Published in final edited form as:
Chem Res Toxicol. 2012 September 17; 25(9): 1964–1974. doi:10.1021/tx300281g.

Directed evolution reveals requisite sequence elements in the functional expression of P450 2F1 in *E. coli*

James B. Y. H. Behrendorff[†], **Chad D. Moore**[‡], **Keon-Hee Kim**[‖], **Dae-Hwan Kim**[‖], **Christopher A. Smith**[†], **Wayne A. Johnston**[§], **Chul-Ho Yun**[‖], **Garold S. Yost**[‡], and **Elizabeth M. J. Gillam**[†,*]

[†]School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Brisbane, Australia 4072

[‡]Department of Pharmacology and Toxicology, University of Utah, Salt Lake City, Utah, 84112-5820, USA

[§]School of Biomedical Sciences, The University of Queensland, St Lucia, Brisbane, Australia 4072

[‖]School of Biological Sciences and Technology and Hormone Research Center, Chonnam National University, Gwangju 500-757, Republic of Korea

Abstract

Cytochrome P450 2F1 (P450 2F1) is expressed exclusively in the human respiratory tract and is implicated in 3-methylindole (3MI)-induced pneumotoxicity via dehydrogenation of 3MI to a reactive electrophilic intermediate, 3-methyleneindolenine (3-MEI). Studies of P450 2F1 to date have been limited by the failure to express this enzyme in *Escherichia coli*. By contrast, P450 2F3, a caprine homologue that shares 84 % sequence identity with P450 2F1 (86 amino acid differences), has been expressed in *E. coli* at yields greater than 250 nmol/L culture. We hypothesized that a limited number of sequence differences between P450s 2F1 and 2F3 could limit P450 2F1 expression in *E. coli*, and that problematic P450 2F1 sequence elements could be identified by directed evolution. A library of P450 2F1/2F3 mutants was created by DNA family shuffling and screened for expression in *E. coli*. Three generations of DNA shuffling revealed a mutant (named JH_2F_F3_1_007) with 96.5 % nucleotide sequence identity to P450 2F1 and which expressed 119 ± 40 pmol (n = 3, mean ± SD) hemoprotein in 1 mL microaerobic cultures. Across all three generations, two regions were observed where P450 2F3-derived sequence was consistently substituted for P450 2F1 sequence in expressing mutants, encoding nine amino acid differences between P450s 2F1 and 2F3: nucleotides 191-278 (amino acids 65-92) and 794-924 (amino acids 265-305). Chimeras constructed to specifically test the importance of these two regions confirmed that P450 2F3 sequence is essential in both regions for expression in *E. coli* but that other non-P450 2F1 sequence elements outside of these regions also improved the expression of mutant JH_2F_F3_1_007. Mutant JH_2F_F3_1_007 catalyzed the dehydrogenation of 3MI to 3-MEI as indicated by the observation of glutathione adducts after incubation in the presence of

[*]Address correspondence to this author at School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Brisbane, Australia, 4072 e.gillam@uq.edu.au.

Supporting information available

S1 Nucleotide sequence data

S2 Peptide sequence data

S3 Oligonucleotide primers for amplification of P450 2F1/2F3 segments for chimera construction

S4 Fe(II)]CO vs. Fe(II) difference spectra of membrane fractions from P450 2F chimeras.

S5 Production of indigo pigment in P450 2F expression cultures.

S6 Predicted effects of mutations on protein stability

glutathione. The JH_2F_F3_1_007 protein differs from P450 2F1 at only 20 amino acids, and should facilitate further studies of the structure(activity relationships of P450s of the 2F subfamily.

Introduction

Cytochrome P450 2F1 is unusual among human P450s in that it is expressed in lung tissue and other tissues of the respiratory tract including trachea and nasal mucosa, with very little expression detected elsewhere.[1-3] Xenobiotic(metabolising enzymes of the respiratory tract are of particular interest because these tissues are directly exposed to exogenous bioactive compounds in the air including environmental pollutants and drugs administered by inhalation. In studies of P450 2F1 expressed in human cell lines, this enzyme has been identified as having a role in the bioactivation of pulmonary toxins including 3-methylindole (3MI), naphthalene, and styrene.[4, 5]

The mechanism by which P450 2F1 converts 3MI into highly mutagenic intermediates has received particular attention.[6] Many mammalian P450s oxidize 3MI, but the 2F subfamily is the only group of P450s known to exclusively dehydrogenate 3MI to the electrophilic intermediate 3-methyleneindolenine (3MEI) without any evidence of oxidation products.[7] This reactive intermediate has been demonstrated to form adducts with proteins, DNA, and glutathione, and to cause mechanism(based inactivation of P450 2F1 in lymphoblastoid microsomes.[8-13]

Investigations of P450 2F1 to date have relied on tissue samples from lung microsomes and recombinant expression in human cell lines.[2, 4, 6, 14, 15] Further, detailed studies of structure-function relationships of P450 2F1 have been hindered by the inability to express the folded P450 2F1 hemoprotein in *E. coli*. P450 2F3, from goat, shares ~ 84 % nucleotide sequence identity with the P450 2F1 coding sequence and has been expressed in *E. coli* in an N-terminally modified form in which the 5′ end of the coding sequence was replaced with that used by Barnes *et al.* for the expression of bovine P450 17 in *E. coli*.[16, 17] However, this and other N-terminal modifications detailed in this work proved unsuccessful in obtaining detectable hemoprotein expression in *E. coli*. Therefore, it was hypothesized that a limited number of sequence differences between P450s 2F1 and 2F3 determine the success or failure to express folded P450 hemoprotein in *E. coli*.

The overall objective of the present study was to determine the basis of the differential expression of P450 2F1 and P450 2F3 in *E. coli*. Our specific aim was to investigate the usefulness of DNA family shuffling to recombine P450 2F1 with the P450 2F3 sequence modified for expression in *E. coli*[18] in an attempt to identify mosaic sequences that expressed as folded hemoprotein in *E. coli*, and to use a directed evolution approach to determine which elements of P450 2F1 sequence prevented successful expression of the native protein.

Materials and methods

Chemicals

Enzymes for molecular biology were purchased from New England Biolabs Inc. (Beverly, MA, USA). SYBR Safe DNA stain and the *Escherichia coli* strain DH5α F’IQ used in expression trials were purchased from Invitrogen (Mulgrave, VIC, Australia). Bacto tryptone and yeast extract were purchased from BD Biosciences (Franklin Lakes, NJ, USA). 3MI, indole(3-carbinol (I-3-C), 3-phenyloxindole (3POI), glutathione (GSH) and NADPH were purchased from Sigma(Aldrich (St. Louis, MO, USA). 3-Methyloxindole (3MOI) was a generous gift from Dr. James Carlson (Washington State University, WA, USA). The

authentic standards for 3-glutathionyl(*S*(methylindole (GS-A1) were synthesized as previously reported for 3-[(*N*(acetylcystein-*S*(yl)methyl]indole (3MINAC)[19]. All other chemicals were obtained from local suppliers at the highest quality available.

Plasmids

The native P450 2F1 coding sequence in the pUC9 plasmid (pUC9/2F1) was generously provided by Dr. Frank Gonzalez (NCI, National Institutes of Health (Bethesda), USA).[2] The P450 2F3 expression plasmid, pCW/2F3, was described in an earlier paper.[17] The plasmids pCW/2C19His/hNPR and pCW/1A2/hNPR, which were used here as vector sources for creating bicistronic expression constructs, were reported previously.[20, 21] The expression vector encoding the GroEL/GroES chaperone system (pGro7) was donated by Professor K. Nishihara (HSP Research Institute, Kyoto, Japan)[22] and the pCW/hNPR expression vector encoding hNPR alone was obtained from Professor F.P. Guengerich (Vanderbilt University, TN, USA).

P450 2F1 and P450 2F3 expression constructs

Bacterial expression plasmids containing cDNAs for P450 2F1 and P450 2F3 in a bicistronic format with human NADPH(cytochrome P450 reductase (hNPR)[21] (designated pCW/<P450>/hNPR) were prepared as follows. The native P450 2F1 sequence was amplified from the pUC9/2F1 plasmid with the following primers: 5′-TTA GGA GGT CAT ATG GAC AGC ATA AGC ACA GCC ATC TTA CTC CTG CTC CTG GCT (3′ -forward primer, which incorporates an NdeI recognition site (underlined) into the start codon) and 5′- TCC TGG GCG TCT AGA TTA GGT CGA CGC GCG TGG GCG CAG GCA CAG CTG GAA A -3′ (reverse primer, which incorporates an XbaI recognition site (underlined) immediately downstream from the stop codon). The bicistronic expression construct was made by ligating the amplified P450 2F1 sequence into the NdeI and XbaI restriction sites of the pCW/2C19His/hNPR expression vector replacing the P450 2C19His coding sequence. The bicistronic expression plasmid for P450 2F3, pCW/2F3/hNPR, was constructed similarly by sub(cloning the NdeI(XbaI fragment from the monocistronic expression plasmid for P450 2F3 into similarly cut pCW/1A2/hNPR.

Codon optimisation and N terminal modification of P450 2F1

P450 2F1 with codon usage optimized for *E. coli* was designed using the DNAworks v2.0 (http://helixweb.nih.gov/dnaworks/), synthesized by recursive overlap(extension PCR of oligonucleotides[23] and verified by DNA sequencing. Nucleotide sequences for native P450 2F1, the previously expressed P450 2F3, and the codon-optimized P450 2F1 (2F1#1) are provided in the supporting information (S1). Three mutants of the codon-optimized P450 2F1 with modified N-termini were constructed: in variant #2 the second codon was replaced with GCT, coding for alanine; in variant #5, the substitutions reported by Barnes *et al.*[16] were introduced with removal of codons 10-13; and in variant #6, the same changes were introduced as for P450 2F1#5 except that codons 10-15 were removed. Modified N-terminal sequences were amplified from the codon-optimized P450 2F1 by PCR using mutagenic primers, then cloned into the codon-optimized P450 2F1 by replacement of sequence between the NdeI recognition site at the beginning of the coding sequence and an EcoRI recognition site at nucleotide position 280. The following mutagenic forward primers were used (NdeI site and start codon underlined, and mutations in bold type): 2F1#2, 5′-TTA GGA GGT CAT ATG G**CT** TCT ATA TCT ACT GCT ATC TTA CTC CTG CTC CTG GCT -3′; 2F1#5, 5′- TTA GGA GGT CAT ATG **GCT CTG TTA TTA GCA GTT TTT** CTG GCT CTC GTC TGT 3′; 2F1#6, 5′- TTA GGA GGT CAT ATG **GCT CTG TTA TTA GCA GTT TTT** CTC GTC TGT CTG CTC CTG -3′. A generic reverse primer

incorporating the EcoRI site (underlined) was used to amplify all three modified N-termini (5′- ATA ATC GCC ACG GCC GCT GAA TTC TTC GCC CTG ATC -3′).

Construction of shuffled libraries

DNA family shuffling was performed according to the general method of Crameri *et al.*[18] except that parental DNA plasmids were fragmented using restriction enzymes[24] according to the general method described previously.[25, 26] Parental cDNAs (pCW/2F1/hNPR, containing the native P450 2F1 nucleotide sequence, and pCW/2F3/hNPR, containing the expressed P450 2F3 sequence (see supporting information)) were fragmented using the following combinations of restriction endonucleases: (i) HpaII plus MlyI and (ii) BamHI plus NlaIII. Each restriction digest (100 YL total digest volume) contained equal quantities of each parental plasmid DNA to a total of 100 Zg with 10 U of each enzyme as appropriate and was undertaken using buffer conditions recommended by the enzyme manufacturer, at 37 °C for 4 h. DNA fragments < 1000 base pairs in length were purified using a method modified from Kadokami and Lewis.[26, 27] Briefly, for each restriction digest, products were stained with 1000X SYBR Safe DNA stain for 15 min at room temperature and separated by electrophoresis on a 1 % agarose gel in parallel with a 100 bp DNA marker ladder (New England Biolabs, Ipswich, MA, USA). DNA fragments > 1000 bp in length were excised from the agarose gel and replaced with a fresh block of 1 % agarose. The polarity on the electrophoresis tank was reversed, and DNA fragments < 1000 bp in length were electrophoresed through the fresh block of agarose until they converged as a single band. Single bands containing DNA fragments < 1000 bp were excised and DNA was extracted using QIAquick Gel Extraction kit (QIAGEN, Doncaster, VIC, Australia) columns according to the manufacturer's instructions. Equal quantities of fragments from each of the two restriction digests were combined in a primerless reassembly PCR containing 50 mM KCl, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, and 0.5 U/ZL *Taq* polymerase in 20 mM Tris(Cl (pH 8.4). Cycling conditions were modified from Abécassis *et al.*[28] and consisted of an initial hot(start at 94 °C while *Taq* polymerase was added followed by 35 cycles of: denaturation at 94 °C for 30 s; a 24-minute hybridization step where the temperature was decreased from 65 °C to 41 °C in 1 °C increments every 60 s; and an extension step at 72 °C for 90 s. This was followed by a final extension at 72 °C for 7 min. Shuffled sequences (5 ZL of products from the reassembly PCR) were amplified in a second PCR with flanking primers. The forward primer (5′ GGAAACAGGATCCATCGATGCTTAGGAGGTCATATG 3′) was designed to anneal immediately upstream of and including the NdeI site and start codon (underlined in the sequence above), while the reverse primer (5′ GAGGTCAATGTCTGAATTTTGGTGAACTCGGGGAC 3′) was designed to anneal downstream of the P450 cDNA, 150 bp into the hNPR open reading frame. Reaction conditions consisted of an initial hot(start at 94 °C while *Taq* polymerase was added followed by ten cycles of: denaturation at 94 °C for 30 s, hybridization at 60 °C for 30 s, and an extension step at 72 °C for 90 s. This was followed by a further 25 cycles of the same conditions except that ten seconds were added on to the extension step time in each successive cycle, and a final extension at 72 °C for 7 min. Amplified products were cloned into the NdeI and XbaI sites of the pCW/2C19His/hNPR bicistronic expression vector replacing the P450 2C19 cDNA. The resultant library of mutant sequences was transformed into competent DH5αF'IQ cells already containing the pGro7 chaperone vector as described previously.[29] Subsequent libraries were created using the same method except that selected mutant sequences were shuffled with the pCW/2F1/hNPR parental sequence. The second-generation library was produced by shuffling pCW/2F1/hNPR with mutants JH_2F_F1_1_064, JH_2F_F1_1_207, and JH_2F_F1_1_217, and the third(generation library was produced by shuffling pCW/2F1/hNPR with mutants JH_2F_F2_1_077 and JH_2F_F2_1_156, as detailed further in the Results section.

Expression of P450 2F mutants and screening for P450 hemoprotein by Fe(II))CO vs. Fe(II) difference spectroscopy

P450 2F mutants from shuffled libraries were expressed in 1 mL micro(aerobic cultures as described previously[29] with the exception that expression cultures were incubated at 28 °C. All expression trials included co(expression of the GroES/EL chaperone encoded in the pGro7 plasmid. *E. coli* transformed with plasmids encoding the parental P450 sequences in bicistronic format with human NADPH(P450 reductase (pCW/2F1/hNPR, pCW/2F3/hNPR) and the pCW/hNPR construct were included in triplicate expression cultures as positive and negative controls. Cultures were screened for expression of hemoprotein using the whole(cell P450 Fe(II)]CO vs. Fe(II) difference spectroscopy method described previously.[30] A sensitivity limit of 50 pmol P450 in 1 mL cultures was determined for the whole(cell P450 screening assay on the basis of the maximum apparent signal observed in negative controls.

Catalytic activity towards indole hydroxylation was monitored in selected mutants and chimeras by observing indigo formation in expression cultures grown in 50 mL modified TB medium under aerobic conditions in 500 mL Erlenmeyer flasks for 48 h, as described previously.[31]

Preparation and characterization of subcellular fractions

P450 2F1 N-terminal variants and selected mutants were expressed in 50 mL cultures and subcellular fractions prepared as described previously.[32] Total protein concentrations in subcellular fractions were determined using the Pierce bicinchoninic acid protein assay reagent kit (Thermo Fisher Scientific, Rockford, IL, USA). Proteins were separated by SDS-PAGE on a 7.5 % acrylamide gel and stained with ammonical silver.[33, 34] Sample volumes were loaded such that insoluble and membrane fractions were equivalent to the material derived from 4 Zg wet mass of harvested *E. coli* cells, and cytosolic fractions were equivalent to 2 Zg wet mass of cells.

Construction of P450 2F1/2F3 chimeras

Five chimeric sequences, as described in figure 5, were constructed to test the importance of defined regions of P450 2F3 for successful hemoprotein expression. The desired segments of P450 2F1 and P450 2F3 were amplified by conventional PCR using oligonucleotide primers designed such that the ends of adjacent segments were complementary. The appropriate segments for the construction of each individual chimera were then combined by overlap extension PCR without primers using the reaction conditions as described above for DNA shuffling**,** and full(length chimeric sequences were amplified with flanking primers and cloned into the pCW/hNPR bicistronic expression vector. Primers used to amplify chimera segments are described in the supporting information (S3).

Sequence analysis of shuffled mutants

Shuffled mutants that expressed P450 hemoprotein at levels greater than the detection limit were selected for DNA sequencing. Plasmid DNA was prepared using a QIAprep Spin Miniprep Kit (QIAGEN, Doncaster, VIC, Australia) according to the manufacturer's instructions. Automated sequencing was performed at the Australian Genome Research Facility (Brisbane node; St Lucia, QLD, Australia). Shuffled mutant sequences were analyzed by alignment to the P450 2F1 and P450 2F3 parental sequences (Vector NTI v. 10.3.0, Invitrogen). Graphical comparisons between shuffled mutants were produced automatically using software described previously[25] to identify which sequence elements were derived from each of the relevant parental sequences.

Structural modelling and thermodynamic stability predictions

Structural predictions of protein sequences were made using the Robetta Full(chain Protein Structure Prediction Server (available at: http://robetta.bakerlab.org).[35] Data files generated were converted to graphical models using the PyMOL Molecular Graphics System, version 1.3r1 (Schrödinger, LLC, available at: http://www.pymol.org).

The effect of amino acid changes on thermodynamic stability were assessed using the Eris Protein Stability Prediction server (available at: http://troll.med.unc.edu/eris)[36] in fixed backbone mode.

Incubations with 3MI

Incubations contained 2.2 mg of protein from bacterial membrane preparations, potassium phosphate buffer (50 mM, pH 7.4), 3MI (100 YM), and GSH (3 mM) in a final volume of 500 YL. Samples were preincubated at 37 °C for 5 min, and the reaction was initiated by the addition of NADPH to 2 mM. The reaction was allowed to proceed for 20 min at 37 °C, and then was terminated by the addition of 500 YL of cold acetonitrile containing 1 pmol of the internal standard, 3POI. Samples were mixed by vortexing, and then centrifuged at 21000 × *g* for 15 min to remove the protein. 3MI and metabolites were extracted using C-18 Sep-Pak cartridges (Waters, Taunton, MA). The resulting eluate was concentrated to dryness by evaporation under nitrogen and reconstituted in 10 % acetonitrile (v/v) for analysis via liquid chromatography(mass spectrometry (LC/MS). LC/MS was conducted with a Finnigan TSQ Quantum AM mass spectrometer (Thermo Fisher Scientific, Waltham, MA), coupled with an LC system consisting of a PerkinElmer Series 200 pumps and autosampler (PerkinElmer, Waltham, MA). Chromatography was conducted with a Phenomenex Luna 5Y C18 (150 × 2.00 mm) reverse(phase column (Phenomenex Inc., Torrance, CA). The mobile phase consisted of solvent A: acetonitrile and solvent B: 0.1 % formic acid (v/v). The mobile phase was 10 % solvent A for 1 min, then increased to 50 % solvent A over 9 min, and finally increased to 60 % solvent A over 10 min. Electrospray ionization (ESI) with positive ionization was utilized for detection. The spray voltage was set to 4000 V, capillary temperature to 295 °C, tube lense offset to 61 V, auxiliary gas (nitrogen) of 50 units, and sheath gas (nitrogen) to 50 units. MS/MS by collision-induced dissociation was achieved using argon gas with collision energy of 26 eV, and Q1 and Q3 peak width set to 0.7 FWHM. Identification of 3MI metabolites was based on product-ion spectra described by D'Agostino *et al.*[12]

Results

N terminal modifications and codon optimisation of P450 2F1

A synthetic gene was made in which codon use was optimized for *E. coli* expression throughout the entire coding sequence of P450 2F1. In addition, the following three N-terminal variants were made from the codon-optimized form: the native peptide sequence except with GCT (encoding alanine) as the second codon (2F1#2, N-terminal peptide sequence MASISTAILLLLLALVC…); replacement of amino acids 10-13 with the N-terminal described by Barnes *et al.*[16] (2F1#5, MALLLAVFLALVC…); replacement of amino acids 10-15 with the Barnes *et al.*[16] sequence (2F1#6, MALLLAVFLVC…). The purpose of changing the second codon to GCT was to provide a favourable second(codon for translation in *E. coli,*[37] while other N-terminal modifications were based on published successes with other mammalian P450 enzymes.[38] However, no P450 hemoprotein was detectable by Fe(II)]CO vs. Fe(II) difference spectroscopy in expression cultures of the codon-optimized P450 2F1 or with any of the N-terminal variants described.

Detection of hemoprotein in first generation shuffled mutants

A library of 252 mutants (corresponding to an archive of three 96-well plates with one row in each plate reserved for controls) was produced by shuffling P450 2F1 and P450 2F3 coding sequences and was screened for expression of folded P450 hemoprotein. Twenty(four mutants were identified that expressed greater than 50 pmol P450 in 1 mL cultures (figure 1), of which 13 were randomly selected for DNA sequencing. Two of these mutants (JH_2F_F1_1_064 and JH_2F_F1_1_185) were later reclassified as non-expressing as they were not reproducibly expressed above the 50 pmol sensitivity limit. Of the expressing mutants, two were reassembled P450 2F3 sequences while the remaining nine sequences were shuffled P450 2F1/2F3 mutants showing 6 ± 2 recombinations and 1 ± 1 point mutations per mutant (mean ± SD) (Figure 2a).

Directed evolution of a P450 2F1 like mutant with successful expression in *E. coli*

Two mutants that expressed successfully and contained a large proportion of P450 2F1-derived sequence were selected for further shuffling. Mutant JH_2F_F1_1_207 expressed 190 pmol/mL folded hemoprotein and shared 89.6 % nucleotide sequence identity with P450 2F1, while JH_2F_F1_1_217 expressed at 80 pmol/mL and was 91.5 % identical to P450 2F1. A third mutant (JH_2F_F1_1_064) which shared 90.2 % identity with P450 2F1 was included in the shuffling reaction, as was the full(length native P450 2F1 sequence. Mutant JH_2F_F1_1_064 had been selected based on apparent expression of P450 hemoprotein close to the detection limit in the first tier screen but subsequent screening failed to confirm detectable P450 expression. Plasmids encoding the three mutants described plus P450 2F1 were combined in equal proportions and shuffled as described in the methods section to produce a second generation of shuffled mutants.

From the second(generation shuffled library, 252 mutants were screened and 95 (37.7 % of the library) were identified that expressed greater than 50 pmol folded P450 (Figure 3). Second generation mutants that expressed above the detection limit had an expression level of 80 ± 30 pmol/mL culture (mean ± SD). Sequence analysis of seventeen expressing mutants revealed 91 ± 2 % (mean ± SD) identity to P450 2F1 (Figure 2b). This represented a statistically significant increase in mean identity to P450 2F1 compared to first(generation shuffled mutants (unpaired Student's *t*(test, $p < 0.001$).

Two second(generation mutants were chosen for a third round of DNA shuffling with P450 2F1 to produce a third(generation library using the same method as for the first two generations. Mutant JH_2F_F2_1_077 was included as it was the second(generation mutant showing the highest sequence identity to P450 2F1 (95.4 % P450 2F1 identity, expression of 110 ± 10 pmol/mL (n = 3, mean ± 1 SD)). Mutant JH_2F_F2_1_156, which showed lower identity to P450 2F1 but greater hemoprotein expression (92.1 % P450 2F1 identity, 220 ± 20 pmol/mL (n = 3, mean ± 1 SD)), was included in the interest of avoiding evolutionary dead-ends.

Screening of the third(generation library revealed mutant JH_2F_F3_1_007 (figure 2c), which exhibited 96.5 % nucleotide identity to P450 2F1 and expressed 120 ± 40 pmol/mL culture (n = 3, mean ± SD) detectable hemoprotein, as determined by Fe(II)]CO vs. Fe(II) difference spectroscopy. Mutant JH_2F_F3_1_007 contained two spontaneous missense mutations at nucleotides 374 and 565 (encoding the mutations R125K and D189Y, respectively) and a silent mutation at nucleotide 522. Across all three generations of shuffled mutants, P450 2F3 sequence was observed consistently between nucleotides 191-278 and between nucleotides 794-924 (figure 2) in mutants that expressed successfully. These two regions, referred to hereafter as the first and second regions of interest respectively, spanned amino acid residues 65-92 (region 1) and 265-305 (region 2). A total of nine amino acids

differ between P450s 2F1 and 2F3 in these two regions, at positions 65, 69, 92, 266, 274, 277, 284, 301 and 304 (figure 4). A single exception was observed in mutant JH_2F_F2_1_001 which contained P450 2F1 sequence from nucleotide 902 (amino acid residue 301) onward. Nucleotide and translated peptide sequence data for P450 2F1, P450 2F3, and mutant JH_2F_F3_1_007 are provided in the supporting information (S1).

The 5′ end of the native P450 2F1 coding sequence was found to be compatible with recombinant expression in *E. coli* as demonstrated by successful expression of shuffled mutants JH_2F_F1_1_217, JH_2F_F2_1_077, JH_2F_F2_1_158, JH_2F_F2_1_169, and JH_2F_F3_1_007. No statistically significant difference could be found between the mean expression levels of mutants with and without P450 2F1 sequence in the first ten codons.

Analysis of the two identified regions of interest

Five chimeric P450 2F1/2F3 mutants (denoted 2FcAexp, 2FcA, 2FcBexp, 2FcB and 2FcC and described in figure 5) were constructed to test the hypothesis that detectable P450 2F1 protein expression is hindered by sequence elements that occur in the two regions where P450 2F1 sequence was not observed in expressing shuffled mutants. The expression of the five chimeras was tested in 1 mL cultures (n = 4) incubated for either 72 h in microaerobic conditions or 48 h in aerobic conditions. None of the chimeras produced sufficient P450 hemoprotein for detection by Fe(II)]CO vs. Fe(II) difference spectroscopy in intact cells in either expression trial (supporting information S4), before or after concentration of cultures, a step shown to enhance the sensitivity of P450 detection in this assay.[30] No accumulation of indigo pigment could be observed for the chimeric forms in 1 mL aerobic cultures. However, minor quantities of indigo were observed in 50 mL aerobic expression cultures of 2FcAexp, but no other chimeras. Aerobic 50 mL cultures of P450 2F3 and mutant JH_2F_F3_1_007 produced large quantities of indigo and turned dark blue within 48 h, but no indigo was observed in P450 2F1 expression cultures (supporting information S5). No clear, characteristic P450 peak could be identified in membrane preparations from 50 mL aerobic cultures of any of the chimeras.

Analysis of apoprotein expression in mutants

Subcellular fractions, namely the 10,000*g* pellet (“insoluble”), 180,000*g* pellet (“membrane”), and 180,000*g* supernatant (“cytosol”), were prepared from expression cultures of cells transformed with pCW/2F1/hNPR, pCW/2F3/hNPR, shuffled mutant pCW/JH_2F_F3_1_007/hNPR, or the negative control, pCW/hNPR. Full(length P450 2F3 protein was observed in each of the three fractions, showing particular abundance in the cytosolic fraction (figure 6). P450 2F1 protein was observed in insoluble and cytosolic fractions but not in the membrane fraction. In the cytosol, P450s 2F1 and 2F3 were present in roughly equal abundance but a larger proportion of the total P450 2F1 protein was present in the cytosolic fraction. Carbon monoxide(binding difference spectroscopy revealed that P450 2F3 in the cytosolic fraction was properly folded, whereas no peak at 450 nm could be observed for cytosolic P450 2F1. The translation product of mutant JH_2F_F3_1_007 was observed in all three fractions but at lower abundance than P450 2F3, and a typical P450 hemoprotein spectrum was only observed in the membrane fraction. The five chimeras constructed to test specific sequence regions of interest expressed detectable apoprotein only in the insoluble and cytosolic fractions (data not shown).

In silico analysis of mutant JH_2F_F3_1_007

A homology model of mutant JH_2F_F3_1_007 was produced using the Robetta Full(chain Protein Prediction Server by performing a structural alignment with P450 2A13, the most similar sequence to P450 2F1 for which a crystal structure is available (51 % amino acid identity, PDB structure 2P85). In this model, the first region of interest (amino acids 65-92)

encompassed the first and second strands of the β1 sheet, and the B' helix, while the second region (amino acids 265-305) encompassed the H helix, the H(I loop, and approximately half of the I helix including the characteristic kink that occurs where the I helix passes the heme centre (figure 7).

The Eris protein stability prediction algorithm[36] was used to assess residues of interest highlighted by this study. Within the first region of interest, substitution of individual P450 2F1 residues into the JH_2F_F3_1_007 model, was predicted to decrease the thermodynamic stability (increase in cG) in all instances (residues 65, 69, and 92; supporting information S6). When the entire first region of interest was substituted with that found in P450 2F1, an increase in cG of > 10 kcal/mol was predicted. Within the second region of interest the most destabilising P450 2F1 substitutions were E301K (cc$G$ 5.14 kcal/mol) and G304S (cc$G$ 2.62 kcal/mol). Other differences between P450s 2F1 and 2F3 in the second region of interest were not predicted to be destabilising (supporting information S6). When the entire second region of interest in the JH_2F_F3_1_007 model was substituted with P450 2F1 sequence, a net increase in cG of 7.42 kcal/mol was predicted. Reversion of the spontaneous mutations K125 and Y189 to the residues found in P450 2F1 (arginine and aspartate), respectively, was predicted to cause an increase in cG of 1.38 kcal/mol.

Metabolism of 3MI by mutant JH_2F_F3_1_007

E. coli membranes from expression cultures of mutant JH_2F_F3_1_007 were incubated with 3MI and the incubations were analyzed for glutathionyl adducts indicative of 3MEI formation (characteristic of P450s 2F1 and 2F3) as well as for metabolites that have been previously identified as products of oxidative 3MI metabolism by other P450s. Three NADPH-dependent 3MI metabolites were detected by LC/MS at 8.73, 9.70 and 10.94 min, with m/z values of 437, 130 and 148 respectively (figure 8). The major metabolite at 8.73 min was found to be the GSH adduct GS(A1, which results from the dehydrogenation of 3MI to the reactive electrophile 3MEI. The identity of GS(A1 was confirmed by comparison of retention time and MS/MS fragmentation with that of an authentic GS(A1 standard. The two minor hydroxylated 3MI metabolites detected at 9.70 and 10.94 min were identified as I-3-C and 3MOI by comparison with authentic standards. The rates of product formation were determined and are reported in table 1.

Discussion

A number of modifications are commonly made to the 5′ end of mammalian P450 coding sequences to achieve expression in *E. coli*. These are designed to minimize mRNA secondary structure formation, to alter codon usage to that favoured by *E. coli,* and/or to truncate the membrane(associated N-terminal region.[38] Substitution of native sequence with the N-terminal used by Barnes *et al*.[16] was successful in producing P450 2F3 hemoprotein in *E. coli*,[17] but failed when applied to P450 2F1 (K. Skordos and G. S. Yost, unpublished data). In the current study, codon optimization of the entire P450 2F1 coding sequence to suit *E. coli* was also unsuccessful, as were N-terminal modifications of the codon-optimized sequence.

Since P450 2F1 protein was observed in the cytosolic subcellular fraction, it is unlikely that poor translation efficiency was the primary problem in producing P450 2F1 in *E. coli.* Rather, the absence of an observable peak at 450 nm in Fe(II)]CO vs. Fe(II) difference spectroscopy suggests that the P450 2F1 protein that was expressed was misfolded or lacked the cysteineligated heme. Moreover, the paucity of P450 2F1 protein in the membrane fraction compared with P450 2F3 suggested a failure of P450 2F1 to associate stably with *E. coli* membranes.

Shuffling of P450s 2F1 and 2F3 was performed with the aim of identifying which sequence regions were associated with successful expression of P450 2F3 compared to P450 2F1. A similar approach had been used previously to reveal which sequence differences in the tomato Cf-4 and Cf-9 proteins were responsible for the recognition of different protein interaction partners.[39] Directed evolution was employed to produce P450 2F mutants that expressed P450 hemoprotein with incremental increases in P450 2F1 sequence content. Across all three generations of mutants produced, two regions were consistently observed where P450 2F1 coding sequence did not occur in mutants that expressed above the detection limit of the assay. Three rounds of directed evolution produced shuffled mutant JH_2F_F3_1_007 which has a P450 2F1-like coding sequence (96.5 % nucleotide sequence identity to P450 2F1; 20 amino acid differences) and was expressed in *E. coli* as properly folded hemoprotein. Additionally, analysis of subcellular fractions by SDS(PAGE revealed that the mutant JH_2F_F3_1_007 protein was associated with the *E. coli* membrane.

None of the five chimeric sequences designed to specifically test the two sequence regions of interest produced sufficient hemoprotein for detection by Fe(II)]CO vs. Fe(II) difference spectroscopy. However the chimera that contained P450 2F3 sequence at residues 65-92 and 266-313 (chimera 2FcAexp) produced a very small amount of indigo in large scale cultures, an activity also seen in aerobic cultures of P450 2F3 and mutant JH_2F_F3_1_007, suggesting very low or unstable hemoprotein expression when P450 2F3 sequence was present in both of these regions. The conservation of the two identified regions of interest across shuffled mutants that expressed as detectable hemoprotein plus the expression of chimera 2FcAexp demonstrated the importance of these sequence regions to stable P450 2F folding in *E. coli*. However, the poor expression of P450 2FcAexp compared with JH_2F_F3_1_007 demonstrates that the presence of other P450 2F3 sequence elements can greatly improve expression. In addition, the observation that chimera 2FcAexp but not 2FcA produced indigo indicates that at least one residue out of Gln266, Glu274, and Glu277 had a deleterious effect on expression and/or indole hydroxylase activity.

A multiple sequence alignment of P450 2F1, P450 2F3, shuffled mutant JH_2F_F3_1_007, and the peptide sequences of all P450s from family 2 for which crystal structures have been obtained (figure 3) was used to identify residues within the regions of interest that may be important to expression in *E. coli*. Within the first region of interest (amino acids 65-92), Met65 in P450 2F1 differed from all other aligned P450s. All other sequences contained valine at the equivalent position, with the exception of P450 2R1, which contained isoleucine. The other two residues that differed between P450s 2F1 and 2F3 (69 and 92) did not appear to be highly conserved and the amino acids found at these positions in both P450 2F1 and JH_2F_F3_1_007 sequences occurred frequently in other isoforms.

Within the second region of interest (amino acids 265-305) six residues differed between P450 2F1 and JH_2F_F3_1_007. Two of these residues were highly conserved in non-P450 2F1 sequences: 266 (glutamine in P450 2F1 versus aspartate in all other isoforms besides P450 2D6, where glutamate occurred at the equivalent position) and 301 (lysine in P450 2F1, glutamate in all other isoforms except P450 2D6, where valine was observed at the equivalent position).

The first region of interest may play an important role in membrane association, as residues of interest 65 and 69 are located in the β1-1 sheet, but the predictive power of modelling is constrained in this instance as these residues lie close to the N-terminal region of the protein that is disordered in crystal structures. Studies of P450 2C5 showed that residues 60-69 in that protein form a hydrophobic membrane(associated surface.[40] These residues align to positions 59-68 in P450 2F1, indicating that Met65 (and possibly His69) in P450 2F1 may influence membrane association in *E. coli*. P450 2F1 residues in the first region of interest

were predicted to be strongly destabilising. None of the chimeras constructed to specifically test the importance of region 1 showed P450 apoprotein in the membrane fraction after SDS-PAGE. Therefore other P450 2F3 residues or missense mutations in JH_2F_F3_1_007 must also play a role in rescuing the membrane(association phenotype in *E. coli*.

Within the second region of interest, residues that differ between P450s 2F1 and 2F3 are generally not strictly conserved, the notable exception being position 301 in P450 2F1. In *in silico*-predicted structures, this residue occurs at approximately the mid(point of the I-helix projecting upward toward the F-helix, and may interfere with side chain packing in the opposing F-helix. Furthermore, Lys301 was predicted to be the most destabilising feature of the second region of interest. P450 2F3 and mutant JH_2F_F3_1_007 have Glu at this position with an opposite charge and slightly shorter side chain. Several studies have reported the influence of individual residues of the H- and I-helices on substrate specificity in other P450s,[41-43] but no studies to date have examined the importance of the residues that differ between P450s 2F1 and 2F3 in this region. The reason why residues predicted here to be destabilizing may be tolerated in mammalian cells but not in *E. coli* is unclear, but may relate to differences between prokaryotic and eukaryotic systems such as in the lipid composition of the endoplasmic reticulum or the activity of chaperone proteins.

The failure of the chimeras to show detectable P450 hemoprotein expression argues for the cumulative effect of multiple amino acid changes in causing the diminution of expression in P450 2F1 relative to P450 2F3, and suggests that expression is unlikely to be rescued by making individual site(directed changes to P450 2F1. However the effect of single mutations may be assessed in the P450 2F3 background; studies towards this goal are in progress.

The shuffled mutant JH_2F_F3_1_007 provides a P450 2F1-like coding sequence that is readily expressed in *E. coli* as folded hemoprotein. Structurally it should serve as a closer approximation to native P450 2F1 than P450 2F3, as the mutant differs from native P450 2F1 by only 20 amino acids and expresses with the native N-terminal sequence (P450 2F3 differs from P450 2F1 in more than 80 amino acids). The use of the native P450 2F1 N-terminal sequence was preferable because N-terminal modifications may exert subtle effects on substrate binding or reconstitution with redox partners.[44, 45]

When mutant P450 JH_2F_F3_1_007 was incubated with 3MI, the major metabolite observed was the GS(A1 glutathione adduct, which results from 3MEI production. The *N*-acetylcysteine adduct 3MINAC was also detected, confirming the production of 3MEI (data not shown). Two minor hydroxylated metabolites, I-3-C and 3MOI, were also detected (figure 8), although rates of product formation for these metabolites were less than 3 % of that for GS(A1 (table 1). Wild type P450 2F1 has previously been shown to predominately bioactivate 3MI to 3MEI. While I-3-C and 3MOI are not produced in significant amounts by P450 2F1[4] or P450 2F3[17], they are produced from 3MI by P450s 1A1, 1A2, 1B1, 2A13 and 2E1.[7, 12] P450s 1B1 and 2E1 exclusively metabolize 3MI to monooxygenated products, while a mixture of monooxygenated products and 3MEI are observed with P450s 1A1, 1A2 and 2A13.[7, 12] The change in the product profile observed with mutant JH_2F_F3_1_007 may have resulted from spontaneous point mutations or a more complex remodeling of the active site caused by recombining the two parental proteins. Kartha *et al.* [46] reported that various single and double point mutations of P450 2F3 located within or near the substrate recognition sites [47] of P450 2F3 were sufficient to increase monooxygenase activity towards 3MI. The introduction of 3MI oxygenase activity could have been due subtle changes caused by recombination of segments of P450s 2F1 and 2F3, or the random mutations observed. Spontaneous missense mutations that arose during construction of shuffled mutant JH_2F_F3_1_007 resulted in two amino acid residues that were not derived from either P450 2F1 or P450 2F3 genetic material (Lys125 and Tyr189, which occur as Arg125 and

Asp189 in both P450s 2F1 and 2F3). Of these point mutations, Lys125 occurs close to the B(C loop region that encloses the active site and so modifications here could alter the packing of side-chains in the active site cavity, affecting substrate specificity. Reversion of these residues to the native amino acids was predicted to have a mild destabilizing effect.

In conclusion, shuffled mutant JH_2F_F3_1_007 provides a new model for studying the structure(activity relationships of P450 2F1. At 96.5 % sequence identity, this mutant has the most similar amino acid sequence to P450 2F1 of any P450 successfully expressed in *E. coli*. Although the mutant produced hydroxylated metabolites of 3MI that were not detected in incubations of 3MI with native P450 2F1, these reaction products were only produced in minor quantities and the dominant reaction remains dehydrogenation of 3MI to 3MEI. Finally, this study highlights the utility of the DNA family shuffling method as a tool for determining which sequence elements are responsible for binary functional differences between highly homologous proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sincere thanks are extended to Dr. Frank Gonzalez (NCI, National Institutes of Health (Bethesda), Professor K. Nishihara (HSP Research Institute, Kyoto, Japan) and Professor F.P. Guengerich (Vanderbilt University, TN, USA) for donating plasmids used in this study and to Dr. James Carlson (Washington State University, WA, USA) for the gift of 3MOI.

Funding information This study was supported by National Heart, Lung, and Blood Institute Grant HL60143 from NIH (GSY) and an Australian Postgraduate Award (to JBYHB).

Abbreviations

3MEI	3-methyleneindolenine
3MI	3-methylindole
3MINAC	3-[(*N*-acetylcystein-*S*-yl-methyl]indole
3MOI	3-Methyloxindole
3POI	3-phenyloxindole
ESI	Electrospray ionization
GS-A1	3-glutathionyl-*S*-methylindole
GSH	reduced glutathione
hNPR	human NADPH-cytochrome P450 reductase
I-3-C	indole-3-carbinol
LC-MS	liquid chromatography-mass spectrometry
P450	cytochrome P450, heme-thiolate protein P450

References

(1). Bieche I, Narjoz C, Asselah T, Vacher S, Marcellin P, Lidereau R, Beaune P, de Waziers I. Reverse transcriptase(PCR quantification of mRNA levels from cytochrome (CYP)1, CYP2 and CYP3 families in 22 different human tissues. Pharmacogenet. Genomics. 2007; 17:731–742. [PubMed: 17700362]

(2). Nhamburo PT, Kimura S, McBride OW, Kozak CA, Gelboin HV, Gonzalez FJ. The human CYP2F gene subfamily: identification of a cDNA encoding a new cytochrome P450, cDNA-directed expression, and chromosome mapping. Biochemistry. 1990; 29:5491–5499. [PubMed: 1974816]

(3). Zhang X, Zhang QY, Liu D, Su T, Weng Y, Ling G, Chen Y, Gu J, Schilling B, Ding X. Expression of cytochrome P450 and other biotransformation genes in fetal and adult human nasal mucosa. Drug Metab. Dispos. 2005; 33:1423–1428. [PubMed: 16014766]

(4). Lanza DL, Code E, Crespi CL, Gonzalez FJ, Yost GS. Specific dehydrogenation of 3-methylindole and epoxidation of naphthalene by recombinant human CYP2F1 expressed in lymphoblastoid cells. Drug Metab. Dispos. 1999; 27:798–803. [PubMed: 10383923]

(5). Nakajima T, Elovaara E, Gonzalez FJ, Gelboin HV, Raunio H, Pelkonen O, Vainio H, Aoyama T. Styrene metabolism by cDNA-expressed human hepatic and pulmonary cytochromes P450. Chem. Res. Toxicol. 1994; 7:891–896. [PubMed: 7696548]

(6). Thornton-Manning JR, Ruangyuttikarn W, Gonzalez FJ, Yost GS. Metabolic activation of the pneumotoxin, 3-methylindole, by vaccinia-expressed cytochrome P450s. Biochem. Biophys. Res. Commun. 1991; 181:100–107. [PubMed: 1958177]

(7). Lanza DL, Yost GS. Selective dehydrogenation/oxygenation of 3-methylindole by cytochrome P450 enzymes. Drug Metab. Dispos. 2001; 29:950–953. [PubMed: 11408359]

(8). Kartha JS, Yost GS. Mechanism-based inactivation of lung-selective cytochrome P450 CYP2F enzymes. Drug Metab. Dispos. 2008; 36:155–162. [PubMed: 17962375]

(9). Ruangyuttikarn W, Skiles GL, Yost GS. Identification of a cysteinyl adduct of oxidized 3-methylindole from goat lung and human liver microsomal proteins. Chem. Res. Toxicol. 1992; 5:713–719. [PubMed: 1446013]

(10). Regal KA, Laws GM, Yuan C, Yost GS, Skiles GL. Detection and characterization of DNA adducts of 3-methylindole. Chem. Res. Toxicol. 2001; 14:1014–1024. [PubMed: 11511175]

(11). Yan Z, Easterwood LM, Maher N, Torres R, Huebert N, Yost GS. Metabolism and bioactivation of 3-methylindole by human liver microsomes. Chem. Res. Toxicol. 2007; 20:140–148. [PubMed: 17226936]

(12). D'Agostino J, Zhuo X, Shadid M, Morgan DG, Zhang X, Humphreys WG, Shu YZ, Yost GS, Ding X. The pneumotoxin 3-methylindole is a substrate and a mechanism-based inactivator of CYP2A13, a human cytochrome P450 enzyme preferentially expressed in the respiratory tract. Drug Metab. Dispos. 2009; 37:2018–2027. [PubMed: 19608696]

(13). Weems JM, Cutler NS, Moore C, Nichols WK, Martin D, Makin E, Lamb JG, Yost GS. 3-Methylindole is mutagenic and a possible pulmonary carcinogen. Toxicol. Sci. 2009; 112:59–67. [PubMed: 19700606]

(14). Thornton-Manning J, Appleton ML, Gonzalez FJ, Yost GS. Metabolism of 3-methylindole by vaccinia-expressed P450 enzymes: correlation of 3-methyleneindolenine formation and protein-binding. J. Pharmacol. Exp. Ther. 1996; 276:21–29. [PubMed: 8558432]

(15). Weems JM, Lamb JG, D'Agostino J, Ding X, Yost GS. Potent mutagenicity of 3-methylindole requires pulmonary cytochrome P450-mediated bioactivation: a comparison to the prototype cigarette smoke mutagens B(a)P and NNK. Chem. Res. Toxicol. 2010; 23:1682–1690. [PubMed: 20795680]

(16). Barnes HJ, Arlotto MP, Waterman MR. Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in *Escherichia coli*. Proc. Natl. Acad. Sci. U S A. 1991; 88:5597–5601. [PubMed: 1829523]

(17). Wang H, Lanza DL, Yost GS. Cloning and expression of CYP2F3, a cytochrome P450 that bioactivates the selective pneumotoxins 3-methylindole and naphthalene. Arch. Biochem. Biophys. 1998; 349:329–340. [PubMed: 9448722]

(18). Crameri A, Raillard SA, Bermudez E, Stemmer WP. DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature. 1998; 391:288–291. [PubMed: 9440693]

(19). Skiles GL, Yost GS. Mechanistic studies on the cytochrome P450-catalyzed dehydrogenation of 3-methylindole. Chem. Res. Toxicol. 1996; 9:291–297. [PubMed: 8924606]

(20). Cuttle L, Munns AJ, Hogg NA, Scott JR, Hooper WD, Dickinson RG, Gillam EM. Phenytoin metabolism by human cytochrome P450: involvement of P450 3A and 2C forms in secondary

metabolism and drug-protein adduct formation. Drug Metab. Dispos. 2000; 28:945–950. [PubMed: 10901705]

(21). Parikh A, Gillam EM, Guengerich FP. Drug metabolism by *Escherichia coli* expressing human cytochromes P450. Nat. Biotechnol. 1997; 15:784–788. [PubMed: 9255795]

(22). Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. Appl. Environ. Microbiol. 1998; 64:1694–1699. [PubMed: 9572938]

(23). Prodromou C, Pearl LH. Recursive PCR: a novel technique for total gene synthesis. Protein Eng. 1992; 5:827–829. [PubMed: 1287665]

(24). Kikuchi M, Ohnishi K, Harayama S. Novel family shuffling methods for the in vitro evolution of enzymes. Gene. 1999; 236:159–167. [PubMed: 10433977]

(25). Huang W, Johnston WA, Hayes MA, De Voss JJ, Gillam EM. A shuffled CYP2C library with a high degree of structural integrity and functional versatility. Arch. Biochem. Biophys. 2007; 467:193–205. [PubMed: 17904094]

(26). Behrendorff JBYH, Johnston WA, Gillam EMJ. DNA shuffling of cytochrome P450 enzymes. Methods Mol. Biol. 2012 In press.

(27). Kadokami Y, Lewis RV. Reverse electrophoresis to concentrate DNA fractions. Anal. Biochem. 1995; 226:193–195. [PubMed: 7785774]

(28). Abecassis V, Pompon D, Truan G. High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. Nucleic Acids Res. 2000; 28:E88. [PubMed: 11024190]

(29). Johnston WA, Huang W, De Voss JJ, Hayes MA, Gillam EM. A shuffled CYP1A library shows both structural integrity and functional diversity. Drug Metab. Dispos. 2007; 35:2177–2185. [PubMed: 17823232]

(30). Johnston WA, Huang W, De Voss JJ, Hayes MA, Gillam EM. Quantitative whole(cell cytochrome P450 measurement suitable for high-throughput application. J. Biomol. Screen. 2008; 13:135–141. [PubMed: 18216397]

(31). Gillam EM, Aguinaldo AM, Notley LM, Kim D, Mundkowski RG, Volkov AA, Arnold FH, Soucek P, DeVoss JJ, Guengerich FP. Formation of indigo by recombinant mammalian cytochrome P450. Biochem. Biophys. Res. Commun. 1999; 265:469–472. [PubMed: 10558891]

(32). Gillam EMJ, Baba T, Kim B-R, Ohmori S, Guengerich FP. Expression of modified human cytochrome P450 3A4 in *Escherichia coli* and purification and reconstitution of the enzyme. Arch. Biochem. Biophys. 1993; 305:123–131. [PubMed: 8342945]

(33). Laemmli UK, Beguin F, Gujer-Kellenberger G. A factor preventing the major head protein of bacteriophage T4 from random aggregation. J. Mol. Biol. 1970; 47:69–85. [PubMed: 5413343]

(34). Wray W, Boulikas T, Wray VP, Hancock R. Silver staining of proteins in polyacrylamide gels. Anal. Biochem. 1981; 118:197–203. [PubMed: 6175245]

(35). Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004; 32:W526–531. [PubMed: 15215442]

(36). Yin S, Ding F, Dokholyan NV. Eris: an automated estimator of protein stability. Nat. Methods. 2007; 4:466–467. [PubMed: 17538626]

(37). Looman AC, Bodlaender J, Comstock LJ, Eaton D, Jhurani P, de Boer HA, van Knippenberg PH. Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in Escherichia coli. EMBO J. 1987; 6:2489–2492. [PubMed: 3311730]

(38). Yun CH, Yim SK, Kim DH, Ahn T. Functional expression of human cytochrome P450 enzymes in Escherichia coli. Curr. Drug Metab. 2006; 7:411–429. [PubMed: 16724930]

(39). Wulff BB, Thomas CM, Smoker M, Grant M, Jones JD. Domain swapping and gene shuffling identify sequences required for induction of an Avr-dependent hypersensitive response by the tomato Cf-4 and Cf-9 proteins. Plant Cell. 2001; 13:255–272. [PubMed: 11226184]

(40). Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. Mol. Cell. 2000; 5:121–131. [PubMed: 10678174]

(41). Wang A, Savas U, Stout CD, Johnson EF. Structural characterization of the complex between alpha-naphthoflavone and human cytochrome P450 1B1. J. Biol. Chem. 2011; 286:5736–5743. [PubMed: 21147782]

(42). Sansen S, Hsu MH, Stout CD, Johnson EF. Structural insight into the altered substrate specificity of human cytochrome P450 2A6 mutants. Arch. Biochem. Biophys. 2007; 464:197–206. [PubMed: 17540336]

(43). 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]

(44). Kim DH, Kim KH, Isin EM, Guengerich FP, Chae HZ, Ahn T, Yun CH. Heterologous expression and characterization of wild-type human cytochrome P450 1A2 without conventional N-terminal modification in Escherichia coli. Protein Expr. Purif. 2008; 57:188–200. [PubMed: 18032064]

(45). Shumyantseva VV, Bulko TV, Alexandrova SA, Sokolov NN, Schmid RD, Bachmann T, Archakov AI. N-terminal truncated cytochrome P450 2B4: catalytic activities and reduction with alternative electron sources. Biochem. Biophys. Res. Commun. 1999; 263:678–680. [PubMed: 10512738]

(46). Kartha JS, Skordos KW, Sun H, Hall C, Easterwood LM, Reilly CA, Johnson EF, Yost GS. Single mutations change CYP2F3 from a dehydrogenase of 3-methylindole to an oxygenase. Biochemistry. 2008; 47:9756–9770. [PubMed: 18717595]

(47). Gotoh O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J. Biol. Chem. 1992; 267:83–90. [PubMed: 1730627]

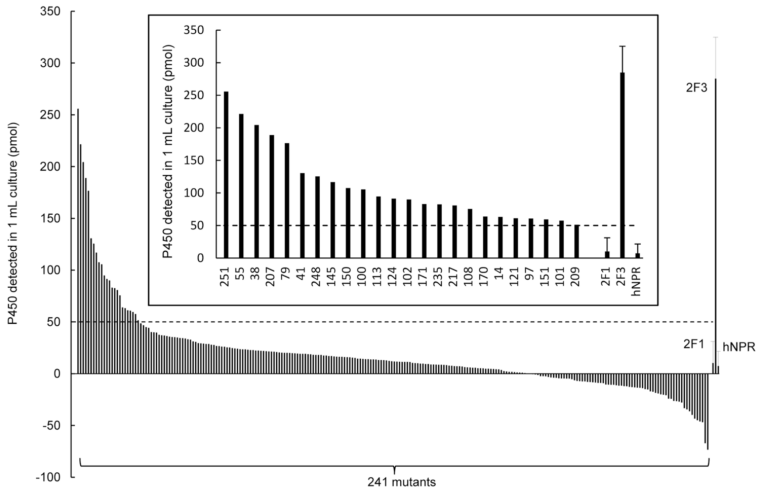

Figure 1. Expression of mutants from the first generation P450 2F1/2F3 shuffled library

The first generation shuffled library (JH_2F_F1_1) was screened for P450 hemoprotein (pmol) produced in 1 mL, 72 h microaerobic cultures as determined by Fe(II)]CO vs. Fe(II) difference spectroscopy. Inset: mutants that expressed greater than 50 pmol P450 per 1 mL culture. Data from the controls P450 2F1, P450 2F3, and hNPR are included in both graphs for reference, and are shown on the main plot in the same order as in the inset. Data from mutants were from single cultures but data from controls represents the mean ± SD of n = 4 independent cultures. The dashed line on the inset plot indicates the sensitivity limit of 50 pmol for detecting P450 hemoprotein.

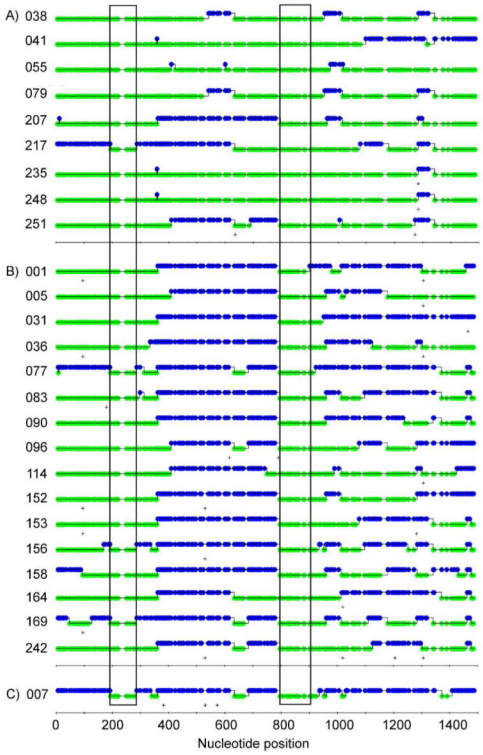

Figure 2. Recombination patterns of shuffled P450 2F mutants

Sequences of mutants from A) first(generation library JH_2F_F1_1, B) second(generation library JH_2F_F2_1, and C) third(generation library JH_2F_F3_1, that were expressed successfully in *E. coli* as detectable hemoprotein were analyzed by alignment to the P450 2F1 and P450 2F3 parental sequences. Software described previously [25] was used to identify which sequence elements were derived from each of the relevant parental sequences. P450 2F1- and P450 2F3-derived sequence is represented by blue and green dots respectively. Where no dot occurs, both parental sequences encode the same base. PCR(derived point mutations are indicated by a cross (+). Boxes indicate regions where P450 2F1 sequence does not occur in any successfully(expressed mutant.

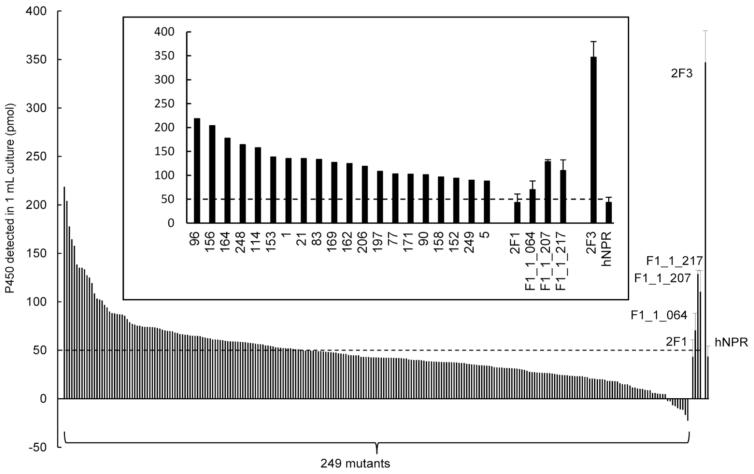

Figure 3. Expression of mutants from the second generation P450 2F1/2F3 shuffled library

The second generation shuffled library (JH_2F_F2_1) was screened for P450 hemoprotein (pmol) produced in 1 mL, 72 h microaerobic cultures as determined by Fe(II)]CO vs. Fe(II) difference spectroscopy. Inset: the twenty mutants with greatest P450 expression. Data from the forms used as parents for the F2 library, P450 2F1, JH_2F_F1_1_064 (abbreviated as F1_1_064 on this plot), JH_2F_F1_1_207 (F1_1_207), and JH_2F_F1_1_217 (F1_1_217), plus controls P450 2F3 and hNPR are included in both graphs for reference (shown in the main plot in the same order as in the inset). F2 mutants were expressed as single cultures. However, results for controls and F1 mutants used as parents for the F2 library are shown as the mean ± SD of n = 4 independent cultures. The dashed line on the inset plot indicates the sensitivity limit of 50 pmol for detecting folded P450.

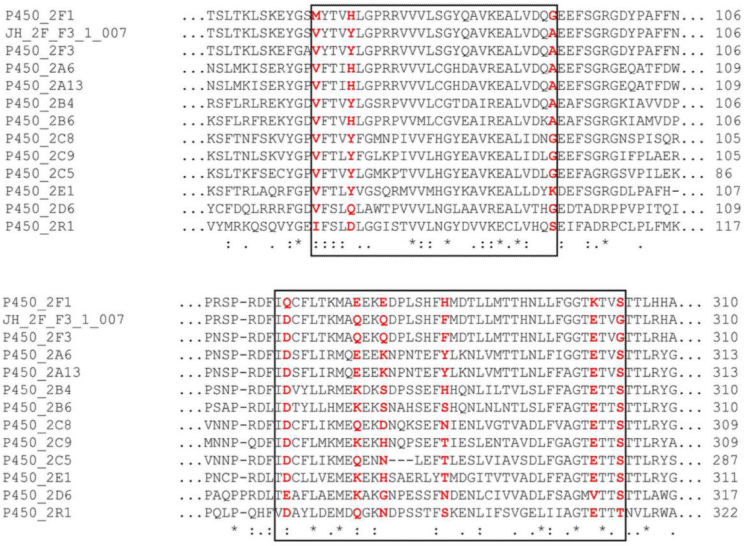

Figure 4. Alignment of P450 2F regions of interest with other P450 2 family P450s that have been expressed in *E. coli*

The amino acid sequences for P450s 2F1, 2F3, and shuffled mutant JH_2F_F3_1_007 were aligned with the sequences of all P450 2 family P450s for which crystal structures had been published to date. The two regions of interest (amino acid residues 65-92 and 266-301) are marked by boxes. Amino acid positions that differ between P450s 2F1 and 2F3 within the regions of interest are marked in red with bold type.

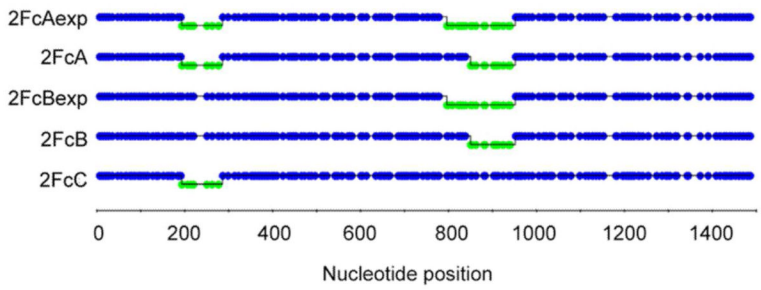

Figure 5. P450 2F1/2F3 chimeras constructed to test the importance of specific sequence regions

P450 2F1- and P450 2F3-derived sequence is represented in blue and green respectively. Chimeras 2FcAexp, 2FcA, and 2FcC contain P450 2F3 sequence between nucleotides 191 and 285. Chimeras 2FcAexp and 2FcBexp contain P450 2F3 sequence between nucleotides 794 and 940. Chimeras 2FcA and 2FcB contain P450 2F3 sequence between nucleotides 848 and 940. The remaining sequence in all chimeras is from P450 2F1.

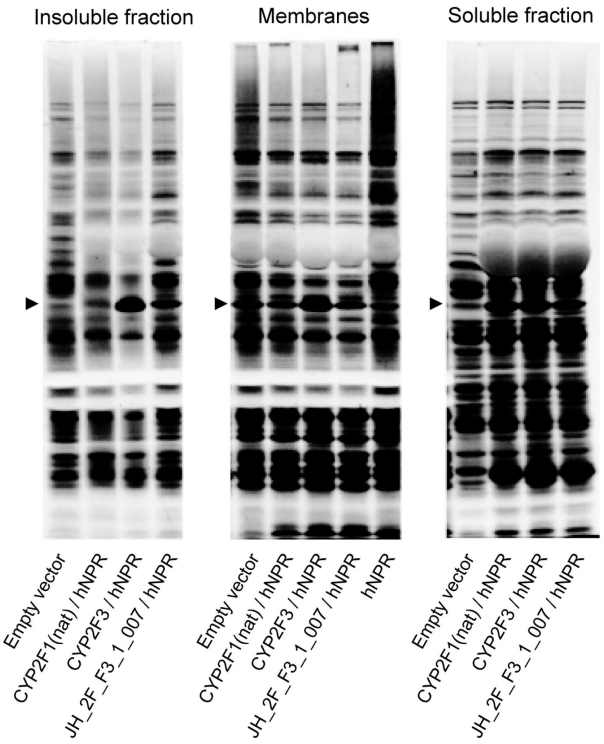

Figure 6. SDS PAGE of subcellular fractions from *E. coli* expressing P450 2F1, P450 2F3, and shuffled mutant JH_2F_F3_1_007

Subcellular fractions were electrophoresed on a 7.5 % polyacrylamide gel and protein was visualized by silver staining. First panel: Insoluble fractions from cultures expressing empty pCW vector (lane 1), pCW/2F1(nat)/hNPR (lane 2), pCW/2F3/hNPR (lane 3), and pCW/JH_2F_F3_1_007/hNPR (lane 4). Second panel: Membrane fractions from cultures expressing the empty pCW vector (lane 5), pCW/2F1(nat)/hNPR (lane 6), pCW/2F3/hNPR (lane 7), pCW/JH_2F_F3_007/hNPR (lane 8), and pCW/hNPR. Third panel: Cytosolic fractions from cultures expressing empty pCW vector (lane 9), pCW/2F1(nat)/hNPR (lane 10), pCW/2F3/hNPR (lane 11), and pCW/JH_2F_F3_1_007/hNPR (lane 12). Protein bands corresponding to P450 from the P450 2F subfamily were identified in lanes 3, 4, 7, 8, 11, 12 and 13, by comparison to negative controls and their position is indicated by the arrowheads. The cell pellet and membrane samples shown were derived from 4 Zg wet mass of harvested cells, while cytosolic samples were derived from 2 Zg wet mass of cells.

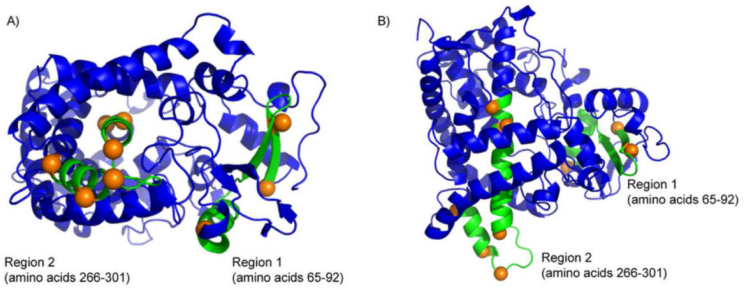

Figure 7. Homology modelling of two P450 2F3 sequence regions associated with successful expression of shuffled mutants

A) A model structure of shuffled mutant JH_2F_F3_1_007 was created using the Robetta Full Chain Protein Prediction server. Two regions where P450 2F3 sequence occurred in all expressing shuffled mutants (region 1: amino acids 65-92, and region 2: amino acids 266-305) are highlighted in green, with amino acids that differ between P450 2F1 and P450 2F3 within these regions marked with orange spheres. B) The structure shown in panel A rotated 90° through the x(axis.

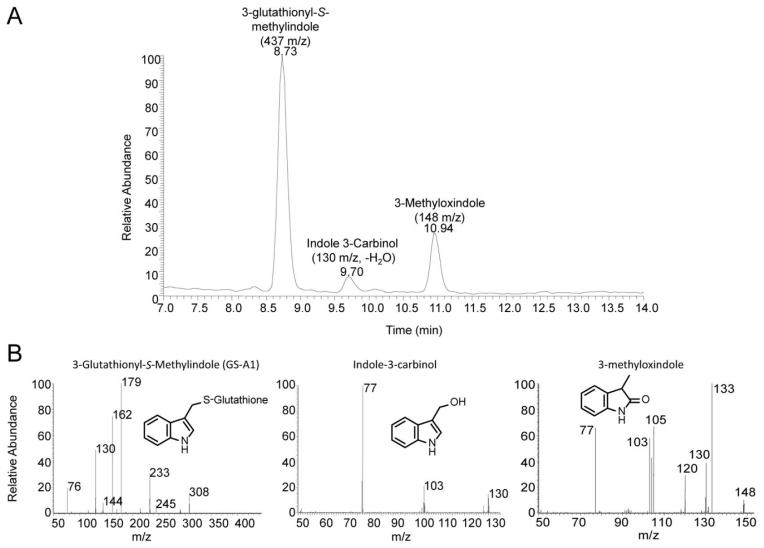

Figure 8. Metabolism of 3MI by shuffled mutant JH_2F_F3_1_007

A) Metabolites detected by LC/MS following incubation of mutant JH_2F_F3_1_007 in *E. coli* membrane preparations with 3MI and glutathione. Products were identified by comparison to authentic standards. B) Mass spectra of identified products.

Table 1

Metabolism of 3-methylindole by shuffled mutant JH_2F_F3_1_007

Metabolite	Metabolite quantity (μmol)[a]	Rate of metabolite formation (nmol/min/mg protein)[a]
Indole-3-carbinol	0.51 ± 0.05	12 ± 1
3-methyloxindole	0.8 ± 1	19 ± 2
3-glutathionyl-*S*-methylindole	33 ± 2	738 ± 73

[a]Mean ± SD from n = 3 identical incubations of 4.4 mg.mL^{-1} of protein from bacterial membrane preparations, with 100 μM 3MI and 3 mM GSH conducted for 20 min.