Biochem. J. (1992) 287, 261–267 (Printed in Great Britain)

Cloning, primary sequence and chromosomal localization of human FMO2, a new member of the flavin-containing mono-oxygenase family

Colin T. DOLPHIN,*§ Elizabeth A. SHEPHARD,† Susan POVEY,‡ Robert L. SMITH§ and Ian R. PHILLIPS*¶

* Department of Biochemistry, Queen Mary & Westfield College, University of London, Mile End Road, London E1 4NS,
† Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT,
‡ MRC Human Biochemical Genetics Unit, The Galton Laboratory, Wolfson House, 4 Stephenson Way, London NW1 2HE,
and § Department of Pharmacology & Toxicology, St Mary's Hospital Medical School, University of London, Norfolk Place, London W2 1PG, U.K.

We have previously reported the cloning of cDNAs for a flavin-containing mono-oxygenase (FMO) of man, designated FMO1 [Dolphin, Shephard, Povey, Palmer, Ziegler, Ayesh, Smith & Phillips (1991) J. Biol. Chem. **266**, 12379–12385], that is the orthologue of pig and rabbit hepatic FMOs. We now describe the isolation and characterization of cDNA clones for a second human FMO, which we have designated FMO2. The polypeptide encoded by the cDNAs is 558 amino acid residues long, has a calculated M_r of 63337, and contains putative FAD- and NADP-binding sites that align exactly with those described in other mammalian FMOs. Human FMO2 has 51–53% primary sequence identity with human FMO1, rabbit pulmonary FMO and rabbit liver FMO form 2, and thus represents a fourth, distinct, member of the mammalian FMO family. The corresponding mRNA is present in low abundance in adult human liver. Southern blot hybridization with single-exon probes demonstrated that human FMO2 and FMO1 are the products of single genes. The gene encoding FMO2 (designated *FMO2*) was mapped, by the polymerase chain reaction, to human chromosome 1, the same chromosome on which *FMO1* is located.

INTRODUCTION

Metabolic oxidation of nitrogen-, sulphur- or phosphorus-containing drugs, pesticides and other xenobiotics may be mediated by the flavin-containing mono-oxygenase (FMO; EC 1.14.13.8) system, the cytochrome *P*-450 mono-oxygenase system, or both. The relative contributions of each of these systems to the metabolism of such compounds is determined predominantly by the nucleophilic nature of the heteroatom, with 'soft' nucleophiles being preferentially oxygenated by FMOs (reviewed by Ziegler, 1980, 1988). FMO is known also to mediate the N-oxidation of the dietary derived amine, trimethylamine (TMA) (Hlavica & Kehl, 1977). This biotransformation exhibits a genetic polymorphism in man (Al Waiz *et al.*, 1987), resulting in a small sub-population unable to effect TMA N-oxidation. Such individuals exhibit the unpleasant symptoms of the associated disorder trimethylaminuria, colloquially known as fish odour syndrome (Humbert *et al.*, 1970; Shelley & Shelley, 1984) due to excretion of the malodorous free amine.

FMO has been purified and characterized from pig (Ziegler & Mitchell, 1972), rat (Kimura *et al.*, 1983), mouse (Sabourin *et al.*, 1984; Tynes *et al.*, 1986) and rabbit (Tynes *et al.*, 1986; Ozols, 1989) liver, mouse (Tynes *et al.*, 1986) and rabbit (Williams *et al.*, 1984; Tynes *et al.*, 1986) lung, and mouse kidney (Venkatesh *et al.*, 1991). Observed differences in substrate specificities, relative M_rs and immunochemical properties indicated that the 'hepatic' and 'pulmonary' FMOs represent different forms of the enzyme. This was subsequently confirmed by the isolation and sequencing of cDNA clones for the hepatic FMOs of pig (Gasser *et al.*, 1990) and rabbit (Lawton *et al.*, 1990), and for the pulmonary FMO of rabbit (Lawton *et al.*, 1990). Ozols (1991) has reported the amino acid sequence of a second FMO purified from rabbit liver, designated rabbit liver form 2, which represents a third form of the enzyme that is distinct from both rabbit hepatic and pulmonary FMOs. Recently, Lawton *et al.* (1991) have investigated the substrate specificities of both rabbit hepatic and pulmonary FMOs by expressing their respective cDNAs in COS cells. Comparisons of kinetic parameters derived from these individually expressed enzymes with those derived from microsome-driven reactions indicate that both rabbit liver and lung each contain at least two forms of the enzyme.

We have recently described the isolation and characterization of cDNA clones that encode the human orthologue of pig and rabbit hepatic FMO (Dolphin *et al.*, 1991), and have proposed the term FMO1 to describe the human protein and its orthologues in other mammalian species. We report here the isolation and characterization of cDNA clones that encode a second form of human FMO, which we have designated FMO2. A comparison of the amino acid sequence, deduced from the cDNAs, with those of other mammalian FMOs, indicates that human FMO2 represents a fourth, distinct, form of the enzyme which has no previously reported animal orthologues. FMO2 is encoded by a single gene which we have mapped to human chromosome 1.

MATERIALS AND METHODS

cDNA library screening

An adult human liver cDNA library constructed in λgt11 (gift from Dr. S. Woo, Baylor College of Medicine, Houston) was plated at a density of 2.5×10^4 plaque-forming units/140 mm-diam. plate and screened with a previously isolated

Abbreviations used: FMO, flavin-containing mono-oxygenase; TMA, trimethylamine.

¶ To whom correspondence should be addressed.

The nucleotide sequence data reported here will appear in the GenBank and EMBL databases under accession no. Z11737.

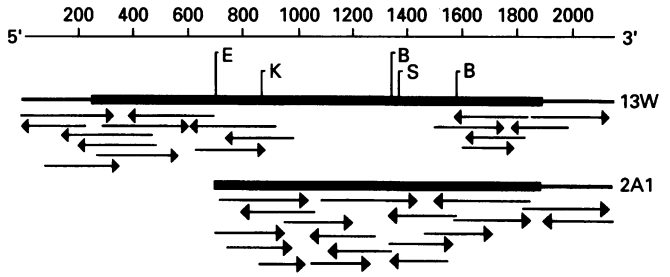

Fig. 1. Strategy for sequencing cDNAs encoding human FMO2

Both clones were isolated from an adult human liver library. Arrows indicate the extent and direction of sequencing reactions. These reactions were primed with oligonucleotides complementary to either vector (pUC19 or pBluescript) or insert sequences. Restriction sites used to generate subclones are indicated: E, *Eco*RI; K, *Kpn*I; B, *Bam*HI; S, *Sac*I.

1.4 kb partial-length cDNA encoding porcine FMO1 (Dolphin *et al.*, 1991). The cDNA was radiolabelled by the oligonucleotide random primer method (Feinberg & Vogelstein, 1983) to a specific radioactivity of approx. 10^9 c.p.m./μg with [α-^{32}P]dCTP (800 Ci/mmol; Amersham International). Duplicate filters were prehybridized, hybridized, washed and subjected to autoradiography as described previously (Dolphin *et al.*, 1991). The final wash was in 0.1 × SSPE (1 × SSPE = 0.18 M-NaCl, 10 mM-sodium phosphate, pH 7.7, 1 mM-EDTA)/0.1 % SDS for 15 min at 37 °C. The same library was subsequently re-screened at higher stringency (final wash of 15 min at 50 °C in 0.1 × SSPE/0.1 % SDS) with a cDNA insert encoding a partial-length human FMO2 clone isolated from the first screen.

DNA sequencing

Restriction fragments of cDNAs were subcloned into either pUC19 or pBluescript vectors and sequenced directly by the dideoxy chain-termination method (Sanger *et al.*, 1977) with sequencing kits that used either T7 polymerase (Sequenase II; United States Biochemicals) or *Taq* polymerase (TaqTrack; Promega). 7-Deaza-dGTP and [α-^{35}S]dATP (> 1000 Ci/mmol; Amersham International) were used with all sequencing protocols. Reactions were primed with either pUC19- or pBluescript-specific primers or with synthetic 17-mer oligonucleotides complementary to sequences within the insert DNA. The complete cDNA sequence was established after sequencing both strands using the strategy shown in Fig. 1. For each reaction the sequence was determined from at least two independent gel readings. Sequence data were analysed using MacVector sequence analysis software (International Biotechnologies).

Southern blot hybridization

Human genomic DNA from a single individual was prepared from whole blood by a modification of the method of Blin & Stafford (1976). DNA samples (10 μg) were digested with either *Eco*RI, *Bam*HI or *Pst*I, electrophoresed in duplicate sets through a 0.8 % agarose gel and capillary-blotted on to a single nylon membrane (Hybond-N; Amersham International). The membrane was cut into two halves to provide duplicate blots which were prehybridized as previously described (Dolphin *et al.*, 1991). One of the duplicate filters was then hybridized with a probe complementary to a region located within the 3′ non-coding region of the cDNA encoding human FMO1 (Dolphin *et al.*, 1991) and the other with a probe complementary to the corresponding region of human FMO2. The probes were generated from cDNA clones by the PCR. The sequences and priming sites of the oligonucleotides used for FMO2 amplification are shown in Fig. 2, and those for FMO1 amplification have been described previously (Dolphin *et al.*, 1991). The oligonucleotides were synthesized on a PCR-MATE DNA synthesizer (model 391; Applied Biosystems). PCR conditions were as previously described (Dolphin *et al.*, 1991). The reaction mixture was incubated (Thermal Rector; Hybaid) for 32 cycles at 94 °C for 45 s, 56 °C for 60 s and 72 °C for 90 s, and then for an additional 5 min at 72 °C. Amplified product was extracted twice with phenol/chloroform (1 : 1, v/v), ethanol-precipitated, resuspended in distilled water, and radiolabelled as described above. Prehybridization and hybridization conditions were as described previously (Dolphin *et al.*, 1991). After hybridization, the membranes were washed in 2 × SSC (1 × SSC = 15 mM-sodium citrate/0.15 M-NaCl, pH 7.0)/0.1 % SDS for 30 min at room temperature, in 0.1 × SSC/0.1 % SDS for 30 min at room temperature, and then at 50 °C in 0.1 × SSC/0.1 % SDS for 15 min. Autoradiography was for 72 h at −78 °C with an intensifying screen.

Chromosomal localization

The hybrid cell lines used in this work (Table 1) have all been described previously, but many have subsequently been re-cloned and re-characterized. Genomic DNA was isolated as described previously (Edwards *et al.*, 1985). The sequences and priming sites of the two 20-mer oligonucleotides used to prime the PCR are indicated in Fig. 2. Amplification reactions were performed using either 1 μg of human or rodent cell DNA or 2 μg of somatic cell hybrid DNA. DNA samples were amplified using the DNA polymerase from *Thermus thermophilus* (TET/Z; Bioline). Reaction conditions were as described previously (Dolphin *et al.*, 1991), except that the amplification was for 35 cycles of 90 °C for 1 min, 50 °C for 1 min and 70 °C for 1 min. PCR products were analysed on a 2.5 % agarose gel.

RESULTS

An initial screen of 2.5×10^5 plaques of an adult human liver cDNA library with a 1.4 kb cDNA insert encoding pig FMO1 yielded 10 positive clones. One of these, which gave a strong hybridization signal, contained a partial-length cDNA insert that encoded human FMO1 and has been described previously (Dolphin *et al.*, 1991). The remaining nine clones gave much weaker hybridization signals. One of these, designated 2A1, contained a 1.4 kb insert (Fig. 1) that encoded a polypeptide, the amino acid sequence of which was approx. 50 % identical with that of human FMO1. Restriction mapping and partial sequencing indicated that the other eight clones contained overlapping cDNAs (results not shown). To obtain longer cDNAs for this polypeptide, 4×10^5 plaques of the library were re-screened at higher stringency with the insert of clone 2A1. Fifteen additional clones were isolated, the largest of which (clone 13W) contained a cDNA insert of over 2.1 kb (Fig. 1). Restriction mapping revealed that the other clones contained shorter versions of the same cDNA (results not shown).

A full-length cDNA sequence was assembled from clones 2A1 and 13W as indicated in Fig. 1, and, together with the amino acid sequence deduced from it, is shown in Fig. 2. The nucleotide sequence contains a 5′ flanking region of 216 bases followed by an open reading frame of 1674 bases, a TGA termination codon, and a 3′ non-coding region of 254 bases that contains a polyadenylation signal (AATAAA) located 17 bases upstream of a polyadenylation site. The cDNA encodes a polypeptide (designated FMO2) of 558 amino acid residues with a calculated M_r of 63337. Potential FAD- and NADP-binding sites, each with the Gly-Xaa-Gly-Xaa-Xaa-Gly consensus sequence found in other forms of FMO (Gasser *et al.*, 1990; Lawton *et al.*, 1990; Ozols,

```
   1 AGACAAACACTTTCCTTGACTTTGAGAAATAATTTAAGTCAAAGAATCTGCTCTATGCTAACCAAGAGATAGAGCACAGCA   81
  82 AAGATCTGCCAGCCCCAGGCCTCTACCTAGTGGCCTGGAAATTCAAGTATTCTTATTGGTGGAGGCCATTTGTTTCTGATT  162
 163 AGAAGCTGTCTAAACCTCCTACTCCTCAACTCAAAGGAAAACACAGAGCATACCATGGCCAAGAAAGTTGCAGTGATTGGA  243
   1                                                    MetAlaLysLysValAlaValIleGly    9

 244 GCTGGTGTGAGTGGCCTCTCCTCCATCAAATGCTGTGTGGATGAGGACCTGGAGCCCACCTGCTTTGAGAGAAGTGATGAC  324
  10 AlaGlyValSerGlyLeuSerSerIleLysCysCysValAspGluAspLeuGluProThrCysPheGluArgSerAspAsp   36

 325 ATTGGGGGATTATGGAAGTTTACTGAATCTTCCAAAGATGGGATGACCAGGGTCTATAAGTCATTAGTGACAAATGTCTGT  405
  37 IleGlyGlyLeuTrpLysPheThrGluSerSerLysAspGlyMetThrArgValTyrLysSerLeuValThrAsnValCys   63

 406 AAGGAAATGTCATGTTACAGTGACTTCCCTTTCCACGAAGATTATCCTAATTTCATGAACCATGAAAAATTTTGGGACTAT  486
  64 LysGluMetSerCysTyrSerAspPheProPheHisGluAspTyrProAsnPheMetAsnHisGluLysPheTrpAspTyr   90

 487 CTCCAAGAATTTGCTGAGCACTTTGACCTCCTGAAATACATTCAGTTTAAGACCACTGTGTGCAGCATAACGAAGCGTCCA  567
  91 LeuGlnGluPheAlaGluHisPheAspLeuLeuLysTyrIleGlnPheLysThrThrValCysSerIleThrLysArgPro  117

 577 GACTTCTCCGAAACTGGTCAGTGGGATGTTGTCACAGAGACAGAGGGCAAGCAAAATAGAGCTGTCTTTGATGCTGTTATG  648
 117 AspPheSerGluThrGlyGlnTrpAspValValThrGluThrGluGlyLysGlnAsnArgAlaValPheAspAlaValMet  144

 649 GTTTGCACTGGACATTTCCTGAATCCCCATTTACCTTTGGAAGCCTTTCCTGGAATTCATAAGTTTAAAGGTCAGATCCTG  729
 145 ValCysThrGlyHisPheLeuAsnProHisLeuProLeuGluAlaPheProGlyIleHisLysPheLysGlyGlnIleLeu  171

 730 CATAGTCAAGAGTACAAGATCCCAGAAGGCTTTCAGGGCAAACGCGTCTTGGTGATTGGTCTTGGGAACACTGGAGGAGAC  810
 172 HisSerGlnGluTyrLysIleProGluGlyPheGlnGlyLysArgValLeuValIleGlyLeuGlyAsnThrGlyGlyAsp  198

 811 ATTGCTGTGGAACTCAGTCGAACGGCAGCTCAGGTACTTCTCAGTACTAGAACTGGTACCTGGGTTCTTGGGCGCTCTTCA  891
 199 IleAlaValGluLeuSerArgThrAlaAlaGlnValLeuLeuSerThrArgThrGlyThrTrpValLeuGlyArgSerSer  225

 892 GATTGGGGCTATCCTTATAATATGATGGTTACAAGAAGATGCTGTAGTTTTATTGCACAAGTTCTGCCTTCACGTTTTCTA  972
 226 AspTrpGlyTyrProTyrAsnMetMetValThrArgArgCysCysSerPheIleAlaGlnValLeuProSerArgPheLeu  252

 973 AACTGGATTCAAGAAAGGAAGTTGAATAAGAGATTTAATCATGAGGATTATGGATTAAGTATTACCAAAGGGAAAAAAGCA 1053
 253 AsnTrpIleGlnGluArgLysLeuAsnLysArgPheAsnHisGluAspTyrGlyLeuSerIleThrLysGlyLysLysAla  279

1054 AAATTCATTGTGAATGATGAGCTGCCAAACTGTATCCTCTGTGGGGCAATCACTATGAAAACCAGCGTGATTGAATTTACA 1134
 280 LysPheIleValAsnAspGluLeuProAsnCysIleLeuCysGlyAlaIleThrMetLysThrSerValIleGluPheThr  306

1135 GAAACCTCTGCTGTCTTTGAAGATGGGACAGTGGAAGAAAACATTGATGTTGTGATCTTCACTACAGGATATACATTTTCT 1215
 307 GluThrSerAlaValPheGluAspGlyThrValGluGluAsnIleAspValValIlePheThrThrGlyTyrThrPheSer  333

1216 TTTCCATTTTTTGAAGAACCTCTTAAAAGCCTCTGTACAAAGAAGATATTTCTATACAAGCAAGTCTTTCCCTTAAACCTA 1296
 334 PheProPhePheGluGluProLeuLysSerLeuCysThrLysLysIlePheLeuTyrLysGlnValPheProLeuAsnLeu  360

1297 GAGAGAGCGACATTAGCCATCATCGGCCTTATCGGCCTTAAAGGATCCATCTTATCAGGCACAGAGCTCCAAGCACGATGG 1377
 361 GluArgAlaThrLeuAlaIleIleGlyLeuIleGlyLeuLysGlySerIleLeuSerGlyThrGluLeuGlnAlaArgTrp  387

1378 GTCACAAGAGTATTCAAAGGACTCTGTAAGATACCTCCATCCCAAAAATTGATGATGGAGGCTACTGAAAAGGAACAGCTC 1458
 388 ValThrArgValPheLysGlyLeuCysLysIleProProSerGlnLysLeuMetMetGluAlaThrGluLysGluGlnLeu  414

1459 ATTAAAAGGGGAGTGTTTAAAGACACCAGCAAAGACAAATTTGACTACATTGCCTACATGGATGATATCGCTGCCTGCATA 1539
 415 IleLysArgGlyValPheLysAspThrSerLysAspLysPheAspTyrIleAlaTyrMetAspAspIleAlaAlaCysIle  441

1540 GGCACAAAGCCCAGCATCCCACTTCTGTTCCTCAAGGATCCCAGACTAGCTTGGGAAGTTTTCTTTGGACCATGTACTCCT 1620
 442 GlyThrLysProSerIleProLeuLeuPheLeuLysAspProArgLeuAlaTrpGluValPhePheGlyProCysThrPro  468

1621 TATCAGTACCGCCTCATGGGCCCTGGAAAATGGGATGGAGCCAGAAATGCCATCCTGACCCAGTGGGACAGAACATTGAAA 1701
 469 TyrGlnTyrArgLeuMetGlyProGlyLysTrpAspGlyAlaArgAsnAlaIleLeuThrGlnTrpAspArgThrLeuLys  495

1702 CCTTTAAAAACTCGAATTGTCCCTGATTCCTCCAAGCCTGCCTCCATGTCACATTATTTAAAAGCCTGGGGGGCACCTGTC 1782
 496 ProLeuLysThrArgIleValProAspSerSerLysProAlaSerMetSerHisTyrLeuLysAlaTrpGlyAlaProVal  522

1783 CTACTTGCCTCTCTTCTACTTATCTGTAAATCTTCACTTTTCTTGAAATTGGTGAGAGATAAACTACAGGACAGAATGTCC 1863
 523 LeuLeuAlaSerLeuLeuLeuIleCysLysSerSerLeuPheLeuLysLeuValArgAspLysLeuGlnAspArgMetSer  549

1864 CCTTACCTAGTAAGTCTTTGGCGAGGATGAACCTGATTGTTACAAGGGTTACACCAAGTCATGCTAATTCTATCTCCAAGT 1944
 550 ProTyrLeuValSerLeuTrpArgGly                                                        558

1945 ATCTTGTGCATCCCTCCTCTGCTCTCCATCATAACTGCTATTAGCCAAATTCAGGCCCAGTCATCTCCTATCTGAATTATT 2025
2026 GTATTATCTTCTTCTTTGTTTTCAGTACCCTCTTTCTTGCCACCCTTTCCAATGCATCTTCTACCCTGCTACCTCAGTGAT 2106
2107 TATTCTAAAATAAATATATATGATATGGTTTAAAAAAAAAA                                          2147
```

Fig. 2. Full-length cDNA sequence encoding human FMO2 and the deduced amino acid sequence

Positions of start and stop codons, a polyadenylation signal (AATAAA), putative FAD- and NADP-binding sites and the priming sites for the two 20-mer oligonucleotides used to amplify a section within the 3′ non-coding region are underlined. An ATG upstream of the initiation codon is overlined.

1990; 1991; Dolphin *et al.*, 1991) are present between residues 9–14 and 191–196 respectively (Fig. 2). The polypeptide contains no Asn-Xaa-(Thr/Ser) potential *N*-glycosylation sites.

The corresponding mRNA has a long 5′ leader sequence that contains a single 'upstream' ATG at positions 55–57 (Fig. 2). The sequence surrounding this ATG (a T three bases upstream and a pyrimidine immediately downstream) is particularly unfavourable for translational initiation (Kozak, 1987). In fact, of the 700 vertebrate mRNA sequences that had been determined at that time, not one functional initiation codon had the flanking sequence TNNATGY. Indeed, this sequence is the one most commonly associated with non-functional upstream ATG codons. Furthermore, the ATG was followed closely by an in-frame termination codon (bases 70–72). In contrast, the sequence surrounding the proposed initiation codon, ACCATGG, is the most favourable for translational initiation (Kozak, 1987).

Fig. 3 shows an alignment of the amino acid sequence of human FMO2 with those of rabbit pulmonary FMO (Lawton *et al.*, 1990), human FMO1 (Dolphin *et al.*, 1991) and rabbit liver FMO form 2 (Ozols, 1991). Examination of this Figure reveals several points. (1) The primary sequence of human FMO2 exhibits global identities of 53, 52 and 51% with rabbit pulmonary FMO, human FMO1 and rabbit liver FMO form 2 respectively. (2) A total of 34% of the amino acid residues are

```
001   MAKKVAVIGAGVSGLSSIKCCVDEDLEPTCFERSDDIGGLWKFTESSKDGMTRVYKSLVTNVCKEMSCYSDF   072
001   ***************I*L******G********TE******R*K*NVE**RASI*Q*VI**TS*****F***   072
001   ***R**IV*******A*****LE*G**********L****R***HVEE*RASL***V*S*S**********   072
001   -G****I****I***A**RS*LE*G*******M**********SDHAEE*RASI*Q*VF**SS***M*FP**   071

073   PFHEDYPNFMNHEKFWDYLQEFAEHFDLLKYIQFKTTVCSITKRPDFSETGQWDVVTETEGKQNRAVFDAVM   144
073   *MP**F***LHNS*LLE*FRI**KK*********Q***I*VK*****ASS***E***QSNS**QS*******   144
073   **P*****YVPNSQ*LE**KMY*N******H*****K***V**CS*SAVS***E***MH*E**ES*I*****   144
072   **PP*F**N*HNS*LQE*ITT**REKN*****E***L*S**K*H****V****Y*A*CRN**KET*******   143

145   VCTGHFLNPHLPLEAFPGIHKFKGQILHSQEYKIPEGFQGKRVLVIGLGNTGGDIAVELSRTAAQVLLSTRT   216
145   **S**HIL*NI**KS****E*****YF**RQ**H*A*LE***I****I**SAS*******KK****YI***K   216
145   ****FLT**Y***DS****NA****YF**RQ**H*DI*KD******M**S*T*****A*HL*EK*F***TG   216
144   I*S**HVK*N**KDS***LKH***KSFRQR***E*GI*K**********S*E***T***H**E**VI*S*S   215

217   GTWVLGRSSDWGYPYNMMVTRRCCSFIAQVLPSRFLNWIQERKLNKRFNHEDYGL-SITKGKKAKFIVNDEL   287
217   *S**MS*I*ED***WD*VFHT*FS*MLRN***RMIVK*MM*QQM*RW****N***APEN*YLMKEPVL**D*   288
217   *G**IS*IF*S***WD*VFMT*FQNMLRNS**TPIVT*LM***I*NWL**AN***IPEDRTQLKE*VL****   288
216   *S**MS*VW*D***WD*LYVT*FQT*LKNN**TAISD*WYVKQM*AK*K**NKS*MPLNGTLRKEPVF**D*   287

288   PNCILCGAITMKTSVI---EFTETSAVFEDGTVEENIDVVIFTTGYTFSFPFFEEPLKSLCTKKIFLYKQVF   356
289   *SR**Y*T*KV*RR*---K*L**SA*I********D***IV*A*****A***L**S*VKIEDNMVS***YM*   357
289   *GR*IT*KVFIRP*I---K*VK*N*VI*NNTSK**P**IIV*A*****A***LD*SVVKVEDGQAS***YI*   357
288   *AR****TVSI*PN*KEFK*******I******F*A**S***A***GYAY**LDDSIIKSENN*VT*F*GI*   359

357   PLNLERATLAIIGLIGLKGSILSGTELQARWVTRVFKGLCKIPPSQKLMMEATEKEQLIKRGVFKDTSKDKF   428
358   *PQ**KS*F*CL***QPL***FPTV******A********SL*SKETM*ADIIKRNE-NRIAL*GESLSQ*L   428
358   *AH*QKP********KPL**MIPTG*T****AV**L**VN*L**P-SV*I*EINARKEN*PSW*GLCYCKAL   428
360   *PQ**KP*M*V***VQSL*AAIPT*D*****AAQ*I**T*TL**VKDM*NDIH**MG-T*LKT*-GKW-ETI   428

429   --DYIAYMDDIAACIGTKPSIPLLFLKDPRLAWEVFFGPCTPYQYRLMGPGKWDGARNAILTQWDRTLKPLK   498
429   QTN**D*L*EL*LE**A**DLVSFLF***K**VKLY****NS*****V***Q*E******F**KQ*I*****   500
429   QS***T*I*ELLTY*NA**NLFSML*T**H**LT******S***F**T*****E******M******F*VI*   500
429   QT***N***EL*SF**V*LN**W***T*****L*******S***F**V*****P***Q********S***M*   500

499   TRIVPDSSKPASMSHYLKAWGAPVLLASLLLICKSSLFLKLVRDKLQDRMSPYLVSLWRG               558
501   **TLKA**N-FPV*FL**FL*LFA*VLAF*FQLQWF                                       535
501   A*V*QE*PS*--FESF**VFSFLA**VAIF**FL                                         532
501   **A*GHLQ***LF*PE*WLLAIA***IAAV*VF                                          533
```

Fig. 3. Primary sequence comparisons of mammalian FMOs

The primary sequence of human FMO2 (line 1) is aligned with those of rabbit pulmonary FMO (Lawton *et al.*, 1990) (line 2), human FMO1 (Dolphin *et al.*, 1991) (line 3) and of rabbit liver FMO form 2 (Ozols, 1991) (line 4). Sequences are given in one-letter code; asterisks indicate an identical amino acid residue with human FMO2. To maximize the alignments, a small number of gaps (–) have been introduced into the sequences. The putative FAD- and NADP-binding sites in human FMO2 are shown underlined from positions 9 to 14 and from 191 to 196 respectively.

absolutely conserved among all four FMOs. Most of these conserved residues (70%) are located within two regions: the *N*-terminal 230 residues which contain both the FAD- and NADP-binding sites, and a section stretching from about residue 450 to residue 500. (3) The putative FAD- and NADP-binding sites within human FMO2 align exactly with those proposed for each of the other FMOs. (4) Human FMO2 is approx. 25 amino acid residues longer than any previously described FMO. These extra residues are present as a single block located at the *C*-terminus of the polypeptide. To check the integrity of the *C*-terminal extension of human FMO2, extensive DNA sequencing was performed in the region between the codon for residue 500 and the stop codon of two independent clones, 2A1 and 13W. Both clones yielded identical nucleotide sequences.

In Fig. 4 the hydropathy profile of human FMO2, determined by the method of Kyte & Doolittle (1982) using a window size of nine residues, is compared with those of rabbit pulmonary FMO (Lawton *et al.*, 1990), human FMO1 (Dolphin *et al.*, 1991) and rabbit liver FMO form 2 (Ozols, 1991). The hydropathy profiles of the FMOs display remarkable similarity even in the region between positions 230 and 450, where only 23% of the residues have been conserved among all four proteins. Human FMO2 contains two hydrophobic regions of about 20 residues in length at positions 363–381 and 517–531. Four additional hydrophobic regions, about 15 residues long, are located at positions 5–18, 138–151, 323–337 and 437–452. However, when the hydropathy profile is calculated using a window width of 19, as recommended by Kyte & Doolittle (1982) and von Heijne (1987) for the

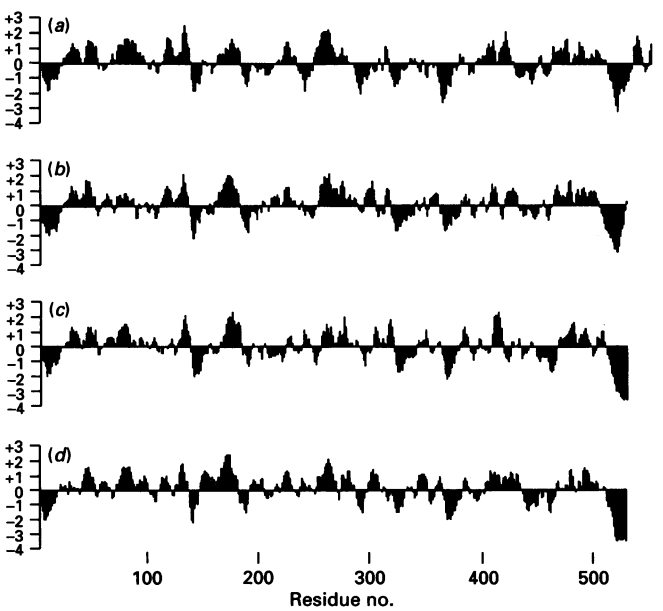

Fig. 4. Hydropathy profiles of mammalian FMOs

Hydropathy profiles of human FMO2 (*a*), rabbit pulmonary FMO (Lawton *et al.*, 1990) (*b*), human FMO1 (Dolphin *et al.*, 1991) (*c*) and rabbit liver FMO form 2 (Ozols, 1991) (*d*) were determined by the method of Kyte & Doolittle (1982) using a window size of 9. Hydrophobic regions are below the line.

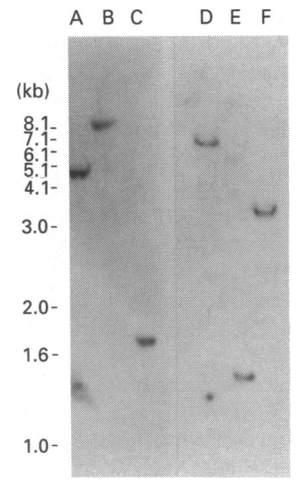

Fig. 5. Southern blot analysis of human genomic DNA

Genomic DNA samples (10 μg) were digested with *Eco*RI (tracks A and D), *Bam*HI (tracks B and E) or *Pst*I (tracks C and F) and hybridized with exon-specific probes for either human FMO1 (tracks A–C) or human FMO2 (tracks D–F). The sizes of DNA standards (1 kb ladder; BRL) are indicated.

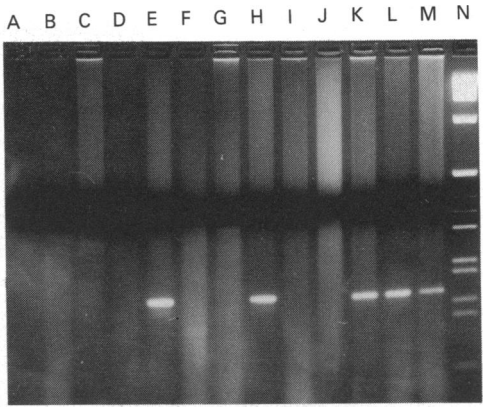

Fig. 6. Chromosomal localization of the human FMO2 gene

DNA products were generated by PCR using the primers indicated in Fig. 2. DNA samples were from mouse cell lines IR (A) and RAG (B), the hamster cell line A23 (C), the rat cell line FAZA (D), human leucocytes (E), and the somatic cell hybrids FST 9/10 (F), CJ9q (G), TWIN 19D12 (H), TWIN 19C5 (I), TWIN 19F6 (J), TWIN 19F9 (K), MOG2C2 (L) and MOG34A4 (M). Size standards (1 kb ladder; BRL) are shown in track N.

identification of membrane-spanning regions, only the sequence between residues 517 and 531 has a mean hydropathy value of greater than 1.6, which is indicative of a potential membrane-spanning segment (results not shown). In other FMOs this hydrophobic region is situated at the extreme *C*-terminus of the polypeptide, but in human FMO2 it is followed by a short stretch of hydrophilic residues that are predicted, by the methods of both Chou & Fasman (1978) and Garnier *et al.* (1978), to adopt an α-helical secondary structure (results not shown).

Although the cDNA clones encoding FMO2 were isolated from an adult human liver library, Northern blot hybridization failed to detect the corresponding mRNA in total RNA isolated from four different adult livers (results not shown). The mRNA was also undetectable in total RNA isolated from adult kidney and lung, and from foetal liver.

Human genomic DNA was analysed by Southern blot hybridization using probes derived from the 3′ non-coding regions of FMO1 and FMO2 cDNAs (Fig. 5). The probes comprise DNA sequences that are contained within single exons of the corresponding genes (Fig. 6; Dolphin *et al.*, 1991). Both cDNA probes hybridized to single fragments in genomic DNA digested by *Eco*RI, *Bam*HI or *Pst*I (Fig. 5). The FMO1 cDNA probe

Table 1. Segregation of human *FMO2* in relation to human chromosomal content in 16 human–rodent somatic cell hybrids

+, Human chromosome present; −, human chromosome not detected; ○, not tested or equivocal result. All equivocal results have been ignored in the summary of concordance and discordance. References: 1, Wong *et al.* (1987); 2, Solomon *et al.* (1979); 3, Andrews *et al.* (1985); 4, Kielty *et al.* (1982); 5, Jones & Kao (1984).

Hybrid	Ref.	PCR product	Chromosome no. 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
Twin 19D12	1	+	+	−	+	+	−	+	−	+	−	−	−	+	−	+	−	+	+	+	−	+	+	+	−
Twin 19F9	1	+	+	+	+	+	−	+	−	+	+	−	−	+	○	+	−	−	+	+	−	+	+	+	−
MOG2E5	1	+	+	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	−	○	+	+	+
MOG2C2	1	+	+	−	+	+	+	+	+	+	+	+	+	−	−	+	+	+	+	+	+	+	+	+	+
MOG 34A4	2	+	+	−	+	+	+	+	−	−	−	+	+	+	−	+	−	−	−	+	+	−	+	+	+
HORL 411 B6P	3	+	+	−	+	−	−	−	−	−	−	−	+	−	+	−	+	−	−	+	−	−	−	+	+
TWIN 19C5	1	−	−	−	+	+	−	−	−	+	−	−	−	+	○	+	−	−	+	+	−	+	+	+	−
TWIN 19F6	1	−	−	+	+	+	−	+	−	+	+	−	−	+	○	+	−	−	+	+	−	+	+	+	−
FST 9/7	4	−	−	−	+	+	−	+	−	+	−	+	−	+	+	+	+	−	−	+	−	+	−	+	+
FST 7	4	−	−	−	−	−	−	−	+	−	−	+	+	○	−	+	−	+	+	+	○	○	−	+	+
FIR 5	1	−	−	−	−	−	○	−	○	−	−	−	−	−	○	−	−	−	−	+	−	−	−	−	+
FG10E8B	1	−	−	+	−	−	+	−	−	+	−	−	−	−	−	−	+	−	−	−	−	−	+	−	−
3WCl 10	1	−	−	−	−	−	−	−	+	−	−	+	+	+	−	+	−	−	+	−	−	−	+	−	+
3W4Cl 5	1	−	−	−	−	−	−	−	+	−	−	○	+	+	−	+	+	−	+	−	−	−	+	−	+
FST9/10	1	−	−	−	+	+	−	+	−	+	○	+	−	+	+	+	+	−	−	+	−	+	−	+	+
640-3a12	5	−	−	−	−	−	−	−	−	−	+q	−	−	−	−	−	−	−	−	−	−	−	−	−	−
FMO/chromosome Concordance		++	6	1	6	5	3	5	2	4	3	3	3	4	2	5	3	3	4	6	2	3	5	6	4
		−−	10	8	6	6	8	7	6	5	7	5	7	3	5	3	6	9	5	4	9	5	5	5	4
Discordance		+−	0	5	0	1	3	1	4	2	3	3	3	2	3	1	3	3	2	0	4	2	1	0	2
		−+	0	2	4	4	1	3	3	5	2	4	3	6	2	7	4	1	5	6	0	4	5	5	6

detected fragments of 5.0 kb (*Eco*RI), 8.5 kb (*Bam*HI) and 1.7 kb (*Pst*I) (Fig. 5, tracks A–C), whereas the FMO2 cDNA hybridized to fragments of 7.1 kb (*Eco*RI), 1.4 kb (*Bam*HI) and 3.5 kb (*Pst*I) (Fig. 5, tracks D–F). These results indicate that both FMO1 and FMO2 are the products of single, distinct, genes.

The chromosomal localization of the human FMO2 gene, designated *FMO2*, was determined by means of a PCR. The oligonucleotide primers used (see Fig. 2) direct the selective amplification from human genomic DNA of a 172 bp fragment (Fig. 6, track E) corresponding to a sequence located entirely within the 3′ non-coding region of the cloned cDNA. No amplification product was obtained from DNA prepared from mouse, rat or hamster donor cell lines (Fig. 6, tracks A–D). Analysis of DNA isolated from 16 human–rodent somatic cell hybrids showed that the 172-bp fragment was amplified only from DNA samples isolated from hybrids that contained human chromosome 1 (Fig. 6 and Table 1). The results shown in Table 1 demonstrate complete correlation between the amplified DNA product (indicative of the human FMO2 gene) and the presence of human chromosome 1. There are at least four cases of discordance for each of the other human chromosomes.

DISCUSSION

We have previously isolated cDNA clones that encode the human orthologue of rabbit and pig hepatic FMOs (designated FMO1) (Dolphin *et al.*, 1991). The primary sequence of human FMO1 displays 86–88 % similarity to its animal orthologues, but less than 60 % similarity to two non-orthologous forms of FMO, namely rabbit pulmonary FMO (Lawton *et al.*, 1990) and rabbit liver FMO form 2 (Ozols, 1991). We now report the isolation of cloned cDNAs that encode a second form of human FMO, which we have designated FMO2. The degree of primary sequence similarity between human FMO2 and other forms of the enzyme (between 51 and 53 %) indicates that it is not orthologous to any of these and thus represents a fourth, distinct, member of the mammalian FMO family.

A comparison of representatives of the four distinct forms of mammalian FMOs identified to date (Fig. 3) reveals that the degree of sequence identity between these proteins is not uniform throughout the polypeptide chain. The relatively high percentage of residues between positions 1 and 230, and between 450 and 500, that are conserved among all four proteins indicates important structural and/or functional roles for these regions, the former of which contains the putative FAD- and NADP-binding sites. Although the region between residues 230 and 450 is less conserved among different FMOs, it contains sections which are highly conserved among FMO1 orthologues (Dolphin *et al.*, 1991), suggesting that sequences within this region may be responsible for inter-form differences in FMO function.

At codons 531 and 532, a position that corresponds very closely to the end of other forms of FMO, human FMO2 has the sequence TGTAAA (Fig. 2). Thus the extra residues at the *C*-terminus of the polypeptide may have resulted from a point mutation in an ancestral FMO2 gene that either changed the third base of a TGA stop codon at position 531 or altered the reading frame of a TAA stop codon, in either case allowing translation to continue until the next in-frame termination codon at position 559.

Korsmeyer *et al.* (1990) have demonstrated that pig FMO1 is glycosylated at one or both of its two possible Asn-Xaa-Thr/Ser *N*-glycosylation sites. Although the glycosylation status of other FMOs has not been determined, all FMO sequences reported to date contain at least one such potential oligosaccharide acceptor site. In contrast, human FMO2 contains no such sites, indicating that it is not subject to modification by N-glycosylation. Peptide sequencing has revealed that the initiation methionine residue is absent from the mature forms of pig (Guan *et al.*, 1990) and rabbit (Ozols, 1990) FMO1 and from rabbit liver FMO form 2 (Ozols, 1991). Although examination of sequences deduced from cDNAs does not allow conclusions concerning the nature of possible *N*-terminal modifications of the mature protein, it is of interest that in the deduced amino acid sequence of human FMO2, as in those of human FMO1 (Dolphin *et al.*, 1991) and rabbit pulmonary FMO (Lawton *et al.*, 1990), the initiation methionine is followed by alanine, a residue that appears to promote the removal of methionine during protein maturation (Flinta *et al.*, 1986).

The hydrophobic stretch of amino acid residues at the *N*-terminus of human FMO2 has some similarities to a signal peptide sequence. However, the presence of a putative FAD-binding site within this region suggests that it may not function as a membrane-insertion domain. Indeed, the only hydrophobic region of the protein that fully meets the criteria of Kyte & Doolittle (1982) for a potential membrane-spanning sequence is that located near the *C*-terminus of the protein. Despite the fact that FMOs exhibit very little sequence identity within this region, the hydrophobic nature of the region has been well conserved and thus it is likely that this domain serves to anchor FMOs in the membranes of the endoplasmic reticulum.

The results of Northern blot hybridization demonstrate that the human FMO2 gene is not expressed at a high level in any of the tissues associated with FMO activities in experimental animals, namely liver, lung and kidney. However, as we were able to isolate cDNA clones for human FMO2 from a library constructed from adult liver RNA, the gene must be expressed in this tissue, albeit at a low level. The frequency of independent FMO2 clones in the library (approx. 1 in 2×10^4) suggests that the abundance of the corresponding mRNA, although very low, is far higher than that of mRNA encoding human FMO1 (which has a cDNA clone frequency of 1 in 7.5×10^5) (Dolphin *et al.*, 1991). In contrast to FMO1, FMO2 is not expressed at a high level in foetal liver.

At present it is unclear which form(s) of human FMO may be responsible for the N-oxidation of TMA and hence may be involved in the metabolic disorder trimethylaminuria. *In vitro* studies to elucidate the relative contributions of human FMO1 and FMO2 to TMA metabolism may help to answer this question.

We have demonstrated that human FMO1 and FMO2 are each encoded by single genes, and have mapped the FMO2 gene (which we have designated *FMO2*) to human chromosome 1. The FMO1 gene is also located on this chromosome (Dolphin *et al.*, 1991), and we have evidence that both the FMO1 and FMO2 genes are situated on the long arm of chromosome 1 (Shephard *et al.*, 1991), raising the possibility that the genes form part of a gene cluster.

Comparison of the sequences of orthologous FMOs indicates that, for these proteins, the time required to establish a 1 % difference in amino acid sequence, known as the unit evolutionary period, is about 6 million years (Dolphin *et al.*, 1991). Based on this rate of evolution and the degree of sequence similarity between non-orthologous forms, it would appear that the known members of the FMO gene family arose from gene duplication events that occurred some 250–300 million years ago.

We thank the Wellcome Trust for financial support, Dr. S. Woo for the adult human liver λgt11 cDNA library, Dr. N. Spurr for some of the somatic cell hybrids and the MRC Tissue Bank for tissue samples.

REFERENCES

Al Waiz, M., Ayesh, R., Mitchell, S. C., Idle, J. R. & Smith, R. L. (1987) Lancet **i**, 634–635

Andrews, P. W., Knowles, B. B., Parkar, M., Pym, M., Stanley, K. & Goodfellow, P. N. (1985) Ann. Hum. Genet. **49**, 31–39
Blin, N. & Stafford, D. W. (1976) Nucleic Acids Res. **3**, 2303–2308
Chou, P. Y. & Fasman, G. D. (1978) Annu. Rev. Biochem. **47**, 251–276
Dolphin, C., Shephard, E. A., Povey, S., Palmer, C. N. A., Ziegler, D. M., Ayesh, R., Smith, R. L. & Phillips, I. R. (1991) J. Biol. Chem. **266**, 12379–12385
Edwards, Y. H., Parkar, M., Povey, S., West, L. F., Parrington, J. M. & Solomon, E. (1985) Ann. Hum. Genet. **49**, 101–109
Feinberg, A. & Vogelstein, B. (1983) Anal. Biochem. **132**, 6–10
Flinta, C., Persson, B., Jornvall, H. & von Heijne, G. (1986) Eur. J. Biochem. **154**, 193–196
Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) J. Mol. Biol. **120**, 97–120
Gasser, R., Tynes, R. E., Lawton, M. P., Korsmeyer, K. K., Ziegler, D. M. & Philpot, R. M. (1990) Biochemistry **29**, 119–124
Guan, S., Falick, A. M. & Cashman, J. R. (1990) Biochem. Biophys. Res. Commun. **170**, 937–943
Hlavica, P. & Kehl, M. (1977) Biochem. J. **164**, 487–496
Humbert, J. R., Hammond, K. B., Hathaway, W. E., Marcoux, J. G. & O'Brien, D. (1970) Lancet **i**, 770–771
Jones, C. & Kao, F. (1984) Cytogenet. Cell Genet. **37**, 499
Kielty, C. M., Povey, S. & Hopkinson, D. A. (1982) Annu. Hum. Genet. **46**, 307–327
Kimura, T., Kodama, M. & Nagata, C. (1983) Biochem. Biophys. Res. Commun. **110**, 640–645
Korsmeyer, K. K., Poulsen, L. L. & Ziegler, D. M. (1990) in Flavins and Flavoproteins: 10th International Conference on Flavins and Flavoproteins, pp. 243–246, Walter de Gruyter & Co., Berlin
Kozak, M. (1987) Nucleic Acids Res. **20**, 8125–8133
Kyte, J. & Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132
Lawton, M. P., Gasser, R., Tynes, R. E., Hodgson, E. & Philpot, R. M. (1990) J. Biol. Chem. **265**, 5855–5861
Lawton, M. P., Kronbach, T., Johnson, E. F. & Philpot, R. M. (1991) Mol. Pharmacol. **40**, 692–698
Ozols, J. (1989) Biochem. Biophys. Res. Commun. **163**, 49–55
Ozols, J. (1990) J. Biol. Chem. **265**, 10289–10299
Ozols, J. (1991) Arch. Biochem. Biophys. **290**, 103–115
Sabourin, P. J., Smyser, B. P. & Hodgson, E. (1984) Int. J. Biochem. **16**, 713–720
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Shelley, E. D. & Shelley, W. D. (1984) J. Am. Med. Assoc. **251**, 253–255
Shephard, E. A., Dolphin, C., Povey, S., Smith, R. L. & Philips, I. R. (1991) Cytogenet. Cell Genetics **58**, 1863
Solomon, E., Swallow, D., Burgess, S. & Evans, L. (1979) Ann. Hum. Genet. **42**, 273–281
Tynes, R. E., Sabourin, P. J., Hodgson, E. & Philpot, R. M. (1986) Arch. Biochem. Biophys. **251**, 654–664
Venkatesh, K., Levi, P. E. & Hodgson, E. (1991) Biochem. Pharmacol. **42**, 1411–1420
von Heijne, G. (1987) Sequence Analysis in Molecular Biology, Academic Press, London
Williams, D. E., Ziegler, D. M., Nordin, D. J., Hale, S. E. & Masters, B. S. S. (1984) Biochem. Biophys. Res. Commun. **125**, 116–122
Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. J. (1987) Ann. Hum. Genet. **51**, 260–288
Ziegler, D. M. (1980) in Enzymatic Basis of Detoxication (Jakoby, W. B., ed.), vol. I, pp. 201–207, Academic Press, New York
Ziegler, D. M. (1988) Drug Metab. Rev. **19**, 1–32
Ziegler, D. M. & Mitchell, C. H. (1972) Arch. Biochem. Biophys. **150**, 116–125

Received 13 February 1992/6 April 1992; accepted 13 April 1992