Published in final edited form as:

Biochim Biophys Acta. 2011 January ; 1814(1): 139–145. doi:10.1016/j.bbapap.2010.05.005.

Approaches to Deorphanization of Human and Microbial Cytochrome P450 Enzymes

F. Peter Guengerich, Zhongmei Tang, Qian Cheng, and S. Giovanna Salamanca-Pinzón

Department of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0146 U.S.A

Abstract

One of the general problems in biology today is that we are characterizing genomic sequences much faster than identifying the functions of the gene products, and the same problem exists with cytochromes P450 (P450). One-fourth of the human P450s are not well-characterized and therefore considered “orphans.” A number of approaches to deorphanization are discussed generally. Several liquid chromatography-mass spectrometry approaches have been applied to some of the human and *Streptomyces coelicolor* P450s. One current limitation is that too many fatty acid oxidations have been identified and we are probably missing more relevant substrates, possibly due to limits of sensitivity.

Keywords

Cytochrome P450; enzyme function; mass spectrometry; metabolomics; secondary pathways

1. Introduction: Human P450s and the orphan problem

Biochemistry is a wonderful field of science and continues to thrive, even with aspects taking new names and their own identities. Although we often talk about biochemical functions of proteins, the reality is that we live in a world in which we employ dyes and model substrates for convenience in assays and often do not know what the functions of our proteins and enzymes really are. Even in *Escherichia coli* we only know the function of ~54% of the genes [1]. Obviously the annotation is much worse in mammals, and even many of the “accepted” human enzyme functions have been assumed based on comparisons with animal models.

Historically two views have dominated the cytochrome P450 (P450) world. One is that the P450s have specific biological functions, even if the true substrates are not known yet. The other is that these enzymes do not have physiological substrates but collectively act as a protective system to remove the natural products that we ingest (from plants and other food sources), e.g. terpenes, alkaloids [2]. With regard to the latter view, modern medicines and many other man-made chemicals resemble the natural products and thus become substrates. For instance, ~75% of the enzyme reactions seen with drugs are attributed to P450s [3].

Address correspondence to: Prof. F. Peter Guengerich, Department of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, 638 Robinson Research Building, 2200 Pierce Avenue, Nashville, Tennessee 37232-0146 U.S.A., Telephone: 1 (615) 322-2261, FAX: 1 (615) 322-3141, f.guengerich@vanderbilt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The growing conclusion is that both views are correct, in part. One way, albeit not perfect, of classifying P450 function is shown in Table 1, first proposed in 2005 [4]. The P450s involved in the metabolism of sterols and vitamins—and P450s 5A1 and 8A1—have important functions and, as shown in mouse models and human inherited diseases, there are very serious consequences if these P450s are deleted or non-functional. The P450s listed under “xenobiotics” and most of those involved in fatty acid oxidation do not seem critical, in mice or humans. We can assume they probably do not catalyze critical reactions, although the possibility exists that epidemiological studies may ultimately show some change in longevity if one of these is absent (or possibly an improvement without one). However, there is clear evidence that the absence of some of these P450s can have major consequences, including death, if certain drugs are used [5,6]. Alternatively, the overexpression of P450 2D6 has been implicated in the unusually rapid conversion of codeine to morphine in a lactating mother and the death of her nursing child [7].

One can view the human P450s in the “Xenobiotics” (substrate) category of Table 1 as having roles as general protective mechanisms. Of interest here are the 13 P450s listed under “Unknown” in Table 1, which are considered the “Orphans.” We will address strategies that we and others have applied to the “deorphanization” of these P450s, i.e. defining their catalytic activities.

2. The general issue of deorphanization

In the biochemistry that developed in the 20th Century, the movement of progress is shown as depicted in the top of the Scheme of Fig. 1. That is, *in vivo* observations (e.g., metabolism) led to *in vitro* assays, which could be used to purify enzymes. With the development of recombinant DNA technology, it became possible to isolate the relevant cDNA and then the gene. Today we know the nucleotide sequences of all of the genes in many species, including humans. How does one work backward then, i.e. from the genomic sequence to function? As shown in the bottom of Fig. 1, a genomic sequence can be used to predict open reading frames (i.e., cDNAs). The next step is heterologous expression of protein, unless this has been done already. This aspect can be problematic, especially if a large amount is desired for structural work or some other uses that require much. We will not address the issue of heterologous expression here but focus on the last step, using proteins to determine substrates and reactions.

3. Targeted strategies for deorphanization

3.1. Sequence comparisons

One approach has been the use of primary sequence comparison to define function. Of course this is a reliable approach when there is a high level of sequence identity and a lack of redundancy; e.g. even if human P450 7A1 had not been studied, its function (first step in bile acid formation) could have been inferred from its homology with animal orthologs. When sequence identities are low, the approach is highly problematic. The positioning of amino acids, e.g. for acid-base catalysis, can lead to catalysis of various types of reactions. For instance, the so-called enolase superfamily contains epimerases, racemases, dehydratases, and enolases [9].

3.2. Structural insight

Another approach involves the use of structural biology. The concept is that elucidation of the 3-dimensional structures of enzymes (e.g. P450s) may allow prediction of their substrates by docking etc. However, there are at least two (major) problems with this approach. The first is that many of the P450s have shown major structural changes between the unbound and substrate-bound forms [10]. The other major problem is that a number of

the P450s have large binding sites in which only part of the available space is used for the substrate [11]. Even in some of the substrate-bound P450 structures the function is not clear. For instance, P450 3A4 has been crystallized with progesterone positioned far from the heme [12]. In the crystal structure of the human P450 1A2- α -naphthoflavone structure [13] the juxtaposition of the substrate is such that the known site of oxidation [14] is inconsistent with the structure, and therefore an alternate conformation must be active.

3.3. Targeted substrates

One approach is to try selected reactions with the recombinant enzyme (P450). When some sequence identity to other proteins (P450s) exists, there is some logic in this approach. For instance, when P450 27C1 was isolated [15], a logical trial involved vitamin D₍₃₎ and cholesterol in light of the known catalytic activities of P450s 27A1 and 27B1 [4]. However, these attempts were unsuccessful [15], and this approach will work only if there is good insight (or luck). Unfortunately, there are many potential reactions to try and life is short, so this approach will have limited success. Demonstrating trace activity of oxidation of fatty acids or model spectrophotometric (or fluorescence) or drug assays is not particularly useful.

3. 4. Trials with small libraries

One approach is to run assays with sets of compounds. Such an approach may provide some advantages. One successful case is the finding that P450 2W1 can activate a variety of chemical carcinogens, detected using a medium-throughput bacterial genotoxicity assay (Fig. 2) [16]. However, similar trials with P450s 2S1 [16] and 4F11 [17] were negative.

We have tried another library approach with a few representatives of each of several classes of potential compounds, e.g. a few fatty acids, a few steroid hormones, cholesterol, etc. The concept is that these can be mixed and used as a substrate, with LC-MS analysis of the mixture. If any of these compounds (e.g. steroids) show activity, then the pool (of steroids) could be expanded with the goal of finding one with better activity. To date this approach has not been successful. One concern is optimizing LC-MS conditions for the entire library of compounds, which is impossible. Nevertheless, this option should not be dismissed.

4. Untargeted searches

Untargeted approaches are used in the absence of preconceived biases about what the substrates should be. In effect there are various types of metabolomics experiments, which may be used in any of several modes. The working hypothesis is that the substrate exists in the tissue in which the P450 is expressed (which may not necessarily be the case, e.g. in an endocrine system).

One approach involves binding of the natural substrate to the P450. The concept is that the protein of interest (e.g. P450) can be immobilized and used as a reagent to pull a ligand from a complex mixture. The ligand can be extracted and characterized. This approach has not been used extensively, although recently Forman and his associates applied this to identifying linoleic acid as a ligand for hepatic nuclear factor 4 α [18]. The inherent problem is that very high affinity is required and washing is difficult. A K_d of 10^{-7} μ M coupled with a “typical” on-rate of 10^7 $M^{-1} s^{-1}$ yields an off-rate of $1 s^{-1}$ ($t_{1/2} \sim 1$ s). We are not aware that this approach has been used successfully with P450s. Of course, a ligand may be an inhibitor and not a substrate.

Both *in vivo* and *in vitro* metabolomic approaches can be applied (Fig. 3). Although NMR could be used, in principle, LC-MS has much greater sensitivity for what are expected to be secondary pathways (that the P450s are involved in). Body fluids and tissues from transgenic animals can be compared directly, or tissues can be analyzed for *in vitro* activity.

Another option presented in Fig. 3 involves the use of recombinant proteins (P450s) with *in vitro* assays, with comparisons of LC-MS profiles.

5. Strategies used with human P450s

The focus of the work done in this laboratory has involved using recombinant human P450s as reagents along with extracts of human tissues as sources of substrates (Fig. 3). The crux of the system is making comparisons of multiple LC-MS data sets. In some cases this can be done with principal component analysis. Another freely available software system is MZmine [19], which can be used to compare two runs, e.g. with and without a P450 added to a tissue extract. The P450 systems can be purified and reconstituted (with NADPH-P450 reductase and phospholipid) or “bicistronic membranes” (membranes from *E. coli* cells in which both the P450 and NADPH-P450 reductase have been expressed [20]). However, we have found in some cases with the *E. coli* membranes that such searches led to the identification of fatty acid substrates in the bacterial membranes as opposed to the tissue extracts.

One of the approaches in Fig. 3 involves the use of $^{18}\text{O}_2$ and the program DoGEX [21], which is elaborated in more detail in Fig. 4. A mixture of the reconstituted P450 system and a tissue extract (corresponding to the localization of the P450) are mixed in a Thunberg tube and air is removed by a series of cycles of vacuum and argon. The sample (under vacuum) is then charged with a 50:50 mixture (v/v) of $^{16}\text{O}_2$ and $^{18}\text{O}_2$. Therefore all oxygenated P450 products should contain parent (m/z) ions at M and $M+2$, with nearly equal intensity (similar to $^{79}\text{Br}/^{81}\text{Br}$ spectra). Elements of the DoGEX program are used for t_R alignment, baseline correction, and peak smoothing (Fig. 4A). The key is the search for the $M/M+2$ doublets (m/z dimension) (Fig. 4B). Peak doublets approaching an $M/M+2$ ratio of unity are shown as dark bands on a light (yellow) background. These can be expanded (Fig. 4C) and displayed in 3-dimensional profiles (Fig. 4D). The example shown in Fig. 4 resulted from the incubation of P450 7A1 with cholesterol to give 7α -hydroxycholesterol, which was then characterized by coincident elution with an authentic standard in this validation experiment [22].

The ^{18}O DoGEX approach was applied to P450s 1A2, 2C8, and 2C9, three “Xenobiotic” P450s already known to oxidize a variety of drugs and some other xenobiotic substrates (Table 1). With the DoGEX analysis a number of fatty acids were identified as substrates [22], and further work with the commercial fatty acids was done to identify the products and rates of formation. Some of the reactions were already known but others were not. The work also establishes these reactions as capable of occurring in these liver extracts, although it does not speak to the physiological relevance.

A similar study with the orphan P450 4F11 (Table 1) also yielded a series of fatty acids as substrates, and the products and catalytic efficiencies of product formation were characterized [17].

6. Current status of human P450 orphans

Current information is summarized in Table 2. First, there is very limited evidence that the CYP2A7 and CYP4A22 genes are expressed (to proteins). With P450 2S1, there have been reports of retinoic acid [24] and naphthalene [25] oxidation that were not subsequently confirmed [16,26]. No other substrates, including carcinogens, were identified in our own work [16]. However, Hankinson’s group has recently reported that P450 2S1 is capable of oxidations of retinoic acid and some carcinogens in the presence of organic hydroperoxides but not the usual NADPH-P450 reductase system [26,27]. The significance of these results is not clear. Although hydroperoxide-supported P450 reactions were first discovered in 1973

[40], the physiological relevance has never been documented, and transgenic mice with very low NADPH-P450 reductase levels show very limited hepatic oxidation of drugs [41]. Nevertheless, cells transfected with P450 2S1 did show metabolism of benzo[*a*]pyrene [26,27].

No substrates have been found for P450 2W1 except several carcinogens [16], except for trace activity toward arachidonic acid [16,30].

P450s 20A1 and 27C1 have been expressed but no substrates have been found [15,39]. P450 27C1 is in a subfamily known to have a mitochondrial localization (at least for P450s 27A1 and 27B1) [4] and presumably uses electrons from adrenodoxin reductase and adrenodoxin. However, attempts to find substrates with these accessory proteins have been negative to date.

The remainder of the P450s in Table 2 have been reported to oxidize fatty acids, (e.g. P450s 2U1, 4F11, 4V2, 4Z1) or anandamide (an amide of arachidonic acid) (P450 4X1) [36]. Although there is a case that some of the fatty acid oxidation products have physiological activities [42], the functions of these P450s should not be considered definite. Many P450s oxidize fatty acids (*vide supra*), and the point can be made that we are seeing these reactions simply because of their sensitivity in the current LC-MS procedures (Fig. 4) [17,22] or simply because investigators have tried these first in targeted searches for substrates [28,35,38]. Recently we have modified dansylation methods for reaction with unactivated alcohols to increase the positive-ion LC-MS sensitivity in DoGEX and other untargeted searches.

7. Microbial and plant P450s

The microbial and plant P450 systems are even more complex than the human and other mammalian systems. One reason is the very large number of P450 genes in plants, e.g. 246 in the model organism *Arabidopsis thaliana* and 328 in rice. A number of efforts are underway to annotate plant P450 genes [43,44] and will not be discussed here. With regard to microorganisms, there are advantages in that knockouts can be readily made and phenotypes (possibly) observed. In some cases, the placement of P450 genes in operons has been instructive [1,45]. However, the number of microorganisms whose genomes have been sequenced is considerable. Many of the more complex bacteria (e.g. Actinomyces) have a substantial number of P450 genes (e.g. 32 in *Streptomyces avermitilis*). Another problem in assessing relevance of P450 reactions is that the “natural” conditions for growth of many of these organisms are unknown (e.g. many are soil bacteria). Finally, the electron transport chains to the P450s are generally not well understood.

8. *Streptomyces coelicolor* P450s

Streptomyces and other Actinomyces figure prominently in the production of antibiotics, being responsible for the production of ~2/3 of these drugs used today. General aspects of the *Streptomyces* P450s have been reviewed elsewhere [45–47] including work on P450s 154C1, 158A2, and 170A1. Two examples of studies of orphan *S. coelicolor* P450s will be presented here.

P450 105D5 was an enigma to study, in that no substrate was known and the electron transfer pathway was unclear [48]. While the only way to deliver electrons to mammalian microsomal P450s is via NADPH-P450 reductase, *S. coelicolor* has four ferredoxin reductases and six ferredoxins, thus providing ≥ 24 potential pathways for delivery to the P450. A break in the problem came with the finding that fatty acids bind to P450 105D5 with relatively high affinity [48]. Catalytic activity (ω -1, ω -2, ω -3 hydroxylation) could be

achieved in the presence of a (heterologous) spinach NADPH-ferredoxin reductase/ferredoxin reduction system or, surprisingly, with *Pseudomonas putida* NADH-putidaredoxin reductase and putidaredoxin. All four *S. coelicolor* ferredoxin reductases and six ferredoxins were expressed (in *E. coli*), purified, and used in combinations. Although the system could be reconstituted with the *P. putida* or spinach reduction systems, only one of the possible *S. coelicolor* systems was highly efficient, involving the pathway $\text{NADH} \rightarrow \text{FDR1} \rightarrow \text{Fdx4} \rightarrow \text{P450 105D5}$ [48]. Also interestingly, this reduction system did not have a higher turnover number than seen observed with the spinach system. Although the significance of fatty acid hydroxylation in *S. coelicolor* is unknown, hydroxy fatty acids can be identified in the organism [48]. One open question is whether this same *S. coelicolor* electron transport system (Fig. 5) would be preferred if another substrate for P450 105D5 were found, thus complicating untargeted searches for more substrates. Another open question is what combination of electron transfer proteins is used for other *S. coelicolor* P450s. In general, the electron transfer pathways to bacterial P450s are not well understood, the very early success with the putidaredoxin system notwithstanding [49].

A different approach has been utilized with *S. coelicolor* P450 154A1 (Fig. 6). Knocking out the *CYP154A1* gene gave a mild phenotypic change. The organism appears to grow normally but the spores are unstable in long-term storage. Extracts of wild-type and *CYP154A1*^{-/-} *S. coelicolor* cells were analyzed by LC-MS, and the program MZmine showed a difference at *m/z* 259 (MH⁺), indicating the accumulation of this compound in *CYP154A1*^{-/-} cells and suggesting that it might be a substrate. When the crude extract was incubated with purified P450 154A1, the LC-MS peak disappeared and at least one new one formed. The chromatographic peak was purified (from *S. coelicolor* 154A1^{-/-} extracts) and characterized, yielding the structure of QCP1 (Fig. 6). Interestingly, the conversion to the new product(s) does not change the molecular mass, and subsequent studies showed that the reduction system (spinach ferredoxin and reductase, NADPH) had no effect. Thus, P450 154A1 appears to be catalyzing a rearrangement, not an oxidation or reduction. This non-redox behavior of P450s is not without precedent but not well understood [50–52].

9. Conclusions

One of the central problems in modern biochemistry (and even biology in general) is defining the functions of genes and their products, i.e. proteins. This is an enormous task and goes well beyond our own capabilities, and we are focused on the problem with the human P450s and some of the microbial P450s.

One of the conclusions about the work to date is that untargeted approaches will be necessary. Fortunately, at this point in time some outstanding LC-MS equipment is available and useful software analysis methods have appeared [19,21,53]. Nevertheless, there is still a need for more sensitivity in the detection of endogenous compounds, as already mentioned. Software systems will need to be even more robust.

Another challenge is to continue to search for substrates for the human P450s. Approaches have been outlined (Fig. 3) but what has been done to date is not very comprehensive. The catalytic activities seen with fatty acids are not necessarily the most relevant reactions. Rather than speculate on why these reactions might be important, the field would be better served by more searches and experimental evidence.

We anticipate that more definitive information about the reactions of the remaining human orphan P450s will be forthcoming in the next decade. However, the enormity of the problem with plant and microbial P450s requires that more time and input will be needed. The challenge of 328 P450s in rice is daunting (and the number in other economically important

plants may even be higher), and many of these P450s are probably critical not only in the normal physiology of these important food plants but also in very practical matters such as pesticide development and natural resistance. Further, there is the challenge in microbial P450 research of moving from intellectually interesting reactions (e.g. Fig. 6) to pathways related to important antibiotics.

Acknowledgments

The work described here was supported in part by U.S. Public Health Service Grants R37 CA090426, R01 GM069970, and P30 ES000267. We thank K. Trisler for her assistance in preparation of the manuscript.

This paper is dedicated to Prof. Dr. Klaus Ruckpaul on the occasion of his 80th birthday. The corresponding author (F. P. G.) first met Prof. Ruckpaul 30 years ago at a meeting in Bratislava (now in Slovakia). Prof. Ruckpaul worked under difficult conditions yet trained some of today's leading figures in German P450 research. He was also the leading force in establishing what are now known as the (biennial) International Conferences on Cytochrome P450 (Biochemistry and Biophysics). Although now held throughout the world and involving many P450 investigators (see <http://www.p450meetings.com> for history and current aspects), when they began (1976) these were one of the few opportunities for Eastern Bloc P450 scientists to meet other researchers and discuss the field. Today many things have changed for the better but we still remember Klaus Ruckpaul's pioneering contributions toward cooperation. At the international P450 symposia (*vide supra*) the Ruckpaul Award is given to the best poster presented by a young scientist at each meeting. Happy birthday, Klaus.

References

- Hanson AD, Pribat A, Waller JC, Crecy-Lagard V. 'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list-and how to find it. *Biochem J* 2010;425:1–11. [PubMed: 20001958]
- Jakoby, WB. Detoxication enzymes. In: Jakoby, WB., editor. *Enzymatic Basis of Detoxication*. Vol. 1. Academic Press; New York: 1980. p. 1-6.
- Williams JA, Hyland R, Jones BC, Smith DA, Hurst S, Goosen TC, Peterkin V, Koup JR, Ball SE. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUC_i/AUC) ratios. *Drug Metab Dispos* 2004;32:1201–1208. [PubMed: 15304429]
- Guengerich, FP. Human cytochrome P450 enzymes. In: Ortiz de Montellano, PR., editor. *Cytochrome P450: Structure Mechanism, and Biochemistry*. Kluwer Academic/Plenum Publishers; New York: 2005. p. 377-530.
- Idle JR, Smith RL. Polymorphisms of oxidation at carbon centers of drugs and their clinical significance. *Drug Metab Rev* 1979;9:301–317. [PubMed: 158499]
- Shah RR, Oates NS, Idle JR, Smith RL, Lockhart JDF. Impaired oxidation of debrisoquine in patients with perhexiline neuropathy. *Brit Med J* 1982;284:295–299. [PubMed: 6277419]
- Willmann S, Edginton AN, Coboeken K, Ahr G, Lippert J. Risk to the breast-fed neonate from codeine treatment to the mother: a quantitative mechanistic modeling study. *Clin Pharmacol Ther* 2009;86:634–643. [PubMed: 19710640]
- Guengerich FP, Tang Z, Salamanca-Pinzon SG, Cheng Q. Characterizing proteins of unknown function: orphan cytochrome P450s as a paradigm. *Mol Interventions*. 2010 in press.
- Babbitt PC, Hasson MS, Wedekind JE, Palmer DRJ, Barrett WC, Reed GH, Rayment I, Ringer D, Kenyon GL, Gerlt JA. The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α -protons of carboxylic acids. *Biochemistry* 1996;35:16489–116501.
- Poulos, TL.; Johnson, EF. Structures of cytochrome P450 enzymes. In: Ortiz de Montellano, PR., editor. *Cytochrome P450: Structure Mechanism, and Biochemistry*. Kluwer Academic/Plenum Publishers; New York: 2005. p. 87-114.
- Yano JK, Wester MR, Schoch GA, Griffin KJ, Stout CD, Johnson EF. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05 Å resolution. *J Biol Chem* 2004;38091–38094. [PubMed: 15258162]

12. Williams PA, Cosme J, Vinkovic DM, Ward A, Angove HC, Day PJ, Vonnrhein C, Tickle IJ, Jhoti H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* 2004;305:683–686. [PubMed: 15256616]
13. Sansen S, Yano JK, Reynald RL, Schoch GA, Griffin KJ, Stout CD, Johnson EF. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *J Biol Chem* 2007;282:14348–14355. [PubMed: 17311915]
14. Sohl CD, Isin EM, Eoff RL, Marsch GA, Stec DF, Guengerich FP. Cooperativity in oxidation reactions catalyzed by cytochrome P450 1A2. Highly cooperative pyrene hydroxylation and multiphasic kinetics of ligand binding. *J Biol Chem* 2008;283:7293–7308. [PubMed: 18187423]
15. Wu ZL, Bartleson CJ, Ham AJL, Guengerich FP. Heterologous expression, purification, and properties of human cytochrome P450 27C1. *Arch Biochem Biophys* 2006;445:138–146. [PubMed: 16360114]
16. Wu ZL, Sohl CD, Shimada T, Guengerich FP. Recombinant enzymes over-expressed in bacteria show broad catalytic specificity of human cytochrome P450 2W1 and limited activity of human cytochrome P450 2S1. *Mol Pharmacol* 2006;69:2007–2014. [PubMed: 16551781]
17. Tang Z, Salamanca-Pinzón SG, Wu ZL, Xiao Y, Guengerich FP. Human cytochrome P450 4F11: heterologous expression in bacteria, purification, and characterization of catalytic function. *Arch Biochem Biophys* 2010;494:86–93. [PubMed: 19932081]
18. Yuan X, Ta TC, Lin M, Evans JR, Dong Y, Bolotin E, Sherman MA, Forman BM, Sladek FM. Identification of an endogenous ligand bound to a native orphan nuclear receptor. *PloS one* 2009;4:e5609. [PubMed: 19440305]
19. Katajamaa M, Miettinen J, Oresic M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* (Oxford, U.K.) 2006;22:634–636.
20. Parikh A, Gillam EMJ, Guengerich FP. Drug metabolism by *Escherichia coli* expressing human cytochromes P450. *Nat Biotechnol* 1997;15:784–788. [PubMed: 9255795]
21. Sanchez-Ponce R, Guengerich FP. Untargeted analysis of mass spectrometry data for elucidation of function of new enzymes. *Anal Chem* 2007;79:3355–3362. [PubMed: 17408243]
22. Tang Z, Martin MV, Guengerich FP. Elucidation of functions of human cytochrome P450 enzymes: Identification of endogenous substrates in tissue extracts by LC-MS metabolomics and isotopic labeling approaches. *Anal Chem* 2009;81:3071–3078. [PubMed: 19301915]
23. Rylander T, Neve EPA, Ingelman-Sundberg M, Oscarson M. Identification and tissue distribution of the novel human cytochrome P450 2S1 (CYP2S1). *Biochem Biophys Res Commun* 2001;281:529–535. [PubMed: 11181079]
24. Smith G, Wolf CR, Deeni YY, Dawe RS, Evans AT, Comrie MM, Ferguson J, Ibbotson SH. Cutaneous expression of cytochrome P450 CYP2S1: individuality in regulation by therapeutic agents for psoriasis and other skin diseases. *Lancet* 2003;361:1336–1343. [PubMed: 12711469]
25. Karlgren M, Miura S, Ingelman-Sundberg M. Novel extrahepatic cytochrome P450s. *Toxicol Appl Pharmacol* 2005;207:57–61. [PubMed: 15987645]
26. Bui PH, Hankinson O. Functional characterization of human cytochrome P450 2S1 using a synthetic gene-expressed protein in *Escherichia coli*. *Mol Pharmacol* 2009;76:1031–1043. [PubMed: 19713358]
27. Bui PH, Hsu EL, Hankinson O. Fatty acid hydroperoxides support cytochrome P450 2S1-mediated bioactivation of benzo [a]pyrene-7,8-dihydrodiol. *Mol Pharmacol* 2009;76:1044–1052. [PubMed: 19713357]
28. Chuang SS, Helvig C, Taimi M, Ramshaw HA, Collop AH, Amad M, White JA, Petkovich M, Jones G, Korczak B. CYP2U1, a novel human thymus- and brain-specific cytochrome P450, catalyzes ω - and (ω -1)-hydroxylation of fatty acids. *J Biol Chem* 2004;279:6305–6314. [PubMed: 14660610]
29. Karlgren M, Backlund M, Johansson I, Oscarson M, Ingelman-Sundberg M. Characterization and tissue distribution of a novel human cytochrome P450-CYP2U1. *Biochem Biophys Res Commun* 2004;315:679–685. [PubMed: 14975754]
30. Karlgren M, Gomez A, Stark K, Svard J, Rodriguez-Antona C, Oliw E, Bernal ML, YCSR, Johansson I, Ingelman-Sundberg M. Tumor-specific expression of the novel cytochrome P450 enzyme, CYP2W1. *Biochem Biophys Res Commun* 2006;341:451–458. [PubMed: 16426568]

31. Agarwal V, Kommaddi RP, Valli K, Ryder D, Hyde TM, Kleinman JE, Strobel HW, Ravindranath V. Drug metabolism in human brain: high levels of cytochrome P450 3A43 in brain and metabolism of anti-anxiety drug alprazolam to its active metabolites. *PLoS one* 2008;3:e2337. [PubMed: 18545703]
32. Cui X, Nelson DR, Strobel HW. A novel human cytochrome P450 4F isoform (CYP4F11): cDNA cloning, expression, and genomic structural characterization. *Genomics* 2000;68:161–166. [PubMed: 10964514]
33. Dhar M, Sepkovic DW, Hirani V, Magnusson RP, Lasker JM. Omega oxidation of 3-hydroxy fatty acids by the human CYP4F gene subfamily enzyme CYP4F11. *J Lipid Res* 2008;49:612–624. [PubMed: 18065749]
34. Li A, Jiao X, Munier FL, Schorderet DF, Yao W, Iwata F, Hayakawa M, Kanai A, Shy Chen M, Alan Lewis R, Heckenlively J, Weleber RG, Traboulsi EI, Zhang Q, Xiao X, Kaiser-Kupfer M, Sergeev YV, Hejtmancik JF. Bietti crystalline corneoretinal dystrophy is caused by mutations in the novel gene CYP4V2. *Am J Hum Genet* 2004;74:817–826. [PubMed: 15042513]
35. Nakano M, Kelly EJ, Rettie AE. Expression and characterization of CYP4V2 as a fatty acid omega-hydroxylase. *Drug Metab Dispos* 2009;37:2119–2122. [PubMed: 19661213]
36. Stark K, Dostalek M, Guengerich FP. Expression and purification of orphan cytochrome P450 4X1 and oxidation of anandamide. *FEBS J* 2008;275:3706–3717. [PubMed: 18549450]
37. Rieger MA, Ebner R, Bell DR, Kiessling A, Rohayem J, Schmitz M, Temme A, Rieber EP, Weigle B. Identification of a novel mammary-restricted cytochrome P450, CYP4Z1, with overexpression in breast carcinoma. *Cancer Res* 2004;64:2357–2364. [PubMed: 15059886]
38. Zollner A, Dragan CA, Pistorius D, Muller R, Bode HB, Peters FT, Maurer HH, Bureik M. Human CYP4Z1 catalyzes the in-chain hydroxylation of lauric acid and myristic acid. *Biol Chem* 2009;390:313–317. [PubMed: 19090726]
39. Stark K, Wu ZL, Bartleson CJ, Guengerich FP. Heterologous expression and mRNA distribution of orphan cytochrome P450 20A1. *Drug Metab Dispos* 2008;36:1930–1937. [PubMed: 18541694]
40. Kadlubar FF, Morton KC, Ziegler DM. Microsomal-catalyzed hydroperoxide-dependent C-oxidation of amines. *Biochem Biophys Res Commun* 1973;54:1255–1261. [PubMed: 4148229]
41. Gu J, Weng Y, Zhang QY, Cui H, Behr M, Wu L, Yang W, Zhang L, Ding X. Liver-specific deletion of the NADPH-cytochrome P450 reductase gene: impact on plasma cholesterol homeostasis and the function and regulation of microsomal cytochrome P450 and heme oxygenase. *J Biol Chem* 2003;278:25895–25901. [PubMed: 12697746]
42. Capdevila, J.; Holla, VR.; Falck, JR. Cytochrome P450 and the metabolism and bioactivation of arachidonic acid and eicosanoids. In: Ortiz de Montellano, PR., editor. *Cytochrome P450 Structure, Mechanism, and Biochemistry*. Kluwer Academic/Plenum Publishers; New York: 2005. p. 531-551.
43. Olry A, Schneider-Belhaddad F, Heintz D, Werck-Reichhart D. A medium-throughput screening assay to determine catalytic activities of oxygen-consuming enzymes: a new tool for functional characterization of cytochrome P450 and other oxygenases. *Plant J* 2007;51:331–340. [PubMed: 17610545]
44. Ehrling J, Sauveplane V, Olry A, Ginglinger JF, Provart NJ, Werck-Reichhart D. An extensive (co-)expression analysis tool for the cytochrome P450 superfamily in *Arabidopsis thaliana*. *BMC Plant Biol* 2008;8:47. [PubMed: 18433503]
45. Lamb DC, Guengerich FP, Kelly SL, Waterman MR. Streptomyces P450s as a model for drug discovery. *Expert Opin Drug Metab Toxicol*. 2006 in press.
46. Lamb DC, Waterman MR, Kelly SL, Guengerich FP. Cytochromes P450 and drug discovery. *Curr Opin Biotechnol* 2007;18:504–512. [PubMed: 18006294]
47. Lamb, DC.; Zhao, B.; Guengerich, FP.; Kelly, SL.; Waterman, MR. *Molecular Biochemistry and Biotechnology*. Horizon Press; 2010. Genomics of Streptomyces cytochrome P450. in press
48. Chun YJ, Shimada T, Sanchez-Ponce R, Martin MV, Lee L, Zhao B, Kelly SL, Waterman MR, Lamb DC, Guengerich FP. Electron transfer pathway for a streptomyces cytochrome P450. Cytochrome P450 105D5-catalyzed fatty acid hydroxylation in streptomyces coelicolor A3(2). *J Biol Chem* 2007;282:17486–17500. [PubMed: 17446171]

49. Katagiri M, Ganguli BN, Gunsalus IC. A soluble cytochrome P450 functional in methylene hydroxylation. *J Biol Chem* 1968;243:3543–3546. [PubMed: 4297783]
50. Guengerich FP. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* 2001;14:611–650. [PubMed: 11409933]
51. Yun CH, Ahn T, Guengerich FP, Yamazaki H, Shimada T. Phospholipase D activity of cytochromes P450 in human liver endoplasmic reticulum. *Arch Biochem Biophys* 1998;367:81–88. [PubMed: 10375402]
52. Zhao B, Lei L, Vassilyev DG, Lin X, Cane DE, Kelly SL, Yuan H, Lamb DC, Waterman MR. Crystal structure of albaflavone monooxygenase containing a moonlighting terpene synthase active site. *J Biol Chem* 2009;367:11–36719. [PubMed: 19858213]
53. Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 2006;78:779–787. [PubMed: 16448051]

Fig. 1.
Classical and “reverse” biochemical approaches to function [8].

Fig. 2.

Screening of P450s 2S1 and 2W1 for abilities to activate libraries of carcinogens using a *umu* bacterial assay [16]. DMBA: 7,12-dimethylbenz[a]anthracene, B[b]FA: benzo[b]fluoranthene, 3-MeO-AAB, 3-methoxy-4-aminoazobenzene, Trp-P-1: 3-amino-1,4-dimethyl-5H-pyrido[4,3-b]indole, MeIQ, 2-amino-3,5-dimethylimidazo[4,5-f]quinoline.

Fig. 3.

Strategies for elucidating substrates and reactions of proteins, including P450s [8]. One strategy utilizes transgenic animals, either missing an ortholog of the protein or with the protein in question over-expressed. Body fluids (or tissues) can be utilized for LC-MS analysis. Alternatively, the recombinant protein can be used *in vitro* with tissue extracts from wild-type (WT) and knockout (KO) animals (or human) tissues. See Figure 4 for an $^{18}\text{O}_2$ labeling strategy. LC-MS analysis is done and the data analyzed using principal component analysis (PCA), DoGEX (for ^{18}O data), MZmine, etc. LC-MS differences are used to provide leads for further study.

Fig. 4.

Use of the program DoGEX in the analysis of substrates of P450 7A1 with human liver extracts, as a validation of the approach [22]. Reconstituted P450 7A1 was incubated with NADPH and a human liver extract under an atmosphere of $^{18}\text{O}_2$ and $^{16}\text{O}_2$ (1:1, v/v). LC-MS analysis of the extract (derivatized with succinic anhydride) yielded the total ion current trace (part A). Analysis with DoGEX software yielded the profile shown in part B, where a dark band indicates a ratio of M and M+2 peaks near 0.95, the target value. The indicated region of the spectrum is expanded in part C and a three dimensional plot of the data is presented in part D, with the m/z 603/601 doublet at t_R 7 min. Confirmation as 7 α -hydroxycholesterol was done with an authentic standard.

Fig. 5. Preferred pathway of electron transport for P450 105D5 in *S. coelicolor* [48]. Fdx: ferredoxin; FDR: Fdx reductase.

Fig. 6. Identification of a substrate (QCP1) of P450 154A1 in *S. coelicolor*. Deletion of the *CYP154A1* gene yielded a phenotype with spores that were unstable to storage. Extracts from wild-type and *CYP154A1*^{-/-} cells were analyzed by LC-MS and compared using the program MZmine [19], which indicated the presence of a compound with *m/z* 259 (MH⁺) only in the *CYP154A1*^{-/-} cells. An extract of the *CYP154A1*^{-/-} cells was incubated with purified P450 154A1, cofactors (NADPH, O₂), and spinach Fdx and FDR. The *m/z* 259 peak was attenuated and a new peak appeared in the P450 154A1 reaction. The structure of QCP1 was elucidated using a battery of spectroscopic methods.

Table 1Classification of Human P450s Based on Major Substrate Class^a

Sterols	Xenobiotics	Fatty acids	Eicosanoids	Vitamins	Unknown
1B1	1A1	2I2	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				

^aThis classification is somewhat arbitrary, e.g. P450s 1B1 and 27A1 could be grouped in either of two different categories [4].

Table 2

Current status of human P450 orphans

P450	Site(s)	Possible substrates
2A7*		
2S1	Skin, liver (also trachea, lung, Stomach, small intestine, spleen) [23,24]	(Retinoic acid, carcinogens?) [16,24–27]
2U1	Thymus, brain [28,29]	Fatty acids (ω , ω -1) [28,29]
2W1	Tumors [16,30]	Carcinogens [16]
3A43	Brain [31]	Alprazolam [31]
4A22*		
4F11	Liver [32]	Fatty acids (ω -1, -2, -3...), β -hydroxy fatty acids [17,33]
4F22		
4V2	Eye [34]	Fatty acids (ω) [35]
4X1	Brain, liver [36]	Anandamide [36]
4Z1	Breast cancer [37]	Fatty acids (ω -2, -3, -4, -5) [38]
20A1	Brain, liver [39]	(Unknown) [39]
27C1	Liver, kidney, spleen, others [15]	(Unknown) [15]

* The literature is not clear that these are expressed.