2U1 in an E. coli *expression system*

Expression of human P450 2U1 has been established in insect cells without pursuit of protein purification (*39*). Expression (180 nM) has also been achieved in an *E. coli* expression system (*37*) with codon-optimized for several amino acids at the N-terminal. To explore the possibility of obtaining active human P450 2U1 in *E. coli* expression systems with a higher yield, an expression vector was constructed with codon-optimized for the entire N-terminal modified coding sequence (*38*). The best expression level (420 nM) was obtained with co-expression of GroES/EL and a N-terminal modification of MAKKTSSKGKLP in TOP10 cells (*38*).

Expression of human 46A1 in an E. coli *expression system*

The expression of the native sequence as described in (*176*) was unsuccessful until codon-optimization was performed with the entire N-terminal

modified coding sequence (*107*). The best yield (600 nM) was obtained with co-expression of GroES/EL in DH5α cells (*107*).

Discussion

To obtain P450 enzymes for future functional studies, heterologous expression systems were established for human P450 46A1 and 2U1 (*E. coli* expression systems) as well as for human P450 4V2 and 4Z1 (baculovirus expression systems). Co-expression of NADPH-P450 reductase and P450 4V2/ 4Z1 was also achieved with bi-cistronic systems in insect cells.

The difficulties in expressing membrane-bound proteins are well recognized. Correct incorporation of heme is another obstacle in P450 expression. Since obtaining active P450 enzymes in high yields is usually the prerequisite for functional studies, many contributing factors have been summarized in a few review articles (*93, 105*). Although most P450 enzymes can be expressed in satisfactory yields with either an *E. coli* expression system or a baculovirus expression system, heterologous expression could still be the bottleneck for many research projects.

Codon-optimization helps protein translation and could increase expression levels significantly in *E. coli* expression systems. Although an increase in protein expression is not guaranteed (i.e., human P450 4Z1), codon-optimization is still recommended as a common practice for the heterologous expression of all P450 enzymes in *E. coli* expression systems. In-house PCR

based gene synthesis is no longer a common practice, as out-sourcing to contract research companies (e.g., Genewiz, Genscript, etc.) is more cost-effective. Although algorithms offered by different vendors (e.g., Genscript, Genewiz, IDT) could differ in their capability to increase protein expression levels, no case has been reported that one algorithm is significantly better than others. Since codon-optimization aids in protein translation, it is plausible to perform codon-optimization for the entire coding sequence. The more codons are optimized, the higher the expression level could be achieved. One good example is the expression of human P450 2U1: N-terminal modified native coding sequence gave an expression level of 50 nM; an optimized N-terminal but not the entire coding sequence gave an expression level of 180 nM; optimizing the entire coding sequence gave an expression level of 420 nM (*37*).

N-Terminal modifications can be designed according to successful examples reported in the literature (*64, 93-95, 97, 98, 100, 103, 109-112*). Although high expression levels have been obtained with many different N-terminal modifications, it is time consuming to test all of these modifications with every new P450 enzyme. Therefore, it is recommended to perform expression trials with two commonly seen modifications, MALLLAVFL and MAKKTSS, plus 3-4 other modifications in parallel.

Co-expression of different chaperones can be tested to increase expression level, but rarely did the use of a chaperone other than GroES/EL gives higher expression level.

Failure in getting any P450 spectrum after testing all of the N-terminal modifications with GroES/EL co-expression could suggest switching from *E. coli* expression systems to other expression systems (e.g., yeast or baculovirus expression systems).

In addition to those contributing factors described above, personal experience suggests that there are other less recognized factors that also affect yields of P450 expression. 1) Aeration: aeration is commonly taken into consideration in large-scale fermentations and is plausible that it also affect yields in small-scale protein expressions commonly used in research laboratories. Aeration was first speculated to effect protein expression levels with the failure to express human P450 2S1. Although the expression method has been established and published (*36*), difficulties have been encountered in obtaining active P450 2S1 with the same method (data not shown). Later trials found that the expression in 50 ml medium (in 250 ml shaking flasks) always gave much more active enzyme than the expression in 500 ml medium (in 2.8 l shaking flasks) (the same result has been observed by Dr. Pavel Souček, personal communication). Another example is the expression of human P450 4F11 (data not shown). The expression at a low shaking speed (120 rpm) gave much more yield then the expression at 200 rpm. 2) Expression temperature: low expression temperature decreases the rate of transcription/translation but could also give more time for protein folding and heme incorporation. The optimal expression temperature has to be determined by trial and error. 3) Bacterial strain: DH5α is the most commonly used one for P450 expression. Many early

publications reported the use of DH5α F'IQ (*36*), but no significant difference in expression levels was found between these two strains. 4) Other unidentified factor(s): although human P450 2S1 was successfully expressed with 50 ml bacterial culture, it is still not clear why the expression in 500 ml culture medium cannot be reproduced. The failure in reproducing protein expression with the same procedure as described in Dr. Zhongliu Wu's notebook (same plasmid, expression temperature, medium, shaking speed, and incubator) may suggest other factors that have not been recognized before. The failure to express human P450 46A1 with the same protocol described in (*176*) also indicates the involvement of that some previously un-recognized factors in successful expression of P450 enzymes.

Unfortunately, not every P450 enzyme can be expressed with an *E. coli* expression system with high yield. Baculovirus expression systems can be used as alternatives since it also has the potential of generating functional P450 enzymes with satisfactory yields. Although insect cell expression systems are becoming more and more user-friendly (i.e., switching from spinner flasks to shaker flasks, the introduction of automated cell counters and serum-free media), it is still not a common practice in many laboratories due to its high cost and technical requirements. Therefore, it is highly recommended to consult with an experienced expert before setting up a baculovirus expression system from scratch.

The detection of P450 4Z1 by immunoblotting in the membrane fraction of 4Z1-expressed *E. coli* cells indicates that the polypeptide was successfully

translated. However, the absence of a P450 spectrum indicates 1) the translated polypeptide never yielded any catalytic activity or 2) the catalytic activity was lost quickly during membrane isolation. Given the fact that P450 spectra can be observed in the membrane fraction collected from 4Z1-expressed insect cells (Figure 26), it is very likely that active P450 4Z1 was never produced in the *E. coli*, and *E. coli* is probably not capable of expressing active P450 4Z1 due to some inherent yet unknown difficulties. Thus, further attempt to express human P450 4Z1 in *E. coli* cells (e.g., other strains, other N-terminal modifications, *etc.*) was not pursued.

In summary, heterologous expression systems for human P450 2U1, 4V2, 4Z1, and 46A1 were established for future functional assays, and contributing factors are summarized to facilitate future protein expression trials.

CHAPTER V

SUMMARY AND CONCLUSIONS

Functional annotation of orphan human P450 2S1, 2W1, and 4X1 by metabolomic approach is described in Chapter II and III. The heterologous expressions of human P450 2U1, 4V2, 4Z1, and 46A1 are described in chapter IV. Functional studies with P450 2S1 indicate its capability to accept electrons from NADPH-P450 reductase (*177*), although no physiological significance has been identified so far. Functional studies with human P450 2W1 revealed a series of lysophospholipids as substrates. No endogenous substrate was identified for human P450 2S1 or 4X1. The heterologous expression of human P450 2U1 led to the identification of endogenous *N*-arachidonoylserotonin as a substrate (*38*). The heterologous expression of P450 46A1 led to the identification of 7-dehydrocholesterol and desmosterol as substrates (*107*). Heterologous expressions for P450 4V2 and 4Z1 in insect cells have been achieved in low yields, and isolated membrane fractions could be used for future functional studies.

The past few years have seen tremendous advances in MS-based metabolomic studies. Advanced instruments with better sensitivity are constantly being introduced to the market, and software systems with better data analysis/visualization capabilities are being developed for public use. As a result, metabolomic studies have been introduced to many fields: the agricultural

industry uses untargeted metabolomics to study flavors (*178, 179*); clinical labs use untargeted metabolomics to identify biomarkers for diagnostic purposes (*180*); pharmaceutical companies use targeted metabolomics to study drug metabolism (*181*).

Despite the exploding number of metabolomic studies found in literature, obtaining comprehensive metabolome coverage is still the bottleneck of many research projects. Breakthrough in this field could greatly benefit modern biomedical studies.

Several MS based methodologies can be used to maximize metabolome coverage: 1) collect data with different instruments (i.e., GC-MS, LC-MS, CE-MS); 2) use multiple ionization methods (i.e., ESI, APCI); 3) use different derivatization methods to increase sensitivity towards certain classes of compounds (e.g., dansylation (*89*)); 4) fractionate metabolome before data collection (i.e., by different extraction methods (*182, 183*)); 5) use different chromatography methods (e.g., longer gradient, reversed-phase, HILIC (*88*)).

A pragmatic approach for metabolomic studies, especially for those involved with large amounts of samples, balances both sample throughput and comprehensive metabolome coverage. One can always increase throughput (e.g., run a faster LC gradient), albeit at the expense of sensitivity and metabolome coverage. Thus, having some knowledge or assumption about the most relevant classes of compounds can tremendously help study designs by allocating recourses in a cost-effective manner.

Most substrates and products of P450 enzymes are non-polar small molecules that can be detected by LC-MS with either ESI or APCI ionization mode and can be well resolved by a reversed-phase column. A dansylation method has also been developed to increase the sensitivity towards unactivated alcohols (*89*), a class of commonly seen P450 products. Thus, our proposed untargeted metabolomic approaches have identified many substrates for orphan human P450 enzymes (*13, 38*).

Two different data analysis approaches were proposed to identify reactions catalyzed by P450 enzymes: 1) identify difference(s) with and without functional enzyme; 2) search for doublet patterns for monooxygenation reactions. LC-MS data deconvolution is the foundation of both approaches. The major disadvantage of the previously reported software system for doublet search, DoGEX, is that it searches for doublets without data deconvolution, which leads to the identification of large amounts of false-positives (i.e., from peak shoulders) that have to be excluded by manual inspections (Table 3).

The doublet search approach was originally proposed to search for a signal intensity ratio M/M+2 close to 0.95 (*14, 85, 184*) or from 0.67 to 1.5 (*13*) in data collected on a Thermo LTQ instrument, which is based on the assumption that little ^{16}O incorporated monooxygenated product can be found in the tissue extract (Figure 32). Although the assumption has been challenged and searches with a wider intensity ratio window is desired, the use of a narrow window continued, mainly to avoid the over-whelming amount of false-positives that could be generated with a wider intensity ratio window. This situation can be

tremendously improved with HRMS (Figure 32): with a Thermo LTQ Orbitrap instrument and a resolving power of 60,000, a much narrower mass difference window (2.001-2.006) can be used to search for doublet patterns and consequently the amount of false-positives can be magnificently decreased. As a result, even if a much wider intensity ratio window (0.01-1.5) is used, the amount of doublet candidates identified is still acceptable (Table 10). Therefore, the chance of identifying a monooxygenation reaction may increase and it is recommended to collect data with the Thermo LTQ Orbitrap instrument for incubations with $^{16}O_2/^{18}O_2$ gas.

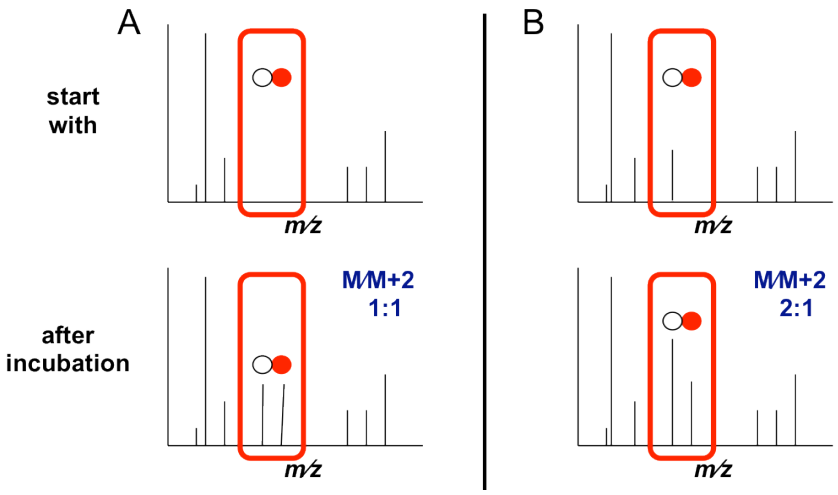


Figure 32. Doublet searches enhanced by HRMS
If there is little monooxygenated product present in the tissue extracts, an intensity ratio about 1:1 can be expected (A); However, if there is already some monooxygenated product in the tissue extracts, the intensity ratio is not necessarily 1:1 (2:1 in this particular case). These doublets can be identified with the help of HRMS (Table 10).

	mass difference window	intensity ratio window	number of candidates identified
LTQ	(1.9, 2.2)	(0.67, 1.5)	~ 200
LTQ-Orbitrap	(2.001, 2.006)	(0.01, 1.5)	~ 200

Table 10. Doublet searches with LTQ and LTQ-Orbitrap instruments.
With the help of HRMS provided by a LTQ Orbitrap instrument, similar amounts of doublet candidates can be identified with a much narrower mass difference window and a much wider intensity ratio window.

Combined with LC-MS data deconvolution and HRMS, it is also possible to make software systems to search for reactions with the addition or loss of certain masses, in an targeted (mass defect (*181*)) or untargeted manner. Untargeted reaction searches would be designed to search for dealkylation reactions, where one can look for two peak features that: 1) with a mass difference of 14.016; 2) decrease of substrate intensity or increase of product intensity after enzymatic incubation; 3) the dealkylated product should be eluted earlier than the substrate due to decreased hydrophobicity (if a reversed-phase column is used). Unfortunately P450-catalyzed dealkylation reactions happen mostly with xenobiotics, and thus the system described above is not applicable in searches for endogenous substrates.

The failure to identify the accumulation of P450 2W1-mediated 18:1 LPC oxidation products by XCMS is very likely due to 1) too vigorous a smoothing method was used 2) the minimal intensity threshold was set too high. Thus, it might be beneficial to revisit and adjust parameters in XCMS. Nevertheless, both

incubation with $^{16}O_2/^{18}O_2$ gas and XCMS analysis are recommended for substrate searches to maximize the chance of identifying an enzymatic reaction.

Even if an enzyme is functionally annotated with a catalytic activity, its physiological significance may still have not been fully revealed. The physiological relevance of the enzymatic activities may be unclear (i.e., fatty acid oxidations by many orphan human P450 enzymes (*39, 40, 42*), lysophospholipids oxidations by P450 2W1 (*13*), *N*-arachidonoylserotonin oxidation by P450 2U1 (*38*)). In addition, other enzymes may be present in the same tissue catalyzing the same reaction. Even for an enzyme whose physiological significance is believed to be fully characterized, new catalytic activities could still be identified afterwards. For example, P450 46A1 was originally known as a cholesterol 24-hydroxylase; however 7-dehydrocholesterol and desmosterol have recently been identified as its substrates as well (*107*). Therefore, any identified catalytic activity does not exclude the possibility that there are other physiological significances yet to be discovered.

The recommended procedure to express a new P450 enzyme of interest begins with codon-optimization of the entire coding sequence for *E. coli* expression systems. Expression trials with several N-terminal modifications are recommended, including MALLLAVFL and MAKKTSS. The best N-terminal modification can be selected in expression trials with cells cultured in 50 ml medium (in 250 ml flasks). The expression of the selected N-terminal can be further optimized by trials with and without GroES/EL co-expression, in both 50 ml medium (in 250 ml flasks) and in 500 ml medium (in 2.8 l flasks). Failure in

obtaining any P450 spectrum at this stage could suggest switching from *E. coli* to other expression systems (e.g., yeast or baculovirus expression systems). If a solid P450 spectrum can be obtained in an *E. coli* expression system, further optimization could be performed (shaking speed, expression temperature, and *E. coli* strains) if needed.

To summarize, I have investigated catalytic activities of orphan human P450 2S1, 2W1, and 4X1. Heterologous expression systems for human P450 2U1, 4V2, 4Z1, and 46A1 have been established for functional studies. LC-MS based metabolomic studies identified a series of lysophospholipids as substrates of human P450 2W1. No endogenous substrate was identified for human P450 2S1 or 4X1. A new software system has been established to search for isotopic patterns in LC-MS data with improved efficiencies.

PUBLICATIONS

Tang, Z., Salamanca-Pinzon, S. G., Wu, Z. L., Xiao, Y., and Guengerich, F. P. (2010) Human cytochrome P450 4F11: heterologous expression in bacteria, purification, and characterization of catalytic function, *Arch Biochem Biophys 494*, 86-93.

Xiao, Y., Shinkyo, R., and Guengerich, F. P. (2011) Cytochrome P450 2S1 is reduced by NADPH-cytochrome P450 reductase, *Drug Metab Dispos 39*, 944-946.

Xiao, Y., and Guengerich, F. P. (2012) Metabolomic analysis and identification of a role for the orphan human cytochrome P450 2W1 in selective oxidation of lysophospholipids, *J Lipid Res 53*, 1610-1617.

Siller, M., Goyal, S., Yoshimoto, F. K., Xiao, Y., Wei, S., and Guengerich, F. P. (2014) Oxidation of endogenous *N*-arachidonoylserotonin by human cytochrome P450 2U1, *J Biol Chem 289*, 10476-10487.

Goyal, S., Xiao, Y., Porter, N. A., Xu, L., and Guengerich, F. P. (2014) Oxidation of 7-dehydrocholesterol and desmosterol by human cytochrome P450 46A1, *J Lipid Res*, in press

REFERENCES

1. Omura, T., and Sato, R. (1962) A new cytochrome in liver microsomes, *J Biol Chem 237*, 1375-1376.

2. Klingenberg, M. (1958) Pigments of rat liver microsomes, *Arch Biochem Biophys 75*, 376-386.

3. Omura, T., and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes. I. evidence for its hemoprotein nature, *J Biol Chem 239*, 2370-2378.

4. Omura, T., and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes. Ii. solubilization, purification, and properties, *J Biol Chem 239*, 2379-2385.

5. Nishida, C. R., Lee, M., and Ortiz de Montellano, P. R. (2010) Efficient hypoxic activation of the anticancer agent AQ4N by CYP2S1 and CYP2W1, *Mol Pharmacol 78*, 497-502.

6. Turpeinen, M., Hofmann, U., Klein, K., Murdter, T., Schwab, M., and Zanger, U. M. (2009) A predominate role of CYP1A2 for the metabolism of nabumetone to the active metabolite, 6-methoxy-2-naphthylacetic acid, in human liver microsomes, *Drug Metab Dispos 37*, 1017-1024.

7. Chowdhury, G., Calcutt, M. W., and Guengerich, F. P. (2010) Oxidation of *N*-nitrosoalkylamines by human cytochrome P450 2A6: sequential oxidation to aldehydes and carboxylic acids and analysis of reaction steps, *J Biol Chem 285*, 8031-8044.

8. Bondon, A., Macdonald, T. L., Harris, T. M., and Guengerich, F. P. (1989) Oxidation of cycloalkylamines by cytochrome P450. Mechanism-based inactivation, adduct formation, ring expansion, and nitrone formation, *J Biol Chem 264*, 1988-1997.

9. Rettie, A. E., Rettenmeier, A. W., Howald, W. N., and Baillie, T. A. (1987) Cytochrome P450--catalyzed formation of delta 4-VPA, a toxic metabolite of valproic acid, *Science 235*, 890-893.

10. Guengerich, F. P. (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity, *Chem Res Toxicol 14*, 611-650.

11. Guengerich, F. P., and Macdonald, T. L. (1984) Chemical mechanisms of catalysis by cytochromes P450: a unified view, *Acc Chem Res 17*, 9-16.

12. Lu, A. Y., and Coon, M. J. (1968) Role of hemoprotein P450 in fatty acid *ω*-hydroxylation in a soluble enzyme system from liver microsomes, *J Biol Chem 243*, 1331-1332.

13. Xiao, Y., and Guengerich, F. P. (2012) Metabolomic analysis and identification of a role for the orphan human cytochrome P450 2W1 in selective oxidation of lysophospholipids, *J Lipid Res 53*, 1610-1617.

14. Tang, Z., Martin, M. V., and Guengerich, F. P. (2009) Elucidation of functions of human cytochrome P450 enzymes: identification of endogenous substrates in tissue extracts using metabolomic and isotopic labeling approaches, *Anal Chem 81*, 3071-3078.

15. Wada, A., Mathew, P. A., Barnes, H. J., Sanders, D., Estabrook, R. W., and Waterman, M. R. (1991) Expression of functional bovine cholesterol side chain cleavage cytochrome P450 ($P450_{scc}$) in *Escherichia coli*, *Arch Biochem Biophys 290*, 376-380.

16. Ryan, K. J. (1958) Conversion of androstenedione to estrone by placental microsomes, *Biochim Biophys Acta 27*, 658-659.

17. Mueller, G. C., and Miller, J. A. (1948) The metabolism of 4-dimethylaminoazobenzene by rat liver homogenates, *J Biol Chem 176*, 535-544.

18. Remmer, H. (1959) The acceleration of evipan oxidation and the demethylation of methylaminopyrine by barbiturates, *Naunyn Schmiedebergs Arch Exp Pathol Pharmakol 237*, 296-307.

19. Ortiz de Montellano, P. R., Ed (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed., Kluwer Academic/Plenum Publishers, New York.

20. Guengerich, F. P., and Cheng, Q. (2011) Orphans in the human cytochrome P450 superfamily: approaches to discovering functions and relevance in pharmacology, *Pharmacol Rev 63*, 684-699.

21. Guengerich, F. P. (2005) Human cytochrome P450 enzymes, In *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R., Ed.) 3rd ed., pp 377-530, Kluwer Academic/Plenum Publishers, New York.

22. Wienkers, L. C., and Heath, T. G. (2005) Predicting *in vivo* drug interactions from *in vitro* drug discovery data, *Nat Rev Drug Discov 4*, 825-833.

23. Williams, J. A., Hyland, R., Jones, B. C., Smith, D. A., Hurst, S., Goosen, T. C., Peterkin, V., Koup, J. R., and Ball, S. E. (2004) Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios, *Drug Metab Dispos 32*, 1201-1208.

24. Chowdhury, G., Shibata, N., Yamazaki, H., and Guengerich, F. P. (2014) Human cytochrome P450 oxidation of 5-hydroxythalidomide and pomalidomide, an amino analogue of thalidomide, *Chem Res Toxicol 27*, 147-156.

25. Wu, S., Moomaw, C. R., Tomer, K. B., Falck, J. R., and Zeldin, D. C. (1996) Molecular cloning and expression of CYP2J2, a human cytochrome P450 arachidonic acid epoxygenase highly expressed in heart, *J Biol Chem 271*, 3460-3468.

26. Travica, S., Pors, K., Loadman, P. M., Shnyder, S. D., Johansson, I., Alandas, M. N., Sheldrake, H. M., Mkrtchian, S., Patterson, L. H., and Ingelman-Sundberg, M. (2013) Colon cancer-specific cytochrome P450 2W1 converts duocarmycin analogues into potent tumor cytotoxins, *Clin Cancer Res 19*, 2952-2961.

27. Wang, K., and Guengerich, F. P. (2012) Bioactivation of fluorinated 2-aryl-benzothiazole antitumor molecules by human cytochrome P450s 1A1 and 2W1 and deactivation by cytochrome P450 2S1, *Chem Res Toxicol 25*, 1740-1751.

28. Edson, K. Z., Prasad, B., Unadkat, J. D., Suhara, Y., Okano, T., Guengerich, F. P., and Rettie, A. E. (2013) Cytochrome P450-dependent catabolism of vitamin K: *ω*-hydroxylation catalyzed by human CYP4F2 and CYP4F11, *Biochemistry 52*, 8276-8285.

29. Gervasini, G., de Murillo, S. G., Ladero, J. M., and Agundez, J. A. (2010) CYP2W1 variant alleles in Caucasians and association of the CYP2W1 G541A (Ala181Thr) polymorphism with increased colorectal cancer risk, *Pharmacogenomics 11*, 919-925.

30. Stenstedt, K., Hallstrom, M., Johansson, I., Ingelman-Sundberg, M., Ragnhammar, P., and Edler, D. (2012) The expression of CYP2W1: a prognostic marker in colon cancer, *Anticancer Res 32*, 3869-3874.

31. Bankovic, J., Stojsic, J., Jovanovic, D., Andjelkovic, T., Milinkovic, V., Ruzdijic, S., and Tanic, N. (2010) Identification of genes associated with non-small-cell lung cancer promotion and progression, *Lung Cancer 67*, 151-159.

32. Nakano, M., Kelly, E. J., Wiek, C., Hanenberg, H., and Rettie, A. E. (2012) CYP4V2 in Bietti's crystalline dystrophy: ocular localization, metabolism of *ω*-3-polyunsaturated fatty acids, and functional deficit of the p.H331P variant, *Mol Pharmacol 82*, 679-686.

33. Pikuleva, I. A., Bjorkhem, I., and Waterman, M. R. (1997) Expression, purification, and enzymatic properties of recombinant human cytochrome P450c27 (CYP27), *Arch Biochem Biophys 343*, 123-130.

34. Wikvall, K. (2001) Cytochrome P450 enzymes in the bioactivation of vitamin D to its hormonal form (review), *Int J Mol Med 7*, 201-209.

35. Guengerich, F. P., Tang, Z., Cheng, Q., and Salamanca-Pinzon, S. G. (2011) Approaches to deorphanization of human and microbial cytochrome P450 enzymes, *Biochim Biophys Acta 1814*, 139-145.

36. Wu, Z. L., Sohl, C. D., Shimada, T., and Guengerich, F. P. (2006) Recombinant enzymes overexpressed in bacteria show broad catalytic specificity of human cytochrome P450 2W1 and limited activity of human cytochrome P450 2S1, *Mol Pharmacol 69*, 2007-2014.

37. Wu, Z. L., Qiao, J., Zhang, Z. G., Guengerich, F. P., Liu, Y., and Pei, X. Q. (2009) Enhanced bacterial expression of several mammalian cytochrome P450s by codon optimization and chaperone coexpression, *Biotechnol Lett 31*, 1589-1593.

38. Siller, M., Goyal, S., Yoshimoto, F. K., Xiao, Y., Wei, S., and Guengerich, F. P. (2014) Oxidation of endogenous *N*-arachidonoylserotonin by human cytochrome P450 2U1, *J Biol Chem 289*, 10476-10487.

39. Chuang, S. S., Helvig, C., Taimi, M., Ramshaw, H. A., Collop, A. H., Amad, M., White, J. A., Petkovich, M., Jones, G., and Korczak, B. (2004) CYP2U1, a novel human thymus- and brain-specific cytochrome P450, catalyzes *ω*- and (*ω*-1)-hydroxylation of fatty acids, *J Biol Chem 279*, 6305-6314.

40. Zöllner, A., Dragan, C. A., Pistorius, D., Muller, R., Bode, H. B., Peters, F. T., Maurer, H. H., and Bureik, M. (2009) Human CYP4Z1 catalyzes the in-chain hydroxylation of lauric acid and myristic acid, *Biol Chem 390*, 313-317.

41. Stark, K., Dostalek, M., and Guengerich, F. P. (2008) Expression and purification of orphan cytochrome P450 4X1 and oxidation of anandamide, *FEBS J 275*, 3706-3717.

42. Nakano, M., Kelly, E. J., and Rettie, A. E. (2009) Expression and characterization of CYP4V2 as a fatty acid *ω*-hydroxylase, *Drug Metab Dispos 37*, 2119-2122.

43. Karlgren, M., Backlund, M., Johansson, I., Oscarson, M., and Ingelman-Sundberg, M. (2004) Characterization and tissue distribution of a novel human cytochrome P450-CYP2U1, *Biochem Biophys Res Commun 315*, 679-685.

44. Downie, D., McFadyen, M. C., Rooney, P. H., Cruickshank, M. E., Parkin, D. E., Miller, I. D., Telfer, C., Melvin, W. T., and Murray, G. I. (2005) Profiling cytochrome P450 expression in ovarian cancer: identification of prognostic markers, *Clin Cancer Res 11*, 7369-7375.

45. Smith, G., Wolf, C. R., Deeni, Y. Y., Dawe, R. S., Evans, A. T., Comrie, M. M., Ferguson, J., and Ibbotson, S. H. (2003) Cutaneous expression of cytochrome P450 CYP2S1: individuality in regulation by therapeutic agents for psoriasis and other skin diseases, *Lancet 361*, 1336-1343.

46. Shrivas, K., Mindaye, S. T., Getie-Kebtie, M., and Alterman, M. A. (2013) Mass spectrometry-based proteomic analysis of human liver cytochrome(s) P450, *Toxicol Appl Pharmacol 267*, 125-136.

47. Saarikoski, S. T., Wikman, H. A., Smith, G., Wolff, C. H., and Husgafvel-Pursiainen, K. (2005) Localization of cytochrome P450 CYP2S1 expression in human tissues by *in situ* hybridization and immunohistochemistry, *J Histochem Cytochem 53*, 549-556.

48. Rylander, T., Neve, E. P., Ingelman-Sundberg, M., and Oscarson, M. (2001) Identification and tissue distribution of the novel human cytochrome P450 2S1 (CYP2S1), *Biochem Biophys Res Commun 281*, 529-535.

49. Rivera, S. P., Saarikoski, S. T., and Hankinson, O. (2002) Identification of a novel dioxin-inducible cytochrome P450, *Mol Pharmacol 61*, 255-259.

50. Savas, U., Hsu, M. H., Griffin, K. J., Bell, D. R., and Johnson, E. F. (2005) Conditional regulation of the human CYP4X1 and CYP4Z1 genes, *Arch Biochem Biophys 436*, 377-385.

51. Li, A., Jiao, X., Munier, F. L., Schorderet, D. F., Yao, W., Iwata, F., Hayakawa, M., Kanai, A., Shy Chen, M., Alan Lewis, R., Heckenlively, J., Weleber, R. G., Traboulsi, E. I., Zhang, Q., Xiao, X., Kaiser-Kupfer, M., Sergeev, Y. V., and Hejtmancik, J. F. (2004) Bietti crystalline corneoretinal dystrophy is caused by mutations in the novel gene CYP4V2, *Am J Hum Genet 74*, 817-826.

52. Wang, P., Mason, P. S., and Guengerich, F. P. (1980) Purification of human liver cytochrome P450 and comparison to the enzyme isolated from rat liver, *Arch Biochem Biophys 199*, 206-219.

53. Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C., and Kraut, J. (1985) The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P450, *J Biol Chem 260*, 16122-16130.

54. Poulos, T. L., Finzel, B. C., and Howard, A. J. (1987) High-resolution crystal structure of cytochrome $P450_{cam}$, *J Mol Biol 195*, 687-700.

55. Williams, P. A., Cosme, J., Sridhar, V., Johnson, E. F., and McRee, D. E. (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity, *Mol Cell 5*, 121-131.

56. Guengerich, F. P., Tang, Z., Salamanca-Pinzon, S. G., and Cheng, Q. (2010) Characterizing proteins of unknown function: orphan cytochrome p450 enzymes as a paradigm, *Mol Interv 10*, 153-163.

57. Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K., Blattner, F. R., Chaudhuri, R. R., Glasner, J. D., Horiuchi, T., Keseler, I. M., Kosuge, T., Mori, H., Perna, N. T., Plunkett, G., 3rd, Rudd, K. E., Serres, M. H., Thomas, G. H., Thomson, N. R., Wishart, D., and Wanner, B. L. (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005, *Nucleic Acids Res 34*, 1-9.

58. Hanson, A. D., Pribat, A., Waller, J. C., and de Crecy-Lagard, V. (2010) 'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list--and how to find it, *Biochem J 425*, 1-11.

59. Wrighton, S. A., Brian, W. R., Sari, M. A., Iwasaki, M., Guengerich, F. P., Raucy, J. L., Molowa, D. T., and Vandenbranden, M. (1990) Studies on the expression and metabolic capabilities of human liver cytochrome P450IIIA5 (HLp3), *Mol Pharmacol 38*, 207-213.

60. Karam, W. G., and Chiang, J. Y. (1992) Polymorphisms of human cholesterol 7*α*-hydroxylase, *Biochem Biophys Res Commun 185*, 588-595.

61. Ridderstrom, M., Zamora, I., Fjellstrom, O., and Andersson, T. B. (2001) Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA, *J Med Chem 44*, 4072-4081.

62. Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies, *Annu Rev Biochem 70*, 209-246.

63. Poulos, T. L., Johnson EF. (2005) Structures of cytochrome P450 enzymes, In *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R., Ed.), pp 87-114, Kluwer Academic/ Plenum Publishers, New York.

64. Tang, Z., Salamanca-Pinzon, S. G., Wu, Z. L., Xiao, Y., and Guengerich, F. P. (2010) Human cytochrome P450 4F11: heterologous expression in bacteria, purification, and characterization of catalytic function, *Arch Biochem Biophys 494*, 86-93.

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

66. Oda, Y., Nakamura, S., Oki, I., Kato, T., and Shinagawa, H. (1985) Evaluation of the new system (*umu*-test) for the detection of environmental mutagens and carcinogens, *Mutat Res 147*, 219-229.

67. Liebler, D. C., and Guengerich, F. P. (2005) Elucidating mechanisms of drug-induced toxicity, *Nat Rev Drug Discov 4*, 410-420.

68. Long, J. Z., Cisar, J. S., Milliken, D., Niessen, S., Wang, C., Trauger, S. A., Siuzdak, G., and Cravatt, B. F. (2011) Metabolomics annotates ABHD3 as a physiologic regulator of medium-chain phospholipids, *Nat Chem Biol 7*, 763-765.

69. Kopp, F., Komatsu, T., Nomura, D. K., Trauger, S. A., Thomas, J. R., Siuzdak, G., Simon, G. M., and Cravatt, B. F. (2010) The glycerophospho metabolome and its influence on amino acid homeostasis revealed by brain metabolomics of GDE1(-/-) mice, *Chem Biol 17*, 831-840.

70. Wikoff, W. R., Nagle, M. A., Kouznetsova, V. L., Tsigelny, I. F., and Nigam, S. K. (2011) Untargeted metabolomics identifies enterobiome metabolites and putative uremic toxins as substrates of organic anion transporter 1 (Oat1), *J Proteome Res 10*, 2842-2851.

71. Saghatelian, A., Trauger, S. A., Want, E. J., Hawkins, E. G., Siuzdak, G., and Cravatt, B. F. (2004) Assignment of endogenous substrates to enzymes by global metabolite profiling, *Biochemistry 43*, 14332-14339.

72. Arnett, D. R., Jennings, J. L., Tabb, D. L., Link, A. J., and Weil, P. A. (2008) A proteomics analysis of yeast Mot1p protein-protein associations: insights into mechanism, *Mol Cell Proteomics 7*, 2090-2106.

73. Yuan, X., Ta, T. C., Lin, M., Evans, J. R., Dong, Y., Bolotin, E., Sherman, M. A., Forman, B. M., and Sladek, F. M. (2009) Identification of an endogenous ligand bound to a native orphan nuclear receptor, *PLoS One 4*, e5609.

74. Vinayavekhin, N., and Saghatelian, A. (2011) Discovery of a protein-metabolite interaction between unsaturated fatty acids and the nuclear receptor Nur77 using a metabolomics approach, *J Am Chem Soc 133*, 17168-17171.

75. Cheng, Q., Lamb, D. C., Kelly, S. L., Lei, L., and Guengerich, F. P. (2010) Cyclization of a cellular dipentaenone by *Streptomyces coelicolor* cytochrome P450 154A1 without oxidation/reduction, *J Am Chem Soc 132*, 15173-15175.

76. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal Chem 78*, 779-787.

77. Katajamaa, M., and Oresic, M. (2005) Processing methods for differential analysis of LC/MS profile data, *BMC Bioinformatics 6*, 179.

78. Katajamaa, M., Miettinen, J., and Oresic, M. (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data, *Bioinformatics 22*, 634-636.

79. Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinformatics 11*, 395.

80. Lommen, A. (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing, *Anal Chem 81*, 3079-3086.

81. Clasquin, M. F., Melamud, E., and Rabinowitz, J. D. (2012) LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine, *Curr Protoc Bioinformatics Chapter 14*, Unit14 11.

82. Parikh, A., Gillam, E. M., and Guengerich, F. P. (1997) Drug metabolism by *Escherichia coli* expressing human cytochromes P450, *Nat Biotechnol 15*, 784-788.

83. Salamanca-Pinzon, S. G., and Guengerich, F. P. (2011) A tricistronic human adrenodoxin reductase-adrenodoxin-cytochrome P450 27A1 vector system for substrate hydroxylation in *Escherichia coli*, *Protein Expr Purif 79*, 231-236.

84. Bansal, S., Leu, A. N., Gonzalez, F. J., Guengerich, F. P., Chowdhury, A. R., Anandatheerthavarada, H. K., and Avadhani, N. G. (2014) Mitochondrial targeting of cytochrome P450 (CYP) 1B1 and its role in polycyclic aromatic hydrocarbon-induced mitochondrial dysfunction, *J Biol Chem 289*, 9936-9951.

85. Sanchez-Ponce, R., and Guengerich, F. P. (2007) Untargeted analysis of mass spectrometry data for elucidation of metabolites and function of enzymes, *Anal Chem 79*, 3355-3362.

86. Zulak, K. G., Weljie, A. M., Vogel, H. J., and Facchini, P. J. (2008) Quantitative ^{1}H NMR metabolomics reveals extensive metabolic reprogramming of primary and secondary metabolism in elicitor-treated opium poppy cell cultures, *BMC Plant Biol 8*, 5.

87. Schmidtke, L. M., Blackman, J. W., Clark, A. C., and Grant-Preece, P. (2013) Wine metabolomics: objective measures of sensory properties of semillon from GC-MS profiles, *J Agric Food Chem 61*, 11957-11967.

88. Ivanisevic, J., Zhu, Z. J., Plate, L., Tautenhahn, R., Chen, S., O'Brien, P. J., Johnson, C. H., Marletta, M. A., Patti, G. J., and Siuzdak, G. (2013) Toward 'omic scale metabolite profiling: a dual separation-mass spectrometry approach for coverage of lipid and central carbon metabolism, *Anal Chem 85*, 6876-6884.

89. Tang, Z., and Guengerich, F. P. (2010) Dansylation of unactivated alcohols for improved mass spectral sensitivity and application to analysis of cytochrome P450 oxidation products in tissue extracts, *Anal Chem 82*, 7706-7712.

90. Johnson, C. H., and Gonzalez, F. J. (2012) Challenges and opportunities of metabolomics, *J Cell Physiol 227*, 2975-2981.

91. Melamud, E., Vastag, L., and Rabinowitz, J. D. (2010) Metabolomic analysis and visualization engine for LC-MS data, *Anal Chem 82*, 9818-9826.

92. Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D., and Wishart, D. S. (2012) MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis, *Nucleic Acids Res 40*, W127-133.

93. Zelasko, S., Palaria, A., and Das, A. (2013) Optimizations to achieve high-level expression of cytochrome P450 proteins using *Escherichia coli* expression systems, *Prot Expr Purif 92*, 77-87.

94. Barnes, H. J., Arlotto, M. P., and Waterman, M. R. (1991) Expression and enzymatic activity of recombinant cytochrome P450 17*α*-hydroxylase in *Escherichia coli*, *Proc Natl Acad Sci U S A 88*, 5597-5601.

95. Guo, Z., Gillam, E. M., Ohmori, S., Tukey, R. H., and Guengerich, F. P. (1994) Expression of modified human cytochrome P450 1A1 in *Escherichia coli*: effects of 5' substitution, stabilization, purification, spectral characterization, and catalytic properties, *Arch Biochem Biophys 312*, 436-446.

96. Sandhu, P., Baba, T., and Guengerich, F. P. (1993) Expression of modified cytochrome P450 2C10 (2C9) in *Escherichia coli*, purification, and reconstitution of catalytic activity, *Arch Biochem Biophys 306*, 443-450.

97. Sandhu, P., Guo, Z., Baba, T., Martin, M. V., Tukey, R. H., and Guengerich, F. P. (1994) Expression of modified human cytochrome P450 1A2 in *Escherichia coli*: stabilization, purification, spectral characterization, and catalytic activities of the enzyme, *Arch Biochem Biophys 309*, 168-177.

98. Scott, E. E., Spatzenegger, M., and Halpert, J. R. (2001) A truncation of 2B subfamily cytochromes P450 yields increased expression levels, increased solubility, and decreased aggregation while retaining function, *Arch Biochem Biophys 395*, 57-68.

99. Wu, Z. L., Bartleson, C. J., Ham, A. J., and Guengerich, F. P. (2006) Heterologous expression, purification, and properties of human cytochrome P450 27C1, *Arch Biochem Biophys 445*, 138-146.

100. Doray, B., Chen, C. D., and Kemper, B. (2001) N-terminal deletions and His-tag fusions dramatically affect expression of cytochrome P450 2C2 in bacteria, *Arch Biochem Biophys 393*, 143-153.

101. Gillam, E. M., Baba, T., Kim, B. R., Ohmori, S., and Guengerich, F. P. (1993) Expression of modified human cytochrome P450 3A4 in *Escherichia coli* and purification and reconstitution of the enzyme, *Arch Biochem Biophys 305*, 123-131.

102. Gillam, E. M., Guo, Z., and Guengerich, F. P. (1994) Expression of modified human cytochrome P450 2E1 in *Escherichia coli*, purification, and spectral and catalytic properties, *Arch Biochem Biophys 312*, 59-66.

103. Sohl, C. D., and Guengerich, F. P. (2010) Kinetic analysis of the three-step steroid aromatase reaction of human cytochrome P450 19A1, *J Biol Chem 285*, 17734-17743.

104. Bauer, S., and Shiloach, J. (1974) Maximal exponential growth rate and yield of *Escherichia coli* obtainable in a bench-scale fermentor, *Biotechnol Bioeng 16*, 933-941.

105. Gillam, E. M. (1998) Human cytochrome P450 enzymes expressed in bacteria: reagents to probe molecular interactions in toxicology, *Clin Exp Pharmacol Physiol 25*, 877-886.

106. Gustafsson, C., Govindarajan, S., and Minshull, J. (2004) Codon bias and heterologous protein expression, *Trends Biotechnol 22*, 346-353.

107. Goyal, S., Xiao, Y., Porter, N. A., Xu, L., and Guengerich, F. P. (2014) Oxidation of 7-dehydrocholesterol and desmosterol by human cytochrome P450 46A1, *J Lipid Res*. In press.

108. Wiedmann, M., Kurzchalia, T. V., Hartmann, E., and Rapoport, T. A. (1987) A signal sequence receptor in the endoplasmic reticulum membrane, *Nature 328*, 830-833.

109. Domanski, T. L., Finta, C., Halpert, J. R., and Zaphiropoulos, P. G. (2001) cDNA cloning and initial characterization of CYP3A43, a novel human cytochrome P450, *Mol Pharmacol 59*, 386-392.

110. Pritchard, M. P., Ossetian, R., Li, D. N., Henderson, C. J., Burchell, B., Wolf, C. R., and Friedberg, T. (1997) A general strategy for the expression of recombinant human cytochrome P450s in *Escherichia coli* using bacterial signal peptides: expression of CYP3A4, CYP2A6, and CYP2E1, *Arch Biochem Biophys 345*, 342-354.

111. Yamasaki, T., Izumi, S., Ide, H., and Ohyama, Y. (2004) Identification of a novel rat microsomal vitamin D3 25-hydroxylase, *J Biol Chem 279*, 22848-22856.

112. Li, Y. C., and Chiang, J. Y. (1991) The expression of a catalytically active cholesterol 7 *α*-hydroxylase cytochrome P450 in *Escherichia coli*, *J Biol Chem 266*, 19186-19191.

113. Richardson, T. H., Jung, F., Griffin, K. J., Wester, M., Raucy, J. L., Kemper, B., Bornheim, L. M., Hassett, C., Omiecinski, C. J., and Johnson, E. F. (1995) A universal approach to the expression of human and rabbit cytochrome P450s of the 2C subfamily in *Escherichia coli*, *Arch Biochem Biophys 323*, 87-96.

114. Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1994) Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding, *Science 265*, 659-666.

115. Berger, I., Fitzgerald, D. J., and Richmond, T. J. (2004) Baculovirus expression system for heterologous multiprotein complexes, *Nat Biotechnol 22*, 1583-1587.

116. Wu, T. Y., Chen, Y. J., Teng, C. Y., Chen, W. S., and Villaflores, O. (2012) A bi-cistronic baculovirus expression vector for improved recombinant protein production, *Bioeng Bugs 3*, 129-132.

117. Chen, W. S., Villaflores, O. B., Jinn, T. R., Chan, M. T., Chang, Y. C., and Wu, T. Y. (2011) Expression of recombinant human interferon-gamma with antiviral activity in the bi-cistronic baculovirus-insect/larval system, *Biosci Biotechnol Biochem 75*, 1342-1348.

118. Karlgren, M., Gomez, A., Stark, K., Svard, J., Rodriguez-Antona, C., Oliw, E., Bernal, M. L., Ramon y Cajal, S., Johansson, I., and Ingelman-Sundberg, M. (2006) Tumor-specific expression of the novel cytochrome P450 enzyme, CYP2W1, *Biochem Biophys Res Commun 341*, 451-458.

119. Gomez, A., Nekvindova, J., Travica, S., Lee, M. Y., Johansson, I., Edler, D., Mkrtchian, S., and Ingelman-Sundberg, M. (2010) Colorectal cancer-specific cytochrome P450 2W1: intracellular localization, glycosylation, and catalytic activity, *Mol Pharmacol 78*, 1004-1011.

120. Gomez, A., Karlgren, M., Edler, D., Bernal, M. L., Mkrtchian, S., and Ingelman-Sundberg, M. (2007) Expression of CYP2W1 in colon tumors: regulation by gene methylation, *Pharmacogenomics 8*, 1315-1325.

121. Renaud, H. J., Cui, J. Y., Khan, M., and Klaassen, C. D. (2011) Tissue distribution and gender-divergent expression of 78 cytochrome P450 mRNAs in mice, *Toxicol Sci 124*, 261-277.

122. Edler, D., Stenstedt, K., Ohrling, K., Hallstrom, M., Karlgren, M., Ingelman-Sundberg, M., and Ragnhammar, P. (2009) The expression of the novel CYP2W1 enzyme is an independent prognostic factor in colorectal cancer - a pilot study, *Eur J Cancer 45*, 705-712.

123. Yoshioka, H., Kasai, N., Ikushiro, S., Shinkyo, R., Kamakura, M., Ohta, M., Inouye, K., and Sakaki, T. (2006) Enzymatic properties of human CYP2W1 expressed in *Escherichia coli*, *Biochem Biophys Res Commun 345*, 169-174.

124. Marek, C. J., Tucker, S. J., Koruth, M., Wallace, K., and Wright, M. C. (2007) Expression of CYP2S1 in human hepatic stellate cells, *FEBS Lett 581*, 781-786.

125. Rivera, S. P., Wang, F., Saarikoski, S. T., Taylor, R. T., Chapman, B., Zhang, R., and Hankinson, O. (2007) A novel promoter element containing multiple overlapping xenobiotic and hypoxia response elements mediates induction of cytochrome P4502S1 by both dioxin and hypoxia, *J Biol Chem 282*, 10881-10893.

126. Rieger, M. A., Ebner, R., Bell, D. R., Kiessling, A., Rohayem, J., Schmitz, M., Temme, A., Rieber, E. P., and Weigle, B. (2004) Identification of a novel mammary-restricted cytochrome P450, CYP4Z1, with overexpression in breast carcinoma, *Cancer Res 64*, 2357-2364.

127. Yu, W., Chai, H., Li, Y., Zhao, H., Xie, X., Zheng, H., Wang, C., Wang, X., Yang, G., Cai, X., Falck, J. R., and Yang, J. (2012) Increased expression of CYP4Z1 promotes tumor angiogenesis and growth in human breast cancer, *Toxicol Appl Pharmacol 264*, 73-83.

128. Tradonsky, A., Rubin, T., Beck, R., Ring, B., Seitz, R., and Mair, S. (2012) A search for reliable molecular markers of prognosis in prostate cancer: a study of 240 cases, *Am J Clin Pathol 137*, 918-930.

129. Bui, P. H., and Hankinson, O. (2009) Functional characterization of human cytochrome P450 2S1 using a synthetic gene-expressed protein in *Escherichia coli*, *Mol Pharmacol 76*, 1031-1043.

130. Bui, P. H., Hsu, E. L., and Hankinson, O. (2009) Fatty acid hydroperoxides support cytochrome P450 2S1-mediated bioactivation of benzo[*a*]pyrene-7,8-dihydrodiol, *Mol Pharmacol 76*, 1044-1052.

131. Bui, P., Imaizumi, S., Beedanagari, S. R., Reddy, S. T., and Hankinson, O. (2011) Human CYP2S1 metabolizes cyclooxygenase- and lipoxygenase-derived eicosanoids, *Drug Metab Dispos 39*, 180-190.

132. Hanna, I. H., Teiber, J. F., Kokones, K. L., and Hollenberg, P. F. (1998) Role of the alanine at position 363 of cytochrome P450 2B2 in influencing the NADPH- and hydroperoxide-supported activities, *Arch Biochem Biophys 350*, 324-332.

133. Guengerich, F. P., Krauser, J. A., and Johnson, W. W. (2004) Rate-limiting steps in oxidations catalyzed by rabbit cytochrome P450 1A2, *Biochemistry 43*, 10775-10788.

134. Patil, P. V., and Ballou, D. P. (2000) The use of protocatechuate dioxygenase for maintaining anaerobic conditions in biochemical experiments, *Anal Biochem 286*, 187-192.

135. Swaine, D. J., Loadman, P. M., Bibby, M. C., Graham, M. A., and Patterson, L. H. (2000) High-performance liquid chromatographic analysis of AQ4N, an alkylaminoanthraquinone *N*-oxide, *J Chromatogr B Biomed Sci Appl 742*, 239-245.

136. Gigon, P. L., Gram, T. E., and Gillette, J. R. (1969) Studies on the rate of reduction of hepatic microsomal cytochrome P450 by reduced nicotinamide adenine dinucleotide phosphate: effect of drug substrates, *Mol Pharmacol 5*, 109-122.

137. Guengerich, F. P., and Johnson, W. W. (1997) Kinetics of ferric cytochrome P450 reduction by NADPH-cytochrome P450 reductase: rapid reduction in the absence of substrate and variations among cytochrome P450 systems, *Biochemistry 36*, 14741-14750.

138. Guengerich, F. P., Ballou, D. P., and Coon, M. J. (1976) Spectral intermediates in the reaction of oxygen with purified liver microsomal cytochrome P450, *Biochem Biophys Res Commun 70*, 951-956.

139. Mansuy, D., Bartoli, J. F., and Momenteau, M. (1982) Alkane hydroxylation catalyzed by metalloporhyrins : evidence for different active oxygen species with alkylhydroperoxides and iodosobenzene as oxidants, *Tetrahedron Lett 23*, 2781-2784.

140. Ortiz de Montellano, P. R. (1995) Oxygen activation and reactivity, In *Cytochrome P450: Structure, Mechanism, and Biochemistry* 2nd ed., Plenum Press, New York.

141. Weiss, R. H., Arnold, J. L., and Estabrook, R. W. (1987) Transformation of an arachidonic acid hydroperoxide into epoxyhydroxy and trihydroxy fatty acids by liver microsomal cytochrome P450, *Arch Biochem Biophys 252*, 334-338.

142. Hecker, M., and Ullrich, V. (1989) On the mechanism of prostacyclin and thromboxane A2 biosynthesis, *J Biol Chem 264*, 141-150.

143. Chang, M. S., Boeglin, W. E., Guengerich, F. P., and Brash, A. R. (1996) Cytochrome P450-dependent transformations of 15*R*- and 15*S*-hydroperoxyeicosatetraenoic acids: stereoselective formation of epoxy alcohol products, *Biochemistry 35*, 464-471.

144. Hughes, T. R., Robinson, M. D., Mitsakakis, N., and Johnston, M. (2004) The promise of functional genomics: completing the encyclopedia of a cell, *Curr Opin Microbiol 7*, 546-554.

145. Carpenter, A. E., and Sabatini, D. M. (2004) Systematic genome-wide screens of gene function, *Nat Rev Genet 5*, 11-22.

146. Yun, C. H., Kim, K. H., Calcutt, M. W., and Guengerich, F. P. (2005) Kinetic analysis of oxidation of coumarins by human cytochrome P450 2A6, *J Biol Chem 280*, 12279-12291.

147. Schoch, G. A., Yano, J. K., Sansen, S., Dansette, P. M., Stout, C. D., and Johnson, E. F. (2008) Determinants of cytochrome P450 2C8 substrate binding: structures of complexes with montelukast, troglitazone, felodipine, and 9-cis-retinoic acid, *J Biol Chem 283*, 17227-17237.

148. Schoch, G. A., Yano, J. K., Wester, M. R., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) Structure of human microsomal cytochrome P450 2C8: evidence for a peripheral fatty acid binding site, *J Biol Chem 279*, 9497-9503.

149. Gillam, E. M., Guo, Z., Martin, M. V., Jenkins, C. M., and Guengerich, F. P. (1995) Expression of cytochrome P450 2D6 in *Escherichia coli*, purification, and spectral and catalytic characterization, *Arch Biochem Biophys 319*, 540-550.

150. Shinkyo, R., and Guengerich, F. P. (2011) Cytochrome P450 7A1 cholesterol 7*α*-hydroxylation: individual reaction steps in the catalytic cycle and rate-limiting ferric iron reduction, *J Biol Chem 286*, 4632-4643.

151. Cuttle, L., Munns, A. J., Hogg, N. A., Scott, J. R., Hooper, W. D., Dickinson, R. G., and Gillam, E. M. (2000) Phenytoin metabolism by human cytochrome P450: involvement of P450 3A and 2C forms in secondary metabolism and drug-protein adduct formation, *Drug Metab Dispos 28*, 945-950.

152. Liu, W., Morrow, J. D., and Yin, H. (2009) Quantification of F2-isoprostanes as a reliable index of oxidative stress in vivo using gas chromatography–mass spectrometry (GC-MS) method, *Free Radic Biol Med 47*, 1101-1107.

153. Rouser, G., Fkeischer, S., and Yamamoto, A. (1970) Two dimensional then layer chromatographic separation of polar lipids and determination of phospholipids by phosphorus analysis of spots, *Lipids 5*, 494-496.

154. Creer, M. H., and Gross, R. W. (1985) Separation of isomeric lysophospholipids by reverse phase HPLC, *Lipids 20*, 922-928.

155. Gao, B., Boeglin, W. E., Zheng, Y., Schneider, C., and Brash, A. R. (2009) Evidence for an ionic intermediate in the transformation of fatty acid hydroperoxide by a catalase-related allene oxide synthase from the cyanobacterium *Acaryochloris marina*, *J Biol Chem 284*, 22087-22098.

156. Wei, X., Sun, W., Shi, X., Koo, I., Wang, B., Zhang, J., Yin, X., Tang, Y., Bogdanov, B., Kim, S., Zhou, Z., McClain, C., and Zhang, X. (2011) MetSign: a computational platform for high-resolution mass spectrometry-based metabolomics, *Anal Chem 83*, 7668-7675.

157. Dong, J., Cai, X., Zhao, L., Xue, X., Zou, L., Zhang, X., and Liang, X. (2010) Lysophosphatidylcholine profiling of plasma: discrimination of isomers and discovery of lung cancer biomarkers, *Metabolomics 6*, 478-488.

158. Pluckthun, A., and Dennis, E. A. (1982) Acyl and phosphoryl migration in lysophospholipids: importance in phospholipid synthesis and phospholipase specificity, *Biochemistry 21*, 1743-1750.

159. Hofer, R., Briesen, I., Beck, M., Pinot, F., Schreiber, L., and Franke, R. (2008) The Arabidopsis cytochrome P450 CYP86A1 encodes a fatty acid omega-hydroxylase involved in suberin monomer biosynthesis, *J Exp Bot 59*, 2347-2360.

160. Capdevila J, H. V., Ralck JR. (2005) Cytochrome P450 and the metabolism and bioactivation of arachidonic acid and eicosanoids, In *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R., Ed.) 3rd ed., pp 531-551, Kluwer Academic/Plenum Publishers, New York.

161. Rabagny, Y., Herrmann, W., Geisel, J., Kirsch, S. H., and Obeid, R. (2011) Quantification of plasma phospholipids by ultra performance liquid chromatography tandem mass spectrometry, *Anal Bioanal Chem 401*, 891-899.

162. Croset, M., Brossard, N., Polette, A., and Lagarde, M. (2000) Characterization of plasma unsaturated lysophosphatidylcholines in human and rat, *Biochem J 345 Pt 1*, 61-67.

163. Makide, K., Kitamura, H., Sato, Y., Okutani, M., and Aoki, J. (2009) Emerging lysophospholipid mediators, lysophosphatidylserine, lysophosphatidylthreonine, lysophosphatidylethanolamine and lysophosphatidylglycerol, *Prostaglandins Other Lipid Mediat 89*, 135-139.

164. Ryborg, A. K., Deleuran, B., Thestrup-Pedersen, K., and Kragballe, K. (1994) Lysophosphatidylcholine: a chemoattractant to human T lymphocytes, *Arch Dermatol Res 286*, 462-465.

165. Yang, L. V., Radu, C. G., Wang, L., Riedinger, M., and Witte, O. N. (2005) Gi-independent macrophage chemotaxis to lysophosphatidylcholine via the immunoregulatory GPCR G2A, *Blood 105*, 1127-1134.

166. Zhao, Z., Xiao, Y., Elson, P., Tan, H., Plummer, S. J., Berk, M., Aung, P. P., Lavery, I. C., Achkar, J. P., Li, L., Casey, G., and Xu, Y. (2007) Plasma lysophosphatidylcholine levels: potential biomarkers for colorectal cancer, *J Clin Oncol 25*, 2696-2701.

167. van Corven, E. J., Groenink, A., Jalink, K., Eichholtz, T., and Moolenaar, W. H. (1989) Lysophosphatidate-induced cell proliferation: identification and dissection of signaling pathways mediated by G proteins, *Cell 59*, 45-54.

168. Moolenaar, W. H., van Meeteren, L. A., and Giepmans, B. N. (2004) The ins and outs of lysophosphatidic acid signaling, *Bioessays 26*, 870-881.

169. Sun, H., Ren, J., Zhu, Q., Kong, F. Z., Wu, L., and Pan, B. R. (2009) Effects of lysophosphatidic acid on human colon cancer cells and its mechanisms of action, *World J Gastroenterol 15*, 4547-4555.

170. Zhang, H., Bialkowska, A., Rusovici, R., Chanchevalap, S., Shim, H., Katz, J. P., Yang, V. W., and Yun, C. C. (2007) Lysophosphatidic acid facilitates proliferation of colon cancer cells via induction of Kruppel-like factor 5, *J Biol Chem 282*, 15541-15549.

171. Shida, D., Kitayama, J., Yamaguchi, H., Okaji, Y., Tsuno, N. H., Watanabe, T., Takuwa, Y., and Nagawa, H. (2003) Lysophosphatidic acid (LPA) enhances the metastatic potential of human colon carcinoma DLD1 cells through LPA1, *Cancer Res 63*, 1706-1711.

172. Choi, J. W., Lim, S., Oh, Y.-S., Kim, E.-K., Kim, S.-H., Kim, Y.-H., Heo, K., Kim, J., Kim, J. K., Yang, Y. R., Ryu, S. H., and Suh, P.-G. (2010) Subtype-specific role of phospholipase C-β in bradykinin and LPA signaling through differential binding of different PDZ scaffold proteins, *Cell Signal 22*, 1153-1161.

173. Lee, S. J., and Yun, C. C. (2010) Colorectal cancer cells - proliferation, survival and invasion by lysophosphatidic acid, *Int J Biochem Cell Biol 42*, 1907-1910.

174. Adachi, J., Asano, M., Yoshioka, N., Nushida, H., and Ueno, Y. (2006) Analysis of phosphatidylcholine oxidation products in human plasma using quadrupole time-of-flight mass spectrometry, *Kobe J Med Sci 52*, 127-140.

175. Karara, A., Dishman, E., Falck, J. R., and Capdevila, J. H. (1991) Endogenous epoxyeicosatrienoyl-phospholipids. A novel class of cellular glycerolipids containing epoxidized arachidonate moieties, *J Biol Chem 266*, 7561-7569.

176. Mast, N., Andersson, U., Nakayama, K., Bjorkhem, I., and Pikuleva, I. A. (2004) Expression of human cytochrome P450 46A1 in *Escherichia coli*: effects of N- and C-terminal modifications, *Arch Biochem Biophys 428*, 99-108.

177. Xiao, Y., Shinkyo, R., and Guengerich, F. P. (2011) Cytochrome P450 2S1 is reduced by NADPH-cytochrome P450 reductase, *Drug Metab Dispos 39*, 944-946.

178. Lee, S. M., Kwon, G. Y., Kim, K. O., and Kim, Y. S. (2011) Metabolomic approach for determination of key volatile compounds related to beef flavor in glutathione-Maillard reaction products, *Anal Chim Acta 703*, 204-211.

179. Mi Lee, S., and Kim, Y.-S. (2013) Application of non-targeted approach in flavor science: a review, In *Nutrition, Functional and Sensory Properties of Foods*, pp 148-155, The Royal Society of Chemistry.

180. Sharma, K., Karl, B., Mathew, A. V., Gangoiti, J. A., Wassel, C. L., Saito, R., Pu, M., Sharma, S., You, Y. H., Wang, L., Diamond-Stanic, M., Lindenmeyer, M. T., Forsblom, C., Wu, W., Ix, J. H., Ideker, T., Kopp, J. B., Nigam, S. K., Cohen, C. D., Groop, P. H., Barshop, B. A., Natarajan, L., Nyhan, W. L., and Naviaux, R. K. (2013) Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease, *J Am Soc Nephrol 24*, 1901-1912.

181. Zhang, H., Zhang, D., Ray, K., and Zhu, M. (2009) Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry, *J Mass Spectrom 44*, 999-1016.

182. Masson, P., Alves, A. C., Ebbels, T. M., Nicholson, J. K., and Want, E. J. (2010) Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS, *Anal Chem 82*, 7779-7786.

183. Saric, J., Want, E. J., Duthaler, U., Lewis, M., Keiser, J., Shockcor, J. P., Ross, G. A., Nicholson, J. K., Holmes, E., and Tavares, M. F. (2012) Systematic evaluation of extraction methods for multiplatform-based metabotyping: application to the Fasciola hepatica metabolome, *Anal Chem 84*, 6963-6972.

184. Kalsotra, A., Turman, C. M., Kikuta, Y., and Strobel, H. W. (2004) Expression and characterization of human cytochrome P450 4F11: Putative role in the metabolism of therapeutic drugs and eicosanoids, *Toxicol Appl Pharmacol 199*, 295-304.