Vol. 195, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 31, 1993 Pages 120-127

IDENTIFICATION OF TWO HUMAN BRAIN ARYL SULFOTRANSFERASE cDNAS

Xiaoyi Zhu[1], Maurice E. Veronese[2], Claude C.A. Bernard[3], Lloyd N. Sansom[4] and Michael E. McManus[1,*]

[1] Department of Physiology and Pharmacology, University of Queensland, Qld 4072, Australia

[2] Department of Clinical Pharmacology, School of Medicine, The Flinders University of South Australia, S.A. 5042, Australia

[3] Department of Neuroimmunology, La Trobe University, Victoria 3083, Australia

[4] School of Pharmacy, University of South Australia, S.A. 5000, Australia

Received July 6, 1993

SUMMARY: A 1,179 bp and a 1,424 bp full-length aryl sulfotransferase cDNAs were isolated from a human brain cDNA library. Their coding domains are 93% identical, each encoding a cytosolic protein of 295 amino acids. Their deduced amino acid sequences of these cDNAs are also 93% identical. The 1179 bp brain cDNA has an identical coding domain to a previously reported human liver aryl sulfotransferase cDNA but it has a different 5' noncoding sequence. Northern blot analysis using a probe specific for the 1,424 bp cDNA identified a 1500 bp band in mRNA of human liver, colon, kidney and lung. In a human hepatocellular carcinoma the same band plus an extra larger band was also recognised. An intron of the gene encoding the 1424 bp cDNA was also identified.
© 1993 Academic Press, Inc.

Sulfotransferases transfer a sulfonate moiety from 3'-phosphoadenosine-5-phosphosulfate to a wide variety of exogenous and endogenous compounds such as phenols, alcohols, aromatic amines, N-hydroxylated aromatic amines and steroids (1-3). In general, this process leads to detoxification of xenobiotics but in the case of N-hydroxylated aromatic amines, it results in the metabolic activation of these compounds to electrophiles that are capable of covalent binding to DNA. At least four distinct human aryl sulfotransferases have been identified (4,5). To date little is known about the structure and function of these human sulfotransferases. The isolation of the cDNAs encoding these enzymes is the initial step towards developing: (1) in vitro heterologous expression systems to study their catalytic function, (2) tissue expression and regulation and (3) to unravel their gene, mRNA and protein structures (3,6). Recently we reported the isolation of a human liver aryl sulfotransferase cDNA (6), which encodes a protein that catalyses the sulfation of

* To whom reprint requests should be addressed.

0006-291X/93 $4.00
Copyright © 1993 by Academic Press, Inc.
All rights of reproduction in any form reserved.

both phenol and dopamine (unpublished data). This paper reports the identification of two human brain aryl sulfotransferase cDNAs.

MATERIALS AND METHODS

Isolation of a DNA probe by the polymerase chain reaction: Forward primer 5' GTGAATTCTCTCTTCTATGAAGACAT and reverse primer 5' TCATCGATTAGTACCTTTCCTCATGA (6) were used to perform the polymerase chain reactions (PCR) with DNA templates isolated from a human brain cDNA library constructed in bacteriophage Lambda-gt10 from mRNA of a 59 year old male who died of internal haemorrhage. The PCR reaction was performed with 0.08 μg templates, 0.05 μg of each primer, 200 μM each of the four deoxynucleotide triphosphates, 2.5 mM $MgCl_2$, 3 units of Taq DNA polymerase and 1 x Taq polymerase buffer (Promega, Australia) in a reaction volume of 100 μl. Thirty-one cycles of the PCR were carried out with denaturation at 94°C for 1 minute, annealing at 45°C for 1 minute and polymerisation at 72°C for 1 minute. The PCR amplified DNA was then purified with the "Prep-A-Gene" kit supplied by BioRad, Australia.

Screening a human brain cDNA library with the PCR amplified DNA: The PCR fragment was [α-^{32}P]dCTP (3000 Ci/mmole, Dupont) labelled with the "sequenase random primed DNA labelling kit" [United States Biochemical (USB), Australia]. The labelled DNA was used to probe the human brain cDNA library and 2.3 x 10^6 plaques at a density of about 1 x 10^5 plaques/150-mm plate were screened. Seven positive plaques were identified and rescreened until positive plaques were well separated. Recombinant Lambda DNAs were then isolated from these seven positives following the method of Benson and Taylor (7).

Subcloning and sequencing the full-length human aryl sulfotransferase cDNAs: The inserts of the isolated recombinant Lambda DNAs were excised with endonuclease *EcoRI* and ligated into *EcoRI*-restricted M13mp18 (Pharmacia, Australia). The ligations were performed at the ratio of 100 ng insert DNA to 20 ng EcoRI-restricted M13mp18 DNA with 1 unit T4 DNA ligase (Promega, Australia) in 10 μl volume at 4°C overnight. The ligations were transformed into JM101 cells (Stratagene, CA). Single-stranded M13mp18 recombinant DNAs were isolated and sequenced from both the 5' end and the 3' end of the inserts by the dideoxynucleotide chain termination method with the "T7 DNA sequencing kit" (Pharmacia, Australia).

Northern blot analysis: Human tissue mRNA was isolated and electrophoresed in a 1% agarose-formaldehyde denaturing gel as previously described (6,8). The use of human tissues in these studies had local research ethics committee approval. The following amounts of human tissue mRNA were added to the gel: 8 μg of liver, colon, kidney and lung mRNA and 6 μg of hepatocellular carcinoma mRNA. The mRNAs were then blotted onto a nylon membrane, which was prehybridized in 5 x SSC, 0.1% SDS, 0.05% sodium pyrophosphate, 10 μg/ml salmon sperm DNA and 1 x Denhardt's reagent at 68°C for more than two hours before being exposed to the radioactive probe. A DNA fragment (PCR-amplified with forward primer 5' GGTAGGATCACAATAGGCCAC and reverse primer 5' GGGTCTTACTAGTCATCCAGGCTGTAGTGA) corresponding to the region from base 1158 to 1380 of the 1424 bp cDNA was used to make the [α-^{32}P]dCTP labelled probe with the "Sequenase Random Primed DNA Labelled Kit". The membrane was hybridised in 5 x SSC, 0.1% SDS, 0.05% sodium pyrophosphate with a probe concentration of 0.1 μg/ml at 68°C for about 16 hours and then washed with 2 x SSC for 30 minutes at 45°C, 2 x SSC containing 1% SDS at 55°C for 30 minutes and 2 x SSC with 1% SDS at 65°C for 30 minutes. The membrane was then rinsed with 2 x SSC and exposed to Kodak X-ray film for about 72 hours.

RESULTS AND DISCUSSION

A 228 bp DNA fragment was isolated from a Lambda gt10 human brain cDNA library by the PCR method using the forward primer 5' GTGAATTCTCTCTTCTATGAAGACAT and the reverse primer 5' TCATCGATTAGTACCTTTCCTCATGA. This probe was then used to screen the above whole brain cDNA library. Seven positive clones were isolated and four of them had an insert of more than one thousand basepairs. These four inserts were subcloned into the M13mp18 vector and sequenced from both ends.

Insert 1 was a 1179 bp full-length cDNA (HAST2; Accession number: L19955) with the same coding domain as the human liver aryl sulfotransferase cDNA [termed HAST

```
HAST2                                                          GGT    3
HAST3     GTGACGGGGAGGCGGTGCCCGGGGCATCTCCGCGGCGGAACTCAGCCTGT   50
HAST3     GAGAAGTCACTGCTTTGGGGAGACCTGATCTGGCTGTGCCAGATGGACAC  100

HAST1     CCGGAGGCACGAGGCCAGGTTCCCAAGAGCTCAGGAACATGGAGCTGATC   50
HAST2     AAGGGAACGGGCCTGGCTCTGGCCCCTGACGCAGGAACATGGAGCTGATC   53
HAST3     TGAGAAAGAAGTAGAAGACTCAGAATTAGAAGAGGAACATGGAGCTGATC  150
             *      *        *              ******************

HAST1     CAGGACACCTCCCGCCCGCCACTGGAGTACGTGAAGGGGGTCCCGCTCAT  100
HAST2     CAGGACACCTCCCGCCCGCCACTGGAGTACGTGAAGGGGGTCCCGCTCAT  103
HAST3     CAGGACACCTCCCGCCCGCCACTGGAGTACGTGAAGGGGGTCCCGCTCAT  200
          **************************************************

HAST1     CAAGTACTTTGCAGAGGCACTGGGGCCCCTGCAGAGCTTCCAGGCCCGGC  150
HAST2     CAAGTACTTTGCAGAGGCACTGGGGCCCCTGCAGAGCTTCCAGGCCCGGC  153
HAST3     CAAGTACTTTGCAGAGGCACTGGGGCCCCTGCAGAGCTTCCAAGCCCGAC  250
          *******************************************.*****.*

HAST1     CTGATGACCTGCTCATCAGCACCTACCCCAAGTCCGGCACCACCTGGGTG  200
HAST2     CTGATGACCTGCTCATCAGCACCTACCCCAAGTCCGGCACCACCTGGGTG  203
HAST3     CTGATGACCTGCTCATCAACACCTACCCCAAGTCTGGCACCACCTGGGTG  300
          ******************.*************** ***************

HAST1     AGCCAGATTCTGGACATGATCTACCAGGGTGGTGACCTGGAGAAGTGTCA  250
HAST2     AGCCAGATTCTGGACATGATCTACCAGGGTGGTGACCTGGAGAAGTGTCA  253
HAST3     AGCCAGATACTGGACATGATCTACCAGGGCGGCGACCTAGAGAAGTGTAA  350
          ********.******************** ** *****.********* *

HAST1     CCGAGCTCCCATCTTCATGCGGGTGCCCTTCCTTGAGTTCAAAGCCCCAG  300
HAST2     CCGAGCTCCCATCTTCATGCGGGTGCCCTTCCTTGAGTTCAAAGCCCCAG  303
HAST3     CCGGGCTCCCATCTACGTACGGGTGCCCTTCCTTGAGGTCAATGATCCAG  400
          ***.**********.*.*.******************.****.*  ****

HAST1     GGATTCCCTCAGGGATGGAGACTCTGAAAGACACACCGGCCCCACGACTC  350
HAST2     GGATTCCCTCAGGGATGGAGACTCTGAAAGACACACCGGCCCCACGACTC  353
HAST3     GGGAACCCTCAGGGCTGGAGACTCTGAAAGACACACCGCCCCCACGGCTC  450
          **...********* ********************** *******.***

HAST1     CTGAAGACACACCTGCCCCTGGCTCTGCTCCCCCAGACTCTGTTGGATCA  400
HAST2     CTGAAGACACACCTGCCCCTGGCTCTGCTCCCCCAGACTCTGTTGGATCA  403
HAST3     ATCAAGTCACACCTGCCCCTGGCTCTGCTCCCTCAGACTCTGTTGGATCA  500
           * ***.************************* *****************

HAST1     GAAGGTCAAGGTGGTCTATGTTGCCCGCAACGCAAAGGATGTGGCAGTTT  450
HAST2     GAAGGTCAAGGTGGTCTATGTTGCCCGCAACGCAAAGGATGTGGCAGTTT  453
HAST3     GAAGGTCAAGGTGGTCTATGTTGCCCGAAACCCAAAGGACGTGGCGGTCT  550
          *************************** *** ******* *****.** *

HAST1     CCTACTACCACTTCTACCACATGGCCAAGGTGCACCCTGAGCCTGGGACC  500
HAST2     CCTACTACCACTTCTACCACATGGCCAAGGTGCACCCTGAGCCTGGGACC  503
HAST3     CCTACTACCATTTCCACCGTATGGAAAAGGCGCACCCTGAGCCTGGGACC  600
          ********** *** ***. ****  **** ********************

HAST1     TGGGACAGCTTCCTGGAGAAGTTCATGGTCGGAGAAGTGTCCTACGGATC  550
HAST2     TGGGACAGCTTCCTGGAGAAGTTCATGGTCGGAGAAGTGTCCTACGGATC  553
HAST3     TGGGACAGCTTCCTGGAAAAGTTCATGGCTGGAGAAGTGTCCTACGGGTC  650
          *****************.**********  *****************.**
```

Figure 1. Alignment of HAST1, 2 and 3 cDNAs. Adenine at base -3 critical to start translation, initiation codon ATG and stop codon TGA and polyadenylation signal AATAAA are underlined.

(6); now termed HAST1] previously isolated in our laboratory. From base 35 to the begining of its 25 bp poly-A tail, HAST2 is 100% identical to HAST1 (fig. 1). The 5' 34 bp noncoding region of the HAST2 cDNA was not obviously homologous to the corresponding 5' region of HAST1.

Inserts 2, 3 and 4 represented a unique cDNA (HAST3; Accession number: L19956) but differed in size. Insert 2 was a 1424 bp full-length cDNA and it was 90%, 48%, 53% homologous to HAST1 (6), the human alcohol/hydroxysteroid (9) and the bovine oestrogen sulfotransferase (10) cDNAs, respectively. HAST3 is therefore a unique aryl sulfotransferase cDNA. Insert 3 was a 1049 bp partial cDNA lacking the 5' 375 bp region of the HAST3. Insert 4 was also a partial HAST3 cDNA of 1711 bp long and it lacked

```
HAST1     CTGGTACCAGCACGTGCAGGAGTGGTGGGAGCTGAGCCGCACCCACCCTG  600
HAST2     CTGGTACCAGCACGTGCAGGAGTGGTGGGAGCTGAGCCGCACCCACCCTG  603
HAST3     CTGGTACCAGCACGTGCAGGAGTGGTGGGAGCTGAGCCGCACCCACCCTG  700
          **************************************************

HAST1     TTCTCTACCTCTTCTATGAAGACATGAAGGAGAACCCCAAAAGGGAGATT  650
HAST2     TTCTCTACCTCTTCTATGAAGACATGAAGGAGAACCCCAAAAGGGAGATT  653
HAST3     TTCTCTACCTCTTCTATGAAGACATGAAGGAGAACCCCAAAAGGGAGATT  750
          **************************************************

HAST1     CAAAAGATCCTGGAGTTTGTGGGGCACTCCCTGCCAGAGGAGACCGTGGA  700
HAST2     CAAAAGATCCTGGAGTTTGTGGGGCACTCCCTGCCAGAGGAGACCGTGGA  703
HAST3     CAAAAGATCCTGGAGTTTGTGGGGCGCTCCCTGCCAGAGGAGACCATGGA  800
          *************************.******************.****

HAST1     CTTCATGGTTCAGCACACGTCGTTCAAGGAGATGAAGAAGAACCCTATGA  750
HAST2     CTTCATGGTTCAGCACACGTCGTTCAAGGAGATGAAGAAGAACCCTATGA  753
HAST3     CTTCATGGTTCAGCACACGTCGTTCAAGGAGATGAAGAAGAACCCTATGA  850
          **************************************************

HAST1     CCAACTACACCACCGTCCCCCAGGAGTTCATGGACCACAGCATCTCCCCC  800
HAST2     CCAACTACACCACCGTCCCCCAGGAGTTCATGGACCACAGCATCTCCCCC  803
HAST3     CCAACTACACCACCGTCCCCCAGGAGCTCATGGACCACAGCATCTCCCCC  900
          ************************* ************************

HAST1     TTCATGAGGAAAGGCATGGCTGGGGACTGGAAGACCACCTTCACCGTGGC  850
HAST2     TTCATGAGGAAAGGCATGGCTGGGGACTGGAAGACCACCTTCACCGTGGC  853
HAST3     TTCATGAGGAAAGGCATGGCTGGGGACTGGAAGACCACCTTCACCGTGGC  950
          **************************************************

HAST1     GCAGAATGAGCGCTTCGATGCGGACTATGCGGAGAAGATGGCAGGCTGCA  900
HAST2     GCAGAATGAGCGCTTCGATGCGGACTATGCGGAGAAGATGGCAGGCTGCA  903
HAST3     GCAGAATGAGCGCTTCGATGCGGACTATGCGGAGAAGATGGCAGGCTGCA 1000
          **************************************************

HAST1     GCCTCAGCTTCCGCTCTGAGCTGTGAGAGGGGCTCCTGGGGTCACTGCAG  950
HAST2     GCCTCAGCTTCCGCTCTGAGCTGTGAGAGGGGCTCCTGGGGTCACTGCAG  953
HAST3     GCCTCAGCTTCCGCTCTGAGCTGTGAGAGGGGCTCCTGGAGTCACTGCAG 1050
          ****************************************.***********

HAST1     AGGGAGTGTGCGAATCAAACCTGACCAAGCGGCTCAAGAATAAAATATGA 1000
HAST2     AGGGAGTGTGCGAATCAAACCTGACCAAGCGGCTCAAGAATAAAATATGA 1003
HAST3     AGGGAGTGTGCGAATCTACCCTGACCAATGGGCTCAAGAATAAAGTATGA 1099
          ****************.* *********. ***************.*****

HAST1     ATTGAGGGCCTGGGACGGTAGGTCATGTCTGTAATCCCAGCAATTTGG-A 1049
HAST2     ATTGAGGGCCTGGGACGGTAGGTCATGTCTGTAATCCCAGCAATTTGG-A 1052
HAST3     TTTTTGAGTCAGGCACAGTGGCTCATGTCTGCAATCCCAGCGATTTGGGA 1149
          .**..*.* *.** **.**.* ********* *********.****** *

HAST1     GGCTGAGGTGGGAGGATCATTTGAGCCCAGGAGTTCGAGACCAACCTGGG 1099
HAST2     GGCTGAGGTGGGAGGATCATTTGAGCCCAGGAGTTCGAGACCAACCTGGG 1102
HAST3     GGTTGAGCTGGTAGGATCACAATAGGCCACGAATTTGAGACCAGCCTGG- 1198
          ** **** ***.******* ...** *** **.** *******.*****
```

Figure 1 - Continued

```
1.   GTGACGGGGAGGCGGTGCCCGGGGCATCTCCGCGGCGGAACTCAGCCTGTGAGAAGTCAC
61   TGCTTTGGGGAGACCTGATCTGGCTGTGCCAGATGGACACTGAGAAAGAAGTAGAAGACT
121  CAGAATTAGAAGAGGAACATGGAGCTGATCCAGGACACCTCCCGCCCGCCACTGGAGTAC
                     M  E  L  I  Q  D  T  S  R  P  P  L  E  Y

181  GTGAAGGGGGTCCCGCTCATCAAGTACTTTGCAGAGGCACTGGGGCCCCTGCAGAGCTTC
     V  K  G  V  P  L  I  K  Y  F  A  E  A  L  G  P  L  Q  S  F

241  CAAGCCCGACCTGATGACCTGCTCATCAACACCTACCCCAAGTCTGGCACCACCTGGGTG
     Q  A  R  P  D  D  L  L  I  N  T  Y  P  K  S  G  T  T  W  V

301  AGCCAGATACTGGACATGATCTACCAGGGCGGCGACCTAGAGAAGTGTAACCGGGCTCCC
     S  Q  I  L  D  M  I  Y  Q  G  G  D  L  E  K  C  N  R  A  P

361  ATCTACGTACGGGTGCCCTTCCTTGAGGTCAATGATCCAGGGGAACCCTCAGGGCTGGAG
     I  Y  V  R  V  P  F  L  E  V  N  D  P  G  E  P  S  G  L  E

421  ACTCTGAAAGACACACCGCCCCCACGGCTCATCAAGTCACACCTGCCCCTGGCTCTGCTC
     T  L  K  D  T  P  P  P  R  L  I  K  S  H  L  P  L  A  L  L

481  CCTCAGACTCTGTTGGATCAGAAGGTCAAGGTGGTCTATGTTGCCCGAAACCCAAAGGAC
     P  Q  T  L  L  D  Q  K  V  K  V  V  Y  V  A  R  N  P  K  D

541  GTGGCGGTCTCCTACTACCATTTCCACCGTATGGAAAAGGCGCACCCTGAGCCTGGGACC
     V  A  V  S  Y  Y  H  F  H  R  M  E  K  A  H  P  E  P  G  T

601  TGGGACAGCTTCCTGGAAAAGTTCATGGCTGGAGAAGTGTCCTACGGGTCCTGGTACCAG
     W  D  S  F  L  E  K  F  M  A  G  E  V  S  Y  G  S  W  Y  Q

661  CACGTGCAGGAGTGGTGGGAGCTGAGCCGCACCCACCCTGTTCTCTACCTCTTCTATGAA
     H  V  Q  E  W  W  E  L  S  R  T  H  P  V  L  Y  L  F  Y  E

721  GACATGAAGGAGAACCCCAAAAGGGAGATTCAAAAGATCCTGGAGTTTGTGGGGCGCTCC
     D  M  K  E  N  P  K  R  E  I  Q  K  I  L  E  F  V  G  R  S

781  CTGCCAGAGGAGACCATGGACTTCATGGTTCAGCACACGTCGTTCAAGGAGATGAAGAAG
     L  P  E  E  T  M  D  F  M  V  Q  H  T  S  F  K  E  M  K  K

841  AACCCTATGACCAACTACACCACCGTCCCCCAGGAGCTCATGGACCACAGCATCTCCCCC
     N  P  M  T  N  Y  T  T  V  P  Q  E  L  M  D  H  S  I  S  P

901  TTCATGAGGAAAGGCATGGCTGGGGACTGGAAGACCACCTTCACCGTGGCGCAGAATGAG
     F  M  R  K  G  M  A  G  D  W  K  T  T  F  T  V  A  Q  N  E

961  CGCTTCGATGCGGACTATGCGGAGAAGATGGCAGGCTGCAGCCTCAGCTTCCGCTCTGAG
     R  F  D  A  D  Y  A  E  K  M  A  G  C  S  L  S  F  R  S  E

1021 CTGTGAGAGGGGCTCCTGGAGTCACTGCAGAGGGAGTGTGCGAATCTACCCTGACCAATG
     L  *

1081 GGCTCAAGAATAAAGTATGATTTTTGAGTCAGGCACAGTGGCTCATGTCTGCAATCCCAG
1141 CGATTTGGGAGGTTGAGCTGGTAGGATCACAATAGGCCACGAATTTGAGACCAGCCTGGT
1201 AAAATAGTGAGACCTCATCTCTACAAAGATGTAAAAAAATTAGCCACATGTGCTGGCACT
1261 TACCTGTAGTCCCAGCTACTTGGGAAGCAGAGGCTGGAGGATCATTTCAGCCCAGGAGGT
1321 TGTGGATACAGTGAGTTATGACATGCCCATTCACTACAGCCTGGATGACAAGCAAGACCC
1381 TCCCTCCAAAGAAAATAAAGCTCAATTAAAATAAAAAAAAAAAA
```

Figure 2. Nucleotide and deduced amino acid sequence of HAST3.

the 5' 171 bp and the 3' 38 bp including the poly-A tail. However, Insert 4 contained a 497 bp intron (named HAST3INT; Accession number: L19957) at the position between base 732 and base 733 of HAST3 cDNA (figure 2). The intron started with a dinucleotide GT and ended with an AG (11; figure 3) and it was located inside the HAST3 coding region. When this intron sequence was compared with the DNA sequences in the Genbank, no obviously homologous sequences were found.

To identify mRNA coding for HAST3 in human tissues, Northern blot analysis was performed with mRNAs isolated from the human liver, colon, kidney, lung and hepatocellular carcinoma tissue. No human brain mRNA was available for analysis. The probe for this Northern blot was a 221 bp DNA fragment amplified from the 3'

```
1   tatgaagacatgaaggagGTGAGACCGACTGTGATGCTTCCCCCCATGTGACACCTGGGG
61  GCAGGCACCTCACAGGGACCCACCAAGGCCACCCAGCCCCGTCCCTGGGCGGCTCCCACA
121 GCAAGCCCGGATTCCCCATCCTACCTCCCTGGCCCAGGCCCCCCCACTGCAGCCCCACCT
181 GGCAGCAGGCTCGGCACAGCTTTCATCTTCTGCACCTGAGTCAGCTGCATGGGTGGCCAC
241 GGATCAGATACTTAGTCCTATTGCTTATCCTCACCAAAGGGTGTGCCACCAGGGCCACAG
301 TCATGGAAGAAGACCATCCCGGTCCTCACCCATAGGCGCCAAGCCCTGTTCATGATGGGA
361 TCACAGGGCAGAGATCAATTCATTTTACTCCAGAGACTAGGGCCCCAGGGGTTGAGGCTC
421 TTTGGGGTTTCTAGGGGAAGTGGCCAGATCCCCTCTGAGGTTAGAGAGGGGGACCCGTTT
481 TGTTTTGCTCCACTGAGGAGCCCTCTGCTGCTCAGaaccccaaaagggagattcaaaaga
```

Figure 3. The intron and its border in one of the HAST3 cDNA clones. Intron sequence starting with GT and ending with AG is in uppercase and sequence in lowercase corresponds to the region from base 715 to 758 of HAST3 in figures 1 & 2.

noncoding region of HAST3 cDNA and it was inside the three HAST3 inserts discussed above. This 221 bp fragment had 68% homology to HAST1 and HAST2. Dotblot experiment indicated that under the same hybridization conditions as the Northern blot, this 221 bp DNA fragment did not hybridize to either HAST1 or HAST2 (photo not shown). This unique probe identified a strong band around 1500 bases of mRNA in human liver, colon, kidney, lung and hepatocellular carcinoma (figure 4) indicating the expression of HAST3 mRNA in these tissues. A additional band of approximately 3,300 bp was also observed in the the hepatocellular carcinoma mRNA (figure 4). The identity and significance of this larger mRNA species in hepatocellular carcinoma tissue is currently unknown.

Figure 1 shows that the translation initiation codon ATG and stop codon TGA were conserved among HAST1, HAST2 and HAST3. The highly homologous regions of HAST1, HAST2 and HAST3 started (with a dinucleotide AG, which seems to be the 3' end of an intron!) from base -6 upstream from the common translation initiation codon and ended around 50 bases after the common translation stop codon (figure 1). Base -3 upstream from the translation initiation codon ATG was an adenine in all the three cDNAs and is critical for the sulfotransferase proteins to be translated from the ATG initiation codon (12; figure 1). The putative polyadenylation signal AATAAA in HAST1

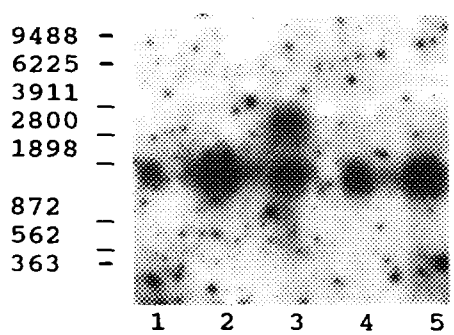

Figure 4. Northern blot analysis of HAST3 mRNA in human colon (1), liver (2), liver cancer (3), kidney (4) and lung (5).

```
HAST2AMI  - MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQARPDDLLISTYPKSG -50
            ||||||||||||||||||||||||||||||||||||||||||||||| ||||||
HAST3AMI  - MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQARPDDLLINTYPKSG -50

HAST2AMI  - TTWVSQILDMIYQGGDLEKCHRAPIFMRVPFLEFKAPGIPSGMETLKDTP -100
            |||||||||||||||||| |||| |||||| || ||| |||||||
HAST3AMI  - TTWVSQILDMIYQGGDLEKCNRAPIYVRVPFLEVNDPGEPSGLETLKDTP -100

HAST2AMI  - APRLLKTHLPLALLPQTLLDQKVKVVYVARNAKDVAVSYYHFYHMAKVHP -150
              ||| | ||||||||||||||||||||||| |||||||||  | | ||
HAST3AMI  - PPRLIKSHLPLALLPQTLLDQKVKVVYVARNPKDVAVSYYHFHRMEKAHP -150

HAST2AMI  - EPGTWDSFLEKFMVGEVSYGSWYQHVQEWWELSRTHPVLYLFYEDMKENP -200
            |||||||||||||| ||||||||||||||||||||||||||||||||||||
HAST3AMI  - EPGTWDSFLEKFMAGEVSYGSWYQHVQEWWELSRTHPVLYLFYEDMKENP -200

HAST2AMI  - KREIQKILEFVGHSLPEETVDFMVQHTSFKEMKKNPMTNYTTVPQEFMDH -250
            |||||||||||| |||||| |||||||||||||||||||||||||| |||
HAST3AMI  - KREIQKILEFVGRSLPEETMDFMVQHTSFKEMKKNPMTNYTTVPQELMDH -250

HAST2AMI  - SISPFMRKGMAGDWKTTFTVAQNERFDADYAEKMAGCSLSFRSEL -295
            |||||||||||||||||||||||||||||||||||||||||||||
HAST3AMI  - SISPFMRKGMAGDWKTTFTVAQNERFDADYAEKMAGCSLSFRSEL -295
```

Figure 5. Alignment of the amino acid sequences of HAST2 (HAST1) and HAST3. 21 amino acid differences were found.

was also conserved in the HAST2 and the HAST3 cDNAs (figure 1). In addition, there were two more putative polyadenylation signals in the HAST3 cDNA following the conserved one. The coding domain of the HAST2 cDNA had 100% identity to that of the HAST1 cDNA (6), whose protein when expressed in COS cells was able to sulfate a number of phenolic substrates (unpublished data). It is highly suggestive that HAST1 and HAST2 were transcribed from the same gene and the differences may be due to splicing. However, the isolation of a HAST2 cDNA from the cDNA library used in this study indicates that this aryl sulfotransferase is expressed in the human brain.

Similar to HAST1 and HAST2 cDNAs (Figure 1), HAST3 contained a coding region of the same length (885 bp), encoding an aryl sulfotransferase (Mr 34196 Da) of 295 amino acids. The coding region of HAST3 is 93% identical to that of HAST1 and HAST2 and single base substitutions accounted for this sequence divergence. HAST3's 5' noncoding region (138 bp) and 3' noncoding region (395 bp), which ended with a poly-A tail (Figure 2), are much longer than those of HAST1 and HAST2 (6; figure 1 & 2). The deduced amino acid sequence of HAST3 is also 93% identical (21 amino acid differences; figure 5) to that of HAST1 and HAST2. Like HAST1 and HAST2, no putative transmembrane hydrophobic domain was found in the HAST3 amino acid sequence (13) and therefore is most probably a cytosolic protein.

REFERENCES

1. Jakoby, W.B., Sekura, R.D., Lyon, E.S., Marcus, C.J. and Wang, J.L. (1980) Sulfotransferases. In Enzymatic Basis of Detoxification (W.B. Jakoby, Ed.), Vol. 2, pp. 199-228, Academic Press, New York.

2. Mulder, G.J. and Jakoby, W.B. (1990) Sulfation. In Conjugation Reactions in Drug Metabolism (G.J. Mulder, Ed.), pp. 107-161, Taylor and Francis Ltd., London.
3. Cruickshank, D., Sansom, L.N., Veronese, M.E., Zhu, X.Y., Mojarrabi, B. and McManus, M.E. (1993) Biochem. Biophys. Res. Commun. 191, 295-301.
4. Whittemore, R.M., Pearce, L.B. and Roth, J.A. (1985) Biochemistry 24, 2477-2482.
5. Whittemore, R.M., Pearce, L.B. and Roth, J.A. (1986) Arch. Biochem. Biophys. 249, 464-471.
6. Zhu, X.Y., Veronese, M.E., Sansom, L.N. and McManus, M.E. (1993) Biochem. Biophys. Res. Commun. 192, 671-676.
7. Benson, S.A. and Taylor, R.K. (1984) Biotechniques pp. 126-127.
8. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) Molecular Cloning-A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
9. Kong, A.N.T., Yang, L., Ma, M., Tao, D. and Bjornsson, T.D. (1992) Biochem. Biophys. Res. Commun. 187, 448-454.
10. Nash, A.R., Glenn, W.K., Moore, S.S., Kerr, J., Thompson, A.R. and Thompson, E.O.P. (1988) Aust. J. Biol. Sci. 41, 507-516.
11 Breathnach, R., Benoist, C., O'Hare, K., Gannon, R. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853-4857.
12. Kozak, M. (1989) J. Cell Biol. 108, 229-241.
13. Watson, M.E.E. (1984) Nucleic Acids Res. 12, 5145-5164.