BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **228,** 134–140 (1996)
ARTICLE NO. 1628

Genomic Organization and DNA Sequences of Two Human Phenol Sulfotransferase Genes (*STP1* and *STP2*) on the Short Arm of Chromosome 16

Thomas P. Dooley[1] and Zimei Huang

Molecular Pharmacology, Southern Research Institute, 2000 Ninth Avenue S., Birmingham, Alabama 35205

Received September 19, 1996

A family of human phenol sulfotransferase genes has been suggested by the cloning of numerous cDNA isolates from different tissues. We have previously cloned and sequenced the *STM* gene encoding the monoamine neurotransmitter-preferring sulfotransferase, M-PST, and a portion of the *STP1* gene encoding the phenol-preferring isozyme, P-PST1 (*BBRC* **205,** 1325–1332; *Genomics* **18,** 440–443). Both genes were mapped to a small region on the short arm of chromosome 16 (*BBRC* **205,** 482–489). Here we report on the sequencing and genomic organization of the *STP1* and *STP2* genes from a single cosmid clone obtained from chromosome 16p12.1-p11.2. *STP1* and *STP2* are 95.9% identical at the amino acid sequence level, whereas the *STM* gene is only 92.9% and 90.5% identical to *STP1* and *STP2*, respectively. Alignment of the genomic sequences indicated that all three genes have 7 coding exons and conserved intron-exon boundaries. These results facilitated the assignment of previously published cDNA isolates as ''alleles'' of the individual *STM*, *STP1*, and *STP2* loci on 16p, and provide to us a greater understanding of the complexity and roles of the phenol sulfotransferase gene family in the metabolism of endogenous and xenobiotic agents.

Phenol/aryl sulfotransferases (PST; EC 2.8.2.1) catalyze the transfer of the sulfonate group from PAPS (phosphoadenosine phosphosulfate) to various phenolic substrates (1, 2). Several PST isozymes are expressed in various human tissues and participate in the metabolism of endogenous and xenobiotic agents. The two major forms of human PST are the phenol-preferring isozyme, P-PST (or TS PST), and the monoamine neurotransmitter-preferring form, M-PST (or TL PST). Purified PSTs can sulfate catecholamines (e.g., dopamine), mono- and polycyclic phenols (e.g., p-nitrophenol, PNP), the hairgrowth stimulant minoxidil, and some tyrosine-containing peptides (1-2). M-PST is the major enzyme responsible for the sulfation of monoamine neurotransmitters in nervous tissue, as it has a K_m for dopamine in the micromolar range and simple phenols in the millimolar range, whereas P-PST exhibits an inverse substrate preference relationship.

Various cDNA isolates from different individuals and tissues have been sequenced over the past several years, suggesting the presence of at least three different gene loci in humans, and likely multiple alleles at each locus (3-9). The cDNAs encoding the human P-PST1 (formerly P-PST) and M-PST enzymes have been cloned. The P-PST1 enzyme is encoded by the *STP1* gene, and several cDNA clones and a portion of the genomic sequence have been reported by us and other groups (3-5, 10). The *STM* gene and corresponding cDNAs have likewise been isolated by several groups (5, 6, 11).

Both *STP1* and *STM* genes map to a small region of human chromosome 16p12.1-p11.2, which is homologous to a portion of mouse chromosome 7 (10, 12, 13). We have recently

[1] Author for correspondence. Fax: (205) 581-2877. E-mail: dooley@sri.org.

0006-291X/96 $18.00

reported that the human *STP1* and *STM* genes are contained within two non-overlapping cosmid contigs from 16p, termed 343.1 and 55.4, respectively (12). We have previously sequenced a 0.5 Kb portion of the *STP1* gene and all of *STM* (10, 11). Here we report the genomic DNA sequences of the human *STP1* gene and another highly-related gene, *STP2,* isolated from a cosmid clone containing both sequences, indicating that *STP2* also maps to chromosome 16p12.1-p11.2.

MATERIALS AND METHODS

In prior studies, the human *STP1* gene (formerly referred to as *STP*) was localized to chromosome 16p12.1-p11.2 (10) and to cosmid contig 343.1 (12). PCR screening of the cosmid clones in this contig with primers designed from the *STP1* cDNA sequence (3) identified positive cosmid clones, 330A9, 39G7, and 304C7 as containing *STP1*-related sequences (12). This cosmid contig was constructed from a chromosome 16-specific cosmid library in sCOS1 vector. Unless stated otherwise all methods have been published (14).

Cosmid 330A9 was digested with PstI (and BamHI), subcloned into pUC19 vector, and were transformed into DH5α cells. The ampicillin-resistant colonies were subsequently cultured in LB medium + 50 μg/ml ampicillin, and plasmid DNAs containing inserts were prepared by alkaline lysis (14). The DNAs were digested with PstI and the fragments were resolved by electrophoresis on a 1.2 % agarose gel. Southern blotting of digested plasmid subclones and parental cosmid clones followed by hybridization with [α-^{32}P]-dATP (Dupont-NEN, Boston, MA) labeled full-length human *STP1* cDNA (3) demonstrated cross-hybridization to various plasmid subclones, and indicated the presence of at least two full-length *STP*-related genes, which are referred to as *STP1* (i.e., the original *STP* gene) and *STP2.*

When necessary, polymerase chain reactions were performed with a Perkin-Elmer Cetus DNA Thermocycler, AmpliTaq DNA Polymerase (Perkin-Elmer, Norwalk, CT), and appropriate oligonucleotide primers (17 - 22 nucleotides in length) based on known phenol sulfotransferase cDNA sequences. The reaction conditions included a Perkin-Elmer Cetus PCR Core Reagent Kit and the following cycles: a single initial melting at 95° C for 5 minutes, denaturation at 94°C for 1 minute, annealing at 57°C for 2 minutes, extension at 72° C for 3 minutes, for a total of 30 cycles, followed by an additional 7 minutes at 72° C. All DNA fragments amplified by PCR were subcloned into pT7Blue(R) Vector (Novagen Inc., Madison WI), transformed into E. coli Nova Blue cells (Novagen), and plated onto LB agar plates containing IPTG, X-gal, and ampicillin. White transformants (i.e., β-galactosidase negative) containing inserts were grown in LB or 2xYT media containing 50 μg/ml of ampicillin.

The nucleotide sequences were determined by double-stranded, manual dideoxynucleotide chain-termination DNA sequencing using either the Sequenase 2.0 or 7-deaza-dGTP Sequencing Kits (USB Cleveland, OH), appropriate oligonucleotide primers, and [α-^{32}P]-dATP or [^{35}S]-dATP (Dupont-NEN). The reactions were resolved on 4-8% polyacrylamide 50% urea gel electrophoresis, followed by overnight exposure of Fuji RX or Kodak BIOMAX film. The sequences were obtained for both strands of DNA. All portions of the genomic sequence of *STP1* spanning the initiating methionine codon (Met^1) to the stop codon were determined, including all introns. All of the coding exons and the intron splice donors and acceptors of *STP2* were confirmed in both directions. However, some portions of the *STP2* introns were subjected to sequencing on only one strand. All sequences were analyzed using the MacVector 4.5 software on a Power Macintosh computer. The sequences were submitted to the GenBank database.

RESULTS AND DISCUSSION

The presence of multiple *STP1*-hybridizing bands on Southern blots of several cosmid clones within cosmid contig 343.1 from chromosome 16p suggested the presence of more than one phenol sulfotransferase gene on individual cosmid clones (i.e., in close proximity to one another; data not shown). Manual dideoxynucleotide DNA sequences were obtained from plasmid subclones of DNA fragments of cosmid 330A9 from contig 343.1, and identified two different genes when assembled and aligned. For comparison, we had previously identified two cosmid clones containing the *STM* gene, which is not present in the 343.1 contig, but is nearby (centromeric) on chromosome 16p11.2 (11, 12). The DNA sequences for the entire coding regions of two phenol sulfotransferase genes, *STP1* (formerly referred to as *STP*) and *STP2* were identified. The presence of *STP2* adjacent to *STP1* on a single cosmid clone (330A9) demonstrates that *STP2* maps to the same region of 16p12.1-p11.2, but slightly removed from the *STM* locus. Further evidence that the two genes (*STP1* and *STP2*) are adjacent to one another was provided by partial sequencing of DNA fragments for an independent partially-

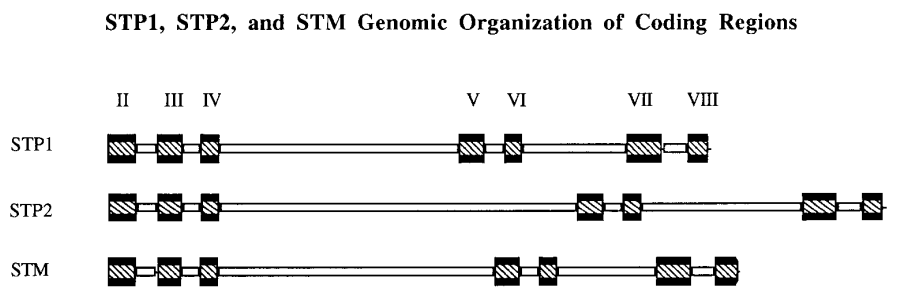

FIG. 1. Genomic organization of the coding regions of human *STP1, STP2,* and *STM* genes from exon II containing the translation initiator Met^1 codon (ATG) to the stop codon (TGA) of exon VIII: Introns are indicated as open boxes and exons are represented as filled boxes. The sizes of each intron and exon are indicated to scale.

overlapping cosmid clone 39G7. No evidence was obtained in support of more than two sulfotransferase genes within cosmids 330A9 or 39G7 from contig 343.1.

The genomic organization of *STP1* and *STP2* are compared to *STM* (11) in Figure 1. Depicted are the regions from the first coding exon (exon II, containing the ATG initiation codon) to the translation stop codon (exon VIII). Seven coding exons are conserved in all three genes. A 0.5 Kb portion near the 5′ end of *STP1* (containing exons II - IV) from another individual had been previously reported by us (10). The human pattern of genomic organization is very similar to that of rat arylsulfotransferase IV (15), the only example of a nonhuman species for which the genomic organization has been reported. The intron-exon boundaries are highly conserved between all three human genes, and each intron utilizes the consensus 5′-GU. . .AG-3′ splice donor/splice acceptor sites (Figure 2). Although the genomic organization is consistent within the gene family, some of the introns vary in length between the three separate loci, all of which were isolated from the same individual.

Portions of the 5′ and 3′ untranslated regions (UTRs) of the three gene were identified, and will be the focus of future studies, as alignment of multiple published cDNAs (3-6, 7-9, 16) to the genomic sequences indicates substantial heterogeneity in the 5′ UTRs (data not shown). These preliminary observations suggest the presence of multiple non-coding exon I and intron I sequences, with alternative splicing and/or alternative promoter utilization in all three genes. Of particular interest will be the identification of putative tissue-specific promoters within each gene.

Sequence alignment of the open reading frames (ORFs) of the three cytosolic sulfotransferase genes are presented in Figure 3. *STP1* and *STP2* are 95.9% identical at the amino acid sequence level, whereas the *STM* gene is only 92.9% and 90.5% identical to *STP1* and *STP2,* respectively. One may speculate that the three highly-related genes arose within the genome by gene duplication and possibly gene conversion. It will be interesting to determine whether other mammalian species share orthologs to all three genes, or whether this presumptive gene duplication occurred only in humans.

Having sequenced the full-length ORFs for these three genomic sequences, we were then able to align the translations of these genes with selected multiple published phenol sulfotransferase cDNA isolates (3-6, 7-9, 16). From this exercise it is clear that the cDNAs isolated in various laboratories worldwide can now be categorized as transcribed ''alleles'' of the three individual loci, *STP1, STP2,* and *STM* (Figure 4). Our results with *STP1, STP2,* and *STM* provide a framework for the classification of the available and subsequently isolated ''normal'' and ''variant'' cDNA and genomic sequences. In Figure 4, we note several examples of

Intron Splice Donor & Acceptor Sites

				MET[1]
INTRON I	STP1		...tgacgc**ag** GAAC	ATG
	STP2		...tgacgc**ag** GAAC	ATG
	STP		...tgatgc**ag** GAAC	ATG
INTRON II	STP1	CCG **gt**aagtga...	...ctctcc**ag** GCA	
	STP2	CCG **gt**aggtga...	...ctctcc**ag** GCA	
	STM	CTG **gt**aagtga...	...ctctcc**ag** GCA	
INTRON III	STP1	CAG **gt**gtgtga...	...ctgctc**ag** GGA	
	STP2	CAG **gt**gtgtgt...	...ctgctc**ag** GGA	
	STM	CAG **gt**gcatgg...	...ctactc**ag** GGC	
INTRON IV	STP1	AAG **gt**gaggca...	...ccctgc**ag** GTG	
	STP2	AAG **gt**gagact...	...ccccgc**ag** GTG	
	STM	AAG **gt**gaggcc...	...ccctgc**ag** GTG	
INTRON V	STP1	AAG **gt**gggttt...	...ttctcc**ag** TGT	
	STP2	AAG **gt**gggctt...	...ttctcc**ag** TGT	
	STM	AAG **gt**gggctt...	...ttctct**ag** TGT	
INTRON VI	STP1	GAG **gt**gagacc...	...ctgctc**ag** AAC	
	STP2	GAG **gt**gagacc...	...ctgctc**ag** AAC	
	STM	GAG **gt**gagacc...	...ctgctc**ag** AAC	
INTRON VII	STP1	AAG **gt**gggtgc...	...tcccccc**ag** GCA	
	STP2	AAG **gt**aggtgc...	...tcccccc**ag** GCA	
	STM	AAG **gt**gggtgc...	...tcccccc**ag** GCA	
CONCENSUS		... **gt**.....	.....**ag** ...	

FIG. 2. Conservation of intron splice donors (5′-GU) and splice acceptors (AG-3′) from introns 1-8: The translation start codon (ATG = Met[1]) is indicated and occurs 4 nucleotides 3′ to the splice acceptor of intron I (within the 5′ non-translated sequence).

translated cDNA sequence variants of the *STP1* and *STP2* loci, but none have yet been identified for the *STM* gene.

The reported *STP1*-encoded P-PST1 cDNA (3) appears to have three amino acid substitutions relative to the genomic sequence, and either contains sequencing errors and/or *bone fide*

Derived Amino Acid Sequences from Coding Exons

```
STP1  MELIQDTSRPPLEYVKGVPLIKYFAEALGPLQSFQARPDDLLISTYPKSGTTWVSQILDMIYQGGDLEKCHRAPIFMRVPFLEFKAPGIPSGMETLKDTP   100
STP2  ..................................................................................V...........N..
STM   .....................................N.....................N....YV......VND..E...L.......

STP1  APRLLKTHLPLALLPQTLLDQKVKVVYVARNAKDVAVSYYHFYHMAKVHPEPGTWDSFLEKFMVGEVSYGSWYQHVQEWWELSRTHPVLYLFYEDMKENP   200
STP2  ..............................................Y.H....E.......A.................................
STM   P...I.S.......................P..........HR.E.A..............A.................................

STP1  KREIQKILEFVGHSLPEETVDFMVQHTSFKEMKKNPMTNYTTVPQEFMDHSISPFMRKGMAGDWKTTFTVAQNERFDADYAEKMAGCSLSFRSEL        295
STP2  ............R........L..E.........T........RR.............................................
STM   ............R......M.........................L.............................................
```

FIG. 3. The translated amino acid sequences derived from the open reading frames for human *STP1*, *STP2*, and *STM* genomic sequences.

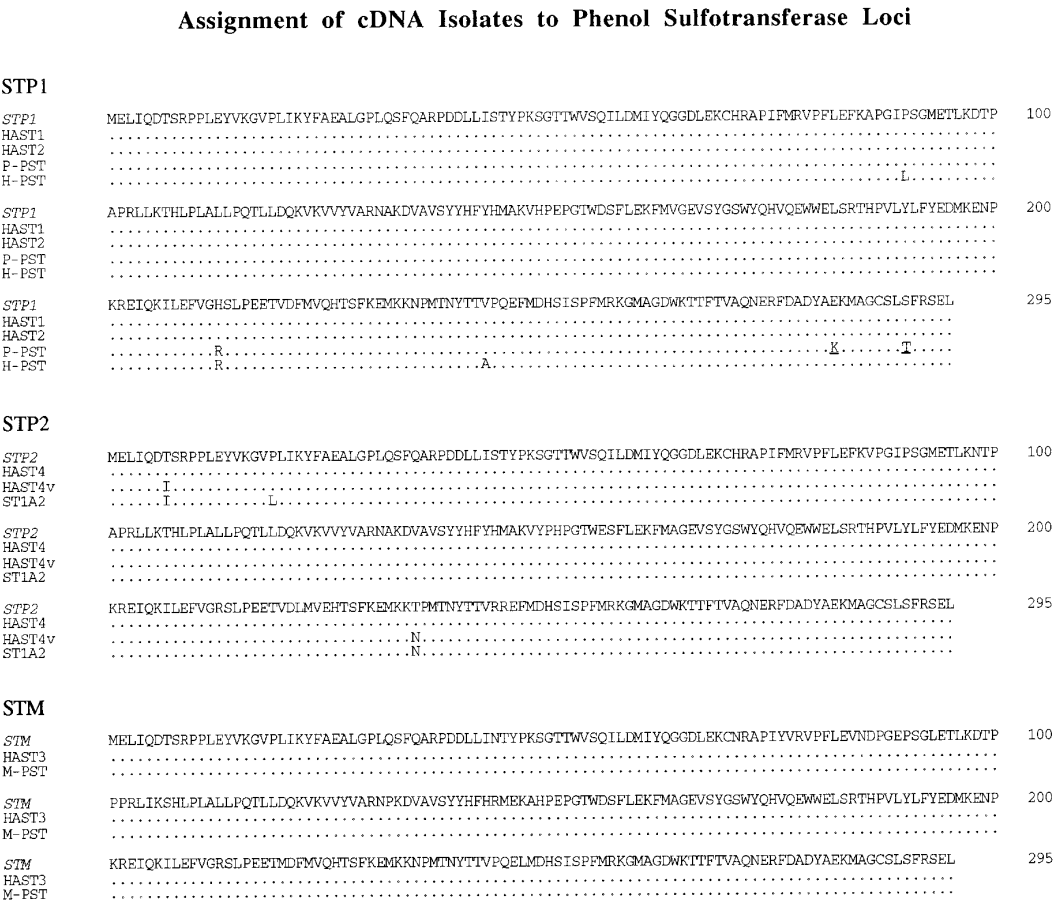

FIG. 4. Correspondence between phenol sulfotransferase genomic loci and selected published cDNA ''alleles'': Genomic sequences - *STP1* and *STP2* (this work), *STM* (11); cDNA isolates - HAST1 (4), HAST2 (5), P-PST (3), H-PST (16), HAST4 and HAST4v (7), ST1A2 (8), HAST3 (5), and M-PST/TL-PST (6). Note that the underlined ''K'' and ''T'' in the P-PST cDNA (3) are the results of sequencing errors and are not actually present in the cDNA clone.

sequence polymorphisms. In an effort to confirm this observation, we re-sequenced the original P-PST1 cDNA plasmid isolate (3), and observed that it contains several sequencing errors. For instance, the reported C-terminal two amino acid substitutions (K^{282} and T^{290} in Figure 4) result from sequencing errors, and are not in fact sequence variants. Thus, the original report (3) contains sequencing errors at both the DNA and protein sequence levels. Another cDNA sequence variant has been reported for *STP1* (16).

Although the *STP1*- and *STM*-encoded cDNA alleles have been well studied in biochemical experiments as recombinant proteins, only recently have the kinetic properties of the newly discovered *STP2*-encoded cDNAs been studied in any detail (7, 8). The *STP2* gene encodes a protein that is only 4% dissimilar to *STP1*, yet is much less active than the *STP1*-encoded protein. The cDNA ''alleles'' termed HAST4, HAST4v (a variant with two substitutions), and ST1A2 (a variant with three substitutions) exhibit higher Km values than *STP1*-encoded proteins against p-nitrophenol as a substrate, and are inactive at sulfating dopamine (7, 8). Therefore, the 4% dissimilarity between *STP1* and *STP2* accounts for substantial kinetic differences

at the enzyme level. And, the 7 - 9.5% differences between STM and the other related genes results in catecholamine substrate selectivity. The reduced activity in *STP2* relative to *STP1* might account for some of the known genetic diversity in humans at the phenol sulfotransferase enzymatic activity level.

The overlapping substrate recognition characteristics of the three phenol sulfotransferase genes confounds efforts to decipher the cytosolic sulfotransferase biochemical activity profiles obtained for human tissue and cell line homogenates. And, since the mRNAs for each gene are so highly related, gene expression studies using Northern blots might not discriminate between the messages for the various isozymes. The sequences reported here will be useful for designing oligonucleotide primers that are unique to each of the highly-related phenol sulfotransferase genes, in order to discriminate between *STP1, STP2, STM,* and other members of the cytosolic sulfotransferase gene superfamily (i.e., *STD* and *STE*) at the mRNA, cDNA, and genomic levels using PCR. The introns unique to each phenol sulfotransferase gene may also be used as probes for Southern blots to determine which genes are associated with known restriction fragment length polymorphisms (RFLPs) detected in human DNA samples, as the full-length human P-PST1 cDNA probe (3) detects both EcoRI and HindIII RFLPs (17, 18), although these polymorphisms have not yet been correlated with altered biochemical activity variants. Furthermore, this region of chromosome 16p has been physically mapped and several highly-informative dinucleotide repeat DNA markers (e.g., D16S298) have been identified in close proximity to the phenol sulfotransferase gene cluster (19). Both types of DNA polymorphisms in the vicinity of these loci should prove to be valuable for molecular genetic or pharmacogenetic studies of the metabolism of xenobiotic and endogenous molecules, and chromosome 16p-linked diseases. For instance, genetic polymorphisms affecting PST expression levels or enzymatic activities might impact on the efficacy of the hairgrowth stimulant, minoxidil (Rogaine^{TM}), which is bioactivated in the skin by sulfation (20).

ACKNOWLEDGMENTS

We are grateful to Drs. Norman Doggett, Sara Mole, and Patricia Munroe for prior collaborative published studies that served as a foundation for the present work; and to Pam Arneson, Shalini Nair, Roy Garcia, and Mindy Neal for technical assistance. This work was supported by the Alabama Power Company Endowed Chair Fund at Southern Research Institute, and a small grant from the Southwest Foundation Forum. During the preparation of this manuscript a report was published by Her et al. (Genomics 33:409-420, 1996) on the genomic organization of *STP2*.

REFERENCES

1. Falany, C. N. (1991) *Trends Pharmacol. Sci.* **2,** 255–259.
2. Campbell, N. R. C., Van Loon, J. A., and Weinshilboum, R. M. (1987) *Biochem. Pharmacol.* **36,** 1435–1446.
3. Wilborn, T. W., Comer, K. A., Dooley, T. P., Reardon, I. M., Heinrikson, R. L., and Falany, C. N. (1993) *Mol. Pharmacol.* **43,** 70–77.
4. Zhu, X., Veronese, M. E., Bernard, C. C. A., Sansom, L. N., and McManus, M. E. (1993) *Biochem. Biophys. Res. Commun.* **192,** 671–676.
5. Zhu, X., Veronese, M. E., Bernard, C. C. A., Sansom, L. N., and McManus, M. E. (1993) *Biochem. Biophys. Res. Commun.* **195,** 120–127.
6. Wood, T. C., Aksoy, I. A., Aksoy, S., and Weinshilboum, R. M. (1994) *Biochem. Biophys. Res. Commun.* **198,** 1119–1127.
7. Zhu, X., Veronese, M. E., Iocco, P., and McManus, M. E. (1996) *Int. J. Biochem. Cell Biol.* **28,** 565–571.
8. Ozawa, S., Nagata, K., Shimada, M., Ueda, M., Tsuzuki, T., Yamazoe, Y., and Kato, R. (1995) *Pharmacogenetics* **5,** s135–s140.
9. Rikke, B. A., and Roy, A. K. (1996) *Biochim. Biophys. Acta* **1307,** 331–338.
10. Dooley, T. P., Obermoeller, R. D., Leiter, E. H., Chapman, H. D., Falany, C. N., Deng, Z., and Siciliano, M. J. (1993) *Genomics* **18,** 440–443.
11. Dooley, T. P., Probst, P., Munroe, P. B., Mole, S. E., Liu, Z., and Doggett, N. A. (1994) *Biochem. Biophys. Res. Commun.* **205,** 1325–1332.

12. Dooley, T. P., Mitchison, H. M., Munroe, P. B., Probst, P., Neal, M., Siciliano, M. J., Deng, Z., Doggett, N. A., Callen, D. F., Gardiner, R. M., and Mole, S. E. (1994) *Biochem. Biophys. Res. Commun.* **205,** 482–489.
13. Aksoy, I. A., Callen, D. F., Apostolou, S., Her., C., and Weinshilboum, R. M. (1994) *Genomics* **23,** 275–277.
14. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
15. Khan, A. S., Taylor, B. R., Chung, K., Etheredge, J., Gonzales, R., and Ringer, D. P. (1993) *Gene* **137,** 321–326.
16. Hwang, S.-R., Kohn, A. B., and Hook, V. Y. H. (1995) *Biochem. Biophys. Res. Commun.* **207,** 701–707.
17. Henkel, R. D., Galindo, L. V., and Dooley, T. P. (1994) *Hum. Mol. Genet.* **3,** 2264.
18. Henkel, R. D., Galindo, L. V., and Dooley, T. P. (1995) *Hum. Genet.* **95,** 245–246.
19. Taschner, P. E. M., de Vos, N., Thompson, A. D., Callen, D. F., Doggett, N., Mole, S., Dooley, T. P., Barth, P. G., and Breuning, M. H. (1995) *Am. J. Hum. Genet.* **56,** 663–668.
20. Dooley, T. P., Walker, C. J., Hirshey, S. J., Falany, C. N., and Diani, A. R. (1991) *J. Invest. Dermatol.* **96,** 65–70.