FLAVIN-CONTAINING MONOOXYGENASE 3 (FMO3): EVOLUTION, PROTEIN STRUCTURE AND CAUSATIVE MUTATIONS OF TRIMETHYLAMINURIA.

Charles Kenneth John Allerston

Department of Biochemistry and Molecular Biology

University College London

A thesis submitted for the degree of Doctor of Philosophy

September 2007

UMI Number: U591392

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U591392
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Statement

I, Charles Kenneth John Allerston confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Flavin-containing monooxygenase 3 (FMO3) is a hepatic, microsomal enzyme important in the Phase I metabolism of xenobiotics. Numerous single nucleotide polymorphisms (SNPs) in FMO3 have been identified and shown to affect the catalytic activity of the enzyme, with some destroying the activity altogether, causing the distressing disorder, Trimethyalminuria (TMAU), in humans.

Genotyping was performed at the *FMO3* locus on several individuals displaying symptoms of TMAU. Two new FMO3 variants, R238Q and R492Q, were discovered in this investigation, with kinetic studies suggesting that R238Q destroys FMO3 catalysis.

Kinetic studies of FMO3 common polymorphisms showed that the E158K/E308G displayed significantly reduced catalytic activity compared with either single variant alone. E158K/V257M also displayed significantly reduced catalytic activity, but not compared to either single variant alone. The pharmacogenetic implications are discussed with particular reference to the recent finding that E158K/E308G has been associated with reduced polyp formation in patients with familial adenomatous polyposis who were treated with sulindac sulphide, an FMO3 substrate.

The evolutionary history of *FMO3* was probed using sequence data of genomic DNA in a Japanese cohort of potential TMAU sufferers and control individuals. Mutational relationships among haplotypes were inferred and time depth of the variation and ages of individual mutations were estimated by coalescent analysis, with test statistics used to detect departure from neutral evolution. A case of balancing selection is proposed at the *FMO3* locus.

In an attempt to understand the structural and biophysical consequences of FMO3 variants, a homology model of FMO3 was generated, refined and validated. Flavin adenine dinucleotide (FAD), an FMO3 cofactor, was also modeled into the FMO3 model and the interactions between the enzyme and cofactor predicted.

Acknowledgements

First and foremost I would like to thank Prof. Elizabeth Shephard for her support, guidance and supervision over the course of my Ph.D. and for giving me the opportunity to explore such an interesting topic. I could not have hoped for a more supportive supervisor.

I wish to thank Dr. Azara Janmohamed for teaching me all the lab basics and day to day skills that allowed me to progress to this stage. Her tireless support, patience and friendship over these years have been very much appreciated. My thanks are also extended to those members of the Shephard lab who have helped me over the years.

I wish to thank Prof. Ian Phillips at Queen Mary University London for all of his help and advice, particularly with the Japanese genetics project.

I am extremely grateful to Prof. Paul Ortiz de Montellano who hosted my Bogue Fellowship at the University of California San Francisco and was very generous with both his time and facilities. My thanks also go to his lab members in San Francisco.

I would like to thank Prof. David Goldstein and Dr. Chris Taylorson for acting as my mentors during my Ph.D.

Thanks go to Mr. Frank Penter and Mr. Peter Cadley at UCL for sending me academic literature during my fellowship and Mr. Brian Watts at UCL for I.T. support.

I offer my extreme thanks to Dr. Andrias O'Reilly of Birkbeck College for introducing me to the MODELLER software and being so generous with his time. I also thank Dr. Abi Jones for preparing the bucal swab kit for the patient study.

I would like to thank the Medical Research Council, the Bogue Fellowship Committee and Bioline Ltd. for financial support. I would also like to thank the Bioline team very much for the invaluable industrial experience I gained.

Finally I would like to thank my wife Kelly for sharing this experience with me, smiling through my rants about bacteria not growing, computer programs crashing etc. and always reminding me about what is really important in life.

Contents

List of Figures

List of Tables

Chapter One: Introduction

Chapter 2: Materials and Methods

Chapter 3: Results and Discussion

List of Abbreviations.

Abbreviation	Definition
Afr	African American
Ala	Alanine
Amp	Ampicillin
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
BBC	Bitish Broadcasting Corporation
BLAST	Basic Local Alignment Search
bp	Base Pairs
BP	Before Present
BSA	Bovine Serum Albumin
Cauc	Caucasian
cDNA	Complemetary DNA
Con	Control group
CYP	Cytochrome P450 Monooxygenase
Cys	Cysteine
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphospate
ddH_2O	double-distilled H_2O
DTT	Dithiotreitol
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DOPE	Discrete optimised protein energy
DOS	Disk Operating System
dsDNA	Double stranded DNA
DTNB	5,5'-Dithio-bis(2-nitrobenzoic acid)
dTTP	Deoxythymidine triphosphate
DX	Diagnosis
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinucleotide
FADHOOH	Hydroperoxyflavin

FMO	Flavin-containing Monooxygenase
G6PD	*Glucose-6-phosphate dehydrogenase*
GHz	Gigahertz
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GP	General Medical Practitioner
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hFMO3	human form of Flavin-containing Monooxygenase (form 3)
His	Histidine
His_6-tag	Hexahistadine-tag
Hisp	Hispanic American
hrs	hours
Ile	Isoleucine
IPTG	Isopropyl bD-thiogalactopyranoside
LB	Luria Broth
LD	Linkage disequilibrium
Leu	Leucine
LINE	Long interspersed nucleotide element
Lys	Lysine
Kyr	Thousand years
MB	Megabyte
MBP	Maltose-Binding Protein
Met	Methionine
MHC	Major histocompatibility complex
MOPSO	3-(N-morpholino)-2-hydroxypropane-sulphonic acid
mRNA	Messenger RNA
NADPH	β-Nicotinamide adenine dinucleotide phosphate (reduced)
ND	Not determined
N_e	Effective Population Size
NT	Not tested.
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PC	Personal Computer

PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
Phe	Phenylalanine
PMSF	Phenylmethylsulfonyl fluoride
Pro	Proline
RAM	Read Access Memory
RM	Reduced Median
RNA	Ribonucleic acid
S.D.	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Ser	Serine
SNP	Single nucleotide polymorphism
S.pombe	*Schizosaccharomyces pombe*
TBS	Tris Buffered Saline
TEMED	N,N,N',N'-Tetramethylethylenediamine
TEV	Tobacco Etch Virus
Thr	Threonine
T_{MRCA}	Time to Most Common Recent Ancestor
TMA	Trimethylamine
TMAU	Trimethylaminuria
Tris	Tris[hydroxymethyl]aminomethane
Trp	Tryptophan
Tyr	Tyrosine
UK	United Kingdom of Great Britain and Northern Ireland
UV	Ultra violet
USA	United States of America
Val	Valine

Chapter 1

Introduction

1. Introduction

Xenobiotic metabolism refers to the biochemical modification *in vivo* of exogenous substances. These modifications are made through a series of enzyme systems, which often convert hydrophobic molecules into more readily excreted, polar metabolites. The primary site of xenobiotic metabolism is the liver, although virtually all tissues carry out xenobiotic metabolism to a greater or lesser extent. Xenobiotic metabolism can be considered in two phases. Phase I metabolism is the stage at which polar groups are either introduced or exposed in the xenobiotic by oxidation by Cytochrome P450 Monooxygenases (CYPs), Flavin-containing Monooxygenases (FMOs), Alcohol and Aldehyde Dehydrogenases, Monoamine Oxidase or Peroxidases or hydrolysis by Esterases, Amidases or Epoxide Hydrolase. Phase II metabolism is concerned with the conjugation of the modified product of Phase I metabolism to produce a water soluble conjugate, readily excreted from the body. The majority of Phase II reactions are catalysed by Transferases e.g. Glutathione *S*-Transferases, Sulfotransferases or UDP Glucuronyl Transferases.

After the CYPs, FMOs are the largest group of enzymes involved in Phase I metabolism of drugs and other xenobiotics and provide the focus for this investigation.

1.1. FMOs

FMOs (EC 1.14.12.8) are β-Nicotinamide adenine dinucleotide phosphate (reduced) (NADPH)-dependent enzymes that catalyse the oxidation of a wide range of compounds. Flavin adenine dinucleotide (FAD) acts as a prosthetic group within the FMO molecule and is reduced by NADPH. This reduced form of FAD readily binds molecular oxygen to generate a hydroperoxyflavin (FMO-

FADHOOH). When in this state, the enzyme acts in a "cocked gun" fashion, ready to attack the soft, nucleophilic centre of chemicals containing nitrogen, sulphur, selenium or phosphorus (Figure 1). This mechanism distinguishes FMOs from other monooxygenases as the activation of oxygen, in the form of the C(4a) hydroperoxide derivative of FAD, does not require binding of the oxygenatable substrate [1]. This is theorised to be the basis of the broad range of substances that are FMO substrates. including hydrazines [2], phosphines [3], iodine boron-containing compounds [4], sulphides [5], selenides [6] and an array of amines [7]. There is evidence of a number of endogenous compounds being FMO substrates, such as methionine [8], cysteamine and cysteine- and homocysteine-*S*-conjugates [1].

In humans, five *FMO* genes are known to encode protein (*FMOs 1-5*) [9]. *FMOs* 1,2,3 and 4 are found in a cluster with *FMO6*, a putative pseudogene, at 1q23-4 [10]. *FMO5* is located further away from this cluster at 1q21. A second cluster on chromosome 1 contains five *FMO* pseudogenes[11], *FMO7P* to *FMO11P*, which are located within a cluster ~4 Mb to the centromeric side of the functional *FMO* gene cluster [11] and appear to have arisen as a result of a locus duplication event. FMOs 1-5 exhibit between 50 to 59% amino acid identity across mammalian species [12] including humans (Table 1). FMOs of other mammalian species have >78% sequence identity to their human orthologues (Table 1). Two GXGXXG motifs, characteristic of FAD- and NADPH-pyrophosphate-binding sites are present at identical positions (residues 9-14 and 191-196, respectively) in mammalian FMOs. Figure 2 shows a primary sequence alignment of the five, functional FMOs in humans. The FAD-binding site is contained within a highly conserved motif that predicts a βαβ secondary

Figure 1. Major steps in the catalytic cycle of FMO [1]

S and SO are the xenobiotic substrate and the oxygenated product respectively. No substrate binding is required for the cycle to begin. (1) FAD-OOH reacts with the nucleophile within the enzyme active site. (2) One atom of molecular O_2 is incorporated into the substrate and the other is released as H_2O. (3) The system regenerates.

Species		Human					Chimpanzee					Dog					Rabbit					Rat				
	FMO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Human	1	-	58	53	52	51	**99**	58	54	53	51	**89**	57	55	52	51	**86**	58	55	52	53	**82**	57	53	51	51
	2	58	-	57	55	55	58	**99**	57	57	56	58	**86**	58	54	56	56	**86**	57	55	56	57	**84**	58	55	57
	3	53	57	-	52	56	54	57	**99**	54	55	55	58	**83**	53	54	53	58	**83**	53	54	54	57	**81**	53	54
	4	52	55	52	-	54	51	55	53	**99**	54	53	55	53	**86**	52	54	54	53	**84**	52	50	54	53	**80**	53
	5	51	56	56	54	-	51	57	55	54	**99**	52	58	52	53	**89**	51	57	52	52	**85**	51	57	54	51	**84**
Chimpanzee	1	**99**	58	54	51	51	-	57	54	50	50	**88**	57	55	53	50	**85**	58	54	52	52	**82**	57	53	51	51
	2	58	**99**	57	55	57	57	-	57	57	57	58	**87**	58	54	56	56	**87**	58	55	56	57	**85**	58	55	57
	3	54	57	**99**	53	55	54	57	-	54	55	55	58	**83**	53	54	54	58	**83**	54	55	54	59	**81**	52	53
	4	53	57	54	**99**	54	50	57	54	-	51	54	53	55	**79**	51	54	56	54	**85**	54	51	56	54	**82**	53
	5	51	56	55	54	**99**	50	57	55	51	-	51	56	53	51	**89**	51	56	52	52	**85**	50	57	51	51	**84**
Dog	1	**89**	58	55	53	52	**88**	58	55	54	51	-	57	56	54	51	**87**	57	55	53	52	**87**	57	56	53	52
	2	57	**86**	58	55	58	57	**87**	58	53	56	57	-	58	54	56	56	**87**	59	57	58	57	**84**	58	55	57
	3	55	58	**83**	53	52	55	58	**83**	55	53	56	58	-	54	54	54	59	**84**	55	54	55	57	**83**	53	54
	4	52	54	53	**86**	53	53	54	53	**79**	51	54	54	54	-		53	54	53	**83**	53	51	54	53	**78**	51
	5	51	56	54	52	**89**	50	56	54	51	**89**	51	56	54		-	51	55	53	52	**84**	51	59	52	50	**83**
Rabbit	1	**86**	56	53	54	51	**85**	56	54	54	51	**87**	56	55	53	51	-	55	54	53	52	**83**	56	53	52	51
	2	58	**86**	58	54	57	58	**87**	58	56	56	57	**87**	53	54	55	55	-	58	53	57	56	**86**	58	52	51
	3	55	57	**83**	53	52	54	58	**83**	54	52	55	59	**84**	53	53	54	58	-	54	56	55	57	**85**	52	54
	4	52	55	53	**84**	52	52	55	54	**85**	52	53	57	55	**83**	52	53	53	54	-		51	55	53	**79**	52
	5	53	56	54	52	**85**	52	56	55	54	**85**	52	58	54	53	**84**	52	57	56	52	-	52	56	56	51	**83**
Rat	1	**82**	57	54	50	51	**82**	57	54	51	50	**87**	57	55	51	51	**83**	56	55	51	52	-	56	53	51	51
	2	57	**84**	57	54	57	57	**85**	59	56	57	57	**84**	57	54	59	56	**86**	57	55	56	56	-	57	55	57
	3	53	58	**81**	53	54	53	58	**81**	54	51	56	58	**83**	53	52	53	58	**85**	53	56	53	57	-	52	52
	4	51	55	53	**80**	51	51	55	52	**82**	51	53	55	53	**78**	50	52	52	52	**79**	51	51	54	52	-	51
	5	51	57	54	53	**84**	51	57	53	53	**84**	52	57	54	51	**83**	51	51	54	52	**83**	51	57	52	51	-

Table 1. The primary sequence identity (%) between mammalian FMOs. **Orthologues** are shown in **bold** and paralogues are shown underlined.

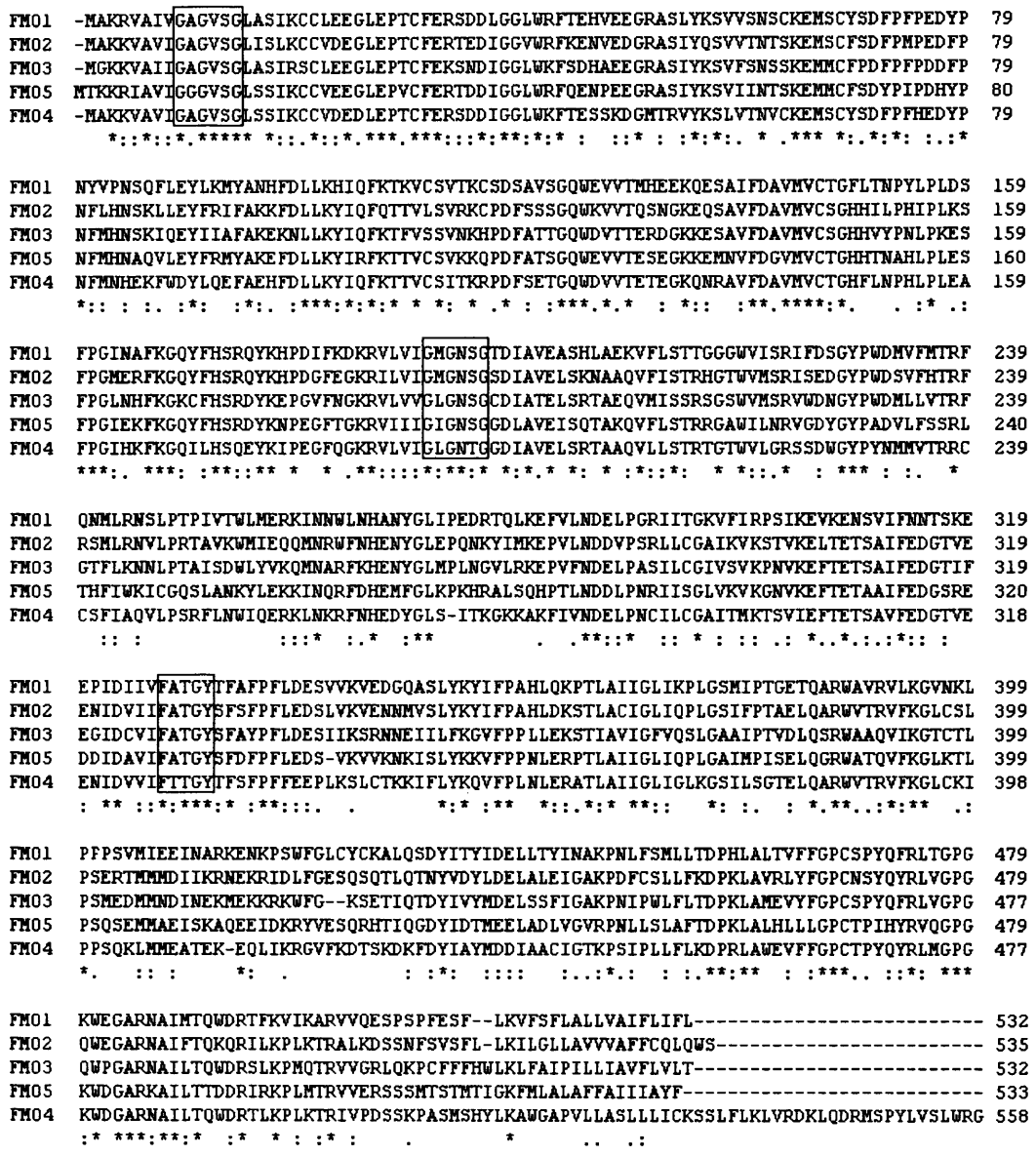


Figure 2. A primary sequence alignment of the five functional FMOs in humans.

The primary sequences of FMOs 1-5 are aligned. An asterix denotes conserved identical residues, a colon indicates conserved similar residues, and a dot indicates highly similar residues. The 'GXGXXG' FAD and NADPH binding moieties are boxed (see text) along with the FMO characteristic 'FATGY' motif.

structure, a Rossman fold, known to be involved in binding dinucleotides [13]. Phylogenetic analysis suggests that the five functional *FMO* genes arose from a common ancestral gene via a series of gene duplications, estimated to have occurred between 210 and 275 million years ago [11]. This is much earlier than the divergence of mammals, some 85 million years ago, and therefore it is predicted that all mammals possess the five functional *FMO* genes 1 to 5 [11].

Unlike *CYPs*, *FMOs* are not inducible [14, 15]. *FMOs* exhibit distinct developmental and tissue-specific expression [14, 16, 17], which are described in the following sections.

FMO1

FMO1, the first FMO purified, was identified as a mixed-function amine oxidase in porcine liver microsomes. Its substrates include nitrogen- and sulphur-containing compounds [18]. A cDNA for the corresponding protein of humans was isolated in 1991 [19] and was formally named as FMO1 in 2004 [12]. *FMO1* encodes a polypeptide of 532 amino-acid residues of molecular mass 60,306.

FMO1 is expressed in abundance in the liver of all mammals examined to date [20-23], with the exception of the adult human liver where the mRNA [9, 14] and the protein [24, 25] is not found. The lack of *FMO1* expression in the human adult liver is thought to be due to the insertion of several long interspersed nucleotide elements (LINEs), just upstream of the proximal promoter [26]. *FMO1* is expressed however in the human foetal liver [9, 14, 16, 24, 25] where the expression is highest in the first trimester (7.8 ± 5.3 pmol/mg microsomal protein), then declines during foetal development and by 3 days after birth is completely extinguished [24]. *FMO1* is also expressed in the foetal human kidney, but this expression is increased after birth [9, 14, 19, 25, 27].

In human, *FMO1* is expressed mainly in the kidney [9, 14, 19, 25] but it is also expressed in the intestinal mucosa, the mammary and salivary glands, and in a number of endocrine tissues, including pancreas, adrenal cortex and medulla, thyroid, thymus and testis [11]. It is thought that FMO1 is likely to be the major contributor to renal xenobiotic metabolism as observed amounts of FMO1 in adult human kidney (47 ± 9 pmol/mg microsomal protein) [25] are not much lower than that observed in the liver for the major hepatic cytochrome P450 (CYP), CYP3A4 (96 ± 51 pmol/mg microsomal protein) [28] and are greater than that of the total content of CYPs in adult human kidney [29].

FMO2

FMO2 is expressed predominantly in the foetal and adult lung of several mammals including our closest evolutionary relative, the chimpanzee [23, 30, 31], but in humans the gene encodes a truncated, non-functional protein [32]. In comparison with FMO2 of the rabbit [23], guinea pig [30] and rhesus macaque [31], the human enzyme lacks 64 residues from its carboxy terminus resulting from a C>T mutation changing a glutamine codon at position 472 to a stop codon. The nonsense mutation, g.23238C>T (Q472X), that gave rise to the truncated polypeptide is not present in non-human primates such as the chimpanzee (*Pan troglodytes*) and the gorilla (*Gorilla gorilla*) [32] and must therefore have arisen in the human lineage some time after the divergence of the *Homo* and *Pan* clades took place some 6 million years ago [33].

Analysis of individuals of different ethnic backgrounds, namely European Caucasians, Orientals (Japanese and Chinese), Africans (including African Americans and UK Afro-Caribbeans), New-Guinea Aboriginals, Indians and Maoris, revealed that the allele encoding the truncated FMO2 occurred at a frequency of 100% in all groups investigated, with the exception of individuals

of recent African descent, in which the ancestral allele, which encodes a full-length FMO2, was present at a frequency of 4% [32]. Analysis of products of heterologously expressed cDNAs revealed that the truncated protein is inactive, whereas the full-length protein is catalytically active [32]

FMO4

FMO4 is 558 amino-acid residues long. Other members of the FMO family, with the exception of the truncated form of FMO2, contain between 532 and 535 residues [9]. Sequence comparisons reveal that the additional residues in FMO4 are contained in a single block located at the C-terminus of the polypeptide [34]. It has been suggested that the additional residues may have arisen as a result of a single point mutation in the termination codon of the ancestral *FMO4* gene. A similar extension is present in FMO4s of other mammals and thus the predicted mutation would have occurred before the radiation of mammals. FMO4 is expressed constitutively at low amounts in the liver, lung and kidney of the adult human [11, 34-37].

FMO5

Despite the gene organization of *FMO5* being similar to the other *FMOs*, it is distinctive as it is not part of the *FMO* gene cluster on human chromosome 1q24.3, but is located ~26Mb closer to the centromere, at 1q21.1 [11], encoding a polypeptide 533 amino acids in length..

FMO5 is expressed in many foetal and adult tissues, but in human its main site of expression is adult liver [11, 16, 36, 37], although it is not present in the same amounts as *FMO3*. *FMO5* expression is the highest of the *FMOs* in the adult skin [16] and in the small intestine [36, 37].

FMO5 is regarded as an atypical FMO with regard to substrate specificity. FMO5 exhibits little catalytic activity towards compounds, such as methimazole, that are good substrates for other FMOs [38-41]. Despite its expression in the adult human liver, FMO5 is not thought to play a significant role in the metabolism of drugs, with the possible exception of esonarimod [42].

1.2. *FMO3*

A cDNA for human FMO3 was identified in 1992 by Lomri et al. [43] and the genomic location of human FMO3 was subsequently mapped by Shephard et al. the following year [10]. FMO3 encodes a polypeptide of 532 amino-acid residues with a molecular mass 60,047.

FMO3 is the major form of FMO present in the adult human liver [9, 14] with expression being switched on at birth [24]. FMO3 mRNA has also been detected in the lung, kidney, adrenal medulla and cortex, pancreas, thyroid, gut and brain [11]. It is thought that *FMO3* expression is switched on after birth in humans [24], however small amounts of FMO3 protein have been detected in the human embryonic, but not foetal, liver. The mRNA for FMO3 is also not detectable in foetal liver [14]. The mechanisms and factors required to express the *FMO3* gene in the embryo, to silence its expression during foetal development and then to re-activate expression of the gene after birth are unknown. Immunochemical analysis of 240 individuals between the ages of 8 weeks gestation and 18 years of age by Koukouritaki et al., (2002) [24] detected small amounts of FMO3 at eight weeks gestation, but not in the foetus (15-40 weeks gestation). FMO3 was undetectable, or present only in small amounts in postnatal samples during the first 3 weeks of life, in most individuals sampled. Subsequently, three distinct phases of *FMO3* expression were defined. Between 3 weeks and 10 months of

age, amounts of FMO3 increased 5-fold, to 4.7 ± 5.9 pmol/mg microsomal protein. Between 10 months and 11 years of age another 3-fold increase in expression was observed (to 12.7 ± 8.0 pmol/mg microsomal protein). Between 11 and 18 years of age, there was a further 2-fold increase (to 26.9 ± 8.6 pmol/mg microsomal protein). The amount of FMO3 was not affected by gender.

The age at which *FMO3* expression is switched on varies from birth to up to 2 years old. However, by 10 months of age the majority of individuals are expressing *FMO3*. This contrasts with the extinction of *FMO1* expression in human liver, which is tightly linked to the process of birth, being switched off within a few days of birth, irrespective of gestational age [24]. This is important in terms of the metabolism of xenobiotics that are substrates of FMO3 as, during the first year of life, most individuals will have no, or very small amounts of FMO, in their liver. The amounts of FMO3 present in adult liver are comparable to those of the most abundant hepatic CYPs, CYP3A4, CYP2C and CYP1A2, with reported specific contents of 96 ± 51, 60 ± 27 and 42 ± 23 pmol/mg microsomal protein, respectively [28].

Inter-individual differences of up to 20-fold in the amount of FMO3 have been reported [24, 44]. Such variation would be expected to have consequences for the ability of individuals to metabolize therapeutic drugs and potentially harmful foreign chemicals that are substrates for the enzyme. FMOs are thought not to be inducible by exogenous chemicals (there have been only two reported cases of induction in the case of *FMO1* and the work is yet to be verified [45, 46]) and, therefore, inter-individual variation in their expression is more likely due to genetic, rather than to environmental, factors. This is supported by the discovery, in the 5'-flanking region of *FMO3*, of single-nucleotide

polymorphisms (SNPs) that affect the transcription of the gene, in some cases effectively silencing it and in others, increasing it as much as 8-fold [24]. However, expression of *FMO3* may be influenced by physiological factors, e.g. FMO3 declines during menstruation [47-49].

FMO3 and Trimethylaminuria

FMO3 is the most abundant FMO in the human liver reaching levels as high as 0.5% of the total microsomal protein in this organ. *FMO3* has previously been implicated in termination of pharmacological activity of primary amines and is therefore important in drug clearance. However, it is the involvement of FMO3 in the *N*-oxidation of the dietary-derived xenobiotic trimethylamine (TMA) that has provided the impetus for the study of this enzyme in relation to the metabolic disorder, Trimethylaminuria (TMAU). TMA is produced by the gut bacteria when they metabolise compounds such as choline and lecithin [50]. TMA is rapidly absorbed. Normally, FMO3 catalyses the *N*-oxidation of dietary derived TMA [51], which smells of rotting fish, into the non-odorous TMA-*N*-oxide, which is readily excreted from the body in the urine. Mutations across the entire length of *FMO3* have been shown to result in the failure to catalyse the *N*-oxidation of TMA into trimethylamine *N*-oxide (Table 2 and Figure 3). TMA is excreted in its raw, odorous form in the bodily fluids of individuals carrying any one of these mutations in a homozygous state (Figure 2). Affected individuals have >40% unmetabolised TMA in their urine compared to 0-9% in unaffected individuals [52]. The first described clinical report of TMAU was in 1970 [53]. TMAU (or fish-odour syndrome) sufferers excrete an offensive smell of rotting fish within their bodily fluids. This condition can not be considered a minor affliction as patients often have trouble integrating into society, many are heavy

Table 2. Mutations in *FMO3* leading to TMAU phenotype

Amino acid change	Nucleic acid change	exon	Reference
	g.-2092-10145del	1-2	[54]
E32K	g.94G>A	2	[55]
I37T	g.110T>C	2	[56]
R51G	g.11145A>G	3	[57]
A52T	g.11148G>T	3	[58]
V58I	g.11166G>A	3	[59]
N61S	g.11177A>G	3	[35, 60]
K64KfsX2**	g. 11185delA	3	[55]
M66I	g.11192G>T	3	[54, 61, 62]
M82T	g.11239T>C	3	[63]
N114S	g.15036A>G	4	[47, 64]
V143E	g.15123T>A	4	[65]
G148X	g.15137G>T	4	[66]
P153L	g.15153C>T	4	[61, 67]
C197fsX	g.15526_15527delTG	5	[68]
D198E	g.15531T>A	5	[69]
I199T	g.15533T>C	5	[70]
T201K*	g.15539C>A	5	[71]
R223Q	g.18177G>A	6	[72]
R238P	g.18225G>C	6	[56]
R238Q	g.18225G>A	6	Unpublished
M260V*	g.18290A>G	6	[71]
E305X	g.21429G>T	7	[58, 73]
E314X	g.21460G>T	7	[62]
R387L	g.21680G>T	7	[58]
W388X	g.21684G>A	7	[64]
K394KfsX11	g.21702delG	7	[56]
M405IfsX $^{+}$	g.23580delG^{+}	8	[56]
M434I	g.24486G>A	9	[35]
Q470X	g.24592C>T	9	[64]
G475D	g.24608G>A	9	[70]
R492W	g.24658C>T	9	[35, 62]
R492Q	g.24658G>A	9	Unpublished
R500X	g.24682C>T	9	[71]

** K64KfsX2 has also been referred to as M66X

*Causative consequences of variant may be considered ambiguous because of variants found in *Trans*.

$^{+}$Although this deletion occurred in exon 7, it caused a frameshift resulting in a truncation at amino acid 405 in exon 8.

Mutation nomenclature follows that recommended by the Human Genome Organization (http://www.hgvs.org/mutnomen/)

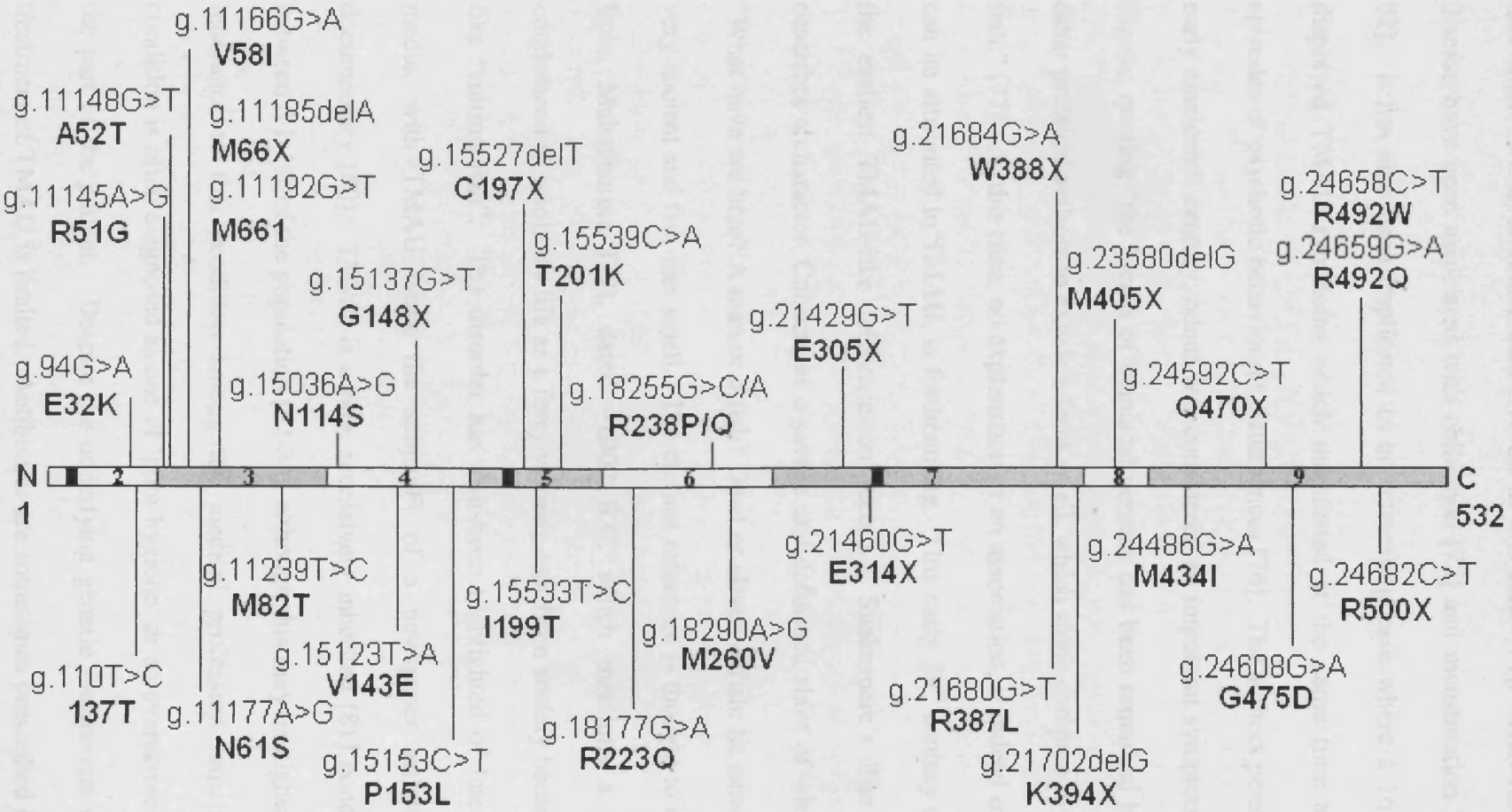


Figure 3.The mutations of human *FMO3*

The genetic and corresponding amino acid substitutions are shown for the mutations causing TMAU within human *FMO3*. Each exon is numbered and the FAD binding moiety within exon 2, the NADPH binding moiety within exon 7 and the FMO characteristic 'FATGY' motif within exon 5 are highlighted as solid, black bars.

smokers in an attempt to mask the smell and in some cases they have been known to take their own lives rather than live on in this state [74]. Transient forms of this disease have been associated with childhood [75] and menstruation in females [48, 52]. It has also been implicated in an interesting case where a 16 year old male displayed TMAU symptoms which manifested at the same time as seizures and episodes of psychotic behavioural disturbances [76]. The authors point out that in the early nineteenth century, odour was considered an important symptom of psychiatric illness, quoting "the breath of maniacal persons has been remarked by Esquirol and other practical authors to exhale a fetid smell, which some compare to that of stinking fish." [77]. At this time, no explanation of an association, if indeed one exists which can be attributed to TMAU, is forthcoming. This early 19th century comment is not the earliest TMAU-like reference on record. Shakespere's The Tempest [78] describes a character, Caliban, as a savage and deformed slave of which was written "What have we here? A man or a fish? Dead or alive? A fish: he smells like a fish; a very ancient and fish-like smell." The earliest reference is thought to be in the Indian Epic, Mahabharata [79], dated ~1000 B.C. which mentions a young woman condemned to a solitary life as a ferry-woman, cast from society because she smelled like "rotting fish". The disorder has also been highlighted of late in the general media, with TMAU being the subjectFt of a newspaper article and a BBC documentary [80]. TMAU is a rare, recessively inherited [81] condition, affecting between 0.1-1% of the population [82-84], although this may be higher as the level of ignorance to this condition among the medical profession remains high. The condition is often diagnosed as one of 'poor hygiene' or an overactive imagination on the part of the patient. Despite the underlying genetic mechanism of the disorder, treatment of TMAU is limited. Antibiotics are sometimes prescribed in an attempt to kill gut flora that produce TMA from dietary precursors such as choline and

lecithin.The dietary supplements copper chlorophyllin and charcoal have been reported to decrease the amount of TMA in the urine [85]. Strict diets that limit the intake of foodstuffs rich in precursors of TMA, such as meat, eggs, soya and fish, have been reported to help control the condition.

Catalytic consequences of the *FMO3* polymorphisms

Although TMAU is a rare condition, polymorphisms in *FMO3* certainly have implications for the wider public. Many xenobiotics are substrates of FMO3 (Table 3) including pesticides, dietary-derived compounds and a diverse range of therapeutic drugs such as atypical antipsychotics, anti-thyroids and anti-fungals. Polymorphisms exist across the entire length of *FMO3*, which are known to affect the catalytic properties of the encoded enzyme (Table 4 and Figure 4), but do not cause TMAU. This could have implications to individuals possessing these polymorphisms. For instance, an individual with a polymorphism in FMO3 that causes a reduction in oxidation activity (e.g. D132H, Table 4 and Figure 4) may be considered a poor metaboliser of FMO3 substrates (Table 3) and therefore may have a heightened risk of overdose. This observation is all the more serious when considering the fact that adverse drug responses are suggested as the fourth leading cause of death in the U.S.A. [86]. Conversely, a polymorphism in FMO3 leading to increased activity of the enzyme (e.g. L360P, Table 4 and Figure 4) may lead to lower efficacy of drugs metabolised by the enzyme and ineffective dose prescription due to a more rapid drug clearance. As displayed in Table 4, work has begun to establish the catalytic consequences of polymorphisms to the FMO3 enzyme with regards to *–N* and *–S* oxidation. These polymorphisms have been generated on the E158 background only. Given that the E158K background is prevalent at a frequency of nearly 50% [87] the work in this thesis considers the catalytic consequences of the polymorphisms found in human populations on both the E158 and K158 backgrounds in order to get an

Table 3. FMO3 substrates.

Substrate	Pharmacological Agent	ref
N-(3R)-1-azabicyclo[2.2.2]oct-3-ylfuro[2,3-c]pyridine-5-carboxamide(1)	Receptor agonist	[88]
1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine	Dopaminergic neurotoxin	[89]
5,6-dimethylxanthenone-4-acetic acid	Anti-neoplastic agent	[90]
Almotriptan	Serotonin agonist	[91]
Amphetamine	CNS stimulant	[92]
Arecoline	Acetylcholine receptor agonist	[93]
Benzydamine	Anti-inflammatory	[94]
Bupivacaine	Anesthetic	[95]
Caffeine	CNS stimulant	[95]
Cevimeline	Muscle relaxant	[96]
Chlorpromazine	Anti-psychotic	[97]
Clozapine	Anti-psychotic	[98]
Dapsone	Anti-inflammatory	[99]
Deprenyl	Type B Monoamine oxidase inhibitor	[92]
K11777	Anti-protozoal	[100]
Ketoconazole	Anti-fungal agent	[101]
Fluphenazine	Anti-psychotic	[102]
Itopride	Dopamine antagonist	[103]
Lidocaine	Anesthetic	[95]
Methamphetamine	CNS stimulant	[92]
MK-0457	Aurora kinase inhibitor	[104]
Olanzapine	Anti-psychotic	[105]
Olopatadine	Anti-inflammatory	[106]
Perazine	Anti-psychotic	[107]
Phenethylamine	Psychotropic agent	[108]
Propranolol	Adrenergic antagonist	[95]
Pyrazoloacridine	Anti-neoplastic agent	[109]
Ranitidine	Gastrointestinal H2 receptor antagonist	[110]
S 16020	Topoisomerase II antagonist	[111]
(*S*)-3-methyl-5-(1-methyl-2-pyrrolidinyl)isoxazole	Anti-anxiety agent	[112]
(*S*)-Nicotine	Neuronal stimulant	[113]
Sulfamethoxazole	Antibacterial	[99]
Tamoxifen	Oestrogen antagonist	[114]
TG100435	Src Kinase inhibitor	[115]
Thioproperazine	Anti-psychotic	[102]
Trifluoperazine	Anti-psychotic	[102]
Trimethylamine	Dietary derived amine	[67]
Tyramine	Dietary derived amine	[116]
Verapamil	Anti-arrhythmia agent	[117]
Xanomeline	Muscarinic agonist	[118]
Zimeldine	Serotonin receptor antagonist	[119]

Table 3.FMO3 substrates

4-chlorophenyl methyl sulphide	Environmental sulphide	[120]
7 alpha-thiomethylspironolactone	Anti-mineralocorticoid	[121]
Aldicarb	Pesticide	[122]
Albendazole	Anti-protozoal	[123]
Cimetidine	Histamine H2 Antagonist	[113]
Cronetron	Pesticide	[122]
Diphenyl sulphide	Environmental sulphide	[120]
Disulfoton	Pesticide	[124]
Ethionamide	Anti-tubercular Agents	[125]
Ethyl methyl sulphide	Environmental sulphide	[120]
Farnesylcysteine	Posttranslationally modified amino acid	[126]
Fenthion	Pesticide	[127]
Fonofos	Pesticide	[128]
Methamphetamine	CNS stimulant	[92, 129]
Methimazole	Anti-thyroid Agents	[130]
Methionine	Essential amino acid	[131]
MK-0767 methyl sulphide	Peroxisome proliferator receptor agonist.	[132]
Phorate	Pesticide	[124]
S-methyl-esonarimod	Cytokine production inhibitor	[42]
Sulindac sulphide	Anti-neoplastic agent	[5]
Tazarotenic acid	Tazarotene (a retinoid) metabolite	[133]
Thiacetazone	Anti-tubercular Agents	[134]
Thiobenzamide	Anti-tubercular Agents	[135]
Triclabendazole	Anti-protozoal	[136]
Diethylphenylphosphine	Pesticide	[3]
Folex	Defoliant	[137]
2-(methylseleno)-cinnamonitrile	Pharmaceutical intermediate	[138]
Benzyl selenide	Selenium salt metabolite	[138]
Dimethylselenide	Selenium salt metabolite	[139]
Ebselen	Anti-inflammatory	[6]
Methyl phenyl selenide (Selenoanisole)	Heame biosynthsis antagonist	[6]
n-[2-(methylseleno)-ethyl] benzamide	Anti-oxidant (Ebselen) metabolite	[138]
(Phenylselenomethyl)-trimethylsilane	Fuel additive	[138]
Seleno-l-methionine	Methionine analogue	[140]

Xenobiotic substrates of FMO3 are listed and highlighted according to the soft nucleophile attacked by the hydroperoxyflavin form of FMO3 (blue – nitrogen, red – sulphur, green – phosphorus and black – selenium).

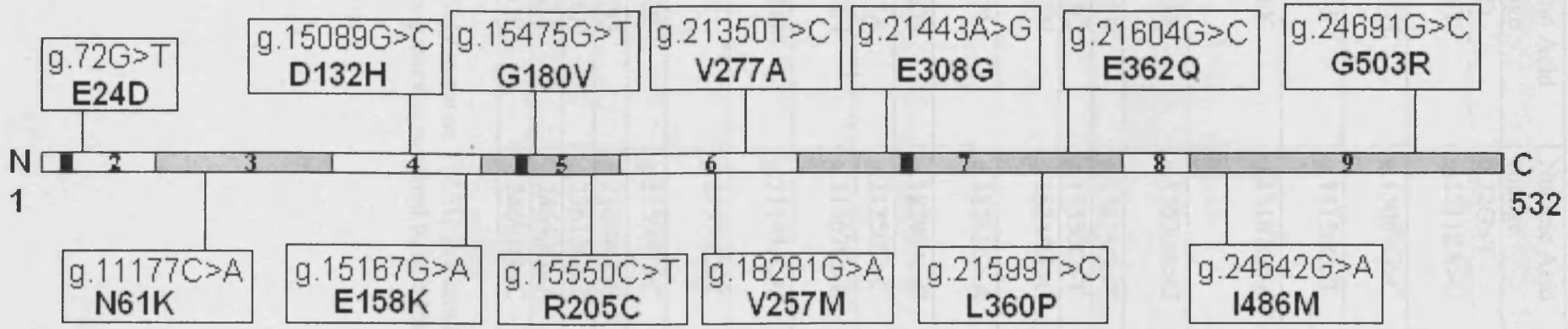


Figure 4.The polymorphisms of human *FMO3*

The genetic and corresponding amino acid substitutions are shown for the nonsynonymous polymorphisms (boxed) of human *FMO3*. Each exon is numbered and the FAD binding moiety within exon 2, the NADPH binding moiety within exon 7 and the FMO characteristic 'FATGY' motif within exon 5 are highlighted as solid, black bars.

Table 4. Polymorphic variants of *FMO3*

Amino Acid change	Nucleic Acid change	exon	Functional consequence	Reference
E24D	g.72G>T	2	Modest reduction	[60]
T108T	g.15112A>C	4	Synonymous. No effect	[141]
D132H	g.15089G>C	4	Substrate dependent decrease	[142, 143]
S147S	g.15136C>T	4	Synonymous. No effect	[141, 142, 144]
E158K	g.15167G>A	4	Limited, substrate-dependent decrease	[62, 67, 70, 73, 142, 145]
K167K	g.15530A>G	5	Synonymous. No effect	[141]
G180V	g.15475G>T	5	No effect	[35]
R205C	g.15550C>T	5	Moderate decrease	[69]
N245N	g.18248G>C	6	Synonymous. No effect	[73]
V257M	g.18281G>A	6	Moderate substrate dependent decrease	[35, 73, 142]
M260V	g.18290A>G	6	n.d.	[71]
V277A	g.21350T>C	7	n.d.	[146]
N285N	g.21467C>T	7	Synonymous. No effect	[141, 142]
E308G	g.21443A>G	7	Moderate substrate dependent decrease	[58, 70, 73, 147]
I350I	g.21570C>T	7	Synonymous. No effect	[142]
L360P	g.21599T>C	7	Increased activity	[143]
E362Q	g.21604G>C	7	n.d.	[142, 146]
K416N	g.23613G>T	8	Modest reduction	[60]
I486M	g.24642G>A	9	n.d.	[146]
G503R	g.24691G>C	9	n.d.	[142]

* Likely to cause TMAU in a homozygous individual (included in this table because its frequency is reported to be >1% [60])

accurate picture of the catalytic effect of polymorphisms in the population. This approach is supported by the finding that a homozygote individual of both the E158K and E308G polymorphisms display a mild TMAU phenotype, whereas the homozygote E158/E308 shows no such phenotype [70].

FMO3 Tertiary structure

The characterisation of any protein sequence is key to understanding the contribution of structure to function. The determination of the tertiary structure of a protein, whether it is via X-ray crystallography or NMR, can be used to learn about the mechanistic processes that underlie a protein's function. Generally, when two proteins are similar at the primary sequence level, they tend to share a similar tertiary structure and thus similar properties and functions [148]. As similarity decreases, so too does the likelihood of inferring similarity in protein function from primary sequence. Currently there are no mammalian FMO structures deposited in the Protein Data Bank (PDB) [149]. However, the possibility exists that a theoretical model could be built of FMO3 using the three-dimensional atomic coordinates of solved protein structures as templates, providing that the templates are of a high enough sequence identity to provide a meaningful guideline. Previously, two models of FMO have been reported [150, 151] based on the crystal structures of glutathione reductase (PDB 1GET, at 1.86 Å resolution [152]) and NADPH peroxidase (PDB code 1NPX, at 2.16 Å resolution [153]). However, these models were generated by merely threading the FMO structure onto the αC-atoms of the very low sequence identity templates (18% and 16% primary sequence identity to FMO3, respectively and 20% identical to each other). Thus their use to infer biophysical significance requires extreme caution.

1.3. Homology Modelling

The idea of generating models of proteins is not new. It was demonstrated in the 1960's that when ribonuclease is denatured, it can refold into its native form and its function is not impaired [154, 155]. This study gave rise to the idea that a protein's tertiary structure may be able to be inferred from the amino acid sequence [155]. Currently, *ab initio* methods of modelling protein structure from primary sequence are extremely unreliable, especially for larger proteins. The most reliable method of protein structure prediction is homology modelling [156-158]. Also known as comparative modelling, the basic principle of homology modelling is the generation of a model of a protein (the target) from a protein of known structure (the template) with a significant degree of primary sequence identity. The scientific defence of this approach is based on the observation that protein tertiary structure changes through evolution far more slowly, relative to the primary sequence of the molecule [159], and indeed the DNA sequence from which it is derived when considering synonymous substitutions.

Homology modelling involves five steps.

1. Identifying a known template structure to the target to be modelled.
2. Aligning the template sequence with the target sequence.
3. Building the model.

The backbone of the model is generated using the coordinates of the template file as a guide. Loops between more highly conserved secondary structural elements are added (it is often the case that gaps in alignment are represented as loops) or are altered in length.

4. Energy Minimisation
5. Assessing the quality of the model.

It is vital that the model quality is assessed. Automated modelling, although convenient, is not advanced enough to generate meaningful results without user input. Pre-calculated models from known structures are available from resources such as ModBase [160], which uses a PSI-BLAST and Modeller [161] (discussed in more detail below) pipeline to generate the results, but they are not validated or refined, nor does the template selected to model from undergo any analysis. Indeed, the ModBase web site itself warns that the database "...contains theoretically calculated models, which may contain significant errors, not experimentally determined structures". There is also a web server version of Modeller, ModWeb [162], that automatically models your target sequence for you but does not assure the quality of alignment, selection of the correct initial template or any sort of assessment of model quality.

To assess the quality of a model, several factors of biophysical attributes are considered. As mentioned above, steric clashes of amino acid side chains are considered by the modelling software, but they still do happen and must be addressed. The bond length and angles between two alpha carbon atoms (αC) on the polypeptide backbone are relatively fixed. The only areas of conformational freedom are around the amino acid side chains from the αC atom to the C atom of the carbon of the carboxyl group and from the αC atom to the N of the amino group known as Psi (Ψ) and Phi (Φ) angles respectively. Normally, the values of Φ are limited to a range of between -60 and -150 degrees and the values of Ψ range between -60 and 120 degrees. The predictability of these angle states led to Gopalasamudram Narayana Ramachandran in 1968 to propose an energy landscape when plotting these angles on a graph. Ramachandran plots [163] (Figure 5) enable the modeller to observe the possible conformations of Φ and Ψ angles for any given protein. The plot can be thought of as areas of allowed and disallowed angle pairs. It is therefore possible, using Ramachandran plots, to identify unreasonable

Figure 5. The Ramachandran Plot

A) Shows the torsion angle between the αC and the N of the amino group, denoted Φ and the torsion angle between the αC and the C atom of the carboxyl group, denoted Ψ. B) A Ramachandran plot made up of the energy landscape when pairs of Φ and Ψ angles are plotted. Three small red areas denote the regions in which Φ/Ψ are 'allowed'. This area is expanded, denoted by the yellow areas on the plot, when the restraints are relaxed by just 0.1 Å. The majority of the plot, represented as white in colour is in the 'disallowed' regions, where dihedral angle pairs are sterically improbable and where one would look for poorly modelled amino acid residues. (Modified from http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html)

torsion pairs of Φ/Ψ. The use of wet lab experimental data in refining homology models can not be understated. The catalytic cycle of FMOs is well defined and contains a regenerative cycle including electron donors and electron carriers (Figure 1). Models that can not explain experimentally observed biochemistry must be viewed critically.

6. Refining of the model.

Processes such as loop modelling and dihedral angle modifications can be made if sterically poor regions are identified. For instance, in the case of dihedral angles, graphical packages such as Swiss-PDBViewer [164], can be used to get a more reasonable representation of sidechains within the polypeptide.

In the past, homology models have been built of CYPs and these have proved useful in generating theories to explain the consequences of polymorphisms, active site dynamics, ligand interactions etc., but it is notable that these have usually been built using templates of reasonably high primary sequence identity [165-168].

It is possible however, to produce homology models of proteins despite only having templates of relatively low primary sequence available. Often, the terms 'midnight' and 'twilight' zones are used to describe the levels of 25% and 30% sequence identity. These thresholds represent the boundaries beyond which structural homology is so low that little useful information could be extrapolated from homology modelling [169], thus any model built using templates below this threshold would be deemed so unreliable that the endeavour would be of questionable value.

Modeller

Modeller [170] has been shown to be among the best modelling software available [157]. Essentially it is a set of Python scripts which perform all of the steps needed to produce and

refine a homology model. Modeller automatically derives restraints from the template structure and the alignment with the target sequence. Modeller uses these restraints to carry out backbone, side chain and loop generation steps. These restraints include atomic distances, angles, dihedral angles, pairs of dihedral angles, rules of secondary structure packing and some other spatial features defined by atoms. Modeller also has a set of restraints that were evaluated empirically from protein X-ray crystal and NMR structures. Energy minimization using simulated annealing refines the special arrangements within the model by simulating a melting of the molecule at a high temperature, allowing the molecule to overcome any local minima and then cooling it slowly until an optimum global minimum is achieved.

A feature of Modeller which is particularly useful in modelling enzymes is the capacity to model heteroatoms such as enzyme cofactors in complex with a target. This can be done by either defining the ligand in the Modeller topology library, which is essential if said ligand is not present in the template structure or, more commonly, by assuming that the ligand interacts in a similar biophysical fashion with the target and the template Modeller can extrapolate all distance restraints from the interaction of the ligand to the template structure and then satisfying these restraints when building the model along with the ligand in the target.

1.4. Evolution of FMO3

The classic genetic dogma that for each gene there are only two alleles, a "wild-type" and a mutant, with the latter title referring to a deleterious state, has long since gone. We now appreciate, if not fully understand, multiple allelic states created by single nucleotide polymorphisms, not only in coding regions but also in upstream genetic elements, promoters and introns. The information presented as the pattern of sequence variation over a single chromosome in a diploid organism, termed the haplotype, is vital to understand the relationship of all the heterozygous positions in a gene within an individual and to understand the product of a haplotype.

The recent explosion of genomic information at both the inter- and intra- species level, both in terms of the extent of coverage and the amount of Single Nucleotide Polymorphism (SNP) data, specifically dbSNP [171] and the HAPMAP initiative [172], has also facilitated this possibility.

Features of datasets can be analysed to gain insight into the evolution of a locus such as SNP frequencies, haplotypes, synonymous and non-synonymous changes, recombinational events, association among variants such as linkage disequilibrium etc. In addition, allele frequency distribution can give indications as to whether natural selection may have had specific influence on the evolution of a locus.

Coalescent Theory

Coalescent theory refers to the idea that all alleles are ultimately inherited from a single ancestor. This individual can be considered as a single most recent common ancestor. When considering a specific allele, this allele can undergo coalescence to return to the most recent common ancestor, often represented graphically as a phylogenetic tree. Coalescence

does not just refer to the arrival at the most common recent ancestor. At any point where two lineages converge it can be said that these lineages have coalesced.

It is possible to reconstruct the evolutionary history of a locus such as the *FMO3* locus using coalescent models provided that sequence data are available over a large enough genomic region. Also, using coalescent based methods, where haplotypes lineages are traced back using mathematical models to the most common recent ancestor [173], it is possible to reconstruct genetic networks which make up distinct haplotypes and also answer questions about the SNPs themselves which make up the haplotypes, such as estimations of age.

Statistical tests using simulations of coalescence

Coalescent simulations can be used to compare values obtained from genetic testing of real data, such as genomic sequencing, to values generated from random samples of the same sample size and level of polymorphism as the observed data. Significant differences between the two values can then be interpreted. Some genetic tests which can be used with coalescent simulations are outlined below.

Watterson's estimator of θ (θ_W) [174] and Nei & Lei's nucleotide diversity (π) [175] are two representations of polymorphism diversity within a sample. They differ in that θ_W uses the number of segregating sites (the number of sites in a stretch of nucleotides which display a polymorphism, e.g. if 2500bp is sequenced in 1000 individuals, 100 individuals possess a polymorphism at 456bp, 449 individuals possess a polymorphism at 1324bp and 20 individuals possess a polymorphism at 1878bp, the number of segregating sites would be 3) as the basis of diversity representation whereas π uses the average number of pairwise differences (in the previous example of 1000 individuals, each individual would be compared to the next and their number of pairwise differences recorded, with an average

of the whole population being taken as representative of the sequence diversity within a population) in a sample as the basis of nucleotide diversity representation.

Tajima's *D* [176] is a test of nucleotide diversity, comparing θ_W and π in the context of the population size. It was proved that, assuming a constant-size neutral population, Tajima's *D* would be near zero as θ_W and π would be about the same. If the locus was undergoing a selective sweep, then a negative selection could be expected as genetic diversity is reduced as fixation of mutations is established. When balancing selection is acting, Tajima's *D* can be expected to be positive owing to the excess of highly polymorphic sites and a deficit of low-frequency polymorphisms, compared to neutrally evolving sites.

Fu and Li's *F** and *D** [177] uses polymorphic data to test the null hypothesis that all mutations are selectively neutral [178]. Fu and Li's *F** is based on the differences between the number of singletons (mutations appearing once) and π whereas Fu and Li's *D** is based on the differences between the number of singletons and the total number of mutations in the population sample.

Fu's F_S statistic can be used to compare the observed number of sequence haplotypes to the number expected under the assumption of an infinite-sites model of neutral mutation with no recombination. The null hypothesis of neutral evolution would generate a value near zero with a paucity of haplotypes leading to a negative value and an intermediate to excess number indicating balancing or directional selection respectively, both discussed in this section.

Harpending's raggedness statistic (r) [179] and Rogers' mismatch distribution test [179] can identify the mode of distribution of pairwise differences with the former describing the variation about the curve generated from the latter. A smooth, unimodal distribution is indicative of population growth or directional selection, both discussed in

this section whereas a ragged distribution is indicative of balancing selection or a population size which has remained constant.

Linkage Disequilibrium (LD), the tendency for alleles to be present in the same haplotype (positive LD) or not (negative LD), can be measured using a number of statistical tests. Wall's B and Wall's Q [180], based on segregating sites, examine each of the possible pairs of adjacent segregating sites. When adjacent sites are found to be the same among individuals within a population, they will be in complete disequilibrium and are referred to as congruent pairs. Wall's tests examine the proportion of segregating site pairs that are congruent. Z_{nS} [181] is another way to measure LD, but this statistic is based on an average of pairwise associations between polymorphisms across the locus within the whole population.

Non-coalescence based statistical tests

A non-coalescent based test of neutrality was proposed by John McDonald and Martin Kreitman in 1991 [182]. The McDonald and Kreitman's test compares synonymous and nonsynonymous variation within and between species because, under neutrality the ratio of replacement to synonymous fixed differences between species should be the same as the ratio of replacement to synonymous differences within the species. When dealing with human samples, it is usual to compare the human locus to that of a close relative such as *Pan Troglodytes* (chimpanzee). This particular comparison has become much more accessible since the completion of the chimpanzee genome sequencing project [183].

To test if genetic differentiation at a locus is due to population-subdivision, Wright's fixation index (F_{ST}) [184] can be used. This test considers the difference between the mean heterozygosity among subdivisions in a population and to the population as a whole, e.g. a Japanese cohort compared to an Irish cohort and a Japanese cohort compared

to the human population as a whole. The fixation index ranges from 0 (indicating no difference at all between subpopulations and the population as a whole) and 1 (meaning complete differentiation).

Natural selection

As mentioned above, variability in SNP distribution can be caused by natural selection. Natural selection can force evolution in different directions, creating different genetic signatures in the form of allele frequencies. In light of this, natural selection can be subdivided into three categories.

Directional selection

Directional selection occurs when a certain allele confers a greater level of fitness than others, resulting in an increase in frequency of the allele. When this selection type is followed through to completion the allele is said to be completely fixed, meaning that the entire population expresses the fitter phenotype. Evidence of this has been presented, for example, in the case of triosephosphate isomerase evolution in bony fish [185] and the factor IX gene locus in humans [186].

Stabilising selection

More commonly a selection event occurs which lowers the frequency of alleles that have a deleterious effect on the phenotype. Selection will continue in this way until the allele is eliminated from the population. Evidence of this has been presented, for example, in the case of immune responsiveness in *Parus caeruleus* (Blue Tit) [187] and transcription factors in humans [188].

Balancing selection

This type of selection maintains alleles at intermediate frequencies in a population. This manifests itself in a diploid species when a heterozygote has a higher fitness than homozygous individuals when considering a relevant loci. Evidence of this has been presented, for example, the major histocompatibility complex (MHC; *HLA* in humans) [189, 190], *Glucose-6-phosphate dehydrogenase* (*G6PD*) in humans [191], the 5' *cis*-regulatory region of *human chemokine* (C-C motif) receptor 5 *CCR5* [192], *PTC* [193], the bitter taste gene in humans, and the *retroviral restriction factor, TRIM5alpha* in Old World monkeys [194].

Given that FMO3 is such an important molecule in terms of the interaction between humans and their chemical environment, it could be speculated that *FMO3* would be a likely target for natural selection. One might envisage a situation where the maintainance of a seemingly deleterious allele in a xenobiotic metabolising enzyme in terms of catalytic activity might be maintained at an intermediate frequency within a certain population, due to a selective advantage this conferred within the population's chemical environment.

1.2. Aims

The aims of this investigation were as follows.

- To investigate the genetic contribution of *FMO3* variation to apparent TMAU sufferers by sequence analysis of the *FMO3* locus in selected individuals and their families where possible.
- To assay the catalytic consequences of FMO3 variation, particularly focussing on combinations of variation involving the E158 site and another site, by comparing kinetic performance of variants using an established enzyme assay for FMO3.
- To attempt to understand the consequences of amino acid variation within FMO3, structurally and mechanistically, by building a homology model of FMO3.
- To attempt to understand how *FMO3* locus variation has evolved to the genotypes evident today, through genetic analysis of a human population cohort.

Chapter 2

Materials and Methods

2.1. *FMO3* TMAU patient screen

2.1.1. DNA extraction from buccal swab sample

The buccal swab kit used to collect a buccal swab sample from a patient was prepared by Dr. Abi Jones (Department of Biology, University College London) and consisted of a sterile, Ultra Violet (UV)-irradiated cotton swab and a solution containing 1 mL of 0.5% (w/v) sodium dodecyl sulphate (SDS) and 0.05M ethylenediaminetetraacetic acid (EDTA) (VWR International, UK). The swab was rubbed against the inside of the patient's cheek and submerged in the SDS/EDTA solution (above) by the GP. This was stored at room temperature and posted to University College London for DNA extraction and analysis.

The buccal swab was added to 800 µL of sterile distilled water containing 8 µg of Proteinase K (Roche, UK). The sample was then incubated at 56°C overnight. 0.6 mL of a phenol (Fisher Scientific, UK) and chloroform (May & Baker Ltd, Dagenham, UK) mixture was aliquoted and stored at -20°C. 0.8 mL of the SDS/EDTA/ solution was transferred from the mouth-swab collection tube into the tube containing the ice-cold phenol/chloroform mixture. The tube was inverted to mix the contents, and then centrifuged at 17,900 x g for 10 min. The aqueous phase was transferred to a fresh microcentrifuge tube containing 0.6 mL chloroform and 30 µL of 5M NaCl (VWR International, UK). The tube was inverted, to mix the contents, and then centrifuged at 17,900 x g for 10 min. The aqueous phase was transferred to a fresh microcentrifuge tube containing 0.7 mL chloroform. The tube was inverted to mix the contents, and then centrifuged at 17,900 x g for 10 min. The aqueous phase was transferred to a fresh microcentrifuge tube containing 0.7ml isopropanol (VWR International, UK), inverted to mix the contents and then incubated at -20°C for 2 hrs. The sample was then centrifuged at 17,900 x g for 15 min. The supernatant was decanted and the tube was placed upside down

at an angle of 45° on tissue paper for approximately 20 min to dry the DNA pellet. 0.8 mL of 70% ethanol was added to the DNA sample, which was then centrifuged at 17,900 x g for 10 min. The DNA pellet was then air-dried, as described above. After 20 min, 400 µL of distilled water was added to the pellet and the sample was incubated at 56°C for 10 min to resuspend the DNA. The sample was then stored at -20°C.

2.1.2. DNA quantification

DNA was quantified using the NanoDrop system (NanoDrop Technologies, USA). The NanoDrop system was calibrated with the elution buffer used to resuspend the DNA and the DNA concentration was calculated automatically utilising pre-programmed extinction coefficients.

DNA absorbs UV light at a wavelength of 260 nm, whereas protein absorbs UV light at 280 nm and aromatic compounds such as phenol and other aromatic compounds absorb UV light at 230 nm. A260/A280 and A260/A230 ratios were compared to assess the quality of the DNA. A level above 1.8 was taken to represent a good quality DNA sample free of contaminants.

2.1.3. Polymerase Chain Reaction (PCR) amplification of *FMO3* exons and upstream loci from human genomic DNA

Upstream elements I (-2310 to -1716, based on assigning the 'A' of the ATG translation transcriptional start codon +1) and II (-2733 to -2162), the non-coding exon 1 (-1906 to -1436) and exons 2-9 of *FMO3* were each amplified by PCR using primers detailed in Table 5 and synthesized by Eurogentec (Belgium). The PCR reaction mixtures contained, in a final volume of 50 µl, 50 ng of patient DNA , 0.5 µL of the appropriate sense (100 mM) and antisense (100 mM) primers, 1 µL of dNTP (10 mM), 5µl of 10X NH_4 Reaction Buffer

Table 5. Primers used to amplify each exon of *FMO3* by PCR

FMO3 sequence region	Primer	Primer Sequence
Upstream element I	Sense	5'-TTCCTGGTACTAATAGATCA-3'
Upstream element I	Antisense	5'-TCTGTGTGTCTACGTCCT-3'
Upstream element II	Sense	5'-ATCCTCTAATCCTTGTTAAA-3'
Upstream element II	Antisense	5'-CCAATAAGGAGGATGACT-3'
Exon 1	Sense	5'-TGGGAGACTGGCCTACAG-3'
Exon 1	Antisense	5'-GAGACGGAGTTTCGCTTTTA-3'
Exon 2	Sense	5'-GTGAGCTACCATACTCAGCCAGTG-3'
Exon 2	Antisense	5'-CACAGTGTGCTCTTATACACTTCCC-3'
Exon 3	Sense	5'-GACCTGATCAGTATACTCATTTACC-3'
Exon 3	Antisense	5'-AGTAGTAGACATACACTTCTTCAGC-3'
Exon 4	Sense	5'-CTTTTCTTTTTTCATACTGTATCTGC-3'
Exon 4	Antisense	5'-AAAAAGAAGACATTATCAAGATATTC-3'
Exon 5	Sense	5'-TATGCTTGGTGTGTTAAAATAGCAC-3'
Exon 5	Antisense	5'-CACACCTTTCAAACGATAATAACTC-3'
Exon 6	Sense	5'-CAGAATATCCACTACAAATGGTCAC-3'
Exon 6	Antisense	5'-GCTTACAGGACATTAAGGGTTGTTG-3'
Exon 7	Sense	5'-GCCTCCATCAATTTGTTCTTCAG-3'
Exon 7	Antisense	5'-CAAAGATCCAAAGTTATTGTCACTG-3'
Exon 8	Sense	5'-GGAAAATTACAGGCTGGTCCTATGC-3'
Exon 8	Antisense	5'-ATAGCTTGTAGTTGTCATTCCAATG-3'
Exon 9	Sense	5'-TTCTCTGTTCTGTTTCTACACAGAG-3'
Exon 9	Antisense	5'-CCCTGTCTGGGTATTGTCAGTAAC-3'

(Bioline Ltd, UK), 0.4 µL BIOTAQ Polymerase (Bioline Ltd, UK) and 1.25 µL (for upstream elements I and II and exons 1,2,3,5,6 and 8), 2 µL (for exon 4) or 2.5 µL (for exons 7 and 9) of $MgCl_2$ (50 mM) in thin-walled 0.5 mL Eppendorf tubes. The tubes were placed in a Techne Genius (Sussex, UK) thermal cycler that has a heated lid.

An initial denaturation step was performed at 95°C for five min. This was followed by 30 cycles of 95°C for 30 s; 58°C (for exons 2, 3, 5, 6 and 8), 55°C (for exon 1), 56°C (for exons 4 and 7) or 50°C (for upstream elements I and II and exon 9) for 30 s, and 72°C for 30 s. Finally, a single step of 72°C for 5 min completed the PCR amplification.

The sizes of the amplicons were confirmed by electrophoresis through a 1% agarose (Invitrogen, UK) SYBR Safe (Molecular Probes, UK) gel in 0.5X TBE.

PCR reactions were cleaned-up using SureClean (Bioline Ltd, UK). An equal volume of SureClean was added to the PCR reaction, the mixture was vortexed briefly, then incubated at room temperature for 10 min. The sample was then centrifuged for 10 min at 17,900 x g. The supernatant was removed and the pellet washed with 70% ethanol (VWR International, UK) and vortexed for 30 s. The sample was centrifuged for 10 min and the supernatant was removed. The pellet was then air-dried.

2.1.4. DNA Precipitation

All chemicals were supplied by BDH (UK) except where stated. 3M sodium acetate, pH 5.2 (0.1X vol) and 2X vol of ethanol (VWR International, UK) was added to the DNA solution. The sample was placed in a -70 °C freezer for 30 min. The mixture was then centrifuged at 17,900 x g using a bench-top centrifuge for 15 min at 4 °C. The supernatant was carefully removed and the DNA pellet was washed with 500 µL of 70% ethanol and centrifuged again at 17,900 x g using a bench-top centrifuge for 15 min at 4 °C. The supernatant was carefully removed and the DNA pellet was air dried for approx 5 min to

allow the residual ethanol to evaporate. The DNA pellet was resuspended in ddH_2O and stored at 4 °C.

2.1.5. DNA Sequencing

The DNA was cleaned by adding 50 µL of SureClean (Bioline, UK) to the reaction mixture, incubating for 10 min and then centrifuging at 17,900 x g for 15 min. The supernatant was removed and the pellet washed with 500 µL of 70% ethanol and the sample was then centrifuged at 17,900 x g. DNA was resuspended in ddH_2O. DNA sequencing was performed by MWG-Biotech (Germany) using the "Value Read" service or by ElimBio (CA, USA), with both requiring 1 µg of plasmid DNA or 20 ng of PCR generated DNA per 100 bases. Sequences were analysed by comparison to the wild-type hFMO3 cDNA. DNA was sequenced in both forward and reverse orientations.

2.1.6. *BsaWI* Restriction Digest of *FMO3* exon 9 PCR amplicon

To detect the presence or destruction of the BswI restriction endonuclease recognition site (5'-WCCGGW-3') by the 1475G>A (R492Q) mutation, PCR was performed as described in section 2.1.3. on *FMO3* exon 9 from patient DNA. 500 ng of PCR amplicons was digested at 37 °C for 2 hrs with 5 U *BsaWI*, in a solution containing 1X NEBuffer 2 (10 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$ and 1 mM Dithiothreitol, pH 7.9 @ 25°C), supplemented with Bovine Serum Albumin (BSA) (100 mg/mL), made up to a total volume of 50 µL with ddH_2O. The digested products were visualised by agarose gel electrophoresis as described in 2.1.3.

2.1.7. CpG site identification

Human FMO3 cDNA was submitted for sequence analysis to the Mutability (version 1.70) web server (http://www.hgu.mrc.ac.uk/Softdata/Mutability/).

2.2. Generation of Polymorphic variants of FMO3

2.2.1. Site-Directed Mutagenesis (SDM)

Site-directed mutagenesis was performed to generate each desired mutation.

Template DNA (5 ng) (hFMO3 cloned into the vector pET-21b(+) (Novagen, UK)); 5 μL Accubuffer (Bioline, UK); a pair of complementary oligonucleotide primers (see Table 6), each containing the desired mutation; 1 μL of 10 mM dNTPs (Bioline, UK); 2.5 U Accuzyme (Bioline, UK) and 41 μL ddH_2O were mixed in a 0.5-mL thin-walled Eppendorf tube and subjected to thermal cycling in a Techne GENIUS unit (Cambridge, UK) featuring a heated lid. An initial denaturation step was performed at 95°C for five min followed by an annealing step of 1 min at 55°C. This was followed by 15 cycles of 72°C for 6 min; 95°C for 30 s and 55°C for 30 s,. Finally, a single step of 72°C for 9 min completed the SDM amplification process.

The DNA was cleaned by adding 50 μL of SureClean (Bioline, UK) to the reaction mixture, incubating for 10 min and then centrifuging at 17,900 x g for 15 min. The supernatant was removed and the pellet washed with 500 μL of 70% ethanol and the sample was then centrifuged at 17,900 x g. DNA was resuspended in 44 μL of ddH_2O and 5 μL of NEB buffer 4 (NEB, UK) and digested with 20 U of *Dpn*I (NEB, UK) for 2 h at 37 °C. *Dpn*I endonuclease specifically digests methylated and hemi-methylated DNA. This means that the template DNA is digested leaving only the newly synthesised DNA, which should contain the desired mutation.

Table 6. Sequence of forward and reverse primers used to generate each mutation in hFMO3 cDNA

Mutant (a.a. level)	Strand	Primer Sequence	Original Codon	Mutated Codon
E158K	Sense	5'-CTACCAAAA**A**AGTCCTTTC-3'	G**A**G	**A**AG
	Antisense	5'-GAAAGGACT**T**TTTTGGTAG-3'		
R238Q*	Sense	5'-TCGTCACTC**A**ATTTGGAACCT-3'	C**G**A	C**A**A
	Antisense	5'-AGGTTCCAAAT**T**GAGTGACGA- 3'		
V257M	Sense	5'-CTGGTTGTAC**A**TGAAGCAGA-3'	G**T**G	**A**TG
	Antisense	5'-TCTGCTTCA**T**GTACAACCAG-3'		
E308G	Sense	5'-GAATTCACAG**G**GACCTCGGC-3'	G**A**G	G**G**G
	Antisense	5'-GCCGAGGTC**C**CTGTGAATTC-3'		

All primers were synthesized by Eurogentec (Belgium) except where marked (*) which were synthesized by ElimBio (CA, USA). The nucleotide change is highlighted in bold and underlined in the primer sequence. The codon changes are also shown.

*Dpn*I-digested amplification product (1 μL) was added to 50 μL of Alpha-Select Gold Efficiency *E.coli* competent cells (deoR endA1 recA1 relA1 gyrA96 hsdR17(rk - mk+) supE44 thi-1 Δ(lacZYA-argFV169) Φ80δlacZΔM15 F-) (Bioline Ltd, UK) in a Falcon 2059 tube and incubated on ice for 20 min. The transformation mixture was heat-shocked at 42 °C for 45 s and then returned to ice for 2 min. 1 mL of SOC (Bacto Tryptone (20 g/L) (BD Diagnostic, MA, USA), Bacto Yeast Extract (5 g/L) (BD Diagnostic, MA, USA), 10 mM NaCl (Sigma-Aldrich, St. Louis, MO, USA), 2.5 mM KCl (Sigma-Aldrich, St. Louis, MO, USA), 10 mM $MgCl_2$ (Sigma-Aldrich, St. Louis, MO, USA), 10 mM $MgSO_4$ (Sigma-Aldrich, St. Louis, MO, USA) and 20 mM glucose (Sigma-Aldrich, St. Louis, MO, USA)) was added and the transformation mixture was incubated for 1hr at 37 °C in a shaking incubator. The mixture was then centrifuged for 5 min at 1200 x g. The bacterial pellet was resuspended in 100 mL of SOC (31 g (31 Capsules) per Litre Containing/litre: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 5 g $MgSO_4 \cdot 7H_2O$, supplemented with 20 mM Glucose (BDH, UK)) (Qbiogene, Inc., Irvine, CA) and plated onto a Luria-Bertani Broth (LB) ampicillin (amp) (50 μg/ml) agar plate (40 g (40 Capsules or 8 Large Capsules) per litre Contents/litre: LB Medium, 15 g Agar-B) (Qbiogene, Inc., Irvine, CA). The plates were incubated at 37 °C for ~16 hrs.

2.2.2. Plasmid isolation

2.2.2.1. Small-scale plasmid isolation

To isolate plasmid DNA from small volumes of bacterial cells, a ChargeSwitch Plasmid ER Mini kit (Invitrogen, CA) was used.

Overnight bacterial culture (1.5 mL) was pipetted into a 1.5 mL thick-walled Eppendorf tube. This was centrifuged at 17,000 x g for 10 min. The supernatant was completely removed. The pellet was resuspended in 300 µL of Resuspension Buffer containing RNase A (5 mg/ml). The mixture was pipetted and vortexed until no cell clumps were visible. 300 µL of ChargeSwitch Lysis Buffer was then added and the tube gently inverted until the solution became opaque. This solution was incubated for 4 min at room temperature. 300 µL of ChargeSwitch Precipitation Buffer was then added and the tube again inverted until no more precipitant was formed. The mixture was centrifuged at 17,900 x g in a bench-top microcentrifuge for 10 min, forming a compact white pellet up the side of the tube. The supernatant was transferred to a fresh tube containing 30 µL of ChargeSwitch Magnetic Beads (25 mg/ml) and incubated for 1 min at room temperature. The tube was applied to the MagnaRack for one min whereupon a compact pellet was formed consisting of DNA bound to the magnetic beads. The supernatant was removed and discarded and the beads were resuspended in 990 µL of ChargeSwitch Wash Buffer. The application to the MagnaRack and wash step was repeated and then the pellet was resuspended in 50-100 µL of ChargeSwitch Elution Buffer (10 mM Tris-HCl, pH 8.5) and incubated at room temperature for 1 min. The tube was applied to the MagnaRack for 1 min and the supernatant, containing the plasmid DNA was recovered.

2.2.2.2. Large-scale plasmid isolation

In order to extract DNA from large volumes of bacterial cells, a QIAGEN QIAfilter™ Plasmid Midi Kit (QIAGEN Inc, USA) was used.

5 mL LB containing 5 µL amp (50 µg/ml) was inoculated with a single colony and grown for 8 hrs at 37°C in a shaking incubator at 225 rpm. 25 µL of culture was used to

inoculate 25 mL of LB and 25 µL amp (50 µg/ml) and grown in a 500 mL Erlenmayer flask for 12-16 h at 37°C in a shaking incubator at 225 rpm.

Bacterial cells were harvested in 400 mL Sorvall centrifugation bottles by centrifugation at 6000 x g for 15 min at 4 °C. The supernatant was removed by inverting the open bottle. The bacterial pellet was resuspended in 4 mL of Buffer P1 containing RNase A (5 mg/ml). The bacterial pellet was resuspended by pipetting and vortexing until no cell clumps were visible. 4 mL of Buffer P2 was then added and the bottle was inverted 6 times. The mixture was incubated at room temperature for 5 min. 4 mL of chilled Buffer P3 was added to the lysate and the bottle was inverted 6 times. The lysate was immediately poured into the barrel of a QIAGEN Midi cartridge and incubated at room temperature for 10 min. During this incubation the QIAGEN-tip 500 was equilibrated by applying 10 mL of Buffer QBT and allowing the column to empty by gravity flow. The cap was removed from the QIAfilter outlet nozzle and the plunger was gently inserted into the cartridge. The lysate was filtered into the previously equilibrated QIAGEN-tip. The cleared lysate was allowed to enter the resin within the QIAGEN-tip by gravity flow. The QIAGEN-tip was washed twice with 10 mL of Buffer QC. The DNA was eluted in 5 mL of Buffer QF (10 mM Tris-HCl, pH 8.5).

The DNA was precipitated by adding 3.5 mL of isopropanol at room temperature. The sample was mixed and centrifuged immediately at 15,000 x g for 30 min at 4 °C. The supernatant was carefully decanted and the DNA pellet was washed with 1 mL of 70% ethanol at room temperature. The ethanol was transferred to a 1.5 mL Eppendorf tube and centrifuged at 15,000 x g for 10 min at 4 °C. The supernatant was removed and the pellet was air dried for approx 10 min to allow residual ethanol to evaporate. The DNA pellet was re-dissolved in 100 µL of ddH_2O, quantified and stored at -20 °C.

2.2.3. Heterologous expression of hFMO3

2.2.3.1. Cell growth and induction of expression

E.coli BL21 cells transformed with the appropriate cDNA were used for the expression of hFMO3 proteins. A modified version of the pET-21b(+) (Novagen, UK) with the His-tag coing region removed was used as the expression vector as this had previously been used in the lab with satisfactory performance. The His-tag was removed due to concerns that it would interfere with catalysis of the recombinant FMO3 enzyme.

A starter culture was grown in 25 mL LB and 5 µL of amp (50 µg/ml) was inoculated with a single colony from a freshly streaked plate containing (E.coli) BL21(DE3) (F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm* (DE3) pLysS (Cm^R)) (Novagen, UK) + hFMO3 transformants and grown in a shaking incubator at 37 °C until the OD_{600} was between 0.6 and 1.0. 1 L of fresh LB medium was inoculated with 10 mL of this starter culture and the cells were incubated at 37 °C until an OD_{600} of 0.4 was reached. Expression of hFMO3 was then induced by adding isopropyl βD-thiogalactopyranoside (IPTG) (Sigma, CA, USA) to a final concentration of 0.5 mM, supplemented with riboflavin (Sigma, CA, USA) to a final concentration of 50 mg/L and incubated overnight in a shaking incubator at 30 °C.

2.2.3.2. Bacterial cell harvest

The culture was transferred to an appropriate container, depending on the amount of culture (small culture – 50 mL Falcon Tube, large culture - Sorvall centrifugation bottle), and centrifuged at 1200 x g for 30 min at 4 °C. The supernatant was removed completely by inverting the tube/bottle.

2.2.3.3. Cell Lysis

Due to the variety of methods used to lyse cells in the stucy of FMOs in the literature, several methods were tested simultaneously, including two methods previously used in the lab 'in house' (methods III and VII below).

2.2.3.3.1. Method I [195]

The pellet was resuspended in 25 mL of lysis buffer (50 mM sodium phosphate buffer (pH7.5) (sodium dihydrogen orthophosphate 1-hydrate ($NaH_2PO_4H_20$) (BDH, UK) was added to 1M di-Sodium hydrogen orthophosphate anhydrous (Na_2HPO_4) (BDH, UK) until a pH of 7.5 was reached), 100 mM KCl (Sigma, CA, USA) and 1mM EDTA (BDH, UK) (a chelating agent added to sequester metal ions in order to minimise proteolysis) and incubated on ice for 20 min. The culture was sonicated for 4 periods of 30 s, with 1 min cooling periods in between.

2.2.3.3.2. Method II [146]

The pellet was resuspended in 25 mL of lysis buffer (50 mM sodium phosphate buffer (pH 7.5) and 0.5 mM phenylmethylsulfonyl fluoride (PMSF) (a serine protease inhibitor)) and incubated on ice for 20 min. The culture was sonicated for 3 periods of 2 min, with 2 min cooling periods in between.

2.2.3.3.3. Method III

The pellet was resuspended in 25 mL of lysis buffer (10% (v/v) glycerol (Sigma, UK), 10 mM sodium phosphate buffer (pH 7.5) and Lyzozyme (Sigma, UK) (0.75 mg/ml) (a enzyme that hydrolyses peptidoglycans in bacterial cell walls, weakening them)) and

incubated on ice for 20 min. The culture was sonicated for 3 periods of 30 s, with 30 s cooling periods in between.

2.2.3.3.4. Method IV [21]

The pellet was resuspended in 25 mL of lysis buffer (50 mM sodium phosphate buffer (pH 7.5), 0.5% Triton X100 (v/v) (Sigma, CA, USA) (a detergent used to solubilise membrane proteins) and 10% (v/v) glycerol) and incubated on ice for 20 min. The resuspended bacterial pellet was sonicated for 3 periods of 2 min, with 2 min cooling periods in between.

2.2.3.3.5. Method V [196]

The pellet was resuspended in 25 mL of lysis buffer (10 mM sodium phosphate buffer (pH 7.5), 10% (v/v) glycerol and Lysozyme (Sigma, UK) (0.75 mg/ml)) and incubated on ice for 20 min. The culture was sonicated for 16 periods of 15 s, with 2 min cooling periods in between

2.2.3.3.6. Method VI [197]

The pellet was resuspended in 25 mL of lysis buffer (100 mM KCl, 50 mM potassium phosphate buffer (pH 7.5) (*di*-Potassium hydrogen orthophosphate (K_2HPO_4) (1 M) and Potassium *di*-Hydrogen orthophosphate (KH_2PO_4) (BDH, UK) 1 M; K_2HPO_4 was added to KH_2PO_4 until pH = 7.5), 20% (v/v) glycerol, 1 mM EDTA and Lysozyme (1 mg/ml) (Sigma, UK)) and incubated on ice for 30 min. The culture was not sonicated.

2.2.3.3.7. Method VII

The pellet was resuspended in 25 mL of lysis buffer (150 μM KCl, 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.5) (Fisher Scientific, NJ, USA), 1 mM EDTA and 20% (v/v) glycerol) and incubated on ice for 20 min. The culture was sonicated for 3 periods of 30 s, with 1 min cooling periods in between.

2.2.3.4. Cell fractionation

Sonicated cultures were centrifuged at 1200 x g for 15 min. The supernatant was recovered and the pellet (Inclusion Bodies) was resuspended in 2 mL 8 M Urea (Fisher Scientific, NJ, USA) and stored at room temperature. The supernatant was centrifuged at 100,000 x g for 1 h at 4 °C. The supernatant (cytoplasm) was recovered and stored at -70 °C. The pellet (cell membrane) was resuspended in 1.5 mL of lysis buffer and stored at -70 °C.

2.2.4. Protein quantification

The method used was based on that of Lowry et al., (1951) [198]. A series of bovine serum albumin (BSA) (Bio-Rad, CA, USA) samples were prepared by diluting the stock solution (1.4mg/ml) with ddH_2O. 100 μL of each standard was pipetted into a 13 mL test tube. Experimental samples were also 100 μL in volume. The DC Protein Assay Kit (Bio-Rad, UK) was used to perform the assay. Reagent A (alkaline copper tartrate solution) (500 μL) was added to both the standards and the samples, which were immediately vortexed. Reagent B (dilute Folin reagent) (4 mL) was added to both the standards and the samples, and again the tubes were vortexed immediately. The tubes were allowed to stand for 15 min after which time the absorbance was stable for 1 hour. The absorbance was read at 750 nm using the GeneQuant *pro* RNA/DNA calculator (Amersham Pharmacia, UK). The

A750 readings of the BSA standards were used to construct a standard curve and the protein concentrations of the experimental samples were extrapolated from this.

2.2.5.1. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

All materials and reagents were supplied by Invitrogen (USA) except where stated.

Protein (20 µg) was placed in a 0.5 µL thin-walled tube along with NuPAGE LDS Sample Buffer (1X), 1 µL of β-mercaptoethanol and ddH_2O to make up to a total volume of 15 µl. The samples were heated at 95°C for 5 min. Samples were then loaded on to a 4-12% NuPAGE Novex Bis-Tris Gel. The gel was run in 50ml of 1X NuPAGE MES SDS Running Buffer at 200 volts, using the XCell *SureLock* Mini-Cell for 35 min. Prestained SDS-PAGE Standard (Broad Range) (3 µL) (Bio-Rad, CA, USA) were run alongside the samples, to provide molecular mass reference.

2.2.5.2. SDS-PAGE Staining

The SDS-PAGE gel was removed from the disposable casings. The gel was placed in a small container and 200 mL of ddH_2O was poured on top of it. As per manufacturer instruction, the gel was microwaved on full power for one min and then incubated for one min on a shaker. The water was removed and this step was repeated twice. The water was removed and 30 mL of SimplyBlue SafeStain (Invitrogen, CA, USA) was poured on top of the gel so the gel was submerged in the solution. The gel was microwaved on full power for 1 min and placed on a shaker for 10 min to incubate. The stain was removed and the gel was rinsed quickly with ddH_2O to remove excess stain. 200 mL of ddH_2O was added to the gel, which was then incubated for 10 min. The ddH_2O was removed and 200 mL of fresh ddH_2O was added and the gel was incubated until the desired result was achieved.

2.2.5.3. Western Blot of SDS-PAGE gel

2.2.5.3.1. Blotting

All materials were supplied by Invitrogen (CA, USA) unless otherwise stated. 200 mL of 1X NuPAGE Transfer Buffer was prepared by combining 10 mL NuPAGE Transfer Buffer, 20 mL Methanol (or 40 mL Methanol if blotting from two gels), 200 µL NuPAGE Antioxidant and making up to the final volume with ddH_2O. One nitrocellulose membrane filter paper sandwich per gel was soaked in the transfer buffer along with an appropriate number of blotting pads. A piece of pre-soaked filter paper was placed on top of the gel and any trapped air bubbles were removed. The gel was turned over and a pre-soaked transfer membrane was placed on the gel and any air bubbles were removed. Another pre-soaked filter paper was placed on top of the transfer membrane and any trapped air bubbles were removed. Two soaked blotting pads were placed into the cathode core of the XCell II Blot Module. The gel/membrane assembly was placed on the blotting pad, ensuring that the gel was closest to the cathode and the membrane closest to the anode. Enough pre-soaked blotting pads were added on top of the membrane assembly to rise 0.5 cm over the rim of the cathode. The anode was placed on top of the pads and the module was slid into the guide rails of the XCell *SureLock* Mini-Cell. The blot module was topped up with transfer buffer to cover the gel/membrane assembly. The outer chamber of the XCell *SureLock* Mini-Cell was filled with 650 mL of ddH_2O. The transfer was performed at 150 mA for 2 hrs.

2.2.5.3.2. Antigen detection

After protein transfer, the membrane was incubated for at least 1 hour at room temperature in 50 mL of blocking buffer (5% powdered non-fat milk (Safeway, CA, USA), 150 mM

NaCl (Sigma, CA) and 25 mM Tris-HCl (pH 7.5) (Sigma, CA, USA)). The blocking solution was removed and replaced with 50 mL of primary antibody solution (WB-FMO3 (BD Biosciences, CA, USA) at 1:5000 dilution in 0.5% powdered non-fat milk, 150 mM NaCl, 25 mM Tris-HCl (pH 7.5)). The membrane was then incubated for 1 hour at room temperature. The nitrocellulose was rinsed 3 times with 50 mL of wash buffer (0.1% Tween 20 (Bio-Rad, CA, USA), 25 mM Tris-HCl (pH 7.5), 150 mM NaCl) for 15 min each. The nitrocellulose was then incubated at room temperature for 1 hour with 50 mL of secondary antibody solution (Horse Radish Peroxidase-conjugated goat anti-rabbit IgG at 1:10,000 dilution (BD Biosciences, CA, USA) in 0.5% powdered non-fat milk, 150 mM NaCl, 25 mM Tris-HCl (pH 7.5)). The nitrocellulose was washed 3 times with 50 mL of wash buffer. The nitrocellulose was then developed using the ECL Western Blotting Analysis System (Amersham, UK). 1 mL of Reagent A was added to 1 mL Reagent B and the mixture was placed on top of the nitrocellulose, using a pipette to ensure that the whole membrane was in contact with the mixture. The blot was then wrapped in Saran wrap and placed in contact with a sheet of Kodak Biomax Light autoradiography film (Kodak, CA, USA). The film was exposed and developed. This procedure was repeated varying the exposure time until a satisfactory level of detection was observed (between 30 s and 10 min).

2.2.6. Immunoquantification of FMO3

Various amounts of quantified FMO3 standards (BD Biosciences, CA, USA) were loaded on a 4-12% NuPAGE Novex Bis-Tris Gel (Invitrogen, CA, USA) along with duplicate samples of 5 µg and 10 µg of bacterial extract containing an unknown amount of expressed recombinant protein. SDS-Page was performed as described in Section 2.2.5 and a western blot was performed on these gels as described in Section 2.2.5.3. The developed western

blot film was scanned using a Hewlett Packard Scanjet 4850 Flatbed Scanner (Hewlett Packard, NJ, USA) and the results were saved in TIFF format. The amount of FMO3 in the recombinant bacterial protein sample was quantified densitometrically, using ImageJ software [199], by comparison with a standard curve of FMO3 standards.

2.2.7. Methimazole Assay of FMO activity [130]

Bacterial membrane protein fractions containing quantified human FMO3 were obtained using cell lysis method VII (2.2.3.3.7.). All other methods of lysis tested yielded protein deemed inactive as no activity was observed spectroscopically. FMO3 was activity was assayed as follows. 2 mL of assay mixture was prepared as follows. A mixture containing 4 mM Dithiothreitol (DTT) (Sigma, CA, USA), 12 mM 5,5'-Dithio-bis(2-nitrobenzoic acid) (DTNB) (Sigma, CA, USA), 20 mM β-Nicotinamide adenine dinucleotide phosphate (reduced form) (NADPH) (Sigma, CA, USA) was made up in Assay Buffer (Tris-HCl (pH 8.4), 1 mM EDTA). The mixture was divided between two clear 2 mL plastic cuvettes (Sarstedt, Germany). All reagents and protein were equilibrated to 37 °C during preparation of the assay. Both cuvettes were placed inside a Cary Split-beam Spectrophotometer 3000 (Varian, CA, USA), with the 1 mL mixture acting as a reference. The mixtures were incubated at 37 °C before the non-reference cuvette was spiked with 2-Mercapto-1-methylimidazole (Methimazole) (Lancaster Synthesis, NH, USA) to a range (1000-3 μM) of final concentrations in a final volume of 1 mL. The absorbance at 412 nm (the disappearance of the yellow coloured TNB to DTNB (colourless)) was followed for 5 min after the baseline had settled. Enzyme velocities were calculated using the TNB extinction coefficient of 13,600 M^{-1} cm^{-1}. The enzyme kinetic parameters, K_M, Vmax were estimated from Hanes-Woolf [200] linear transformations of the Michaelis-Menton equation using the Enzyme Kinetics module (v. 1.3.) of the SigmaPlot (v 10.0.) program

(Systat Software Inc., CA, USA). This transformation has been shown to be superior to others such as the Eadie-Hofstee [201] and Lineweaver-Burk [202], because it applies the same weights to all substrate concentrations, unlike the Eadie-Hofstee and Lineweaver-Burk which place a high weight on low concentration values causing a bias in the data towards reflecting the enzyme behaviour at these lower concentrations as representative of the enzyme behaviour as a whole [203]. K_M, *K*cat and *K*cat/ K_M values were calculated for each FMO3 variant tested and these were compared with those of wild type FMO3 (n=3 for each protein). Statistical significance was assessed using an unpaired t-test performed using GraphPad Prism 4.0 (GraphPad Software, CA, USA).

2.3. Generation of a homology model of FMO3

2.3.1. Template Identification

The blastp function of the Basic Local Alignment Search Tool (BLAST) [204] was used to search the Protein Data Bank (PDB) for homologues similar at a global primary sequence level to human FMO3. GenTHREADER [205] was used to look for structural homologues similar at a fold-level, inferred from sequence-sequence alignment against library of known folds derived from solved protein structures. This latter step was especially important in this case as it has been shown that as the sequence identity drops into the twilight zone (see section 1.3.), there may be insufficient signal in the sequence-based methods (e.g. BLAST) to detect distant relationships [206]. Threading methods (e.g. GenTHREADER) have the ability to detect common folds that may be present even in the absence of significant sequence homology [207].

2.3.2. Homology Modelling

Homology Modelling was performed using MODELLER [161].

2.3.2.1. Homology Modelling hFMO3

The FMO3 (target) sequence and the template sequence(s) were aligned using the ClustalW algorithm [208] web server at the European Bioinformatics Institute (EBI) (http://www.ebi.ac.uk/clustalw). The Blosum scoring matrix [209] was used with penalty values set at 10 for opening or closing a gap, and 0.05, for extending a gap. The output was saved in .pir format. The 'model-default.py' python script was modified to insert the location of the .ali file (the renamed PIR file), the location of the template PDB file and the number of models to build. The 'model-default.py' python script was executed using a 1700 Advent PC with a Mobile Intel Pentium processor 4 (3.20GHz) and 512 MB of RAM within the DOS command line, to generate the models. The 'model-loop.py' python script was used in an attempt to improve the accuracy of loop predictions generated by the previous script.

2.3.2.2. Model scoring

Regions of probable inaccuracy were defined by positive ProSa [210] profiles or negative VERIFY_3D [211] profiles. The stereochemical quality of the models was also assessed using WHAT_CHECK [212] to assess the side-chain environment and PROCHECK [213] to validate the backbone and side-chain conformation and Swiss-PDBViewer v.3.7 [164] was used to generate Ramachandran plots (Figure 5, section 1.3.).

2.3.2.3. Model refinement

The 'dope_loopmodel.py' python script was modified to direct it to the models generated and then executed using a 1700 Advent PC with a Mobile Intel Pentium 4 processor (3.20GHz) and 512 MB of RAM within the DOS command line. This method uses the more advanced 'Discrete Optimised Protein Energy' (DOPE) method to assess the spacing of the atoms in the model and refine them at the cost of computational time rather than simply using 'loopmodel.py'. Additionally, the regions of probable inaccuracy identified in Section 2.3.2.2., which corresponded to loop regions, were targeted specifically for loop modelling. Several versions of any model are returned with various loop arrangements, the one with the lowest energy being taken as the 'best' orientation (see Section 1.3.).

Dihedral angles within the polypeptide model were modified using the Ramachandran plot within Swiss-PDBViewer v.3.7, which has click-and-drag capabilities allowing side chains to be moved within the plot, from unlikely positions to more favourable positions as determined by their proximity to the favoured regions of the ramachandran plot (see Figure 5), resulting in a more robust biophysical model.

2.3.2.4. Modelling FAD into the FMO3 model

In the case of the FMO3 model generated using the structure "Crystal structure of a protein with similarity to flavin-containing monooxygenases and to mammalian dimethylalanine monooxygenases" (PDB code, 1VQW) [214], the FAD molecule present in the template was modelled into the putative active site of the FMO3 model. The atomic coordinates of the FAD molecule within the PDB file of 1VQW were assigned a chain identifier manually using a text editor. A 'blk' character was added to the .ali file at the end of both the target and template sequences to represent the ligand. This effort was performed in an attempt to

gain insight into the possible interactions of the FAD molecule within the FMO3 homology model.

The 'model-ligand.py' script was executed using a 1700 Advent PC with a Mobile Intel Pentium 4 processor (3.20GHz) and 512 MB of RAM within the DOS command line to generate a model that included the FAD molecule in the structure.

2.3.2.5. Putative Active Site assignment

Models based on the PDB crystal structure of "Phenylacetone Monooxygenase, a Baeyer-Villiger Monooxygenase" (PDB code, 1W4X) [215] were examined for hydrophobic pockets, initially using Q-SiteFinder [216] which uses the interaction energy between the protein and a van der Waals probe to locate energetically favourable binding sites within the molecule. FAD- and NADPH- binding motifs were highlighted on the model, which provided evidence, when found in the proximity of known enzyme catalytic residues identified using the Catalytic Site Atlas [217], that the model made sense in terms of the FMO3 catalytic cycle (Figure 1). Finally, the CAVER [218] module of the PyMol visualisation tool was used to identify buried cavities within the model to show the possible physical contact between substrates within the molecule.

Models based on 1VQW were assessed by eye as the active site of the ligand (methimazole) bound state of the template model had been deposited in the PDB, named "Crystal structure of flavin-containing monooxygenase (FMO)from S.pombe and substrate (methimazole) complex" (PDB code, 2GVC).

2.3.2.6. Examination of FAD-FMO3 interactions

The hydrophobic and hydrophilic interactions between the modelled FAD molecule within the FMO3 homology model was examined using the Ligand Explorer tool within the Molecular Biology Toolkit (MBT) suite [219].

2.3.2.7. Tertiary Model Alignment

Tertiary models of FMO3 generated by Modeller were aligned using FATCAT [220] (http://fatcat.burnham.org/), for pairwise alignment, or POSA [221] (http://fatcat.burnham.org/POSA/), for multiple structure alignment.

2.3.3. Visualization of molecules

Molecules were visualised using PyMol (www.pymol.org).

2.5. Evolution studies of the *FMO3* locus

Previously, DNA had been obtained by the Yamazaki group at Showa Pharmaceutical University, Japan, from 23 unrelated Japanese (18 male and 5 female ranging in age from 19 to 52 years old) who had responded to an internet article because of self-perception of TMAU symptoms, specifically a body odour smelling strongly of rotten fish. Urinary TMA/TMA *N*-oxide was examined by the Yamazaki group and DNA was also obtained from 45 Japanese non-TMAU sufferers. Coding exons, together with 3.4 kb of upstream sequence and 2.9 kb of flanking intron and 3' untranslated gene regions were amplified by PCR and sequenced. Haplotypes were inferred by the Yamazaki group. For details of the haplotypes inference and PCR conditions, see Allerston et al., 2007 [222].

2.5.1. Data Analysis

2.5.1.1. Mutational relationship between haplotypes

All pairwise comparisons between SNPs were analysed for evidence of linkage disequilibrium (measured as R^2) using DnaSP (version 4.10.3) [223]. Any significant associations between SNPs were identified using χ^2 tests with a Bonferroni correction for multiple comparisons. The minimum number of recombination events (Rm) within the samples was estimated by the four-gamete test [224]. To visualise mutational relationships between haplotypes, the Reduced Median (RM) Network algorithm [225], as implemented by Network 4.1.1.2 (http://www.fluxus-engineering.com), was used.

2.5.1.2. Probing the time depths of *FMO3* variants

In an attempt to estimate the time to the most recent common ancestral sequence (T_{MRCA}) and the ages of some of the mutations in our sample, we used the program GENETREE (version 9.0), which uses maximum-likelihood coalescent analysis [226]. The method is based on a standard coalescent model, which assumes an infinite-sites model of mutation (a model where each new mutation occurs at a site that has not mutated before) with no recombination. Thus, only infinite-sites-compatible data were used for this analysis. Under the coalescent model, mutations are assumed to occur along ancestral lineages according to a Poisson process of rate $\theta/2$, where θ (the population mutation parameter) = $4N_e\mu$. N_e is the effective population size (the size of a theoretical, randomly mating, isolated and stationary population, which would have the same distribution and type of polymorphism as the population examined) and μ is the neutral mutation rate per locus per generation, which can be calculated from the equation $\mu = \nu gL$, where ν is the neutral mutation rate per nucleotide per year, g is the generation time in years and L is the number of silent sites in

the sequence. An estimate of the parameter θ was obtained through the use of the mL function of GENETREE and was refined by using the surface option of the program. This, together with our estimate of the neutral locus mutation rate, enabled N_e to be determined, from the equation $\theta = 4N_e\mu$. Coalescent simulations performed using GENETREE gave estimates for T_{MRCA} and the ages of SNPs in coalescent units of $2N_e$ generations, which were converted to years by using our estimate of N_e and a generation time of 20 years.

2.5.1.3. Testing *FMO3* for departures from neutrality

DnaSP was used to calculate two estimates of sequence diversity, Watterson's estimator of θ (θ_W) [174] and nucleotide diversity (π) [175], and to determine a number of test statistics: Tajima's *D* [176], Fu and Li's *F** and *D** [177], the expected number of haplotypes (h) [227], Fu's F_S [228], raggedness (r) [179], Z_{nS} [181] and Wall's *B* and *Q* [180]. Significance values for the test statistics were assessed by comparison with a distribution of estimates obtained from coalescent simulations of 5,000 random samples of the same sample size and level of polymorphism as the observed data. Simulations were performed assuming constant population size and no recombination (making tests conservative) or for the level of recombination in the observed data. DnaSP was also used to analyse the allele frequency spectrum and the pairwise differences among *FMO3* haplotypes, via Rogers' mismatch test [179], performed using Arlequin (version 3.01) [229] (http://lgb.unige.ch/arlequin/). The ratio of replacement to silent polymorphisms within humans was compared with the ratio of replacement to silent fixed differences between humans and chimpanzee, by means of a McDonald-Kreitman test [182], using a Fisher's exact test of independence. Wright's fixation index (F_{ST}) [184] was used to measure population subdivision by comparing genetic diversity within subpopulations to that of the

whole population by comparing the *FMO3* SNPs of the Japanese cohort to population panels provided in dbSNP [171].

Chapter 3

Results and Discussion

3. Results and discussion

3.1. TMAU Patient studies

Introduction

This section describes the analysis of the DNA from several suspected TMAurics. In each case the individual concerned had been diagnosed, through TMA:TMA *N*-Oxide ratio anaysis through urine testing, as having suspected primary TMAuria. The *FMO3* locus was examined for mutations that would provide evidence for a TMAuria diagnosis and provide further insight into the genetics and molecular biology of this disease. Exons 2 to 9 were sequenced along with three flanking upstream regions of *FMO3* analysed (-2310 to -1716, -2733 to -2162 and -1906 to -1436 (containing the non-coding exon 1) where the 'A' of the ATG start codon of FMO3 is considered +1) by sequencing in any of the suspected patients.

Subject one

A medical doctor from Northern Norway contacted us after the parents of a female child (2 years old) had complained about a strong bodily odour, reminiscent of rotten fish, emanating periodically from the child. Neither parent complained of a similar problem, but after further questioning, the mother did report of an uncle who lived alone, has never married and she remembered that he had an odd smell. The parents were not consanguineous.

The parents were reluctant to bleed the child and so their blood was obtained first in order to check their DNA, after which time, if anything interesting was found, the child would then be bled for DNA extraction. In the meantime the mother contacted her uncle and introduced him to the doctor. His DNA was also obtained. DNA samples were sent

from the Norwegian doctor to UCL, where the sequencing analysis was of the *FMO3* locus (mentioned above and described in detail in 2.1.).

A previously unreported mutation was found in exon 6 in a heterozygous state in the mother and in the subject. This same mutation was found in a homozygous state in the great uncle confirming a Primary TMAU diagnosis for this individual.

A relatively common *FMO3* haplotype was discovered in the father, but this was not found to have been inherited by the subject and so did not contribute to the TMAU phenotype.

No other mutations were observed in the regions amplified.

Mother of subject one

The sequencing trace indicated that the mother was homozygous for the E158 background (Figure 6 A). The trace also indicated that the mother was a heterozygote for the 713G>A (R238Q) mutation (Figure 6 B).

Father of subject one

The sequencing trace indicates that the father is heterozygous for the 472G>A (E158K) polymorphism (Figure 7 A) and for the 923A>G (E308G) (Figure 7 B). From this information it was not possible to deduce if these polymorphisms were in *cis* or *trans*.

Great Uncle of subject one

The sequencing trace indicated that the great uncle was homozygous for the 713G>A (R238Q) polymorphism (Figure 8).

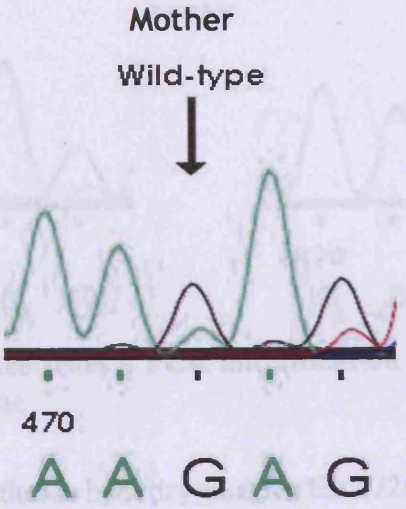


Figure 6 A). A sequence trace from a PCR amplification of exon 4 of the *FMO3* gene from the mother of subject one

The trace indicated that the mother of subject one was homozygous for the Glu^{158} background.

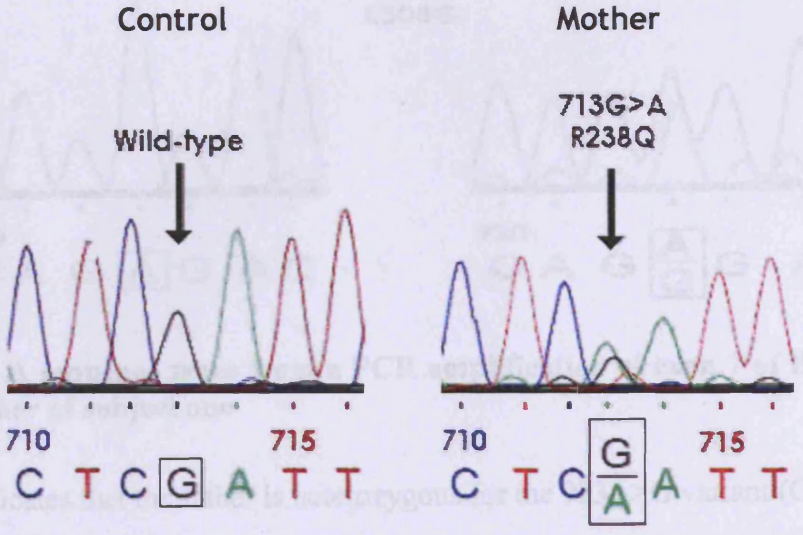


Figure 6 B). A sequence trace from a PCR amplification of exon 6 of the *FMO3* gene from the mother of subject one

The trace indicates that the mother is heterozygous for the 713G>A ($Arg^{238}Gln$) polymorphism.

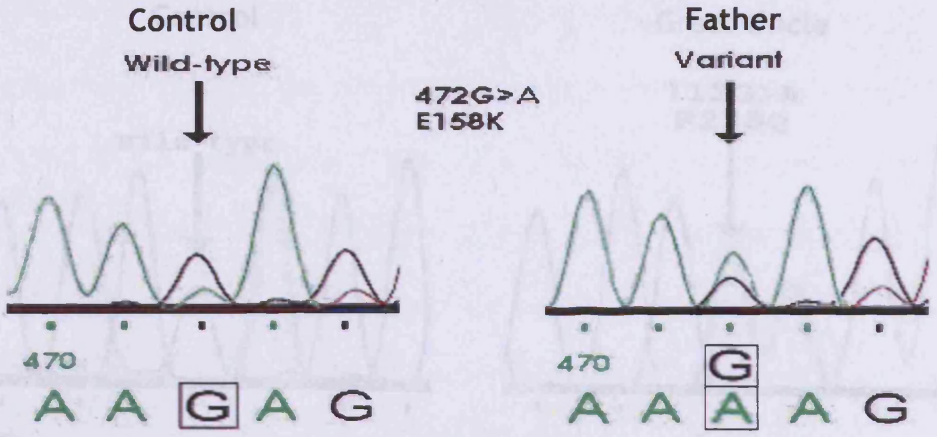


Figure 7 A). A sequence trace from a PCR amplification of exon 4 of the *FMO3* gene from the father of subject one

The trace indicates that the father is heterozygous for the 472G>A variant ($Glu^{158}Lys$).

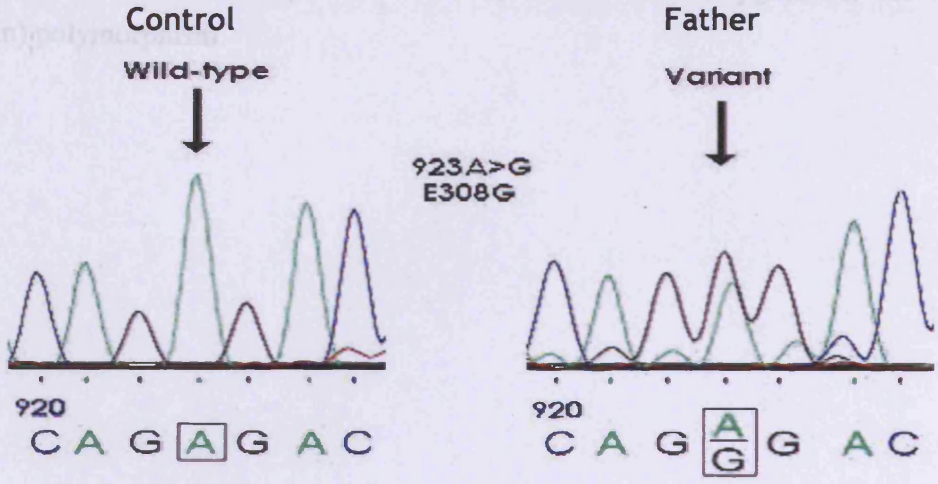


Figure 7 B). A sequence trace from a PCR amplification of exon 7 of the *FMO3* gene from the father of subject one

The trace indicates that the father is heterozygous for the 923A>G variant ($Glu^{308}Gly$).

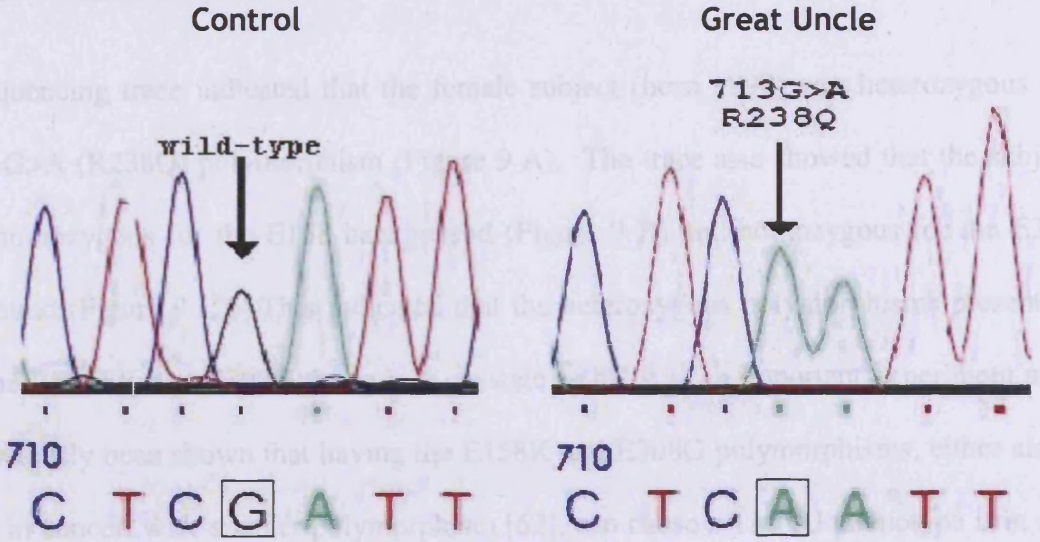


Figure 8. A sequence trace from a PCR amplification of exon 6 of the *FMO3* gene from the great uncle of subject one

The trace indicated that the great uncle of subject one is homozygous for the 713G>A ($Arg^{238}Gln$) polymorphism.

Subject one - Norwegian child

The sequencing trace indicated that the female subject (born 1999) was heterozygous for the 713G>A (R238Q) polymorphism (Figure 9 A). The trace also showed that the subject was a homozygous for the E158 background (Figure 9 B) and homozygous for the E308 background (Figure 9 C). This indicated that the heterozygous polymorphisms present in the father (E158K and E308G) were in a *cis* state. This was an important experiment as it has previously been shown that having the E158K and E308G polymorphisms, either alone [70] or in concert with another polymorphism [62], can cause a TMAU phenotype if in *cis*. Both polymorphisms had previously been shown to be present in the paternal *FMO3* in a heterozygous state, but the phase of these polymorphisms was unknown. This experiment showed that the polymorphisms were in *cis* and that this allele had not been inherited by the subject.

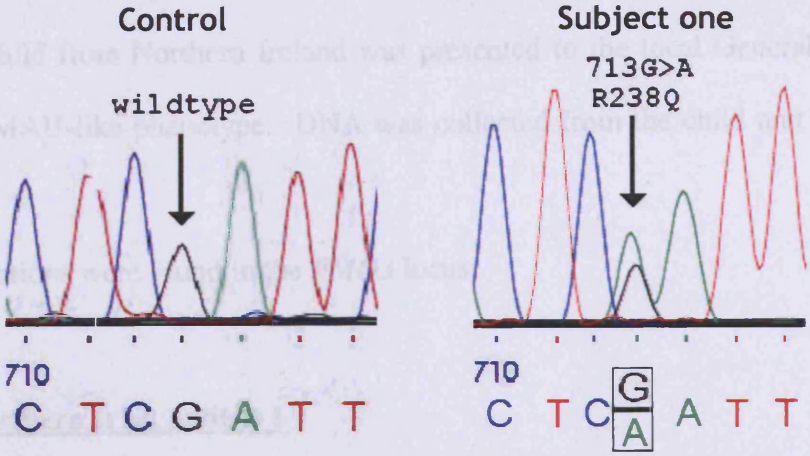


Figure 9 A). A sequence trace from a PCR amplification of exon 6 of the *FMO3* gene from subject one
The trace indicates that subject one is heterozygous for the 713G>A ($Arg^{238}Gln$) polymorphism.

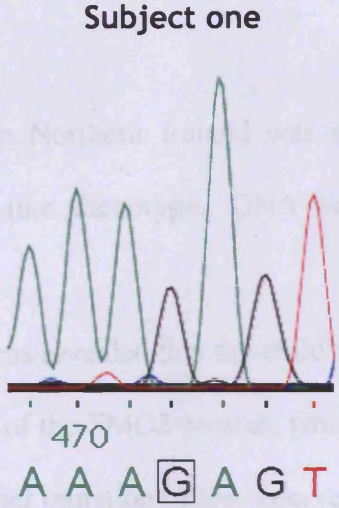


Figure 9 B). A sequence trace from a PCR amplification of exon 4 of the *FMO3* gene from subject one
The trace indicates that subject one is homozygous for the Glu^{158} variant.

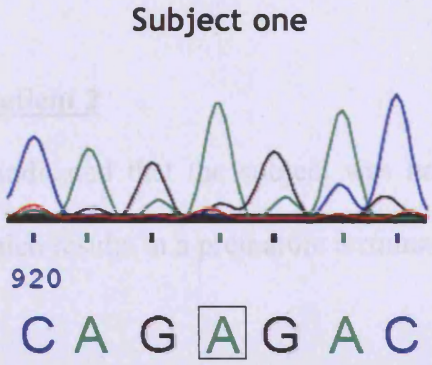


Figure 9 C). A sequence trace from a PCR amplification of exon 7 of the *FMO3* gene from subject one
The trace indicates that subject one is homozygous for the Glu^{308} variant.

Subject two

A 1 year old child from Northern Ireland was presented to the local General Practitioner displaying a TMAU-like phenotype. DNA was collected from the child and sent to UCL for analysis.

No mutations were found in the *FMO3* locus.

Subject 2 - Northern Irish patient 1

Analysis revealed there were no mutations in the subject's *FMO3* gene or in the upstream regions sequenced.

Subject three

A 4 year old female child from Northern Ireland was presented to the local General Practitioner displaying a TMAU-like phenotype. DNA was collected from the child and sent to UCL for analysis.

Analysis of the *FMO3* locus revealed that the child was homozygous for a mutation in exon 7 leading to a truncation of the FMO3 protein, which has previously been reported to destroy FMO3 activity. No other mutations were observed in the regions amplified.
A previously unreported mutation was found in exon 9 in a heterozygous state. No other mutations were observed in the regions amplified.

Subject 3 - Northern Irish patient 2

The DNA sequencing trace indicated that the subject was homozygous for the 913G>T polymorphism (Figure 10) which results in a premature termination at codon 305 in exon 7.

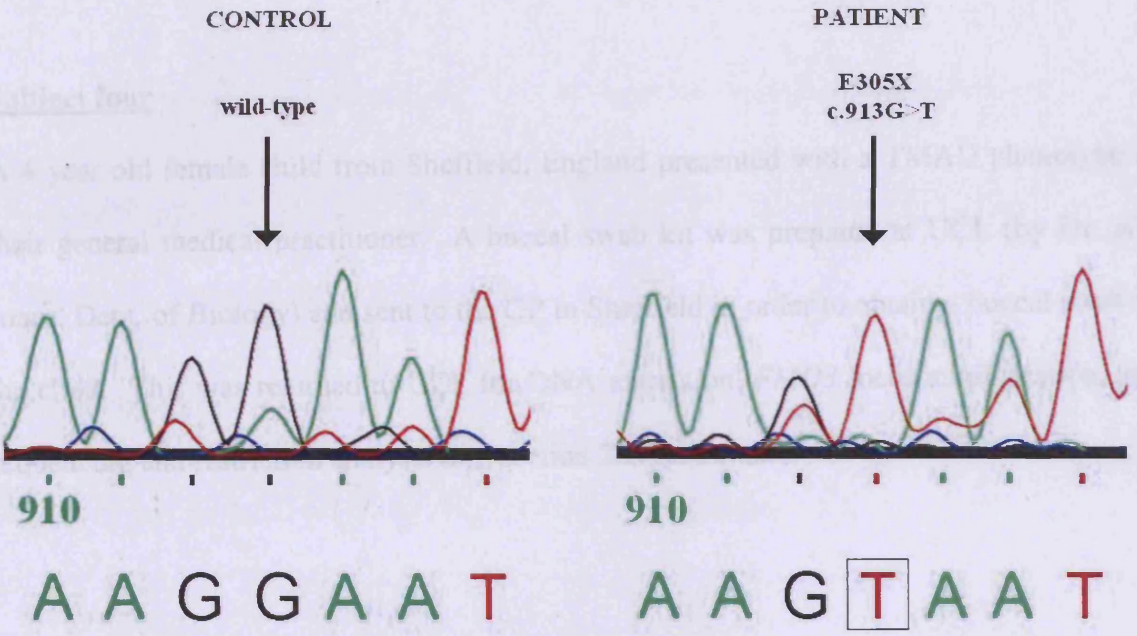


Figure 10. A sequence trace from a PCR amplification of exon 7 of the *FMO3* gene from subject three

The trace indicated that the subject is homozygous for the 913G>T ($Glu^{305}X$) polymorphism.

Subject 4 - Sheffield, England patient

The sequencing traces indicated that the 4 year old subject was heterozygous for a 1475G>A (R492Q) polymorphism (Figure 11). A restriction digest assay of a PCR amplified exon 9 of subject FMO3 also indicated that this was this case (Figure 12).

Subject four

A 4 year old female child from Sheffield, England presented with a TMAU phenotype to their general medical practitioner. A buccal swab kit was prepared at UCL (by Dr. Abi Jones, Dept. of Biology) and sent to the GP in Sheffield in order to obtain a buccal swab of the child. This was returned to UCL for DNA extraction, *FMO3* locus amplification, and sequencing and restriction analysis (see section 2.1. for details).

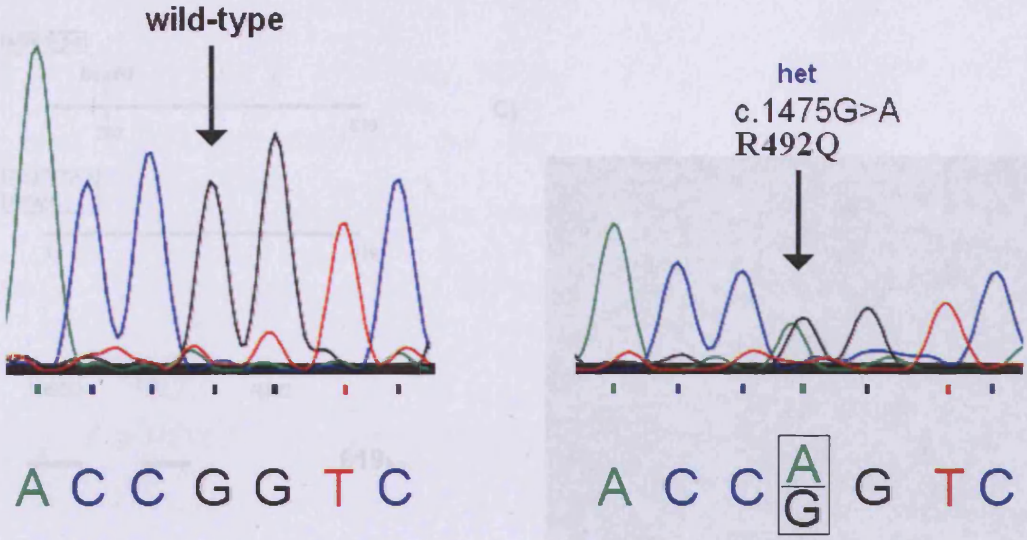


Figure 11. Sequencing trace showing the region of *FMO3* exon 9 containing the 1475G>A polymorphism of subject four

When comparing sequencing trace of subject four (right) to the wild-type trace (left) a distinct G/A double peak can be observed in the subject. This indicated that the subject was heterozygous for the 1475G>A (Arg^{492} Gln) polymorphism.

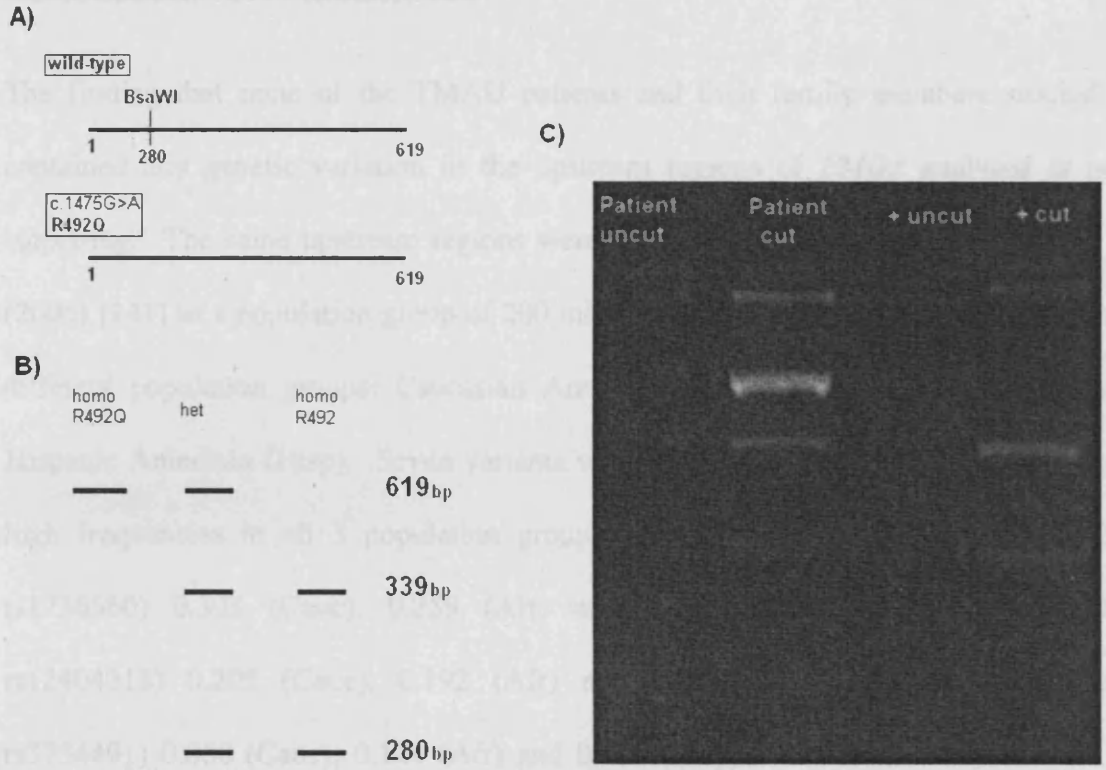


Figure 12. Restriction digest analysis of *FMO3* exon 9 from subject 4 DNA with *BsaWI*

A) A wild-type PCR product of FMO3 exon 9 yield a 619bp product containing a single restriction site 280bp from the start of the exon, whereas the 1475G>A mutant contains no such restriction site. B) A *BsaWI* digest of a wild-type homozygote would therefore yield a 339bp and a 280bp fragment. A 1475G>A homozygous mutant would not cut at all and a heterozygote digest yield a 339bp fragment, a 280bp fragment and a 619bp uncut fragment. C) The subject's restriction digest generated restriction fragments suggesting that subject four was indeed a heterozygote for the 1475G>A (Arg^{492} Gln) mutation.

TMAU Patient studies - discussion

The finding that none of the TMAU patients and their family members studied (n=7) contained any genetic variation in the upstream regions of *FMO3* analysed is perhaps surprising. The same upstream regions were analysed in a study by Koukouritaki et al., (2005) [141] in a population group of 200 individuals (400 chromosomes) each from three different population groups; Caucasian American (Cauc), African American (Afr) and Hispanic American (Hisp). Seven variants were discovered with 3 occurring at relatively high frequencies in all 3 population groups; -2650C>G (dbSNP [230] Build 127 id# rs1736560) 0.303 (Cauc), 0.259 (Afr) and 0.438 (Hisp); -2543T>A (dbSNP id# rs12404218) 0.205 (Cauc), 0.192 (Afr) and 0.291 (Hisp); -2177G>C (dbSNP id# rs3754491) 0.068 (Cauc), 0.124 (Afr) and 0.286 (Hisp). The same study showed that -2650C>G on its own and together with 2543T>A has been shown to increase luciferase reporter expression, with the double variant resulting in an 8-fold increase in reporter activity. It is conceivable that the lack of these variants in our patient cohort may be a compounding, were other mutations or polymorphisms were found within the coding regions, or causative, where no mutations were found with regards to the TMAU phenotypes reported. However, our own investigation of the sequencing information from the 22 putative TMAU sufferers (TMAU) (44 chromosomes) and 45 control individuals (Con) (90 chromosomes) from a Japanese population featured all three of the upstream SNPs at relatively high frequencies in both groups; -2650C>G 0.545 (TMAU) and 0.578 (Con); --2543T>A 0.114 (TMAU) and 0.100 (Con); -2177G>C 0.114 (TMAU) and 0.110 (Con) (Table 10, section 3.4.). It is possible that the lack of variants in the Norwegian family cohort (see section 3.1.) is due to population differences. This seems unintuitive though, both because of the otherwise global occurrence at high frequency detailed above and due to the subject and her family being of European ancestry, perhaps represented

genetically in the Caucasian American group of Koukouritaki et al's study and recent data from the HapMap [172] project. SNP information deposited from this project in dbSNP found the allele frequencies for the 3 SNPs in a European panel of 120 chromosomes to be 0.308 (-2650C>G), 0.100 (-2543T>A) and 0.100 (-2177G>C).

Sequencing a control group from specific populations from which the patients belong to in tandem with a patient population could shed more light on this observation.

Norwegian Cohort – Subject one

Initially, an infant with a phenotype typical of TMAU was presented to a genetic councillor (Dr. Hildegunn Hoeberg Vetti, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Norway). The parents wanted to confirm this TMAU via a genetic screen, but were reluctant to bleed the child. Therefore, both parents agreed to be screened instead, from which the genotype of the child could be inferred. The results (detailed in sections 3.1. and 3.1.) were interesting. The father is heterozygous for the E158K and E308G variants in *cis* (see section 3.1.2.1.2 and Figure 13). When found in *cis*, these relatively common polymorphisms can result in a reduction in FMO3 activity [21, 24, 147]. When the mother was shown to be heterozygous for a new variant, R238Q (see section 3.1. and Figure 13), a residue shown to cause TMAU when changed to tryptophan [56], it led to the postulation that the child had inherited the E158K and E308G variants in *cis* allele from the father and the allele containing the R238Q variant from the mother, reducing levels of FMO3 activity to cause the observed TMAU phenotype. To confirm this idea, DNA was extracted from the child and analysed. In the meantime the mother of subject one had relayed information to the genetic councillor regarding her uncle whom she remembered as always having a strong smell of fish and had led a somewhat

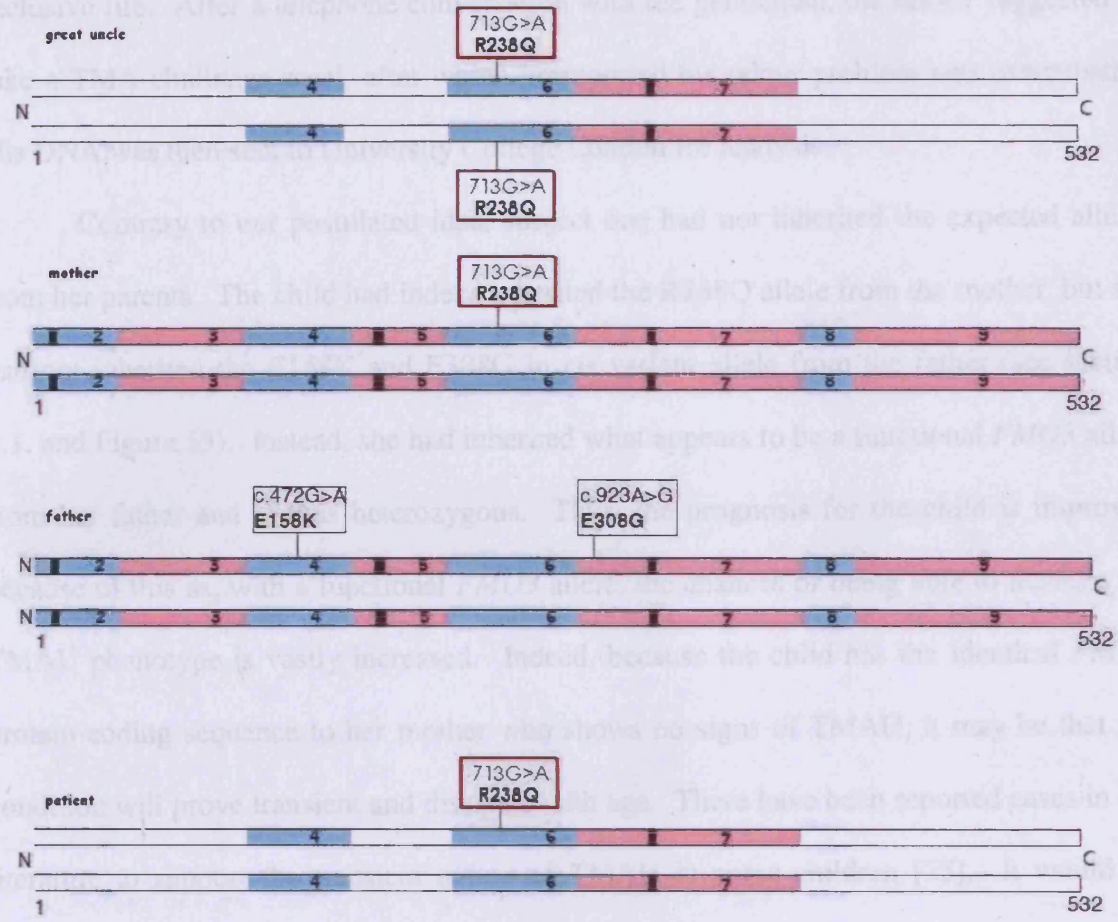


Figure 13. A representation of the 2 alleles of *FMO3* present in the great uncle, mother, father and subject one, with the mutations highlighted

The great uncle is homozygous for the R238Q mutation. The mother is heterozygous for this mutation. The father is heterozygous the E158K and E308G polymorphisms in *cis*. Subject one has inherited the R238Q mutation from the mother and the allele without the polymorphism in exon 4 and exon 6 from the father. The white sections indicate regions, in subject one and great uncle, which have not been sequenced.

reclusive life. After a telephone conversation with the gentleman, the doctor suggested he take a TMA-challenge meal, after which he reported his odour problem was exacerbated. His DNA was then sent to University College London for analysis.

Contrary to our postulated idea, subject one had not inherited the expected alleles from her parents. The child had indeed inherited the R238Q allele from the mother, but she had not inherited the E158K and E308G in *cis* variant allele from the father (see section 3.1. and Figure 13). Instead, she had inherited what appears to be a functional *FMO3* allele from her father and is thus heterozygous. Thus, the prognosis for the child is improved because of this as, with a functional *FMO3* allele, the chances of being able to manage the TMAU phenotype is vastly increased. Indeed, because the child has the identical *FMO3* protein-coding sequence to her mother who shows no signs of TMAU, it may be that the condition will prove transient and dissipate with age. There have been reported cases in the literature to support the transient nature of TMAU in some children [75]. It would be interesting to analyse the urine TMA/TMA *N*-oxide ratio of the mother of subject one to confirm whether the effect of being heterozygous for the variant R238Q is indeed devoid of a TMAU phenotype as she reports.

In contrast, the great uncle of the child is homozygous for the R238Q variant of FMO3 and as such would be considered as a TMAU sufferer (see section 3.1. and Figure 13). The great uncle has since received genetic counselling.

The genetic pedigree of this family paints an interesting genetic lineage not least because it contained a consanguineous relationship in the family; with the great uncle being the offspring of two first cousins (see Figure 14). The documented high levels of consanguinity in northern Norway [231], where the family are from, may have given rise to the prevalence of this, otherwise rare allele (it is not present in dbSNP) in this isolated population.This investigation has revealed a previously unreported mutation in *FMO3*,

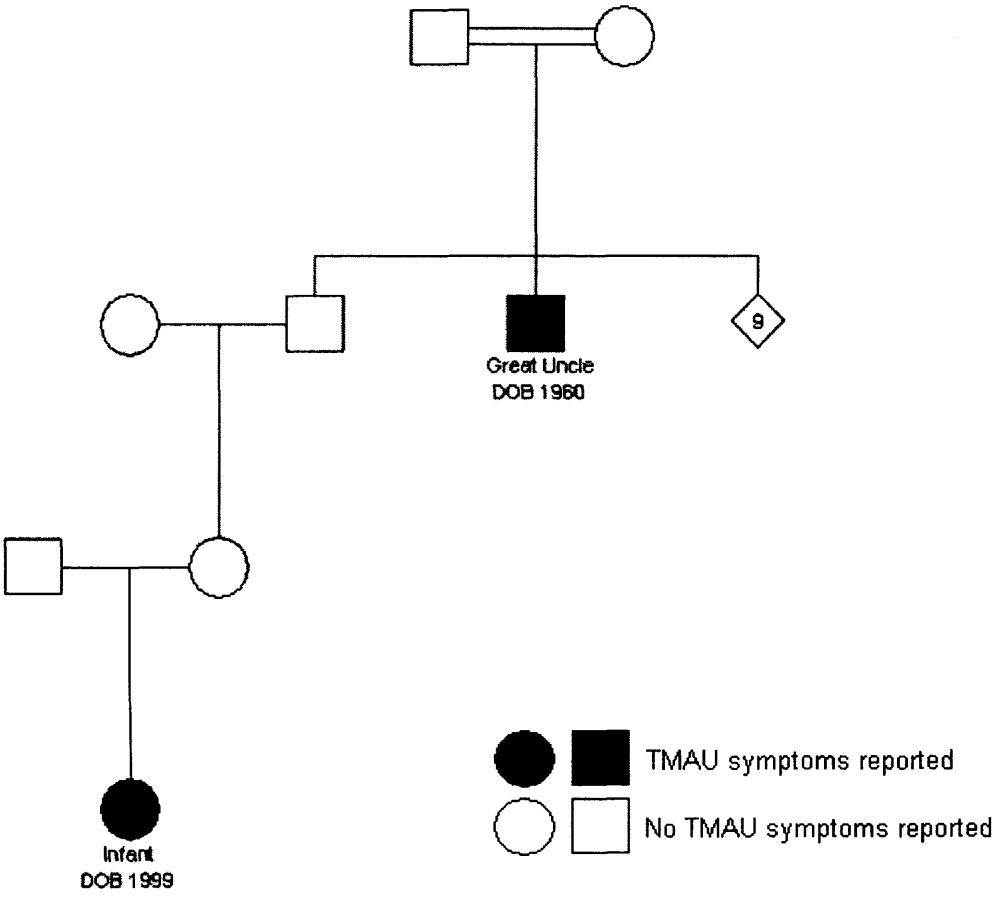


Figure 14. A genetic pedigree of the Norwegian family in this study

The great uncle, offspring of consanguineous relationship (first cousins), is homozygous for the 713G>A (R238Q) polymorphism and displays TMAU symptoms. Subject one displays TMAU symptoms despite being merely heterozygous for the 713G>A (R238Q) polymorphism. The mother does not report TMAU symptoms despite being genotypically identical to her daughter at the *FMO3* locus.

713G>A (R238Q) which reportedly causes a TMAU phenotype. As mentioned above, this represents the second reported case of TMAU caused by the mutation of R238, and the first time two different mutations have been implicated in TMAU at the same nucleotide leading to two separate amino acid changes. Previously, Teresa et al., (2006) [56] had reported an Italian cohort that featured a TMAU patient with the R238P amino acid change, resulting from a 713G>C mutation. The importance of this highly conserved residue is thus highlighted further. Possible structural changes associated with the residue are discussed later (section 4) as are the catalytic consequences of the variant (section 3.2.).

Subject two - Northern Irish Patient one

The lack of mutations in the *FMO3* gene of subject two (section 3.1) means that primary TMAU is an unlikely diagnosis. Subject two may have Secondary Trimethylaminuria, which can be transient and so may disappear with age [75]. This is unlikely however as reported cases of this usually manifest in children with heterozygous variations in *FMO3*. It is possible that the child has a problem involving gut flora, rendering digestion of TMA rich foods more difficult, leading to a TMAU-like phenotype. Current treatment regimes include dietary restrictions to cut out choline (a TMA precursor) and TMA rich foods and treatment with antibiotics such as Neomycin and Metronidazole, which has previously proved moderately successful in TMAU patients [232, 233]. The development of antibiotic resistance has been suggested as a problem which may be overcome by bi-weekly therapy of different antibiotics [233]. Management of TMA production and sequestration of free TMA by consuming the dietary supplement chlorophyllin-copper complex is a suggested alternative therapy to alleviate symptoms of TMAU [47, 85]. Individual patients report mixed responses to these different treatments. It has also been suggested that the use of

soaps with lower pH (5.5-6.5) could be used to try and neutralise the odorous TMA from the skin surface.

Subject three - Northern Irish Patient two

The DNA analysis indicates that the patient is homozygous for the 913G>T (E305X) polymorphism (section 3.1.) (Figure 10). This polymorphism encodes a truncated form of FMO3 which does not include the final 227 amino acids of the 532 amino acid molecule. This stretch of the enzyme is predicted to contain important residues involving linkage between NADPH and active site domains, a binding site for FAD flavin moiety, the FAD domain side of a substrate cleft-isoalloxazine ring, an active site to active site linkage of monomers, a dimer interface (important only if FMO3 exists as a dimer, it is not known in what state the enzyme exists *in vivo*) and active site of the substrate cleft and a predicted membrane anchoring domain [40]. The loss of such protein domains may be academic as a truncation of this size would most probably have destructive consequences in terms of protein folding and protein localisation.

The 913G>T polymorphism has been reported before [73] in patients displaying TMA *N*-oxide/ TMA ratios indicating Primary Trimethylaminuria (Table 7). The polymorphism was found in a homozygous state in a single individual (Family 2 in the paper) who, although having TMA concentrations close to that of the control level, had extremely elevated TMANO/TMA ratios. It has been shown *in vitro* that the E305X truncation totally destroys the catalytic activity of FMO3 in terms of *N*-oxidation of TMA [73].

Where this polymorphism was found in the other individuals it was found in a heterozygous state along with 551C>T (P153L) which has been shown to destroy FMO3 catalytic activity in terms of *N*-oxidation [62, 67, 73]. Furthermore, a polymorphism has been reported

Table 7. - Clinical characterization of index cases with TMAU (adapted from Treacy et al., (1998) [73])

Case	Ancestry	Age at DX (years)	TMA[a]	TMANO/TMA[b]	Genotype
1	English-Irish	15	60.4 ± 11.7	21.9/78.1	P153L\E305X
2	English-Irish	14	19.7 ± 2.8	11.8/88.2	E305X\E305X
5	English-Irish German-Spanish	58	501.2 ± 280.3	NT	P153L\E305X
6	English-Irish	4	40.2 ± 7.8	NT	P153L\E305X
7	English-Scottish	18	48.1 ± 7.5	NT	P153L\E305X

[a]The normal level is <18 μmol/mmol creatinine.

[b]The normal ratio is >97:3.

NT, not tested.

DX, Diagnosis

leading to a premature termination codon only nine residues downstream from E305X, E314X (940G>T) which has also been shown *in vitro* to destroy catalytic activity in terms of *N*-oxidation [62].

At this stage there is no genotypic evidence to suggest Secondary Trimethylaminuria and the presence of a mutation in a homozygous state, previously shown to destroy FMO3 catalysis of *N*-oxidation, indicate that a Primary Trimethylaminuria phenotype would be expected.

Subject four – Sheffield, England subject.

Subject four was found to be heterozygous for the1475G>A (R492Q) variant (section 3.1.). Thus, two different mutations have been implicated in TMAU leading to two separate amino acid changes, the other being 1474C>T (R492W). This is the first time two different mutations have been implicated in TMAU at the same codon, but not due to a mutation at the same nucleotide, leading to two separate amino acid changes

Previously the R492W has been reported in a compound heterozygote TMAU patient. Akerman et al., (1999) [62] first described the 1474C>T mutation (leading to a R492W amino acid change) in a compound heterozygote with another novel mutation, 198G>T (M66I). Akerman et al., (1999) [58] also investigated an 8 year old French Canadian girl displaying a classic TMAU biochemical phenotype, who was a compound heterozygote for the R492W mutation along with a P153L (458C>T) mutation. This latter mutation was shown to segregate with TMAU [67] and has been shown to destroy the activity of FMO3 in a recombinant protein. Dolphin et al., (2000) [35] investigated a patient with compound heterozygosity for the R492W and a previously unreported mutation, M434I (1302G>A). Both mutations were generated individually in a baculovirus-mediated insect cell protein expression system and assayed for activity. The

M434I mutation was found to severely decrease FMO3 activity and the R492W mutation was found to destroy enzyme function completely.

It is probable that the structural change produced by the R492W substitution destroys (or greatly reduces) the protein activity, due to the proximity of the amino acid 492 to the predicted active site of the substrate cleft [40]. As subject four is heterozygous for this mutation, it is possible that TMAU will prove transient. The subject was 4 years old when she was first examined by the GP. Transient TMAU has been reported previously with the TAMU phenotype disappearing completely in these cases as the child reaches 7-8 years old [75]. Here, the transient TMAU phenotype was postulated to have been caused by "...a transient overproduction in the gut or an overloading of the gut-generated substrate overwhelming the hepatic enzymes' [FMO3] oxidizing capacity." However, it is more probable that rather than being the "cause", the gut-overload was compounded because the child had only a single *FMO3* allele able to generate a functional FMO3 enzyme.

Because this subject possesses one functional *FMO3* gene, it would be reasonable to predict that the TMAU phenotype will be a transient one, and once adult expression of FMO3 is established, the phenotype would be predicted to dissipate (discussed in section 1.2.).

3.2. Kinetic parameters of FMO3 polymorphic variants

Introduction

To assess the catalytic consequences of FMO3 amino acid variation, kinetic parameters were assessed using an established assay for FMO activity (described in section 2.2.7.). Of particular interest were two double amino acid variants, E158K/V257M and E158K/E308G. It has been shown previously that in combination with E158K, the E308G variant was catalytically impaired, compared with either variant present alone. Also, the R238Q variant discovered in the TMAU Norwegian cohort discussed earlier (section 3.1.) was also assayed for FMO3 activity.

Kinetic parameters of FMO3 polymorphic variants - Results

Variant FMO3 cDNAs were created as described in section 2.1.1. Analysis of the DNA sequencing traces, for both strands, confirmed the presence of the mutation in the cDNA on the E158 (Figure 15) and K158 (Figure 16) backgrounds. The DNA traces were also used to confirm the absence of undesired mutations that might have been inadvertently introduced, that would cause premature stop codons or unwanted amino acid variants.

All FMO3 variants were readily expressed in bacteria according to the protocol detailed in section 2.2.3.1. Only cell lysis method VII (detailed in section 2.2.3.3.7.) was amenable to maintaining catalytically active FMO3 during the cell protein preparation protocol.

FMO3 activity was assessed by the Methimazole assay, detailed in section 2.2.7. A hyperbolic Michaelis-Menton relationship between enzyme rate and substrate concentration was observed for all variants of FMO3 tested, except the R238Q form of the enzyme which was catalytically inactive. A Hanes-Woolf linear transformation of the Michaelis-Menton

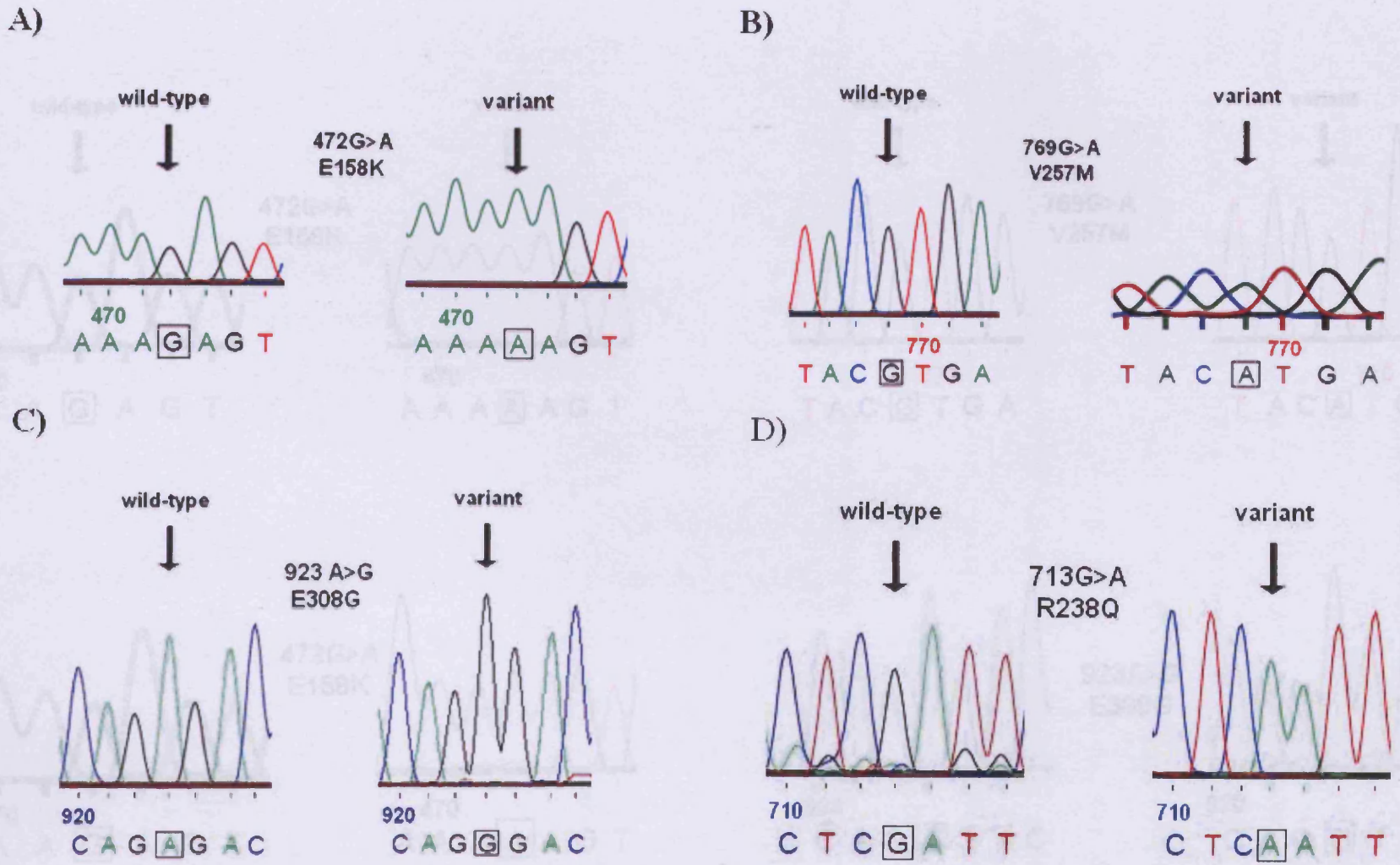


Figure 15. DNA sequencing traces of FMO3 cDNA on the E158 background
Sequencing traces of FMO3 cDNA generated via site-directed mutagenesis on the E158 background are shown. A) E158K, B) V257M, C) E308G and D) R238Q.

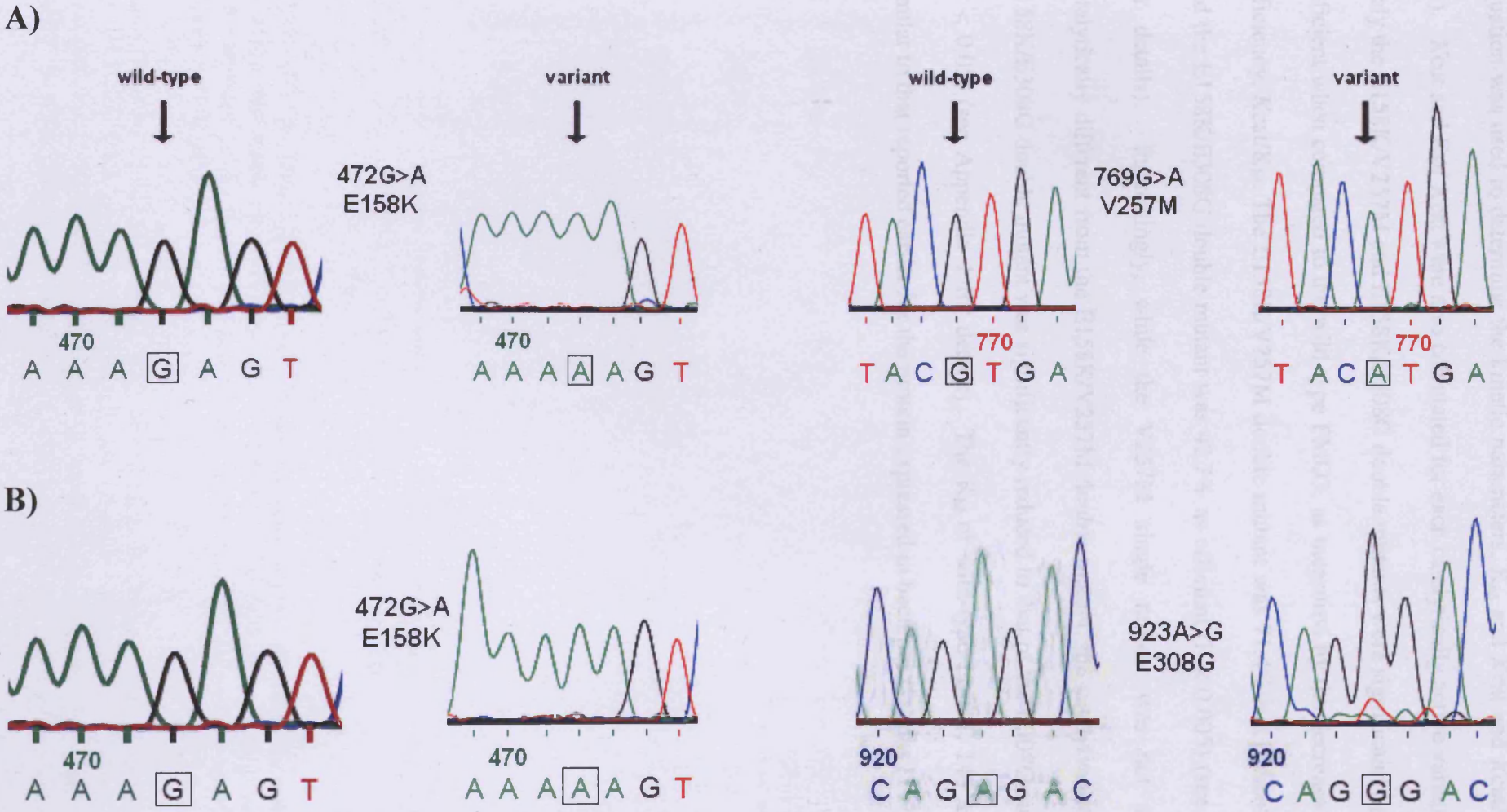


Figure 16. DNA sequencing traces of FMO3 cDNA variants on the K158 background
Sequencing traces of FMO3 cDNA generated via site directed mutagenesis on the K158 background are shown. A) The E158K/V257M double variant and B) the E158K/E308G double variant.

equation was used to determine the kinetic parameters, K_M and *K*cat and *K*cat/K_M (Figure 17). *K*cat and K_M/ *K*cat were also calculated for each catalytically active variant (Table 8). Only the E158K/V257M and E158K/E308G double mutants were significantly catalytically deficient when compared to the wild-type FMO3, as measured by the decrease in catalytic efficiency, *K*cat/K_M. The E158K/V257M double mutant was 71.4% as efficient ($P < 0.05$) and the E158K/E308G double mutant was 42.7% as efficient ($P < 0.005$) (see appendix 1 for details). Interestingly, while the V257M single mutant was not significantly catalytically different from the E158K/V257M double mutant, the catalytic activity of the E158K/E308G double mutant was significantly reduced to that of the E308G single mutant ($P < 0.05$) (see Appendix 1 for details). The K_M of wild-type FMO3, 143 ± 17 μM, is similar to that reported earlier for the protein expressed in bacterial systems [197].

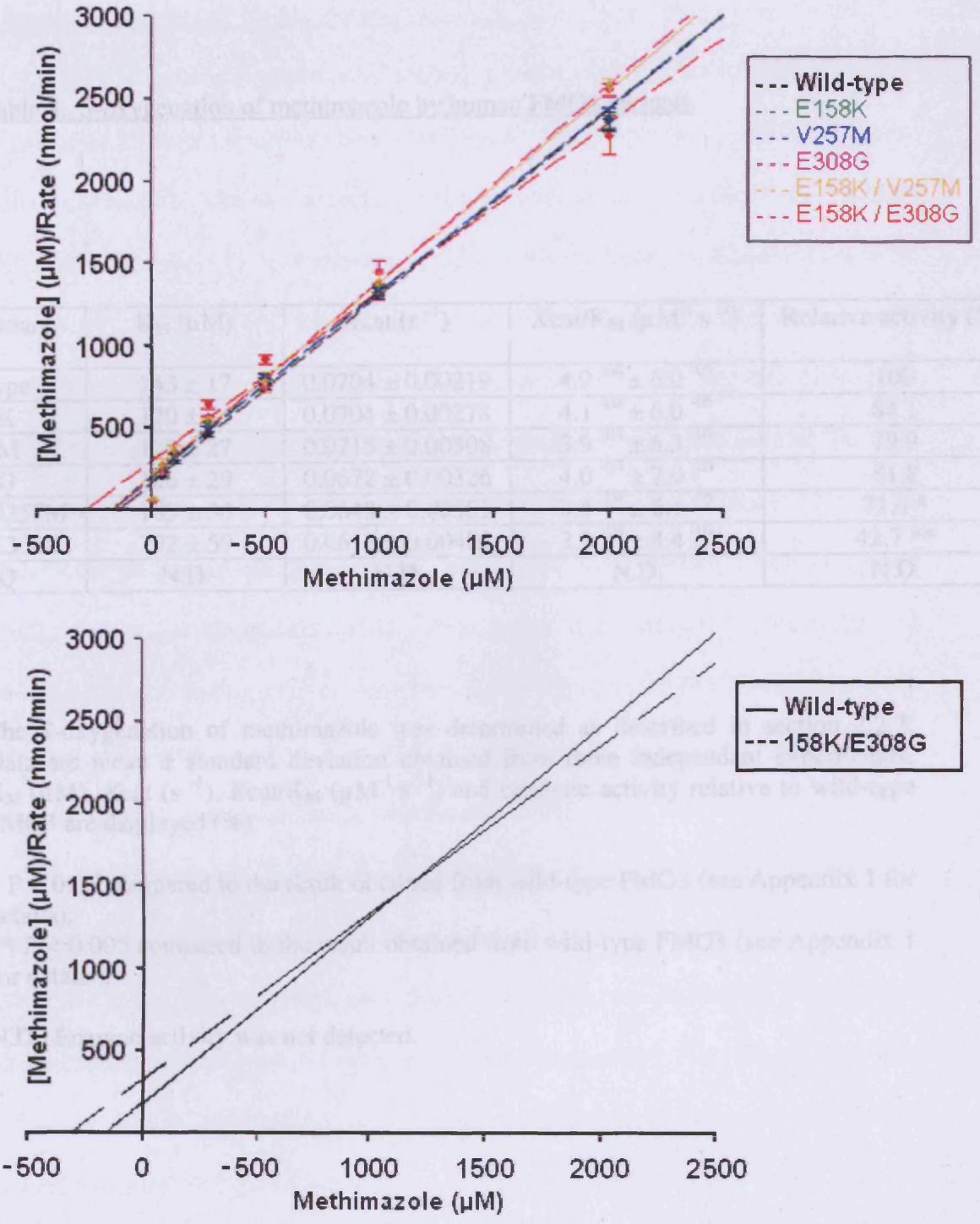


Figure 17. A Hanes-Woolf linear transformation plot of the Michaelis-Menton equation for each catalytically active variant of FMO3, measuring methimazole concentration (µM)/ Assay product formed (nmol/min) as a function of Methimazole concentration (µM)

Wild-type (black), E158K (Green), V257M (blue), E308G (pink), E158K/V257M (orange) and E158K/E308G (red) FMO3 variants are plotted. A plot showing ony wild-type (unbroken line) and E158K/E308G (dashed line) is also presented for clarity. All data display a good linearity, indicative of obeying classical Michaelis-Menton kinetics.

Table 8. *S*-oxygenation of methimazole by human FMO3 variants

FMO3 variant	K_M (μM)	*K*cat (s^{-1})	*K*cat/K_M ($\mu M^{-1} s^{-1}$)	Relative activity (%)
Wild type	143 ± 17	0.0704 ± 0.00219	$4.9^{-04} \pm 6.0^{-05}$	100
E158K	170 ± 24	0.0704 ± 0.00278	$4.1^{-04} \pm 6.0^{-05}$	84.1
V257M	179 ± 27	0.0715 ± 0.00308	$3.9^{-04} \pm 6.3^{-05}$	79.9
E308G	166 ± 29	0.0672 ± 0.00326	$4.0^{-04} \pm 7.0^{-05}$	81.8
E158K / V257M	183 ± 30	0.0643 ± 0.00301	$3.5^{-04} \pm 6.1^{-05}$	71.4 *
E158K / E308G	292 ± 59	0.0615 ± 0.00402	$2.1^{-04} \pm 4.4^{-05}$	42.7 **
R238Q	N.D.	N.D.	N.D.	N.D.

The *S*-oxygenation of methimazole was determined as described in section 2.2.7. Data are mean ± standard deviation obtained from three independent experiments. K_M (μM), *K*cat (s^{-1}), *K*cat/K_M ($\mu M^{-1} s^{-1}$) and catalytic activity relative to wild-type FMO3 are displayed (%).

* $P < 0.05$ compared to the result obtained from wild-type FMO3 (see Appendix 1 for details).
** $P < 0.005$ compared to the result obtained from wild-type FMO3 (see Appendix 1 for details).

N.D. Enzyme activity was not detected.

Kinetic parameters of FMO3 polymorphic variants - Discussion

The impact of pharmacogenetics of monooxygenases on the drug metabolism and distribution of certain drugs has been established. Variant alleles of *CYP2C8*3* have been implicated in altered oxidation of the mitotic inhibitor paclitaxel [234, 235]. Patients carrying amino acid variants of CYP2C9 have been implicated in warfarin sensitivity, e.g. CYP2C9*2 R144C and CYP2C9*3 I359L, requiring lower doses of warfarin [236, 237] due to lowered catalytic activity of the enzyme [238-240]. CYP2D6 polymorphism has been implicated in the impaired metabolism of the anti-hypertensive debrisoquine [241-243] and the anti-convulsant mephenytoin [244, 245]. The allelic frequency of *CYP* variants vary widely from population to population [246], which is important when considering the clinical significance of pharmacogenetic studies. Given the extremely broad range of FMO3 substrates, any frequent polymorphisms that alter FMO3 catalytic efficiency would be relevant to pharmacogenetic considerations for therapeutic strategies.

E158K is the most common FMO3 polymorphism with an allelic frequency ranging from 39-50% depending on the population sampled. [65, 66, 70, 87, 143, 147, 222, 247-251]. Heterozygotes have shown no significant increase in free TMA in urine [70]. *In vitro* studies by other groups have also shown the lack of significant effect on FMO3 catalysis by this common polymorphism, as confirmed in this investigation (section 3.2.) [248, 252].

The allelic frequency of V257M has been reported to range from 3-25% depending on the ethnicity of the population [35, 142, 222, 247-249, 253, 254]. The lack of significant reduction in recombinant FMO3 enzyme efficiency due to the V257M polymorphism presented in this investigation (section 3.2.) has also been

demonstrated by other groups [35, 255]. Another group has also reported a lack of significant reduction, but in a substrate specific manner [256].

E308G has been shown to be a relatively common polymorphism with an allelic frequency of 6-26% depending on the population sampled [62, 70, 147, 248-251, 257]. *In vivo* [249, 258] and *in vitro* [255, 259] methods have shown the variant to have little consequence catalytically to FMO3 efficiency and this is supported by the work presented in this investigation with an insignificant reduction in enzyme efficiency, as a measure of Kcat/K_M.

Little data on the frequency of E158K and V257M being found in *cis* exists. E158K and V257M were found in *cis* at an allelic frequency of 4-6% in a Han Chinese population [254]. In a different study, E158K/V257M was found at an allelic frequency of 1% in a study of Hispanic Americans, but was not present in any other population in the study (Caucasian American, African American and Asian American) [248]. Surprisingly, this E158K/V257M double mutant was found with the E308G mutation in *cis* also, with the other allele of the individual unknown. Because the source of the DNA was a blood bank, it was not possible to test the individual for symptoms of TMAU and, to date, this E158K/V257M/E308G triple mutant has not been assayed and it may be of interest to see what the catalytic consequence would be to FMO3. This investigation suggests that the E158K/V257M double mutant does cause a significant reduction in FMO3 catalytic efficiency ($P < 0.05$) with efficiency dropping to 71.48% of that of the wild-type enzyme. It is interesting to note that the E158K/V257M double mutant has relatively similar catalytic parameters, as does the single V257M mutation on its own.

E158K and E308G occur in *cis* at an allelic frequency ranging from 1-16% depending on the source population [248, 250, 254]. The E158K/E308G double

variant when found in *cis* (featured earlier in the Norwegian cohort, section 3.2.) has previously been shown to reduce FMO3 activity by around 50% *in vivo* (as a measure of free TMA in urine samples compared with control subjects) [58, 70]. The activity of *in vitro* expressed FMO3 have shown a dramatic reduction in catalytic efficiency of the E158K/E308G variant when compared to either the E158K or E308G single variants [21, 143, 147]. The kinetic study presented as part of this investigation supports a reduced catalytic efficiency in the E158K/E308G double mutant of 42.7% compared to wild-type FMO3, which represents a significant reduction ($P < 0.005$). That the E158K/E308G double mutant significantly disrupts FMO3 catalysis, compared with either the E158K or E308G single mutants, is made all the more interesting when we consider that the E158K/V257M double mutant has a relatively similar catalytic activity to either of the single mutations alone. FMO3-substrate metabolism shown to be affected by the reduction in catalytic efficiency of the E158K/E308G variant occurring in *cis* includes the histamine H_2-receptor antagonist ranitidine [147], the anti-tubercular agent thiobenzamide [147] and the non-narcotic, non-steroidal anti-inflammatory drug, sulindac [5]. Interestingly, in the latter case, the E158K/E308G variants were found to have a protective effect on the development of polyps in familial adenomatous polyposis patients who received sulindac as a means of primary chemoprevention, presumably reducing the ability of the enzyme to inactivate the drug, leading to prolonged exposure to the active form of sulindac, sulindac sulphide [260, 261]. Therefore, the relationship between the E158K and E308G variants has the potential to be important from a pharmacogentic standpoint when considering FMO3 substrates as therapeutic treatments.

The finding that the R238Q variant of FMO3 was catalytically inactive is consistent with the finding in this investigation that individuals with a *FMO3* allele

containing this variant have compromised TMA metabolism. In chapter 3, a case study was presented of a Norwegian cohort that was screened for TMAU-associated mutations at the *FMO3* locus (see section 3.2. for details). A child showing symptoms of TMAU was found to be heterozygous for the 713G>A (R238Q) mutation. Her great uncle was found to be homozygous for this mutation and consequently received genetic counselling after being diagnosed as having TMAU. With no other mutations present, R238Q was suggested as the cause of the TMAU phenotype. Evidence of importance of arginine at amino acid 238 for FMO3 catalysis is also presented by Teresa et al., (2006) who describe the mutation 713 G>C (R238P) in an Italian TMAU cohort with a TMAU phenotype [56].

3.3. Generating a homology model of FMO3

Introduction

To theorise about the structural consequences of the variants of FMO3, homology modelling of FMO3 was performed.

Modelling FMO3

The two closest homologues to human FMO3 in the PDB at present, in terms of identity at the primary sequence level, are the crystal structure of Baeyer-Villiger Monooxygenase from the moderate thermophilic bacterium, *Thermobifida fusca* (*T.fusca*) [215] (26% identity) at 2.40 Å resolution (PDB accession code 1W4X) and the crystal structure of FMO from *Schizosaccharomyces pombe* (*S.pombe*) [214] (29% identity) at 1.70 Å resolution (PDB accession code 1VQW) (see Figure 18 for ClustalW primary sequence alignments). The next nearest homologues are of extremely low sequence identity (<15%).

Using the homology modelling software 'Modeller' it was possible to generate models of FMO3 from the crystal structure templates 1W4X and 1VQW.

Homology model of FMO3 using 1VQW as a template

The crystal structure 1VQW was co-crystallised with ligands and the coenzyme FAD. Using a python script within the suite of scripts that make up Modeller, it was possible to model this coenzyme along with the rest of the target from the template.

The model displays two main bundles of secondary structure (Figure 19). The larger of the two seems to be where the FAD and NADPH molecules are predicted to bind. It consists of a bundle of six α-helixes, one, four-stranded parallel β-sheet and

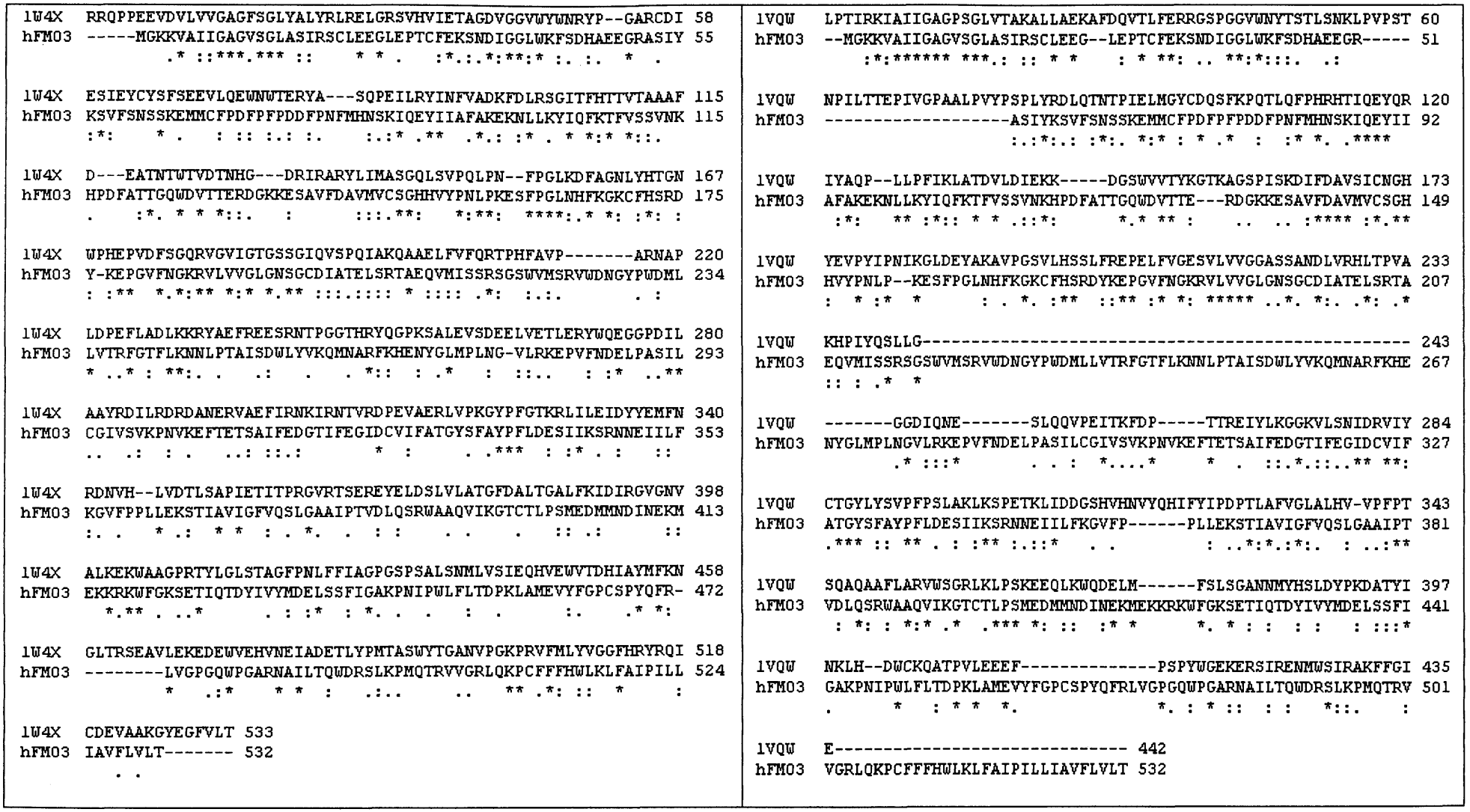


Figure 18. ClustalW alignments of 1W4X vs hFMO3 and 1VQW vs hFMO3.

An asterix denotes conserved identical residues, a colon indicates conserved similar residues, and a dot indicates highly similar residues.

Figure 19. A cartoon representation of the homology model, based on 1VQW, of FMO3

The model displays two main bundles of secondary structure. The larger of the two seems to be where the FAD (shown in blue) and NADPH (not shown) molecules are predicted to bind. It consists of a bundle of six α-helixes (red), one, four-stranded parallel β-sheet (yellow) and one, three stranded anti-parallel β-sheet. In the model, the flavin moiety is bound within the α-helix bundle, with the adenine dinucleotide moiety of the molecule penetrating deep into the protein, poking between an anti-parallel β-sheet and with the tip of the moiety residing near a parallel β-sheet.

one, three stranded anti-parallel β-sheet. In the model, the flavin moiety is bound

within the α-helix bundle, with the adenine dinuceotide moiety of the molecule

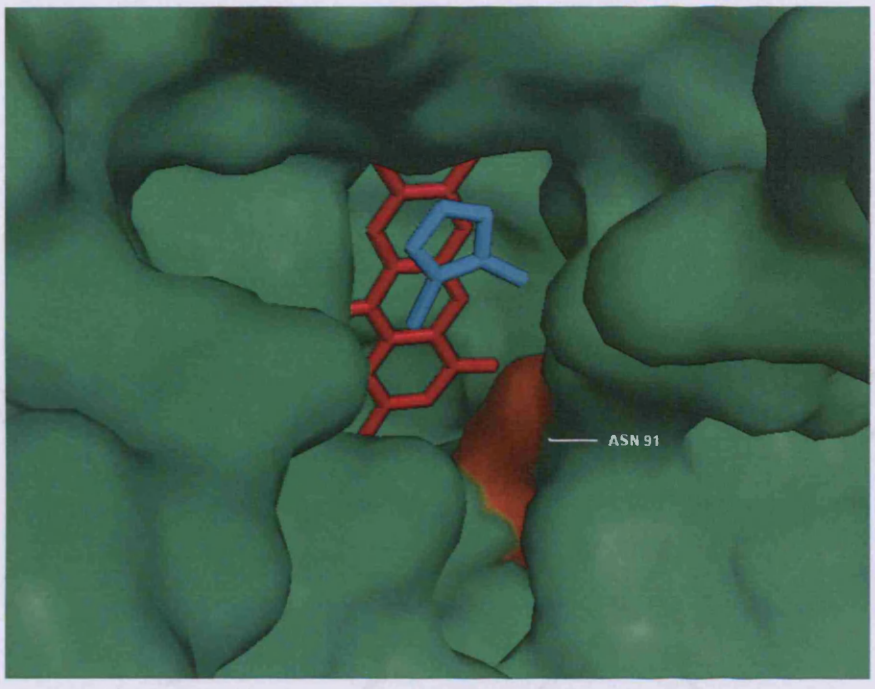
penetrating deep into the protein, poking between an anti-parallel β-sheet and with the tip of the moiety residing near a parallel β-sheet.

The crystal structure of 1VQW [214] have helped to give further insights into the mechanism of FMO action. The correct orientation of the isoalloxazine ring of the FAD molecule is predicted to be vital in order for it to act as an efficient prosthetic group, namely to accept a hydride ion from NADPH, accept molecular oxygen and for oxygenation of the substrate (Figure 20 A). The flavin moiety is visible from the surface image of the protein, residing in a pocket, readily accessible and solvent-exposed. The orientation of the FAD molecule within the FMO3 model was examined and shown to be positioned in a reasonable orientation (Figure 20 B) with the Asn 61 residue in close proximity to the FAD isoalloxazine ring. This residue is equivalent to Asn 91 in the 1VQW structure, which has been shown to be a key residue in binding molecular oxygen, and is thought to be the only FMO residue directly involved in substrate catalysis.

In the FMO3 model, the adenine dinucleotide moiety of FAD is buried deep within the protein (Figure 21 A). There are several residues known to be involved in TMAU, where FMO3 catalysis is severely disrupted or destroyed, that reside proximal to the modelled FAD molecule within the FMO3 model, namely E32K, A52T, V58I and N61S (Figure 21 B). The residues predicted to interact with the FAD molecule within the FMO3 homology model were investigated further using MBT Ligand Explorer (Figure 22). Hydrophobic and hydrophilic interactions were predicted, measured and tabulated (Table 9).

A)

A)



B)

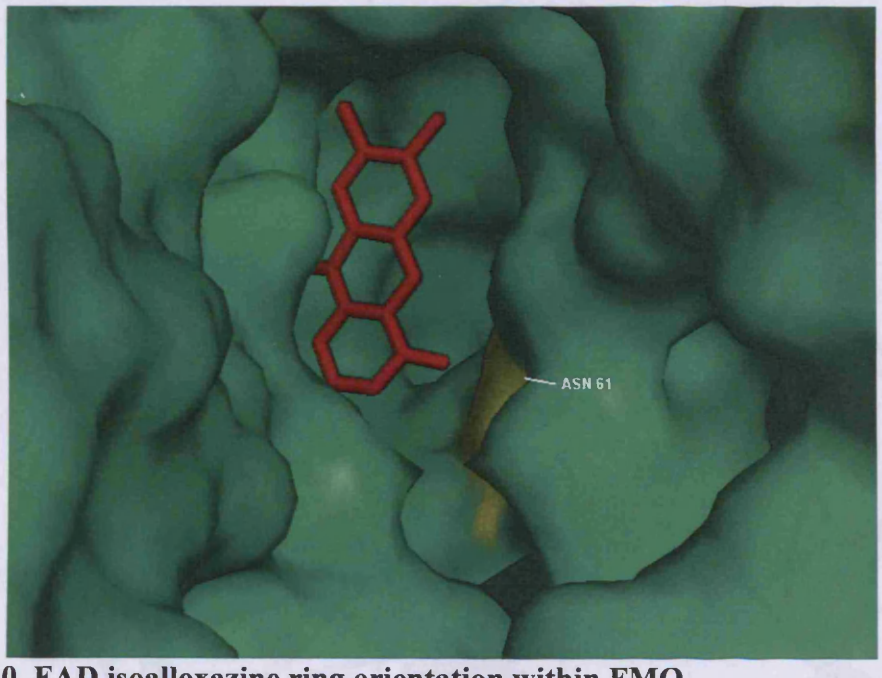


Figure 20. FAD isoalloxazine ring orientation within FMO

A) The isoalloxazine ring of the FAD molecule within the active site of *S.pombe* FMO solved crystal structure (1VQW) is shown (red) along with a molecule of FMO substrate Methimazole (blue). Asn 91 is highlighted in orange. Asn 91 in 1VQW has been shown to be a key residue in binding molecular oxygen and is thought to be the only FMO residue directly involved in substrate catalysis.
B) The isoalloxazine ring of the modelled FAD molecule within the FMO3 model based on 1VQW is shown (red). The Asn 61 residue is highlighted in yellow, which is the equivalent residue to Asn91 in 1VQW.

A)

B)

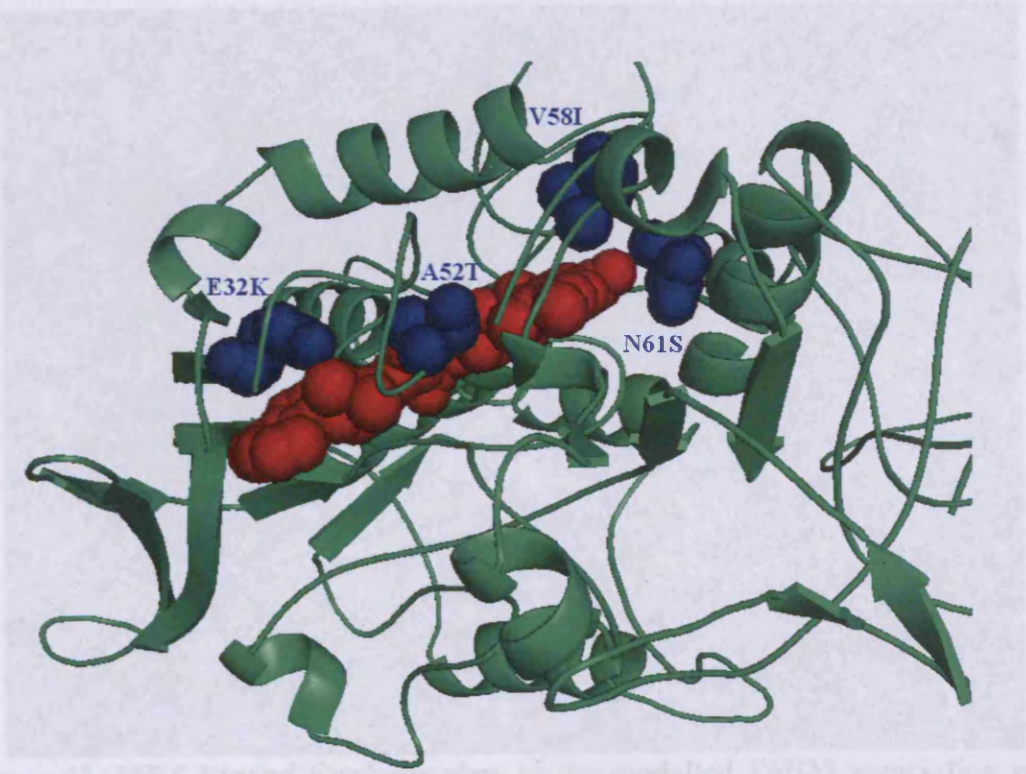
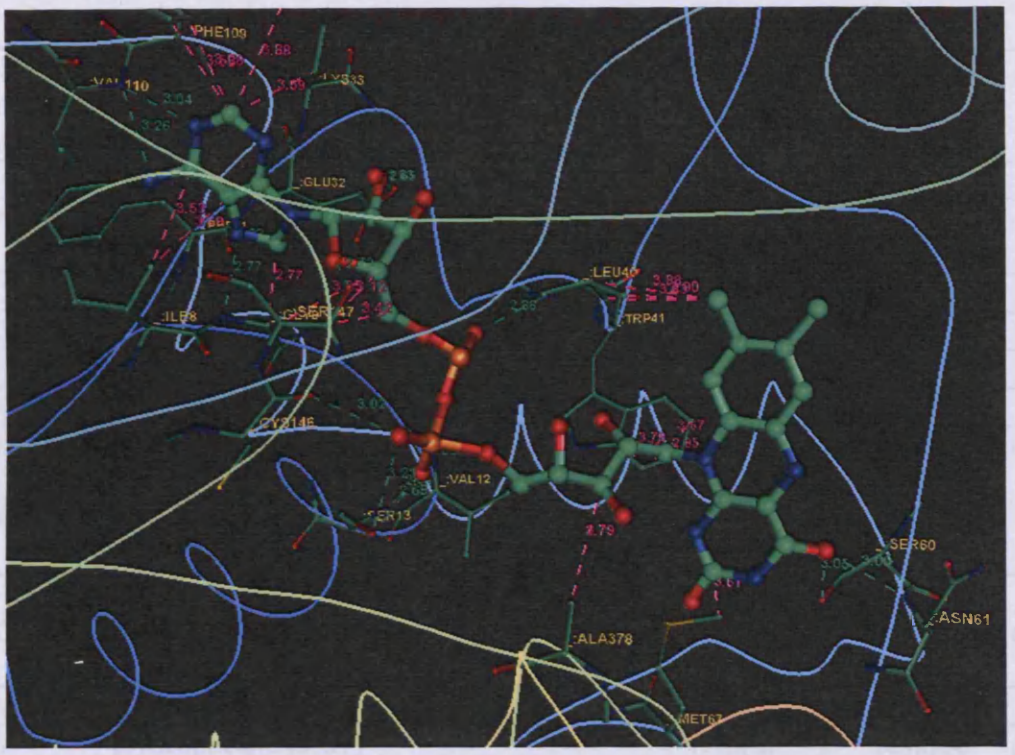


Figure 21. FAD modelling within the homology Model of FMO3 based on 1VQW

A) A cartoon representation of FMO3 (green) with FAD represented as red spheres. B) A close up view of FAD within FMO3, with the local amino acid residues known to be associated with TMAU represented by blue spheres.

A)



B)

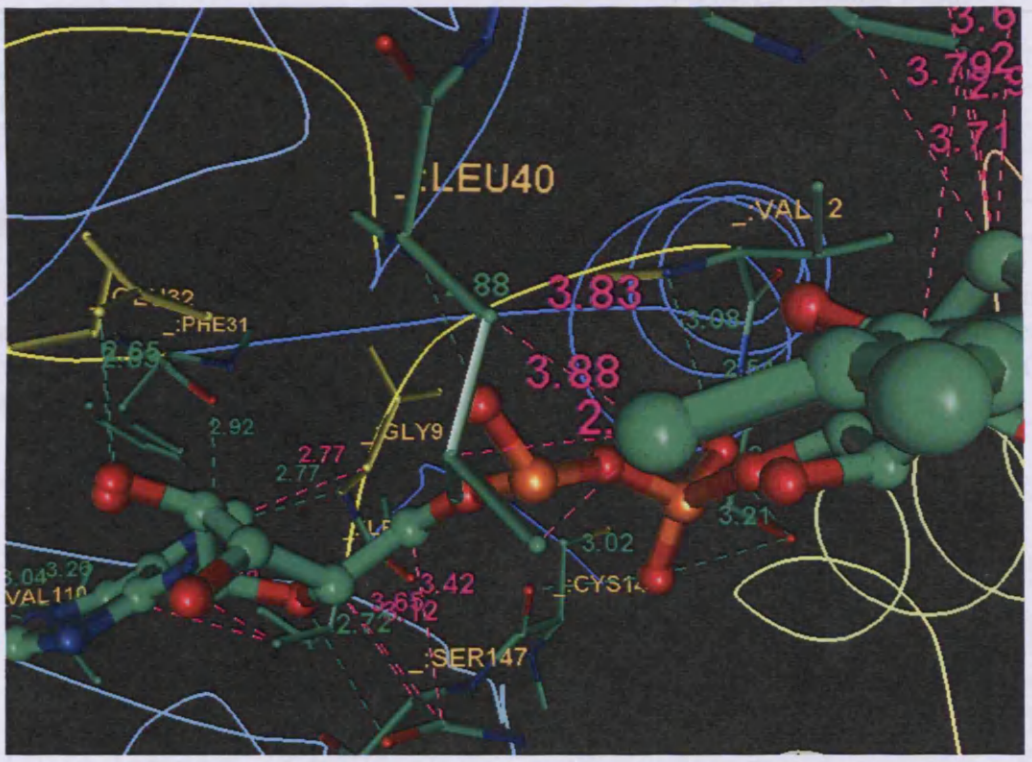


Figure 22. MBT Ligand Explorer view of the modelled FMO3 association with FAD

A) A cartoon representation of FMO3 with FAD shown as ball-and-sticks. FMO3 residues predicted to interact with FAD are labelled and their hydrophobic (purple) and hydrophilic (green) interactions are shown as dashed lines along with their distance, measured in Angstroms, which is labelled.

B) A close up view.

Table 9. A tabulation of all interactions predicted from the model of FMO3 along with an incorporated FAD molecule, by MBT Ligand Explorer

FMO3 amino acid residue	Interaction(s) (distance (Å))	Catalytic effect reported at residue
Ile8	Hydrophobic (3.48, 3.53)	
Gly 9	Hydrophilic (2.77) Hydrophobic (2.77)	
Val 12	Hydrophilic (2.54, 3.08)	
Ser 13	Hydrophilic (2.68, 3.21)	
Phe 31	Hydrophilic (2.92)	
Glu 32	Hydrophilic (2.65, 2.83)	TMAU E32K [55]
Lys 33	Hydrophobic (3.59, 3.88)	Within 1 residue of TMAU causing E32K
Leu 40	Hydrophobic (2.90, 3.83, 3.88) Hydrophilic (2.88)	
Trp 41	Hydrophobic (2.85, 2.92, 3.67, 3.72, 3.79)	
Ser 60	Hydrophilic (3.05)	Within 1 residue of TMAU causing N61S
Asn 61	Hydrophilic (3.0)	TMAU N61S [35, 60]
Met 67	Hydrophobic (3.61)	Within 1 residue of TMAU causing M66I
Phe 109	Hydrophobic (3.08, 3.56)	
Val 110	Hydrophilic (3.04, 3.26)	
Cys 146	Hydrophilic (3.02)	
Ser 147	Hydrophilic (2.72) Hydrophobic (3.12, 3.42, 3.65)	
Ala 378	Hydrophobic (3.79)	

The FMO3 residue is detailed along with any hydrophobic and/or hydrophilic interactions, with their distances measured in Angstroms shown in parentheses. Hydrophilic interactions are H-H measurements and hydrophobic interactions are H-C measurements. Also shown is any implication in TMAU of an amino acid change.

Seventeen FMO3 residues were predicted to contribute to the binding of FAD, including 2 of the residues associated with TMAU above, E32K and N61S. Met 67, which is one residue away from Met 66 and shown to be a causative mutation in TMAU when mutated to isoleucine [54, 61, 62], was shown to interact with the FAD molecule.

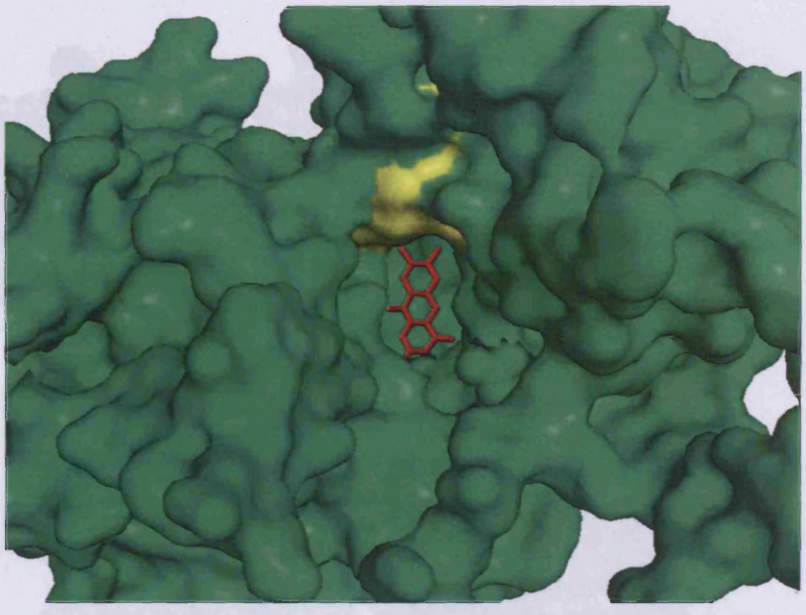
FMOs are known to have a hydrophobic 'FATGY' sequence motif common to most FMOs (it is 'FTTGY in FMO4'), and an 'identifying sequence' FXGXXXHXXXF. Despite being known for quite some time, the actual role of these two highly conserved motifs has remained elusive. Both motifs were mapped onto the FMO3 model to try to gain some insight into their possible roles. The FATGY motif was shown to map to a region around the rim of the pocket containing the putative active site of FMO3 (Figure 23 A). The FMO identifying sequence maps to the 'top' of the FMO3 model (Figure 23 B).

The catalytic site atlas did not identify any probable catalytic residues from its database of enzyme active sites, within the putative active site of the FMO3 model based on 1VQW.

Mutations and Polymorphisms of FMO3

In an attempt to try and understand possible insights that may be gleaned from a structural standpoint, the residues known to be variant in FMO3 were mapped on to the model (Figure 24). The variants included all residues known to destroy (or severely disrupt) FMO3 activity, those residues known to alter FMO3 catalytic activity and the two residues for which novel mutations were discovered in this investigation (see section 3.2.). The amino acid residues affected are shown in Figure 24 and the consequences for these changed are discussed later in this chapter.

A)



B)

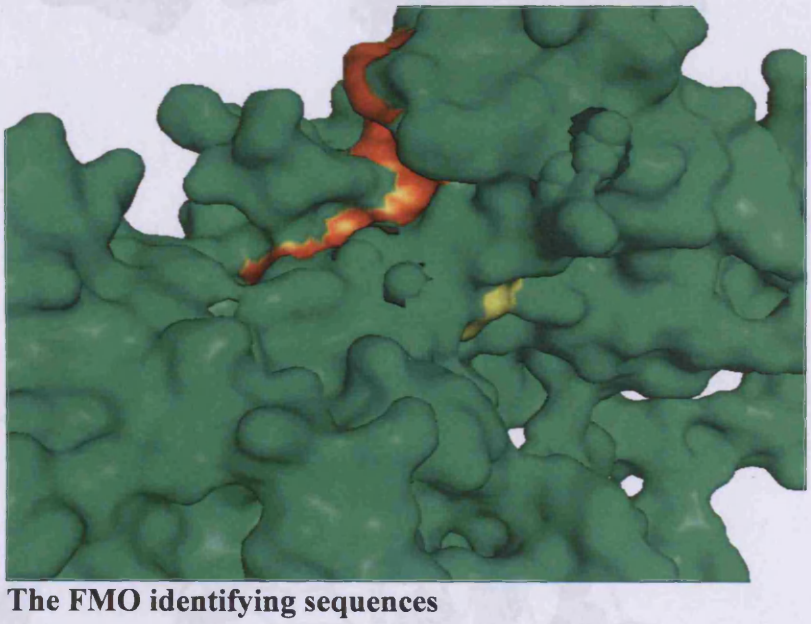


Figure 23. The FMO identifying sequences

FMOs are known to have a hydrophobic sequence motif common to all FMOs, 'FATGY' and an 'identifying sequence' FXGXXXHXXXF.
A) The FATGY motif is shown (yellow) to map to a region around the 'mouth' of the pocket containing the putative active site of FMO3. A FAD molecule can be seen within the pocket (red).
B) The FXGXXXHXXXF FMO identifying sequence (orange) maps to the 'top' of the FMO3 model, with the FATGY motif shown in yellow to emphasise the relative position of the two motifs.

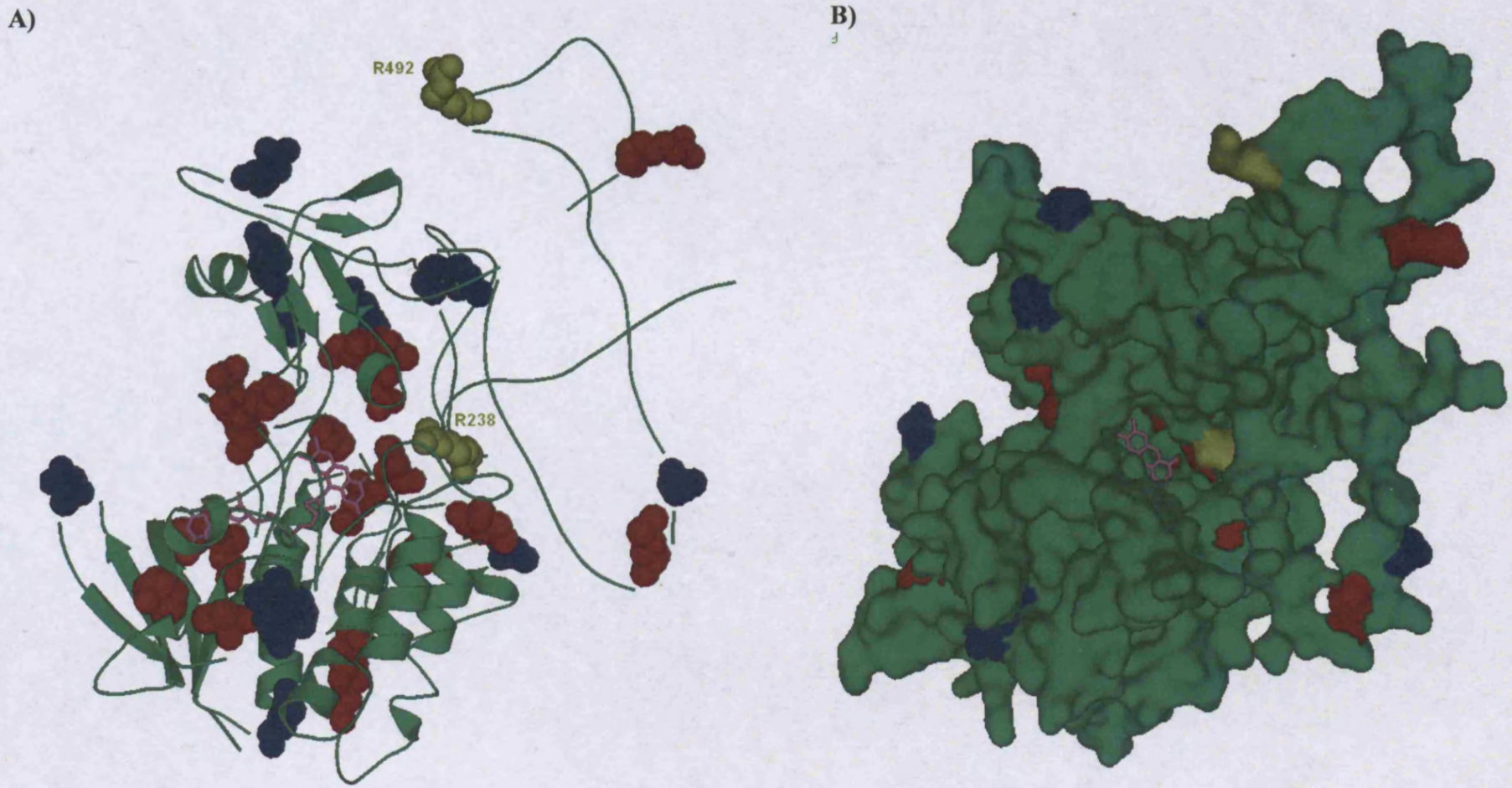


Figure 24. The homology model, based on 1VQW, of FMO3 with amino acid variants highlighted
The model is shown as a cartoon (A) and as a surface view (B) with residues known to be variant in FMO3 were mapped on to the model. Residues known to destroy (or severely disrupt) FMO3 activity are shown as red spheres, residues known to alter but not severely disrupt FMO3 catalytic activity are shown in blue and the two sites of allelic variation shown to destroy catalytic activity where two novel variants were discovered in this investigation (R238Q and R492Q) are shown as yellow spheres and labelled.

Model scoring FMO3 homology model based on 1VQW

The programme Verify 3D considers the 3D compatibility of the protein model to its 1D primary sequence. The residues in the 3D model are characterized by the environment surrounding it, defined by the area of the residue that is buried, the proportion of side-chain area covered by polar atoms and the local secondary structure. The likelihood of finding a specific amino acid in the environment defined is then scored, e.g. finding a charged residue buried in a non-polar environment would be assigned a poor score reflecting the improbable nature of such an occurrence [262]. Models that are largely wrong have a low overall score, whereas models that contain small numbers of improperly built segments can be detected by low scoring regions in the profile plot. A single region (residue 223 to residue 267) was defined as a low scoring region, and can be considered to be an improperly built segment of the FMO3 model based on 1VQW.

ProSA analyses the energy distribution in proteins as a function of sequence position, with native structures typically having a balanced energy distribution devoid of high energy peaks or 'strained' areas littering the protein, such features being typical of unlikely regions in terms of amino acid 3D position. The ProSA plot of residue scores identified two areas of concern in the model (Figure 25 A). One area was the C-terminal region and the other was the region around 223 - 267 that verify 3D also identified. Despite these regions of uncertainty, ProSA indicated that the model generated was of good overall quality with a z-score within the range of scores typically found for native proteins of a similar size with solved structures in the PDB (Figure 25 B).

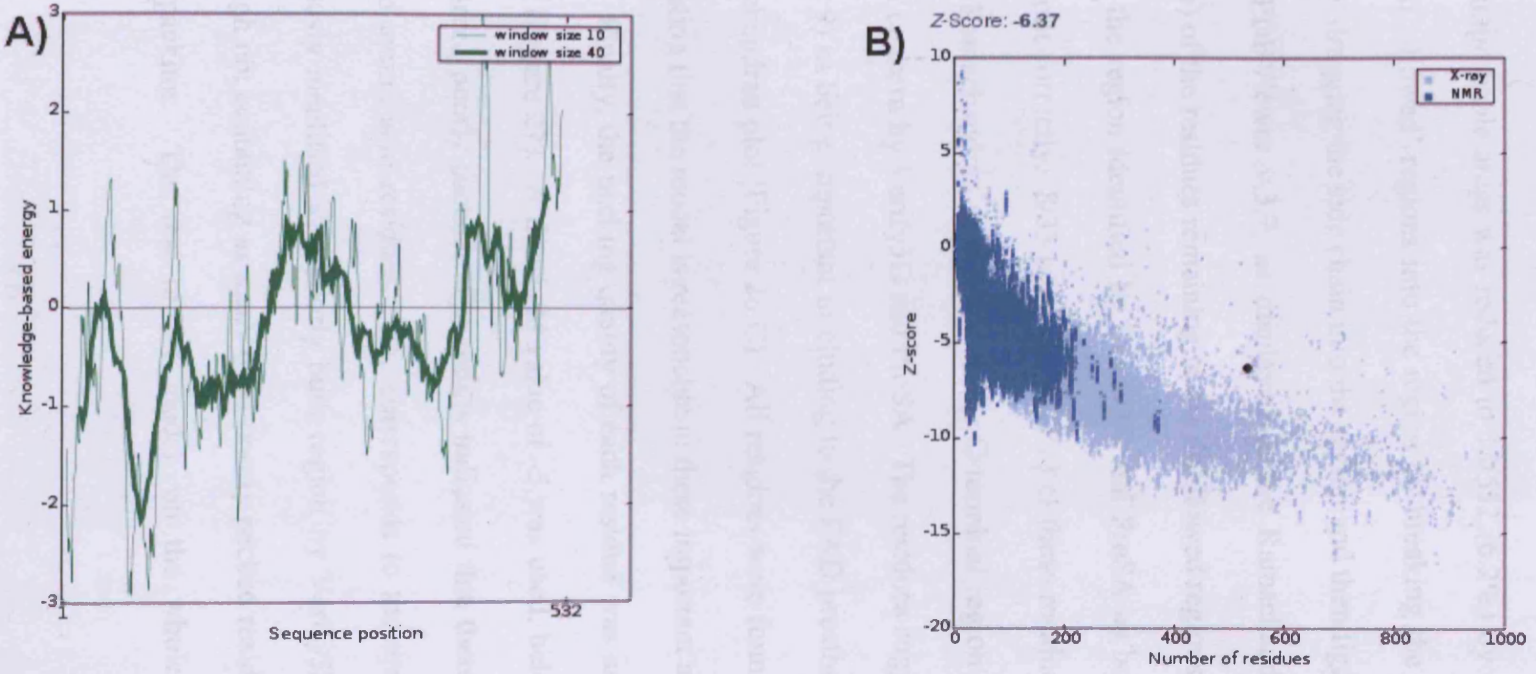


Figure 25. ProSA output for FMO3 model generated using 1VQW as a template

A) An energy distribution plot of FMO3. The thin green line considers the energy distribution over a 10-residue stretch, whereas the thick green line considers the energy distribution over a 40-residue stretch. The plot suggests two areas of concern with the model, indicated by peaks of positive energy, at the C-terminus and in the middle of the protein around 220-270 amino acids.
B) A plot of the *z*-scores of all solved structures currently in the PDB as a function of their length, solved either by NMR (dark blue) or X-ray crystallography (light blue). The position of the *z*-score of the FMO3 model as a function of the length of the model is indicated by a black dot. This shows that the model is generally within the range of the *z*-score expected for a molecule of that size and indicates that the model is of a general good quality.

The Ramachandran plot for the FMO3 model, generated using 1VQW as a template, showed 41/532 (7.7%) of amino acid residues in sterically improbable regions of the energy landscape (Figure 26 A). One measure of a good model is having more than 90% of the amino acide residues in the sterically 'allowed' regions [263]. The number of amino acids in the improbable areas was reduced to 33/532 (6.2%) by moving residue sidechains very close to 'allowed' regions into the region by breaking the αC-backbone either side of the residue, dragging the side chain into the region and then ligating the αC-backbone within Swiss-PDBViewer v.3.7, as displayed in the Ramachandran plot (Figure 26 B). 6/33 (18.1%) of the residues remaining in the disallowed regions of the Ramchandran plot reside within the region identified by Verify3D and ProSA as being a region that may not have been built correctly. 8/33 residues (24.2%) of these residues within the disallowed regions of the Ramachandran plot reside in the C-terminal region of the model, highlighted as an area of concern by Verify3D and ProSA. The residues suggested by MBT Ligand Explorer (Table 9) as being important in binding to the FAD prosthetic group were highlighted on a Ramachandran plot (Figure 26 C). All residues were found to be in the 'allowed' regions, suggesting that the model is reasonable in these important areas.

Finally, the packing quality of each residue was assessed using the WHATIF web server (Figure 27). A threshold value of -5 was used, below which level a residue can be considered poorly packed. The results indicated that there were two main areas of poorly packed amino acid residues. One corresponds to the area around amino acids 220-270, previously identified as a poorly built region by Verify3D and ProSA. The C-terminus, although not containing as many or as poorly packed residues, still stands out as an area of poor packing. The rest of the model, on the whole, was reasonably well packed.

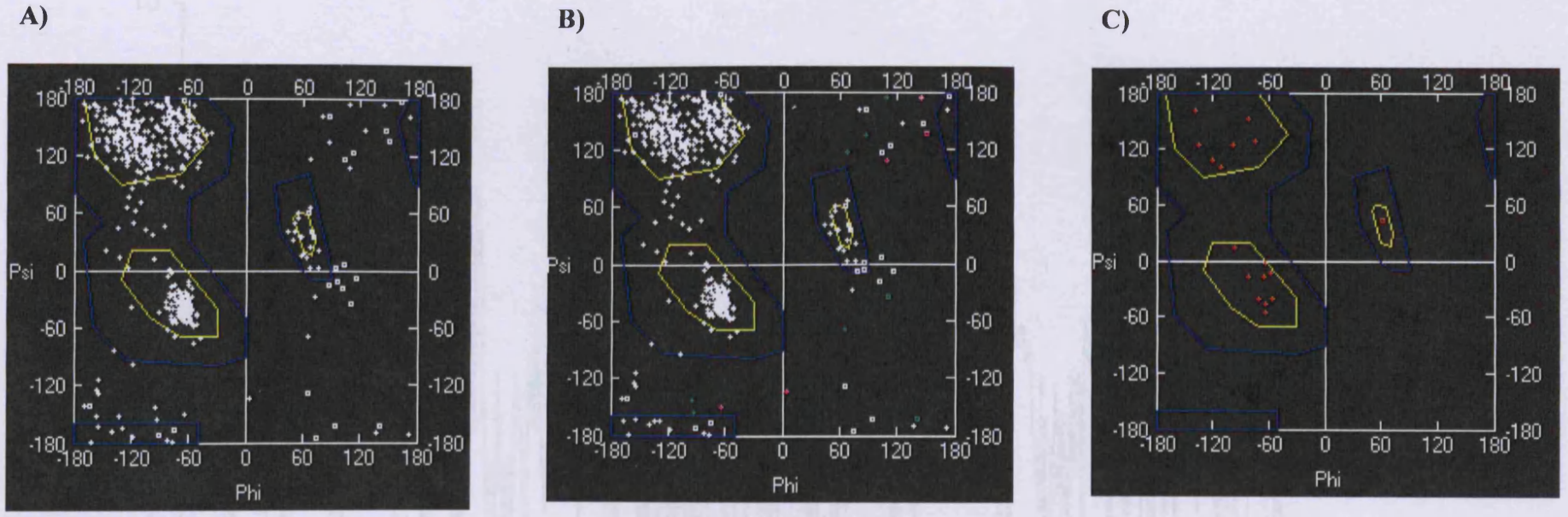


Figure 26. Ramachandran plots of the FMO3 model generated using 1VQW as a template

A) The Ramachandran plot for the FMO3 model generated using 1VQW as a template showing 41/532 (7.7%) of amino acid residues in sterically improbable regions of the energy landscape (points outside the defined rings boundries of allowed regions, shown in blue and yellow). B) The number of amino acids in the improbable areas was reduced to 33/532 (6.2%) by moving residue sidechains very close to 'allowed' regions into the region by breaking the αC-backbone either side of the residue, dragging the side chain into the region and then ligating the αC-backbone within Swiss-PDBViewer v.3.7, as displayed in the Ramachandran plot. 6/33 (18.1%) residues are within the region identified by Verify3D and ProSA as being a region that may not have been built correctly (see section), highlighted in pink. 8/33 residues (24.2%) of these residues within the disallowed regions of the Ramachandran plot reside in the C-terminal region of the model, highlighted as an area of concern by Verify3D and ProSA and are highlighted in green.
C) The residues suggested by MBT Ligand Explorer (see Table 9) as being important in binding to the FAD prosthetic group are highlighted as red dots in the Ramachandran plot. All residues were found to be in the 'allowed' regions, suggesting that the model is reasonable in these important areas

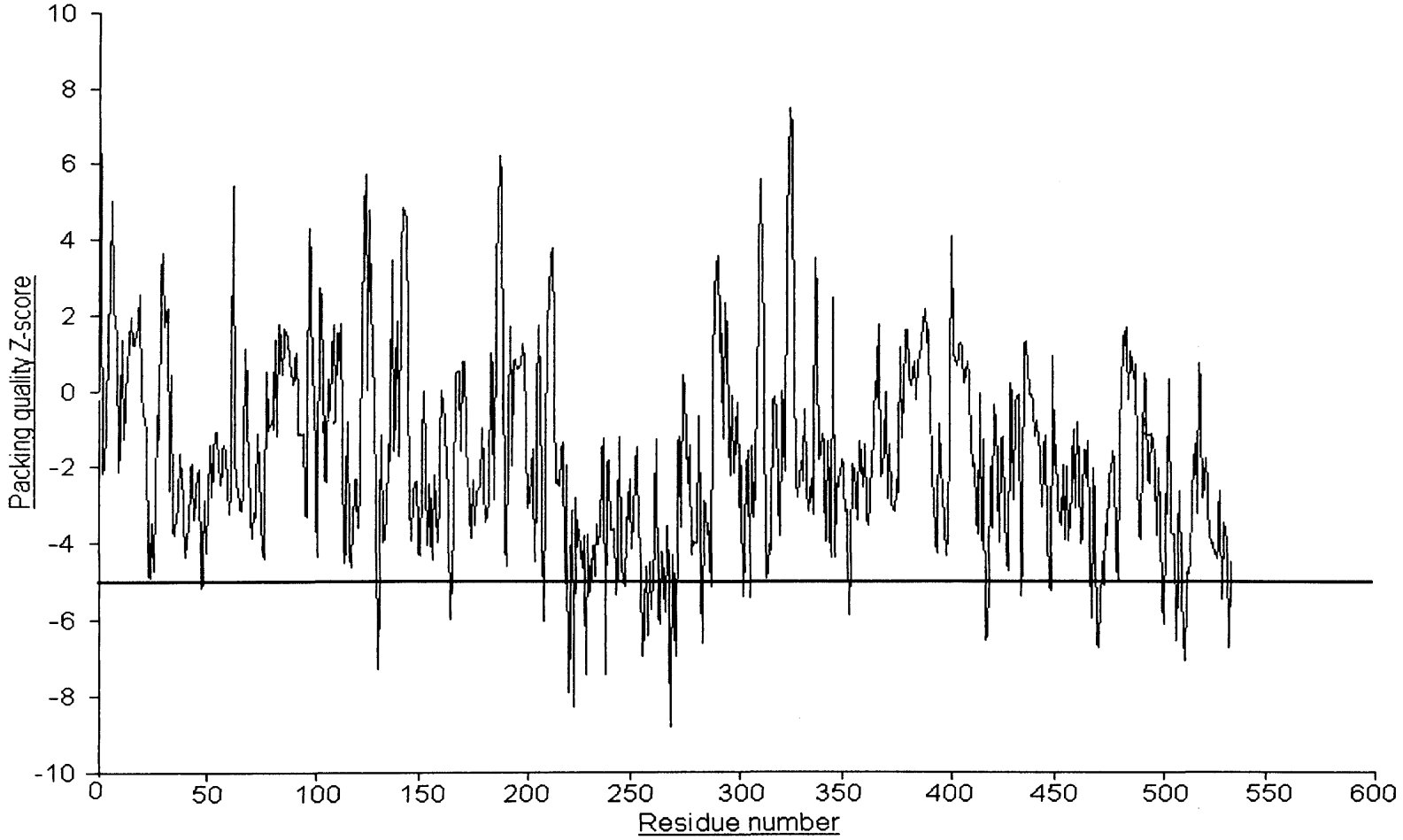


Figure 27. WHATIF assessment of the amino acid packing quality of the FMO3 model generated using 1VQW as a template
A threshold value of -5 was used, below which level a residue can be considered poorly packed. For clarity, a horizontal, black bar indicates the threshold. It is apparent that there are two main areas of poorly packed amino acid residues. One corresponds to the area around amino acids 220-270, previously identified as a poorly built region by Verify3D and ProSA. The C-terminus, although not containing as many or as poorly packed residues, still stands out as and area of poor packing.

Homology model of FMO3 using 1W4X as a template

A homology model of FMO3 was also built using the Baeyer-Villiger Monooxygenase from the moderate thermophilic bacterium, *Thermobifida fusca* (1W4X) as a template. The hydrophobic network highlighted in the model (Figures 28 A and B) based on 1W4X was shown to consist of two channels, one containing predicated binding motifs for FAD and the other a predicted binding motif for NADPH. Indeed, the crystal structure of 1W4X included FAD bound to the protein corresponding to the predicted FAD binding site derived from the motifs. A Q-site finder probing of the model (Figure 28 C) showed various hydrophobic sites, but the deepest hydrophobic pocket (incidentally, where NADPH was predicted to bind) contained several hits classified as 'Probable' (E-value < 1e-5) from the Catalytic Site Atlas (Figure.28 D).

The FMO3 model built using 1W4X was aligned with the model generated using 1VQW as a template and compared using the FATCAT server. The alignment of the two models generated using 1W4X and 1VQW showed that they were significantly similar ($P < 0.05$) with an RMSD of 3.22 Å (Figure 29), with the majority of the difference occurring in the difficult C-terminal area where the template coverage was extremely low. Indeed, when the last 37 amino acids are removed from both homology models, the pairwise alignment lowers to 2.3 Å which is significantly similar ($P < 0.005$). Despite this, for reasons discussed later, it was decided not to continue with more detailed studies of the model, built using 1W4X as a template.

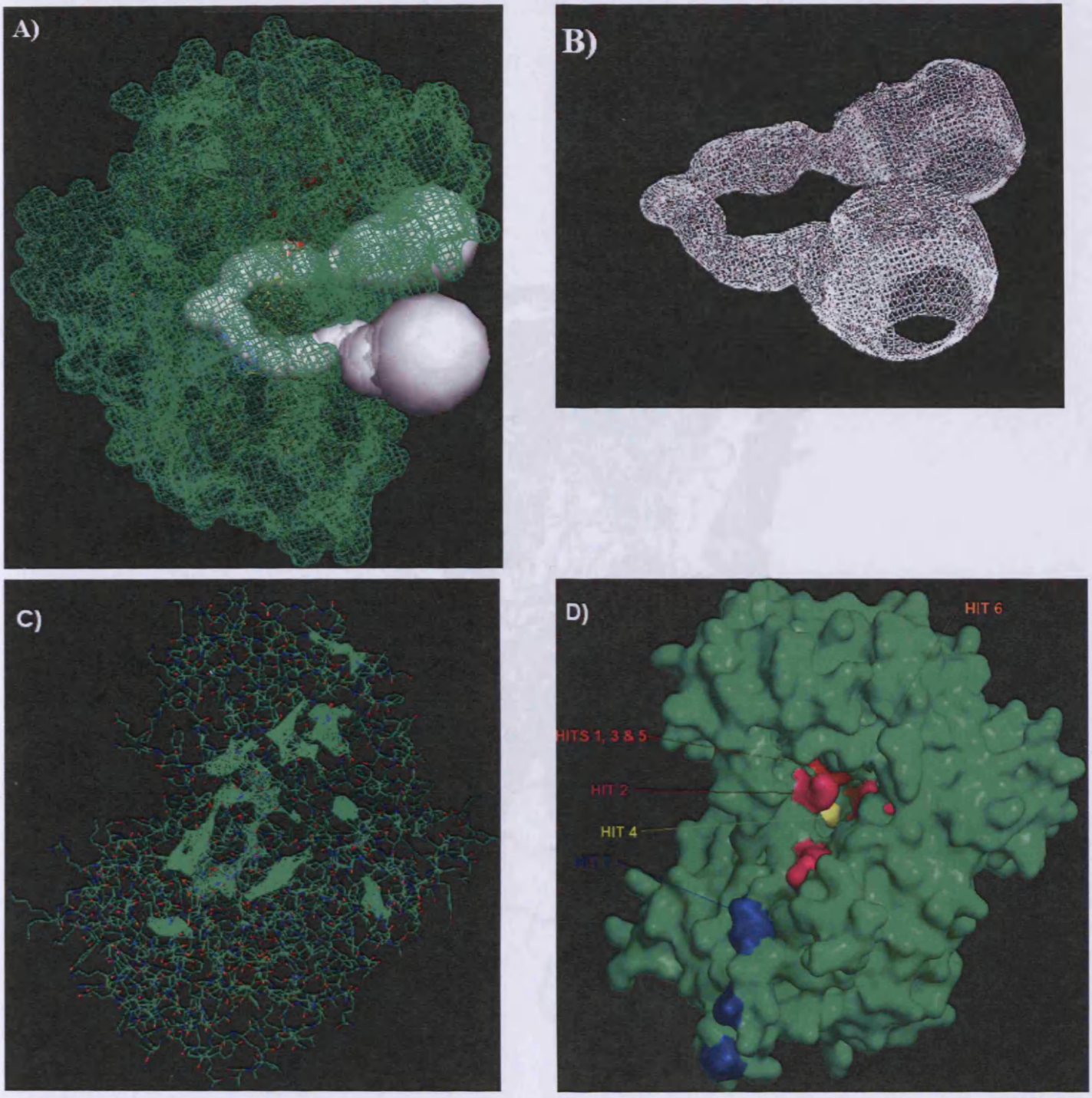


Figure 28. Buried cavities within the homology model of FMO3 based on 1W4X

A) Two buried cavities are highlighted (white) within the model of FMO3 generated using 1W4X as a template. B) The buried cavities are shown with the FMO3 model stripped away to emphasise their association. C) The meshed features within the model are Q-site Finder predictions of predicted buried protein-ligand interaction sites. D) The highlighted residues on the FMO3 model represent the top hits reported by the Catalytic Site Atlas (1 being the top hit).

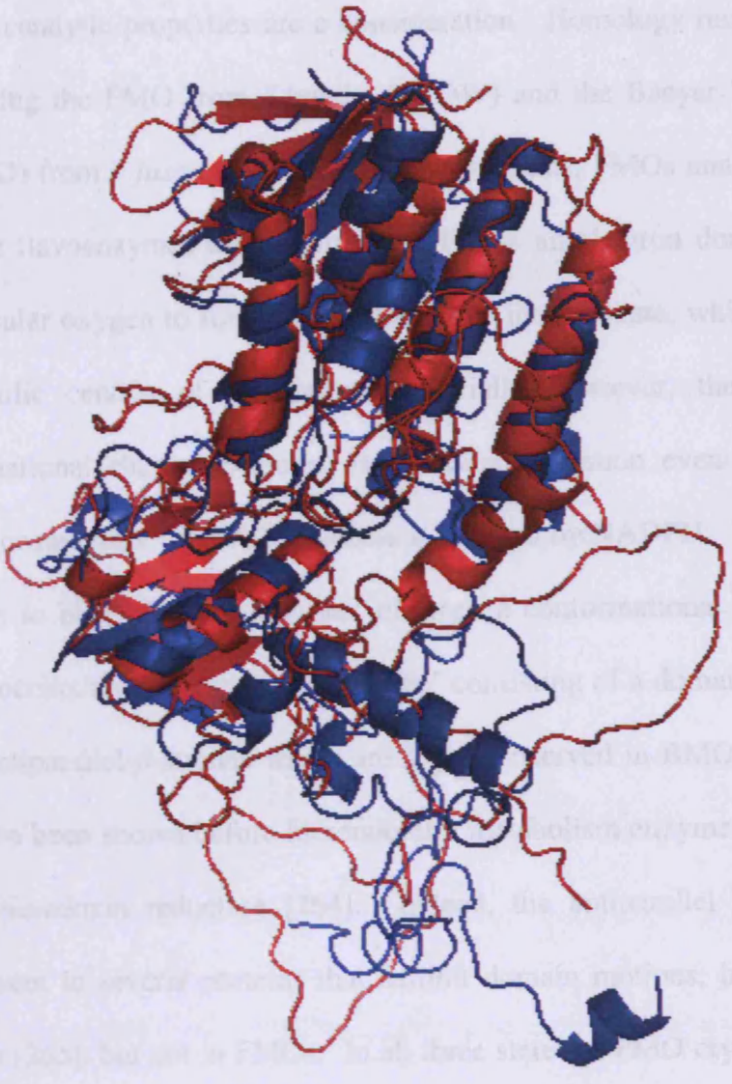


Figure 29. A superimposition of the two models of FMO3 generated by homology modelling

A cartoon view of the two models of FMO3 are shown, generated using 1W4X as a template (blue) and the model of FMO3 generated using 1VQW a template (red). The models are significantly similar ($P < 0.05$) with an RMSD of 3.22 Å.

Homology Modelling of FMO3 - Discussion

When employing homology modelling in an attempt to understand protein structure and biophysical consequences of structural changes, it is important to understand the molecular biology of the protein template. This is especially important when dealing with enzymes, where catalytic properties are a consideration. Homology models of FMO3 were made using the FMO from *S.pombe* (1VQW) and the Baeyer-Villiger monooxygenase (BMO) from *T.fusca* (1W4X). In many respects, FMOs and BMOs are similar. Both are flavoenzymes that require NADPH as an electron donor and both react with molecular oxygen to form a flavin-peroxide intermediate, which then attacks soft-nucleophilic centres of substrates. Crucially, however, the BMO undergoes a conformational change executed via a domain rotation event during catalysis [215]. This occurs once the FAD molecule is reduced by NADPH. In order for molecular oxygen to bind, the enzyme must undergo a conformational change. The domain rotation occurs around a molecular 'hinge' consisting of a domain linker in the form of two antiparallel β-strands which are higly conserved in BMOs [215]. Domain rotations have been shown before in xenobiotic metabolism enzymes during catalysis, such as thioredoxin reductase [264]. Indeed, the antiparallel β-strand domain linker is present in several proteins that exhibit domain motions, including thioredoxin reductase [265], but not in FMOs. In all three states of FMO crystallised (with no ligand bound, with substrate bound and with NADPH bound), no conformational change is present [214]. Despite the high degree of similarity of the two homology models generated, due to the different mode of action of catalysis by the two enzymes, only the model generated using the FMO from *S.pombe* (1VQW) was selected for further study.

Homology modelling of FMO3 based on the FMO from *S.pombe* (1VQW)

Despite domain assignment software such as DomPred [266] consistently predicting FMO3 existing as a single domain unit, the homology model of FMO3 displays two distinct regions of secondary structure bundles. Although these regions may well not qualify as being true protein domains (independent or, at the least, semi-independent units, able to fold and, in some cases, retain function if separated from the parent chain [267]) their seemingly distinct nature is interesting. It may well be that these regions are true domains, with automatic domain assignment methods being fallible due to the complexity of the task [268] and the the paucity of domain definitions already in the databases for FMOs. The C-terminal region of the FMO3 model exists essentially as a long, floppy loop, devoid of more complex secondary structure. In many ways this region is meaningless due to the lack of coverage by the template protein and therefore very little structural information can be gained from analysis of this region.

The larger of the two domains is where the NADPH and FAD molecules are predicted to bind and where a molecule of the latter is modelled. In the model, the flavin moiety is bound within the α-helix bundle, with the adenine dinuceotide moiety of the molecule penetrating deep into the protein, poking between an anti-parallel β-sheet and with the tip of the moiety residing near a parallel β-sheet. It is perhaps not unreasonable to speculate that this region of secondary structure may be involved in maintaining core stability around this vital region of catalysis. Despite the proposed catalytic region residing in a large, deep pocket, we theorise that the actual metabolism of substrates takes place relatively near the solvent exposed opening of the pocket, where the isoalloxazine ring of the flavin moiety of the FAD molecule resides.

The crystal structures of 1VQW [214] have helped to give further insights into the mechanism of FMO action. The correct orientation of the isoalloxazine ring of the FAD molecule is predicted to be vital in order for it to act as an efficient prosthetic group, namely to accept a hydride ion from NADPH, accept molecular oxygen and for oxygenation of the substrate. Therefore, it was judged to be important for the modelled FAD molecule within the FMO3 model to be in a reasonable orientation. A cursory examination of the FAD molecule orientation seemed to suggest that the ring was in a reasonable orientation when compared to the 1VQW template. Further examination of the model revealed that the relative distance and orientation of Asn 61 to the FAD isoalloxazine ring (Figure 20 and Table 9) was comparable to Asn 91 and the FAD isoalloxazine ring within 1VQW. This is encouraging as Asn 61 is the human FMO3 equivalent to *S.pombe* Asn 91, and Asn 91 has been shown to be a key residue in binding molecular oxygen and is thought to be the only FMO residue directly involved in substrate catalysis. Protein sequence alignment of all mammalian FMOs available at the time, an FMO from *Saccharomyces cerevisiae* (*S.cerevisiae*) and four putative FMOs from *Caenorhabditis elegans* (*C.elegans*) showed that the asparigine at this position was conserved in all sequences. This calls into question the previous theory of Asn 61 involvement in a conserved secondary structure of a conserved membrane interaction domain [141].

FAD is a relatively large molecule and, within the FMO3 model, most of the molecule would not be visible, with the adenine dinucleotide moiety of FAD buried deep within the protein (Figures 21 and 24). Several residues are known to be involved in TMAU, where FMO3 catalysis is severely disrupted or destroyed, that reside proximal to the modelled FAD molecule within the FMO3 model, namely E32K, A52T, V58I and N61S (Figure 21 B). It was therefore decided to probe the

involvement of other FMO3 residues in FAD binding. The hydrophobic and hydrophilic interactions were predicted, measured and tabulated using MBT Ligand Explorer (Figure 22 and Table 9). 17 FMO3 residues were predicted to contribute to the binding of FAD (Table 9), including two of the residues associated with TMAU, E32K and N61S (Figure 30). Asn 61, discussed above, was predicted to be vital for the catalytic cycle of FMO3 with the other residues associated with FAD binding, and was presumably important in FAD stability and molecule accessibility. Other interacting residues reported by MBT Explorer, although, not associated with FMO3 catalytic reduction to date, were proximal to residues known to be so. Lys 33 (Figure 30), which is one residue away from Glu 32 (Figure 30) and known to be associated with TMAU, was predicted to have hydrophilic interactions with FAD. Ser 60 is one residue away from the important Asn 61 (Figure 30), discussed above, and associated with TMAU, and is predicted to have a hydrophilic association with FAD. Met 67 (Figure 30) is one residue away from Met 66 (Figure 30), known to be associated with TMAU, and is predicted to have a hydrophobic association with FAD.

In order to probe the involvement and importance of each residue predicted to be involved with FAD binding and stability *in vitro* FMO3 codons could be altered by mutagenesis, the protein expressed and relative FAD content assayed, to see how much FAD binding is disrupted. Recently, a group showed that the M66I FMO3 variant failed to incorporate or retain FAD, despite expressing to comparable protein levels [259]. M66 (Figure 30) is highlighted on the homology model of FMO3 in Figure 24. Interestingly, a variant of the 492 residue (Figure 30), R492W was also shown to lack FAD. This residue is also highlighted in Figure 24. This residue was

Figure 30. Homology model of FMO3 with residues implicated in FAD incorporation and/or retention.

The FMO3 homology model is shown with Glu 32 (orange), Lys 33 (pink), Ser 60 (magenta), Asn 61 (cyan), Met 67 (yellow) and R492 (blue) displayed as spheres. The FAD molecule is displayed in as sticks (red).

found to be variant in the English subject studied in this investigation (see section 3.1.), albeit in the Q492 state. The finding that this residue caused the enzyme to be unable to incorporate or retain FAD is curious, as the residue does not reside near the catalytic core near the FAD molecule. Unfortunately, very little information about this region is forthcoming from the homology model due to lack of template coverage, discussed above. No other residues predicted to be involved with FAD incorporation or stability (Table 9) were investigated by this group, although other variants shown to be involved in TMAU were and had FAD levels comparable to wild-type enzyme, showing that FAD presence required but not sufficient for FMO3 catalytic efficacy.

Mapping the two FMO identifying sequences, 'F(A/T)TGY' and 'FXGXXXHXXXF', onto the model does not reveal an obvious reason for their high level of conservation. Due to the highly conserved nature of the 'FATGY' motif, it would be reasonable to assume that this motif is important to the molecule. A 'fishy taint' in the eggs of chickens has been associated with a mutation in *Fmo3* which changes the threonine to a serine [269]. The position of the motif on the model of FMO3 does not instantly give a clue as to the role of this motif. The 'FATGY' motif on the FMO3 model maps to just above the putative active site pocket (Figure 23 A). Undoubtedly, the integrity of this pocket would be very important to the maintenance of an environment suitable for redox reactions to take place, but any specific reason as to why this region is so highly conserved is open to speculation. This is also the case for the 'FXGXXXHXXXF' motif, perhaps even more so, as it maps to a position in the larger domain but with little obvious function (Figure 23 B). It may be that these motifs are important in order for proper protein folding to take place or for maintenance of structural stability.

The catalytic site atlas did not identify any probable catalytic residues from its database of enzyme active sites, within the putative active site of the FMO3 model based on 1VQW. This may be due to the Atlas' bias towards enzymes that have had their mechanisms solved; hence any enzyme without a family member in the PDB would be less likely to be represented.

Mutations and Polymorphisms of FMO3

The residues known to be variant in FMO3 were mapped on to the homology model of FMO3 to try and understand any structural consequences of their variance. The variants included all residues known to destroy (or severely disrupt) FMO3 activity, those residues known to alter FMO3 catalytic activity and the two residues for which novel mutations were discovered in this investigation.

The majority of the amino acid variants shown to be destructive in FMO3 seem to cluster in the larger domain, with a significant concentration around the FAD coenzyme (Figure 24). The majority of polymorphic sites shown to affect, but not destroy, FMO3 activities seem to cluster in the smaller bundle of secondary structure (Figure 24). Other than disrupting protein folding or stability, it is not obvious as to how many of the variants that map away from the catalytic core are responsible for the lack of FMO3 activity associated with their presence. R238 is positioned near, but not within the putative active site. A variant of this residue, R238Q, was discovered by a patient study in this investigation (see section 3.1.). As shown in this investigation, when this residue is mutated to glutamine, FMO3 activity is destroyed (see section 3.2.). It has also been shown that when this residue is found as a proline, FMO3 activity is destroyed [56]. The proximity of the residue to the active site courts speculation that the variants may prevent proper access to the active site for the

coenzyme, electron donor, -substrate or a combination of the three. Any and all of these situations would prevent FMO3 catalysis. The second variant, discovered for the first time in this investigation, was R492Q (see section 3.1.). It has been shown previously to destroy FMO3 activity when changed to a tryptophan at this position [35, 62]. It is difficult to hypothesise about the possible structural consequences of this mutation, as it resides on the poorly modelled C-terminal region, discussed in more detail below.

Homology model validation

The observations so far assume that the model of FMO3 resembles the actual structure of human FMO3. There is no way of knowing this until the structure is solved by crystallographic methods, but the quality of the homology model was assessed to try and give some sort of indication of how much faith one can place in this model of FMO3.

Four well-established and widely cited methods of structure validation were used to assess the homology model of FMO3; Verify 3D (assesses residue packing environment), ProSA (assesses packing based on special separation of Cβ residues), WHATIF (assesses residue packing quality) and Ramachandran Plot (examines the psi and phi backbone conformational angles for each residue in a protein.) analyses, generated using Swiss-PDB Viewer.

Generally, Verify 3D, ProSA and WHATIF agreed that the model of FMO3 was a good one in terms of residue environment, energy distribution and atomic packing quality. However, all three methods identified two areas of concern with the model, the first around a 50 amino acid stretch ranging from around amino acids 220-270 and the second region located at the C-terminal portion of the model. It is not

surprising that the C-terminal region was flagged as being unrealistically built. This region was always going to be difficult to model due to the lack of template coverage (see Appendix 2 for details). A possible way of improving the model of this C-terminal region into a more realistic conformation would be to consider this tail region in isolation, BLAST the PDB for more suitable templates, and then build a chimeric model incorporating the top hit. This would require a reasonable level of primary sequence homology to expect an improvement. Such homology templates are not available in the PDB at this time. It is not clear as to why the region around amino acids 220-270 has not been built correctly. The template provided to Modeller as a basis of structural comparison had coverage in the majority of this area (see Appendix 2 for details). Despite the classification of this region as being poorly modelled, the general topology of the model would be expected to be generally correct, so one can draw some sort of gross observations, such as, the proximity of the residue to the active site may possibly of interest. Both regions of poor quality do not seem to be directly involved in the catalytic core of the molecule and do not contain many mutations and/or polymorphisms known to destroy FMO3 activity, although the R238 residue mentioned above does reside in this area. We can not dismiss the importance of the C-terminal region of FMO3, with several truncation mutations and three residue mutations in this region known to perturb FMO3 activity (Table 2), including the residue R492, mentioned earlier. Mutations in this area may have more to do with the importance of the C-terminal region with regards to cellular localisation, rather than catalytic activity of protein folding/stability.

One measure of a good model is having more than 90% of the amino acid residues in the sterically 'allowed' regions of a Ramachandran plot [263]. This was found to be the case in the FMO3 homology model before refinement and was

subsequently improved during refinement. Almost half of the residues in poorly defined regions of the Ramachandran plot had also been previously identified as being present in the two areas of concern discussed above. Crucially, the residues identified by MBT Ligand Explorer (Table 9) as being important in binding to the FAD prosthetic group were found to be in the 'allowed' regions, suggesting that the model is reasonable in these important areas.

Comparison of FMO3 model built with others recently published

Recently, Borbás et al., (2006) [270] described a model of human FMO3 that they had created to help describe the consequences of the L360P polymorphism, discovered in their lab. This model was generated using the SWISS-MODEL system [164] which is a fully automated modelling program with absolutely no human intervention during model building or model scoring or refinement afterwards, much like ModWeb (see section 1.3.). More recently, Yeung et al., (2007) [259] presented a homology model of FMO3 using proprietary software (Molecular Operating Environment (available from http://www.chemcorp.com)) which was not refined or scored.

The general topology of the two recently published models of FMO3 and the one built in this investigation are similar (Figure 31). This is unsurprising as all three use the same template, 1VQW, to generate the majority of the molecule, with the Yeung et al., (2007) [259] model having the added feature of a C-terminal tail segment modelled on cytochrome C oxidase from *Rhodobactor sphaeroides* [271] (PDB code 1M56). The value of this feature could be questioned however, considering the low template coverage in terms of primary sequence identity being only 22% to human FMO3, below the 'midnight zone' cut-off point (see section 1.3.).

All three models seem to have two bundles of secondary structure, with the largest forming the core, catalytic centre with the FAD binding domain. More detailed comparison of the structures would require the PDB coordinates of all three, not presently available.

A)

B)

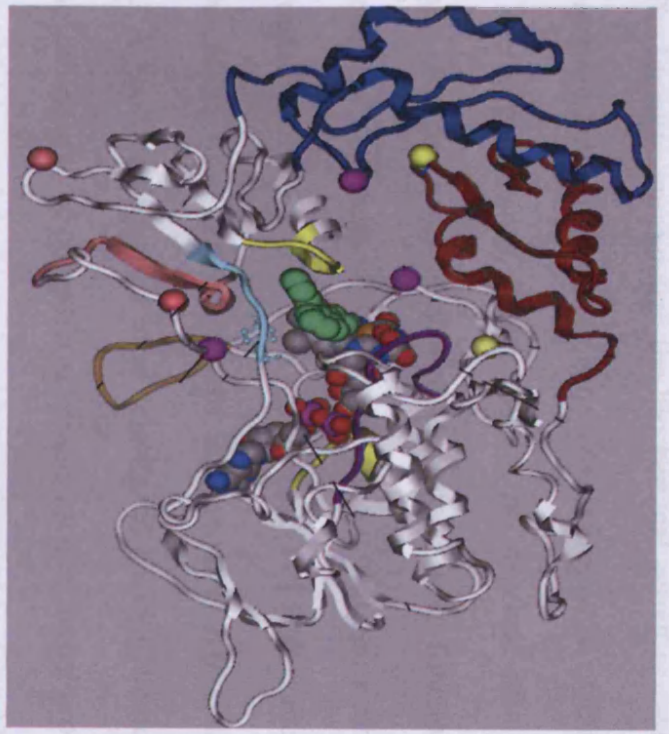


Figure 31. Homology models recently published of FMO3
Both models are based on the *S.pombe* FMO, 1VQW, the template used to generate the homology model in this investigation. A) Borbas et al., (2006) [270] presented a model highlighting the location of FAD (pink) and NAPDH (purple) binding sites. The variant site 360 is also highlighted (yellow). B) Yeung et al., (2007) [259] presented a model highlighting several variant sites (single spheres), regions most similar to yeast FMO (white), additional residues at C terminus (red), FATGY motif (cyan), nucleotide binding motifs (yellow) and FMO 'identifying' and signature motifs (purple).

FMO3 model – Active site of FMO3

The literature surrounding the biochemical activities catalysed by the active site of FMOs has been built up since the early pioneering work of Daniel Ziegler in the 1960's right up to the present day. Since there is no structure of a mammalian FMO and no homologue of high enough sequence identity to infer function, the active site and the consequences of alterations to the active site, i.e. through amino acid substitutions, have been shrouded in mystery. Having a refined model of FMO3 means it would be possible to probe the biophysical environment of any putative active site through protein-ligand docking. A good starting point in protein-ligand docking would be AutoDock, which is a free-to-academia suite of computer programs used to model protein-ligand interactions. It is the most highly cited docking algorithm in the peer-reviewed literature (correct as of 10th October, 2006) [272]. AutoDock has been shown to be more accurate than other free-to-academia, protein-ligand software such as FlexX [273] and Dock [274] in benchmarking tests [275]. AutoDock uses a protein molecule (such as the homology model of FMO3) and a ligand (such as FMO3 substrate, methimazole or FMO3 cofactor, FAD) as input files. The program then probes a user-defined space, e.g. the putative FMO3 active site, using simulated annealing techniques for a low energy conformation. This process is an iterative one and the results can be viewed by the user. Models can also be mutated and then subsequently subjected to docking studies in an attempt to understand the biophysical consequences of amino acid substitutions at the tertiary level. This sort of study could give fresh insight into the impact of the mutations underlying TMAU and the polymorphisms known to cause changes in the catalytic capacity of FMO3.

FMO3 model – insight to the structural changes which lead to TMAU

The mutations leading to TMAU are more likely to cause perturbations in FAD binding or orientation, leading to FAD being completely displaced from the molecule or bound in a state which renders the prosthetic capacity of the molecule inefficient, or indeed, lost. This could be assayed *in vitro* by generating mutations predicted to perturb FAD binding by site directed mutagenesis and then assessing FAD content compared to wild-type protein. This approach is supported by the recent work by Yeung et al., (2007) [259] who showed that M66I, a residue known to cause TMAU and one residue away from M67 which was predicted in section 3.3. Table 9, to interact with FAD, failed to incorporate FAD forming the FMO3 holoenzyme *in vitro* [259]. It would be interesting, and possible, to systematically generate the FMO3 variants predicted to interact with FAD and then assay the molecules for FAD incorporation. To complement this, having a refined FMO3 homology model allows protein-ligand docking studies to be performed. Where docking studies suggested that FAD may well be able to bind to FMO3 but does so in such a manner that the correct orientation of the isoalloxazine ring can not be achieved, then FAD content would be expected to be maintained at a wild-type level, but catalysis would be perturbed.

Other variant sites are more difficult to theorise as to the causes of FMO3 activity perturbation, other than protein folding, structural integrity maintenance and perhaps cell localisation.

3.4. Evolution of the *FMO3* locus

The work in this section has been published.

Allerston, C.K., Shimizu, M., Fujieda, M., Shephard, E. A., Yamazaki, H. and Phillips, I. R. (2007) Molecular evolution and balancing selection in the flavin-containing monooxygenase 3 gene (*FMO3*). *Pharmacogenetics and Genomics*. 17(10):827-839.

FMO3 nucleotide sequence variation

6,274 bp of the *FMO3* gene were sequenced from 23 Japanese potential TMAU sufferers. This comprised 3,358 bp of contiguous 5' flanking sequence (extending from -5144 to -1786, relative to the A of the ATG translational initiation codon), the entire coding sequence (1,599 bp, 533 codons, including the stop codon, in exons 2-9) and 1,317 bp of partial intronic and 3'-untranslated region. The sequencing was performed by the Yamazaki group in Japan. Haplotype phases were inferred by the Yamazaki group using the software packages SNPAlyze (DYNACOM, Chiba, Japan) and HAP [276]. Results from one individual were ambiguous and thus were excluded from the data set. Sixteen diallelic SNPs were identified, 12 in the 5'-flanking region and 4 in the coding region (Table 10). Twelve of the SNPs were transitions and 4 were transversions. Three of the SNPs, all in the 5'flanking region, at -4600, -4488 and -3788, are novel. All of the coding-region SNPs were nonsynonymous substitutions: at positions +15167G>A in exon 4 (Glu>Lys change at amino-acid residue 158), +15549C>T in exon 5 (Arg>Cys change at residue 205), +18281G>A in exon 6 (Val>Met change at residue 257) and +21443A>G in exon 7 (Glu>Gly change at residue 308). The SNPs at codons 158, 257 and 308 are polymorphic variants that individually have a limited effect on enzyme activity [35, 67, 255]. None of the SNPs were present as a singleton, i.e. occurred in only one sequence of the sample, and only one (+15549C>T) was present as a doubleton. Comparison with the *FMO3* sequence of chimpanzee (*Pan troglodytes*) (Genbank accession no. NM_001009092.1) showed that for each diallelic SNP observed in humans, one of the two alleles was present at the

Table 10. SNPs and their frequencies

dbSNP Accession Number		Allele		Frequency of derived variant in	
				Potential TMAU sufferers[c] (n=44)	Controls[d] (n=90)
	Position	Ancestral[a]	Derived[b]		
rs6608461	-5109	G	C	0.114 (0.051, 0.206)	0.100 (0.061, 0.151)
	-4600	T	C	0.114 (0.051, 0.206)	0.100 (0.061, 0.151)
rs12402693	-4488	C	T	0.114 (0.051, 0.206)	0.100 (0.061, 0.151)
	-3788	T	C	0.114 (0.051, 0.206)	0.100 (0.061, 0.151)
rs1736554	-3606	G	A	0.114 (0.051, 0.206)	0.211 (0.163, 0.266)
rs3754487	-3549	C	T	0.318 (0.228, 0.419)	0.267 (0.216, 0.321)
rs1736555	-3548	A	G	0.545 (0.446, 0.643)	0.578 (0.524, 0.631)
rs3754489	-3544	C	T	0.114 (0.051, 0.206)	0.100 (0.061, 0.151)
rs12404183	-2854	T	C	0.114 (0.051, 0.206)	0.100 (0.061, 0.151)
rs1736560	-2650	C	G	0.545 (0.446, 0.643)	0.578 (0.524, 0.631)
rs12404218	-2543	T	A	0.114 (0.051, 0.206)	0.100 (0.061, 0.151)
rs3754491	-2177	G	C	0.114 (0.051, 0.206)	0.100 (0.061, 0.151)
rs2266782	15167(E158K)	G	A	0.364 (0.270, 0.464)	
rs28363549	15549(R205C)	C	T	0.045 (0.009, 0.128)	
rs1736557	18281(V257M)	G	A	0.068 (0.021, 0.155)	
rs2266780	21443(E308G)	A	G	0.318 (0.228, 0.419)	

[a]The positions of the SNPs are given relative to the 'A' of the 'ATG' translational initiation codon. For each codon-region SNP, the codon position is also given.
[b]The ancestral state of each SNP was inferred from the chimpanzee sequence.
[c]Numbers in parenthesis are 95% confidence intervals.

corresponding position in chimpanzee (Table 11). The ancestral state of each SNP was inferred from the chimpanzee sequence. With the exception of SNPs at –3548 and –2650, the more common human variant corresponded to the inferred ancestral allele.

Calculations of measures of nucleotide diversity were based on the number of effectively silent sites in the sequenced region. For the potential TMAU sufferers this was estimated to be 5,065 and is the sum of all 5'flanking and 3'-untranslated sites, plus all intronic sites (with the exception of the two conserved splice sites at the 5' and 3' end of each intron) and exon sites that are silent. Although the 5'-flanking region undoubtedly contains some sites that are functionally constrained, the identity of these is not known and intuitively they would constitute a very small proportion of the 3,358 bp of 5'-flanking region sequenced. For the control group, all 3,358 bp of 5'-flanking region were considered silent.

Two estimates of sequence diversity were calculated: the average expected per-site nucleotide heterozygosity (θ_w), estimated from the observed number of polymorphic sites [174], and nucleotide diversity (π), a direct estimate of per-site heterozygosity, derived from the average pairwise sequence difference between two random sequences in a sample [227]. For the potential TMAU sufferers, θ_w is 0.00054 and π is 0.00065 (Table 12). For the control group, both estimates of diversity are higher, θ_w being 0.00070 and π 0.00095. The estimates of nucleotide sequence diversity (π) for *FMO3* are similar to the average value of π (0.00075) for human genetic loci [277, 278]. There were no mutations at silent sites in the coding region.

Table 11. Haplotypes and their estimated frequencies in the potential TMA patient group

	Nucleotide position																
	-5109	-4600	-4488	-3788	-3606	-3549	-3548	-3544	-2854	-2650	-2543	-2177	+15167	+15549	+18281	+21443	
													Codon				
Haplotype													158	205	257	308	Frequency
Chimp	G	T	C	T	G	C	A	C	T	C	T	G	G	C	G	A	
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.364
2	-	-	-	-	A	-	G	-	-	G	-	-	-	-	-	-	0.045
3	C	C	T	C	-	-	G	T	C	G	A	C	-	-	-	-	0.114
4	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	0.045
5	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	0.045
6	-	-	-	-	-	T	G	-	-	G	-	-	A	-	-	G	0.318
7	-	-	-	-	A	-	G	-	-	G	-	-	-	-	A	-	0.068

The positions of the SNPs are given relative to the 'A' of the 'ATG' translational initiation codon. For each codon-region SNP, the codon position is also given. Bases identical to the chimpanzee sequence are indicated with a dash and are inferred as the ancestral state. Haplotype 1 is identical at each SNP position and thus represents the ancestral haplotype. Polymorphisms are derived from this state and the base change is indicated.

Table 12. Diversity estimates and neutrality tests for *FMO3*

	Potential TMAU sufferers		Control group	
Sample summaries				
n[a]	44		90	
S[b]	16		12	
s[c]	0		0	
R_m[d]	1		0	
		P values		P values
h[e]	7	< 0.05*	4	< 0.01
Expected no. of haplotypes[f]	11.6		9.6	
Parameter estimates				
θ_W[g]	0.00055		0.0070	
π[h]	0.00066		0.00095	
Test statistics				
Tajima's *D*[176]	0.66	-	0.92	-
Fu and Li's *F**[177]	1.59	< 0.05	1.47	< 0.001
Fu and Li's *D**[177]	1.51	< 0.05	1.52	< 0.05
Fu's *Fs*[228]	3.50	< 0.05*	7.38	< 0.01
Raggedness (r)[179]	0.16	-	0.37	< 0.02
Z_{nS}[181]	0.32	< 0.05*	0.48	< 0.02
Wall's *B*[180]	0.33	< 0.05*	0.45	< 0.05
Wall's *Q*[180]	0.38	-	0.50	< 0.02

A total of 6,274 sites (5,033 silent sites) were analysed in potential TMAU sufferers and 3,358 silent sites in control individuals.

[a]Number of chromosomes surveyed

[b]Number of segregating sites

[c]Number of singleton sites

[d]Minimum number of recombination events, based on the four-gamete test [224]

[e]Number of haplotypes

[f]Given the extent of nucleotide diversity in the sample and assuming no recombination events have taken place

[g]Expected heterozygosity per nucleotide [174]

[h]Average pair-wise sequence difference per nucleotide [227]

P values are given within the potential TMAU sufferer group when significant ($P < 0.05$) and are based on comparisons to estimates obtained from coalescent simulations performed assuming constant populations size and no recombination, except where marked "*", which were significant only when simulations were performed for the level of recombination observed in the sample.

Haplotypes

The 16 SNPs in the sample of 44 chromosomes were found to segregate as 7 distinct haplotypes. These haplotypes, together with the haplotype composed of the ancestral state of each allele, inferred from chimpanzee, are shown in Table 11. The most frequent haplotype in the sample (haplotype 1), with a frequency of 0.363, is identical to the chimpanzee sequence at all 16 SNP positions and thus represents the inferred ancestral haplotype. One of the haplotypes (haplotype 4) appears to have arisen as the result of a recombination event (see section 3.4.).

Although individuals in the TMAU cohort were self-selected on the basis of an apparent fishy body odour, only seven were diagnosed as suffering from mild TMAU, as judged by urinary secretion of 70-90% of total TMA as TMA *N*-oxide, with two more being classed as borderline (excretion of 90% of total TMA as the *N*-oxide) [85]. Of the seven mild TMAU sufferers, six have a haplotype pair that is the same as that present in one of the unaffected individuals (Table 13). If only the 5'-flanking region is considered the 12 SNPs present in this region segregate as four distinct haplotypes. The identity of these haplotypes is the same, and their frequencies are similar, in the potential TMAU sufferers and control groups. This is also the case for the haplotype pairs in the two groups. Therefore, with the exception of R205C (see section 3.4.), none of the SNPs or haplotypes identified in the 44 chromosomes of the potential TMAU group is apparently causative of the disorder. Some affected individuals may possess mutations, in unsequenced regions of the gene, which would compromise gene expression or RNA processing. Alternatively, the mild TMAU may be the result of dietary intake of relatively large amounts of TMA precursors, e.g. choline, or to factors that affect the metabolism of such compounds, for instance, gut flora composition or functional polymorphisms of endogenous enzymes.

Table 13. Genotypes and their occurrences in potential TMAU sufferers

Genotype (haplotype pair)	Occurrences in		
	Diagnosed TMAU sufferers	Borderline cases	Non-affected individuals
1-1	1		1
1-2	1		1
1-3	1		1
1-5	1		1
1-6	2		4
3-4			1
3-6			2
4-7	1		
6-6		2	
6-7			2
Total	7	2	13

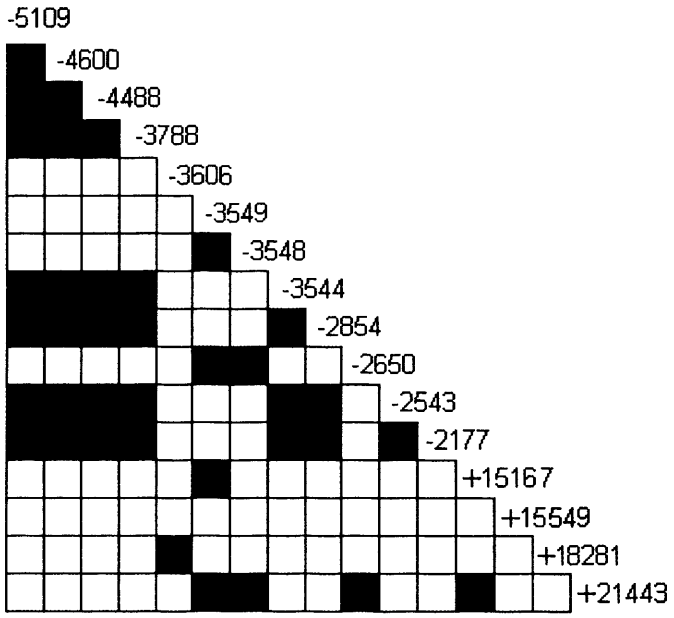
The number of ocurrances of each haplotype pair in diagnosed TMAU sufferers, borderline cases and non-affected individuals is presented. Of note is the finding that six of the individuals diagnosed as suffering for TMAU (albeit mild cases) have a haplotype pair that is also found in non-affected individuals.

Recombination and linkage disequilibrium

The presence in haplotypes of diallelic SNPs of pairs of sites with four gamete combinations is indicative of recombination (crossing-over between the two sites or gene conversion) or repeated mutation. The probability of repeated mutation at the same site is very small. The four-gamete test (Table 12) [224] reveals that only 2 out of 120 pairs of SNP sites compared (-3548, 158 and -2650, 158) have all four gametes. The presence of these two four-gamete site pairs can be accounted for by a single recombination event (between variant sites -2650 and 158).

Of the 120 pairwise comparisons of variant sites, 37 show significant LD ($P < 0.001$) (Figure 32 A). A plot of pairwise LD (measured as R^2) against physical distance indicates that strong LD extends to a distance of 25 kb (Figure 32 B), e.g. between the SNP at –3549 and that at 308, which lies 21.5 kb downstream of the translation initiation codon.

A)



B)

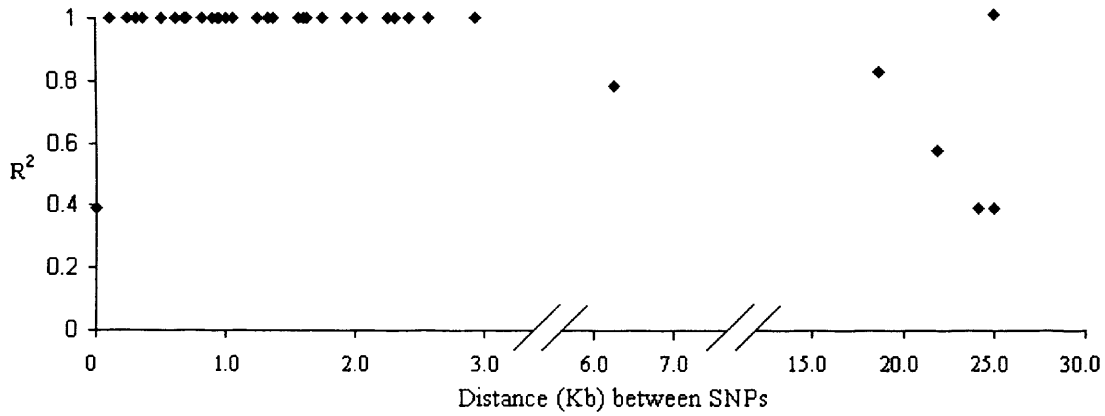


Figure 32. Pairwise linkage disequilibrium

A) A plot of the association among 16 SNPs within 22 Japanese individuals. Blackened boxes indicate 37 out of 120 possible associations that display significant LD at the 0.001% level by a χ^2 test and Bonferroni correction. Numbers denote the nucleotide positions of the SNPs relative to the A of the ATG translational initiation codon.
B) The relationship between pairwise LD, measured as R^2, and physical distance, in kilobase pairs (kb), for the 37 associations significant at the 0.001% level.

Evolutionary history of *FMO3* allelic divergence

The mutational relationship among the haplotypes was visualised by construction of a RM network [225] (Figure 33). In a RM network, circles correspond to distinct haplotypes. The size of each circle is proportional to the relative frequency of the haplotype in the sample. Haplotypes are connected by branches (lines) along which mutations are indicated. Where multiple mutations occur along a single branch, e.g. between the central node and haplotype 3, the order in which these occurred is not known and hence their sequence along the branch is arbitrary.

Haplotype 1 corresponds to the ancestral sequence and thus is the root haplotype of the network. By tracing the connections that link the route haplotype to the other haplotypes, the evolutionary history of genetic changes at the *FMO3* locus can be inferred. The *FMO3* haplotypes fall into four major clades. One of these comprises the ancestral haplotype (haplotype 1) and haplotype 5, which are separated from each other by a single cSNP that causes an Arg>Cys substitution at residue 205 of the protein. The three other haplotype clades comprise of phylogenetically distinct lineages radiating from a central node, separated from the ancestral haplotype by two mutations (at -2650 and -3548): haplotype 3 has undergone eight further mutations in the 5'-flanking region; haplotypes 2 and 7 both have an additional single mutation (at -3606) and are separated from each other by a single SNP that causes a Val > Met substitution at residue 257; haplotype 6 is separated from the central node by three mutations, one (at -3549) in the 5'-flanking region and two SNPs that cause Glu > Lys and Glu > Gly substitutions at residues 158 and 308, respectively.

The placement of one haplotype, haplotype 4, is ambiguous, as indicated by a reticulation in the network. This is due to homoplasy (the inferred occurrence of multiple independent evolutionary events giving rise to the same allelic state at a variable site) that

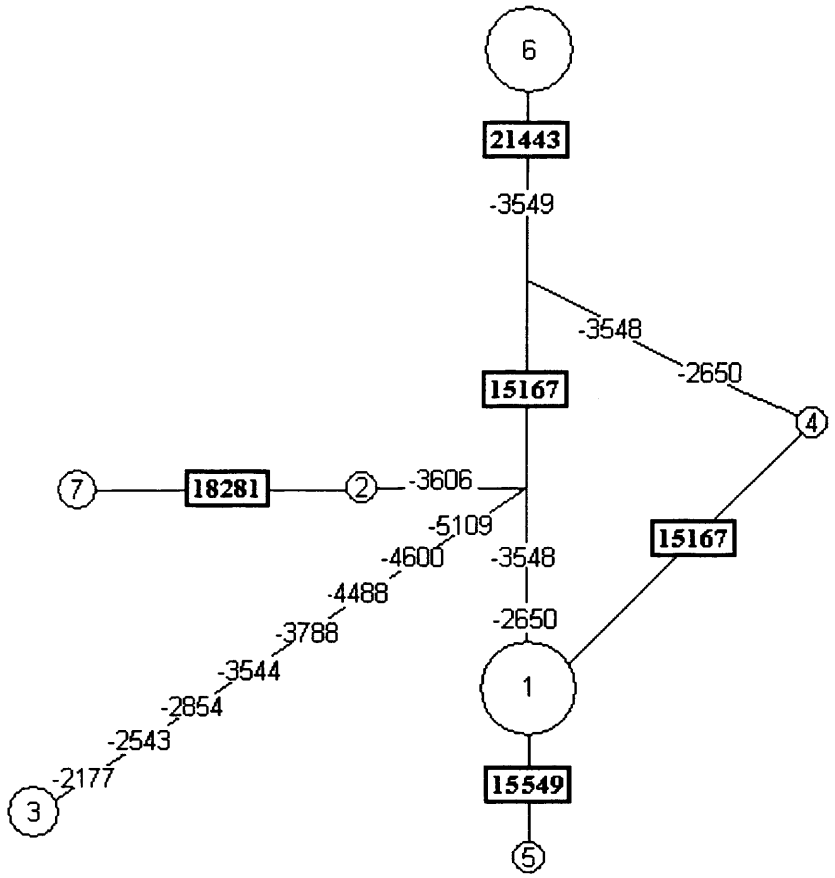


Figure 33. Reduced-median network of *FMO3* haplotypes.

Each of the seven unique haplotypes is represented by a circle, the size of which is proportional to the relative frequency of the haplotype. Mutational relationships are indicated by lines linking the haplotypes. Mutational differences between haplotypes are indicated on the branches of the network. Mutations in the 5'-flanking region of the gene are identified by their nucleotide position relative to the A of the ATG translational initiation codon. Coding-region mutations leading to an amino acid change, 15167 (E158K), 15549 (R205C), 18281 (V257M) and 21443 (E308G), are boxed. Where mutations occur along a branch, their order is arbitrary.

affects sites -3548, -2650 and 158.

Time depth of *FMO3* variation and ages of mutations

To investigate the length of time over which the sequence differences have arisen and of the ages of individual mutations, something which can't be done using a RM network, the sequence haplotypes were analysed through the use of the program GENETREE, which uses maximum-likelihood coalescent analyses [226], under assumptions of neutrality and an infinite-sites model of evolution, i.e. no recombination. Thus haplotype 4, which appears to have arisen via a recombination event, was excluded from this analysis. The method uses all of the information in the infinite-sites-compatible sequence data set.

The maximum-likelihood estimate of the population mutation parameter (θ_{ML}) per locus, conditional on the *FMO3* gene tree and determined by using the GENETREE program, is 3.30. This is slightly larger than θ_W, the estimate of the parameter from the number of polymorphic sites, which is 2.74 for the locus.

The gene tree for *FMO3* is shown in Figure 34. To estimate the neutral mutation rate of *FMO3* we used the net silent-site sequence divergence between humans and chimpanzee (i.e. divergence between species minus divergence within species) and a human-chimp divergence time of ~6 million years before present (BP) [33]. The net sequence divergence was estimated as 47.31. This gave a neutral mutation rate (v) of 7.78 x 10^{-10} per site per year. Assuming a generation time of 20 years, the locus-specific neutral mutation rate (μ) is 7.88 x 10^{-5} per locus per generation. Using the GENETREE maximum likelihood estimate of the parameter θ and our estimate of the locus-specific mutation rate μ, the effective population size (N_e) was calculated, from the relationship $\theta = 4N_e\mu$, as 10,463.

The estimated time to the most recent common ancestral sequence (T_{MRCA}), i.e. the

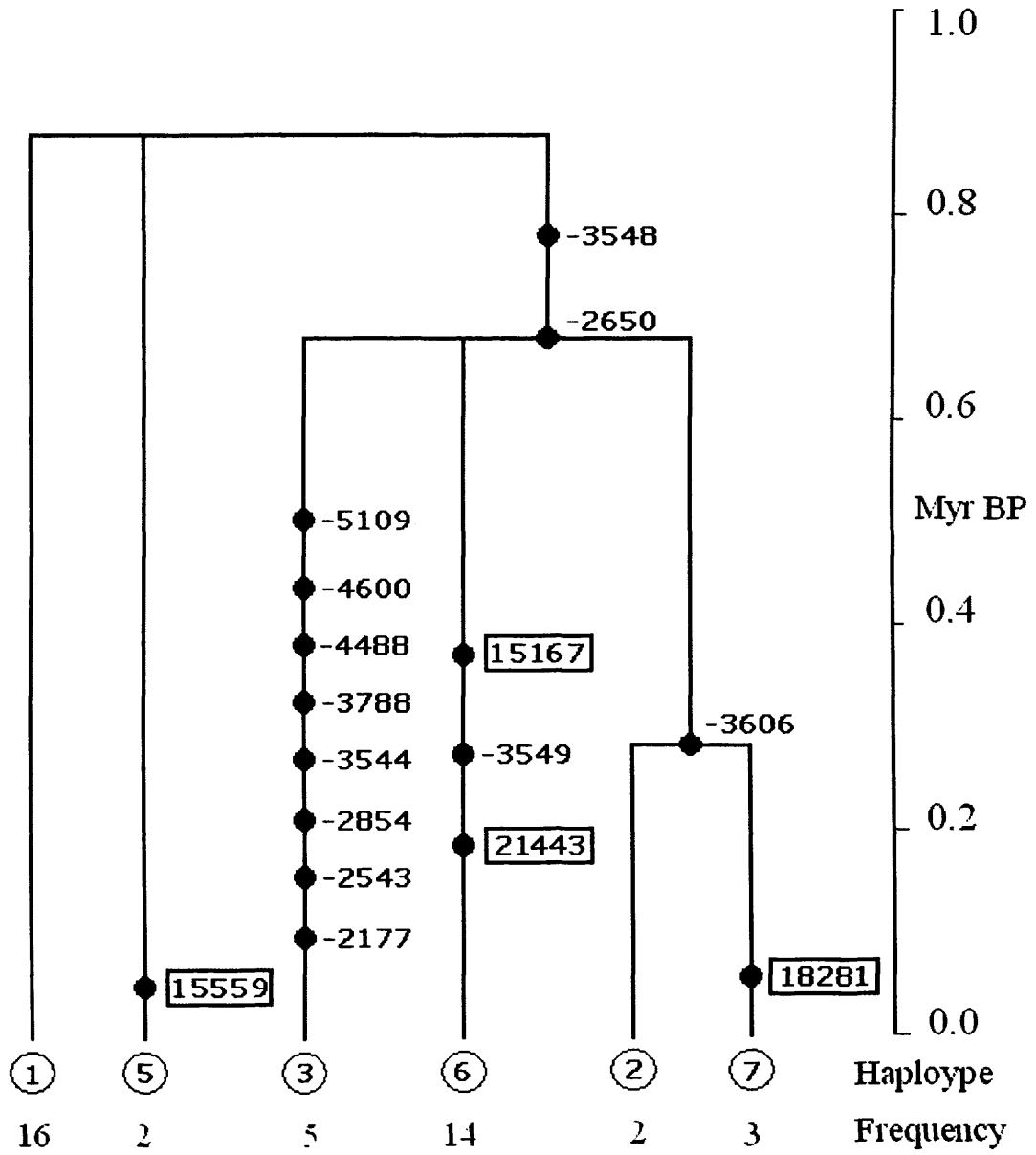


Figure 34. A genetree for FMO3, estimated through maximum-likelihood coalescent simulation

The tree shows the inferred genealogical history of the 6 haplotypes compatible with the infinite sites model of mutation. Mutations in the 5'-flanking region of the gene are identified by their nucleotide position relative to the A of the ATG translational initiation codon. Coding-region mutations leading to an amino acid change, 15167 (E158K), 15549 (R205C), 18281 (V257M) and 21443 (E308G), are boxed. Where branches contain multiple mutations, with the exception of 15167 (see text), the order of mutations through time is arbitrary. The scale is presented as millions of years before present (Myr BP).

total time depth of the tree, derived from the maximum-likelihood coalescent analysis, was calculated as 2.08 in units of $2N_e$ generations. This value was substituted using the estimate of N_e and assuming 20 years per generation, to years. This corresponds to 872 thousand years (kyr) BP (SD = 226 kyr BP). This corresponds closely to the coalescence time expected for a neutrally evolving human nuclear locus, assuming random mating and a long-term constant effective population size, i.e. $4N_e$ generations ago [279]. In a coalescent model assuming constant effective population size, average pairwise divergence is dominated by the time taken for the last two lineages to coalesce, which is expected to take $2N_e$ generations, or half the total time back to the most recent common ancestral sequence. From the net sequence divergence between human and chimp *FMO3* genes (47.31) and the average pairwise sequence difference among human haplotypes in the same region (3.32), the average pairwise age of sequence diversity is estimated as $6 \times 10^6 \times 3.32/47.31 = 421$ kyr BP. This indicated a T_{MRCA} of 842 kyr BP.

The maximum-likelihood coalescent analysis of GENETREE also enables estimation of the ages of individual mutations and the coalescent times of various haplotypes. The coalescent time for haplotypes 2, 3, 6 and 7, which corresponds to the point at which these evolutionary lineages coalesce to a common ancestor, is estimated as 674 kyr BP (SD = 182 kyr BP). Similarly, the coalescent time of haplotypes 2 and 3 is estimated as 283 kyr BP. In cases where multiple mutations have occurred along a single branch, e.g. that leading to haplotype 3, the order in which they arose is unknown and consequently the dates of individual mutations can not be estimated. However, in the case of the branch leading to haplotype 6, which contains three mutations, the single recombination event that, from the RM network (Figure 33) and the four gamete test, was predicted to give rise to haplotype 4 (see above), indicates that the mutation at 158 predated those at –3549 and 308. This allows the SNP at 158 (Gly > Lys) to be dated. The mutation

is estimated to have occurred 371 kyr BP (SD = 132 kyr BP). Although the order in which mutations at −3549 and 308 occurred is not known, the SNP at 308 (Glu > Gly) is estimated to have occurred some 200 kyr BP and thus it, too, is relatively old. The SNPs at 257 (Val > Met) and 205 (Arg > Cys) occurred more recently, with estimated ages of 62 kyr (SD = 52 kyr) and 52 kyr (SD = 59 kyr), respectively.

Evidence for balancing selection at the FMO3 locus

To examine the possibility that balancing selection, suggested by the relatively long branch lengths separating intermediate-frequency haplotypes on both the RM network (Figure 33) and the genetree (Figure 34), has acted on the *FMO3* locus, several tests were performed. These tests fell into four classes based on (1) the frequency spectrum of alleles within the *FMO3* locus, (2) the number and diversity of haplotypes in the sample, (3) LD and finally (4) interspecific comparison of sequence variation. To address the possibility of bias that might arise from analysis of sequences derived from potential TMAU sufferers, all tests were also performed independently, on 3,356 bp of 5 –flanking region of FMO3 from 45 unaffected Japanese individuals.

A summary of the tests carried out on the *FMO3* population samples is presented in Table 12.

Allele frequency spectrum

Under neutrality, and assuming random mating and a constant population size, the two estimates of nucleotide variation, θ_W and π, should be equal and Tajima's *D*, which compares these values, should be zero. The two neutrality tests are not equal in both the potential TMAU group and the control group leading to a positive Tajima's *D* value.

Fu and Li's F^* and D^* statistics were used to compare the observed number of singleton polymorphic sites with that expected under a neutral model. Fu and Li's F^* and D^* were positive for both the potential TMAU sufferers ($P > 0.05$) and the control group ($P > 0.001$ and $P > 0.05$, respectively).

Wrights fixation index, F_{ST}, which measures population subdivision by comparing genetic diversity within subpopulations to that of the whole population, was used to analyse the *FMO3* SNPs. Only Japanese individuals were sequenced in this investigation, but SNP coverage in the genotyped panels of dbSNP contained details of four of the SNPs featured in the Japanese group, -3606, -2854, 15167 (E158K) and 18281 (V257M), in panels of European, Asian and African origin. The average F_{ST} value four the four SNPs among these populations were found to be 0.050.

The allele frequency spectrum (Figure 35 A) showed an excess of intermediate-frequency SNPs compared to the number expected under a neutral model of evolution.

The Number and diversity of haplotypes

The number and diversity of haplotypes was examined to investigate departures from neutrality. Fu's F_S statistic was used to compare the observed number of sequence haplotypes to the number expected under the assumption of an infinite-sites model of neutral mutation with no recombination. The significantly positive value of F_S ($P > 0.02$) (Table 12) indicated an excess of intermediate-frequency haplotypes.

As mentioned earlier, analysis of DNA sequences from the potential TMAU sufferers indicated that the sample has undergone a single recombination event. When coalescent simulations were performed for the level of recombination observed in this data set, the number of haplotypes and F_S for the TMAU group were significantly different from those expected under neutrality ($P < 0.05$)

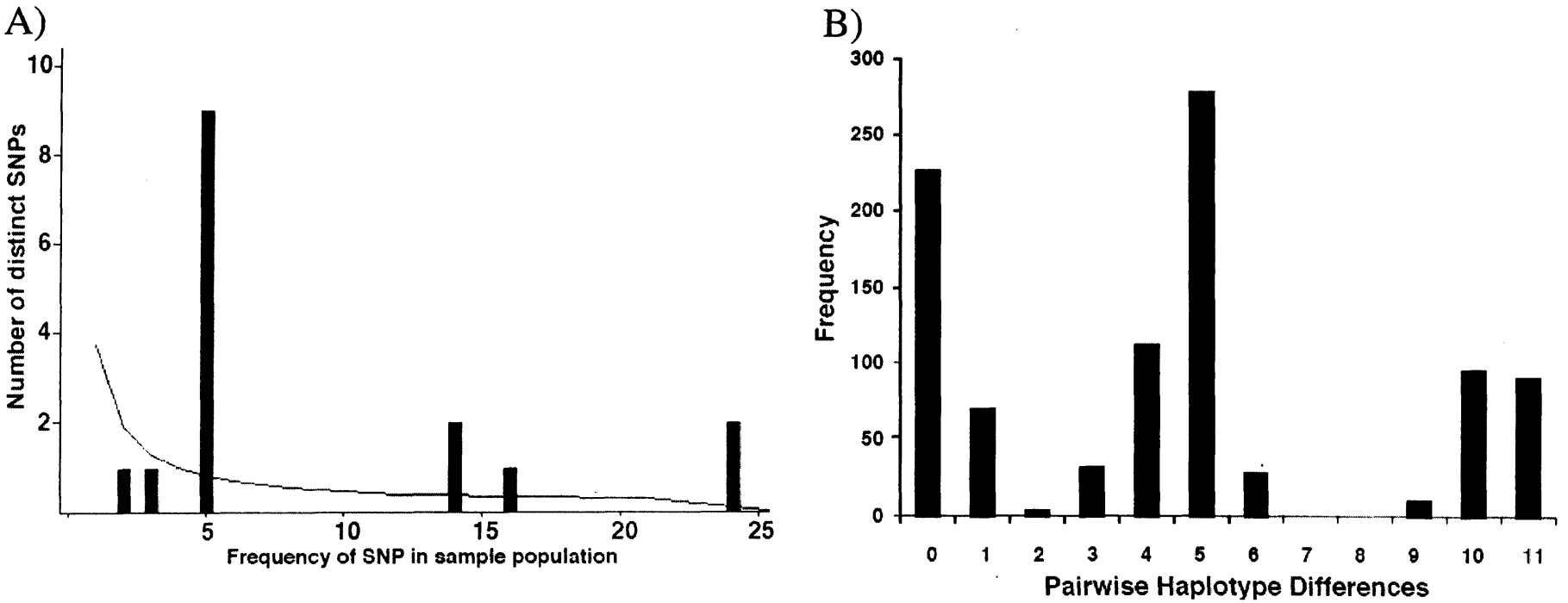


Figure 35. SNP frequency spectrum and Pairwise mismatch distribution histograms

A) A frequency spectrum of SNP alleles. Bars represent the observed frequency of SNP alleles and the line indicates the frequency expected under a model of neutral evolution.
B) Pairwise mismatch distribution among *FMO3* haplotypes. Bars show the observed frequencies of the numbers of differences between all pairs of the 44 haplotypes in the Japanese potential TMAU sufferers.

The pairwise difference distribution among FMO3 haplotypes was examined using the Rogers' mismatch distribution test [179]. The result showed a multimodal, also known as ragged, distribution (Figure 35 B) with a raggedness index of 0.16 for the potential TMAU sufferer group and 0.37 ($P < 0.02$) for the control group (Table 12).

Linkage disequilibrium

Z_{nS} [181], Wall's *B* and Wall's *Q* [180], were used to assess whether the strong LD which occurred throughout the *FMO3* locus (Figure 32) was consistent with a neutral model of evolution. In the control group, values of Z_{nS}, Wall's *B* and Wall's *Q* were all significantly higher than expected under neutrality (Table 12). Z_{nS} and Wall's *B* for the potential TMAU sufferers was significantly high when compared with estimates obtained from simulations performed for the level of recombination observed in this group ($P < 0.05$).

Inter-specific comparisons of the *FMO3* locus

Comparison of the sequenced regions of the human and chimpanzee (Pan troglodytes) (Ensembl [280] genebuild (August 2006 PanTro 1.0)) *FMO3* locus revealed 49 fixed differences. Eight of these were in coding regions of which two were non-synonymous variants (247G>A in exon four generating a R116H variant and 1421A>T in exon 9 generating a D474V variant). The average nucleotide divergence per silent site between human and chimpanzee within the sequenced regions of the *FMO3* locus was found to be 0.010.

The McDonald-Kreitman test, a test of neutrality that is not dependent on assumptions about population history, was used to compare the ratio of replacement to silent polymorphisms within humans with the ratio of replacement to silent fixed differences between humans and chimpanzee (Table 14). The test indicated that the

Table 14. McDonald-Kreitman test of neutrality

	Silent	Replacement	Fisher's exact test
Fixed	47	2	P<0.05
Polymorphic	12	4	

Numbers of polymorphisms and fixed difference founds at both silent and replacement (those that change amino acid residue) sites in the 44 chromosomes are contrasted with a 2 x 2 Fisher's exact test of independence.

number of replacement SNPs in human *FMO3* is significantly more than would be expected under a neutral model of evolution ($P > 0.05$).

Evolution of the *FMO3* locus - Discussion

FMO3 nucleotide sequence variation

There was a marked paucity of low-frequency variance with regards to the SNPs present in the population sample of *FMO3*. The finding that of the seven mild TMAU sufferers, six have a haplotype pair that is the same as present in unaffected individuals is perhaps not as surprising when one considers that those six individuals had haplotype 1 as one of the haplotypes. This haplotype is the ancestral haplotype and as such can be considered to code for a wild-type FMO3. TMAU is a recessive condition, where severely affected individuals have both of their FMO3 alleles containing mutations leading to perturbation or destruction of FMO3 activity. The individuals in this study were assessed biochemically and diagnosed as having borderline or mild TMAU (Table 15).

Evidence that the sample has undergone only one recombination event (see 3.4.) is indicative of strong linkage disequilibrium (LD).

Evolutionary History of *FMO3*

The presence of homoplasic sites is indicative either of recurrent mutations or recombination events. Haplotype 4 may have arisen via a recurrent mutation of the ancestral haplotype at position 158. However, the mutation at the 15167 (G>A) (encoding amino acid 158) site is not a CpG site and thus would not be subject to a markedly higher rate of mutation. A more likely origin is via a single recombination event that took place between sites -2650 and 15167 on an ancestral haplotype and a haplotype corresponding to

Table 15. Genotype and biochemical phenotype of individuals potentially suffering from TMAU

Genotype (haplotype pair)	Total TMA (mmol/mol creatinine)	Free TMA (mmol/mol creatinine)	TMA *N*-oxide (% of total TMA)	Diagnosis
1-1	123.7	4.7	96	Unaffected
1-1	88.7	11.0	88	Mild
1-2	70.1	12.9	82	Mild
1-2	44.0	2.6	94	Unaffected
1-3	71.7	21.0	71	Mild
1-3	69.0	4.0	94	Unaffected
1-5	60.7	1.8	97	Unaffected
1-5	85.1	22.2	74	Mild
1-6	127.5	13.0	90	Borderline
1-6	32.8	1.6	95	Unaffected
1-6	156.0	0.8	99	Unaffected
1-6	70.0	14.0	80	Mild
1-6	36.9	1.0	97	Unaffected
1-6	110.0	1.1	99	Unaffected
3-4	116.5	10.6	91	Borderline
3-6	64.6	0.8	99	Unaffected
3-6	157.1	2.2	99	Unaffected
4-7	121.2	16.0	87	Mild
6-6	119.4	13.0	89	Mild
6-6	109.8	10.1	91	Borderline
6-7	117.0	6.2	95	Unaffected
6-7	49.1	0.5	99	Unaffected

Free TMA and TMA *N*-oxide were measured in the first void urine of the morning of each individual. The haplotype pair of each individual is shown along with the diagnostic urine analysis, which was carried out by the Yamazaki Group, Showa Pharmaceutical University, Japan.

the node located on the branch linking the central node to haplotype 6. The haplotype corresponding to this node would have contained a mutation at site 158, but not sites -3549 or 21448 (encoding amino acid 308), indicating that the mutation at site 15167 (encoding amino acid 158) pre-dated the occurrence of the latter two mutations. The recombination event inferred from the RM network corresponds to that predicted by the four-gamete test (see above). Data from dbSNP [230], which indicate that the SNP at codon 158 occurs more frequently than that at codon 308, supports the conclusion that the former mutation occurred before the latter.

Time depth

The net sequence divergence was estimated as 47.31. This gave a neutral mutation rate (ν) of 7.78×10^{-10} per site per year which is similar to estimates for many other human nuclear loci [281]. The effective population size (N_e) was calculated, from the relationship $\theta = 4N_e\mu$, as 10,463 (very close to the average value determined from other genes [281-286]). The calculated value of the T_{MRCA} of 842 kyr BP from the coalescent model was very close to the estimate obtained from the maximum-likelihood coalescent analysis of 872 thousand years (kyr) BP (S.D. = 226 kyr BP). Thus two independent methods of calculation give very similar estimates for the time-depth of the *FMO3* gene tree, which correspond to that expected for nuclear genes. Thus most of the mutations in the sample, including the SNPs at 158 and 308, probably arose before the origin of modern humans some 150-200 kyr ago [281, 287] and almost all probably predated the migration out of Africa, estimated to have occurred 50-100 kyr ago [281, 287]. Of the identified SNPs that are present in dbSNP (-3606, -2854, 158 and 257), all occur in African, European and Asian populations, as does the SNP at 308 [248], which supports their pre-out-of-Africa origin. The estimated age of

the most recent mutation in this study at nucleotide 15549 (R205C) indicated that it may have arisen after migration out of Africa and thus may be restricted to Eurasian or Asian populations. This is supported by the presence of the mutation in Asian, but not African or European, populations.

Tests for departure from neutrality – a case for balancing selection

A higher value of θ_W and π, for both the potential TMAU sufferers and control groups (Table 12) and, consequently, a positive Tajima's *D*, implies an excess of moderate frequency compared to low-frequency variants, though the excess is not statistically significant.

Positive Fu and Li's *F** and *D** values indicate a significantly high proportion of 'old' versus 'young' mutations, suggesting balancing selection or population subdivision. The F_{ST} value of 0.050 among the populations in the dbSNP panels for the four SNPs covered indicated that only 5% of the observed variation was due to differences among populations. This value is lower than that typically observed for nuclear genes ~0.15 [281, 288-290], and is lower than the values reported for 80% of 25,549 SNPs [291]. This trend of FMO3 haplotype diversity has been reported recently in African Americans, non-Latino whites and Hispanics of Mexican descent [60]. The low F_{ST} values estimated for the *FMO3* SNPs suggests that little genetic differentiation of *FMO3* has occurred among continental populations and suggests that the positive Fu and Li's *D** and *F** are more likely due to balancing selection than to population subdivision.

The excess of intermediate-frequency SNPs, in both the patient and control groups, compared to the number expected under a neutral model of evolution are suggestive of balancing selection. Most genes studied to date differ from this in that they contain an excess of rare alleles, which has been interpreted as either being due to positive selection or

population growth [292]. The significantly positive value of Fu's F_S (P>0.02) indicated an excess of intermediate-frequency haplotypes, supporting the earlier findings that this may be due to balancing selection acting on the *FMO3* locus. This paucity of haplotypes was significant (P>0.01), considering the extent of nucleotide diversity of the sample, which futher supports this interpretation.

The multimodel distribution observed by the Rogers' mismatch test and Harpending's Raggedness test is typical for a distribution observed when balancing selection is in effect or where the population size has remained constant over a long period [179]. This is in contrast to a unimodel, also known as smooth, distribution, which is typical of a population growth or directional selection [179] and which is exhibited by the majority of genes studied to date [292].

The results of significant Z_{nS}, Wall's B and Wall's Q tests are compatible with balancing selection acting on the *FMO3* locus, which would be expected to result in LD among selectively neutral polymorphic sites closely linked to the sites under selection.

The average nucleotide divergence per silent site between human and chimpanzee within the sequenced regions of the *FMO3* locus was consistent with that of other genes [191, 288, 293].

The McDonald-Kreitman test, indicating that the number of replacement SNPs in human *FMO3* is significantly more than would be expected under a neutral model of evolution, could be construed as balancing selection or slightly deleterious selection. In the latter case, although purifying selection would prevent a slightly deleterious amino acid variant from attaining a high frequency, it may be weak enough to allow the variant to be maintained at a low frequency. However, replacement polymorphisms of *FMO3* are not all rare, with two of the four having a moderate frequency (>30%), which would favour balancing rather than slightly deleterious selection.

The analysis of the *FMO3* samples from the two cohorts of Japanese individuals generated results based on phylogenetics, allele frequency spectrum, the number, frequency and mismatch distribution of haplotypes, linkage disequilibrium and inter-specific sequence comparisons, which all provide evidence for the *FMO3* locus undergoing a period of balancing selection. The failure of several of the test statistics used in this investigation to detect selection where there is independent, compelling evidence of such (as in the cases of the β-globin [282] and Duffy blood group loci [294]) underscored the strength of the signal of detection for the departure from neutrality in the *FMO3* locus where significant values were attained for all tests. It has been shown that the vast majority of human genes investigated display negative Tajima's *D* [295], an allele frequency spectrum skewed to the left due to an excess of rare alleles [296] and a haplotype mismatch distribution that is unimodal [296], interpreted as evidence of the expansion of the human population, which would obscure any affects of balancing selection. Therefore, our results represent strong evidence for balancing selection at the FMO3 locus.

Balancing selection can maintain variation in the population for longer than would be expected under a neutral model of evolution. It is characterised by an excess of alleles at intermediate frequency and a phylogeny in which ≥ 2 lineages are separated by relatively large branching, giving the network an old coalescence. These characteristics have been presented describing the *FMO3* locus. The variants being maintained would be expected to confer a selective advantage in order to be maintained in a population, of which the determination of the identity and phenotypic consequences are extremely difficult to determine. In the case of the FMO3 locus, two haplotypes, 1 and 6, have been maintained at an intermediate frequency over long periods. Haplotype 1 is the ancestral form whereas Haplotype 6 contains three SNPs in the 5'-flanking region and two coding region SNPs (corresponding to amino acid variants E158K and E308G) (Figures 33 and 34). Both

coding region SNPs are relatively old, with ages estimated as 363,000 years and 200,000 years respectively (see section 3.4.). The coincidence of the two variants is interesting as they have been reported clinically and in the FMO community for several years. As mentioned earlier (section 3.2.), although E158K and E308G individually have little affect, when found in *cis* this study has shown (and others in the past have also shown) that this can result in a reduction in FMO3 activity [21, 147]. The FMO3 substrate metabolism shown to be affected by this double variant also mentioned earlier (see section 3.2., ranitidine [147], thiobenzamide [147] sulindac [5]), is of great interest. Of particular interest is the sulindac case, where the E158K/E308G variants were found to have a protective effect on the development of polyps in familial adenomatous polyposis patients who received sulindac as a means of primary chemoprevention, presumably reducing the ability of the enzyme to inactivate the drug, leading to prolonged exposure [260]. This instance gives rise to the possibility that the maintenance of haplotypes that encode versions of enzymes displaying substrate-dependent differences in catalytic activity may confer an evolutionary advantage, enabling polymorphic individuals to retain the ability to detoxify harmful xenobiotics while reducing their metabolic capacity to inactivate a natural compound that has a beneficial effect.

Although not present at such a high frequency as Haplotypes 1 and 6, Hapotype 3 coalesced with other haplotypes some 675,000 years ago. Evidence exists for three of the ten upstream SNPs that make up Haplotype 3, that their presence increases *FMO3* promotion 8-fold [141]. These SNPs (-2650, -2543 and -2177) presumably lie in the FMO3 promoter region, which is poorly defined at present. Sequences upstream of -2714 were not investigated by these authors, so the effect on promoter activity of the other SNPs in Haplotype 3 is not known at this time. However, it is still reasonable to speculate that individuals possessing Haplotype 3 would express greater amounts of FMO3 and thus, on

the one hand are able to detoxify harmful foreign chemicals more quickly but on the other hand also metabolise therapeutic agents that are substrates of this enzyme more quickly, possibly lowering drug efficacy and thus therapeutic benefit.

Chapter 4

Summary

4. Summary

Evolution is shaped by differences at a genetic level between and among species, with SNPs contributing a significant potion of these differences. This investigation has considered the SNPs of *FMO3* and their importance to the molecule in terms if its evolution, biophysical properties, structure and catalytic efficiency. This is of importance to human health not only because of the importance of FMO3 in clearing potentially harmful xenobiotics from the body, but also because of the variation in SNP and allelic frequencies of the haplotypes they make up, between populations. In the case of therapeutic drugs, the likely efficacy of such drugs in patients with perturbed xenobiotic metabolism would depend on the population with which you are dealing. On an individual case basis, the genotype of the individual could be used to predict the likely outcome of administering a drug. Such pharmacogenetic considerations are becoming more important in the development phase of drug-development [297, 298] and any information that can be learned in this area is of interest.

This study considered the rare mutations in *FMO3* that cause the disease TMAU. The distress and strain that this disorder puts on the individual and their loved ones can not be understated and research into this disease is important. This investigation presented patient studies including one family study which resulted in the discovery of two novel variants of FMO3, R492Q and R238Q. R238Q was generated, assessed *in vitro* for catalytic efficiency and found to be catalytically inactive.

This study also considered the relevance of commonly occurring polymorphisms of FMO3. Of particular interest was of the occurrence of two common polymorphism in *cis*, E158K and E308G. When found together, these variants cause a significant reduction in enzyme efficiency, when compared to when found alone and also compared to E158K/V257M, another relatively common polymorphism pairing found in *cis*. The

outcome of the kinetic investigation of this study supported the consequences of the double mutant. This is of pharmacogenetic interest due to the case of the E158K/E308G variant having a protective effect on the development of polyps in familial adenomatous polyposis patients who received sulindac as a chemotherapy agent, presumably by reducing the ability of FMO3 to catalyse the metabolism of sulindac *S*-sulphide to the sulindac *S*-sulphoxide thus reducing the capacity of FMO3 to inactivate the drug. The maintenance of such a seemingly destructive haplotype in the population at relatively high frequencies is consistent with this investigation's proposal that *FMO3* is under balancing selection, presented in section 3.4., along with an in-depth evolutionary study of the *FMO3* locus.

To try and understand the mechanistic consequences of the amino acid changes caused by these common SNPs and to understand the FMO3 molecule as a whole, a homology model was generated. Figure 36 shows the homology model with the variants tested for catalytic activity highlighted. The physical proximity of the amino acid residues at sites 158 and 308 compared to that of 158 and 257 may offer some kind of clue as to why the consequences of there being variations in the amino acids at both sites have such different consequences. Both polymorphisms have been shown to cause limited catalytic change individually, but when both are changed to K158 and G308, it could be imagined that the physical proximity of the two residues would become critically close to a threshold that a double mutation breaches, lowering catalytic activity. Why the double variant E158K/V257M seemingly caused a significant lowering in catalytic activity is not obvious. The two residues are extremely far apart on the model generated in this study. The exact nature of the V257 residue's contribution is difficult to show from this model as this area was of low coverage in the template used to build the model. The same can be said of R238. Why this residue, when changed to R238Q, seems to destroy catalytic activity, shown *in vivo* and *in vitro* in this investigation, is not clear.

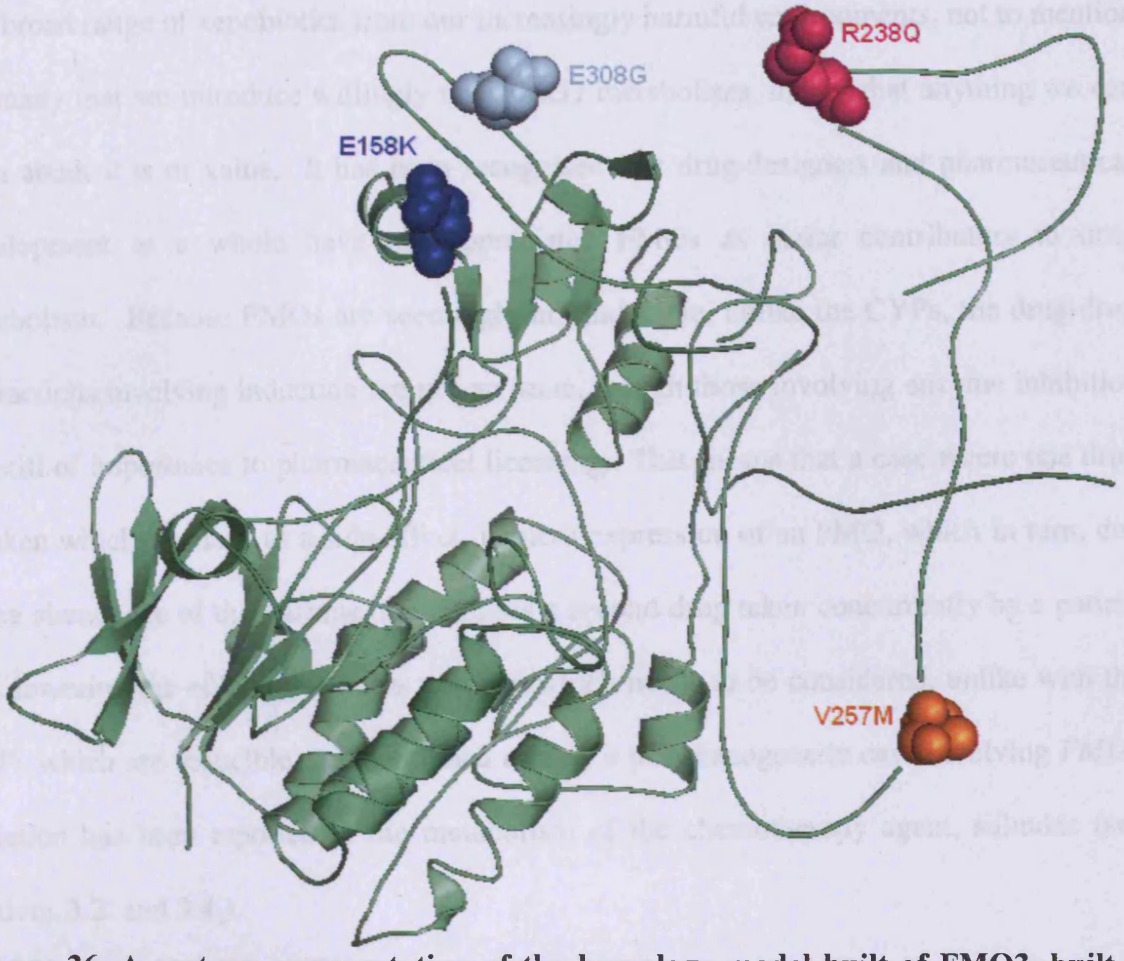


Figure 36. A cartoon representation of the homology model built of FMO3, built as part of this investigation

The FMO3 polymorphisms E158K (blue), V257M (orange) and E308G (cyan), generated and probed as part of a kinetic study (section 3.2.), featured within both an evolutionary (section 3.4.) and TMAU patient screen (section 3.1.) are highlighted as spheres. Also highlighted as spheres is the R238Q variant (pink) which is a novel FMO3 variant described in section.3.1. and also investigated by kinetic studies, in section 3.2.

FMO3 is a fascinating molecule. Such an important molecule, found at abundant levels in the liver, yet with so much ancient variation at a genetic level, contributes to a diverse genotypic and phenotypic distribution spread among the populations of the planet. The broad range of xenobiotics from our increasingly harmful environments, not to mention the many that we introduce willingly that FMO3 metabolises, means that anything we can learn about it is of value. It has been recognised that drug-designers and pharmaceutical development as a whole have not appreciated FMOs as major contributors to drug metabolism. Because FMOs are seemingly not inducible, unlike the CYPs, the drug-drug interactions involving induction are not an issue, though those involving enzyme inhibition are still of importance to pharmaceutical licensing. This means that a case where one drug is taken which, perhaps as a side-effect, induces expression of an FMO, which in turn, due to the abundance of the enzyme, metabolises a second drug taken concurrently by a patient thus lowering the efficacy, is not a scenario which needs to be considered, unlike with the CYPs which are inducible. As mentioned earlier, a pharmacogenetic case involving *FMO3* variation has been reported in the metabolism of the chemotherapy agent, sulindac (see sections 3.2. and 3.4.).

Pharmacogenetic information has the potential to improve therapeutic efficacy as we learn more about the genetic contribution to drug metabolism. Whether this potential is realised remains to be seen.

Chapter 5

References

1. Ziegler, D.M., *Recent studies on the structure and function of multisubstrate flavin-containing monooxygenases.* Annu Rev Pharmacol Toxicol, 1993. **33**: p. 179-99.
2. Prough, R.A., M.A. Stalmach, P. Wiebkin, and J.W. Bridges, *The microsomal metabolism of the organometallic derivatives of the group-IV elements, germanium, tin and lead.* Biochem J, 1981. **196**(3): p. 763-70.
3. Smyser, B.P. and E. Hodgson, *Metabolism of phosphorus-containing compounds by pig liver microsomal FAD-containing monooxygenase.* Biochem Pharmacol, 1985. **34**(8): p. 1145-50.
4. Jones, K.C. and D.P. Ballou, *Reactions of the 4a-hydroperoxide of liver microsomal flavin-containing monooxygenase with nucleophilic and electrophilic substrates.* J Biol Chem, 1986. **261**(6): p. 2553-9.
5. Hamman, M.A., B.D. Haehner-Daniels, S.A. Wrighton, A.E. Rettie, and S.D. Hall, *Stereoselective sulfoxidation of sulindac sulfide by flavin-containing monooxygenases. Comparison of human liver and kidney microsomes and mammalian enzymes.* Biochem Pharmacol, 2000. **60**(1): p. 7-17.
6. Ziegler, D.M., P. Graf, L.L. Poulsen, W. Stahl, and H. Sies, *NADPH-dependent oxidation of reduced ebselen, 2-selenylbenzanilide, and of 2-(methylseleno)benzanilide catalyzed by pig liver flavin-containing monooxygenase.* Chem Res Toxicol, 1992. **5**(2): p. 163-6.
7. Ziegler, D., *Enzymatic Basis of Detoxification*, ed. W. Jakoby. 1980, New York: Academic Press. 201-277.
8. Duescher, R.J., M.P. Lawton, R.M. Philpot, and A.A. Elfarra, *Flavin-containing monooxygenase (FMO)-dependent metabolism of methionine and evidence for FMO3 being the major FMO involved in methionine sulfoxidation in rabbit liver and kidney microsomes.* J Biol Chem, 1994. **269**(26): p. 17525-30.
9. Phillips, I.R., C.T. Dolphin, P. Clair, M.R. Hadley, A.J. Hutt, R.R. McCombie, R.L. Smith, and E.A. Shephard, *The molecular biology of the flavin-containing monooxygenases of man.* Chem Biol Interact, 1995. **96**(1): p. 17-32.
10. Shephard, E.A., C.T. Dolphin, M.F. Fox, S. Povey, R. Smith, and I.R. Phillips, *Localization of genes encoding three distinct flavin-containing monooxygenases to human chromosome 1q.* Genomics, 1993. **16**(1): p. 85-9.
11. Hernandez, D., A. Janmohamed, P. Chandan, I.R. Phillips, and E.A. Shephard, *Organization and evolution of the flavin-containing monooxygenase genes of human and mouse: identification of novel gene and pseudogene clusters.* Pharmacogenetics, 2004. **14**(2): p. 117-30.
12. Lawton, M.P., J.R. Cashman, T. Cresteil, C.T. Dolphin, A.A. Elfarra, R.N. Hines, E. Hodgson, T. Kimura, J. Ozols, I.R. Phillips, et al., *A nomenclature for the mammalian flavin-containing monooxygenase gene family based on amino acid sequence identities.* Arch Biochem Biophys, 1994. **308**(1): p. 254-7.
13. Wierenga, R., M. De Maeyer, and W. Hol, *Interactions of Pyrophospate Moieties with alpha-Helixes in Dinucleotide Binding Proteins.* Biochemistry, 1985. **24**: p. 1346-1357.
14. Dolphin, C.T., T.E. Cullingford, E.A. Shephard, R.L. Smith, and I.R. Phillips, *Differential developmental and tissue-specific regulation of expression of the genes encoding three members of the flavin-containing monooxygenase family of man, FMO1, FMO3 and FM04.* Eur J Biochem, 1996. **235**(3): p. 683-9.
15. Shehin-Johnson, S.E., D.E. Williams, S. Larsen-Su, D.M. Stresser, and R.N. Hines, *Tissue-specific expression of flavin-containing monooxygenase (FMO) forms 1 and 2 in the rabbit.* J Pharmacol Exp Ther, 1995. **272**(3): p. 1293-9.

16. Janmohamed, A., C.T. Dolphin, I.R. Phillips, and E.A. Shephard, *Quantification and cellular localization of expression in human skin of genes encoding flavin-containing monooxygenases and cytochromes P450.* Biochem Pharmacol, 2001. **62**(6): p. 777-86.
17. Janmohamed, A., D. Hernandez, I.R. Phillips, and E.A. Shephard, *Cell-, tissue-, sex- and developmental stage-specific expression of mouse flavin-containing monooxygenases (Fmos).* Biochem Pharmacol, 2004. **68**(1): p. 73-83.
18. Ziegler, D.M., L.L. Poulsen, and E.M. McKee, *Interaction of primary amines with a mixed-function amine oxidase isolated from pig liver microsomes.* Xenobiotica, 1971. **1**(4): p. 523-31.
19. Dolphin, C., E.A. Shephard, S. Povey, C.N. Palmer, D.M. Ziegler, R. Ayesh, R.L. Smith, and I.R. Phillips, *Cloning, primary sequence, and chromosomal mapping of a human flavin-containing monooxygenase (FMO1).* J Biol Chem, 1991. **266**(19): p. 12379-85.
20. Stevens, J.C., R.J. Melton, M.J. Zaya, and L.C. Engel, *Expression and characterization of functional dog flavin-containing monooxygenase 1.* Mol Pharmacol, 2003. **63**(2): p. 271-5.
21. Lattard, V., C. Longin-Sauvageon, J. Lachuer, P. Delatour, and E. Benoit, *Cloning, sequencing, and tissue-dependent expression of flavin-containing monooxygenase (FMO) 1 and FMO3 in the dog.* Drug Metab Dispos, 2002. **30**(2): p. 119-28.
22. Gasser, R., R.E. Tynes, M.P. Lawton, K.K. Korsmeyer, D.M. Ziegler, and R.M. Philpot, *The flavin-containing monooxygenase expressed in pig liver: primary sequence, distribution, and evidence for a single gene.* Biochemistry, 1990. **29**(1): p. 119-24.
23. Lawton, M.P., R. Gasser, R.E. Tynes, E. Hodgson, and R.M. Philpot, *The flavin-containing monooxygenase enzymes expressed in rabbit liver and lung are products of related but distinctly different genes.* J Biol Chem, 1990. **265**(10): p. 5855-61.
24. Koukouritaki, S.B., P. Simpson, C.K. Yeung, A.E. Rettie, and R.N. Hines, *Human hepatic flavin-containing monooxygenases 1 (FMO1) and 3 (FMO3) developmental expression.* Pediatr Res, 2002. **51**(2): p. 236-43.
25. Yeung, C.K., D.H. Lang, K.E. Thummel, and A.E. Rettie, *Immunoquantitation of FMO1 in human liver, kidney, and intestine.* Drug Metab Dispos, 2000. **28**(9): p. 1107-11.
26. Shephard, E.A., P. Chandan, M. Stevanovic-Walker, M. Edwards, and I.R. Phillips, *Alternative promoters and repetitive DNA elements define the species-dependent tissue-specific expression of the FMO1 gene of human and mouse.* Biochem J, 2007. **406**(3): p. 491-9.
27. Krause, R.J., K.M. Stettler, L.H. Lash, and A.A. Elfarra, *Human Kidney Flavin-Containing Monooxygenases (FMOs) and Their Potential Roles in Cysteine Conjugate Metabolism and Nephrotoxicity.* FASEB Journal, 2002. **16**: p. A574.
28. Shimada, T., H. Yamazaki, M. Mimura, Y. Inui, and F.P. Guengerich, *Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians.* J Pharmacol Exp Ther, 1994. **270**(1): p. 414-23.
29. Jakobsson, S.V. and D.L. Cintig, *Studies on the cytochrome P-450-containing mono-oxygenase system in human kidney cortex microsomes.* J Pharmacol Exp Ther, 1973. **185**(2): p. 226-34.
30. Nikbakht, K.N., M.P. Lawton, and R.M. Philpot, *Guinea pig or rabbit lung flavin-containing monooxygenases with distinct mobilities in SDS-PAGE are allelic variants that differ at only two positions.* Pharmacogenetics, 1992. **2**(5): p. 207-16.

31. Yueh, M.F., S.K. Krueger, and D.E. Williams, *Pulmonary flavin-containing monooxygenase (FMO) in rhesus macaque: expression of FMO2 protein, mRNA and analysis of the cDNA.* Biochim Biophys Acta, 1997. **1350**(3): p. 267-71.
32. Dolphin, C.T., D.J. Beckett, A. Janmohamed, T.E. Cullingford, R.L. Smith, E.A. Shephard, and I.R. Phillips, *The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein.* J Biol Chem, 1998. **273**(46): p. 30599-607.
33. Brunet, M., F. Guy, D. Pilbeam, H.T. Mackaye, A. Likius, D. Ahounta, A. Beauvilain, C. Blondel, H. Bocherens, J.R. Boisserie, et al., *A new hominid from the Upper Miocene of Chad, Central Africa.* Nature, 2002. **418**(6894): p. 145-51.
34. Dolphin, C.T., E.A. Shephard, S. Povey, R.L. Smith, and I.R. Phillips, *Cloning, primary sequence and chromosomal localization of human FMO2, a new member of the flavin-containing mono-oxygenase family.* Biochem J, 1992. **287 (Pt 1)**: p. 261-7.
35. Dolphin, C.T., A. Janmohamed, R.L. Smith, E.A. Shephard, and I.R. Phillips, *Compound heterozygosity for missense mutations in the flavin-containing monooxygenase 3 (FM03) gene in patients with fish-odour syndrome.* Pharmacogenetics, 2000. **10**(9): p. 799-807.
36. Cashman, J.R. and J. Zhang, *Human flavin-containing monooxygenases.* Annu Rev Pharmacol Toxicol, 2006. **46**: p. 65-100.
37. Zhang, J. and J.R. Cashman, *Quantitative analysis of FMO gene mRNA levels in human tissues.* Drug Metab Dispos, 2006. **34**(1): p. 19-26.
38. Overby, L.H., A.R. Buckpitt, M.P. Lawton, E. Atta-Asafo-Adjei, J. Schulze, and R.M. Philpot, *Characterization of flavin-containing monooxygenase 5 (FMO5) cloned from human and guinea pig: evidence that the unique catalytic properties of FMO5 are not confined to the rabbit ortholog.* Arch Biochem Biophys, 1995. **317**(1): p. 275-84.
39. Rettie, A.E., M.P. Lawton, A.J. Sadeque, G.P. Meier, and R.M. Philpot, *Prochiral sulfoxidation as a probe for multiple forms of the microsomal flavin-containing monooxygenase: studies with rabbit FMO1, FMO2, FMO3, and FMO5 expressed in Escherichia coli.* Arch Biochem Biophys, 1994. **311**(2): p. 369-77.
40. Krueger, S.K. and D.E. Williams, *Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism.* Pharmacol Ther, 2005. **106**(3): p. 357-87.
41. Cherrington, N.J., J.G. Falls, R.L. Rose, K.M. Clements, R.M. Philpot, P.E. Levi, and E. Hodgson, *Molecular cloning, sequence, and expression of mouse flavin-containing monooxygenases 1 and 5 (FMO1 and FMO5).* J Biochem Mol Toxicol, 1998. **12**(4): p. 205-12.
42. Ohmi, N., H. Yoshida, H. Endo, M. Hasegawa, M. Akimoto, and S. Higuchi, *S-oxidation of S-methyl-esonarimod by flavin-containing monooxygenases in human liver microsomes.* Xenobiotica, 2003. **33**(12): p. 1221-31.
43. Lomri, N., Q. Gu, and J.R. Cashman, *Molecular cloning of the flavin-containing monooxygenase (form II) cDNA from adult human liver.* Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1685-9.
44. Overby, L.H., G.C. Carver, and R.M. Philpot, *Quantitation and kinetic properties of hepatic microsomal and recombinant flavin-containing monooxygenases 3 and 5 from humans.* Chem Biol Interact, 1997. **106**(1): p. 29-45.
45. Chung, W.G., C.S. Park, H.K. Roh, and Y.N. Cha, *Induction of flavin-containing monooxygenase (FMO1) by a polycyclic aromatic hydrocarbon, 3-methylcholanthrene, in rat liver.* Mol Cells, 1997. **7**(6): p. 738-41.

46. Kisanga, E.R., L.L. Moi, J. Gjerde, G. Mellgren, and E.A. Lien, *Induction of hepatic drug-metabolising enzymes and tamoxifen metabolite profile in relation to administration route during low-dose treatment in nude rats.* J Steroid Biochem Mol Biol, 2005. **94**(5): p. 489-98.
47. Yamazaki, H. and M. Shimizu, *Genetic Polymorphism of the Flavin-Containing Monooxygenase 3 (FMO3) Associated with Trimethylaminuria (Fish Odor Syndrome): Observations from Japanese Patients.* Curr Drug Metab, 2007. **8**(5): p. 487-91.
48. Zhang, A.Q., S.C. Mitchell, and R.L. Smith, *Exacerbation of symptoms of fish-odour syndrome during menstruation.* Lancet, 1996. **348**(9043): p. 1740-1.
49. Shimizu, M., J.R. Cashman, and H. Yamazaki, *Transient trimethylaminuria related to menstruation.* BMC Med Genet, 2007. **8**: p. 2.
50. Mitchell, S.C., *Trimethylaminuria: susceptibility of heterozygotes.* Lancet, 1999. **354**(9196): p. 2164-5.
51. Lang, D.H., C.K. Yeung, R.M. Peter, C. Ibarra, R. Gasser, K. Itagaki, R.M. Philpot, and A.E. Rettie, *Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes: selective catalysis by FMO3.* Biochem Pharmacol, 1998. **56**(8): p. 1005-12.
52. Cashman, J.R., K. Camp, S.S. Fakharzadeh, P.V. Fennessey, R.N. Hines, O.A. Mamer, S.C. Mitchell, G.P. Nguyen, D. Schlenk, R.L. Smith, et al., *Biochemical and clinical aspects of the human flavin-containing monooxygenase form 3 (FMO3) related to trimethylaminuria.* Curr Drug Metab, 2003. **4**(2): p. 151-70.
53. Humbert, J.A., K.B. Hammond, and W.E. Hathaway, *Trimethylaminuria: the fish-odour syndrome.* Lancet, 1970. **2**(7676): p. 770-1.
54. Forrest, S.M., M. Knight, B.R. Akerman, J.R. Cashman, and E.P. Treacy, *A novel deletion in the flavin-containing monooxygenase gene (FMO3) in a Greek patient with trimethylaminuria.* Pharmacogenetics, 2001. **11**(2): p. 169-74.
55. Zhang, J., Q. Tran, V. Lattard, and J.R. Cashman, *Deleterious mutations in the flavin-containing monooxygenase 3 (FMO3) gene causing trimethylaminuria.* Pharmacogenetics, 2003. **13**(8): p. 495-500.
56. Teresa, E., F. Lonardo, A. Fiumara, C. Lombardi, P. Russo, C. Zuppi, G. Scarano, S. Musumeci, and F. Gianfrancesco, *A spectrum of molecular variation in a cohort of Italian families with trimethylaminuria: identification of three novel mutations of the FM03 gene.* Mol Genet Metab, 2006. **88**(2): p. 192-5.
57. Mazon Ramos, A., A. Gil-Setas, S. Berrade Zubiri, T. Bandres Echeverri, R. Wevers, U. Engelke, and J. Zschocke, *[Primary trimethylaminuria or fish odor syndrome. A novel mutation in the first documented case in Spain].* Med Clin (Barc), 2003. **120**(6): p. 219-21.
58. Akerman, B.R., H. Lemass, L.M. Chow, D.M. Lambert, C. Greenberg, C. Bibeau, O.A. Mamer, and E.P. Treacy, *Trimethylaminuria is caused by mutations of the FMO3 gene in a North American cohort.* Mol Genet Metab, 1999. **68**(1): p. 24-31.
59. Kubota, M., Y. Nakamoto, K. Nakayama, P. Ujjin, S. Satarug, T. Mushiroda, T. Yokoi, M. Funayama, and T. Kamataki, *A mutation in the flavin-containing monooxygenase 3 gene and its effects on catalytic activity for N-oxidation of trimethylamine in vitro.* Drug Metab Pharmacokinet, 2002. **17**(3): p. 207-13.
60. Koukouritaki, S.B., M.T. Poch, M.C. Henderson, L.K. Siddens, S.K. Krueger, J.E. VanDyke, D.E. Williams, N.M. Pajewski, T. Wang, and R.N. Hines, *Identification and functional analysis of common human flavin-containing monooxygenase 3 genetic variants.* J Pharmacol Exp Ther, 2007. **320**(1): p. 266-73.

61. Cashman, J.R., Y.A. Bi, J. Lin, R. Youil, M. Knight, S. Forrest, and E. Treacy, *Human flavin-containing monooxygenase form 3: cDNA expression of the enzymes containing amino acid substitutions observed in individuals with trimethylaminuria.* Chem Res Toxicol, 1997. **10**(8): p. 837-41.
62. Akerman, B.R., S. Forrest, L. Chow, R. Youil, M. Knight, and E.P. Treacy, *Two novel mutations of the FMO3 gene in a proband with trimethylaminuria.* Hum Mutat, 1999. **13**(5): p. 376-9.
63. Murphy, H.C., C.T. Dolphin, A. Janmohamed, H.C. Holmes, H. Michelakakis, E.A. Shephard, R.A. Chalmers, I.R. Phillips, and R.A. Iles, *A novel mutation in the flavin-containing monooxygenase 3 gene, FM03, that causes fish-odour syndrome: activity of the mutant enzyme assessed by proton NMR spectroscopy.* Pharmacogenetics, 2000. **10**(5): p. 439-51.
64. Shimizu, M., S. Tomioka, N. Murayama, and H. Yamazaki, *Missense and Nonsense Mutations of the Flavin-containing Monooxygenase 3 Gene in a Japanese Cohort.* Drug Metab Pharmacokinet, 2007. **22**(1): p. 61-4.
65. Basarab, T., G.H. Ashton, H.P. Menage, and J.A. McGrath, *Sequence variations in the flavin-containing mono-oxygenase 3 gene (FMO3) in fish odour syndrome.* Br J Dermatol, 1999. **140**(1): p. 164-7.
66. Park, C.S., W.G. Chung, J.H. Kang, H.K. Roh, K.H. Lee, and Y.N. Cha, *Phenotyping of flavin-containing monooxygenase using caffeine metabolism and genotyping of FMO3 gene in a Korean population.* Pharmacogenetics, 1999. **9**(2): p. 155-64.
67. Dolphin, C.T., A. Janmohamed, R.L. Smith, E.A. Shephard, and I.R. Phillips, *Missense mutation in flavin-containing mono-oxygenase 3 gene, FMO3, underlies fish-odour syndrome.* Nat Genet, 1997. **17**(4): p. 491-4.
68. Yamazaki, H., H. Fujita, T. Gunji, J. Zhang, T. Kamataki, J.R. Cashman, and M. Shimizu, *Stop codon mutations in the flavin-containing monooxygenase 3 (FMO3) gene responsible for trimethylaminuria in a Japanese population.* Mol Genet Metab, 2006.
69. Fujieda, M., H. Yamazaki, M. Togashi, T. Saito, and T. Kamataki, *Two novel single nucleotide polymorphisms (SNPs) of the FMO3 gene in Japanese.* Drug Metab Pharmacokinet, 2003. **18**(5): p. 333-5.
70. Zschocke, J., D. Kohlmueller, E. Quak, T. Meissner, G.F. Hoffmann, and E. Mayatepek, *Mild trimethylaminuria caused by common variants in FMO3 gene.* Lancet, 1999. **354**(9181): p. 834-5.
71. Shimizu, M., H. Fujita, T. Aoyama, and H. Yamazaki, *Three novel single nucleotide polymorphisms of the FMO3 gene in a Japanese population.* Drug Metab Pharmacokinet, 2006. **21**(3): p. 245-7.
72. Preti, G. and J.J. Leyden, *Body odor in dermatologic diagnosis.* Cutis, 2002. **69**(4): p. 315; author reply 316.
73. Treacy, E.P., B.R. Akerman, L.M. Chow, R. Youil, C. Bibeau, J. Lin, A.G. Bruce, M. Knight, D.M. Danks, J.R. Cashman, et al., *Mutations of the flavin-containing monooxygenase gene (FMO3) cause trimethylaminuria, a defect in detoxication.* Hum Mol Genet, 1998. **7**(5): p. 839-45.
74. Todd, W.A., *Psychosocial problems as the major complication of an adolescent with trimethylaminuria.* J Pediatr, 1979. **94**(6): p. 936-7.
75. Mayatepek, E. and D. Kohlmuller, *Transient trimethylaminuria in childhood.* Acta Paediatr, 1998. **87**(11): p. 1205-7.

76. McConnell, H.W., S.C. Mitchell, R.L. Smith, and M. Brewster, *Trimethylaminuria associated with seizures and behavioural disturbance: a case report.* Seizure, 1997. **6**(4): p. 317-21.
77. Man Burrows, C., *Commentaries on insanity*, in *Madness and Morals. Ideas on Insanity in the 19th Century*, V. Skultanes, Editor. 1975, Rutlidge and Keegan Paul: London. p. 78.
78. Shakespeare, W., *The Tempest.* Barnes and Noble Shakespeare. 1603, New York: Barnes and Noble.
79. *Mahābhārata.* 2005, New York: New York Univ. Press
80. BBC3, *Help! I Smell Of Fish.* 2007, BBC3: UK. p. 1 hour.
81. Al-Waiz, M., R. Ayesh, S.C. Mitchell, J.R. Idle, and R.L. Smith, *Trimethylaminuria ('fish-odour syndrome'): a study of an affected family.* Clin Sci (Lond), 1988. **74**(3): p. 231-6.
82. Mitchell, S.C. and R.L. Smith, *Trimethylaminuria: the fish malodor syndrome.* Drug Metab Dispos, 2001. **29**(4 Pt 2): p. 517-21.
83. Hadidi, H.F., S. Cholerton, S. Atkinson, Y.M. Irshaid, N.M. Rawashdeh, and J.R. Idle, *The N-oxidation of trimethylamine in a Jordanian population.* Br J Clin Pharmacol, 1995. **39**(2): p. 179-81.
84. Mitchell, S.C., A.Q. Zhang, T. Barrett, R. Ayesh, and R.L. Smith, *Studies on the discontinuous N-oxidation of trimethylamine among Jordanian, Ecuadorian and New Guinean populations.* Pharmacogenetics, 1997. **7**(1): p. 45-50.
85. Yamazaki, H., M. Fujieda, M. Togashi, T. Saito, G. Preti, J.R. Cashman, and T. Kamataki, *Effects of the dietary supplements, activated charcoal and copper chlorophyllin, on urinary excretion of trimethylamine in Japanese trimethylaminuria patients.* Life Sci, 2004. **74**(22): p. 2739-47.
86. Lazarou, J., B.H. Pomeranz, and P.N. Corey, *Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies.* Jama, 1998. **279**(15): p. 1200-5.
87. Dolphin, C.T., J.H. Riley, R.L. Smith, E.A. Shephard, and I.R. Phillips, *Structural organization of the human flavin-containing monooxygenase 3 gene (FMO3), the favored candidate for fish-odor syndrome, determined directly from genomic DNA.* Genomics, 1997. **46**(2): p. 260-7.
88. Shaffer, C.L., M. Gunduz, R.J. Scialis, and A.F. Fang, *METABOLISM AND DISPOSITION OF A SELECTIVE {alpha}7 NICOTINIC ACETYLCHOLINE RECEPTOR AGONIST IN HUMANS.* Drug Metab Dispos, 2007.
89. Chiba, K., K. Kobayashi, K. Itoh, S. Itoh, T. Chiba, T. Ishizaki, and T. Kamataki, *N-oxygenation of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine by the rat liver flavin-containing monooxygenase expressed in yeast cells.* Eur J Pharmacol, 1995. **293**(1): p. 97-100.
90. Zhou, S., P. Kestell, and J.W. Paxton, *6-methylhydroxylation of the anti-cancer agent 5,6-dimethylxanthenone-4-acetic acid (DMXAA) by flavin-containing monooxygenase 3.* Eur J Drug Metab Pharmacokinet, 2002. **27**(3): p. 179-83.
91. Salva, M., J.M. Jansat, A. Martinez-Tobed, and J.M. Palacios, *Identification of the human liver enzymes involved in the metabolism of the antimigraine agent almotriptan.* Drug Metab Dispos, 2003. **31**(4): p. 404-11.
92. Szoko, E., T. Tabi, T. Borbas, B. Dalmadi, K. Tihanyi, and K. Magyar, *Assessment of the N-oxidation of deprenyl, methamphetamine, and amphetamine enantiomers by chiral capillary electrophoresis: an in vitro metabolism study.* Electrophoresis, 2004. **25**(16): p. 2866-75.

93. Giri, S., K.W. Krausz, J.R. Idle, and F.J. Gonzalez, *The metabolomics of (+/-)-arecoline 1-oxide in the mouse and its formation by human flavin-containing monooxygenases.* Biochem Pharmacol, 2006.
94. Lang, D.H. and A.E. Rettie, *In vitro evaluation of potential in vivo probes for human flavin-containing monooxygenase (FMO): metabolism of benzydamine and caffeine by FMO and P450 isoforms.* Br J Clin Pharmacol, 2000. **50**(4): p. 311-4.
95. Wu, R.F., C.X. Liao, S. Tomita, Y. Ichikawa, and L.S. Terada, *Porcine FAD-containing monooxygenase metabolizes lidocaine, bupivacaine and propranolol in vitro.* Life Sci, 2004. **75**(8): p. 1011-9.
96. Washio, T., K. Kohsaka, H. Arisawa, and H. Masunaga, *Pharmacokinetics and metabolism of the novel muscarinic receptor agonist SNI-2011 in rats and dogs.* Arzneimittelforschung, 2003. **53**(1): p. 26-33.
97. Yamada, H., K. Yuno, K. Oguri, and H. Yoshimura, *Multiplicity of liver microsomal flavin-containing monooxygenase in the guinea pig: its purification and characterization.* Arch Biochem Biophys, 1990. **280**(2): p. 305-12.
98. Tugnait, M., E.M. Hawes, G. McKay, A.E. Rettie, R.L. Haining, and K.K. Midha, *N-oxygenation of clozapine by flavin-containing monooxygenase.* Drug Metab Dispos, 1997. **25**(4): p. 524-7.
99. Vyas, P.M., S. Roychowdhury, F.D. Khan, T.E. Prisinzano, J. Lamba, E.G. Schuetz, J. Blaisdell, J.A. Goldstein, K.L. Munson, R.N. Hines, et al., *Enzyme-mediated protein haptenation of dapsone and sulfamethoxazole in human keratinocytes: I. Expression and role of cytochromes P450.* J Pharmacol Exp Ther, 2006. **319**(1): p. 488-96.
100. Jacobsen, W., U. Christians, and L.Z. Benet, *In vitro evaluation of the disposition of A novel cysteine protease inhibitor.* Drug Metab Dispos, 2000. **28**(11): p. 1343-51.
101. Rodriguez, R.J. and C.L. Miranda, *Isoform specificity of N-deacetyl ketoconazole by human and rabbit flavin-containing monooxygenases.* Drug Metab Dispos, 2000. **28**(9): p. 1083-6.
102. Sofer, S.S. and D.M. Ziegler, *Microsomal mixed-function amine oxidase. Oxidation products of piperazine-substituted phenothiazine drugs.* Drug Metab Dispos, 1978. **6**(3): p. 232-9.
103. Mushiroda, T., R. Douya, E. Takahara, and O. Nagata, *The involvement of flavin-containing monooxygenase but not CYP3A4 in metabolism of itopride hydrochloride, a gastroprokinetic agent: comparison with cisapride and mosapride citrate.* Drug Metab Dispos, 2000. **28**(10): p. 1231-7.
104. Ballard, J.E., T. Prueksaritanont, and C. Tang, *Hepatic metabolism of MK-0457, a potent Aurora kinase inhibitor: Interspecies comparison and role of human cytochrome P450 and flavin-containing monooxygenase.* Drug Metab Dispos, 2007.
105. Ereshefsky, L., *Pharmacokinetics and drug interactions: update for new antipsychotics.* J Clin Psychiatry, 1996. **57 Suppl 11**: p. 12-25.
106. Kajita, J., K. Inano, E. Fuse, T. Kuwabara, and H. Kobayashi, *Effects of olopatadine, a new antiallergic agent, on human liver microsomal cytochrome P450 activities.* Drug Metab Dispos, 2002. **30**(12): p. 1504-11.
107. Stormer, E., J. Brockmoller, I. Roots, and J. Schmider, *Cytochrome P-450 enzymes and FMO3 contribute to the disposition of the antipsychotic drug perazine in vitro.* Psychopharmacology (Berl), 2000. **151**(4): p. 312-20.
108. Lin, J. and J.R. Cashman, *N-oxygenation of phenethylamine to the trans-oxime by adult human liver flavin-containing monooxygenase and retroreduction of phenethylamine hydroxylamine by human liver microsomes.* J Pharmacol Exp Ther, 1997. **282**(3): p. 1269-79.

109. Reid, J.M.W., D.L. Miller, J.K. Benson, L.M. Tomlinson, A.J. Naylor, S. Blajeski, A.L. LoRusso, P.M. Ames, M.M., *The metabolism of pyrazoloacidine (NSC 366140) by cytochromes p450 and flavin monooxygenase in human liver microsomes.* Clin Cancer Res, 2004. **10**(4): p. 1471-80.
110. Chung, W.G., C.S. Park, H.K. Roh, W.K. Lee, and Y.N. Cha, *Oxidation of ranitidine by isozymes of flavin-containing monooxygenase and cytochrome P450.* Jpn J Pharmacol, 2000. **84**(2): p. 213-20.
111. Pichard-Garcia, L., R.J. Weaver, N. Eckett, G. Scarfe, J.M. Fabre, C. Lucas, and P. Maurel, *The olivacine derivative s 16020 (9-hydroxy-5,6-dimethyl-N-[2-(dimethylamino)ethyl)-6H-pyrido(4,3-B)-carbaz ole-1-carboxamide) induces CYP1A and its own metabolism in human hepatocytes in primary culture.* Drug Metab Dispos, 2004. **32**(1): p. 80-8.
112. Rodrigues, A.D., M.J. Kukulka, J.L. Ferrero, and J.R. Cashman, *In vitro hepatic metabolism of ABT-418 in chimpanzee (Pan troglodytes). A unique pattern of microsomal flavin-containing monooxygenase-dependent stereoselective N'-oxidation.* Drug Metab Dispos, 1995. **23**(10): p. 1143-52.
113. Cashman, J.R., *Human flavin-containing monooxygenase: substrate specificity and role in drug metabolism.* Curr Drug Metab, 2000. **1**(2): p. 181-91.
114. Mani, C., E. Hodgson, and D. Kupfer, *Metabolism of the antimammary cancer antiestrogenic agent tamoxifen. II. Flavin-containing monooxygenase-mediated N-oxidation.* Drug Metab Dispos, 1993. **21**(4): p. 657-61.
115. Kousba, A.A., R.M. Soll, S. Yee, and M.B. Martin, *Cyclic conversion of the Novel SRC Kinase Inhibitor TG100435 and its N-oxide metabolite by Flavin-containing Monoxygenases and Cytochromes P450 Reductase.* Drug Metab Dispos, 2007.
116. Lin, J. and J.R. Cashman, *Detoxication of tyramine by the flavin-containing monooxygenase: stereoselective formation of the trans oxime.* Chem Res Toxicol, 1997. **10**(8): p. 842-52.
117. Cashman, J.R., *Enantioselective N-oxygenation of verapamil by the hepatic flavin-containing monooxygenase.* Mol Pharmacol, 1989. **36**(3): p. 497-503.
118. Ring, B.J., S.A. Wrighton, S.L. Aldridge, K. Hansen, B. Haehner, and L.A. Shipley, *Flavin-containing monooxygenase-mediated N-oxidation of the M(1)-muscarinic agonist xanomeline.* Drug Metab Dispos, 1999. **27**(10): p. 1099-103.
119. Cashman, J.R., J. Proudfoot, D.W. Pate, and T. Hogberg, *Stereoselective N-oxygenation of zimeldine and homozimeldine by the flavin-containing monooxygenase.* Drug Metab Dispos, 1988. **16**(4): p. 616-22.
120. Nnane, I.P. and L.A. Damani, *Involvement of cytochrome P450 and the flavin-containing monooxygenase(s) in the sulphoxidation of simple sulphides in human liver microsomes.* Life Sci, 2003. **73**(3): p. 359-69.
121. Cashman, J.R. and S. Pena, *S-oxygenation of 7 alpha-thiomethylspironolactone by the flavin-containing monooxygenase.* Drug Metabol Drug Interact, 1988. **6**(3-4): p. 337-48.
122. Hajjar, N.P. and E. Hodgson, *Flavin adenine dinucleotide--dependent monooxygenase: its role in the sulfoxidation of pesticides in mammals.* Science, 1980. **209**(4461): p. 1134-6.
123. Rawden, H.C., G.O. Kokwaro, S.A. Ward, and G. Edwards, *Relative contribution of cytochromes P-450 and flavin-containing monoxygenases to the metabolism of albendazole by human liver microsomes.* Br J Clin Pharmacol, 2000. **49**(4): p. 313-22.
124. Henderson, M.C., S.K. Krueger, L.K. Siddens, J.F. Stevens, and D.E. Williams, *S-oxygenation of the thioether organophosphate insecticides phorate and disulfoton*

by human lung flavin-containing monooxygenase 2. Biochem Pharmacol, 2004. **68**(5): p. 959-67.

125. Henderson, M.C. *S-OXIDATION OF ETHIONAMIDE BY BACULOVIRUS-EXPRESSED HUMAN FMO2 AND FM03 AND SELECTED VARIANTS.* in *16th International Symposium on Microsomes and Drug Oxidations.* 2006. Budapest, Hungary.
126. Park, S.B., W.N. Howald, and J.R. Cashman, *S-oxidative cleavage of farnesylcysteine and farnesylcysteine methyl ester by the flavin-containing monooxygenase.* Chem Res Toxicol, 1994. **7**(2): p. 191-8.
127. Furnes, B. and D. Schlenk, *Evaluation of xenobiotic N- and S-oxidation by variant flavin-containing monooxygenase 1 (FMO1) enzymes.* Toxicol Sci, 2004. **78**(2): p. 196-203.
128. Furnes, B. and D. Schlenk, *Extrahepatic metabolism of carbamate and organophosphate thioether compounds by the flavin-containing monooxygenase and cytochrome P450 systems.* Drug Metab Dispos, 2005. **33**(2): p. 214-8.
129. Cashman, J.R., Y.N. Xiong, L. Xu, and A. Janowsky, *N-oxygenation of amphetamine and methamphetamine by the human flavin-containing monooxygenase (form 3): role in bioactivation and detoxication.* J Pharmacol Exp Ther, 1999. **288**(3): p. 1251-60.
130. Dixit, A. and T.E. Roche, *Spectrophotometric assay of the flavin-containing monooxygenase and changes in its activity in female mouse liver with nutritional and diurnal conditions.* Arch Biochem Biophys, 1984. **233**(1): p. 50-63.
131. Dever, J.T. and A.A. Elfarra, *In Vivo Metabolism of L-Methionine in Mice: Evidence for Stereoselective Formation of Methionine-d-Sulfoxide and Quantitation of Other Major Metabolites.* Drug Metab Dispos, 2006.
132. Karanam, B.V., C.E. Hop, D.Q. Liu, M. Wallace, D. Dean, H. Satoh, M. Komuro, K. Awano, and S.H. Vincent, *In vitro metabolism of MK-0767 [(+/-)-5-[(2,4-dioxothiazolidin-5-yl)methyl]-2-methoxy-N-[[(4-trifluoromet hyl) phenyl]methyl]benzamide], a peroxisome proliferator-activated receptor alpha/gamma agonist. I. Role of cytochrome P450, methyltransferases, flavin monooxygenases, and esterases.* Drug Metab Dispos, 2004. **32**(9): p. 1015-22.
133. Attar, M., D. Dong, K.H. Ling, and D.D. Tang-Liu, *Cytochrome P450 2C8 and flavin-containing monooxygenases are involved in the metabolism of tazarotenic acid in humans.* Drug Metab Dispos, 2003. **31**(4): p. 476-81.
134. Qian, L. and P.R. Ortiz de Montellano, *Oxidative activation of thiacetazone by the Mycobacterium tuberculosis flavin monooxygenase EtaA and human FMO1 and FMO3.* Chem Res Toxicol, 2006. **19**(3): p. 443-9.
135. Falls, J.G., N.J. Cherrington, K.M. Clements, R.M. Philpot, P.E. Levi, R.L. Rose, and E. Hodgson, *Molecular cloning, sequencing, and expression in Escherichia coli of mouse flavin-containing monooxygenase 3 (FMO3): comparison with the human isoform.* Arch Biochem Biophys, 1997. **347**(1): p. 9-18.
136. Virkel, G., A. Lifschitz, J. Sallovitz, A. Pis, and C. Lanusse, *Assessment of the main metabolism pathways for the flukicidal compound triclabendazole in sheep.* J Vet Pharmacol Ther, 2006. **29**(3): p. 213-23.
137. Tynes, R.E.a.H., E., *Magnitude of involvement of the mammalian flavin-containing monooxygenase in the microsomal oxidation of pesticides.* J. Agric. Food Chem., 1985. **33**: p. 471-479.
138. Chen, G.P. and D.M. Ziegler, *Liver microsome and flavin-containing monooxygenase catalyzed oxidation of organic selenium compounds.* Arch Biochem Biophys, 1994. **312**(2): p. 566-72.

139. Goeger, D.E. and H.E. Ganther, *Oxidation of dimethylselenide to dimethylselenoxide by microsomes from rat liver and lung and by flavin-containing monooxygenase from pig liver.* Arch Biochem Biophys, 1994. **310**(2): p. 448-51.
140. Krause, R.J., S.C. Glocke, A.R. Sicuri, S.L. Ripp, and A.A. Elfarra, *Oxidative Metabolism of Seleno-l-methionine to l-Methionine Selenoxide by Flavin-Containing Monooxygenases.* Chem Res Toxicol, 2006. **19**(12): p. 1643-1649.
141. Koukouritaki, S.B., M.T. Poch, E.T. Cabacungan, D.G. McCarver, and R.N. Hines, *Discovery of novel flavin-containing monooxygenase 3 (FMO3) single nucleotide polymorphisms and functional analysis of upstream haplotype variants.* Mol Pharmacol, 2005. **68**(2): p. 383-92.
142. Furnes, B., J. Feng, S.S. Sommer, and D. Schlenk, *Identification of novel variants of the flavin-containing monooxygenase gene family in African Americans.* Drug Metab Dispos, 2003. **31**(2): p. 187-93.
143. Lattard, V., J. Zhang, Q. Tran, B. Furnes, D. Schlenk, and J.R. Cashman, *Two new polymorphisms of the FMO3 gene in Caucasian and African-American populations: comparative genetic and functional studies.* Drug Metab Dispos, 2003. **31**(7): p. 854-60.
144. Cashman, J.R., *Human flavin-containing monooxygenase (form 3): polymorphisms and variations in chemical metabolism.* Pharmacogenomics, 2002. **3**(3): p. 325-39.
145. Brunelle, A., Y.A. Bi, J. Lin, B. Russell, L. Luy, C. Berkman, and J. Cashman, *Characterization of two human flavin-containing monooxygenase (form 3) enzymes expressed in Escherichia coli as maltose binding protein fusions.* Drug Metab Dispos, 1997. **25**(8): p. 1001-7.
146. Cashman, J.R. and J. Zhang, *Interindividual differences of human flavin-containing monooxygenase 3: genetic polymorphisms and functional variation.* Drug Metab Dispos, 2002. **30**(10): p. 1043-52.
147. Park, C.S., J.H. Kang, W.G. Chung, H.G. Yi, J.E. Pie, D.K. Park, R.N. Hines, D.G. McCarver, and Y.N. Cha, *Ethnic differences in allelic frequency of two flavin-containing monooxygenase 3 (FMO3) polymorphisms: linkage and effects on in vivo and in vitro FMO activities.* Pharmacogenetics, 2002. **12**(1): p. 77-80.
148. Doolittle, R.F., *Similar amino acid sequences: chance or common ancestry?* Science, 1981. **214**(4517): p. 149-59.
149. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
150. Ziegler, D.M., *An overview of the mechanism, substrate specificities, and structure of FMOs.* Drug Metab Rev, 2002. **34**(3): p. 503-11.
151. Cashman, J., *Flavin monooxygenases*, in *Enzyme Systems that Metabolize Drugs and Other Xenobiotics*, C. Ioannides, Editor. 2002, John Wiley and Sons, Ltd.: London. p. 67-94.
152. Mittl, P.R., A. Berry, N.S. Scrutton, R.N. Perham, and G.E. Schulz, *Anatomy of an engineered NAD-binding site.* Protein Sci, 1994. **3**(9): p. 1504-14.
153. Stehle, T., S.A. Ahmed, A. Claiborne, and G.E. Schulz, *Structure of NADH peroxidase from Streptococcus faecalis 10C1 refined at 2.16 A resolution.* J Mol Biol, 1991. **221**(4): p. 1325-44.
154. Anfinsen, C.B., E. Haber, M. Sela, and F.H. White, Jr., *The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain.* Proc Natl Acad Sci U S A, 1961. **47**: p. 1309-14.
155. Anfinsen, C.B., *Principles that govern the folding of protein chains.* Science, 1973. **181**(96): p. 223-30.

156. Baker, D. and A. Sali, *Protein structure prediction and structural genomics.* Science, 2001. **294**(5540): p. 93-6.
157. Wallner, B. and A. Elofsson, *All are not equal: a benchmark of different homology modeling programs.* Protein Sci, 2005. **14**(5): p. 1315-27.
158. Hillisch, A., L.F. Pineda, and R. Hilgenfeld, *Utility of homology models in the drug discovery process.* Drug Discov Today, 2004. **9**(15): p. 659-69.
159. Lesk, A.M. and C. Chothia, *How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins.* J Mol Biol, 1980. **136**(3): p. 225-70.
160. Pieper, U., N. Eswar, H. Braberg, M.S. Madhusudhan, F.P. Davis, A.C. Stuart, N. Mirkovic, A. Rossi, M.A. Marti-Renom, A. Fiser, et al., *MODBASE, a database of annotated comparative protein structure models, and associated resources.* Nucleic Acids Res, 2004. **32**(Database issue): p. D217-22.
161. Marti-Renom, M.A., A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali, *Comparative protein structure modeling of genes and genomes.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 291-325.
162. Eswar, N., B. John, N. Mirkovic, A. Fiser, V.A. Ilyin, U. Pieper, A.C. Stuart, M.A. Marti-Renom, M.S. Madhusudhan, B. Yerkovich, et al., *Tools for comparative protein structure modeling and analysis.* Nucleic Acids Res, 2003. **31**(13): p. 3375-80.
163. Ramachandran, G.N. and V. Sasisekharan, *Conformation of polypeptides and proteins.* Adv Protein Chem, 1968. **23**: p. 283-438.
164. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.* Electrophoresis, 1997. **18**(15): p. 2714-23.
165. Lewis, D.F., B.G. Lake, M. Dickins, and P.S. Goldfarb, *Molecular modelling of CYP2B6 based on homology with the CYP2C5 crystal structure: analysis of enzyme-substrate interactions.* Drug Metabol Drug Interact, 2002. **19**(2): p. 115-35.
166. Jean, P., J. Pothier, P.M. Dansette, D. Mansuy, and A. Viari, *Automated multiple analysis of protein structures: application to homology modeling of cytochromes P450.* Proteins, 1997. **28**(3): p. 388-404.
167. Lozano, J.J., E. Lopez-de-Brinas, N.B. Centeno, R. Guigo, and F. Sanz, *Three-dimensional modelling of human cytochrome P450 1A2 and its interaction with caffeine and MeIQ.* J Comput Aided Mol Des, 1997. **11**(4): p. 395-408.
168. Lewis, D.F., P.T. Bailey, and L.K. Low, *Molecular modelling of the mouse cytochrome P450 CYP2F2 based on the CYP102 crystal structure template and selective CYP2F2 substrate interactions.* Drug Metabol Drug Interact, 2002. **19**(2): p. 97-113.
169. Rost, B., *Twilight zone of protein sequence alignments.* Protein Eng, 1999. **12**(2): p. 85-94.
170. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993. **234**(3): p. 779-815.
171. Smigielski, E.M., K. Sirotkin, M. Ward, and S.T. Sherry, *dbSNP: a database of single nucleotide polymorphisms.* Nucleic Acids Res, 2000. **28**(1): p. 352-5.
172. Thorisson, G.A., A.V. Smith, L. Krishnan, and L.D. Stein, *The International HapMap Project Web site.* Genome Res, 2005. **15**(11): p. 1592-3.
173. Hudson, R.R., *Gene genealogies and the coalescent process.* Oxf. Surv. Evol. Biol., ed. D.F.a.J. Antonovics. Vol. 7. 1990: Oxford University Press, New York.
174. Watterson, G.A., *On the number of segregating sites in genetical models without recombination.* Theor Popul Biol, 1975. **7**(2): p. 256-76.

175. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol, 1987. **4**(4): p. 406-25.
176. Tajima, F., *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.* Genetics, 1989. **123**(3): p. 585-95.
177. Fu, Y.X. and W.H. Li, *Statistical tests of neutrality of mutations.* Genetics, 1993. **133**(3): p. 693-709.
178. Kimura, M., *The neutral theory of Molecular Evolution.* 1983 Cambridge, Massachusetts: Cambridge University Press.
179. Rogers, A.R. and H. Harpending, *Population growth makes waves in the distribution of pairwise genetic differences.* Mol Biol Evol, 1992. **9**(3): p. 552-69.
180. Wall, J., *Recombination and the power of statistical tests of neutrality.* Genet Res, 1999. **74**(1): p. 65-79.
181. Kelly, J.K., *A test of neutrality based on interlocus associations.* Genetics, 1997. **146**(3): p. 1197-206.
182. McDonald, J.H. and M. Kreitman, *Adaptive protein evolution at the Adh locus in Drosophila.* Nature, 1991. **351**(6328): p. 652-4.
183. Consortium, C.S.a.A., *Initial sequence of the chimpanzee genome and comparison with the human genome.* Nature, 2005. **437**(7055): p. 69-87.
184. WRIGHT, S., *The genetical structure of populations.* Ann. Eugenics., 1951. **15**: p. 323-354.
185. Merritt, T.J. and J.M. Quattro, *Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase.* Genetics, 2001. **159**(2): p. 689-97.
186. Harris, E.E. and J. Hey, *Human populations show reduced DNA sequence variation at the factor IX locus.* Curr Biol, 2001. **11**(10): p. 774-8.
187. Raberg, L. and M. Stjernman, *Natural selection on immune responsiveness in blue tits Parus caeruleus.* Evolution Int J Org Evolution, 2003. **57**(7): p. 1670-8.
188. Gilad, Y., A. Oshlack, G.K. Smyth, T.P. Speed, and K.P. White, *Expression profiling in primates reveals a rapid evolution of human transcription factors.* Nature, 2006. **440**(7081): p. 242-5.
189. Parham, P. and T. Ohta, *Population biology of antigen presentation by MHC class I molecules.* Science, 1996. **272**(5258): p. 67-74.
190. Hedrick, P.W. and P.S. Miller, *Rare alleles, MHC and captive breeding.* Exs, 1994. **68**: p. 187-204.
191. Verrelli, B.C., J.H. McDonald, G. Argyropoulos, G. Destro-Bisol, A. Froment, A. Drousiotou, G. Lefranc, A.N. Helal, J. Loiselet, and S.A. Tishkoff, *Evidence for balancing selection from nucleotide sequence analyses of human G6PD.* Am J Hum Genet, 2002. **71**(5): p. 1112-28.
192. Bamshad, M.J., S. Mummidi, E. Gonzalez, S.S. Ahuja, D.M. Dunn, W.S. Watkins, S. Wooding, A.C. Stone, L.B. Jorde, R.B. Weiss, et al., *A strong signature of balancing selection in the 5' cis-regulatory region of CCR5.* Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10539-44.
193. Wooding, S., U.K. Kim, M.J. Bamshad, J. Larsen, L.B. Jorde, and D. Drayna, *Natural selection and molecular evolution in PTC, a bitter-taste receptor gene.* Am J Hum Genet, 2004. **74**(4): p. 637-46.
194. Newman, R.M., L. Hall, M. Connole, G.L. Chen, S. Sato, E. Yuste, W. Diehl, E. Hunter, A. Kaur, G.M. Miller, et al., *Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha.* Proc Natl Acad Sci U S A, 2006. **103**(50): p. 19134-9.

195. Atta-Asafo-Adjei, E., M.P. Lawton, and R.M. Philpot, *Cloning, sequencing, distribution, and expression in Escherichia coli of flavin-containing monooxygenase 1C1. Evidence for a third gene subfamily in rabbits.* J Biol Chem, 1993. **268**(13): p. 9681-9.
196. Fukuda, K., H. Morioka, S. Imajou, S. Ikeda, E. Ohtsuka, and T. Tsurimoto, *Structure-function relationship of the eukaryotic DNA replication factor, proliferating cell nuclear antigen.* J Biol Chem, 1995. **270**(38): p. 22527-34.
197. Lawton, M.P. and R.M. Philpot, *Functional characterization of flavin-containing monooxygenase 1B1 expressed in Saccharomyces cerevisiae and Escherichia coli and analysis of proposed FAD- and membrane-binding domains.* J Biol Chem, 1993. **268**(8): p. 5728-34.
198. Lowry, O.H., N.J. Rosebrough, A.L. Farr, and R.J. Randall, *Protein measurement with the Folin phenol reagent.* J Biol Chem, 1951. **193**(1): p. 265-75.
199. Abramoff, M.D., P.J. Magelhaes, and S.J. Eram, *Image Processing with ImageJ.* Biophotonics International, 2004. **268**(8): p. 5728-34.
200. Hanes, C.S., *Studies on plant amylases: The effect of starch concentration upon the velocity of hydrolysis by the amylase of germinated barley.* Biochem J, 1932. **26**(5): p. 1406-21.
201. Hofstee, B.H., *Non-inverted versus inverted plots in enzyme kinetics.* Nature, 1959. **184**: p. 1296-8.
202. Lineweaver, H. and D. Burk, *The Determination of Enzyme Dissociation Constants.* J. Amer. Chem. Soc, 1934. **56**: p. 658-66.
203. Wilkinson, G.N., *Statistical estimations in enzyme kinetics.* Biochem J, 1961. **80**: p. 324-32.
204. Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
205. Jones, D.T., *GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences.* J Mol Biol, 1999. **287**(4): p. 797-815.
206. Lindahl, E. and A. Elofsson, *Identification of related proteins on family, superfamily and fold level.* J Mol Biol, 2000. **295**(3): p. 613-25.
207. Godzik, A., *Fold recognition methods.* Methods Biochem Anal, 2003. **44**: p. 525-46.
208. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
209. Henikoff, S. and J.G. Henikoff, *Amino acid substitution matrices from protein blocks.* Proc Natl Acad Sci U S A, 1992. **89**(22): p. 10915-9.
210. Sippl, M.J., *Recognition of errors in three-dimensional structures of proteins.* Proteins, 1993. **17**(4): p. 355-62.
211. Luthy, R., J.U. Bowie, and D. Eisenberg, *Assessment of protein models with three-dimensional profiles.* Nature, 1992. **356**(6364): p. 83-5.
212. Hooft, R.W., G. Vriend, C. Sander, and E.E. Abola, *Errors in protein structures.* Nature, 1996. **381**(6580): p. 272.
213. Laskowski, R.A., J.A. Rullmannn, M.W. MacArthur, R. Kaptein, and J.M. Thornton, *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J Biomol NMR, 1996. **8**(4): p. 477-86.
214. Eswaramoorthy, S., J.B. Bonanno, S.K. Burley, and S. Swaminathan, *Mechanism of action of a flavin-containing monooxygenase.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9832-7.

215. Malito, E., A. Alfieri, M.W. Fraaije, and A. Mattevi, *Crystal structure of a Baeyer-Villiger monooxygenase.* Proc Natl Acad Sci U S A, 2004. **101**(36): p. 13157-62.
216. Laurie, A.T. and R.M. Jackson, *Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites.* Bioinformatics, 2005. **21**(9): p. 1908-16.
217. Porter, C.T., G.J. Bartlett, and J.M. Thornton, *The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data.* Nucleic Acids Res, 2004. **32**(Database issue): p. D129-33.
218. Damborsky, J., M. Petrek, P. Banas, and M. Otyepka, *Identification of tunnels in proteins, nucleic acids, inorganic materials and molecular ensembles.* Biotechnol J, 2007. **2**(1): p. 62-7.
219. Moreland, J.L., A. Gramada, O.V. Buzko, Q. Zhang, and P.E. Bourne, *The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications.* BMC Bioinformatics, 2005. **6**: p. 21.
220. Ye, Y. and A. Godzik, *FATCAT: a web server for flexible structure comparison and structure similarity searching.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W582-5.
221. Ye, Y. and A. Godzik, *Multiple flexible structure alignment using partial order graphs.* Bioinformatics, 2005. **21**(10): p. 2362-9.
222. Allerston, C.K., M. Shimizu, M. Fujieda, E.A. Shephard, H. Yamazaki, and I.R. Phillips, *Molecular evolution and balancing selection in the*
flavin-containing monooxygenase 3 gene (FMO3). Pharmacogenetics and Genomics, 2007. **In Press**.
223. Rozas, J., J.C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas, *DnaSP, DNA polymorphism analyses by the coalescent and other methods.* Bioinformatics, 2003. **19**(18): p. 2496-7.
224. Hudson, R.R. and N.L. Kaplan, *Statistical properties of the number of recombination events in the history of a sample of DNA sequences.* Genetics, 1985. **111**(1): p. 147-64.
225. Bandelt, H.J., P. Forster, B.C. Sykes, and M.B. Richards, *Mitochondrial portraits of human populations using median networks.* Genetics, 1995. **141**(2): p. 743-53.
226. Tavare, S., D.J. Balding, R.C. Griffiths, and P. Donnelly, *Inferring coalescence times from DNA sequence data.* Genetics, 1997. **145**(2): p. 505-18.
227. Nei, M., *Molecular evolutionary genetics.* 1987, New York: Columbia University Press. x, 512 p.
228. Fu, Y.X., *Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection.* Genetics, 1997. **147**(2): p. 915-25.
229. Schneider, S., D. Roessli, and L. Excoffier, *Arlequin: a software for population genetics data analysis. Ver 3.01.* 2000, Genetics and Biometry Lab, Dept Anthropology, University of Geneva: Geneva.
230. Sherry, S.T., M.H. Ward, M. Kholodov, J. Baker, L. Phan, E.M. Smigielski, and K. Sirotkin, *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.
231. Bittles, A.H. and I. Egerbladh, *The influence of past endogamy and consanguinity on genetic disorders in northern Sweden.* Ann Hum Genet, 2005. **69**(Pt 5): p. 549-58.
232. Fraser-Andrews, E.A., N.J. Manning, G.H. Ashton, P. Eldridge, J. McGrath, and P. Menage Hdu, *Fish odour syndrome with features of both primary and secondary trimethylaminuria.* Clin Exp Dermatol, 2003. **28**(2): p. 203-5.

233. Treacy, E., D. Johnson, J.J. Pitt, and D.M. Danks, *Trimethylaminuria, fish odour syndrome: a new method of detection and response to treatment with metronidazole.* J Inherit Metab Dis, 1995. **18**(3): p. 306-12.
234. Dai, D., D.C. Zeldin, J.A. Blaisdell, B. Chanas, S.J. Coulter, B.I. Ghanayem, and J.A. Goldstein, *Polymorphisms in human CYP2C8 decrease metabolism of the anticancer drug paclitaxel and arachidonic acid.* Pharmacogenetics, 2001. **11**(7): p. 597-607.
235. Bahadur, N., J.B. Leathart, E. Mutch, D. Steimel-Crespi, S.A. Dunn, R. Gilissen, J.V. Houdt, J. Hendrickx, G. Mannens, H. Bohets, et al., *CYP2C8 polymorphisms in Caucasians and their relationship with paclitaxel 6alpha-hydroxylase activity in human liver microsomes.* Biochem Pharmacol, 2002. **64**(11): p. 1579-89.
236. Scott, S.A., L. Edelmann, R. Kornreich, and R.J. Desnick, *Warfarin pharmacogenetics: CYP2C9 and VKORC1 genotypes predict different sensitivity and resistance frequencies in the Ashkenazi and Sephardi Jewish populations.* Am J Hum Genet, 2008. **82**(2): p. 495-500.
237. Kirchheiner, J. and J. Brockmoller, *Clinical consequences of cytochrome P450 2C9 polymorphisms.* Clin Pharmacol Ther, 2005. **77**(1): p. 1-16.
238. Haining, R.L., A.P. Hunter, M.E. Veronese, W.F. Trager, and A.E. Rettie, *Allelic variants of human cytochrome P450 2C9: baculovirus-mediated expression, purification, structural characterization, substrate stereoselectivity, and prochiral selectivity of the wild-type and I359L mutant forms.* Arch Biochem Biophys, 1996. **333**(2): p. 447-58.
239. Rettie, A.E., L.C. Wienkers, F.J. Gonzalez, W.F. Trager, and K.R. Korzekwa, *Impaired (S)-warfarin metabolism catalysed by the R144C allelic variant of CYP2C9.* Pharmacogenetics, 1994. **4**(1): p. 39-42.
240. Crespi, C.L. and V.P. Miller, *The R144C change in the CYP2C9*2 allele alters interaction of the cytochrome P450 with NADPH:cytochrome P450 oxidoreductase.* Pharmacogenetics, 1997. **7**(3): p. 203-10.
241. Wolf, C.R., C.A. Smith, A.C. Gough, J.E. Moss, K.A. Vallis, G. Howard, F.J. Carey, K. Mills, W. McNee, J. Carmichael, et al., *Relationship between the debrisoquine hydroxylase polymorphism and cancer susceptibility.* Carcinogenesis, 1992. **13**(6): p. 1035-8.
242. Wolf, C.R. and N.K. Spurr, *Identification of genetic differences in debrisoquine hydroxylase activity.* Hepatology, 1992. **15**(2): p. 360.
243. Tucker, G.T., J.H. Silas, A.O. Iyun, M.S. Lennard, and A.J. Smith, *Polymorphic hydroxylation of debrisoquine.* Lancet, 1977. **2**(8040): p. 718.
244. De Morais, S.M., G.R. Wilkinson, J. Blaisdell, U.A. Meyer, K. Nakamura, and J.A. Goldstein, *Identification of a new genetic defect responsible for the polymorphism of (S)-mephenytoin metabolism in Japanese.* Mol Pharmacol, 1994. **46**(4): p. 594-8.
245. de Morais, S.M., G.R. Wilkinson, J. Blaisdell, K. Nakamura, U.A. Meyer, and J.A. Goldstein, *The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans.* J Biol Chem, 1994. **269**(22): p. 15419-22.
246. van Schaik, R.H., *Cancer treatment and pharmacogenetics of cytochrome P450 enzymes.* Invest New Drugs, 2005. **23**(6): p. 513-22.
247. Cashman, J.R., B.R. Akerman, S.M. Forrest, and E.P. Treacy, *Population-specific polymorphisms of the human FMO3 gene: significance for detoxication.* Drug Metab Dispos, 2000. **28**(2): p. 169-73.
248. Cashman, J.R., J. Zhang, J. Leushner, and A. Braun, *Population distribution of human flavin-containing monooxygenase form 3: gene polymorphisms.* Drug Metab Dispos, 2001. **29**(12): p. 1629-37.

249. Sachse, C., S. Ruschen, M. Dettling, J. Schley, S. Bauer, B. Muller-Oerlinghausen, I. Roots, and J. Brockmoller, *Flavin monooxygenase 3 (FMO3) polymorphism in a white population: allele frequencies, mutation linkage, and functional effects on clozapine and caffeine metabolism.* Clin Pharmacol Ther, 1999. **66**(4): p. 431-8.
250. Hao, D., J. Sun, B. Furnes, D. Schlenk, M. Li, S. Yang, and L. Yang, *Allele and genotype frequencies of polymorphic FMO3 gene in two genetically distinct populations.* Cell Biochem Funct, 2007. **25**(4): p. 443-53.
251. Kang, J.H., W.G. Chung, K.H. Lee, C.S. Park, J.S. Kang, I.C. Shin, H.K. Roh, M.S. Dong, H.M. Baek, and Y.N. Cha, *Phenotypes of flavin-containing monooxygenase activity determined by ranitidine N-oxidation are positively correlated with genotypes of linked FM03 gene mutations in a Korean population.* Pharmacogenetics, 2000. **10**(1): p. 67-78.
252. Koukouritaki, S.B. and R.N. Hines, *Flavin-containing monooxygenase genetic polymorphism: impact on chemical metabolism and drug development.* Pharmacogenomics, 2005. **6**(8): p. 807-22.
253. Lee, J.K., J.H. Kang, Y.N. Cha, W.G. Chung, and C.S. Park, *Alteration of Substrate Specificity by Common Variants, E158K/E308G and V257M, in Human Hepatic Drug-metabolizing Enzyme, Flavin-containing Monooxygenase 3.* Korean J Physiol Pharmacol., 2003. **7**(3): p. 157-162.
254. Hao, D.C., J. Sun, B. Furnes, D. Schlenk, Z.F. Hou, Y.P. Zhang, S.L. Yang, and L. Yang, *Haplotype frequency distribution and linkage disequilibrium analysis of single nucleotide polymorphisms at the human FMO3 gene locus.* Biochem Genet, 2006. **44**(7-8): p. 391-407.
255. Stormer, E., I. Roots, and J. Brockmoller, *Benzydamine N-oxidation as an index reaction reflecting FMO activity in human liver microsomes and impact of FMO3 polymorphisms on enzyme activity.* Br J Clin Pharmacol, 2000. **50**(6): p. 553-61.
256. Shimizu, M., H. Yano, S. Nagashima, N. Murayama, J. Zhang, J.R. Cashman, and H. Yamazaki, *Effect of genetic variants of the human flavin-containing monooxygenase 3 on N- and S-oxygenation activities.* Drug Metab Dispos, 2007. **35**(3): p. 328-30.
257. Dolan, C., D.C. Shields, A. Stanton, E. O'Brien, D.M. Lambert, J.K. O'Brien, and E.P. Treacy, *Polymorphisms of the Flavin containing monooxygenase 3 (FMO3) gene do not predispose to essential hypertension in Caucasians.* BMC Med Genet, 2005. **6**: p. 41.
258. Lambert, D.M., O.A. Mamer, B.R. Akerman, L. Choiniere, D. Gaudet, P. Hamet, and E.P. Treacy, *In vivo variability of TMA oxidation is partially mediated by polymorphisms of the FMO3 gene.* Mol Genet Metab, 2001. **73**(3): p. 224-9.
259. Yeung, C.K., E.T. Adman, and A.E. Rettie, *Functional characterization of genetic variants of human FMO3 associated with trimethylaminuria.* Arch Biochem Biophys, 2007.
260. Hisamuddin, I.M., M.A. Wehbi, A. Chao, H.W. Wyre, L.M. Hylind, F.M. Giardiello, and V.W. Yang, *Genetic polymorphisms of human flavin monooxygenase 3 in sulindac-mediated primary chemoprevention of familial adenomatous polyposis.* Clin Cancer Res, 2004. **10**(24): p. 8357-62.
261. Hisamuddin, I.M., M.A. Wehbi, B. Schmotzer, K.A. Easley, L.M. Hylind, F.M. Giardiello, and V.W. Yang, *Genetic polymorphisms of flavin monooxygenase 3 in sulindac-induced regression of colorectal adenomas in familial adenomatous polyposis.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(10): p. 2366-9.

262. Bowie, J.U., R. Luthy, and D. Eisenberg, *A method to identify protein sequences that fold into a known three-dimensional structure.* Science, 1991. **253**(5016): p. 164-70.
263. Laskowski, R.A., D.S. Moss, and J.M. Thornton, *Main-chain bond lengths and bond angles in protein structures.* J Mol Biol, 1993. **231**(4): p. 1049-67.
264. Williams, C.H., L.D. Arscott, S. Muller, B.W. Lennon, M.L. Ludwig, P.F. Wang, D.M. Veine, K. Becker, and R.H. Schirmer, *Thioredoxin reductase two modes of catalysis have evolved.* Eur J Biochem, 2000. **267**(20): p. 6110-7.
265. Hayward, S., *Structural principles governing domain motions in proteins.* Proteins, 1999. **36**(4): p. 425-35.
266. Marsden, R.L., L.J. McGuffin, and D.T. Jones, *Rapid protein domain assignment from amino acid sequence using predicted secondary structure.* Protein Sci, 2002. **11**(12): p. 2814-24.
267. Holm, L. and C. Sander, *Parser for protein folding units.* Proteins, 1994. **19**(3): p. 256-68.
268. Holland, T.A., S. Veretnik, I.N. Shindyalov, and P.E. Bourne, *Partitioning protein structures into domains: why is it so difficult?* J Mol Biol, 2006. **361**(3): p. 562-90.
269. Honkatukia, M., K. Reese, R. Preisinger, M. Tuiskula-Haavisto, S. Weigend, J. Roito, A. Maki-Tanila, and J. Vilkki, *Fishy taint in chicken eggs is associated with a substitution within a conserved motif of the FMO3 gene.* Genomics, 2005. **86**(2): p. 225-32.
270. Borbas, T., J. Zhang, M. Cerny, I. Liko, and J.R. Cashman, *Investigation of structure and function of a catalytically efficient variant of the human flavin-containing monooxygenase form 3 (FMO3).* Drug Metab Dispos, 2006.
271. Svensson-Ek, M., J. Abramson, G. Larsson, S. Tornroth, P. Brezezinski, and S. Iwata, *Structure of cytochrome c oxidase from Rhodobactor sphaeros (Wtild Type).* Journal of Molecular Biology, 2002. **321**: p. 329-339.
272. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *Protein-ligand docking: current status and future challenges.* Proteins, 2006. **65**(1): p. 15-26.
273. Kramer, B., M. Rarey, and T. Lengauer, *Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking.* Proteins, 1999. **37**(2): p. 228-41.
274. Shoichet, B.K., D.L. Bodian, and I.D. Kuntz, *Molecular docking using shape descriptors.* Journal of Computational Chemistry, 1992. **13**(3).
275. Park, H., J. Lee, and S. Lee, *Critical assessment of the automated AutoDock as a new docking tool for virtual screening.* Proteins, 2006. **65**(3): p. 549-54.
276. Halperin, E. and E. Eskin, *Haplotype reconstruction from genotype data using Imperfect Phylogeny.* Bioinformatics, 2004. **20**(12): p. 1842-9.
277. Bamshad, M. and S.P. Wooding, *Signatures of natural selection in the human genome.* Nat Rev Genet, 2003. **4**(2): p. 99-111.
278. Sachidanandam, R., D. Weissman, S.C. Schmidt, J.M. Kakol, L.D. Stein, G. Marth, S. Sherry, J.C. Mullikin, B.J. Mortimore, D.L. Willey, et al., *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.* Nature, 2001. **409**(6822): p. 928-33.
279. Takahata, N., *Allelic genealogy and human evolution.* Mol Biol Evol, 1993. **10**(1): p. 2-22.
280. Hubbard, T.J., B.L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, et al., *Ensembl 2007.* Nucleic Acids Res, 2007. **35**(Database issue): p. D610-7.

281. Tishkoff, S.A. and B.C. Verrelli, *Patterns of human genetic diversity: implications for human evolutionary history and disease.* Annu Rev Genomics Hum Genet, 2003. **4**: p. 293-340.
282. Harding, R.M., S.M. Fullerton, R.C. Griffiths, and J.B. Clegg, *A gene tree for beta-globin sequences from Melanesia.* J Mol Evol, 1997. **44 Suppl 1**: p. S133-8.
283. Fullerton, S.M., A.G. Clark, K.M. Weiss, D.A. Nickerson, S.L. Taylor, J.H. Stengard, V. Salomaa, E. Vartiainen, M. Perola, E. Boerwinkle, et al., *Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism.* Am J Hum Genet, 2000. **67**(4): p. 881-900.
284. Wooding, S.P., W.S. Watkins, M.J. Bamshad, D.M. Dunn, R.B. Weiss, and L.B. Jorde, *DNA sequence variation in a 3.7-kb noncoding sequence 5' of the CYP1A2 gene: implications for human population history and natural selection.* Am J Hum Genet, 2002. **71**(3): p. 528-42.
285. Rieder, M.J., S.L. Taylor, A.G. Clark, and D.A. Nickerson, *Sequence variation in the human angiotensin converting enzyme.* Nat Genet, 1999. **22**(1): p. 59-62.
286. Clark, A.G., K.M. Weiss, D.A. Nickerson, S.L. Taylor, A. Buchanan, J. Stengard, V. Salomaa, E. Vartiainen, M. Perola, E. Boerwinkle, et al., *Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase.* Am J Hum Genet, 1998. **63**(2): p. 595-612.
287. Cavalli-Sforza, L.L. and M.W. Feldman, *The application of molecular genetic approaches to the study of human evolution.* Nat Genet, 2003. **33 Suppl**: p. 266-75.
288. Przeworski, M., R.R. Hudson, and A. Di Rienzo, *Adjusting the focus on human variation.* Trends Genet, 2000. **16**(7): p. 296-302.
289. Schneider, J.A., M.S. Pungliya, J.Y. Choi, R. Jiang, X.J. Sun, B.A. Salisbury, and J.C. Stephens, *DNA variability of human genes.* Mech Ageing Dev, 2003. **124**(1): p. 17-25.
290. Watkins, W.S., A.R. Rogers, C.T. Ostler, S. Wooding, M.J. Bamshad, A.M. Brassington, M.L. Carroll, S.V. Nguyen, J.A. Walker, B.V. Prasad, et al., *Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms.* Genome Res, 2003. **13**(7): p. 1607-18.
291. Akey, J.M., G. Zhang, K. Zhang, L. Jin, and M.D. Shriver, *Interrogating a high-density SNP map for signatures of natural selection.* Genome Res, 2002. **12**(12): p. 1805-14.
292. Jorde, L.B., W.S. Watkins, and M.J. Bamshad, *Population genomics: a bridge from evolutionary history to genetic medicine.* Hum Mol Genet, 2001. **10**(20): p. 2199-207.
293. Nachman, M.W., *Single nucleotide polymorphisms and recombination rate in humans.* Trends Genet, 2001. **17**(9): p. 481-5.
294. Hamblin, M.T. and A. Di Rienzo, *Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus.* Am J Hum Genet, 2000. **66**(5): p. 1669-79.
295. Stephens, J.C., J.A. Schneider, D.A. Tanguay, J. Choi, T. Acharya, S.E. Stanley, R. Jiang, C.J. Messer, A. Chew, J.H. Han, et al., *Haplotype variation and linkage disequilibrium in 313 human genes.* Science, 2001. **293**(5529): p. 489-93.
296. Ryman, N. and P.E. Jorde, *Statistical power when testing for genetic differentiation.* Mol Ecol, 2001. **10**(10): p. 2361-73.
297. Ingelman-Sundberg, M., S.C. Sim, A. Gomez, and C. Rodriguez-Antona, *Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic,*

pharmacoepigenetic and clinical aspects. Pharmacol Ther, 2007. **116**(3): p. 496-526.
298. Reynolds, G.P., *The impact of pharmacogenetics on the development and use of antipsychotic drugs.* Drug Discov Today, 2007. **12**(21-22): p. 953-9.

Molecular evolution and balancing selection in the flavin-containing monooxygenase 3 gene (*FMO3*)

Charles K. Allerston[a], Makiko Shimizu[c], Masaki Fujieda[d], Elizabeth A. Shephard[a], Hiroshi Yamazaki[c] and Ian R. Phillips[b]