cDNA Cloning of the Mouse Bilirubin/Phenol Family of UDP-Glucuronosyltransferase (*mUGTbr2-like*)

Qian Chu,[1] Wei Lei,[1] Krishnamurthy Gudehithlu,[1] George A. Matwyshyn,[1] Maurizio Bocchetta,[2] and Ah-Ng Tony Kong[1,3]

Received November 4, 1996; accepted February 6, 1997

KEY WORDS: UDP-glucuronosyltransferases; UGT; UDPGT; UGT family 1 gene; glucuronidation; drug metabolism; bilirubin conjugation; phenol conjugation; Phase II metabolism.

INTRODUCTION

Glucuronidation, catalyzed by UDP-glucuronosyltransferase (UGT), is an important process in the metabolism of many drugs, carcinogens, xenobiotics, and endogenous substances such as bilirubin, steroids, and biogenic amines (1,2,3). UGT also plays an important role in the biosynthesis of proteoglycans, which confer important recognition properties on cell surfaces that influence cell adhesion and migration as well as the binding of growth factors and enzyme inhibitors (4). In humans, deficiency in glucuronidation have significant clinical and toxicological consequences. For example, inability to glucuronidate bilirubin may lead to its accumulation in plasma and tissues and result in jaundice. In more severe cases, patients may be at risk for bilirubin encephalopathy (kernicterus). These clinical sequalae have been observed in patients with Crigler-Najjer syndrome where there is an absence or partial deficiency in bilirubin UGT activity (5).

UGTs belong to a multigene family which comprises at least 2 families, *UGT1* and *UGT2* based on the similarities of their predicted amino acid sequences (3). The *UGT1* family of isozymes glucuronidate bilirubin and small planar phenols and the genomic structures of human and rat *UGT1* have been recently elucidated (6,7). It was found that there are structural similarities between the bilirubin UGTs and phenol UGTs between and within two species. Specifically, the 3′ end of both the bilirubin and phenol UGT are identical. To date, the structure of the mouse *UGT1* gene is not known. Since the mouse is one of the most widely utilized animal models for pharmacologic, as well as toxicological testing of pharmaceutical compounds, it is relevant to study the mouse genes. Previously, we have reported the cloning of two cDNAs encoding the mouse bilirubin (*mUGTbr1*) and phenol (*mUGTp4*) (8). Recently, additional *UGT1* cDNAs have been cloned, a mouse phenol UGT cDNA (*UGT1-06*) (9,10), and a mouse bilirubin UGT cDNA (*UGTbr2*) (11). To further isolate and characterize the mouse *UGT1* gene, we obtained *UGT1* genomic clone by screening a mouse P1 library, and a new cDNA sequence which belongs to the bilirubin subfamily of *UGT1* family, designated *mUGTbr2-like*, was found. It consists of at least 2468 basepairs (bp) and contains 1590-nucleotide open reading frame which encodes 529 amino acid residues. It possesses an identical *C*-terminus but variable *N*-terminus with *mUGTbr1*, *mUGTp4*, *mUGT1-06*, and *mUGTbr2* (7–10). From the amino acid sequence alignment of the *N*-terminus of *mUGTbr2-like*, it shares 75%, 70%, 68%, 67%, 62%, 61%, 59.2%, 58.8%, 58.5%, 57.4%, 45.4%, 43.3% and 43.6% sequence similarities with *rUGTB5* (7), *mUGTbr2* (11), *rUGTB2* (7), *rUGT1*0* (12), *hUGT1D* (br2) (13), *hUGT1C* (13), *rUGTB3* (7), *hUGT1E* (13), *lUGT1-4* (14), *rUGTB4* (7), *hUGT1A* (br1) (13), *rUGTB1* (br1) (7) and *mUGTbr1* (9-1) (8), respectively.

MATERIALS AND METHODS

Screening of a λgt11 cDNA Library and a λZAP cDNA Library

A mouse adult liver cDNA library in phage lambda gt11 (λgt11) [Clontech Laboratory, Palo Alto, CA] was screened with a 426-bp KpnI/XbaI fragment of the human 2351-bp full length *HUG-Br1* (*hUGT1A*) cDNA (6) as described previously in our laboratory (8) according to methods described in Ausubel *et al.* (15). Positive plaques were replated and rescreened twice until two isolated positive clones were obtained. The inserted cDNA in λ gt11 was subcloned into plasmid Bluescript SK+ vector (Stratagene, La Jolla, CA). The sequences were determined by the dideoxynucleotide chain termination method of Sanger (16). It was demonstrated that the insert cDNA sequences of the two positive clones were identical and contained 1202 bp without an open reading frame (ORF). Therefore, they were partial cDNAs. One of this partial cDNA (clone 7a) was then used to re-screen a mouse liver λZAP cDNA library (Stratagene, Lo Jolla, CA) with methods described above. After the third round screening, five independent clones were selected and *in vivo* excision of pBluescript SK vector containing the cDNA inserts using R408 helper phage according to the protocols provided by the manufacturer (Stratagene). The complete nucleotide sequence of the cDNA inserts were determined as above.

Screening of Mouse P1 Library and Sequencing of P1 Clone

One of the clones 6-3-2, contained 1974 base pairs (bp) and without ORF, was highly homologous to human *UGT1D* (br2) (13). Two primers, oligo RP-7 (5′-ACACTTCTCTCTGAAAAA-3′) and oligo UP-2 (5′-TCCACCACCTGAATTCCA-3′) which will amplify a 316 bp polymerase chain reaction (PCR) product, were used to PCR screen a mouse ES (c129/

[1] Department of Pharmaceutics and Pharmacodynamcis, Center for Pharmaceutical Biotechnology, College of Pharmacy, University of Illinois, Chicago, Illinois 60612-7173.

[2] Department of Medicinal Chemistry and Pharmacognosy, University of Illinois, Chicago, Illinois.

[3] To whom correspondence should be addressed. (email: kongt@uic.edu)

0724-8741/97/0500-0662$12.50/0

SvJ strain) P1 library (Genome Systems, St. Louis, MO). Two P1 clones (1518,1519) were obtained. Both clones showed positive PCR amplification and Southern blot hybridization. One of the clone, 1519 was further purified and both strands of the DNA were sequenced. The first sequencing primer, RP-9 (5′-AACGGTGGACATGTTTCT-3′) was designed from the 5′-end of clone 6-3-2. DNA sequencing of P1 clone was performed by Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL.

Isolation of RNA, Reverse Transcription and RT-PCR Analysis

In order to confirm that the above *mUGTBr2-like* gene product transcribes into mature mRNA, RT-PCR was performed

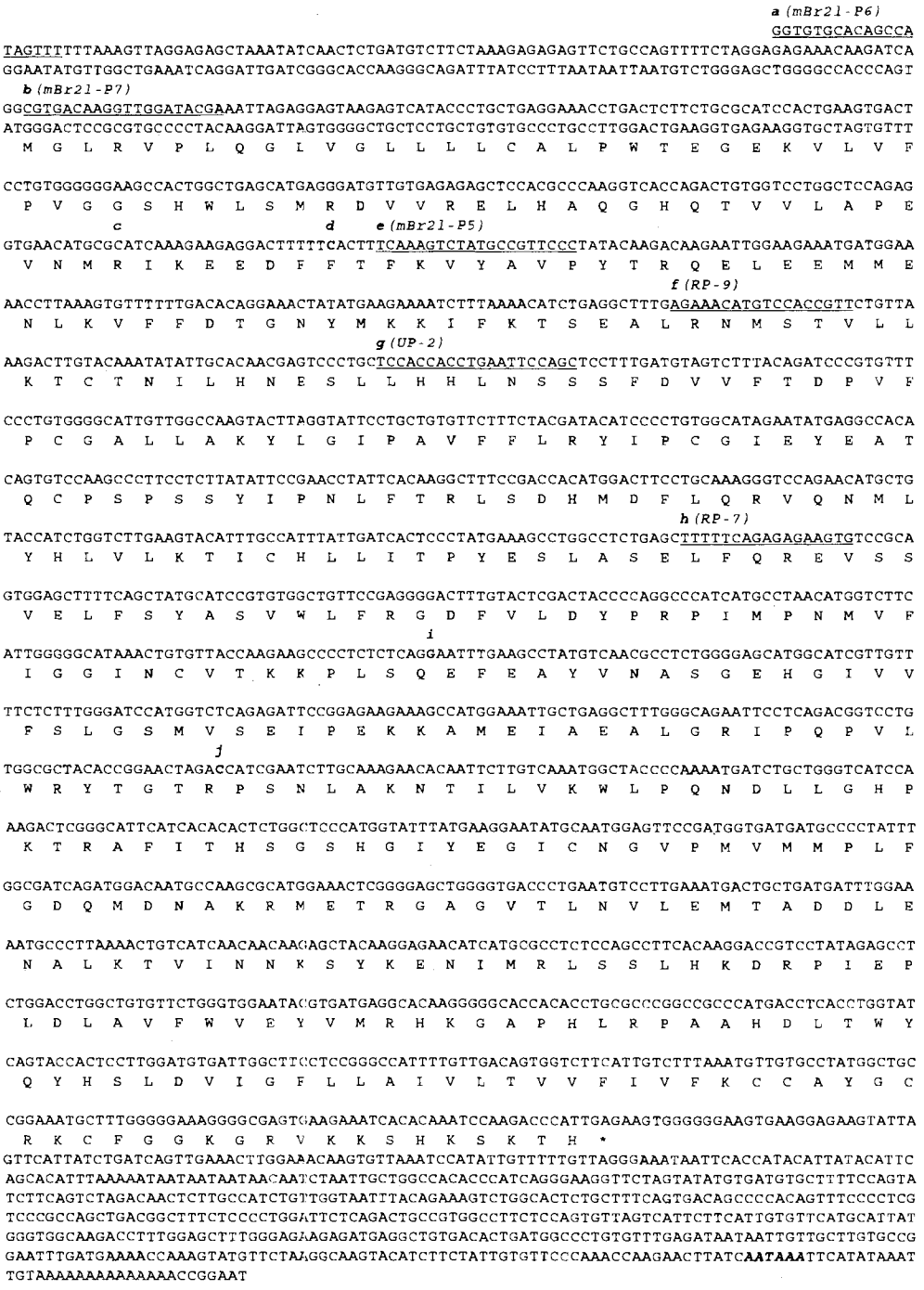

Fig. 1. Nucleotide and deduced amino acid sequences of mouse *UGTBr2-like* cDNA. The following symbols indicating: *a*–primer mBr2l-P6; *b*–primer mBr2l-P7; *c*–T instead of C in RT-PCR product; *d*–partial cDNA clone 6-3-2 begins; *e*–primer mBr2l-P5; *f*–primer RP-9; *g*–primer UP-2; *h*–primer RP-7; *i*–common region begins; *j*–clone 7a begins; and *–stop codon. Consensus polyadenylation signal is denoted by boldface italics.

with total RNA isolated from adult (20–25 g) male Swiss-Webster mice livers (Harlan, Madison, WI) using DNA sequences obtained from the P1 clone above. The RT-PCR would cover the truncated 5′-end of the cDNA clone 6-3-2. Based on the genomic sequence of P1 clone, two forward oligos mBr2l-P6 (5′-GGTGTGCACAGCCATAGTTT-3′), and mBr21-P7 (5′-CGTGACAAGGTTGGATACGAAA-3′) were designed. The reverse primer mBr2l-P5 (5′-GGGAACGGCAT AGACTTTGA-3′) is located in the 5′-end of clone 6-3-2 (Fig. 1). Total RNA was isolated with TRIzol Reagent (Gibco) and cDNA was synthesized using SuperScript Preamplification System (Gibco BRL). PCR reaction consisted of a 5-min "hot start" with AmpliWax PCR Gem 100 (Perkin Elmer) at 80°C, followed by 30 cycles of amplification (94°C, 30 sec; 58°C, 1 min; 72°C, 1 min). PCR product was cloned into TA vector (Invitrogen), and positive clone were sequenced with M13 and T7 primers.

Phylogenetic Methods

Amino acid sequences were aligned with the CLUSTAL V (17) software package with a gap penalty of 8.0 as described previously (18). Unrooted phylogenetic trees were constructed using the programs CONSENSUS, FITCH, PROTDIST, PROTPARS and SEQBOOT implemented in the Phylogeny Inference Package (PHYLIP) version 3.5c (19). Evolutionary distances were calculated invoking available options of the program PROTDIST. Bootstrap confidence levels for internal branches were estimated using either the SEQBOOT and CONSENSE program of PHYLIP.

RESULTS AND DISCUSSION

A 1202-bp partial cDNA clone (clone 7a) which was demonstrated to be a member of *UGT1* family was obtained by screening of a mouse liver λgt11 cDNA library. Using this partial cDNA as a probe, we rescreened another mouse liver cDNA library (λZAP) and obtained three cDNA clones 9-1 (*mUGTbr1*), A-1 (*mUGTp4*) and 6-3-2 (*mUGTbr2-like*). Two of these clones, 9-1 and A-1 have been published (8). These three clones have identical 3′-end which encode identical *C*-termini. Clone 6-3-2 contained 1974 bp without an intact ORF, therefore, it is a partial cDNA. In order to find the full length cDNA clone of 6-3-2, a mouse P1 genomic library was screened using sequences obtained from this partial cDNA. A P1 genomic clone, 1519 with about 70 kilobase insert was obtained, and designated as *mUGT1*. Based on the 5′-end nucleotide sequence of clone 6-3-2, sequencing primer, RP-9 was designed and used to sequence 1519, first towards the 5′ direction, and subsequently reverse direction. Fig. 1 shows the nucleotide sequence and deduced amino acid sequence of full-length clone 6-3-2. The full length cDNA encodes 529 amino acid polypeptides with a molecular weight of around 60 kDa. Amino acid sequence comparison with *mUGTbr1* and *mUGTp4* shows 36% and 37% similarity at the *N*-terminus, but with identical *C*-termini (8). Comparison with *mUGTbr2* reported by Koiwai *et al.* (11) shows 70% similarity at the *N*-terminus, but identical *C*-terminal region. Therefore, this cDNA clone is a member of the *UGT1* bilirubin subfamily, and was designated *mUGTbr2-like*. In order to ascertain that *mUGTbr2-like* indeed transcribes into mature mRNA transcript, two forward primers (mBr21-P6 and mBr2l-P7)—located at the 5′ end, and one reverse primer (mBr21-P5)—located at the 3′ end, near the beginning of the partial cDNA clone 6-3-2, were designed to carry out two sets of RT-PCR reactions with mouse liver RNA. Fig. 2 shows a 325 bp RT-PCR amplification product with primers mBr21-P5 and mBr21-P7 (lane 2), and a 521 bp product with primers mBr2l-P5 and mBr21-P6 (lane 3). When the P1 DNA (clone 1519) was used as the template, the same size of PCR products were amplified with same primers (lanes 4 and 5). The RT-PCR products were subcloned into TA vector and sequenced in both direction as described above. The nucleotide sequences was identical to the genomic sequence of P1 clone with the exception of one nucleotide at position 476, where T was found instead of C (Fig. 1). However, there is no difference in the amino acid encoded. The difference in the nucleotide is not clear, but could be due to fidelity of PCR reaction, or differences in the strains of mice.

Alignment of the deduced amino acid of the *N*-terminus of *mUGTbr2-like* with other *UGT1* bilirubin subfamily shows overall similarities of: 75% to *rUGTB5* (7), 70% to *mUGTbr2* (11), 68% to *rUGTB2* (7), 67% to *rUGT1*0* (br) (12), 62% to *hUGT1D* (br2) (13), 61% to *hUGT1C* (13), 59.2% to *rUGTB3* (7), 58.8% to *hUGT1E* (13), 58.5% to *lUGT1-4* (14), 57.4% to *rUGTB4* (7), 45.4% to *hUGT1A* (br1) (13), 43.3% to *rUGTB1* (br1) (7), and 42.6% to *mUGT9-1* (br1) (8). It is of interest to note that our *mUGTbr2-like* cDNA clone is more homologous to the rat *UGTB5* (7), whereas the cDNA clone obtained by Koiwai et al. (11), *mUGTbr2* is more homologous to *rUGTB2* (7) and *rUGT1*0* (12). When our clone, *mUGTbr2-like* and *mUGTbr2* (11) are compared to human *UGT1D* (br2) (13), the

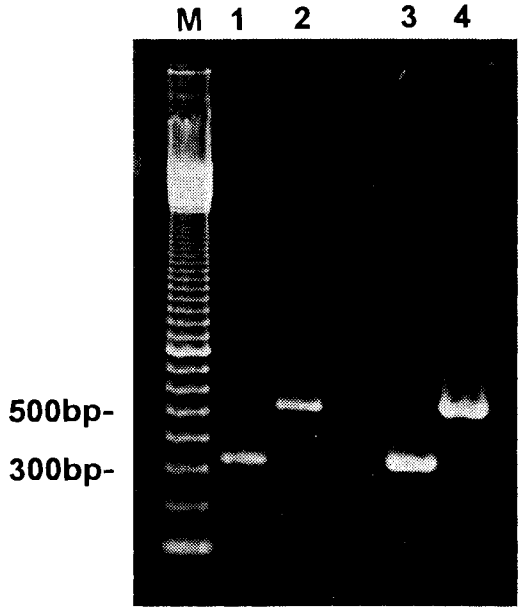

Fig. 2. RT-PCR. 1 μg of total RNA from mouse liver tissue was reverse transcriped into cDNA. Two sets of RT-PCR reactions were carried out with primers mBr2l-P5/mBr2l-P7; mBr2l-P5/mBr2l-P6. 5 μl of the 50 μl PCR reaction mixture was electrophoresed in 1.2% agarose gel. M: 100 bp ladder; Lane 1: 325 bp of RT-PCR product with primers mBr2l-P5 and mBr2l-P7; Lane 2: 521 bp of RT-PCR product with primers mBr2-P5 and mBr2l-P6. Lane 3: 325 bp of PCR product of P1 genomic DNA with same primers as lane 1; and Lane 4: 521 bp of PCR product of P1 genomic DNA with same primers as lane 2.

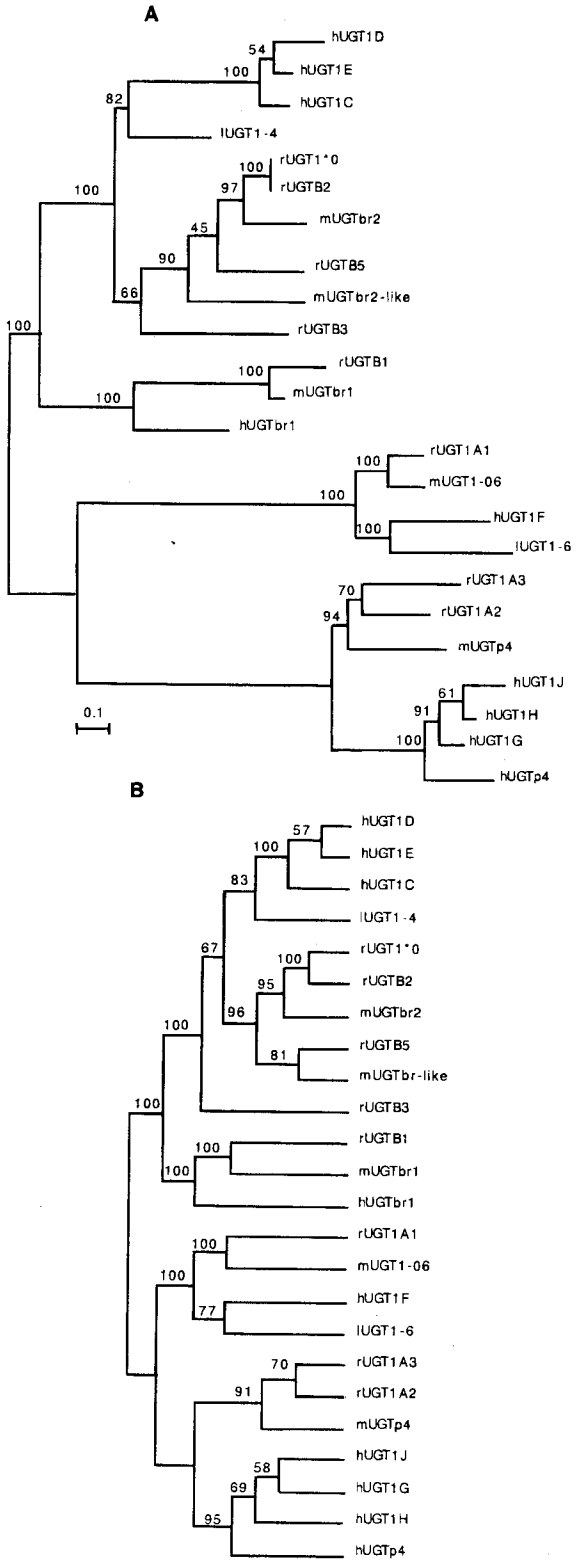

Fig. 3. (A) Distance-matrix tree (program FITCH) inferred from the amino acid sequences based on distances calculated by the "Dayhoff" option of program PRODIST (17,18). The same topology was obtained by using the "category" and "Kimura" methods (program PRODIST). Scale bar is in unit of amino acid substitution per sequence position. (B) Maximum parsimony tree. The lengths of the horizontal branches do not indicate distances. Numbers near the internal branches refer to the boostrap replications (out of 100 resamplings) confirming the grouping of the sequences of the right of the branch.

overall similarities are 62% and 60.1%, respectively, indicating that *mUGTbr2-like* is more homologous to *hUGT1D* (br2) than *mUGTbr2*. Future functional expression of these cDNA clones in cell lines would yield insights into their substrates specificity.

The family 1 of UGT (*UGT1*) and *mUGTbr2-like* protein sequences were multiply aligned with the entire inventory of available homologues in the Genebank (NCBI gene bank; http:/ ncbi.nlm.nih.gov). Unrooted trees were inferred by various methods from a restricted version of the alignment positions considered reliably alignable among all sequences of the *N*-terminal portion of *UGT1*, i.e., excluding the highly conserved alternatively-spliced *C*-terminal common region (complete and restricted alignments are available upon request from ATK).

Fig. 3 shows the evolutionary trees of *UGT1* protein sequences based on the evolutionary distances (A), and maximum parsimony (B). According to both methods, *UGT1* sequences reproducibly segregate into two major subfamilies, represented by the bilirubin *UGT* (upper half), and the phenol *UGT* (lower half). The phylogenetic analysis shows that most of the *UGT1* subfamilies arose from gene duplications predating the diversification of the rodentia, legamorphs and apes. The phenol *UGT1* subfamily can be divided into two groups; one comprising the sequences of *rUGT1A1* (7), *mUGT1-06* (9,10), *hUGT1F* (13), and *lUGT1-6* (10), and the other represented by the sequences of *rUGT1A3* (7), *rUGT1A2* (7), *mUGTp4* (8), and human *UGTs* (*1G*, *1H*, *1J*, and *p4*) (13). Similarly, the bilirubin *UGT1* subfamily cluster is divided into two groups; separating the *rUGTB1* (7), *mUGTbr1* (8), and the *hUGTbr1* (13) sequences from the other bilirubin *UGTs*. Whether the above classification accurately reflects the evolutionary history of the *UGT1* genes depends on the actual position of the root in the phylogenetic reconstruction. The addition of outgroup sequences (i.e., *UGT1* homologues from lower vertebrates) to the analysis should further clarify this issue in the future.

ACKNOWLEDGMENTS

We thank Dr. Ida S. Owens (NIH) for stimulating discussion on the evolutionary phylogenetic tree. Supported in part by grants GM 49172 and ES 06887 from the National Institutes of Health, a PhRMA Foundation Research Grant (Washington, D.C.). A.T.K. was the recipient of the 1994 American Association of Pharmaceutical Scientists (AAPS) Young Investigator Award in Pharmacokinetics, Pharmacodynamics and Drug Metabolism (PPDM).

REFERENCES

1. G. J. Dutton. *Glucuronidation of Drugs and Other Compounds*, CRC Press, Boca Raton, FL, 1980, pp. 3–78.
2. T. R. Tephly and B. Burchell. *Trends Pharmacol. Sci.* **11**:276–279 (1990).
3. B. Burchell, D. W. Nebert, D. R. Nelson, K. W. Bock, T. Iyanagi, P. L. M. Jansen, D. Lancet, G. J. Mulder, J. R. Chowdhury, G. Siest, T. R. Tephly, and P. I. Mackenzie. *DNA Cell Biol.* **10**:487–494 (1991).
4. T. N. Wight and R. P. Mecham. *Biology of Proteoglycans*, Academic Press, Orlando, FL, 1987.
5. A. W. Wolkoff, J. R. Chowdhury, and I. M. Arias. In C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (eds), *The Metabolic Basis of Inherited Disease*, McGraw Hill Book Co., New York, 1987, pp. 1367–1408.

6. J. K. Ritter, F. Chen, Y. Y. Sheen, H. M. Tran, S. Kimura, M. T. Yeatman, and I. S. Owens. *J. Biol. Chem.* **267**:3257–3261, 1992.
7. Y. Emi, S. Ikushiro, and T. Iyanagi. *J. Biochem.* **117**:392–399, 1995.
8. A. T. Kong, M. Ma, D. Tao, and L. Yang. *Pharm. Res.* **10**:461–465, 1993.
9. S. F. Reuter, V. Vasiliou, A. Puga, and D. W. Nebert. *Toxicologist* **14**:410, 1994.
10. J. G. Lamb, P. Straub, and R. H. Tukey. *Biochem.* **33**:10513–10520, 1994.
11. O. Koiwai, K. Hasada, Y. Yasui, Y. Sakai, H. Sato, and T. Watanabe. *Biochem. Gen.* **33**:111–122, 1995.
12. H. Sato, O. Koiwai, K. Tanabe, and S. Kashiwamata. *Biochem. Biophys. Res. Commun.* **169**:260–264, 1990.
13. I. S. Owens, and J. K. Ritter. *Pharmacogenetics* **2**:93–108, 1992.
14. T. Philips, M. Durazzo, C. Trautwein, B. Alex, E. F. Johnson, J. G. Straub, P. Straub, R. H. Tukey, and M. Manns. *Lancet* **344**:578–581, 1994.
15. F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (eds), *Current Protocols in Molecular Biology*, Greene Publishing Associates, New York, 1989.
16. F. Sanger, S. Nicklen, and A. R. Coulson. *Pro. Natl. Acad. Sci. USA* **74**:5463–5467 (1977).
17. D. G. Higgins, E. J. Bleasby, and R. Ruchs. *Comput. Appl. Biosci.* **8**:198–191, 1992.
18. M. Boccheta, E. Ceccarelli, R. Creti, A. M. Sanagelantoni, O. Tiboni, and P. Cammarano. *J. Mol. Evol.* **41**:803–812, 1995.
19. J. Felsenstein. PHYLIP (Phylogenetic Inference Package), Version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.