Nucleic Acids Research

Close linkage of the human cytochrome P450IIA and P450IIB gene subfamilies: implications for the assignment of substrate specificity

John S.Miles, Wendy Bickmore[1], J.David Brook[2], Aileen W.McLaren, Richard Meehan[1+] and C.Roland Wolf*

Imperial Cancer Research Fund, Laboratory of Molecular Pharmacology, University Department of Biochemistry, Hugh Robson Building, George Square, Edinburgh EH8 9XD, [1]Medical Research Council, Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU and [2]Institute of Medical Genetics, University of Wales College of Medecine, Heath Park, Cardiff CF4 4XN, UK

ABSTRACT

We have isolated from human liver libraries two cytochrome P450 cDNA clones (λMP14 and λMP3) which are highly similar (83% over the coding region) to mouse testosterone 15α hydroxylase and are therefore part of the cytochrome P450IIA gene subfamily. The P450IIA (*CYP2A*) gene subfamily was found to be closely linked to the P450IIB (*CYP2B*) subfamily and their chromosomal location could not be distinguished using somatic cell hybrids containing fragments of chromosome 19 between 19q12 and 19q13.2. Pulsed field gel electrophoresis indicates that both gene subfamilies are contained within 350-kb genomic DNA fragments, but were separated using various restriction enzymes. Northern blot analysis identified three P450IIA mRNAs each showing a wide inter-individual variation in their levels in the liver. High levels of P450IIA transcript were associated with high levels of P450IIB transcript suggesting that common factors may influence the expression of genes within these subfamilies. Genetic analysis has suggested previously that a member of the P450IIB subfamily is responsible for coumarin hydroxylase activity in the mouse. We discuss the possibility, based on our findings of tight linkage of the human P450IIA and IIB subfamilies, that a member of the IIA subfamily is a better candidate for this enzyme activity.

INTRODUCTION

Cytochrome P450-dependent monooxygenases (P450s) are a supergene family of enzymes that catalyze the oxidation of lipophilic chemicals through the insertion of one atom of molecular oxygen into the substrate. As a consequence they play a central role in the metabolism of a wide variety of endogenous molecules as well as foreign compounds (1,2). Mammalian P450s are encoded by several multigene families which have been classified on the basis of their sequence homology, gene organization and chromosomal localization (3,4,5). In general, genes within the same family exhibit at least 40% sequence identity, but less than 40% with members from other families, and in those cases where it has been determined, they have the same intron-exon organization (5). Members of the same gene subfamily share greater than 59% identity and are clustered in the genome, but different subfamilies may have different chromosomal locations. For example, members of the P450IIC subfamily are located on chromosome 10 in humans and chromosome 19 in mouse; the P450IID subfamily is on chromosome 22 in humans and chromosome 15 in mouse; and the human P450IIE subfamily is on chromosome 10 and the mouse P450IIE subfamily on chromosome 7 (4).

In contrast the human P450IIB gene subfamily (*CYP2B* locus) has been located close to the P450IIA (*CYP2A*) subfamily at 19q12 → 13.2 (6,7,8,9). The mouse P450IIB gene subfamily has been located on chromosome 7 and is associated with the *Coh* locus encoding a cytochrome P450 with coumarin hydroxylase activity (10). A preliminary report suggests

that the P450IIA gene subfamily is also on chromosome 7 in mouse (11), and this would be consistent with the finding that many genes on mouse chromosome 7 have homologues on chromosome 19 in man. In the mouse some of the genes within the P450IIA subfamily encode enzymes involved in testosterone 15α hydroxylase activity (12).

In this paper we report the sequence of a human P450IIA cDNA which is highly homologous to mouse testosterone 15α hydroxylase but distinct from that reported by Phillips *et al.* (13) and show that in man the P450IIA and P450IIB gene subfamilies are very tightly linked. We also show that there is considerable inter-individual variation in the expression of P450IIA genes, and that the expression of genes from both IIA and IIB subfamilies may be influenced by common factors.

MATERIALS AND METHODS
Preparation of radioactive probes
Restriction fragments for use as probes were radioactively-labelled with [α-^{32}P]dCTP (3000 Ci mmol^{-1}) by either nick-translation (14) or by random primer extension (15).
Isolation of human cytochrome P450IIA cDNA clones.
A cDNA clone (R17) encoding the C-terminal 211 amino acid residues of the rat P450IIB1 gene (16) was used to screen a human liver cDNA library cloned in λgt11 (Clontech, CA) by the method of Benton and Davis (17). Hybridization was in 6×SSC, 5×Denhardt's solution and 0.1% SDS. Filters were washed in 2×SSC (1×SSC=0.15M NaCl, 0.015M Na$_3$ citrate; 1×Denhardt's solution is 0.2% (w/v) each of Ficoll, polyvinylpyrrolidone, bovine serum albumin). Two strongly positive clones, λMP1 and λMP2, were isolated and subsequently shown to be from the human P450IIB subfamily (6). A third weakly positive clone, λMP3, was also identified. The 990-bp *Eco*RI fragment from λMP3 was subcloned into pUC19 to give pMP12. As the λMP3 insert did not encode a full length P450, a second human liver cDNA library (18) was screened with the 990-bp *Eco*RI fragment. 50000 plaques were screened and eighteen positive clones identified. The largest cDNA insert (λMP14; 1.76kb) was subcloned into pUC19 giving pMP81 for further analysis.
Nucleotide sequence analysis
The 0.99-kb *Eco*RI fragment of pMP12, and the 1.76-kb *Eco*RI fragment of pMP81 were subcloned into M13mp18 and template DNA prepared (19). The nucleotide sequence of each clone was determined using the dideoxy chain-termination method with [α-^{35}S]thio-dATP and a series of synthetic oligonucleotides (20,21). Sequences were compiled and analysed using Staden Plus software implemented on a DCS286 computer (22). The nucleotide sequences pMP12 and pMP81 have been deposited in the EMBL Data Library with accession numbers X06401 and X13897, respectively.
Cell lines
The somatic cell hybrids containing fragments of chromosome 19 used in this study have been documented elsewhere (9) except J640 (Brook *et al.*, in preparation).
Isolation of DNA and restriction endonuclease analysis
DNAs from somatic cell hybrids containing fragments of human chromosome 19 were isolated by standard methods (23). *Hind*III-digested DNA was fractionated electrophoretically on agarose gels, and was transferred to Hybond-N (Amersham) filters as described by Southern (24).
Pulsed-field gel electrophoresis (PFGE)
DNA was prepared for PFGE in agarose blocks. Cells, resuspended in PBS at 2.0×10^7

cells/ml, were mixed with an equal volume of 1% molten agarose (BRL Ultrapure), aliquoted into moulds and allowed to set. The resulting blocks were then incubated at 50°C in NDS (0.5M EDTA, 10mM Tris, 1% lauryl sarcosine, pH 9.5) and 0.5mg/ml proteinase K for 3 days with one change of NDS and proteinase K after 24h. Blocks were stored in NDS at 4°C.

The PFGE apparatus of Southern *et al.* (25) was used. 1% agarose gels were made and run in 0.5×TBE at 10°C and 5V/cm for 48h with a pulse time of 90s. Blocks were soaked in 5ml TE (10mM TrisHCl pH7.5, 1mM EDTA) plus 5μl (20mg/ml) PMSF at 50°C for 30min, washed twice in TE at 4°C and then equilibrated in restriction enzyme buffer for 1h before digestion with 20U of enzyme overnight. Size markers were chromosomes of *Saccharomyces cerevisiae* strain YPH148 (26).

Northern blot analysis

Total RNA was isolated from human liver samples obtained from kidney transplant donors by the guanidine HCl method (27). 15μg samples were run on denaturing agarose gels and then transferred to Hybond-N (Amersham International; 28).

Hybridization analysis

The human intron 5-containing P450IIB6 probe, pMP10 has been described previously (6,29). The rat P450IIB1 cDNA probe (R17; 15) was a kind gift of Dr. M. Adesnik; pMP17 is a cDNA clone containing exons 6,7 and 8 of a human P450 gene sharing approximately 96% identity with P450IIB6, but containing no *Alu* repeat sequences (unpublished observation), and the human P450IIA probe pMP12 is described in this paper. A mouse γ-actin processed pseudogene (30); a human P450IIE cDNA (phP450j; 31), a gift of Dr F.J. Gonzalez; and a human P450III (nifedipine oxidase) cDNA (32) were also used as probes.

DNA was labelled with ^{32}P and hybridized to Southern blots as described by Hill *et al.* (33). Filters were washed for 30min at 65° in 1×SSC in the case of the somatic cell hybrids, in 0.1×SSC in the case of the pulsed-field blots and in 2×SSC for the Northern blots. Bands were visualized by exposure to X-ray film for 2−14 days. Radioactive DNA was removed from the Southern filters by boiling in 0.1% NaPP$_i$, 0.2% SDS for 10 min and from Northern filters by washing in 5mM TrisHCl, pH8; 2mM Na$_2$EDTA; 0.1× Denhardt's for 1h at 65°. The filters were then exposed to X-ray film to confirm that the probe had been removed.

Reagents

Restriction enzymes, T4 DNA ligase and DNA polymerase I (Klenow fragment) were from Boehringer Ltd, Bethesda Research Laboratories and New England Biolabs. [α-^{32}P]dCTP and [α-^{35}S]thio-dATP were from Amersham International. Other reagents were from the usual suppliers. Oligonucleotides were made on an Applied Biosystems 380A DNA synthesizer.

RESULTS

Nucleotide sequence analysis

The nucleotide sequences of pMP12 and pMP81 are shown in Fig. 1. The two sequences were identical except for two nucleotide substitutions in the coding region, a 3-nucleotide insertion immediately preceding the poly(A) sequence, and that pMP81 extended a further 770-bp in the 5′ direction.

An open reading frame of 1467bp was detected in the pMP81 sequence extending from a Met codon at position 11 to a stop codon at position 1477. Comparison of the pMP81

```
           (G) M  L  L  V  A  L  L  V  C  L  T  V  M  V  L  M  S  V  W  Q  Q  R  K  N  K  G  K
GAATTCCGGGATGCTTCTGGTGGCCTTGCTGGTCTGCCTGACTGTAATGGTCTTGATGTCTGTTTGGCAGCAGAGGAAGAACAAGGGGAA      90

    L  P  P  G  P  T  P  L  P  F  I  G  N  Y  L  Q  L  N  T  E  Q  M  Y  N  S  L  M  K  I  S
GCTGCCTCCGGGACCCACCCCATTGCCCTTCATTGGAAACTACCTGCAGCTGAACACAGAGCAGATGTACAACTCCCTCATGAAGATCAG     180

    E  R  Y  G  P  V  F  T  I  H  L  G  P  R  R  V  V  V  L  C  G  H  D  A  V  R  E  A  L  V
TGAGCGCTATGGCCCCGTGTTCACCATTCACTTGGGGCCCCGGCGGGTCGTGGTGCTGTGTGGACATGATGCCGTCAGGGAGGCTCTGGT     270

    D  Q  A  E  E  F  S  G  R  G  E  Q  A  T  F  D  W  V  F  K  G  Y  G  V  V  F  S  N  G  E
GGACCAGGCTGAGGAGTTCAGCGGGCGAGGCGAGCAAGCCACCTTCGACTGGGTCTTCAAAGGCTATGGCGTGGTATTCAGCAACGGGGA     360

    R  A  K  Q  L  R  R  F  S  I  A  T  L  R  D  F  G  V  G  K  R  G  I  E  E  R  I  Q  E  E
GCGCGCCAAGCAGCTCCGGCGCTTCTCCATCGCCACCCTGCGGGACTTCGGGGTGGGCAAGCGAGGCATCGAGGAGCGCATCCAGGAGGA     450

                           R  L  R                                  L     I        F
    A  G  F  L  I  D  A  L  R  G  T  G  G  A  N  I  D  P  T  F  F  L  S  R  T  V  S  N  V  I
GGCGGGCTTCCTCATCGACGCCCTCCGGGGC ACTGGCGGCGCCAATATCGATCCCACCTTCTTCCTGAGCCGCACAGTCTCCAATGTCAT     540
                              CGACTGC                              C        C T        T

                   pHP450(1)                                   R  L     F  I     V     V P V H
    S  S  I  V  F  G  D  R  F  D  Y  K  D  K  E  F  L  S  L  L  R  M  M  L  G  I  F  Q  F  T
CAGCTCCATTGTCTTTGGGGACCGCTTTGACTATAAGGACAAAGAGTTCCTGTCACTGTTGCGCATGATGCTAGGAATCTTCCAGTTCAC     630
                               G     C     C           T        T           T   TG T CAG TCA

    V  N                                                  Q
    S  T  S  T  G  Q  L  Y  E  M  F  S  S  V  M  K  H  L  P  G  P  Q  Q  Q  A  F  Q  L  L  Q
GTCAACCTCCACGGGGCAGCTCTATGAGATGTTCTCTTCGGTGATGAAACACCTGCCAGGACCACAGCAACAGGCCTTTCAGTTGCTGCA     720
CGTC A                                                 A

                         N  T  P  -
    G  L  E  D  F  I  A  K  K  V  E  H  K  Q  R  T  L  D  P  N  S  P  R  D  F  I  D  S  F  L
AGGGCTGGAGGACTTCATAGCCAAGAAGGTGGAGCACAAACAGCGCACGCTGGATCCCAATTCCCCACGGGACTTCAT TGACTCCTTTCT     810
                         CAC  CA---                                        pMP12

                                           E     K
    I  R  M  Q  E  E  E  K  N  P  N  T  E  F  Y  L  K  N  L  V  M  T  T  L  N  L  F  I  G  G
CATCCGCATGCAGGAGGAGGAGAAGAACCCCAACACGGAGTTCTACTTGAAAAACCTGGTGATGACCACGTTGAACCTCTTCATTGGGGG     900
                               T                  G     A                    GT

                         I        G           K
    T  E  T  V  S  T  T  L  R  Y  G  F  L  L  L  M  K  H  P  E  V  E  A  Q  V  H  E  E  I  D
CACCGAGACCGTCAGCACCACCCTGCGCTATGGCTTCTTGCTGCTCATGAAGCACCCAGAGGTGGAGGCCCAGGTCCATGAGGAGATTGA     990
                         G  C                 GA  C        A

                                                                    M
    R  V  I  G  K  N  R  Q  P  K  F  E  D  R  A  K  M  P  Y  M  E  A  V  I  H  E  I  Q  R  F
CAGAGTGATCGGCAAGAACCGGCAGCCCAAGTTTGAGGACCGGGCCAAGATGCCCTACATGGAGGCAGTGATCCACGAGATCCAAAGATT    1080
                                                                    A              G

                   I  W  P  G                        L
    G  D  V  I  P  M  S  L  A  R  R  V  K  K  D  T  K  F  R  D  F  F  L  P  K  G  T  E  V  Y
TGGAGACGTGATCCCCATGAGTTTGGCCCGCAGAGTCAAAAAGGACACCAAGTTTCGGGATTTCTTCCTCCCTAAGGGCACCGAAGTGTA    1170
                   T  G  C  G                        A               T        G

                            I        L     K                           T  E  L  E     A
    P  M  L  G  S  V  L  R  D  P  S  F  F  S  N  P  Q  D  F  N  P  Q  H  F  L  N  E  K  G  Q
CCCTATGCTGGGCTCTGTGCTGAGAGACCCCAGTTTCTTCTCCAACCCCCAGGACTTCAATCCCCAGCACTTCCTGAATGAGAAGGGGCA    1260
                            A     T  A     G  A        T              T     ACTG ATTAG A    GC

    P                               Q  P
    F  K  K  S  D  A  F  V  P  F  S  I  G  K  R  N  C  F  G  E  G  L  A  R  M  E  L  F  L  F
GTTTAAGAAGAGTGATGCTTTTGTGCCCTTTTCCATCGGAAAGCGGAACTGTTTCGGAGAAGGCCTGGCCAGAATGGAGCTCTTTCTCTT    1350
ACC                         T  C     C  T

                                                                                P  R
    F  T  T  V  M  Q  N  F  R  L  K  S  S  Q  S  P  K  D  I  D  V  S  P  K  H  V  G  F  A  T
CTTCACCACCGTCATGCAGAACTTCCGCCTCAAGTCCTCCCAGTCACCTAAGGACATTGACGTGTCCCCCAAACACGTGGGCTTTGCCAC    1440
                                                                                C  ACG

    F
    I  P  R  N  Y  T  M  S  F  L  P  R  *
GATCCCACGAAACTACACCATGAGCTTCCTGCCCCGCTGAGCGAGCGGCGGTGTGCCGGTGCAGGTCTGGTGGGCGGGGCCAGGGAAAGGGCA    1530
AT                    T        G  C     TG                 AG  A  G C GG

GGGCCAAGACCGGGCTTGGGAGAGGGGCGCAGCTAAGACTGGGGGCAGGATGGCGGAAAGGAAGGGGCGTGGTGGCTAGAGGGAAGAGAA    1620
 TAGGGC  GGGTT  GCG  AAGAC  C   GTATAAG  T  GGG  AA  ATGC  GGAAA      G  GC T  T GCTA AGGGA  A     A C

GAAACAGAAGCGGCTCAGTTCACCTTGATAAGGTGCTTCC GAGCTGGGATGAGAGGAAGGAAACCCTTACATTATGCTATGAAGAGTAGT    1710
AG    GG  CTAA  TTA  TTA  A  GG  GCT  CGA  CTGGATC  pHP450(1)

AATAATAGCAGCTCTTATTTCCTG AGC AAAAAAAAAAAAAGGAATTC                                          1756
```

sequence to that of pHP450(1), the sequence of a partial human P450IIA3 cDNA (Fig. 1; 13,4), shows that they are 91% identical over the comparable coding region (position 482−1660; Fig. 2) when a 3-nucleotide gap at position 767 is made but only 45% identical in the 3' non-coding region. However, a greater degree of sequence alignment was possible if various insertions/deletions (Fig. 1). These data suggest that pMP81 and pHP450(1) are derived from distinct human P450IIA genes.

Figure 2 shows the alignment of the deduced amino acid sequence from pMP81 with other P450 sequences. The greatest similarity is seen with mouse male-specific testosterone 15α hydroxylase (P450IIA3; 12) which shares 83% identity at both the amino acid and nucleotide sequence levels. Rat testosterone 7α hydroxylase (P450IIA1; 34) and rat testosterone 15α hydroxylase (P450IIA2; 35) share 67% and 65% amino acid sequence identity, respectively, with the pMP81 sequence, but only after the insertion of single gaps at positions 27 and 287 in the rat sequences (Fig. 2). Human P450IIB6 (6,29) exhibits 52% amino acid sequence identity and 62% nucleotide sequence identity with the pMP81 sequence.

Although the pMP81 sequence has a potential ATG initiation codon (Fig. 2), comparison with mouse testosterone 15α hydroxylase (Fig. 2) suggests that there may be an additional fifteen 5' coding nucleotides (i.e. 5 amino acids). This possibility will only be clarified by amino-terminal sequence analysis of the corresponding protein.

Linkage between the human CYP2A and CYP2B loci

(a)*Somatic cell hybrid analysis.* The *CYP2A* and *CYP2B* loci map to the same region of chromosome 19, namely 19q12 → 13.2 (8,9,6,7). A series of somatic cell hybrids containing fragments of chromosome 19 between q12 and q13.2 were used to establish how tightly linked these loci are. Table I shows the segregation of *CYP2B* compared with other loci which map to 19q12→19q13.2. It can be seen that *CYP2B* has the same segregation pattern as *CYP2A* and MSK37 (9).

(b)*Pulsed-field gel analysis.* The close linkage of *CYP2A* and *CYP2B* suggested by the somatic cell hybrid analysis was confirmed and further defined by pulsed-field gel electrophoresis followed by Southern blot analysis (Fig. 3). Fig. 3a shows an autoradiograph of the filter probed with pMP12 (*CYP2A*) and washed in 0.1×SSC at 65°C. Fig. 3b shows the same filter stripped of radioactivity, probed with pMP17 (*CYP2B*; a cDNA clone containing exons 6,7 and 8 of a human P450 gene sharing approximately 96% identity with P450IIB6, but containing no *Alu* repeat sequences, unpublished observations), and washed in 0.1×SSC at 65°C. Under these conditions we have shown previously that the P450IIA and P450IIB genes do not cross-hybridize (6). It can be seen that both the P450IIA and P450IIB probes hybridize to single fragments generated by *Mlu*I, *Not*I and *Sst*II (350kb, 520kb and 350kb respectively), indicating that the human P450IIA and P450IIB gene subfamilies are very closely linked and present within a maximum of 350kb of DNA. However, the two gene subfamilies can be distinguished from each other with the enzymes

Fig.1. Nucleotide sequence of pMP12 and pMP81 and comparison with pHP450(1). The nucleotide sequence of pMP81 and its translation are shown. The nucleotide sequence of pMP12 starts at position 799 and is identical to pMP81 except for two single base substitutions and a 3-bp deletion (indicated by boxes). The sequence of pHP450(1) (13) is shown underneath that of pMP81 starting at position 482 and extending to position 1660; only nucleotide differences are indicated. A 3-bp deletion in pHP450(1) is required for optimal alignment (position 767). Amino acid changes are shown above those for pMP81.

```
A  GMLLVALLVC LTVMVLMSVW QQRKNKGKLP PGPTPLPFIG NYLQLNTEQM YNSLMKISER  64
B  L    AVAF S L        K  LS              V  F                    Q
C  L   VI AS S L V L      - IR R                    KDV  S ITQL
D  ************************PCTHDR       R   LL   L  MDRRGL LK FLRFR K

A  YGPVFTIHLG PRRVVVLCGH DAVREALVDQ AEEFSGRGEQ ATFDWVFKGY GVVFSNGERA 124
B        Y   S I     Q E K                            L         IA  S
C                    Y Y   K                       YNTL         A   S
D   D  V      P M   V E I      K   A      KI MV PF R       I A  N W

A  KQLRRFSIAT LRDFGVGKRG IEERIQEEAG FLIDALRGTG GANIDPTFFL SRTVSNVISS 184
B    S                                   SF K N  F    Y
C     L              V    L       Y KM Q  C     P     IY  K
D   V    VT M    M  S V       Q C  EE  KSK  LM        LF QSITA I C

A  IVFGDRFDYK DKEFLSLLRM MLGIFQFTST STGQLYEMFS SVMKHLPGPQ QQAFQLLQGL 244
B        E                  SL   A  M  V                        KE
C    E   E  T       Q    GQMNR AAS P     D H      Y        IIKVT K
D    K  H Q  Q    KM NL FYQT SLI S VF    F L    GFL YF  AH R VYKN  EI

A  EDFIAKKVEH KQRTLDPNSP RDFIDSFLIR MQEEEKNPNT EFYLKNLVMT TLNLFIGGTE 304
B    T       N                    L  K        M    L  Q  FA
C    MIE  RQ NHS        N             - G S    HM         S  FA S
D  NAY GHS  K HRE    SA K L  TY LH  EK KS AHS   SHQ  NLN   S  FA

A  TVSTTLRYGF LLLMKHPEVE AQVHEEIDRV IGKNRQPKFE DRAKMPYMEA VIHEIQRFGD 364
B  G                 Y DI  K          R    Y    M    T             A
C    S              D   K      EQ     R   QY   HM    TQ    N       SN
D  T              ML Y H A ER YR  EQ   PH P ELH       T    Y        S

A  VIPMSLARRV KKDTKFRDFF LPKGTEVYPM LGSVLRDPSF FSNPQDFNPQ HFLNEKGQFK 424
B  L  G       T      L        F       K  K     K    K    DD
C  LA LGIP I I N T  G        A D F I  LMT  K   PS K  D  N  DD   L
D  LL  GVPHI TQH S  GYI I  D    FLI  STA H  HY  EK DA   D    DAN AL

A  KSDAFVPFSI GKRNCFGEGL ARMELFLFFT TVMQNFRLKS SQSPKDIDVS PKHVGFATIP 484
B            Y                  L  NI    HF   T A Q       RL  V
C  NA  L  T    F L D     K    LL   IL    F F  PMKLE  NE    PL  TR I
D  TE  I  L    I L   I   A          IL    SMA PVA E    LT QEC  VGK

A  RNYTMSFLPR 494
B  PT     S
C  PK    M I
D  PT QIR
```

Fig. 2. Amino acid sequence comparisons. The deduced amino acid sequences for pMP81 (A); mouse male-specific testosterone 15α hydroxylase (B); rat testosterone 7α hydroxylase (C); and human P450IIB6 (D) have been aligned. Dashes (−) indicate amino acid insertions required for optimal alignment, and asterisks (*) indicate putative amino acids missing from a partial length P450IIB6 cDNA (29). Only amino acid sequence differences are indicated for B, C and D. Numbers correspond to the mouse testosterone 15α hydroxylase sequence.

*Bss*HII, *Nae*I, *Nar*I, *Sfi*I and *Sma*I, where differently-sized fragments hybridize to P450IIA compared to P450IIB (Fig. 3).

Northern blot analysis

Three sizes of transcript (1.9kb, 2.7kb and 3.8kb) are identified when Northern blots of RNA from six individual human livers are probed with pMP12 (P450IIA; Fig. 4a). The major transcript (1.9-kb) is present in all individuals and corresponded well with the size expected from the cDNA clone pMP81 (1.76kb plus poly[A]). A marked variation (up to 10-fold) in the levels of this transcript between different individuals was observed. However, even more striking was the inter-individual variation (> >100-fold) in the amounts of the two larger transcripts. Whether these represent the products of different P450IIA genes, or arise from the use of alternative polyadenylation sites remains to be established.

The inter-individual variability in the levels of transcripts detected by three other human cytochrome P450 cDNAs, P450IIB, P450IIE, and P450III is also shown (Fig. 4). Livers 1 and 4 which had the highest levels of the 1.9-kb P450IIA mRNA also had the highest

Table I

Chromosome 19 Locus	TVBID	J640	Cell lines WILFM2	WILFM6	G175C5	G175C4
q12						
D19S7	+	+	−	+	+	+
GPI	+	+	−	+	+	N/T
D19S9	+	+	−	+	+	−
TGFB	+	+	+	−	−	−
MSK37	+	+	+	−	+	−
CYP2A	+	+	+	−	+	−
ATPA2	+	−	+	−	+	+
D19S8	+	−	+	−	+	+
APOC2	+	−	+	−	+	−
APOE	+	N/T	+	−	+	−
CKM	+	−	+	−	+	−
q13.2						
CYP2B	+	+	+	−	+	−

Segregation of *CYP2B* and other loci which map to 19q12→19q13.2, in hybrids containing fragments of chromosome 19. A + represents presence of marker; − represents absence of marker; and N/T is not tested. The *CYP2B* probe, pMP10, has been described previously (6), as have the somatic cell hybrids (9) except J640 (Brook *et al.*, in preparation).

levels of the 3.0-kb P450IIB mRNA suggesting that there may be common factors which cause an increase in the levels of these transcripts. This relationship is not seen with either the P450IIE or P450III transcripts, although liver 1 which contained a very high level of P450III mRNA also had the highest level of P450IIA-related message.
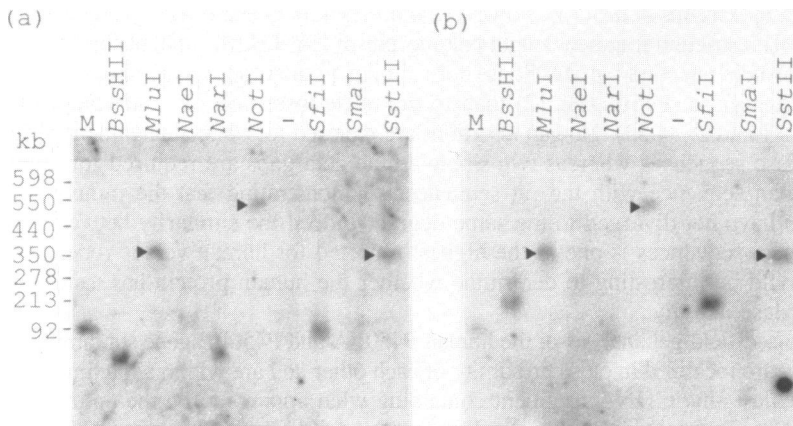
Fig. 3. Human DNA pulsed-field gel analysis. Autoradiographs of a Southern blot of human DNA digested with a variety of rare-cutting restriction endonucleases probed with (a) pMP12 (P450IIA) and (b) the same filter stripped of radioactivity and reprobed with pMP17 (P450IIB). The filter was washed finally in 0.1×SSC at 65° for 30min in each case and exposed to X-ray film for 5 days. DNA fragments hybridizing to both P450IIA and P450IIB probes are indicated by arrows. Sizes of the yeast chromosome markers are indicated in kb; the chromosome contains pBR322-derived sequence. The region of the gel between 50kb and 600kb is shown. No other hybridizing fragments were detected.
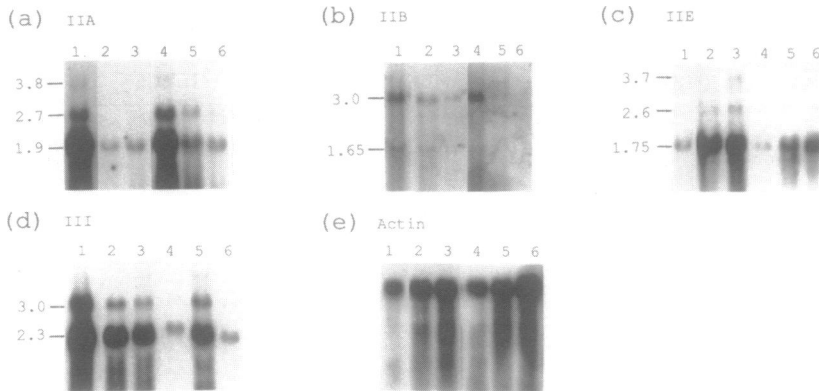
Fig. 4. Northern blot analysis of human liver RNA. Six individual liver RNA samples probed with (a) pMP12 [P450IIA]; (b) pMP17 [P450IIB]; (c) phP450j [P450IIE]; (d) pMP26 [P450III] and (e) mouse actin to confirm equal loading. The same filter is shown in (b) and (c) and was stripped of radioactivity between probes. The size of transcripts are indicated in kb. Livers 1 and 4 express the highest levels of both P450IIA and P450IIB but not of the other P450s, suggesting that common factors may cause an increase in the levels of the IIA and IIB transcripts.

DISCUSSION

On the basis of sequence similarity the cDNA clones described here appear to be members of the P450IIA gene subfamily. The nucleotide substitutions found between pMP81 and pMP12 (i.e. two nucleotide substitutions in the coding region and a 3-nucleotide insertion immediately preceding the poly[A] sequence) are consistent with the clones being derived from allelic variants. This gene is different from that giving rise to the cDNA, pHP450(1) reported by Phillips *et al.* (13), however both pHP450(1) and pMP12 can detect the same *Sst*I (*Sac*I) restriction fragment length polymorphism (36; J.S.M., unpublished observations).

The protein encoded by pMP81 is more similar to the mouse male-specific testosterone 15α-hydroxylase (P450IIA3; 12) than to the rat testosterone 7α- and 15α-hydroxylases (P450IIA1 and 2; 34,35). No gap insertions are required for optimal alignment of the human and mouse sequences whereas two single amino acid gaps are required for alignment of the human sequence with the rat sequences, demonstrating that the mouse and human proteins have not diverged to the same degree. Indeed the similarity between the mouse and human sequences is one of the highest reported for human versus rodent P450s and so it would be interesting to determine whether the human protein has testosterone 15α hydroxylase activity.

The pulsed-field gel analysis of the human P450IIA and P450IIB gene subfamilies indicates that they are localized in close proximity of each other and are within a maximum of 350kb. The smallest single DNA fragment containing what appears to be the entire P450IIA or P450IIB gene subfamilies are the *Nar*I fragments of 50kb and 90kb, respectively. Estimates based on normal gel electrophoresis and Southern blot analysis suggest that there are two or three genes in the P450IIB subfamily and a similar number in the P450IIA subfamily (6,13). The isolation and analysis of a human P450IIB6 gene indicates that it is > 14.5kb in length (29), but there is no information available on the size of the P450IIA genes. Assuming that the P450IIA and P450IIB genes are similar in size to other P450II genes (ca. 10−20kb) then the *Nar*I fragments are of sufficient size to contain all of the genes

within each of the subfamilies, implying that they are clustered together rather than interspersed with each other.

The enzymes MluI (ACGCGT), NotI (GCGGCCGC) and SstII (CCGCGG) each contain two CpGs in their recognition sequences and NotI and SstII are known to be associated strongly with CpG islands, which in turn are associated with the starts of genes (37). These enzymes generate single fragments of 350kb, 520kb and 350kb, respectively, which contain both subfamilies. In the case of MluI and SstII it is likely that they are cutting within the same CpG islands to generate the same-sized fragments.

It has been established that a group of linkage markers on human chromosome 19 including GPI and PEPD (38,9) have homologues on chromosome 7 in mouse (Gpi-1, Pep-4; 39,40). A locus encoding a cytochrome P450 with coumarin hydroxylase activity (Coh) has been mapped to this region of mouse chromosome 7 (39), and the P450IIB gene subfamily was localized within 4.4 centimorgans of Coh, using recombinant inbred lines (10). The close linkage homology between the corresponding regions of human chromosome 19 and mouse chromosome 7 suggests that the mouse P450IIA and P450IIB gene subfamilies are also very closely linked. A report has indicated that two mouse P450IIA genes (testosterone 15α hydroxylases) are present on chromosome 7 (11). We have preliminary data showing that the P450IIA and P450IIB gene subfamilies are indeed very closely linked in mouse (unpublished).

In rat the P450IIB genes (P450b,e locus) are located on chromosome 1 along with pink-eyed dilution, albino and haemoglobin β chain genes. Linkage homology between these genes and homologous genes on mouse chromosome 7 has been described previously (41); however the rat homologues of Gpi-1 and Pep-4 are found on rat chromosome 7. It will be interesting to know whether the rat P450IIA gene subfamily is linked to the P450IIB gene cluster on rat chromosome 1.

Analysis of recombinant inbred (RI) mouse lines has led to the suggestion that a P450 from the IIB subfamily is responsible for coumarin hydroxylase activity in mouse despite there being a discordant genotype between the DNA polymorphism detected by the rat P450IIB1 probe and coumarin hydroxylase activity in the CBA/CaJ mouse strain (10,3). If the P450IIA and P450IIB gene subfamilies are as tightly linked in mouse as they are in man then the P450IIA gene subfamily could be an equally good candidate for the Coh locus (encoding coumarin hydroxylase) in mouse (and in man). A P450 (M_r=50000) with coumarin hydroxylase activity has been purified from mouse (42). Polyclonal antiserum raised against this enzyme cross-reacts strongly with a protein (M_r=48000), and is capable of effectively inhibiting coumarin hydroxylase activity in human liver microsomes (43). It seems unlikely that the mouse protein is a member of the P450IIB subfamily as antibodies raised to the rat P450IIB1 protein (PB$_{3a}$; 44) recognise protein(s) in mouse liver microsomes with a Mr >54700 (45,44), quite distinct from the apparent Mr of 50000 for coumarin hydroxylase. Indeed, Raunio et al. (46) have recently shown that a monoclonal antibody directed against the rat P450IIB1 protein cross-reacts strongly to a mouse protein which is distinct from coumarin hydroxylase and fails to inhibit this activity. Similarly they show that antiserum against coumarin hydroxylase is not capable of detecting the mouse P450IIB-like protein nor inhibiting its associated activity (pentoxyresorufin O-dealkylase).

Based on this genetic and biochemical evidence, we propose that coumarin hydroxylase in both human and mouse is encoded by a member of the P450IIA gene subfamily. We are currently hoping to confirm this hypothesis by expressing our P450IIA cDNA in a

suitable cell system and testing whether it generates coumarin (and/or testosterone 15α-) hydroxylase activity.

ACKNOWLEDGMENTS

*To whom correspondence should be addressed

+Present address: Institute of Molecular Pathology, Dr Bohr Gasse 7, A1030 Wein, Austria

REFERENCES
1. Adesnik, M & Atchison, M. (1986) CRC Crit. Rev. Biochem. *19*, 247–305.
2. Wolf, C.R. (1986) Trends Genet, 2, 209–214.
3. Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B., Levin, W., Phillips, I.R., Sato, R. & Waterman, M.R. (1987) DNA, 6, 1–12.
4. Nebert, D.W., Nelson, D.R., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B., Levin, W., Phillips, I.R., Sato, R. & Waterman, M.R.(1989) DNA, 8, 1–14.
5. Nebert, D.W. & Gonzalez, F.J. (1987) Ann. Rev. Biochem. 56, 945–993.
6. Miles, J.S., Spurr, N.K., Gough, A.C., Jowett, T., McLaren, A.W., Brook, J.D. & Wolf, C.R. (1988) Nucl. Acids. Res. 16, 5783–5795.
7. Santisteban, I., Povey, S., Shephard, E.A. & Phillips, I.R. (1988) Ann. Hum. Genet. 52, 129–135.
8. Davis, M.B., West, L.F., Shephard, E.A. & Phillips, I.R. (1986) Ann. Hum. Genet. 50, 237–240.
9. Brook, J.D., Skinner, M., Roberts, S.H., Rettig, W.J., Almond, J.W. & Shaw, D.J. (1987) Genomics 1, 320–328.
10. Simmons, D.L. & Kasper, C.B. (1983) J. Biol. Chem. 258, 9585–9588.
11. Burkhart, B., Skow, L., Lalley, P. & Negishi, M. (1986) Fed. Proc. 454, 1853.
12. Squires, E.J. & Negishi, M. (1988) J. Biol. chem. 263, 4166–4171.
13. Phillips, I.R., Shephard, E.A., Ashworth, A. & Rabin, B.R. (1985) Proc. Natl. Acad. Sci. USA 82, 983–987.
14. Rigby, P.W.J., Dieckman, M., Rhodes, C. & Berg, P. (1977) J. Mol. Biol. 113, 237–251.
15. Feinberg, D.P. & Vogelstein, B. (1983) Anal. Biochem. 136, 6–13.
16. Kumar, A., Raphael, C. & Adesnik, M. (1983) J. Biol. Chem. 258, 1120–11284.
17. Benton, W.D. & Davies, R.W. (1977) Science, 196, 180–187.
18. Kwok, S.C.M., Ledley, F.D., DiLella, A.G., Robson, K.J.H. & Woo, S.L.C. (1985) Biochemistry 24, 556–561.
19. Norrander, J., Kempe, T. & Messing, J. (1983) Gene 26, 101–106.
20. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
21. Biggin, M.D., Gibson, T.J. & Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA. 80, 3963–3965.
22. Staden, R. (1986) Nucl. Acids. Res. 13, 217–231.
23. Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) in 'Molecular Cloning, a laboratory manual', Cold Spring Harbor Laboratory.
24. Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
25. Southern, E.M., Anand, R., Brown, W.R.A & Fletcher, D.S. (1987) Nucl. Acids Res. 15, 5925–5943.
26. Carle, K.F. & Olson, M.V. (1985) Proc. Natl. Acad. Sci. USA 82, 3756- 3760.
27. Cox, R.A. (1968) Meth. Enzymol. 12B, 120–129.
28. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201–5205.
29. Miles, J.S., McLaren, A.W. & Wolf, C.R. (1989) (submitted for publication).
30. Leader, D.P., Gall, I. & Lehrach, H. (1985) Gene 36, 369–374.
31. Song, B.-J., Gelboin, H.V., Park, S.-S., Yang, C.S. & Gonzalez, F.J. (1986) J. Biol. Chem. 261, 16689–16697.
32. Spurr, N.K., Gough, A.C., Stevenson, K., & Wolf, C.R. (1989), Hum. Genet. Vol 81 p.171–174.
33. Hill, R.E., Shaw, P.H., Barth, R.K. & Hastie, N.D. (1985) Mol. Cell. Biol. 5, 2114–2122.

34. Nagata, K., Matsunaga, T., Gillette, J., Gelboin, H.V. & Gonzalez, F.J. (1987) J. Biol. Chem. *262*, 2787–2793.
35. Matsunaga, T., Nagata, K., Holsztynska, E.J., Lapenson, D.P., Smith, A., Kato, R., Gelboin, H.V., Waxman, D.J.& Gonzalez, F.J. (1988) J. Biol. Chem. *263*, 17995–18002.
36. Wainwright, B.J., Watson, E.K., Shephard, E.A. & Phillips, I.R. (1985) Nucl. Acids Res. *13*, 4610.
37. Lindsay, S. & Bird, A.P. (1987) Nature *327*, 336–338.
38. Davis, M.B., Schonk, D., Monteiro, M., Oerlmans, F., Povey, S., & Wieringa, B. (1987) Ann. Hum. Genet. *51*, 195–199.
39. Wood, A.W. & Taylor, B.A. (1979) J. Biol. Chem. *254*, 5647–5651.
40. Skow, L.C. (1981) Biochem. Genet. *19*, 695–700.
41. Rampersaud, A. & Walz, F.G.,jr (1987) J. Biol. Chem. *262*, 5649–5653.
42. Juvonen, R., Shkumatov, V. & Lang, M. (1988) Eur. J. Biochem. *171*, 205- 211.
43. Raunio, H., Syngalma, T., Pasanen, M., Juvonen, R., Honkakoski, P., Kairaluoma, M., Sotaniemi, E., Lang, M A. & Pelkonen, O. (1988) Biochem. Pharmacol. *37*, 3889–3895.
44. Wolf, C.R., Seilman, S., Oesch, F., Mayer, R.T. & Burke, M.D. (1986) Biochem. J. *240*, 27–33.
45. Meehan, R.R., Forrester, L.M., Stevenson, K., Hastie, N.D., Buchmann, A., Kunz, H.W. & Wolf, C.R. (1986) Biochem. J. *254*, 780–797.
46. Raunio, H., Kojo, A., Juvonen, R., Honkakoski, P., Jarvinen, P., Lang, M.A., Vahankangas, K., Gelboin, H.V., Park, S.-S. & Pelkonen, O. (1988) Biochem. Pharmacol. 37, 4141–4147.