Alternative promoters and repetitive DNA elements define the species-dependent tissue-specific expression of the *FMO1* genes of human and mouse

Elizabeth A. SHEPHARD*[1], Pritpal CHANDAN*, Milena STEVANOVIC-WALKER†, Mina EDWARDS* and Ian R. PHILLIPS†[1]

*Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, U.K., and †School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, U.K.

In humans, expression of the *FMO1* (flavin-containing mono-oxygenase 1) gene is silenced postnatally in liver, but not kidney. In adult mouse, however, the gene is active in both tissues. We investigated the basis of this species-dependent tissue-specific transcription of *FMO1*. Our results indicate the use of three alternative promoters. Transcription of the gene in fetal human and adult mouse liver is exclusively from the P0 promoter, whereas in extra-hepatic tissues of both species, P1 and P2 are active. Reporter gene assays showed that the proximal P0 promoters of human (h*FMO1*) and mouse (m*Fmo1*) genes are equally effective. However, sequences upstream (−2955 to −506) of the proximal P0 of m*Fmo1* increased reporter gene activity 3-fold, whereas h*FMO1* upstream sequences (−3027 to −541) decreased reporter gene activity by 75%. Replacement of the upstream sequence of human P0 with the upstream sequence of mouse P0 increased activity of the human proximal P0 8-fold. Species-specific repetitive elements are present immediately upstream of the proximal P0 promoters. The human gene contains five LINE (long-interspersed nuclear element)-1-like elements, whereas the mouse gene contains a poly A region, an 80-bp direct repeat, an LTR (long terminal repeat), a SINE (short-interspersed nuclear element) and a poly T tract. The rat and rabbit *FMO1* genes, which are expressed in adult liver, lack some (rat) or all (rabbit) of the elements upstream of mouse P0. Thus silencing of *FMO1* in adult human liver is due apparently to the presence upstream of the proximal P0 of L1 (LINE-1) elements rather than the absence of retrotransposons similar to those found in the mouse gene.

Key words: alternative promoter, flavin-containing mono-oxygenase 1 (FMO1), gene regulation, long-interspersed nuclear element (LINE), promoter, retrotransposon.

INTRODUCTION

Mammalian FMOs (flavin-containing mono-oxygenases; EC 1.14.13.8) are microsomal enzymes that catalyse the NADPH-dependent mono-oxygenation of numerous foreign chemicals including therapeutic drugs and environmental pollutants [1]. These enzymes thus constitute an important interface between the organism and its chemical environment. In human and in mouse, the *FMO1*, *2*, *3*, *4* and *6* genes are clustered on chromosome 1 [2,3]. The *FMO5* gene lies outside the cluster; in humans it is located on chromosome 1 [2] and in mouse it is on chromosome 3, in a region of synteny between mouse and human [3]. A second *FMO* gene cluster, also located on chromosome 1 of both species, encodes five pseudogenes in human and, in the mouse, three genes that are not known to be expressed [3]. Although functional *FMO* genes show evolutionary conservation with respect to both sequence and organization, changes have occurred in individual genes, which markedly influence the species-specific expression of the *FMO1*, *2* and *3* genes [3–6].

The reasons for these species-specific differences in *FMO* gene expression differ. For instance, in most humans the *FMO2* gene encodes a non-functional protein, because of a C > T mutation that converts a glutamine residue codon at position 472 into a stop codon [5]. In mouse and all other species examined to date, including the chimpanzee, the *FMO2* gene encodes a glutamine residue at position 472 and hence a full-length FMO2 protein [5,7]. In humans, with the exception of individuals who suffer from trimethylaminuria, the *FMO3* gene is most highly expressed in the liver [8–10], whereas in male mice expression of the *Fmo3* gene in liver is switched off 5 weeks after birth [6,11]. The hepatic silencing of the *Fmo3* gene in male mice has been shown to be mediated by hormonal factors [6].

In humans, expression of the *FMO1* gene in liver is switched off shortly after birth, but the gene continues to be expressed in adult kidney [2,4,9]. Silencing of *FMO1* gene expression in adult liver is specific to humans. In all other mammals studied, e.g. pig [12], rat [13], rabbit [14], mouse [11,15] and dog [16], the gene continues to be expressed in the liver after birth.

A consequence of this species difference in *FMO1* expression is that, in adult humans, the contribution of this protein to detoxification is extra-hepatic. In contrast, in laboratory animals used in drug metabolism studies, FMO1 is a major form of the enzyme present in adult liver. An increasing number of therapeutic drugs, including tamoxifen, itopride, benzydamine, olopatidine and xanomeline, have been shown to be substrates for human FMO1 [1]. In such cases, extrapolation of drug metabolism data derived from experimental animals and *in vitro* systems requires careful consideration.

In the present paper, we report the use of functional transcription assays and comparative gene analyses to identify DNA sequences that play a role in the species-specific extinction of *FMO1* gene expression in human liver. We show that the presence of species-specific repetitive DNA elements and the use of alternative tissue-specific promoters in liver and kidney can account for the

Abbreviations used: FMO, flavin-containing mono-oxygenase; HepG2 cells, human hepatocellular carcinoma G2 cells; HNF, hepatocyte nuclear factor; HFH, HNF3 homologue; LINE, long-interspersed nuclear element; L1, LINE-1; LTR, long terminal repeat; ORF, open reading frame; RT, reverse transcriptase; SINE, short-interspersed nuclear element; TAP, tobacco acid pyrophosphatase; TAP-RLPCR, TAP reverse ligation-mediated PCR; TSD, target-site domain; 3′-UTR, 3′-untranslated region.

[1] Correspondence may be addressed to either of these authors (email e.shephard@ucl.ac.uk or i.r.phillips@qmul.ac.uk).

differential developmental and tissue-specific expression of the *FMO1* gene in human and mouse.

EXPERIMENTAL

Identification of transcriptional start sites

RNA was isolated from liver and kidney of 8-week-old female C57BL/6 mice. The *Fmo1* exon 2 primer 5′-gggtgttaacggtgagcg-aa-3′ (Eurogentec, Hampshire, U.K.) was end-labelled with [γ-^{32}P]ATP and mixed with 12 μg of total RNA. A primer extension reaction was carried out for 1 h with AMV RT (reverse transcriptase) (Roche, Lewes, East Sussex, U.K.). For RT–PCR reactions, RNA from liver and kidney was reverse-transcribed by using the *Fmo1* exon 5 primer 5′-cccttaaaagttagtatacct-3′, and amplified with Taq DNA polymerase (Roche Molecular Biochemicals) using the same primer and one located in intron 1 of the gene (5′-gcacaccacacagatagtct-3′). In addition, RNA was reverse-transcribed by using the exon 2 primer described above and amplified with this primer and one located in exon 0 (5′-gctctgggat-cctaattgtgt-3′). The amplified products were cloned and sequenced.

The transcriptional start sites of the human *FMO1* gene were determined by TAP-RLPCR [TAP (tobacco acid pyrophosphatase) reverse ligation-mediated PCR], as described by Fromont-Racine et al. [17]. Fetal liver and adult kidney samples were obtained from the MRC Tissue Bank (Royal Marsden Hospital, London, U.K.) and St Mary's Hospital (London, U.K.) respectively as described in [8]. Total RNA was extracted by the use of TRIzol® (Invitrogen, Paisley, Renfrewshire, Scotland, U.K.). RNA was incubated with calf intestinal alkaline phosphatase (New England Biolabs, Hitchin, Herts., U.K.). This removed phosphate groups from the 5′-end of partially degraded, i.e. non-capped, mRNAs. The RNA was then treated with TAP (5 units; Epicenter Technologies, Madison, WI, U.S.A.), which hydrolysed the 5′–5′-phosphodiester-linked cap structure from full-length mRNAs. An RNA linker 5′-gggcauaggcugacccucgcugaaa-3′ was synthesized from a partially double-stranded DNA comprising a 17-nt sequence 5′-taatacgactcactata-3′ bound to a 42-nt sequence 3′-attatgctgagtgatatcccgtatccgactgggagcgacttt-5′. The 17-bp double-stranded region of the DNA contained a T7 promoter, and the 25-nt single-stranded region provided a template for RNA synthesis. An RNA copy of the single-stranded region was produced by transcription with T7 RNA polymerase [18]. The RNA linker was ethanol-precipitated and electrophoresed through a 12 % (w/v) polyacrylamide gel containing 8 M urea. Nucleic acids were visualized by UV shadowing and the band corresponding to the 25-mer linker was excised from the gel, eluted from the polyacrylamide and purified by reverse-phase chromatography on a Sep-Pack C18 column (Waters Associates, Milford, MA, U.S.A.).

RNA (1 μg), treated as described above, was ligated to 100 ng of RNA linker with T4 RNA ligase. Ligated RNAs were reverse-transcribed by using the *FMO1* primer +203/+229 (5′-cttctggg-aatggaaagtctgagtaac-3′) (numbers are relative to the A of the ATG translational initiation codon). The resulting product was used as a template for PCR by using the linker-specific primer DNAPr-1 (5′-gggcataggctgaccctcgctg-3′) and an *FMO1* primer (−71 to −90, relative to the A of the ATG) (5′-atcagtatgagccagtgctg-3′) labelled at its 5′-end with [γ-^{32}P]ATP (> 5000 Ci/mmol) through the use of a 5′-end-labelling kit (Amersham Biosciences). The PCR was catalysed using ThermoZyme DNA polymerase (Invitrogen).

For sequence analysis of transcriptional start sites, RNA from human fetal liver or adult kidney was ligated to the RNA linker and reverse-transcribed with random primers. The products were amplified by semi-nested PCR using the linker-specific primer (DNA-Pr1) and *FMO1* primers +373/+399 (5′-ctcttcatgcatagtg-accacctccca-3′) and +203/+229. DNAs were amplified by 'touchdown' PCR [19] in a GeneAmp PCR System 96 thermal cycler (PerkinElmer, Norwalk, CT, U.S.A.). Amplified products were cloned into pCR4-TOPO cloning vector (Invitrogen) and sequenced.

Mining cDNA sequences

Human and mouse FMO1 cDNAs were identified from BLASTn analyses (http://www.ncbi.nlm.nih.gov/entrez/) using clones M64082 [4] and NM_010231 [15] respectively as query sequences. Human cDNAs identified were BC047129 [IMAGE consortium (Integrated Molecular Analysis of Genomes and their Expression consortium); http://www.image.llnl.gov] and AK097039 [20]. Mouse cDNAs identified were D16215 [13], BC011229 [21], BF532824, AI115B9, AA245076, AI255718, AA238774 and BI247068 (IMAGE consortium), all derived from liver mRNAs, and BF784152, CB954312, CB599568, AI118998 and CB955318 (IMAGE consortium), derived from kidney mRNAs.

Promoter-reporter gene constructs

The parent plasmid for each construct was pGL3 Basic (Promega, Southampton, U.K.). Oligonucleotides used to prime amplification of promoter sequences of *FMO1* genes of human and mouse are given below. Restriction sites, for insertion of amplified products into the parent plasmid, were included in the primers and are shown in bold-face. Human sequences were amplified from human genomic DNA by using the reverse primer +27, 5′-c**aagctt**ccccagcacagtggataaac-3′, and forward primer −544, 5′-c**gagctc**ccactcgatcatgcctattt-3′, or −3027, 5′-c**gagctc**-gccctgctcatcacattca-3′, to produce plasmids pGL-544(H) and pGL-3027(H) respectively. Mouse sequences were amplified from mouse genomic DNA by using the reverse primer +50, 5′-c**aagctt**gggagttccctgcacacaggat-3′, and forward primer −431, 5′-c**gagctc**gccaggactcatcatgacttcgaa-3′, or −2955, 5′-c**gagctc**gg-catggcatgaaaggaaaa-3′, to produce plasmids pGL3 −431 (M) and pGL3 −2955(M) respectively. The construct pGL −544(H)/−2955(M) was prepared by cloning an amplified mouse product (forward primer −2955 and reverse primer −506, 5′-c**gagctc**-gggaatgcaagacagatgtgtg-3′) upstream of the human proximal promoter in pGL-544(H). The products were amplified by using BIO-X-ACT Short or Long DNA polymerase (Bioline, London, U.K.) as appropriate.

Cell transfection and reporter gene assays

HepG2 cells (passage 9–13) were obtained from the European Collection of Animal Cell Culture. Cells were cultured in Williams' E medium (Sigma–Aldrich) supplemented with gentamicin (50 μg/ml) and 10 % (v/v) fetal calf serum. Cells were transfected at 70–75 % confluency with the reporter constructs shown in Figure 5. Each 60-mm plate was transfected with 5 μg of a reporter gene construct and 0.25 μg of pRL-TK (Promega, Madison, WI, U.S.A.), as a control for transfection efficiency, using Tfx20 reagent (Promega) in the ratio of 3:1 (TfX20/DNA). Luciferase reporter gene activity was measured after 48 h by using the Dual-luciferase reporter system (Promega) according to the manufacturer's recommendations.

Comparative sequence analyses

Human and mouse *FMO1* genomic sequences were downloaded from the Wellcome Trust Sanger Centre, U.K., at http://www.

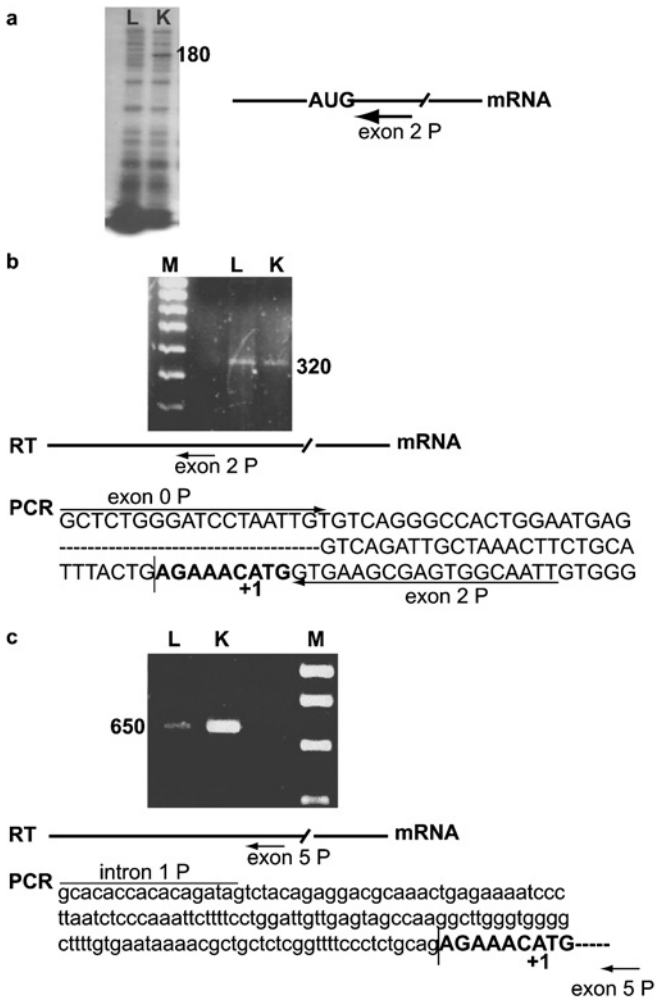

Figure 1 Analysis of transcriptional start sites of mouse *Fmo1* in adult liver and kidney

(**a**) A ^{32}P-labelled *Fmo1*-specific primer, located in exon 2, downstream of the AUG translation initiation codon, was used to prime synthesis from adult mouse liver (L) or kidney (K) total RNA. Extended products were electrophoresed and autoradiographed. The most abundant extended product (180 nt) obtained from kidney RNA is indicated. (**b**) RT–PCR analysis of RNA isolated from liver (L) and kidney (K). Reverse transcription was primed with the same exon 2 primer used in (**a**). PCR was primed with the exon 2 primer and a primer located in exon 0. (**c**) RT–PCR analysis of RNA isolated from liver (L) and kidney (K). Reverse transcription was primed with a primer located in exon 5. PCR was primed with the exon 5 primer and one located in intron 1. In both (**b**) and (**c**), sequences determined from the amplified products are shown. In the former, sequences at the 5′- and 3′-ends are shown. In the latter, only the sequence at the 5′-end is given. In each case, the vertical line represents the 5′ boundary of exon 2 and '+1' indicates the A of the translation initiation codon.

ensembl.org/Homo_sapiens/ and http://www.ensembl.org/Mus_musculus/ (Ensemblv37). Rat *FMO1* sequences were downloaded from http://genome.ucsc.edu/ (release June 2003). Rabbit *FMO1* promoter sequence was from [22]. Sequences were aligned using the dot matrix facility of MacVector 6.5.3 and VISTA (http://www.gsd.lbl.gov/vista) [23,24]. Transcription factor sequences were analysed by using MacVector 6.5.3 and AliBaba2 (http://www.alibaba2.com/) [25]. Conserved motifs were identified using DNA Footprinter (http://bio.cs.washington.edu/software.html) [26] and ConSite (http://mordor.cgb.ki.se/cgi-bin/ConSite/consite/) [27].

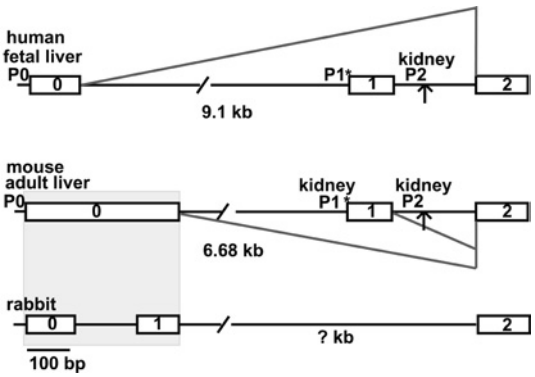

Figure 2 Alternative promoters are used in the transcription of the *FMO1* gene in human and mouse

The positions in the *FMO1* genes of human and mouse of three promoters, P0, P1 and P2, are shown. The grey lines indicate the favoured (predominant) alternative splicing patterns used in liver and kidney. Boxes represent exons. The vertical arrow indicates the transcriptional start site located in intron 1; '*' indicates that the transcriptional start site for P1 is inferred from cDNA clones. The shaded grey area indicates that mouse exon 0 is equivalent to the combined exon 0, intron 0 and exon 1 of rabbit [3]. It has previously been shown that, in rabbit, two alternative promoters (upstream of exon 0 and exon 1 respectively) can be used in the adult liver and that the 5′-most promoter is the most active [22].

RESULTS

Tissue-specific use of alternative promoters in the mouse *Fmo1* gene

Transcriptional start sites of *Fmo1* were determined by primer extension. An *Fmo1*-specific primer complementary to the sequence +26 to +7, relative to the A of the AUG translation initiation codon, was used to prime synthesis from adult mouse liver and kidney RNAs. A complex pattern of extended products was obtained from each of the two tissues (Figure 1a). Although these patterns are similar, a product of 180 nt is clearly more abundant in kidney than in liver, indicating a potential tissue-specific difference in the transcriptional start of *Fmo1*.

Analysis of available mouse liver FMO1 cDNA clones (NM_01023, D16215, BC011229, BF532824, AI115B9, AA245076, AI255718, AA238774 and BI247068) shows that although their leader sequences differ in length, all are derived from what we now call exon 0 spliced to sequences derived from exon 2 (Figure 2). RT–PCR and sequence analysis of the amplified products confirm that mouse liver FMO1 mRNAs are derived from the splicing of exon 0 to exon 2 (Figure 1b). These two exons are separated by a distance of ~6.68 kb. Thus, in mouse liver, transcription of *Fmo1* is initiated from a promoter (P0) located upstream of exon 0. Of the five available kidney FMO1 cDNAs, three, BF784152, CB954312 and CB599568, have leader sequences derived from the 3′-end of intron 1 and exon 2. RT–PCR and sequence analysis of the amplified products confirmed that, in kidney, transcription can occur from within intron 1, from a promoter designated P2 (Figure 1c). The two other kidney cDNAs, AI118998 and CB955318, are the products of a splicing event between exon 2 and a novel exon we now call exon 1 (Figure 2). These two exons are separated by a short intron of 231 bp. Transcription of these mRNAs starts from a promoter (P1) located upstream of exon 1 (Figure 2). Thus the results of the RT–PCR experiments confirm the use of alternative promoters and promoter-driven splicing events in the expression of the mouse *Fmo1* gene.

Although the results of the RT–PCR experiments suggest that the P0 and P2 promoters can be used in both liver and kidney, there is a preference for P0 in liver and a marked preference for

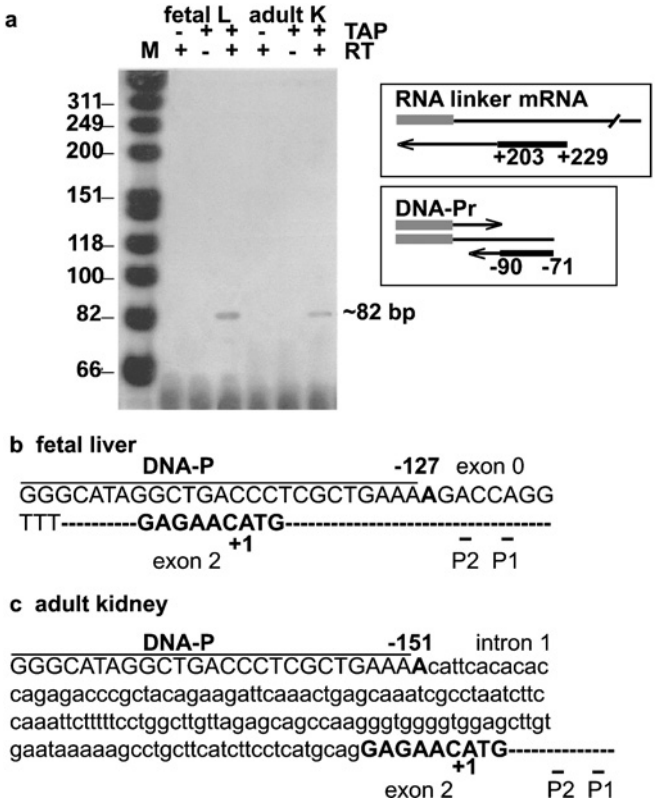

Figure 3 Mapping of the transcriptional start site of the human *FMO1* gene in fetal liver and adult kidney

(**a**) TAP-RLPCR of RNA isolated from human fetal liver (L) and adult kidney (K). The strategies for the extension and subsequent amplification reactions are described in the Experimental section and are illustrated in the right hand panels. Reactions were carried out in the absence (−) or presence (+) of TAP or RT, as indicated. (**b**, **c**) Products of TAP reverse ligation amplified by a semi-nested PCR were cloned and the products were fully sequenced. Partial sequences of products obtained from fetal liver (**b**) and adult kidney (**c**) RNAs are shown. Horizontal lines indicate primer sequence; upper-case letters that are not part of the primer represent exon sequences; the A in boldface corresponds to the major transcriptional start site; '+1' refers to the A of the translation initiation codon. The fetal liver mRNA has a leader sequence of 127 nt, derived mainly from exon 0, whereas the kidney mRNA has a leader of 151 nt, derived mainly from intron 1 sequence.

P2 in kidney (Figures 1b and 1c). This is supported strongly by the finding that all nine of the liver-derived cDNAs are transcribed from P0, whereas none of the five kidney-derived cDNAs is transcribed from this promoter. Therefore, taken together, the results indicate a marked *in vivo* tissue-specific preferential, if not exclusive, use of the P0 promoter in liver and the P1 and P2 promoters in kidney.

Tissue-specific use of alternative promoters in the human *FMO1* gene

Primer extension analysis of mouse FMO1 mRNA yields a complex pattern of extended products (see above). A similar pattern was reported previously for the rabbit [22] and human *FMO1* mRNAs [28]. Consequently, to define the start sites of transcription for the human *FMO1* gene in fetal liver and adult kidney, we used TAP-RLPCR [29]. This method is advantageous because only mRNAs that are capped at their 5′-ends (and hence are full-length) are amplified.

Human fetal liver and adult kidney RNAs were reverse-transcribed and then amplified by PCR (Figure 3a). A PCR product of approx. 82 bp was obtained from fetal liver and, in less abundance, also from adult kidney RNAs (Figure 3a), indicative of an mRNA that extends 127 nt upstream of the AUG translation initiation codon. To define the transcriptional start site, a second TAP-RLPCR was performed using semi-nested PCR with *FMO1* gene-specific primers corresponding to sequences within exons 4 and 3 respectively. A 380-bp fragment was amplified from fetal liver and adult kidney (results not shown). An additional 400-bp fragment was amplified only from adult kidney. The amplified products were cloned and sequenced. Of ten clones of the 380-bp product, eight corresponded to a leader sequence extending 127 nt upstream of the A of the AUG initiation codon and two to a leader extending 125 nt upstream of the AUG. The 5′-ends of these mRNAs correspond to transcriptional start sites located −9594 bp and −9592 bp respectively upstream of the A of the ATG translation initiation codon. These start sites are used in fetal liver and adult human kidney. In both cases, exon 0 of the *FMO1* gene is spliced to exon 2. Sequence analysis of clones corresponding to the 400-bp adult kidney-specific PCR product showed an alternative transcriptional start site 151 bp upstream of the translation initiation codon (Figure 3c). The leader sequence of these mRNAs is derived from intron 1 sequence and in these mRNAs exon 2 is extended at the 5′-end.

Therefore, in the human, as in the mouse, different start sites can be used to transcribe the *FMO1* gene (Figures 1–3). In fetal liver and adult kidney, transcription begins with exon 0, using the promoter P0. In the kidney, transcription can begin also from within intron 1 (Figure 3), from promoter P2. Very few human FMO1 cDNAs have been identified. The only liver-derived cDNA (M64082) is transcribed from P0. A cDNA (BC047129) derived from a pool of colon, kidney and stomach RNAs is transcribed from P2. A single FMO1 cDNA (AK097039) has been isolated from human small intestine [20]. The 5′-end of this cDNA is derived from an exon located between exons 0 and 2, which we call exon 1, and is thus transcribed from promoter P1. The position of exon 1 in human *FMO1* is equivalent to that of the corresponding exon in the mouse gene (Figure 2) [3]. The small intestine cDNA has sequence derived from exons 1–9, but also contains intronic sequences from between exons 2 and 3 and between exons 7 and 8. Thus it is not expected that this cDNA would encode a functional protein.

Analysis of transcriptional start sites, RT–PCR and sequence analysis of the amplified products, and a survey of the sequences of available cDNA clones, revealed that in both the human and mouse *FMO1* genes, transcription can begin from three different promoters: P0 (upstream of exon 0), P1 (upstream of exon 1) and P2 (upstream of exon 2).

Species-specific repetitive elements and the transcription of the *FMO1* gene from P0

RT–PCR analysis showed that, in human fetal liver, FMO1 mRNAs are derived from a splicing event between exon 0 and exon 2. Thus, in this tissue, *FMO1* is transcribed exclusively from the P0 promoter. The P0 promoter remains active in adult mouse liver (Figure 1), but is switched off in adult human liver. To identify sequences that may be responsible for the specific silencing of the *FMO1* gene in adult human liver, we investigated 5′-flanking sequences upstream of P0 by a combination of sequence analysis and reporter gene assays.

A comparison of sequences upstream of the P0 promoter of *FMO1* of human and mouse, and of two other species, rat and rabbit, in which the gene is expressed in adult liver, revealed considerable sequence identity extending for ∼450 bp upstream of the P0 transcriptional start site (Figure 4). In this region, the human sequence is 66, 69 and 70 % identical respectively with

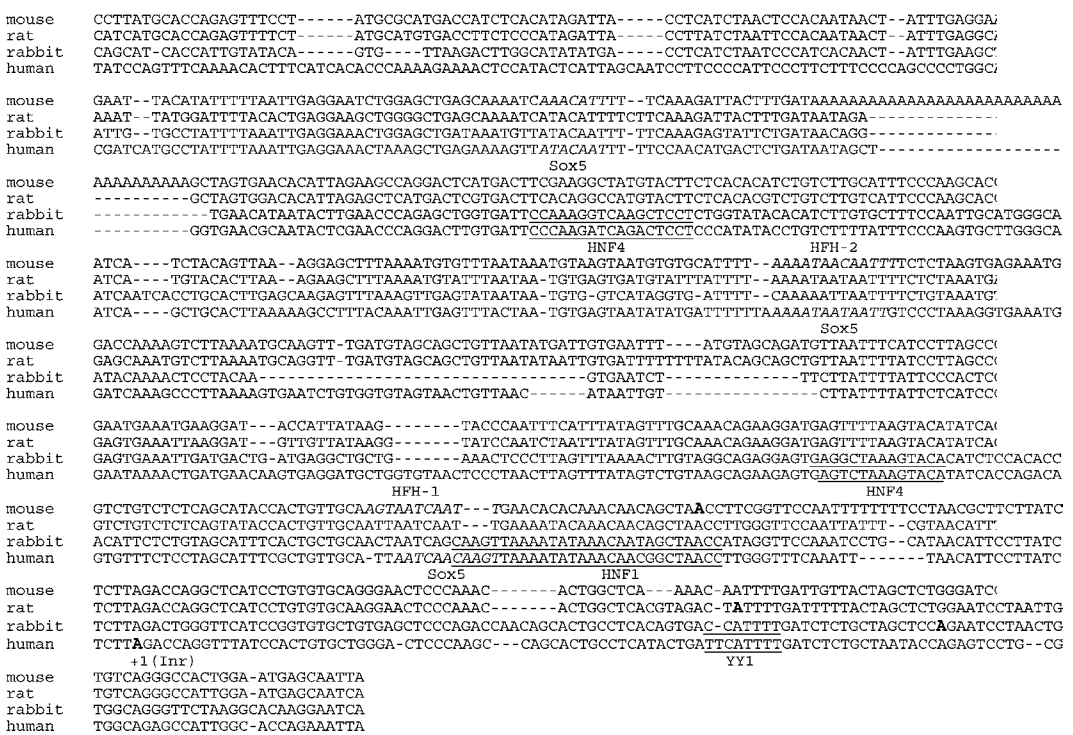

Figure 4 The proximal region of the *FMO1* P0 promoter of human, mouse, rat and rabbit

Boldface letters indicate the most 5′ transcriptional start site used in each species. The '+1' refers to the start site of human *FMO1*. Human (the present study) and rat (M. Stevanovic-Walker and I. R. Philips, unpublished work) start sites were determined by TAP-RLPCR and the mouse start site by primer extension (the present study). The preferred rabbit start is from [22] and the positions of YY1 (Yin and Yang 1), HNF1 and HNF4 sites (underlined) are from [28]. Potential Sox-5 and HFH sites (italics) were identified by ConSite analysis. 'inr' indicates a consensus initiator sequence.

the sequences of mouse, rat and rabbit, and the mouse sequence is 88 and 60 % identical respectively with the sequences of rat and rabbit. Analysis of the proximal region of the P0 promoter, with the phylogenetic footprinting program ConSite [27], identified five motifs that were conserved in all four species. Of these, three were predicted as binding sites for Sox-5, a protein known to be involved in developmental processes [30], and two as binding sites for the liver-specific transcription factors HFH1 [HNF3 (hepatocyte nuclear factor 3) homologue 1] and HFH2 [31] (Figure 4). The program Footprinter 2.1 [26] identified 13 evolutionarily conserved footprints in the four species. Analyses of these conserved motifs, both manually and by the transcription factor identification programs MacVector 6.5.3 and AliBaba 2.1 [32], failed to identify any nucleotide differences that would create or destroy a predicted protein-binding site (results not shown).

Thus the proximal region of the P0 promoter of *FMO1* of human, mouse, rat and rabbit is well conserved. HepG2 cells (human hepatocellular carcinoma G2 cells) were transfected with a reporter gene under the control of the proximal region of the P0 promoter of *FMO1* of human, pGL[−544 to +27(H)], or mouse, pGL[−431 to +50(M)]. No significant differences were observed in reporter gene activity directed by the proximal P0 of the two species (Figure 5). Therefore the transcriptional activity of the proximal region of the P0 promoter of the *FMO1* gene of human is very similar to that of the mouse.

Although the proximal region of the P0 promoter of human and mouse *FMO1* share considerable sequence identity, dot matrix analyses showed that upstream of approx. −470 there is little sequence similarity between the human and mouse genes (Figure 6). This breakdown in sequence similarity is due to the presence of species-specific repetitive DNA elements that have inserted upstream of the proximal region of P0 of both human and mouse. Region −541 to −1835 of the human *FMO1* gene contains three L1 [LINE (long-interspersed nuclear element)-1]-like elements, which lie 3′–5′ with respect to the transcriptional start site of the P0 promoter (Figure 6). We have named these L1a, L1b and L1c. The longest of these, L1c (Figure 6), which lies between −1310 and −1835 of P0, was identified by a BLAST search and its characteristics were assigned by alignment to the query sequence human transposon L1.2 (accession no. M80343) [32]. L1c is truncated at its 5′-end and contains an incomplete ORF2 (open reading frame 2) with several stop codons, a 3′-UTR (3′-untranslated region) (220 bp), and an AT-rich region (38 bp). L1a (−707 to −541) and L1b (−1218 to −720) have been identified by the UCSC Genome Browser [33]. They have 47 % identity to each other and 41 % identity to L1c. Neither L1b nor L1a has an A-rich tail.

A further two L1 elements, L1d and L1e, were identified upstream of L1c (Figure 6). L1d (−4386 to −3189) is truncated at its 5′-end and comprises a 967-bp ORF2, a 203-bp 3′-UTR and a 26-bp A-rich sequence. It is flanked by an almost perfect (13/14) TSD (target-site domain). None of the other four L1 elements has clearly identifiable TSDs. L1d and L1c are inverted with respect to one another and are separated by 1874 bp of host *FMO1* gene sequence. L1e lies 688 bp upstream of L1d. The sequence separating L1e and L1d elements is host *FMO1* sequence. L1e (−11369 to −5080) represents a full-length L1 element. It has a 5′-leader, a 1024-bp ORF1, a 63-bp spacer, a 3767-bp ORF2, a 152-bp 3′-UTR and a 26-bp A-rich region. Both ORF1 and ORF2 have several frameshifts and stop codons. The features of L1d and L1e were identified by comparison with human transposon L1.2 (accession no. M80343). Of the five L1-like elements, L1d and L1c have the greatest sequence identity (85 %). These two elements have an identity of approx. 67 % to L1e.

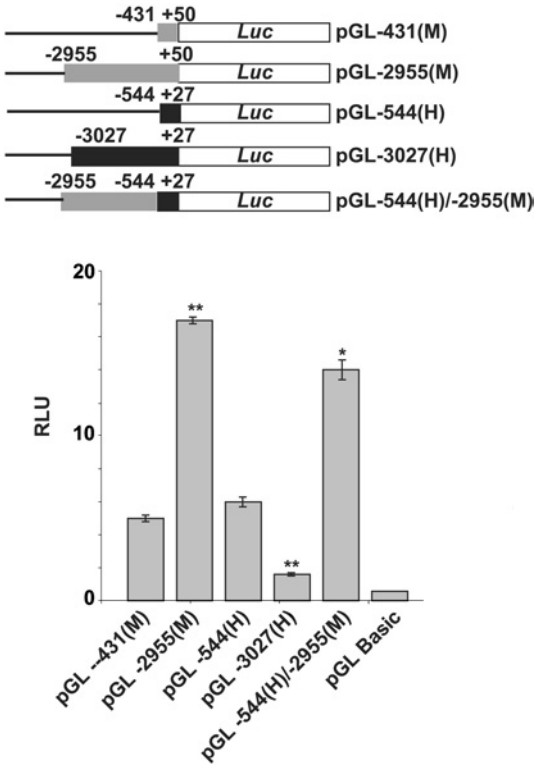

Figure 5 Transcriptional activity of human and mouse *FMO1* P0 reporter gene constructs

The upper panel shows the *FMO1* P0 reporter gene constructs used to transfect HepG2 cells. Promoter sequences were derived from mouse (M) or human (H). Open boxes represent the luciferase reporter gene, grey boxes mouse P0 sequences, and black boxes human P0 sequences. Horizontal lines represent vector sequence. Numbers indicate the position of nucleotides relative to the transcriptional start sites of P0, indicated in Figure 4. The lower panel shows luciferase reporter gene activities of the different constructs. Luciferase activity was measured as relative light units (RLU). Each experiment was carried out in triplicate. Error bars indicate standard deviation. The transcriptional activity of construct pGL-2955(M) is significantly different from that of pGL-431(M) (**$P < 0.01$). The activity of pGL-3027(H) is significantly different from that of pGL-544(H) (**$P < 0.01$) and the activity of construct pGL-544(H)/−2955(M) is significantly different from that of pGL-544(H) (*$P < 0.05$).

Analysis of the 5′-flanking sequence of the mouse *Fmo1* P0 promoter revealed a stretch of 35 adenine residues between nt −486 and −450 that is not present in the human flanking sequence (Figure 4). A BLAST query using the sequence upstream of this poly A stretch identified the *Fmo1* gene itself and a cDNA clone (AK046657) from a 4-day adipose tissue library, which is identical in sequence with a region extending from −2994 to −450 of the mouse *Fmo1* promoter. The poly A stretch originally noted in the mouse *Fmo1* gene (see Figures 4 and 6) corresponds to T residues at the 5′-end of the cDNA, indicating that this region can be transcribed, but in the opposite direction to that of the *Fmo1* gene. The RIKEN project states that the cDNA has an unclassifiable product. The annotations of the cDNA (http://fantom.gsc.riken.jp/db/link/cloneid.cgi?id=B430304E15) [34] show that it contains a SINE (short-interspersed nuclear element) B2 element. The SINE element lies between −2822 and −2616 of *Fmo1* and is flanked by a 6-bp TSD (GGAGAT). An LTR (long terminal repeat)-like sequence (RLTR13D), identified in the cDNA, was found to lie between −2312 and −1593 of *Fmo1*. The LTR is flanked by a 6-bp TSD (CTAAAG). The presence of the LTR-like sequence in mouse *Fmo1* was confirmed by PCR amplification of genomic DNA using primers that flanked the LTR. A product of the expected size (670 bp) was obtained and its identity was verified by DNA sequencing (results not shown). Between −822 and −743 of *Fmo1*, there is an almost perfect 80-bp direct repeat. Thus, within approx. 2.5 kb, the 5′-flanking sequence of mouse *Fmo1* has a number of different retrotransposable elements, which are flanked at the 5′-end by a 33-bp T-rich sequence and at the 3′-end by the 35-bp stretch of A residues. Figure 6(a) shows a dot matrix comparison of the *FMO1* P0 sequences of human and mouse. The regions of difference in the matrix correspond to the insertion of species-specific retrotransposons.

The effect of the species-specific retrotransposons on transcriptional activity of *FMO1* P0 promoters was investigated (Figure 5). Transfection of HepG2 cells with pGL[−3027/+27(H)], a construct that contains the L1a, b and c elements that lie upstream of the human *FMO1* P0 promoter, resulted in a 75% reduction in reporter gene expression compared with cells transfected with the proximal P0 construct pGL[−544/+27(H)], which contains no L1 elements. In contrast, reporter gene activity was 3-fold higher when cells were transfected with the mouse *Fmo1* construct pGL[−2955/+50(M)] compared with cells transfected with pGL[−431/+50(M)] (the mouse P0 proximal promoter construct). We then tested the transcriptional activity of a chimaeric construct, pGL[−2955(M)/−544(H)], in which the human P0 proximal promoter (−544 to +27) was placed under the control of the mouse upstream retrotransposon sequence (−2955 to −373). Reporter gene activity from this construct was 8-fold higher than that from a construct, pGL[−3027/+27(H)], in which the human P0 proximal promoter was controlled by the human L1-enriched sequence upstream of P0. The results of these reporter gene assays show that mouse upstream 5′-flanking sequences enhance transcription from P0, whereas human upstream 5′-flanking sequences act to repress transcription from P0.

To determine whether other species that express the *FMO1* gene in adult liver contain retrotransposon elements similar to those found in the mouse, we analysed the P0 5′-flanking sequences of the *FMO1* genes of rat and rabbit. Dot matrix (Figure 6b) and VISTA alignments (results not shown) of the rabbit and mouse *FMO1* flanking sequences show that there is little identity between the two species in the 2.5-kb region containing the mouse retrotransposon elements. In contrast, there are some similarities between rat and mouse in this region. For instance, the rat gene contains the SINE element found in the mouse *Fmo1* gene and has a TG-rich sequence in the equivalent position to that of the mouse T-rich sequence (Figure 6e). However, the rat gene lacks the 35-bp poly A tract present in mouse *Fmo1* (Figure 4) and it possesses only one copy of the 80-bp direct repeat sequence of the mouse. The rat *FMO1* flanking sequence does not contain the LTR sequence found in the mouse (Figure 6e). The absence of the LTR was confirmed by PCR amplification of rat genomic DNA and DNA sequencing (results not shown). The 6-bp TSD that flanks the mouse LTR is present in the corresponding position in the rat gene as a single copy. This suggests that the rat has not lost the LTR, but instead the mouse has gained an additional retrotransposon sequence in its *Fmo1* gene after the speciation of rodents. The poly A stretch, which is seen in the mouse, but not rat (Figure 4), may have arisen by a transposition event that split the 5′-end of a transposable element from its 3′-end (3′-transduction) [35].

Dot matrix comparison of the region upstream of the *FMO1* P0 promoter of mouse, rat and rabbit with that of human showed that, in each case, the break in the matrix was due to the insertion of the L1 elements in the human gene (Figures 6a, 6c and 6d).

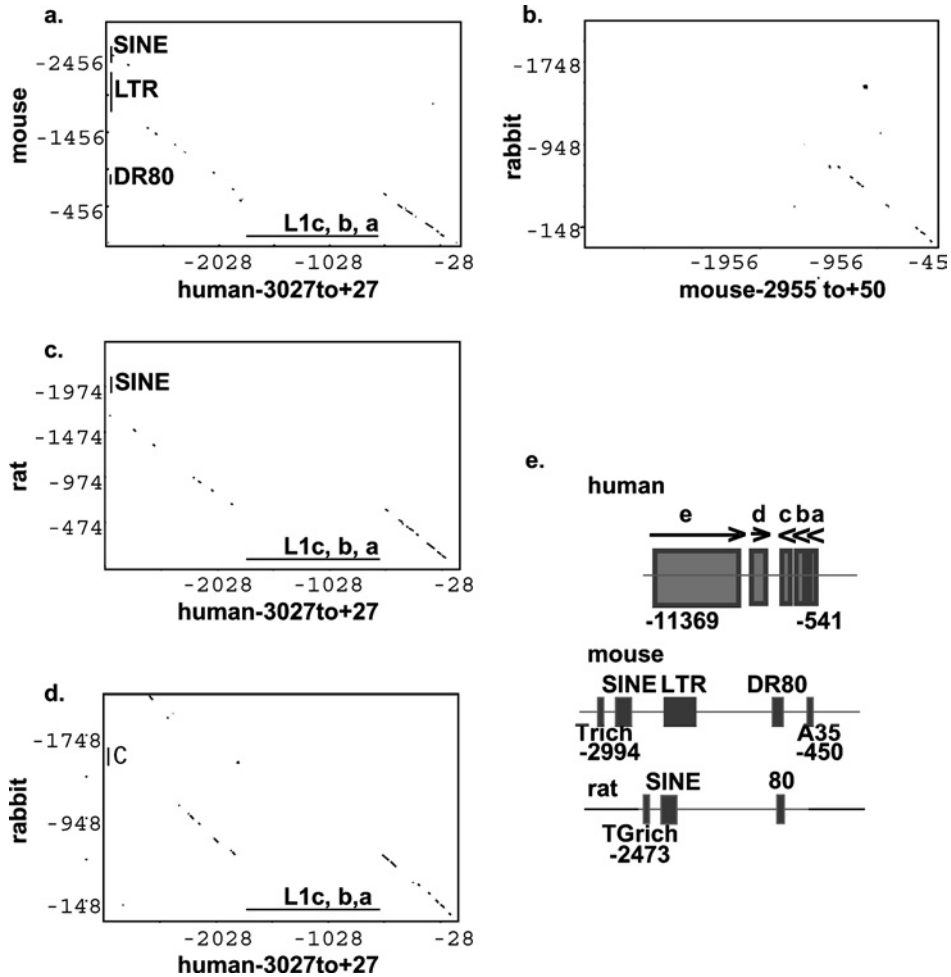

Figure 6 Transposable elements upstream of P0 in the *FMO1* gene of human, mouse, rat and rabbit

(**a**–**d**) The sequences upstream of P0 were compared using dot matrix analyses, with a setting of 55 % sequence similarity. Numbers indicate position relative to the transcriptional start site of P0. The positions of species-specific repetitive elements are shown: human LINE elements L1a, L1b and L1c (**a**, **c**, **d**); mouse DR(80), LTR and SINE (**a**); rat SINE (**c**); rabbit C element (**d**). (**e**) The retrotransposon elements located upstream of P0 in the *FMO1* gene of human and mouse. The equivalent region of the rat *FMO1* gene is shown for comparison. Arrows and arrowheads indicate the direction of LINE elements (5′ to 3′).

The alternative *FMO1* promoters P1 and P2

Although transcription from P0 is silenced in adult human liver, the *FMO1* gene continues to be expressed in adult kidney [4]. A reason for this is the use in kidney, but not in liver, of an alternative promoter, P2, located downstream of P0 (Figure 3). We used a bioinformatic approach to explore the P1 and P2 regions of the human and mouse *FMO1* genes. The P2 transcriptional start site of human *FMO1* lies within a consensus INR (initiator) site [36], TC**A**CAT (base + 1 is indicated in boldface), located 151 bp upstream of the ATG translation initiation codon. An identical INR sequence is present in the mouse gene 153 bp upstream of the ATG codon. In both species, the sequence TTAAC is located approx. 30 bp upstream of the transcriptional start site and may represent a binding site for the TATA-binding protein. The sequence GGGCGG, a potential SP-1-binding site, is present between − 49 and − 54 of the human gene. A similar sequence, GGGTGG, is present between − 57 and − 52 of the mouse gene. No downstream promoter element consensus sequence is evident in either species. In the mouse, transcription of the *Fmo1* gene in the kidney can start also at exon 1, and in this case P1 is used (see above). The location of P1 can be inferred from analysis of cDNA clones from mouse kidney and human small intestine.

Phylogenetic footprinting was used to map evolutionarily conserved transcription factor-binding motifs within a 3.0-kb region upstream of P1 and P2 of human and mouse *FMO1* genes. Eight conserved footprints, identified by ConSite [27], lie upstream of exon 2 and are indicated in Figure 7. Two footprints are located in the region of P2 in the short intronic sequence between exons 1 and 2 (intron lengths are 238 bp in mouse and 231 bp in human). Five conserved footprints are positioned within a ∼500-bp region located just upstream of exon 1, in the region of P1. The eighth conserved phylogenetic footprint is located further upstream, approx. 1500 and 1300 bp from the ATG, in mouse and human respectively. No conserved potential transcription factor footprints were identified within exon 1 of either the human or mouse gene.

DISCUSSION

We have identified three promoters, P0, P1 and P2, that are used in the transcription of the *FMO1* gene in both human and mouse. Different promoters can be used in different tissues: in adult mouse and fetal human liver transcription is exclusively from the P0 promoter, P1 and P2 are used in mouse kidney and P2

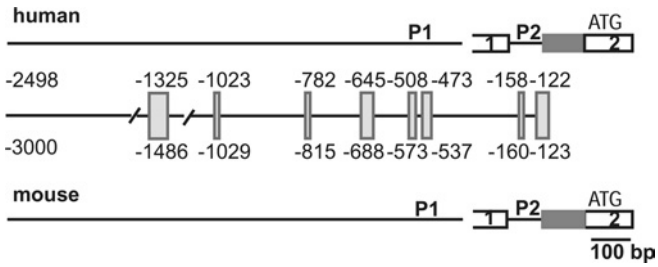

Figure 7 Phylogenetic footprint analyses of the intronic sequences encompassing P1 and P2 promoters of the *FMO1* gene of human and mouse

Vertical grey boxes indicate conserved footprints upstream of P1 and P2 in human and mouse *FMO1* genes. The numbers refer to the positions of the footprints relative to the A of the ATG translation initiation codon in human (above) and mouse (below). Exons 1 and 2 are shown. The horizontal shaded area shows the extension of exon 2, at its 5′-end, when P2 is used.

is used in human kidney. In addition to the tissue-specific use of *FMO1* promoters, there is a species-specific developmental silencing of the P0 promoter in adult human liver. The continued expression of *FMO1* in adult human kidney can be explained by the use of an alternative downstream promoter, P2. Whichever promoter is used, FMO1 protein-coding sequences are derived from the splicing of eight constitutive exons, exons 2–9. The use of alternative promoters results in the inclusion in the mRNA of additional leader sequences derived from the mutually exclusive, regulated, cassette exons 0 and 1 or from intron 1. FMO1 mRNA transcribed from P0 contains exon 0. mRNA from P1 has a leader sequence derived from exon 1 and the 5′-end of exon 2. When P2 is used, the leader contains sequences derived from intron 1 and the 5′-end of exon 2. None of the different leader sequences contains protein-coding information. Thus mRNAs derived from the use of different *FMO1* promoters encode identical proteins. Humans display up to 10-fold inter-individual variation in amounts of FMO1 [9,37]. FMO1 is not known to be subject to induction by endogenous or exogenous compounds. Thus the observed differences in FMO1 amounts are more likely to be a consequence of genetic variations that influence the strength of one or more of the promoters from which the gene is transcribed, rather than of environmental factors.

Reporter gene assays revealed that the proximal regions of the P0 promoter of human and mouse *FMO1* have very similar strengths. However, inclusion in reporter gene constructs of additional human *FMO1* upstream sequences was found to markedly down-regulate the activity of the P0 promoter. In contrast, inclusion of the corresponding upstream region of the mouse *Fmo1* gene increased the activity of the human P0 core promoter.

Comparisons of sequences upstream of the proximal P0 promoter of human and mouse *FMO1* showed that both have undergone species-specific insertion of DNA elements into this region during evolution. In the mouse, a 2.5-kb region upstream of the core P0 promoter contains a collage of different transposable elements, none of which is present in the human P0 promoter. However, most or all of these elements are also absent from the corresponding regions of the *FMO1* gene of organisms such as rat and rabbit, in which the *FMO1* gene is expressed in adult liver. Thus, although such elements may have a positive effect on gene transcription, their absence from the human P0 promoter cannot account for the negative effect of upstream sequences on the activity of the human P0 promoter. Instead, the repressive effect of upstream sequences on the P0 promoter is most likely due to the unique presence in the human *FMO1* gene of L1 elements.

In the human, the bulk of the 11-kb region upstream of the P0 promoter is composed of L1 elements (see Figure 6). Inclusion in a reporter construct of the three most proximal of these, L1a, b and c, down-regulates expression from P0. One explanation for the effect of the L1 elements on the activity of the human P0 promoter is that their insertion during evolution may have led to the separation from the core promoter of a regulatory element that is essential for transcription from P0 in adult human liver, but not in fetal liver. Alternatively, the L1 elements may have a more direct effect on activity of the P0 promoter. For instance, L1 elements can be heavily methylated, a mechanism that is thought to protect our genome from spurious transcription of these sequences [38]. As methylation spreads, it leads to heterochromatin formation and transcriptional repression. Thus, although methylation is unlikely to explain the action of these elements in a reporter gene assay, silencing of the *FMO1* gene in adult human liver may be the result of methylation of the battery of L1 elements that lie upstream of the liver-specific promoter P0. The P1 and P2 promoters, which are active in non-hepatic tissues, may be sufficiently removed (approx. 9 kb) from the L1 elements to be unaffected by the methylation.

Although the P0 promoter of the *FMO1* gene is inactive in adult human liver, transcription of the gene in fetal liver occurs from this promoter (see Figure 3). This may be due to the presence in the P0 promoter, proximal to the L1 elements, of binding sites for the developmentally regulated transcription factor Sox-5 (see Figure 4). Large-scale analysis of the human transcriptome (HG-U95A) (http://www.ncbi.nlm.nih.gov/projects; gene expression omnibus GDS181) reveals that the mRNA Sox-5 is expressed in fetal, but not adult, human liver.

In addition, methylation of L1 elements is an epigenetic modification, which would occur during development. Thus, at the fetal stage, the L1 elements may be insufficiently methylated to inhibit transcription from the P0 promoter, but later during development, methylation may have progressed sufficiently to switch off transcription from the promoter. The continued expression of the *FMO1* gene in human kidney after birth can be explained by the use of the downstream promoter P2. The insertion of retrotransposons into promoter regions of genes may be one of the factors that drive tissue-specific switches in promoter use.

The region upstream of the P0 promoter of the *FMO1* gene appears to be a hotspot for retrotransposition, given the independent insertion of various transposable elements into the same region of DNA in three species, human, mouse and rat (Figure 6e). The collage of repetitive elements accumulated upstream of P0 by the mouse *Fmo1* gene increases transcription and the gene continues to be expressed in the liver after birth. In contrast, in the case of the human, the presence of inserted L1-like elements has a deleterious effect on the expression of the *FMO1* gene and appears to be a major factor contributing to the silencing of the gene in adult human liver.

This work was funded by a grant to E.A.S. and I.R.P. from the Wellcome Trust, U.K. (053590). P.C. and M.S.-W. were the recipients of Ph.D. studentships from the MRC, U.K., and Bart's and the Royal London School of Medicine and Dentistry respectively.

REFERENCES

1 Krueger, S. K. and Williams, D. E. (2005) Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. Pharmacol. Ther. **106**, 357–387
2 Phillips, I. R., Dolphin, C. T., Clair, P., Hadley, M. R., Hutt, A. J., McCombie, R. R., Smith, R. L. and Shephard, E. A. (1995) The molecular biology of the flavin-containing monooxygenases of man. Chem. Biol. Interact. **96**, 17–32
3 Hernandez, D., Janmohamed, A., Chandan, P., Phillips, I. R. and Shephard, E. A. (2004) Organization and evolution of the flavin-containing monooxygenase genes of human and mouse: identification of novel gene and pseudogene clusters. Pharmacogenetics **14**, 117–130

4 Dolphin, C., Shephard, E. A., Povey, S., Palmer, C. N., Ziegler, D. M., Ayesh, R., Smith, R. L. and Phillips, I. R. (1991) Cloning, primary sequence, and chromosomal mapping of a human flavin-containing monooxygenase (FMO1). J. Biol. Chem. **266**, 12379–12385

5 Dolphin, C. T., Beckett, D. J., Janmohamed, A., Cullingford, T. E., Smith, R. L., Shephard, E. A. and Phillips, I. R. (1998) The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein. J. Biol. Chem. **273**, 30599–30607

6 Cherrington, N. J., Cao, Y., Cherrington, J. W., Rose, R. L. and Hodgson, E. (1998) Physiological factors affecting protein expression of flavin-containing monooxygenases 1, 3 and 5. Xenobiotica **28**, 673–682

7 Karoly, E. D. and Rose, R. L. (2001) Sequencing, expression, and characterization of cDNA expressed flavin-containing monooxygenase 2 from mouse. J. Biochem. Mol. Toxicol. **15**, 300–308

8 Dolphin, C. T., Cullingford, T. E., Shephard, E. A., Smith, R. L. and Phillips, I. R. (1996) 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. **235**, 683–689

9 Koukouritaki, S. B., Simpson, P., Yeung, C. K., Rettie, A. E. and Hines, R. N. (2002) Human hepatic flavin-containing monooxygenases 1 (FMO1) and 3 (FMO3) developmental expression. Pediatr. Res. **51**, 236–243

10 Overby, L. H., Carver, G. C. and Philpot, R. M. (1997) Quantitation and kinetic properties of hepatic microsomal and recombinant flavin-containing monooxygenases 3 and 5 from humans. Chem. Biol. Interact. **106**, 29–45

11 Janmohamed, A., Hernandez, D., Phillips, I. R. and Shephard, E. A. (2004) Cell-, tissue-, sex- and developmental stage-specific expression of mouse flavin-containing monooxygenases (Fmos). Biochem. Pharmacol. **68**, 73–83

12 Gasser, R., Tynes, R. E., Lawton, M. P., Korsmeyer, K. K., Ziegler, D. M. and Philpot, R. M. (1990) The flavin-containing monooxygenase expressed in pig liver: primary sequence, distribution, and evidence for a single gene. Biochemistry **29**, 119–124

13 Itoh, K., Kimura, T., Yokoi, T., Itoh, S. and Kamataki, T. (1993) Rat liver flavin-containing monooxygenase (FMO): cDNA cloning and expression in yeast. Biochim. Biophys. Acta **1173**, 165–171

14 Shehin-Johnson, S. E., Williams, D. E., Larsen Su, S., Stresser, D. M. and Hines, R. N. (1995) Tissue-specific expression of flavin-containing monooxygenase (FMO) forms 1 and 2 in the rabbit. J. Pharmacol. Exp. Ther. **272**, 1293–1299

15 Cherrington, N. J., Falls, J. G., Rose, R. L., Clements, K. M., Philpot, R. M., Levi, P. E. and Hodgson, E. (1998) Molecular cloning, sequence, and expression of mouse flavin-containing monooxygenases 1 and 5 (FMO1 and FMO5). J. Biochem. Mol. Toxicol. **12**, 205–212

16 Lattard, V., Longin-Sauvageon, C., Lachuer, J., Delatour, P. and Benoit, E. (2002) Cloning, sequencing, and tissue-dependent expression of flavin-containing monooxygenase (FMO) 1 and FMO3 in the dog. Drug Metab. Dispos. **30**, 119–128

17 Fromont-Racine, M., Bertrand, E., Pictet, R. and Grange, T. (1993) A highly sensitive method for mapping the 5′ termini of mRNAs. Nucleic Acids Res. **21**, 1683–1684

18 Grange, T., Bertrand, E., Espinas, M. L., Fromont-Racine, M., Rigaud, G., Roux, J. and Pictet, R. (1997) *In vivo* footprinting of the interaction of proteins with DNA and RNA. Methods **11**, 151–163

19 Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. and Mattick, J. S. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. **19**, 4008

20 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K. et al. (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat. Genet. **36**, 40–45

21 Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F. et al. (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc. Natl. Acad. Sci. U.S.A. **99**, 16899–16903

22 Luo, Z. and Hines, R. N. (1996) Identification of multiple rabbit flavin-containing monooxygenase form 1 (FMO1) gene promoters and observation of tissue-specific DNase I hypersensitive sites. Arch. Biochem. Biophys. **336**, 251–260

23 Bray, N., Dubchak, I. and Pachter, L. (2003) AVID: a global alignment program. Genome Res. **13**, 97–102

24 Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res. **32**, W273–W279

25 Grabe, N. (2002) AliBaba2: context specific identification of transcription factor binding sites. *In Silico* Biol. **2**, S1–S15

26 Blanchette, M. and Tompa, M. (2003) FootPrinter: a program designed for phylogenetic footprinting. Nucleic Acids Res. **31**, 3840–3842

27 Sandelin, A., Wasserman, W. W. and Lenhard, B. (2004) ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res. **32**, W249–W252

28 Luo, Z. and Hines, R. N. (2001) Regulation of flavin-containing monooxygenase 1 expression by ying yang 1 and hepatic nuclear factors 1 and 4. Mol. Pharmacol. **60**, 1421–1430

29 Hu, M. C. and Davidson, N. (1993) Mapping transcription start points with T4 DNA polymerase. Methods Enzymol. **217**, 446–458

30 Wunderle, V. M., Critcher, R., Ashworth, A. and Goodfellow, P. N. (1996) Cloning and characterization of SOX5, a new member of the human SOX gene family. Genomics **36**, 354–358

31 Overdier, D. G., Porcella, A. and Costa, R. H. (1994) The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. Mol. Cell. Biol. **14**, 2755–2766

32 Santos, F. R., Pandya, A., Kayser, M., Mitchell, R. J., Liu, A., Singh, L., Destro-Bisol, G., Novelletto, A., Qamar, R., Mehdi, S. Q. et al. (2000) A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome. Hum. Mol. Genet. **9**, 421–430

33 Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J. et al. (2003) The UCSC Genome Browser Database. Nucleic Acids Res. **31**, 51–54

34 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H. et al. (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature **420**, 563–573

35 Szak, S. T., Pickeral, O. K., Makalowski, W., Boguski, M. S., Landsman, D. and Boeke, J. D. (2002) Molecular archeology of L1 insertions in the human genome. Genome Biol. **3**, research0052

36 Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S. T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol. Cell. Biol. **14**, 116–127

37 Yeung, C. K., Lang, D. H., Thummel, K. E. and Rettie, A. E. (2000) Immunoquantitation of FMO1 in human liver, kidney, and intestine. Drug Metab. Dispos. **28**, 1107–1111

38 Yu, F., Zingler, N., Schumann, G. and Stratling, W. H. (2001) Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. Nucleic Acids Res. **29**, 4493–4501

Received 17 April 2007/30 May 2007; accepted 4 June 2007
Published as BJ Immediate Publication 4 June 2007, doi:10.1042/BJ20070523