Molecular Pharmacology Fast Forward. Published on April 27, 2005 as doi:10.1124/mol.105.012062

MOL 12062

Discovery of Novel Flavin-Containing Monooxygenase 3 (*FMO3*) Single Nucleotide Polymorphisms and Functional Analysis of Upstream Haplotype Variants

Sevasti B. Koukouritaki, Mark T. Poch, Erwin T. Cabacungan,
D. Gail McCarver, and Ronald N. Hines

Departments of Pediatrics and Pharmacology/Toxicology, Medical College of Wisconsin, Milwaukee, WI, USA 53226

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Running Head: *FMO3* promoter genetic variation

Corresponding Author:

Ronald N. Hines
Department of Pediatrics
Medical College of Wisconsin
8701 Watertown Plank Rd.
Milwaukee, WI 53226, U.S.A.
Tel: (414) 456-4322
Fax: (414) 456-6651
E-mail: rhines@mail.mcw.edu

Number of Text Pages:	20
Number of Tables:	7
Number of Figures:	4
Number of References:	40
Number of Words in Abstract:	247
Number of Words in Introduction:	579
Number of Words in Discussion:	1361

Non-Standard Abbreviations: 24PDR, 24 sample subset of the Coriell Polymorphism Discovery Resource; FMO, Flavin-containing monooxygenase; PCR, polymerase chain reaction; SBE, single-base extension; SNP, single-nucleotide polymorphism; TMAU, trimethylaminuria

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Abstract

The flavin-containing monooxygenases (FMOs) are important for xenobiotic metabolism. FMO3, the predominant FMO enzyme in human adult liver, exhibits significant interindividual variation that is poorly understood. This study was designed to identify common *FMO3* genetic variants and determine their potential for contributing to interindividual differences in *FMO3* expression. *FMO3* single nucleotide polymorphism (SNP) discovery was accomplished by re-sequencing DNA samples from the Coriell Polymorphism Discovery Resource. Population-specific SNP frequencies were determined by multiplexed, single-base extension using DNA from 201 Hispanic-American (Mexican descent), 201 African-American, and 200 Caucasian-American (northern European descent) subjects. Haplotypes were inferred and population frequencies estimated using PHASE V2.1. Multi-site-directed mutagenesis was used to introduce inferred upstream haplotypes into an *FMO3*/luciferase construct for functional analysis in HepG2 cells. Sequence analysis revealed seven *FMO3* upstream SNPs, eleven exon SNPs, and twenty-two intron SNPs. Five of the latter fell within consensus splice sites. A g.72G>T variant (E24D) is predicted to impact the structure of the Rossmann fold involved in FAD binding, while a g.11177C>A variant (N61K) is predicted to disrupt the secondary structure of a conserved membrane interaction domain. Seven common (>1%) promoter region haplotypes were inferred in one or more of the study populations that differed in estimated frequency among the groups. Haplotype 2 resulted in an 8- fold increase in promoter activity, while haplotype 8 and 15 exhibited a near complete loss of activity. In conclusion, *FMO3*

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

promoter haplotype variants modulate gene function and likely contribute to interindividual differences in *FMO3* expression.

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

The flavin-containing monooxygenases (FMOs) (EC 1.14.13.8) are a family of NADPH- and oxygen-dependent microsomal enzymes involved in the oxidative metabolism of many nucleophilic nitrogen-, sulfur- and phosphorous-containing drugs and toxicants (Cashman, 2002). Multiple human *FMO* genes have been identified: a five gene cluster at 1q24.3 (*FMO1-4* and *FMO6p*) that encodes four active enzymes (FMO1-4), a second cluster of five genes at 1q24.2 (*FMO7p-11p*), all representing pseudogenes, and a single gene, *FMO5*, at 1q21.1, that also encodes an active enzyme (FMO5) (Hines *et al.*, 2002; Hernandez *et al.*, 2004). Each of the human FMO enzymes exhibit a distinct but unusually broad and overlapping substrate specificity that is partly attributable to the unique FMO catalytic mechanism (Ziegler, 2002). Individual *FMO* genes exhibit a highly tissue-, species- and developmental-specific expression pattern (Shehin-Johnson *et al.*, 1995; Yeung *et al.*, 2000; Koukouritaki *et al.*, 2002). For example, human FMO3 is essentially nondetectable in fetal liver, but is observed in most individuals by 1 to 2 years of age (Koukouritaki *et al.*, 2002) and is expressed in the adult human liver, but not other tissues, at levels that approach those reported for CYP3A4 (Overby *et al.*, 1997; Wrighton *et al.*, 1990). However, FMO3 represents a minor hepatic enzyme in most other mammalian species with the possible exception of the female mouse.

Ten- to twenty-fold differences in interindividual *FMO* expression have been reported that may contribute to an individual's susceptibility to toxicants and/or response to drugs (Overby *et al.*, 1997; Yeung *et al.*, 2000; Koukouritaki *et al.*, 2002). For the cytochromes P450, the superfamily of proteins considered most important for phase I xenobiotic metabolism, both genetic diversity and differential exposure to xenobiotics capable of inducing gene expression contribute to inter-subject expression differences. However, the *FMO* genes are not known to

respond to such environmental influences and as such, genetic differences are thought to contribute more heavily to differences in expression between individuals. Further, differences in variant allelic frequencies among ethnic and/or racial groups also contribute to inter-population differences in xenobiotic metabolism.

Thirty genetic variants in human *FMO3* coding sequences have been characterized. Thirteen of these are single nucleotide changes that result in loss of function and represent rare alleles causative for trimethylaminuria (TMAU) or "fish-odor syndrome" (for review, see Cashman, 2004). In addition, a homozygous deletion of exons 1 and 2 was reported as causative in an Australian of Greek ancestry with TMAU (Forrest *et al.*, 2001). However, the substantial differences in FMO3 expression observed within or among different populations (Overby *et al.*, 1997; Yeung *et al.*, 2000; Koukouritaki *et al.*, 2002) cannot be attributed to such rare alleles. Rather, common variants encoding an altered, but functional FMO3 must contribute. Three of the 30 known human *FMO3* genetic variants (E158K, V257M, and E308G) are common (*i.e.*, allelic frequencies >1%), two of which are associated with reduced FMO3 activity (for review Cashman, 2004). However, considering the percent loss of activity and their frequency, these variants would not fully explain observed interindividual differences in *FMO3* expression.

The potential for variation within regulatory sequences to contribute to interindividual differences in gene expression is increasingly being recognized (Hoogendoorn *et al.*, 2003; Pastinen and Hudson, 2004). Yet little has been reported regarding *FMO3* promoter variation and its possible contribution to expression differences. The current study's objectives were to identify common *FMO3* genetic variants, determine the allelic frequencies of the promoter

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

variants in several populations and infer haplotypes, and finally, examine the potential for haplotype variants to impact *FMO3* promoter function.

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Materials and Methods

Materials. Cell culture medium and fetal bovine serum were purchased from Sigma-Aldrich (St. Louis, MO). DNAzol reagent, *Taq* DNA polymerase and Lipofectamine 2000 were purchased from Invitrogen (Carlsbad, CA). FlexiGene DNA kits were obtained from Qiagen (Valencia, CA). The ExoSAP-IT mix, containing both exonuclease I and shrimp alkaline phosphatase, as well as shrimp alkaline phosphatase alone, were purchased from USB Corp. (Cleveland, OH). ABI Prism Big Dye Terminator Cycle Sequencing kits were obtained from Applied Biosystems (Foster City, CA) and CEQ SNP-Primer Extension and Dye Terminator Cycle Sequencing kits from Beckman Coulter, Inc. (Fullerton, CA). HepG2 hepatoblastoma cells were a gift of Dr. Barbara Knowles (Jackson Laboratories, Bar Harbor, ME). The luciferase reporter plasmid, pGL3Basic, and luciferase reporter assay kit were purchased from Promega (Madison, WI), whereas the Luminescent β-Galactosidase Detection System was obtained from BD Biosciences (Palo Alto, CA). Custom oligonucleotides were synthesized by MWG Biotech (High Point, NC). For single nucleotide polymorphism (SNP) discovery, the 24 sample subset of the Polymorphism Discovery Resource (24PDR) (Collins *et al.*, 1998) was obtained from the Coriell Institute (Camden, NJ). All other reagents were obtained from commercial sources at the purest grade available.

Patient Recruitment and DNA Isolation. Individuals representing various ethnic and/or racial groups were recruited to provide DNA samples as described previously (McCarver *et al.*, 1997; Whetstine *et al.*, 2000; Zheng *et al.*, 2003). In all instances, ethnicity and/or race was self-reported. After obtaining informed consent, blood was obtained by venipuncture, and

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

DNA was extracted using the FlexiGene DNA kit according to the manufacturer's instructions. Using DNAzol reagent, DNA samples from individuals of known ethnicity also were isolated from tissue obtained from the University of Miami and University of Maryland Brain and Tissue Banks for Developmental Disorders that are under contract with the National Institute for Child Health and Development, NOI-HD-8-3284 and NOI-HD-8-3283, respectively. Finally, a panel of 100 DNA samples from unrelated, self-declared Caucasians and distinct from the 24PDR was obtained from the Coriell Institute. These three DNA sources resulted in 201 samples from individuals of Hispanic-American (Mexican) descent, 201 samples from individuals of African-American descent and 200 samples from individuals of Caucasian-American (northern European) descent. Research protocols were approved by all Institutional Review Boards involved.

DNA Amplification and Sequence Analysis for SNP Discovery. Common *FMO3* genetic variants were identified by sequencing approximately 1.0 kbp of *FMO3* 5'-flanking information along with each of 9 exons and a minimum of 64 bases of exon flanking sequences using the Coriell 24PDR. Reactions were performed in a final volume of 50 μL and contained 0.4 μg of genomic DNA, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM each deoxyribonucleotide triphosphate, 0.5 μM of each primer and 1.25 U of *Taq* DNA polymerase. Polymerase chain reaction (PCR) amplification primers and individual reaction conditions are shown in Table 1. Cycle sequencing reactions were performed with the same primer pairs used for amplification. Thus, for each amplicon, both strands were sequenced and any low frequency SNPs (allelic frequency of 0.021 in the 24PDR) were re-sequenced to confirm their identification.

Multiplex Single Base Extension (SBE) Assay. Genotype frequencies of common *FMO3* upstream variants in different populations were determined using a multiplexed SBE assay (Lindblad-Toh *et al.*, 2000) following the recommended protocol included in the CEQ SNP-primer extension kit. Template amplification was performed as described above (SNP discovery) and Table 1, but using 0.2 μM of the upstream and downstream primers from sets 2 and 1, respectively, 100 ng genomic DNA, and 2.5 U *Taq* DNA polymerase in a 50 μL reaction volume. After amplification, excess primer and unincorporated deoxynucleotides were removed by incubating with ExoSAP-IT (USB Corp.) at 37ºC for 60 min followed by enzyme inactivation at 80ºC for 15 min. An aliquot of the amplification reaction was fractionated by agarose gel electrophoresis both to verify the fidelity of the amplification reaction and quantify yield. The SBE reactions were performed in a 20 μL volume containing 20 fmol DNA template, 80 mM Tris-HCl pH 8.9, 2 mM $MgCl_2$, 0.1 μM of each fluorescent-tagged dideoxynucleotide triphosphate, 1 μL of the proprietary DNA polymerase supplied with the CEQ SNP Primer Extension kit (Beckman-Coulter), and SBE primers at concentrations optimized to give uniform signal strength (see Table 2). After 30 cycles of 96ºC for 30 sec, 50ºC for 30 sec, and 72ºC for 30 sec, unincorporated dideoxynucleotides were eliminated by incubating with 2 U shrimp alkaline phosphatase (USB Corp.) at 37ºC for 60 min followed by enzyme inactivation at 75ºC for 15 min. The SBE reaction product was diluted 1:400 in SLS buffer (Beckman-Coulter) and analyzed by capillary electrophoresis using a CEQ8000 Genetic Analysis System along with the Size Standard 80 standards (Beckman-Coulter). For a positive control, 20 fmol of linearized pRNH707 (reference sequence), pRNH858 (containing all queried SNPs), or an equal mixture of both were used as templates in the SBE reaction.

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Plasmids. Human *FMO3* upstream sequences from position -2965 to +42 (relative to the transcription start site) were isolated, cloned and verified in a similar fashion as that described earlier for human *FMO1* (Hines *et al.*, 2003), resulting in the plasmid, pRNH659. These same *FMO3* sequences were subsequently excised using *Bam*HI/*Srf*I and cloned into the same restriction endonuclease sites within the pGL3basic vector to generate pRNH694. Finally, pRNH694 was digested with *Sma*I/*Bst*1071, the excised fragment discarded, and the remaining DNA re-ligated, resulting in pRNH707 containing *FMO3* sequences from position -988 to +42 directing luciferase expression. Sequence analysis of pRNH707 confirmed this plasmid represented the *FMO3* reference sequence (accession no. AL021026). Site-directed mutagenesis (QuikChange® Multi Site-Directed Mutagenesis kit, Stratagene) was used to introduce various single base changes into pRNH707 representing the inferred *FMO3* haplotypes: pRNH817, -2650C>G; pRNH830, -2650C>G, -2543T>A, and -2177G>C; pRNH859, -2543T>A; pRNH862, -2650C>G and -2543T>A; pRNH863, -2589C>T and -2106G>A; pRNH864, -2106G>A and -1961T>C. As a control for the genotyping assays, pRNH858 also was created containing all of the upstream SNPs listed in Table 5. In all instances, the nucleotide changes and the fidelity of the non-targeted information were verified by DNA sequence analysis.

Cell Culture and Transient Expression Reporter Assays. HepG2 hepatoblastoma cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics as described by Boucher *et al.* (1993). The cells were subcultured (1:4 ratio) every 3 to 4 days and used in the present study at passages 12 to 20. For transfection studies 1.2 x 10^4 cells were subcultured in each well of a 24-well plate, and 24 hours later, were transfected using a mixture of 3 μg Lipofectamine 2000, 0.8 μg test luciferase reporter plasmid

and 0.2 μg pCMVβgal. After incubation for 16 h at 37°C, transfection medium (OPTI-MEM supplemented with 10% fetal bovine serum) was replaced with growth medium and the cells were incubated for an additional 48 h. The cells were processed and luciferase assays performed according to the manufacturer's instructions. Data were normalized with respect to β-galactosidase specific activity to correct for differences in transfection efficiency and luciferase activity expressed as relative luciferase units. The data reported are from two independent transfection assays with each experiment performed in triplicate.

Data Analysis. The potential ability of the identified intron variants to form cryptic splice sites or alter existing putative cryptic splice sites or normal splice site strength was assessed using the Automatic Splicing Mutation Analysis software (https://splice.cmh.edu). The information content (R_i) value of the natural or potential cryptic splice site was calculated based on a weight matrix from a scan of the complete human genome (April, 2003). A similar approach was used to assess the impact of both synonymous and non-synonymous variants on exon splice enhancer motifs for SRp40, SF2/ASF, and SC35 (Liu *et al.*, 1998; Liu *et al.*, 2000; Rogan *et al.*, 1998) (Peter K. Rogan, personal communication). The hypothetical default minimum value for functional binding ($R_{i,min}$) is 0.0 bits. However, for splice site acceptor and donor sites, a more comprehensive analysis of known splice sites identified in the human genome resulted in a revision of $R_{i,min}$ to 1.6 bits, which was the value utilized in the current study (Rogan *et al.*, 2003). A sequence variation causing a decrease in information content that remains above the $R_{i,min}$ may lead to a leaky site. However, the average decrease in R_i needed to reduce levels of normal splicing was reported to be 2.9 bits for donor sites and 4.3 bits for

acceptor sites (Rogan *et al.*, 1998). Thus, these minimum values were used to assess variation in splice donor and acceptor sites identified in the current study.

The impact of non-synonymous exon variants on predicted protein structure was assessed using the Protein Structure and Hydrophilicity Analysis modules within the Clone Manage Suite 7.1 software (Scientific and Educational Software, Cary, NC) which in turn are based on the methods of Garnier *et al.* (1978) and Kyte and Doolittle (1982), respectively. Pairwise alignment and progressive assembly of *FMO3* sequences from various species using Neighbor-Joining Phylogeny was accomplished using the Align program V5.0 within Clone Manager Suite 7.1 (Scientific and Educational Software).

Individual haplotypes and their estimated population frequencies were inferred using the PHASE program, V2.1, with all parameters set at the default values (Stephens *et al.*, 2001; Stephens and Donnelly, 2003). Allelic frequencies for individual sequence variants were compared using Fisher's exact test. A Bonferroni adjustment for the comparisons between the three groups was used, reducing the accepted α from 0.05 to 0.016. Differences in inferred mean haplotype frequencies were compared using unpaired t-tests with a Welch correction or by one-way ANOVA with a Bonferroni post-hoc test. Functional differences among the different haplotypes were assessed using transient expression assays and compared by one-way ANOVA with a Dunnett's post-hoc test. Except where the Bonferroni adjustment was applied, a $P < 0.05$ was accepted as indicating a significant difference. Software used was GraphPad Instat V3.05 (San Diego, CA).

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Results

***FMO3* SNP Discovery.** SNP discovery was accomplished by sequencing approximately 1 kbp upstream of the *FMO3* 5' transcription start site, as well as all nine exons and flanking splice-sites using DNA samples from 24 unrelated individuals obtained from the Coriell PDR (Camden, NJ). A total of 40 SNPs were identified. Eleven SNPs were identified within exon sequences, of which five were reported previously in the literature and/or build 124 of the dbSNP database (Table 3) (Cashman *et al.*, 1997; Treacy *et al.*, 1998) (http://www.ncbi.nlm.nih.gov). Two previously unreported *FMO3* exon variants shown in Table 3, g.15019A>C and g.15437A>G, are synonymous changes (T108T and K167K, respectively) and as such, would not alter FMO3 catalytic activity. Three novel non-synonymous exon variants also were identified: g.72G>T (E24D), g.11177C>A (N61K), and g.23613G>T (K416N). A total of 22 SNPs were identified within intron sequences, five of which fall within *FMO3* splice donor or acceptor sites (Table 4). Six of the identified intron SNPs were previously described (g.15399G>A, Basarab *et al.*, 1999) and/or have been reported in dbSNP build 124 (http://www.ncbi.nlm.nih.gov). *FMO3* sequence analysis revealed 7 upstream variants, none of which have been reported in the literature, but 5 of which were reported in build 124 of dbSNP (Table 5) (http://www.ncbi.nlm.nih.gov).

Predicted Impact of *FMO3* Exon and Intron SNPs. All three of the *FMO3* non-synonymous exon variants fall within conserved domains that are part of predicted, functional FMO motifs (Figure 1). Analysis of the predicted impact of these variants revealed that the E24D substitution would disrupt the α-helical structure at the C-terminus of the conserved

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Rossmann fold involved in FAD binding and as such, may impact FMO3 catalytic activity (Figure 2). The N61K substitution falls within a less well confirmed dimer-loop structure thought to be involved in membrane interactions. This substitution has little or no effect on the predicted hydrophilicity of this peptide domain, but is predicted to significantly impact the peptides secondary structure, disrupting a strong β-sheet structure and favoring an α-helix (Figure 3). In contrast, the K416N variant is predicted to have minimal impact on either hydrophilicity or protein structure.

Both synonymous and non-synonymous polymorphisms can potentially effect RNA splicing by altering the efficiency of exon splice enhancer motifs (Fairbrother *et al.*, 2002). Of the *FMO3* variants reported in the present study, three result in an increase in R_i value for potential splice site enhancer motifs: the 15019A>C (T108T) variant results in the creation of a potential SC35 site in exon 4 at 15016 (R_i from -3.4 to 2.7), the 15136C>A (S147S) variant results in the strengthening of a potential SRP40 site in exon 4 at 15134 (R_i from 1.2 to 3.8), and the 21443A>G (E308G) variant in exon 7 results in the creation of a potential SF2/ASF site at 21438 (R_i from 0.1 to 6.8), a potential SRP40 site at 21439 (R_i from 0.6 to 6.7) and a potential SC35 site at 21443 (R_i from -0.5 to 3.2). However, the observation that the R_i values for the splice donor and acceptor sites flanking exon 4 and exon 7 are high (8.6 and 11.3 for exon 4 and 11.8 and 10.1 for exon 7, respectively), suggests that the creation of new splice site enhancers and/or the strengthening of existing enhancers may not significantly alter splicing efficiency. Two of the variants result in the abolishment of potential splice site enhancer motifs: the -1732G>T variant in exon 1 eliminates a potential SRP40 site at -1730 (R_i from 6.6 to 1.9), a potential SF2/ASF site at -1729 (R_i from 6.1 to 0.3) and a potential SC35 site at -1732 (R_i from

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

2.5 to -2.2). The 72G>T (E24D) variant abolishes a potential SF2/ASF site in exon 2 at position 68 (R_i from 3.3 to -2.5). The likelihood of these sequence variations impacting splicing efficiency is somewhat higher, given the lower Ri values of the flanking splice sites (8.8 for the intron 1 splice donor site and 6.4 and 5.7 for the intron 1 splice acceptor site and intron 2 splice donor site, respectively).

Of the 17 variants falling outside the reported *FMO3* splice donor or acceptor sites, two resulted in changes in information content sufficient to consider their potential to create a cryptic site, or strengthen an existing cryptic site (*i.e.*, a potential splice site with an R_i value >1.6 bits) (Rogan *et al.*, 2003). The 154T>G intron 2 variant results in an R_i value change in a potential cryptic donor splice site at position +150 from 1.0 to 4.9 bits. Use of the +150 cryptic donor site would result in a 17 b insertion into the 5'-leader sequence of the processed transcript, but would not alter FMO3 coding information. The 4.9 bit R_i value of the potential +150 cryptic site approaches the 5.7 bit R_i value of the reference sequence donor site at +133 and as such, may lead to alternative splicing. The 14982T>G intron 3 variant results in an R_i value change from -5.6 bits to 3.1 bits, consistent with the creation of a cryptic splice acceptor site which would result in a 34 bp insertion and a frame-shift. However, the reference sequence splice acceptor site at 15016 has an R_i value of 8.6 bits and as such, remains a much stronger site.

Of the five variants falling within splice sites, three (11322T>G within the intron 3 splice donor site; 15399G>A within the intron 4 splice acceptor site; and 24416G>T within the intron 8 splice acceptor site) result in an increase in information content, suggesting they strengthen the respective splice site. The remaining two SNPs result in a decrease in information content. The

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

23532C>A variant decreases the intron 7 splice acceptor site R_i value from 13.9 to 12.6 bits while the 23628G>T variant decreases the intron 8 splice donor site R_i value from 9.3 to 8.3 bits. A variant causing a decrease in information content that remains above the minimum for splice site recognition (1.6 bits) (Rogan *et al.*, 2003) may lead to a leaky splice site. However, given that neither the 23532C>A nor 23628G>T variants result in R_i changes approaching the average necessary decrease (see Materials and Methods) (Rogan *et al.*, 1998), it is unlikely they would cause leaky splicing. Further, there are no cryptic sites within a 1000 bp window on either side of these splice sites that offer an alternative with significantly higher information content.

Analysis of *FMO3* Promoter Variants. Given the paucity of known information regarding molecular mechanisms regulating *FMO3* expression, it was impossible to predict whether or not any of these variants might contribute to changes in transcription efficiency. As such, these variants were given the highest priority for further pursuit experimentally. To define *FMO3* promoter region haplotypes and elucidate their potential role in modulating gene function, the allelic frequency of the seven common *FMO3* upstream variants was determined in Caucasian-American (northern European descent), Hispanic-American (Mexican descent) and African-American populations using a multiplexed SBE assay (Table 6). The pattern of ethnic distribution differed among the different SNPs. The frequencies of the g.-2650C>G and the g.-2543T>A variants were greater among Hispanic Americans compared to either the Caucasian- or African-Americans, whereas they did not differ in the latter two populations. The g.-2177G>C SNP, although most common within the Hispanic-American population, also was more common among African- than Caucasian-Americans. Two variants were more common

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

among Caucasian-Americans. The g.-2589C>T variant was more frequent in this population compared to the other two populations whose frequency did not differ from each other. Finally, the -2106G>A variant was more common in the Caucasian-American population than the Hispanic-American population, whereas the frequency among African-Americans did not differ from either of these groups. The g.-2099A>G SNP was not observed in any of the study populations, suggesting that this variant will be present in one of the other population groups represented in the 24PDR DNA panel, *i.e.*, Native-Americans or Asian-Americans.

Based on the determined allelic frequencies of the upstream variants, a total of fifteen haplotypes were inferred using the Phase V2.1 program, seven of which were common (>1%) in one or more of the study populations with haplotype 1 matching the reference sequence (Table 7). Haplotype 2 (g.-2650C>G, g.-2543T>A and g.-2177G>C) was the most common variant haplotype in the Hispanic-American and African-American populations, while haplotype 3 (g.-2650C>G) was the most common inferred haplotype variant in the Caucasian-American population. However, the frequencies of both haplotype 2 and 3 were significantly different from one population group to another (ANOVA, Bonferroni post hoc, $P<0.001$). Four additional haplotypes were inferred in the African-American and Caucasian-American study populations that were not inferred in the Hispanic-American study group. Haplotype 4 (g.-2650C>G and g.-2543T>A) was estimated to occur at essentially the same frequency in both the African-American and Caucasian-American study populations. In contrast, haplotype 11 (g.-2543T>A) was estimated to occur at a significantly higher frequency in the Caucasian-American versus African-American study groups (unpaired t-test, $p<0.001$). Further, haplotype 11 was estimated to be the second most abundant in the Caucasian-American population.

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Finally, haplotype 8 (g.-2589C>T, g.-2106) was unique to the Caucasian-American group while haplotype 15 (g.-2106G>A, g.-1961T>C) was unique to the African-American population.

To explore the potential impact of the human *FMO3* upstream variants on promoter activity, the different haplotypes inferred in the three study groups were introduced into a reporter construct containing *FMO3* position -988 (-2762 relative to the ATG start codon) to +42 (-1732 relative to the ATG start codon) directing luciferase expression. Separate studies in our laboratory have suggested that most, if not all of the regulatory elements important for controlling *FMO3* expression are located within these sequences (D. Klick and R.N. Hines, unpublished data). The luciferase expression plasmids were subsequently analyzed by transient expression in HepG2 cells (Figure 4). Relative to the reference sequence (pRNH707), no difference in promoter activity was observed with haplotypes 3 (pRNH817, g.-2650C>G), 4 (pRNH862, g.-2650C>G and -2543T>A) or 11 (pRNH859, g.-2543T>A). In contrast, haplotype 2 (pRNH830), consisting of the g.-2650C>G, g.-2543T>A, and g.-2177G>C SNPs, exhibited an 8-fold increase in *FMO3* promoter activity (ANOVA, Dunnett's post hoc, $P<0.01$). Strikingly, both haplotype 8 (pRNH863) (g.-2589C>T and g.-2106G>A) and haplotype 15 (pRNH864) (g.-2106G>A, g.-1961T>C) (pRNH864) exhibited a near complete loss of promoter activity (ANOVA, Dunnett's post hoc, $P<0.01$).

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Discussion

The current study focused on identifying common *FMO3* variants that might contribute to observed interindividual differences in expression. A total of 40 SNPs were identified in the multiethnic, multiracial 24 sample Coriell PDR. Thirteen of these SNPs have been reported previously in the literature (see Cashman, 2004 for review) or in build 124 of the dbSNP database (http://www.ncbi.nlm.nih.gov). Of the previously reported common SNPs, g.18281G>A, resulting in a V257M substitution, does not appear to have any effect on enzyme activity (Dolphin *et al.*, 2000), although an exhaustive analysis of substrates has not been reported. In contrast, both the individual E158K and E308G variants, and the compound variant at these same two codons results in reduced FMO3 catalytic activity that appears to vary depending on the substrate (Cashman *et al.*, 1997; Park *et al.*, 2002; Lattard *et al.*, 2003). Such substrate-dependent impact of polymorphisms also has been observed for the well-studied *CYP2D6* gene (Yu *et al.*, 2002). Of the three novel SNPs resulting in non-synonymous changes in the FMO3 coding sequences, two are predicted to impact protein structure, and as such, may have a significant impact on catalytic function and/or protein stability. The E24D variant is predicted to disrupt the α-helical structure that is an integral part of the Rossmann fold known to be important for FAD binding while the N61K disrupts the β-sheet structure of a proposed membrane interaction domain. Of significant interest, substitution of a serine for asparagine at this same position results in a null trimethylamine N-oxidation phenotype causative for trimethylaminuria, although methimazole S-oxidation by this same FMO3 variant appeared normal (Dolphin *et al.*, 2000). Further, comparing the effect of both the serine and lysine

substitution on the predicted secondary structure of this peptide domain suggests a more dramatic effect of the latter amino acid (Figure 3).

Although 22 common sequence variants were identified within 50 to 100 bp of *FMO3* splice donor or acceptor sites, the predicted inability of these polymorphisms to significantly change existing splice sites or create cryptic splice sites is consistent with the observations of Lattard *et al.* (2004) who reported only two *FMO3* splice variants, *i.e.* exon 3 skipping or exon 7 skipping. Although detectable using a highly sensitive reverse-transcriptase coupled PCR DNA amplification technique, the variants reported by this group appeared to be rare relative to the abundance of the normal FMO3 transcript in all tissues examined. The observations of Lattard *et al.* (2004) also would argue against a significant impact of the variants resulting in the loss of splice site enhancer motifs in exons 1 and 2.

Given the paucity of information regarding the possible impact of *FMO3* promoter variants, much of this study focused on the 7 SNPs identified within 1.0 kbp of the transcription start site and the haplotypes inferred in three separate study populations, *i.e.*, Hispanic-Americans of Mexican descent, African-Americans, and Caucasian-Americans of northern European descent. Haplotype 2 (g.-2650C>G, g.-2543T>A, and g.-2177G>C), exhibited an 8-fold increase in *FMO3* promoter activity, and, combined with its estimated frequency, would be expected to have a significant impact on *FMO3* expression and FMO3-dependent xenobiotic metabolism within all three populations. Although haplotype 2's frequency differed among these groups (Table 7), assuming conformity to the Hardy-Weinberg equilibrium, it is anticipated 49.4%, 20.6% and 12.2% of Hispanic-Americans of Mexican descent, African-Americans, and Caucasian-Americans of northern European descent, respectively, would

possess at least one of the haplotype 2 alleles. Because haplotype 2 is a compound variant consisting of 3 SNPs, g.-2650C>G, -2543T>A and -2177G>C, consideration of this haplotype in isolation is not informative with regards to determining which SNP, or combination of SNPs is responsible for the altered activity. However, haplotype 3 consists of only the g.-2650C>G variant and exhibited promoter activity no different than the references sequence. Further, haplotype 11 consists of only the g.-2543T>A SNP and exhibited only a small, statistically insignificant reduction in promoter activity. Thus, these observations would be consistent with the g.-2177G>C SNP being responsible for the increased promoter activity observed with haplotype 2. However, two other promoter constructs were examined that were inferred as rare (<1%) haplotypes, but involved the g.-2177G>C variant: haplotype 5 consisting of the g.-2177G>C SNP alone and haplotype 7 consisting of the g.-2543T>A and g.-2177G>C SNPs. When these haplotypes were examined in *FMO3*/luciferase transient expression assays, both resulted in a 2- to 3-fold increase in promoter activity (data not shown). Although these data are consistent with the proposed role of the g.-2177G>C SNP, the quantitative difference between these two haplotypes and haplotype 2 suggests a combinatorial effect of all three SNPs in the latter.

In contrast to the increased promoter activity observed with haplotype 2, a nearly complete loss of promoter function was observed with haplotypes 8 and 15. Interestingly, haplotypes 8 and 15 were only estimated as common in the Caucasian-American and African-American study populations, respectively. Again assuming conformity to the Hardy-Weinberg equilibrium, at least one haplotype 15 allele would be observed in 3.0% of African-Americans while the haplotype 8 allele would be observed in 7.3% of Caucasian-Americans. The

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

frequencies of these functionally significant alleles in the three population groups strongly suggests that genetic diversity within the *FMO3* promoter contributes substantially to observed interindividual differences in *FMO3* expression levels. Further, the apparent absence of the loss of function haplotypes 8 and 15 in the Hispanic-American study population combined with the relative abundance in this group of the gain of function haplotype 2 allele would be consistent with *FMO3* mean expression being higher in Hispanic-Americans. Both haplotype 8 and haplotype 15 share a common SNP, i.e., g.-2106G>A, suggesting that this transition may be responsible for the dramatic loss of promoter activity. Further support for this supposition was obtained with another SNP combination that was inferred as a rare haplotype in all three populations, *i.e.*, haplotype 9 consisting of g.-2650C>G, g.-2589C>T, and g.-2106G>A. The *FMO3* reporter construct representing haplotype 9 also resulted in a near complete loss of promoter activity (data not shown). When the sequences surrounding the g.-2106G>A SNP were examined for potential transcription factor binding sites using the Match program in conjunction with the TRANSFAC Professional V8.3 database (http://www.biobase.de) (Matys *et al.*, 2003), no binding site matrices were identified in which the change at position -2106 caused a critical change in a core transcription factor binding element. Thus, insight into the molecular mechanisms responsible for the change in promoter function observed with haplotypes 8 and 15, as well as haplotype 2, will require further study.

The magnitude of loss of function for haplotypes 8 and 15, suggests both these alleles might contribute to the incidence of trimethylaminuria in the Caucasian-American and African-American populations, respectively. Studies on individuals heterozygous for structural *FMO3* null variants suggest that a 50% loss of metabolic capacity to N-oxidize trimethylamine does not

result in an overt trimethylaminuria phenotype, but only renders such individuals susceptible to a trimethylamine challenge (Zschocke *et al.*, 1999). Thus, assuming conformity to the Hardy-Weinberg Equilibrium, one would predict that only individuals homozygous for either haplotype 8 at a frequency of 0.1% in the Caucasian-American population or haplotype 15 at a frequency less than 0.1% in the African-American population would present with symptoms for this disorder. In contrast, a 50% loss of metabolic activity in the more frequent heterozygotes may significantly impact FMO3-dependent drug metabolism.

In summary, a total of 40 *FMO3* SNPs have been identified, 27 of which are novel. Seven of the identified SNPs were located within 1.0 kbp of the transcription start site and were used to infer seven common haplotypes based on individual allelic frequencies in Hispanic-American, African-American, and Caucasian-American study populations. Three of the inferred haplotypes significantly altered promoter function based on transient expression of *FMO3* reporter constructs. Further, substantial differences were observed in estimated haplotype frequencies among the three population groups studied. These observations suggest that genetic variation within *FMO3* regulatory sequences will contribute to differences in FMO3 metabolic capacity both within and among different populations. Further, these differences may well contribute to differential susceptibility to environmental toxicants and adverse drug reactions both on an individual and population basis.

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Acknowledgments

The authors thank Dr. Pippa M. Simpson, Director of Statistics, Department of Pediatrics, University of Arkansas Medical Center, Little Rock, AR, USA, for advice and assistance with the statistical analysis of data.

References

Basarab T, Ashton G H S, Menage H D and McGrath J A (1999) Sequence Variations in the Flavin-Containing Monooxygenase 3 Gene (*FMO3*) in Fish Odour Syndrome. *Brit J Dermatol* 140:164-167.

Boucher PD, Ruch R J and Hines R N (1993) Specific Nuclear Protein Binding to a Negative Regulatory Element on the Human *CYP1A1* Gene. *J Biol Chem* 268:17384-17391.

Cashman JR (2004) The Implications of Polymorphisms in Mammalian Flavin-Containing Monooxygenases in Drug Discovery and Development. *Drug Discov Today* 9:574-581.

Cashman JR (2002) Human Flavin-Containing Monooxygenase (Form 3): Polymorphisms and Variations in Chemical Metabolism. *Pharmacogenomics* 3:325-339.

Cashman JR, Bi Y A, Lin J, Youil R, Knight M, Forrest S and Treacy E (1997) Human Flavin-Containing Monooxygenase Form 3: CDNA Expression of the Enzymes Containing Amino Acid Substitutions Observed in Individuals With Trimethylaminuria. *Chem Res Toxicol* 10:837-841.

Collins FS, Brooks L D and Chakravarti A (1998) A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation. *Genome Res* 8:1229-1231.

Dolphin CT, Janmohamed A, Smith R L, Shephard E A and Phillips I R (2000) Compound Heterozygosity for Missense Mutations in the Flavin-Containing Monooxygenase 3 (FMO3) Gene in Patients With Fish-Odour Syndrome. *Pharmacogenetics* 10:799-807.

Fairbrother WG, Yeh R-F, Sharp P A and Burge C B (2002) Predictive Identification of Exonic Splicing Enhancers in Human Genes. *Science* 297:1007-1013.

Forrest SM, Knight M, Akerman B R, Cashman J R and Treacy E P (2001) A Novel Deletion in the Flavin-Containing Monooxygenase Gene (FMO3) in a Greek Patient With Trimethylaminuria. *Pharmacogenetics* 11:169-174.

Garnier J, Osguthorpe D J and Robson B (1978) Analysis of the Accuracy and Implications of Simple Methods for Predicting the Secondary Structure of Globular Proteins. *J Mol Biol* 120:97-120.

Hernandez D, Janmohamed A, Chandan P, Phillips I R and Shephard E A (2004) Organization and Evolution of the Flavin-Containing Monooxygenase Genes of Human and Mouse: Identification of Novel Gene and Pseudogene Clusters. *Pharmacogenetics* 14:117-130.

Hines RN, Hopp K A, Franco J, Saeian K and Begun F P (2002) Alternative Processing of the Human Hepatic *FMO6* Gene Renders Transcripts Incapable of Encoding a Functional Flavin-Containing Monooxygenase. *Mol Pharmacol* 62:320-325.

Hines RN, Luo Z, Hopp K A, Cabacungan E T, Koukouritaki S B and McCarver D G (2003) Genetic Variability at the Human *FMO1* Locus: Significance of a Yin Yang 1 Element Polymorphism (*FMO1*6*). *J Pharmacol Exp Ther* 306:1210-1218.

Hoogendoorn B, Coleman S L, Guy C A, Smith K, Bowen T, Buckland P R and O'Donovan M C (2003) Functional Analysis of Human Promoter Polymorphisms. *Hum Mol Genet* 12:2249-2254.

Koukouritaki SB, Simpson P, Yeung C K, Rettie A E and Hines R N (2002) Human Hepatic Flavin-Containing Monooxygenase 1 (*FMO1*) and 3 (*FMO3*) Developmental Expression. *Pediatric Res* 51:236-243.

Krueger SK and Williams D E (2005) Mammalian Flavin-Containing Monooxgenases: Structure/Function, Genetic Polymorphisms and Role in Drug Metabolism. *J Pharmacol Ther*, in press

Kyte J and Doolittle R F (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. *J Mol Biol* 157:105-132.

Lattard V, Zhang J and Cashman J R (2004) Alternative Processing Events in Human FMO Genes. *Mol Pharmacol* 65:1517-1525.

Lattard V, Zhang J, Tran Q, Furnes B, Schlenk D and Cashman J R (2003) Two New Polymorphisms of the FMO3 Gene in Caucasian and African-American Populations: Comparative Genetic and Functional Studies. *Drug Metab Dispos* 31:854-860.

Lindblad-Toh K, Winchester E, Daly M J, Wang D G, Hirschhorn J N, Laviolette J-P, Ardlie K, Reich D E, Robinson E, Sklar P, Shah N, Thomas D, Fan J-B, Gingeras T, Warrington J, Patil N, Hudson T J and Lander E S (2000) Large-Scale Discovery and Genotyping of Single-Nucleotide Polymorphisms in the Mouse. *Nat Genet* 24:381-386.

Liu H-X, Chew S L, Cartegni L, Zhang M Q and Krainer A R (2000) Exonic Splicing Enhancer Motif Recognized by Human SC35 Under Splicing Conditions. *Mol Cell Biol* 20:1063-1071.

Liu H-X, Zhang M and Krainer A R (1998) Identification of Functional Exonic Splicing Enhancer Motifs Recognized by Individual SR Proteins. *Genes Dev* 12:1998-2012.

Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel A E, Kel-Margoulis O V, Kloos D U, Land S, Lewicki-Potapov B, Michael H, Munch R,

Reuter I, Rotert S, Saxel H, Scheer M, Thiele S and Wingender E (2003) TRANSFAC: Transcriptional Regulation, From Patterns to Profiles. *Nucleic Acids Res* 31:374-378.

McCarver DG, Thomasson H, Martier S S, Sokol R J and Li T-K (1997) Alcohol Dehydrogenase-2*3 Allele Protects Against Alcohol-Related Birth Defects Among African Americans. *J Pharmacol Exp Ther* 283:1095-1101.

Overby LH, Carver G C and Philpot R M (1997) Quantitation and Kinetic Properties of Hepatic Microsomal and Recombinant Flavin-Containing Monooxygenases 3 and 5 From Humans. *Chem Biol Interact* 106:29-45.

Park C-S, Kang J-H, Chung W-G, Yi H-G, Pie J-E, Park D-K, Hines R N, McCarver D G and Cha Y-N (2002) Ethnic Differences in Allelic Frequency for Two Flavin-Containing Monooxygenase 3 (*FMO3*) Polymorphisms: Linkage and Effects on in Vivo and in Vitro FMO Activities. *Pharmacogenetics* 12:77-80.

Pastinen T and Hudson T J (2004) Cis-Acting Regulatory Variation in the Human Genome. *Science* 306:647-650.

Rogan PK, Faux B M and Schneider T D (1998) Information Analysis of Human Splice Site Mutations. *Hum Mutat* 12:153-171.

Rogan PK, Svojanovsky S and Leeder J S (2003) Information Theory-Based Analysis of *CYP2C19*, *CYP2D6* and *CYP3A5* Splicing Mutations. *Pharmacogenetics* 13:1-12.

Shehin-Johnson SE, Williams D E, Larsen-Su S, Stresser D M and Hines R N (1995) Tissue-Specific Expression of Flavin-Containing Monooxygenase (FMO) Forms 1 and 2 in the Rabbit. *J Pharmacol Exp Ther* 272:1293-1299.

Stephens M and Donnelly P (2003) A Comparison of Bayesian Methods for Haplotype Reconstruction From Population Genotype Data. *Am J Hum Genet* 73:1162-1169.

Stephens M, Smith N J and Donnelly P (2001) A New Statistical Method for Haplotype Reconstruction From Population Data. *Am J Hum Genet* 68:978-989.

Treacy EP, Akerman B R, Chow L M L, Youil R, Bibeau C, Lin J, Bruce A G, Knight M, Danks D M, Cashman J R and Forrest S M (1998) Mutations of the Flavin-Containing Monooxygenase Gene (*FMO3*) Cause Trimethylaminuria, a Defect in Detoxication. *Hum Mol Genet* 7:839-845.

Whetstine JR, Yueh M-F, Hopp K A, McCarver D G, Williams D E, Park C-S, Kang Y-N, Cha Y-N, Dolphin C T, Shephard E A, Phillips I R and Hines R N (2000) Ethnic Differences in Human Flavin-Containing Monooxygenase 2 (*FMO2*) Polymorphisms: Detection of Expressed Protein in African-Americans. *Toxicol Appl Pharmacol* 168:216-224.

Wrighton SA, Brian W R, Sari M-A, Iwasaki M, Guengerich F P, Raucy J L, Molowa D T and Vandenbranden M (1990) Studies on the Expression and Metabolic Capabilities of Human Liver Cytochrome P450IIIA5 (HLp3). *Mol Pharmacol* 38:207-213.

Yeung CK, Lang D H, Thummel K E and Rettie A E (2000) Immunoquantitation of FMO1 in Human Liver, Kidney, and Intestine. *Drug Metab Dispos* 28:1107-1111.

Yu A, Kneller B M, Rettie A E and Haining R L (2002) Expression, Purification, Biochemical Characterization and Comparative Function of Human Cytochrome P450 2D6.1, 2D6.2, 2D6.10, and 2D6.17 Allelic Forms. *J Pharmacol Exp Ther* 303:1291-1300.

Zheng Y-M, Henne K R, Charmley P, Kim R B, McCarver D G, Cabacungan E T, Hines R N and Rettie A E (2003) Genotyping and Site-Directed Mutagenesis of a Cytochrome P450

Meander Pro-X-Arg Motif Critical to CYP4B1 Catalysis. *Toxicol Appl Pharmacol* 186:119-126.

Ziegler DM (2002) An Overview of the Mechanism, Substrate Specificities, and Structure of FMOs. *Drug Metab Rev* 34:503-511.

Zschocke J, Kohlmüeller D, Quak E, Meissner T, Hoffmann G F and Mayatepek E (1999) Mild Trimethylaminuria Caused by Common Variants in *FMO3* Gene. *Lancet* 354:834-835.

Footnotes

This work was supported in part by U.S. Public Health Service grant number CA53106 from the National Cancer Institute, as well as funds from the Children's Hospital of Wisconsin Foundation.

Send reprint requests to: Ronald N. Hines, Ph.D., Department of Pediatrics, Medical College of Wisconsin, 8701 Watertown Plank Rd., Milwaukee, WI 53226. E-mail: rhines@mail.mcw.edu

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

Figure Legends

Fig. 1. Potential Impact of previously unreported Human FMO3 Nonsynonymous Structural Variants. Pairwise alignment of FMO3 protein sequences from *Arabidopsis thaliana* (BT000473), *Gallus gallus* (AJ431490), *Bos taurus* (NM_174057), *Homo sapiens* (NM_006894), *Canis familias* (AF384054), *Oryctolagus cuniculus* (L10391), *Mus musculus* (U87147), and *Rattus norvegicus* (NM_053433) and progressive assembly using Neighbor-joining phylogeny was accomplished using the Align program in Clone Manager Suite V7.1 (Scientific and Educational Software). Protein motifs were compiled by Krueger and Williams (Krueger and Williams, 2005). The position of the amino acid changes resulting from the identified SNPs are highlighted.

Fig. 2. Predicted FMO3 secondary structure changes resulting from the E24D variant. The conserved FMO3 Rossmann fold domain involved in FAD binding (Figure 1) was analyzed for predicted secondary structure changes caused by the E24D variant using the algorithm developed by Garnier *et al.* (1978). Tendency toward α-helix, β-sheet, and turns are shown as line plots for both the E24 reference (solid line) and D24 variant (dashed line) peptides. The resulting predicted secondary structure for both peptides is shown in boxes above the plots.

Fig. 3. Predicted FMO3 secondary structure changes resulting from the N61K and N61S variants. The conserved FMO3 peptide, proposed to form a dimer loop, membrane interaction domain was analyzed for predicted secondary structure changes caused by the N61K variant

identified in this study, as well as the N61S rare variant reported to result in a trimethylamine N-oxidation null phenotype (Dolphin *et al.*, 2000). The algorithm developed by Garnier *et al.* (1978) was used. Tendency toward α-helix, β-sheet, and turns are shown as line plots for both the N61 reference (solid line), K61 variant (dotted line), and S61 variant (dashed line) peptides. The resulting predicted secondary structure for all three peptides is shown in boxes above the plots.

Fig. 4. Functional Analysis of Human FMO3 Upstream Variants. Haplotype variants were introduced into a reporter construct containing *FMO3* position -988 (-2762 relative to ATG start codon) to +42 (-1732 relative to ATG start codon) directing luciferase expression. Plasmids were used to transfect HepG2 cells which were subsequently analyzed for transient luciferase expression. Individual plasmid numbers and the haplotypes they represent (see Materials and Methods) are shown on the ordinate. Luciferase activities were normalized for transfection efficiency using co-transfected β-galactosidase activity and within each experiment, normalized to the activity observed with the reference haplotype 1. Data represent the mean ± SD from a minimum of 6 experiments. (** indicates $P < 0.01$; ANOVA, Dunnett's post test).

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

TABLE 1

FMO3 Template Amplification Primers for SNP Discovery

Primer Set	Primer Sequence	*FMO3* Coordinates[a]	PCR[b]	Target
1	5' TTCCTGGTACTAATAGATCA 3'	-2310 to -2291	A	Upstream
	5' TCTGTGTGTCTACGTCCT 3'	-1716 to -1733		
2	5' ATCCTCTAATCCTTGTTAAA 3'	-2733 to -2714	A	Upstream
	5' CCAATAAGGAGGATGACT 3'	-2145 to -2162		
3	5' TGGGAGACTGGCCTACAG 3'	-1906 to -1889	B	Exon 1
	5' GAGACGGAGTTTCGCTTTTA 3'	-1455 to -1436		
4	5' ATTAAGCCAAAGAGCGAAAT 3'	-102 to -83	B	Exon 2
	5' TTTCAAAGCCCATTGTCTAA 3'	+334 to +315		
5	5' GCCCTGACCATGATCAGTAT 3'	+11062 to +11081	B	Exon 3
	5' GAGAGGCCACACTGTTAAA 3'	+11442 to +11424		
6	5' ATATTTTTCTTAACCCAC 3'	+14920 to +14937	C	Exon 4
	5' TAATTTAACCTGCTTATACT 3'	+15255 to +15274		
7	5' CATCTATTCACAAGGTCGC 3'	+15336 to +15354	B	Exon 5
	5' CCCCACATTTCATATCACAC 3'	+15646 to +15627		
8	5' CTGCAGCTGGGGTAATAG 3'	+18032 to +18049	B	Exon 6

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

	5' CCAGCAGGCATATCACGTTT 3'	+18371 to +18352		
9	5' ATTTATATATGGACCAAT 3'	+21276 to +21293	D	Exon 7
	5' GACCTTGTAACTAGGATTAT 3'	+21796 to +21777		
10	5' TAGAACTGAATTTGGTGTCT 3'	+23412 to +23431	A	Exon 8
	5' GCTTGTAGTTGTCATTCC 3'	+23743 to +23726		
11	5' TGTTGTATGTCAGGGTAGT 3'	+24288 to +24306	B	Exon 9
	5' CTTTCAGAAATCCTAGAGAA 3'	+24810 to +24791		

[a] Coordinates are based on assigning the "A" of the ATG translation initiation codon +1.

[b] All PCR conditions (A, B, C, D) had an initial denaturation step at 94°C for 2 min followed by 35 cycles of: A, denaturation at 94°C for 30 s, annealing at 50°C for 30 s, and elongation at 72°C for 1 min; B, denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 1 min; C, denaturation at 94°C for 30 s, annealing at 47°C for 30 s, and elongation at 72°C for 1 min; D, denaturation at 94°C for 30 s, annealing at 48°C for 30 s, and elongation at 72°C for 1 min. All PCR conditions (A, B, C, D) had a final elongation step at 72°C for 2 min.

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

TABLE 2.

Single Base Extension Primers

Primer	SNP Queried	Quantity in Assay (pmol)
5' $(C)_{28}$TCCCTATTTTATTAATGAGTAAATCAA 3'	-2650C>G	20.0
5' $(C)_{8}$CATCTCTGAAGTGTGAGTATGA 3'	-2589C>T	1.0
5' $(C)_{4}$AATTTTGAGGAATTACCCTTC 3'	-2543T>A	6.0
5' $(C)_{20}$GGAGGATGACTTTTCTCTGCTCTCT 3'	-2177G>C	1.0
5' $(C)_{13}$AGCTAAGTTTCTACAGTATTTG 3'	-2106G>A	1.0
5' $(C)_{23}$CCTAGTTCTTTTGTGGGTAGCTCATAG 3'	-2099A>G	0.1
5' $(C)_{18}$TGGTAAACACAAGCTGCTGGTT 3'	-1961T>C	30.0

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

TABLE 3

Human *FMO3* SNP Discovery - Exon Sequences

Position[a]	SNP and Sequence Context	Location	Accession No.[b]	Amino Acid Change
-1732	GGACA [G>T] GACGT	Exon 1		--
72	GAAGA [G>T] GGGCT	Exon 2		E24D
11177	TCCAA [C>A] TCTTC	Exon 3		N61K
15019	TAGAC [A>C] TTTGT	Exon 4		T108T
15136	TGTTC [C>T] GGACA	Exon 4	rs1800822	S147S
15167[c]	CAAAA [G>A] AGTCC	Exon 4	rs2266782	E158K
15437	TTTAA [A>G] GGCAA	Exon 5		K167K
18281[c]	TGTAC [G>A] TGAAG	Exon 6	rs1736557	V257M
21375	TTTAA [C>T] GATGA	Exon 7	rs909530	N285N
21443[c]	CACAG [A>G] GACCT	Exon 7	rs2266780	E308G
23613	AAAAA [G>T] CGCAA	Exon 8		K416N

[a] Coordinates based on assigning the "A" of the ATG translation initiation codon +1.

[b] Accession number reported in dbSNP build 124

[c] Previously reported (Cashman et al., 1997; Treacy et al., 1998; Cashman et al., 2000; Dolphin et al., 2000; Park et al., 1999).

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

TABLE 4

Human *FMO3* SNP Discovery - Intron Sequences

Position[a]	SNP and Sequence Context	Location[b]	Accession No.[c]
-1639	CAGGA [G>T] CTTGA	Intron 1	
-1592	CCAAA [G>A] ATACA	Intron 1	rs1736561
-1536	TGGGA [G>T] GCTGA	Intron 1	
154	TGTAA [T>G] AGACA	Intron 2	
177	GGGAA [G>T] TGTAT	Intron 2	
187	TAAGA [G>T] CACAC	Intron 2	
191	AGCAC [A>G] CTGTG	Intron 2	rs2064074
11322	TAAGA [T>G] GTTAT	Intron 3 Splice Donor (+6)	
14916	TAATC [A>C] TTAAA	Intron 3	
14951	TTTTC [A>C] TACTG	Intron 3	
14982	TAGCA [T>G] AGAAA	Intron 3	
15366	TTAAA [G>A] TCTTT	Intron 4	rs2066529
15399[d]	GCCAT [G>A] TATTT	Intron 4 Splice Acceptor (-21)	rs1920149
15573	TACTC [C>G] CCGGG	Intron 5	rs2066534
15610	GAAGA [G>T] TTATT	Intron 5	

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

18088	CCAGA [delA] TATCC	Intron 5	
18102	ACAAA [T>G] GGTCA	Intron 5	
21738	GAAGA [T>C] GAATG	Intron 7	rs909531
23525	TTACC [A>C] TCGTG	Intron 7	
23532	CGTGT [C>A] TTTCC	Intron 7 Splice Acceptor (-16)	
23628	TAAGA [G>T] TACCT	Intron 8 Splice Donor (+6)	
24416	CCACA [G>T] TGGTG	Intron 8 Splice Acceptor (-24)	

[a] Coordinates based on assigning the "A" of the ATG translation initiation codon +1

[b] Location of SNP within the splice donor or acceptor consensus sequence (Rogan et al., 1998)

[c] Reported in dbSNP build 124

[d] Previously reported (Basarab et al., 1999)

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

TABLE 5

Human *FMO3* SNP Discovery – 5' Flanking Sequences

Position[a]	SNP and Sequence Context	Accession No.[b]
-2650	ATCAA [C>G] GATTG	rs1736560
-2589	AATCT [C>T] TCATA	
-2543	AGGCA [T>A] GAAGG	rs12404218
-2177	GAAAA [G>C] AGAGA	rs3754491
-2106	ATTTG [G>A] AGGAC	rs16864006
-2099	GGACT [A>G] CTATG	
-1961	GAGTA [T>C] AACCA	rs16864007

[a] Coordinates based on assigning the "A" of the ATG translation initiation codon +1.

[b] Reported in dbSNP build 124

Molecular Pharmacology Fast Forward. Published on April 27, 2005 as DOI: 10.1124/mol.105.012062
This article has not been copyedited and formatted. The final version may differ from this version.

TABLE 6

FMO3 Upstream Genetic Variation - Population-Specific Allelic Frequencies

SNP	Variant Allelic Frequency (95% CI)		
	Hispanic American (n = 402)	African-American (n = 402)	Caucasian-American (n = 400)
-2650C>G	0.438 (0.388,0.488)[a,b]	0.259 (0.216,0.305)[c]	0.303 (0.258,0.350)[c]
-2589C>T	0.007 (0.000,0.016)[b]	0.007 (0.000,0.016)[b]	0.040 (0.023,0.064)[a,c]
-2543T>A	0.291 (0.247,0.338)[a,b]	0.192 (0.155,0.234)[c]	0.205 (0.167,0.248)[c]
-2177G>C	0.286 (0.243,0.334)[a,b]	0.124 (0.094,0.161)[b,c]	0.068 (0.045,0.097)[a,c]
-2106G>A	0.012 (0.002,0.023)[b]	0.027 (0.014,0.048)	0.040 (0.023,0.064)[c]
-2099A>G	0.000 (0.000,0.000)	0.000 (0.000,0.000)	0.000 (0.000,0.000)
-1961T>C	0.012 (0.002,0.023)	0.020 (0.009,0.039)	0.005 (0.000,0.012)

[a] Different from African-American study population, $P<0.016$ (Fisher's exact test, Bonferroni adjustment)

[b] Different from Caucasian-American study population, $P<0.016$ (Fisher's exact test, Bonferroni adjustment)

[c] Different from Hispanic-American study population, $P<0.016$ (Fisher's exact test, Bonferroni adjustment)

TABLE 7

Common (>1%) *FMO3* Upstream Haplotypes

Haplotype		Mean Frequency ± SEM		
Number	Nucleotide Changes	Hispanic American (n=402)	African-American (n=402)	Caucasian American (n=400)
1	Reference Sequence	54.1 ± 0.2	68.9 ± 0.6[a]	55.4 ± 0.7
2	-2650C>G, -2543T>A, -2177G>C	28.9 ± 0.2[a]	10.9 ± 0.4[a]	6.3 ± 0.4[a]
3	-2650C>G	15.3 ± 0.2[a]	8.3 ± 0.5[a]	19.8 ± 0.7[a]
4	-2650C>G, -2543T>A	R[b]	4.7 ± 0.4	4.3 ± 0.6
8	-2589C>T, -2106G>A	R	N[c]	3.7 ± 0.3
11	-2543T>A	N	2.9 ± 0.4[a]	9.2 ± 0.6[a]
15	-2106G>A, -1961T>C	N	1.5 ± 0.3	N

[a] Significantly different from other population groups, $p<0.001$

[b] Haplotype inferred, but rare (<1%)

[c] Haplotype not inferred

Binding site for ADP moiety of FAD

```
                                                  e24d
                                                   ▼
Arabidopsis     11 khvavigagaaglvtarelrregh
Gallus           3 rrvavvgagisglaatkccleegl
Bos taurus       3 kkvaiigagisglasirncleegl
Homo             3 kkvaiigagvsglasirscleegl
Canis            3 krvaiigagvsglasirscleegl
Oryctolagus      3 kkvaiigagisglasirscleegl
Mus              3 kkvaiigagvsglaairscleegl
Rattus           3 rkvavigagvsglaairscleegl
```

Dimer loop, possible membrane interaction

```
                          n61k
                           ▼
Arabidopsis     74 yeslrtnlprecmgftdfpfvprihdisrd
Gallus          55 yrtvftnsckemmcypdfpf-------pdd
Bos taurus      55 yrsvftnsskemtcfpdfpf-------pdd
Homo            55 yksvfsnsskemmcfpdfpf-------pdd
Canis           55 yqsvftnsskemmcfpdfpy-------pdd
Oryctolagus     55 yqsvftnsskemmcfpdfpf-------pdd
Mus             55 yqsvftnsskemmcfpdfpy-------pdd
Rattus          55 yqsvftnsskemmcfpdfpy-------pdd
```

FMO isoform-specific substrate cleft

```
                                                  k416n
                                                   ▼
Arabidopsis    351 rvilpsqdkmmediiewyatldvlgipkrhthklgkisce
Gallus         397 lctlpsvnemlediee------kkrnkirwfgtsntlqtd
Bos taurus     396 tcplpsvkdmmndide------kmgkklklfgksdtiqtd
Homo           396 tctlpsmedmmndine------kmekkrkwfgksetiqtd
Canis          396 tctlpsvtdmmndidk------kregklkwfgtsetvqtd
Oryctolagus    396 tctlppvkdmmndihe------kmgtklkwfgksetiqtd
Mus            396 tctlpsvndmmddide------kmgekfkwygnsttiqtd
Rattus         396 tcilpsvndmmddide------kmgkklkwfgnsttiqtd
```

Fig. 2

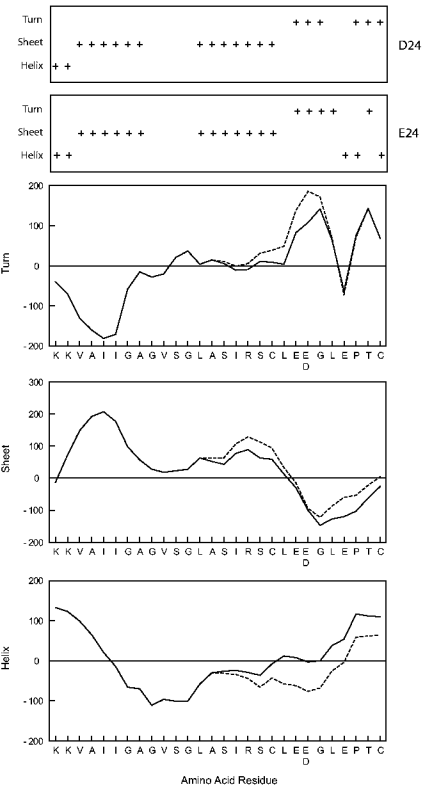

Turn
Sheet
Helix
D24
Turn
Sheet
Helix
E24
Turn
200
100
0
-100
-200
K K V A I I G A G V S G L A S I R S C L E E G L E P T C
D
Sheet
300
200
100
0
-100
-200
K K V A I I G A G V S G L A S I R S C L E E G L E P T C
D
Helix
200
100
0
-100
-200
K K V A I I G A G V S G L A S I R S C L E E G L E P T C
D
Amino Acid Residue

Fig. 3

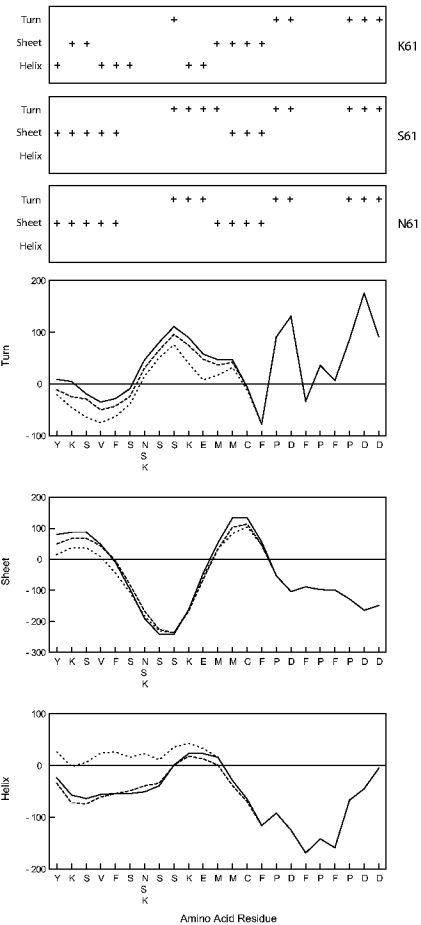
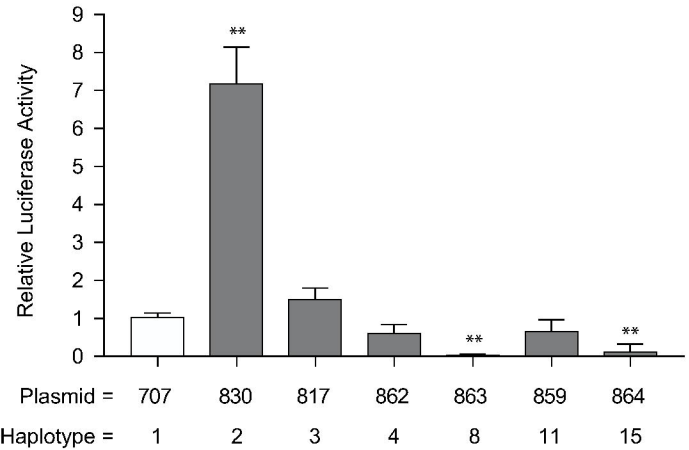

Turn
Sheet
Helix
K61
S61
N61
Turn
Sheet
Helix
Y K S V F S N S S K S K E M M C F P D F P F P D D
Amino Acid Residue

Fig. 4


Relative Luciferase Activity
9
8
7
6
5
4
3
2
1
0
**
**
**
Plasmid = 707 830 817 862 863 859 864
Haplotype = 1 2 3 4 8 11 15