0026-895X/03/6403-659–669$7.00
MOLECULAR PHARMACOLOGY
Copyright © 2003 The American Society for Pharmacology and Experimental Therapeutics
*Mol Pharmacol* 64:659–669, 2003

Vol. 64, No. 3
2396/1084480
*Printed in U.S.A.*



# Genetic Polymorphism of CYP1A2 in Ethiopians Affecting Induction and Expression: Characterization of Novel Haplotypes with Single-Nucleotide Polymorphisms in Intron 1

ELENI AKLILLU, JUAN ANTONIO CARRILLO, EYASU MAKONNEN, KARIN HELLMAN, MARIÀ PITARQUE, LEIF BERTILSSON, and MAGNUS INGELMAN-SUNDBERG

*Division of Molecular Toxicology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden (E.A., M.P., M.I.-S.); Department of Pharmacology, Medical School, University of Extremadura, Badajoz, Spain (J.A.C.); Department of Pharmacology, Faculty of Medicine, Addis Ababa University, Addis Ababa, Ethiopia (E.M.); and Division of Clinical Pharmacology, Department of Medical Laboratory Sciences and Technology at Karolinska Institutet, Huddinge University Hospital, Stockholm, Sweden (K.H., L.B.)*

Received January 13, 2003; accepted May 19, 2003

This article is available online at http://molpharm.aspetjournals.org

## ABSTRACT

CYP1A2 polymorphism has been well studied in white persons and Asians but not in Africans. We performed CYP1A2 genotype and phenotype analysis using caffeine in Ethiopians living in Ethiopia ($n = 100$) or in Sweden ($n = 73$). We sequenced the *CYP1A2* gene using genomic DNA from 12 subjects, which revealed a novel intron 1 single-nucleotide polymorphism (SNP), −730C>T. We developed SNP-specific polymerase chain reaction-restriction fragment length polymorphism genotyping and molecular haplotyping methods for the intron 1 SNPs, and four different haplotypes were identified: *CYP1A2*1A* (wild-type for all SNPs), *CYP1A2*1F* (−164A), *CYP1A2*1J* (−740G and −164A), and *CYP1A2*1K* (−730T, −740G, and −164A), having frequencies of 39.9, 49.6, 7.5, and 3.0%, respectively. The frequency of *CYP1A2*1J* and *CYP1A2*1K* among Saudi Arabians ($n = 136$) was 5.9% and 3.6%, and among Spaniards ($n = 117$) 1.3% and 0.5%, respectively. Functional significance of the different intron 1 haplotypes was analyzed. Subjects with *CYP1A2*1K* had significantly decreased CYP1A2 activity in vivo, and reporter constructs with this haplotype had significantly less inducibility with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in human B16A2 hepatoma cells. Electrophoretic mobility shift assay using nuclear extracts from B16A2 cells revealed a specific DNA binding protein complex to an Ets element. Efficient competition was obtained using oligonucleotide probes carrying the wt sequence and Ets consensus probe, whereas competition was abolished using probes with the −730C>T SNP alone or in combination with −740T>G (*CYP1A2*1K*). The results indicate a novel polymorphism in intron 1 of importance for Ets-dependent CYP1A2 expression in vivo and inducibility of the enzyme, which might be of critical importance for determination of interindividual differences in drug metabolism and sensitivity to carcinogens activated by CYP1A2.

CYP1A2, a hepatic enzyme inducible by smoking, metabolizes various chemical procarcinogens, such as food-derived heterocyclic and aromatic mutagens, *N*-heterocyclics found in tobacco smoke, and difuranocoumarins, to reactive carcinogens (McManus et al., 1990). It is also involved in the metabolism of several drugs such as paracetamol, theophylline, caffeine, and clozapine (Bertilsson et al., 1994). Endogenous substrates of CYP1A2 include estradiol and uroporphyrinogen. The enzyme has a significant role in chemical carcinogenesis (Eaton et al., 1995) and is induced by its substrates, and a polymorphism in its capacity to activate procarcinogens has been indicated (Minchin et al., 1985).

Hepatocellular carcinoma is a common neoplasm, especially in Africa, and is to a great extent caused by the intake of dietary aflatoxin (Uwaifo and Bababunmi, 1984). CYP1A2 has been reported to play a more important role than CYP3A4 in the bioactivation of aflatoxin at low concentrations in human liver microsomes (Gallagher et al., 1996). The antiparasitic drug oltipraz, which is currently on phase II human clinical trials for its cancer-chemopreventive effect in humans, especially with respect to aflatoxin-associated hepa-

This study was supported by grants from The Swedish Research Council and from the National Institutes of Health (National Institute of General Medical Sciences Grant 1-R01-GM60548-01A2).

**ABBREVIATIONS:** SNP, single-nucleotide polymorphism; bp, base pair(s); PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; wt, wild type; mut, mutated; PBS, phosphate-buffered saline; DTT, dithiothreitol; PMSF, phenylmethylsulfonyl fluoride; EMSA, electrophoretic mobility shift assay; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; GABP-$\alpha$, GA binding protein-$\alpha$; PEA3, polyomavirus enhancer A binding protein 3; XRE, xenobiotic response element; MR, metabolic ratio; ANOVA, analysis of variance; 95% CI, 95% confidence interval; PCR, polymerase chain reaction; AFMU, 5-acetylamino-6-formylamino-3-methyluracil; bp, base pair(s); kb, kilobase(s). The intron 1 SNPs in the *CYP1A2* gene are defined as follows: *CYP1A2*1A*, wt; *CYP1A2*1F* carries −164C>A; *CYP1A2*1J* carries −164C>A and −740T>G; *CYP1A2*1K* carries −164C>A, −730C>T, and −740T>G.

tocarcinogenesis, has been shown to be a potent inhibitor of CYP1A2 (Sofowora et al., 2001). Subjects with higher CYP1A2 activity and exposed to dietary aflatoxin B1 might thus be at a higher risk for developing hepatocellular carcinoma. Individual differences in CYP1A2 activity may thus influence individual susceptibility to cancer and the therapeutic efficacy of some drugs.

Several studies have indicated the presence of wide interindividual and ethnic differences in CYP1A2 activity when caffeine has been used as a probe drug. Striking differences (greater than 15-fold) in levels of CYP1A2 mRNA expression from human liver (Ikeya et al., 1989) and polymorphic metabolism of procarcinogens by human liver microsomes have been reported (Minchin et al., 1985). Unlike other drug-metabolizing cytochromes P450, such as CYP2D6 and CYP2C19, no nucleotide differences that could clearly explain the phenotypic variability in *CYP1A2* gene expression or inducibility have been identified.

The human *CYP1A2* gene, located on chromosome 15, spans about 7.8 kb and contains seven exons. The coding region starts at nucleotide 10 of exon 2 (Ikeya et al., 1989). Exon 2-6 is highly conserved among human, mouse, and rat. In these species, regions of high conservation have also been found in intron 1 of CYP1A2 (Ikeya et al., 1989), suggesting a possible regulatory role of this intron. Two single-nucleotide polymorphic sites (SNPs), −164G>A and −740T>G, have previously been reported in the intron 1 of human *CYP1A2* gene (see http://www.imm.ki.se/CYPalleles/cyp1a2.htm). The −164C>A SNP has been suggested to be associated with higher enzyme inducibility by smoking among white persons (Sachse et al., 1999), whereas the −740T>G has not been functionally characterized.

Previous studies on the human *CYP1A2* gene regulation in HepG2 cells have identified two regions of importance for basal expression: a proximal region containing a GC box, a CCAAT box, and a TATA box, and a distal region, named "1A2 enhancer," which contains two activator protein-1 sites, a xenobiotic-responsive element, and a hepatic nuclear factor 1 site and a second TATA box (Quattrochi et al., 1994, 1998; Chung and Bresnick, 1995, 1997). The aryl hydrocarbon receptor null mice show significant decrease in CYP1A2 expression in the liver, suggesting that the xenobiotic-responsive elements may be involved in the regulation of the basal expression (Schmidt et al., 1996). It is possible that several of these factors, including nuclear factor 1, participate in the tissue-selective activation of *CYP1A2* gene expression and that the absence of any single component may abolish or down-regulate gene expression.

In the present investigation we have evaluated interindividual variability in CYP1A2 activity in an African population and compared the activity between Ethiopians living in Sweden and Ethiopia to investigate any environmental influence. We have found new CYP1A2 haplotypes with SNPs in intron 1, which affect binding of nuclear proteins and inducibility in reporter gene systems, and correlate to the CYP1A2 activity monitored in vivo using caffeine as a probe drug.

## Materials and Methods

**Subjects.** Unrelated Ethiopians living in Ethiopia ($n$ = 100) or living in Sweden ($n$ = 73) participated in this study. The Ethiopians living in Sweden have been described previously (Aklillu et al., 2002a). Participants from Sweden were recruited among two groups: subjects adopted as small children by Swedish parents and now about 20 to 30 years of age ($n$ = 5) and subjects who left their home country and lived in Sweden for more than 10 years ($n$ = 30), 5 to 10 years ($n$ = 31), and 3 to 5 years ($n$ = 7). In addition, all subjects gave information about 1) their ethnic group, 2) time of arrival to Sweden, 3) health status, 4) dietary habits, 5) drug intake, and 6) smoking habits using a detailed questionnaire. The subjects were of a mixed Ethiopian origin and were of the Oromo, Amhara, Tigriyan, and Gurage ethnic groups. The Human Ethics Committees at Huddinge University Hospital, Karolinska Institutet, Stockholm, Sweden and the National Ethics committees at Ethiopian Science and Technology commission, Addis Ababa, Ethiopia, approved the study. The Saudi Arabians ($n$ = 136) and Spaniards ($n$ = 117) genotyped for CYP1A2*J and CYP1A2*K have been described in McLellan et al. (1997) and Oscarson et al. (1998), respectively.

**Caffeine Phenotyping.** The caffeine urinary test was performed according to the method of Carrillo et al. (2000). Briefly, the subjects were refrained from taking coffee, tea, Coca-Cola, chocolate, or any caffeine-containing beverage for at least 24 h before and throughout the study. Written information about the volunteer's habit of smoking as well as coffee, tea, or other caffeine-containing beverage intake was obtained from a detailed questionnaire. Subjects received a 100-mg oral dose of caffeine (Koffein; ACO AB, Helsingborg, Sweden) before bedtime, and 0- to 8-h urine was collected. The volume and pH of the urine collected were measured, pH was adjusted to 3.5 with 0.1 M HCl, and 20-ml aliquots were stored at −20°C until analysis. The molar concentrations of caffeine and its metabolites (micromoles per liter) were analyzed twice in one run and in duplicate by high-performance liquid chromatography. The urinary caffeine *N*-3-demethylation index, calculated and used as probe for CYP1A2 activity, included the following metabolites (AFMU + 1U + 1X + 17U + 17X)/137X) (Carrillo et al., 2000), where AFMU is 5-acetylamino-6-formylamino-3-methyluracil, 1U is 1-methyluracil, 1X is 1- methylxanthine, 17U is 1,7-dimethyluric acid, 17X is 1,7-dimethylxanthine (paraxanthine), and 137X is 1,3,7-trimethylxanthine (caffeine). The reliability and reproducibility of 0- to 8-h urinary caffeine (AFMU + 1U + 1X + 17U + 17X)/137X) metabolic ratio for estimation of in vivo CYP1A2 activity in humans has been evaluated (Carrillo et al., 2000).

**Isolation of DNA.** A 10-ml venous blood sample was taken from each subject into an EDTA-containing Vacutainer tube, and DNA was isolated from peripheral leukocytes using a guanidinium-isothiocyanate method.

**Genomic Sequencing of the Human *CYP1A2* Gene.** The *CYP1A2* gene was PCR-amplified into two different fragments from 12 different subjects previously phenotyped with caffeine and found to have a very low or very high caffeine metabolic ratio. A 5307-bp fragment (fragment 1) from −1593 at the 5′-flanking regions down to part of intron 6 was amplified by long PCR using a forward primer 1A2F and reverse primer 1A2R. Fragment 1 was also used as a template for intron 1 SNP genotyping and haplotyping after 10× dilution of PCR product with water. The second fragment, 2104 bp long, covering exon 7 to a part of the 3′-flanking region, was amplified by long PCR using primer seq-ex7FA and seq-ex7RB. The long PCR condition for sequencing is described in Aklillu et al. (2002b). The entire coding regions as well as intron-exon junctions were sequenced in both directions using fragments 1 and 2 as a template with forward and reverse sequence primers. Primer sequences are listed in Table 1. DNA sequencing was performed using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA) and analyzed on an ABI Prism 377 DNA sequencer.

**Genotyping of the CYP1A2 Intron 1 SNPs.** Genotyping was carried out for −164C>A according to the method of Nordmark et al. (2002) using allele-specific primers and fluorescent-labeled reporter probes. Real-time PCR was performed using Taqman Universal PCR

Master mix, and amplification and detection were performed in an ABI 7700 sequence detection system (Taqman). Genotyping for the −740T>G and the newly identified SNP, −730C>T in intron 1, was performed using a PCR-RFLP method. A 168-bp region covering the two polymorphic sites in intron 1 was amplified from fragment 1 by PCR using primers, 1A2promF and 1A2promR followed by RFLP. The PCR I condition is described in Aklillu et al. (2002b), using 1 mM $MgCl_2$ and annealing at 54°C. The −740T>G polymorphism and the −730C>T abolishes the *Ava*II and *Nci*I restriction sites, respectively. Fourteen microliters of the PCR product was therefore analyzed for the presence of −740T>G and −730C>T by digestion at 37°C overnight in 20 μl of total reaction mixture containing 1× NEBuffer 4 with 8 U of *Ava*II and *Nci*I (New England Biolabs, Beverly, MA), respectively. The products were subsequently separated on a 3% agarose gel. *Ava*II digestion gives 49- and 119-bp fragments and *Nci*I digestion gives 61- and 107-bp fragments.

**CYP1A2 Intron 1 SNPs Mapping and Determination of Haplotypes.** All subjects heterozygous for −740 and −164 SNPs in intron 1 were further reanalyzed for linkage disequilibrium in the following manner. The two alleles were first separated by allele-specific PCR with respect to the presence or the absence of the −740 SNP followed by genotyping of each individual allele for the −164 SNP. Part of *CYP1A2* amplified by long PCR (fragment 1) was used as a template in the haplotype-specific PCR II analysis after a 10-fold dilution in water. In brief, a 665-bp fragment containing part of intron 1 was amplified from fragment 1 using primers 1A2-740 wt F or 1A2-740 mut F, separately, and a common reverse primer 1A2-164R. DNA samples being homozygous wt and homozygous mutated for −740 SNP were used as controls to ensure the absence of non-specific amplification. The PCR II condition was according to the method of Aklillu et al. (2002b). The PCR products amplified by the wt primer as well as with the mutant specific primer were subjected subsequently for −164 genotyping by RFLP separately. The −164C>A abolishes the *Apa*I restriction site. Fourteen microliters of the PCR II product was subsequently analyzed for the presence of −164C>A by digestion with 8 U of *Apa*I (Invitrogene) at 25°C overnight in 20 μl of total reaction mixture containing 1× React buffer IV. The products were subsequently separated on a 3% agarose gel. *Apa*I digests the −164 wt allele giving 594- and 71-bp fragments.

To analyze the linkage disequilibrium of the −740T>G and the −730C>T, 14 μl of the PCR product used for genotyping of these SNPs was subjected for simultaneous digestion with 8 U of *Ava*II and *Nci*I (New England Biolabs) at 37°C overnight in 20 μl of total reaction mixture containing 1× NEBuffer 4.

**Nuclear Extract Preparation.** Crude nuclear extracts used for electrophoretic mobility shift assay were prepared according to the method of Zhu and Pfaff (1994). The human hepatoma cell line, B16A2, used for nuclear extract preparation and cell transfection assay, was kindly provided by Dr. Laurent Corcos and Dr. Andre Guillouzo (University of Rennes, Rennes, France). In brief, cells were grown in a 100-mm plate to 100% confluence and were kept for 3 to 4 additional weeks before preparing the nuclear extracts. The cells were washed with ice-cold PBS twice and collected in 1 ml PBS, centrifuged at 500*g* for 5 min at 4°C. The pellet was washed once with PBS and resuspended in 5× pellet volume of hypotonic buffer A (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 10 μg/ml leupeptin), kept on ice for 10 min, and centrifuged for 3 min at 4°C. The cells were resuspended in 200 μl of hypotonic buffer A and pooled together, and 1 ml of hypotonic buffer A was added, followed by homogenization. The cell lysate was centrifuged for 20 min at 4°C; the nuclear pellet was resuspended in 1 to 1.5 volumes of ice-cold salt buffer C (20 mM HEPES, pH 7.9, 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% glycerol, 0.5 mM DTT, 0.5 mM PMSF, 10 μg/ml leupeptin), kept on ice for 30 min, and centrifuged at 6000*g* for 30 min at 4°C, and the supernatant was aliquoted and stored at −70°C until use.

**Electrophoretic Mobility Shift Assay.** The sequences of the oligonucleotide probes used in EMSA are listed in Table 2. The double-stranded oligonucleotide probes for EMSA were prepared as follows. The sense and antisense oligonucleotides for the respective probe were annealed by incubation at 88°C for 2 min, 65°C for 10

TABLE 1

Sequences of primers used for genomic DNA sequencing and CYP1A2 genotyping and haplotyping

The letters F and R at the end of the primer name indicate that the primer is a forward and reverse primer, respectively. Location of primers is numbered according to GenBank accession number AF253322 (this sequence has −164C>A SNP).

| Primer Name | Location[a] | Primer Sequence (5′ to 3′) |
|---|---|---|
| 1A2–1F | 33349–33370 | CATCTAATCTCCAGTCCGTGCT |
| 1A2-1R | 38634–38655 | GCCATCCTAGTTGATTCCTGTG |
| Seq-1F | 33512–33531 | GGCCTCTCTTTAGGATGCAA |
| Seq-ex1F | 33927–33946 | GCCAAGAGTTGATCCTTCCA |
| Seq-ex1R | 33600–34619 | AAGGACAGCACCCTTGTCAG |
| Seq-ex2FA | 34507–34527 | TGGTCCCTTGGGTATATGGA |
| Seq-ex2RA | 35361–35380 | AAGGTGTTGAGGGCATTCTG |
| Seq-ex2FB | 34886–34905 | TGTCTCCATCTCAACCCTCA |
| Seq-ex2RB | 35946–35965 | GGCTTCAAGGATGAGGAAAC |
| Seq-ex3F | 36270–36289 | AGAGACCAAGTTGGGAGGAT |
| Seq-ex3R | 36665–36684 | ACTGGGAATTGGTGGCATAG |
| Seq-ex4F | 36862–36881 | GGCCAGAGAAAGCTAATGCT |
| Seq-ex4R | 37181–37200 | TCCCACATCTTCTAGGCAAG |
| Seq-ex5F | 37177–37196 | TGCCCTTGCCTAGAAGATGT |
| Seq-ex5R | 37520–37501 | GCTGAGAAGCCAGGAAGAGA |
| Seq-ex6F | 38211–38230 | AGAACAGGACGCAAGGAAGA |
| Seq-ex6R | 38650-38631 | CCTAGTTGATTCCTGTGCAC |
| Seq-ex7FA | 39852–39871 | AGTTTGGTTCCTTCCCACCT |
| Seq-ex7RA | 40866–40885 | GGGTCAGGAGTTCGAGACAA |
| Seq-ex7FB | 40717–40737 | ATCTCAGCTCACTGCAACCT |
| Seq-ex7FC | 41110–41129 | CTCCCATGTTCAAGCGATTC |
| Seq-ex7RB | 41936–41955 | GAAGGAGGTGCCTATCAACG |
| 1A2 prom F | 34153–34170 | TGGAAGCTAGTGGGGACA |
| 1A2 prom R | 34303–34320 | TTGTGCTAAGGGGGAAGC |
| 1A2-740 wtF | 34186–34203 | TGGGCTAGGTGTAGGGGT |
| 1A2-740 mutF | 34186–34203 | TGGGCTAGGTGTAGGGGG |
| 1A2-164R | 34831–34850 | CTGATGCGTGTTCTGTGCTT |

TABLE 2

List of oligonucleotide probes and their sequences used in EMSA

wt probe 1 spans the −740 and −730 SNP sites, and wt probe 2 spans the −164 SNP site. The polymorphic sites are indicated in bold in the wt oligonucleotide probes and underlined in the respective mutant oligonucleotide probe.

| Name | Oligonucleotide Sequence from 5′ to 3′ |
|---|---|
| wt probe 1 | GGG CTA GGT GTA GGG **GT**C CTG<br>AGT TC**C** GGG CTT TGC TAC CC<br>CCC GAT CCA CAT CCC C**A**G GAC<br>TCA AG**G** CCC GAA ACG ATG GG |
| −740 mut | GGG CTA GGT GTA GGG G**G**C CTG<br>AGT TCC GGG CTT TGC TAC CC<br>CCC GAT CCA CAT CCC C**C**G GAC<br>TCA AGG CCC GAA ACG ATG GG |
| −730 mut | GGG CTA GGT GTA GGG GTC CTG<br>AGT TC**T** GGG CTT TGC TAC CC<br>CCC GAT CCA CAT CCC CAG GAC<br>TCA AG**A** CCC GAA ACG ATG GG |
| −740 and −730 mut | GGG CTA GGT GTA GGG G**G**C CTG<br>AGT TC**T** GGG CTT TGC TAC CC<br>CCC GAT CCA CAT CCC C**C**G GAC<br>TCA AG**A** CCC GAA ACG ATG GG |
| wt probe 2 | GAG CTC TGT GGG C**C**C AGG ACG<br>CAT GGT<br>CTC GAG ACA CCC G**G**G TCC TGC<br>GTA CCA |
| 164 mut | GAG CTC TGT GGG CAC AGG ACG<br>CAT GGT<br>CTC GAG ACA CCC G**T**G TCC TGC<br>GTA CCA |
| Ets-consensus[a] | TCG AGG GGA GGA AAT GGG TGT<br>CGA<br>TCG ACA CCC ATT TCC TCC CCT<br>CGA |

[a] Ets consensus binding site for Ets transcription factor family (Laumonnier et al., 2000).

min, 37°C for 10 min, and 25°C for 5 min. The double-stranded wt probe for the −740 and −730 SNP (probe wt1) or another wt probe for −164 polymorphic site (probe wt2) was labeled with [$\gamma$-$^{32}$P]ATP by T4 polynucleotide kinase, and gel retardation assays were carried out by incubating the labeled probe with 7 $\mu$g of a nuclear extract from B16A2 cells in buffer C (20 mM HEPES, pH 7.9, 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% glycerol, 0.5 mM DTT, 0.5 mM PMSF, 10 $\mu$g/ml leupeptin) and 1 $\mu$g of poly(dI-dC) (Roche Diagnostics, Indianapolis, IN) in the presence of a 100 mM salt concentration. Unlabeled competitor oligonucleotide probe was first added at 100× molar excess concentration followed by addition of nuclear extract and incubated for 20 min at 37°C before adding labeled wt probe and incubated for another 30 min. Samples were loaded on a prerun 4% polyacrylamide gel and electrophoresis was carried out at 4°C in 0.5× Tris-borate/EDTA buffer at 180 V for 3 h. For the supershift-interference assay with antibodies, 1 $\mu$g of antibody was added to the reaction mixture and incubated further on ice for 45 min, and then applied on the gel for electrophoresis. Alternatively, the antibody was added to the reaction mixture before the addition of radiolabeled probe, incubated 15 min at room temperature according to the instruction provided. Polyclonal antibodies against Ets1, Ets1/2, Pu.1, Elk-1, GABP-$\alpha$, and monoclonal antibody against PEA3 were purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA). All gels were vacuum dried and subjected to autoradiography.

**Construction of Promoter-Luciferase Gene Plasmids.** A 1137-bp fragment 5′ upward region from the translation start site (Fig. 1) was amplified by PCR from a wt genomic DNA (*CYP1A2*1A*) using a forward primer1A2-consF (5′-ATG CAC GCG TAC CCT GAA CCC TAA AGA CAG C-3′) and 1A2-ConsR (5′-ATG CCT CGA GCT GTA CCA ACT GCA GGG AAA-3′), which contained *Mlu*I and *Xho*I restriction sites, respectively. The respective restriction sites for the enzymes are underlined (Fig. 1) and were introduced immediately upstream and downstream of the amplified sequences, respectively. The long PCR conditions were as described above under *Genomic Sequencing of the Human CYP1A2 Gene.* The amplified PCR products were purified using the Wizard PCR Preps DNA Purification kit (Promega, Madison, WI), digested with the respective restriction enzymes, and cloned into the *Mlu*I/*Xho*I site of the promoterless pGL3-Basic vector (Promega), upstream of the firefly luciferase reporter gene.

The various *CYP1A2* intron 1 haplotypes were generated by site-directed mutagensis. The *CYP1A2*1A* promoter plasmid was used as a template to introduce the −164C>A SNP to generate *CYP1A2*1F* promoter plasmid using the forward primer 5′-GGG TGA GCT CTG TGG GC**A** CAG GAC GCA TGG TAG ATG-3′, the reverse primer 5′-CAT CTA CCA TGC GTC CTG **T**GC CCA CAG AGC TCA CCC-3′, and the QuikChange Site-Directed Mutagenesis Kit (Stratagene, LA Jolla, CA) according to the manufacturer's instructions. The *CYP1A2*1F* promoter plasmid was used as template to introduce the −740T>G to generate the *CYP1A2*1J* promoter plasmid, using the forward primer 5′-TGG GCT AGG TGT AGG GG**G** CCT GAG TTC CGG GCT TTG-3′ and the reverse primer 5′-CAA AGC CCG GAA CTC AGG **C**CC CCT ACA CCT AGC CCA-3′. The *CYP1A2*1J* promoter plasmid was used as template to introduce the −730C>T and generate the *CYP1A2*1K* promoter plasmid using the forward primer 5′-GTA GGG GTC CTG AGT TC**T** GGG CTT TGC TAC CCA GC-3′ and the reverse primer 5′-GC TGG GTA GCA AAG CCC **A**GA ACT CAG GAC CCC TAC-3′. To mutate the three core nucleotides of the putative XRE intron 1 sequence, CACGC, the CYP1A2 wt plasmid, was used as a template using the forward primer 5′-GCT GGG AGC CAA GCA CAG AAC A**TC T**AT CAG TGT TTA TCA AAT GAC TG-3′ and the reverse primer, 5′-CAG TCA TTT GAT AAA CAC TGAT AGA TGT TCT GTG CTT GGC TCC CAGC-3′. We also deleted the three nucleotides from the putative XRE sequence using the forward primer 5′-GCT GGG AGC CAA GCA CAG AAC $A_{cgc}$AT CAG TGT TTA TCA AAT GAC TG-3′ and the reverse primer, 5′-CAG TCA TTT GAT AAA CAC TGA $T_{gcg}$TG TTC TGT GCT TGG CTC CCA GC-3′. The SNP intended to be introduced in the respective plasmid construct is shown in bold, and nucleotide bases to be deleted are illustrated by underscore. The *CYP1A2*1F* promoter plasmid contains −164C>A; the *CYP1A2*1J* promoter plasmid contains −164C>A and −740T>G. The *CYP1A2*1K* plasmid contains −164C>A, and −740T>G and −730C>T. All plasmids were sequenced using the ABI Prism BigDye Terminator Cycle Sequencing kit and analyzed with ABI Prism 377 DNA sequencer to ensure the correct constructs and to exclude any potential PCR artifacts.

**Cell Culture and Transfection.** All DNA plasmids used in transient transfection studies were purified using QIAGEN plasmid Maxi kits (QIAGEN, Valencia, CA). The human hepatoma cell line, B16A2, was used for transfection with the plasmid constructs. The culture conditions for B16A2 cells have been previously established (Le Jossic et al., 1996). In brief, cells were grown in 150-ml flasks to 100% confluence using Williams' medium E with Glutamax-1 supplemented with 5% fetal bovine serum, 1 $\mu$g/ml insulin, 1% penicillin streptomycin (all products from Invitrogen, Carlsbad, CA), and 0.25 $\mu$g/ml hydrocortisone (Sigma-Aldrich, St. Louis, MO). Cells were cultured in 12-well plates and, after reaching 100% confluence, were kept for 3 to 4 additional weeks by changing the medium, before

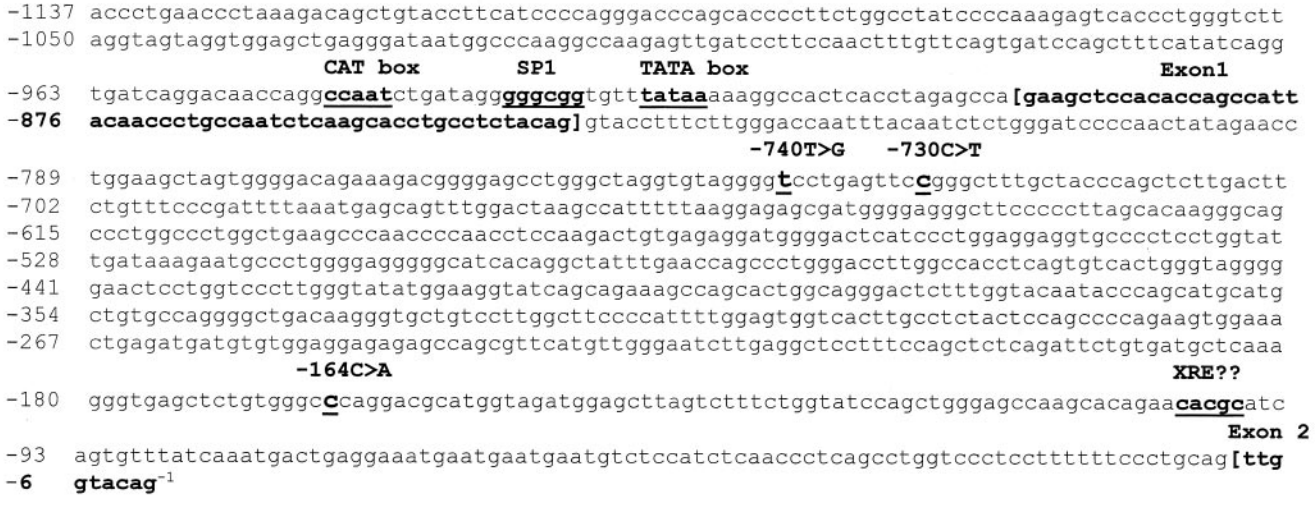


**Fig. 1.** Nucleotide sequences of −1137-bp upstream regions of *CYP1A2* inserted in pGL3 basic vector containing a luciferase gene as a reporter used to analyze the effect of intron 1 SNPs on transcriptional regulation and inducibility of the enzyme by TCDD. The CAT box, TATAA box, SP1 binding sites (Chung and Bresnick, 1995), exon 1, putative XRE, and intron 1 polymorphic sites are indicated in bold and are underlined.

being used for transfection. The transfection assay was done in quadruplicate, and each transfection was performed using 5 μg of the different plasmid constructs, cotransfected with 0.1 μg of pRL-TK plasmid containing the *Renilla reniformis* luciferase reporter gene (Promega), to provide an internal control of the transfection efficiency, and DMRIE-C as a transfection reagent (Invitrogen). The cells were incubated with the DNA-lipid complexes at 37°C for 5 h, and subsequently, 1 ml of Williams' medium E supplemented with 10% serum was added and the cells were incubated further at 37°C for 24 h. The medium was removed and replaced with 2 ml of fresh Williams' medium E supplemented with 5% serum and was kept for an additional 24 h at 37°C. Cells were rinsed with phosphate-buffered saline and lysed by treating with 1× passive lysis buffer (Promega) with gentle shaking at room temperature for 20 min. The firefly and *R. reniformis* luciferase activities were measured using the Dual Luciferase reporter assay system (Promega), and the firefly/*R. reniformis* ratio was determined.

**Statistics.** All statistical tests were carried out using the computer program STATISTICA, version 5.5 (StatSoft, Tulsa, OK). The normality of distribution was checked by Shapiro-Wilks' W test. Comparisons of caffeine log MR between Ethiopians living in Ethiopia and Ethiopians living in Sweden having the same genotype were performed using two-way ANOVA and by calculating the 95% CI for the difference. Comparisons of caffeine MR between CYP1A2 intron 1 haplotypes were carried out using Kruskal-Wallis ANOVA, median test, and Mann-Whitney *U* test. Allele and phenotype frequency differences were calculated with the two-tailed Fisher's exact test. Observed and expected genotype frequency was compared using the chi square test. Comparisons of data from the transfection assay were done using an independent *t* test, and $p < 0.05$ was considered a statistically significant difference.

## Results

**Phenotyping with Caffeine.** Caffeine was given to 100 Ethiopians living in Ethiopia and to 73 subjects living in Sweden, and the CYP1A2 activity was determined using the (AFMU + 1U + 1X + 17U + 17X)/137X) index for CYP1A2 activity. The individual caffeine MR ranged from 9.47 to 339 (log value 0.97–2.53) among smokers and 0.85 to 168 (log value −0.07 to −2.22) among nonsmokers. A histogram and probit plot for the CYP1A2 activity between the two groups of Ethiopians are given in Fig. 2A. We found no significant difference in the CYP1A2 activity between the two groups using Kruskal-Wallis ANOVA, median test, and Mann-Whitney *U* test ($p > 0.05$). This finding is evidently illustrated by a comparison of probit plots (Fig. 2B). The 95% CI for the difference between the two means was (−0.073;0.017). As shown from the log caffeine MR probit plot, CYP1A2 activity was bimodally distributed in Ethiopians, and using the antimode at 0.96, we could estimate the incidence of slow metabolizers to be 7.5%. Figure 2C illustrates the frequency distribution of log (AFMU + 1U + 1X + 17U + 17X)/137X) in smokers ($n = 20$) and nonsmokers ($n = 153$). The caffeine MR and, hence, CYP1A2 activity was significantly higher among smokers ($p < 0.0001$, independent *t* test). However, not all smokers had higher caffeine MR, indicting the presence of nonresponder phenotypes with respect to smoking habit. No gender-related difference in CYP1A2 activity was observed.

**Analysis of CYP1A2 Polymorphism in the Ethiopian Population.** The important interindividual differences in activity seen prompted us to examine new polymorphic sites in the *CYP1A2* gene, with possible influence on the constitutive as well as inducible levels of the enzyme. We sequenced the 5′-flanking region (up to −1593 bp) and all exons and introns (except for a part of intron 6) using genomic DNA from 12 different individuals having very low or very high caffeine MR. The new SNPs identified by genomic DNA sequencing are shown in Table 3. The results confirmed the presence of the previously described intron 1 SNPs, −164C>A and −740T>G. A novel intron 1 SNP, −730C>T was identified in a subject with low caffeine MR (see Table 3). Two silent mutations, 30G>A and 5347T>C, were identified in exon 2 and exon 7, respectively. Furthermore, 1589G>T and 1611G>A in intron 3; 2159A>G, 2848C>A, 3614T>C, and 6676C>G in intron 4,5,6; and noncoding regions of exon 7 were identified, respectively. The 2159 A>G in the intron 4 and 5347T>C, the last nucleotide just before the stop codon in exon 7, were the more frequent ones.

**Genotyping for CYP1A2 Intron 1 Polymorphism.** The frequencies of the −164C>A as well as the −740T>G and −730C>T mutations were analyzed in the 173 Ethiopians using real-time PCR or allele-specific PCR-RFLP. The frequencies of the −164C>A, −740T>G, and −730C>T SNPs were 60, 10, and 3%, respectively. As expected, there was no significant difference in SNP frequency between the two Ethiopian groups. No individual was found to be homozygous for the −730C>T, and only one subject was homozygous for the −740T>G. Saudi Arabians ($n = 136$) and Spaniards ($n = 117$) were genotyped for the −740T>G and −730C>T SNP, and the SNP frequencies were 9.6% and 3.7% in Saudi Arabians and 1.7% and 0.5% in Spaniards, respectively. Those subjects positive for the −740T>G and −730C>T were further analyzed for −164C>A SNP. Similar to Ethiopians, among Saudi Arabians and Spaniards, no individual was found to be homozygous for the −730C>T, and only one individual was homozygous for the −740T>G among Saudi Arabians.

**SNP Mapping and Determination of the CYP1A2 Intron 1 Haplotypes.** We considered it important to determine haplotypes of the *CYP1A2* intron 1 SNPs. All subjects being heterozygous for two or more SNP sites were further analyzed for linkage disequilibrium by a PCR-RFLP method. All individuals with −740T>G were either homozygous or heterozygous for −164C>A. Subjects heterozygous for −740T>G and −164C>A were further subjected to haplotype analysis using a SNP-specific PCR-RFLP method as described under *Materials and Methods*. First, the two chromosomes were separated with respect to the −740T>G SNP using a −740 wt or mutant specific forward primer and a common reverse primer at the end of intron 1 (Fig. 3A). Then, the PCR product amplified by the −740 wt or mutant specific primer was further subjected to −164C>A genotyping using *Apa*I RFLP separately (Fig. 3B). In these individuals, the PCR product amplified by the −740 wt primer was completely digested by *Apa*I. By contrast, the other allele, amplified by the mutant specific primer, was resistant toward *Apa*I, indicating that the −740T>G polymorphism is always in linkage disequilibrium with −164C>A on the same allele, yielding a new *CYP1A2* allele (*CYP1A2*1J*).

To study linkage distribution analysis of −730C>T and −740T>G in individuals heterozygous for the two SNPs, we first amplified a region covering the two polymorphic sites followed by simultaneous *Ava*II and *Nci*I digestion of the PCR fragment. The results showed that one of the alleles remained undigested by the two restriction enzymes, whereas the other one was completely digested, indicating

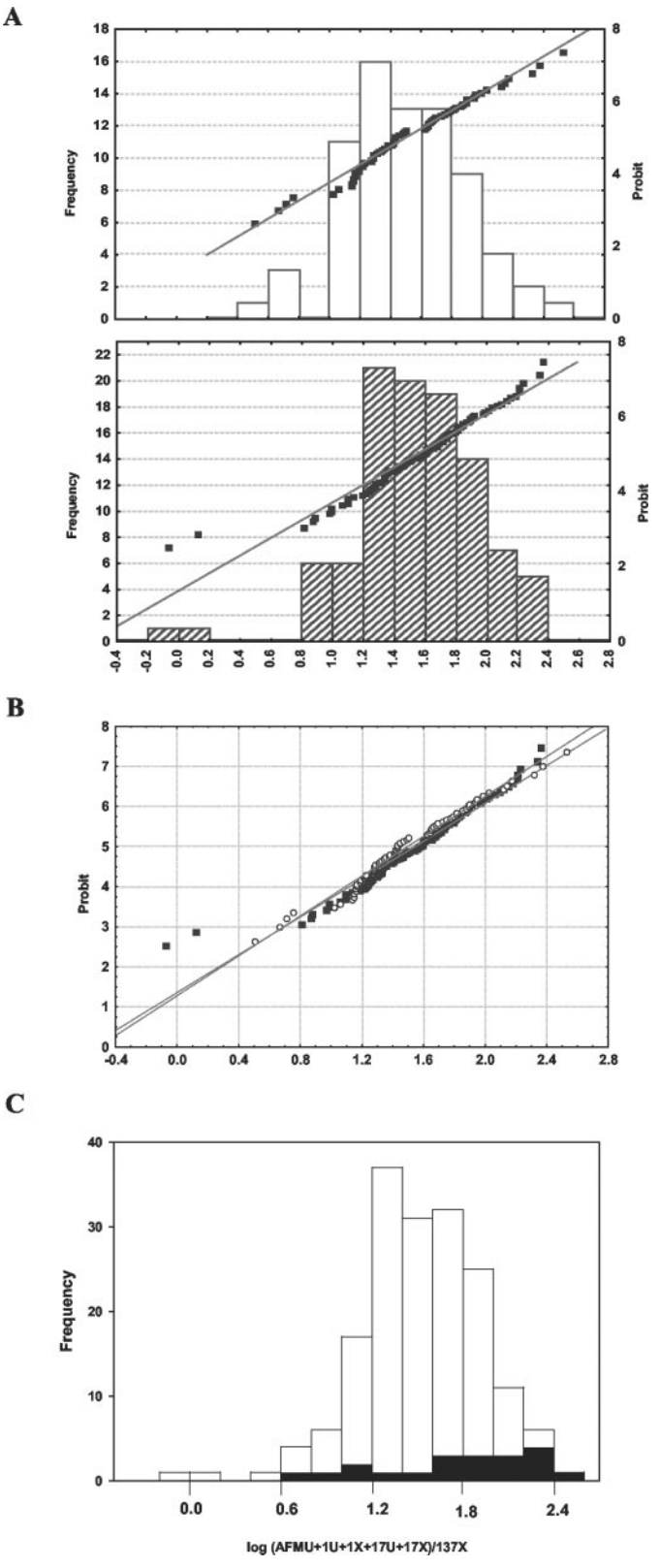


**Fig. 2.** A, comparisons of frequency distribution of CYP1A2 activity index, log (AFMU + 1U + 1X + 17U + 17X)/137X, among Ethiopians living in Sweden (upper panel, $n$ = 73) and in Ethiopia (lower panel, $n$ = 100). B, comparisons of probit plots of CYP1A2 activity index in Ethiopians living in Sweden (○) and in Ethiopia (■). C, frequency distribution of CYP1A2 activity index in smokers (black bar, $n$ = 20) and nonsmokers (white bar, $n$ = 153).

that −730C>T is always linked with −740T>G. Since the −740T>G is in linkage disequilibrium with −164C>A (see above), the −730C>T polymorphism always exists in combination with −740C>T and −164C>A, giving a new intron 1 *CYP1A2* haplotype (*CYP1A2*1K*). Likewise, in Saudi Arabians and Spaniards, the −730C>T was linked to the −740T>G and the −164C>A, and the −740T>G was linked with the −164C>A. Thus the frequency of *CYP1A2*1J* and *CYP1A2*1K* among Saudi Arabians was 5.9% and 3.6% and among Spaniards, 1.3% and 0.5%, respectively.

With the three SNPs in intron 1 of the *CYP1A2* gene, theoretically, $2^3 = 8$ possible different haplotypes could exist, whereas our haplotype analysis revealed that only four were present. The haplotypes and their frequencies are listed in Table 4. Eight different combinations of these four haplotypes were present among the individuals examined, of which the **1A/*1F* combination was found to be the most frequent one (Table 5). The observed genotype distribution pattern was consistent with that expected according to the Hardy-Weinberg law using chi square test ($p > 0.05$).

**Influence of the Haplotypes on Caffeine Metabolism.** We examined the in vivo effects of the four haplotypes from the caffeine phenotype analysis among smokers and nonsmokers (see Table 5). There was no significant difference in caffeine MR among subjects carrying *CYP1A2*1A*, *CYP1A2*1F* (−164 C>A), and *CYP1A2*1J* (−164C>A and −740T>G) haplotype, indicating that −164C>A alone or in combination with −740T>G does not influence the CYP1A2 activity significantly in vivo. No significant difference was observed between subjects with *CYP1A2*1A/*1A* and *CYP1A2*1F/*F* among smokers or nonsmokers. However, nonsmoker subjects heterozygous for *CYP1A2*1K* (−164C>A, −740T>G, and −730C>T) haplotype had significantly lower CYP1A2 activity compared with subjects with the *CYP1A2*1A*, *CYP1A2*1F*, or *CYP1A2*1J* haplotype ($p < 0.02$, two-way Kruskal-Wallis ANOVA). There was only one individual who was heterozygous for *CYP1A2*1K* among smokers, and he had a lower caffeine MR in this group. The only additional SNP in *CYP1A2*1K*, as compared with the others, is −730C>T, and, thus, the presence of this SNP is apparently critical for a decreased CYP1A2 activity in vivo.

**EMSA Analysis.** The results from the in vivo analysis indicating a functional role of the −730C>T mutation prompted us to perform in vitro experiments analyzing the protein binding and the in vitro transcription activity of this allele compared with the others. Computer analysis of the TRANSFAC database, using the TFSEARCH program (with threshold set at 85.0), for intron 1 polymorphic sites to identify possible transcription factor binding sites indicated the −730 region as core binding sites for Ets transcription factors on the antisense strand. By contrast, no specific transcription factor was identified whose binding could be affected by the presence of −164C>A.

EMSAs were performed using a 41-bp oligonucleotide probe (probe 1) spanning the −740 and −730 SNP site (Fig. 4). Two DNA-protein complexes were observed (Fig. 4, lane 2), one of which corresponded to a specific binding of nuclear proteins as revealed by efficient competition in the presence of 100 M excess cold wt probe (Fig. 4, lane 3) in the reaction, but not by a nonspecific oligonucleotide probe (Fig. 4, lane 7) or consensus sequence for octamer 1 (Fig. 4, lane 8), demonstrating binding specificity. The effect of the two SNP sites

TABLE 3

List of SNPs identified by genomic DNA sequencing from 12 Ethiopians phenotyped using caffeine and with high or low caffeine MR

Numbering is from the translation site as recommended by the cytochromes P450 nomenclature committee. Smoking habit is designated as S and NS, referring to smokers and non-smokers respectively.

| −164 C>A | −730C>T | −740T>G | 30G>A (Silent) exon 2 | 1589G>T intron 3 | 1611 G>A intron 3 | 2159 A>G intron 4 | 2848 C>A intron 5 | 3614 T>C intron 6 | 5347 T>C (Silent) exon 7 | 6676C>G exon 7 Noncoding | Caffeine MR | Caffeine Log MR | Smoking Habit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C/C | | | | | | G/G | | | | | 20 | 1.3 | S |
| C/A | | | | | | A/G | | | C/C | | 12 | 1.1 | S |
| A/A | | T/G | | G/T | | A/G | C/A | T/C | C/C | C/G | 21 | 1.3 | S |
| C/A | | | G/A | | | A/G | | | | | 1 | 0 | NS |
| A/A | | | | | | | | | | | 3 | 0.5 | NS |
| C/C | | | | | | G/G | | | | | 5 | 0.7 | NS |
| C/A | | | | | | A/G | | | C/C | | 6 | 0.8 | NS |
| A/A | C/T | T/G | | | | A/G | | T/C | T/C | C/G | 8 | 0.9 | NS |
| C/A | | | | | | A/G | | | T/C | | 10 | 1.0 | NS |
| A/A | | | | | | A/G | | | T/C | | 230 | 2.4 | S |
| A/A | | | | | | | | | | | 218 | 2.3 | S |
| A/A | | T/G | | | G/A | | | | | | 11 | 1.0 | NS |

was analyzed by using unlabeled mutant oligonucleotides as competitors in 100-fold molar excess. The formation of protein-DNA complex was not affected by a probe carrying the −740T>G SNP, since a 100-fold molar excess of the mutated probe competed efficiently with the wt probe for the binding of nuclear protein (Fig. 4, lane 4). However, a probe carrying the −730C>T SNP alone (Fig. 4, lane 5) or in combination with the −740T>G SNP (Fig. 4, lane 6) failed to inhibit binding of the nuclear protein and did not compete with the wt probe for protein binding. This indicates that the −730C>T SNP but not the −740T>G abolishes the protein binding site and might be of importance for control of gene transcription. A probe carrying a consensus binding site for Ets transcription factor family competed with the wt oligonucleotide probe for the protein binding (Fig. 4, lane 9), indicating that the protein binding to the polymorphic site is a member of the Ets family of transcription factors.

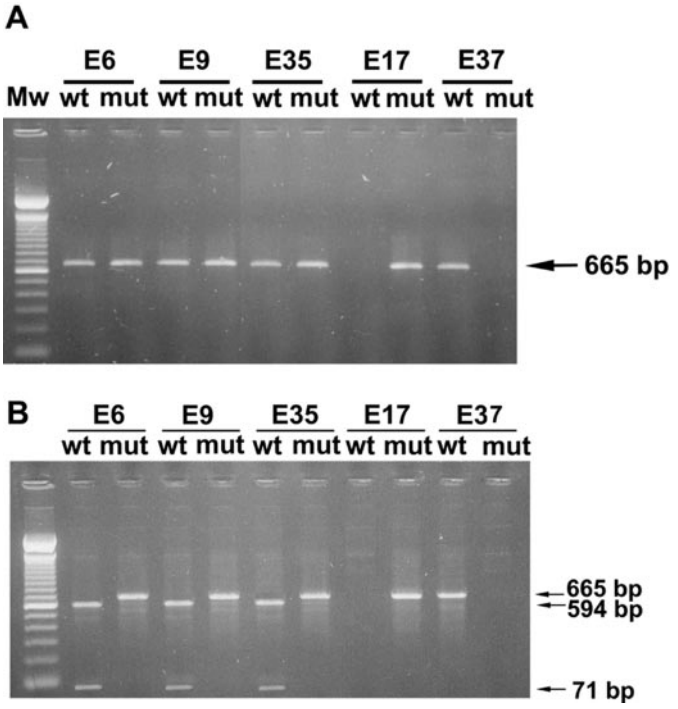


**Fig. 3.** Linkage analysis of −740T>G and −164C>A. A, alleles were separated with respect to the presence of −740T>G by allele-specific forward primers 1A2-740Fwt or 1A2-740Fmt and a common reverse primer, 1A2-164R. The 100-bp plus DNA ladder (Invitrogene) was used as a marker. Individual E17, homozygous mutated for both SNPs, and E37, homozygous wt for −740T>G and homozygous mutated for −164C>A, were used as controls. E9, E10, and E40 were heterozygous for both SNPs. B, RFLP analysis of PCR fragments from A for the presence of −164C>A. Fragments from PCR II (A) were subjected to *Apa*I digestion. The −164C>A mutation eliminates the *Apa*I restriction site. All PCR product amplified with the 1A2 −740wt primer from individuals heterozygous for both SNPs was completely digested, whereas the PCR product amplified with the 1A2 −740mut was not, indicating that the −740T>G is linked with −164C>A on the same allele.

To identify which member of the family could be involved in the formation of DNA-protein complex, we performed a supershift or interference assay using poyclonal or monoclonal antibodies specific for Ets1, Ets1/Ets2, PU.1, PEA3, Elk-1, and GABP-$\alpha$. However, the antibodies failed to modify the nucleoprotein complex formed with the wt probe (data not shown). Thus, the DNA-protein complex is likely to result from binding of another Ets family member. Currently, the Ets family contains more than 30 different members, and the lack of available antibodies specific for Ets family members in general restricts us from further attempts to identify the transcription factor responsible. There was no specific DNA-protein complex formation that could be seen using probe 2, indicating no specific protein binding to the area around the −164 SNP (data not shown).

**Cell Transfection Experiments.** Transfection studies were performed to determine the effect of intron 1 SNPs on the transcriptional regulation of *CYP1A2* using reporter gene assays. A 1137-bp upward fragment from the translation site of *CYP1A2* (Fig. 1) was amplified by PCR and introduced into the pGL3 basic vector. The different intron 1 SNPs were introduced by site-directed mutagenesis. The inserted sequence contains a TATA box, a CCAAT box, and a putative responsive element for SP1 (Chung and Bresnick, 1995), but protein binding to these potential sites has not been demonstrated. The effects of the different SNPs on transcriptional activity and enzyme inducibility were investigated by treating the B16A2 cells with vehicle or with 10 nM TCDD, respectively (see Fig. 5). There was no significant difference among the *CYP1A2*1A* wt, *CYP1A2*1F* (containing −164C>A), *CYP1A2*1J* (containing −164C>A and −740T>G), and *CYP1A2*1K* plasmids (containing −164C>A, −740T>G, and −730C>T) on the constitutive transcriptional activity ($p > 0.05$, using independent $t$ test). But upon TCDD treatment, the luciferase activity was increased more than 2-fold when cells were transfected with the *CYP1A2*1A*, *CYP1A2*1F*, or *CYP1A2*1J* plasmids, and there was no significant difference in induction among these three

TABLE 4
*CYP1A2* intron 1 haplotypes and their frequency in Ethiopians, Saudi Arabians, and Spaniards.

| Allele | Nucleotide Changes | Allele Frequency | | |
|---|---|---|---|---|
| | | Ethiopians ($n$ = 173) | Saudi Arabians ($n$ = 136) | Spaniards ($n$ = 117) |
| *CYP1A2*1A* | None | 0.399 (95%CI,0.33–0.47) | N.D. | N.D. |
| *CYP1A2*1F* | −164 C>A | 0.496 (95%CI,0.42–0.57) | N.D. | N.D. |
| *CYP1A2*1J*[a] | −164C>A;−740 T>G | 0.075 (95%CI,0.04–0.11) | 0.059 (95%CI,0.02–0.10) | 0.013 (95%CI,0.0–0.10) |
| *CYP1A2*1K*[a] | −164C>A;−740T>G;−730C>T | 0.03 (95%CI,0.00–0.06) | 0.036 (95%CI,0.00–0.07) | 0.005 (95%CI,0.0–0.07) |

N.D., not determined.
[a] Newly identified *CYP1A2* alleles.

haplotypes. However, there was 40% less induction of cells transfected with *CYP1A2*1K*, and this difference was significant compared with cells treated with *CYP1A2*1A*, **1F*, or **1J* ($p < 0.01$ using independent $t$ test), indicating that the *CYP1A2*1K* haplotype is less inducible.

Since we observed induction with TCDD, we searched for a possible XRE sequence in our *CYP1A2*1A* wt construct. XRE contains an invariant CACGC core sequence that is recognized by the aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator heterodimer, and is known to be present in most, if not all, 3-methylcholanthrene-inducible genes (Sogawa et al., 1995). This sequence is present in the CYP1A2 intron 1 further downstream of the −164C>A polymorphic site (see Fig. 1). Transient transfection analyses of B16A2 cells with reporter genes that contain deletions or triple mutations of the core putative XRE site of CYP1A2 intron 1 did not alter the basal expression or inducibility with TCDD compared with the wt plasmid (data not shown). Further work is required to characterize the XRE sites responsible for the observed induction with TCDD.

## Discussion

The results indicate that polymorphism in intron 1 of the *CYP1A2* gene might be critical for CYP1A2 inducibility and that transcriptional factors of the Ets family are of importance in this respect. By contrast, the data do not provide evidence for an important influence of environmental factors different between Sweden and Ethiopia for the control of CYP1A2 expression, in contrast to what has been seen for CYP2D6 (Aklillu et al., 2002a). Three different SNPs were found in intron 1, and we identified four haplotypes present in Ethiopians, of which two are novel, *CYP1A2*1J* and *CYP1A2*1K*. Subjects with the *CYP1A2*1K* allele had significantly reduced CYP1A2 activity as compared with those carrying *CYP1A2*1A, CYP1A2*1F*, or *CYP1A2*1J*. Thus, the −164C>A alone (*CYP1A2*1F*) or in combination with −740T>G (*CYP1A2*1J*) has apparently no influence on CYP1A2 activity in vivo, in accordance with other studies showing no significance of the −164 SNP on CYP1A2 activity (Schulze et al., 2001; Nordmark et al., 2002; Shimoda et al., 2002).

Our finding was further supported by in vitro analysis using EMSA and reporter gene assay. In EMSA, analysis of protein binding to an intron 1 element was completely dependent on the wt C at position −730, whereas oligonucleotides with −730T were ineffective in competing for protein binding. No influence of the −740T>G or −164C>A SNPs was found. The −730 SNP site was indicated as a putative binding site in the antisense strand for Ets family transcription factors; indeed, we observed a clear competition using a 100-fold molar excess of consensus Ets transcription factor family oligonucleotide. The −730C>T affects the core-binding site for Ets transcription factor. Despite evaluating the effect of polyclonal or monoclonal antibodies against Ets1, Ets1/Ets2, PU.1, PEA3, Elk-1, and GABP-$\alpha$ for protein binding, the isoform of the 30 different known members of the Ets family could not be identified. Hence, further work is required to specifically identify the Ets protein that binds to the functional −730 polymorphic site.

Members of the Ets family have been shown to be involved in the regulation of rat *cyp24* (Dwivedi et al., 2000), *cyp27* (Mullick et al., 2001), *cyp2d-9* (Yokomori et al., 1995), and human cytochrome *c* oxidase gene expression. The present study represents the first report in which Ets transcription factors have been suggested to be involved in the regulation of *CYP1A2*. The Ets family members are important transcription factors involved in development and are implicated in several types of cancer and other human diseases (Dittmer and Nordheim, 1998). Members of the PEA3 group of Ets transcription factors are commonly expressed, and their expression is associated with invasion capacity of human cancerous mammary epithelial cells. Prostate-derived Ets factor is over-expressed in human breast tumors and has been suggested to be a candidate tumor marker and breast tumor antigen (Ghadersohi and Sood, 2001). Ets1 activates transcription of genes that have crucial roles in hepatocarcinogenesis and progression, and its over-expression is suggested to contribute to the invasion and metastasis of hepatocellular carcinoma (Ozaki et al., 2000).

It appears that there is interracial difference in the frequency of *CYP1A2* intron 1 SNPs and haplotypes. There was no significant difference between Ethiopian (60%) and white persons, 68% in Germans (Sachse et al., 1999); or Asian, 61% in Japanese (Chida et al., 1999) populations in the allele frequency of the −164C>A SNP. However, the −740T>G SNP is frequent in Ethiopians (10%), Saudi Arabians (9.6%), and Japanese (8%) (Chida et al., 1999) but is significantly lower in Spaniards (1.7%). In the present study we found the novel intron 1 SNP, −730C>T, and two haplotypes, *CYP1A2*1J* and *CYP1A2*1K*, in Ethiopians and further investigated the presence of these haplotypes in Saudi Arabians and Spaniards. The allele frequency of *CYP1A2*1J* and *CYP1A2*1K* is similar between Ethiopians and Saudi Arabians, but the frequency is much lower in Spaniards. Likewise, Ethiopians and Saudi Arabians show resemblance in having higher frequency of *CYP2D6* gene duplication and *CYP1B1* haplotype frequency (McLellan et al., 1997; Aklillu et al., 2002b). In an attempt to study population history and natural selection in shaping genetic diversity in CYP1A2, Wooding et al. (2002) sequenced a 3.7-kb 5′ region of *CYP1A2*, not including intron 1, in 113 individuals from Africa, Asia, and

TABLE 5

Intron 1 SNPs and genotype combinations, showing their observed frequencies, expected frequencies, the respective median caffeine MR, and mean log caffeine MR ± S.D. among smokers and nonsmokers

| Smoking Habit | −164 C>A | −740T>G | −730C>T | Genotype | Number of Subjects | Total Observed Frequency[a] | Total Expected Frequency[b] | Median MR | Median log MR | Mean Log MR ± S.D. | 95% CI of Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Nonsmokers | C/A | T/T | C/C | *1A/*1F | 62 | 69 | 68.5 | 33 | 1.52 | 1.50 ± 0.41 | 1.39–1.61 |
| | A/A | T/T | C/C | *1F/*1F | 38 | 42 | 42.6 | 40 | 1.60 | 1.52 ± 0.30 | 1.42–1.61 |
| | C/C | T/T | C/C | *1A/*1A | 21 | 27 | 27.5 | 32 | 1.51 | 1.50 ± 0.31 | 1.37–1.60 |
| | A/A | T/G | C/C | *1F/*1J | 12 | 14 | 12.9 | 47 | 1.67 | 1.57 ± 0.30 | 1.37–1.76 |
| | C/A | T/G | C/C | *1A/*1J | 11 | 11 | 10.4 | 42 | 1.62 | 1.66 ± 0.29 | 1.47–1.90 |
| | A/A | T/G | C/T | *1F/*1K | 5 | 5 | 5.1 | 18* | 1.26* | 1.20 ± 0.24 | 0.82–1.48 |
| | C/A | T/G | C/T | *1A/*1K | 4 | 4 | 4.1 | 21* | 1.32* | 1.34 ± 0.15 | 0.86–1.52 |
| Smokers | C/C | T/T | C/C | *1A/*1A | 6 | | | 116 | 2.06 | 1.91 ± 0.54 | 1.33–2.49 |
| | C/A | T/T | C/C | *1A/*1F | 7 | | | 82 | 1.91 | 1.84 ± 0.44 | 1.38–2.30 |
| | A/A | T/T | C/C | *1F/*1F | 4 | | | 135 | 2.13 | 1.85 ± 0.64 | 0.81–2.88 |
| | A/A | T/G | C/C | *1F/*1J | 2 | | | 73 | 1.86 | 1.71 ± 0.54 | −0.18 to −5.0 |
| | A/A | G/G | C/T | *1J/*1K | 1 | 1 | 0.8 | 22 | 1.34 | 1.35 | |

[a, b] Total observed and expected frequency of subjects having the respective genotype regardless of smoking habit. The observed genotype distribution pattern was consistent with the expected frequencies according to the Hardy-Weinberg equilibrium ($P > 0.05$, $\chi^2$ test).

* $p < 0.02$, Kruskal-Wallis ANOVA, median test. Nonsmoker subjects with −730 SNP, carrying the *CYP1A2*1K* haplotype, had significantly lower log caffeine MR value compared to subjects with *CYP1A2*1A*, **1F*, or **1J*.

Europe and reported that the African population had the highest level of nucleotide diversity and the lowest level of linkage disequilibrium. Of the 17 haplotypes found in this study, 12 were found in the African sample, 8 were found in Indians, 5 were found in non-Indian Asians, and 5 were found in Europeans. Haplotypes found outside Africa were mostly a subset of those found within Africa (Wooding et al., 2002).

In the present study we investigated a possible influence of environmental factors different between Sweden and Ethiopia, such as diet and infections, by examining the CYP1A2 activity among Ethiopians living in Sweden or in Ethiopia. The dietary differences between Ethiopia and Sweden as well as plant constituents specifically used in Ethiopia for dietary or nondietary purposes has previously been discussed (Aklillu et al., 2002a), in a study aimed at investigating environmental differences on CYP2D6 and CYP2C19 enzyme activity. In contrast to the important differences we found for CYP2D6, there was no significant difference between the two populations of Ethiopians for caffeine metabolism, as was also registered for CYP2C19. Thus, we assume that environmental factors such as dietary habits have small effects on CYP1A2 activity and cannot primarily explain interethnic differences in activity. This is in accordance with a recent study on twins, which indicated that the CYP1A2 activity is mainly governed by genetic factors (Rasmussen et al., 2002).

Smoking induces CYP1A2 activity and thus significantly lowers the plasma concentration of antipsychotic drugs including haloperidol. Many schizophrenic patients are smokers. Among smokers, CYP1A2 MR shows a trend toward a bimodal distribution (i.e., with the existence of a "nonresponder" phenotype concerning CYP1A2 induction by compounds present in tobacco smoke) in both white persons and Asians (Nakajima et al., 1994; Schrenk et al., 1998). The −164C>A SNP has been suggested to be associated with higher enzyme inducibility in white smokers (Sachse et al., 1999) and the −3858G>A (*CYP1A2*1C*) with decreased enzyme inducibility in Japanese smokers (Nakajima et al., 1999). Several studies have been conducted to investigate the importance of the −164C>A polymorphism. We did not observe significant difference in enzyme activity between *1A/*1A and *1F/*1F in either smokers or nonsmokers. In accordance with our results, a lack of impact of the −164 SNP genotypes on the plasma concentration of haloperidol in smoking male Japanese with schizophrenia (Shimoda et al., 2002) and in smoking Swedish pregnant women (Nordmark et al., 2002) has recently been reported. A study in U.S. schizophrenic patients showed subjects with −164 C/C genotype to be at increased risk to develop more severe tardive dyskinesia than the A/C or A/A genotype (Basile et al., 2000), whereas another study on schizophrenic smoking patients in a German population reported lack of association between the −164 C>A genotypes and severity of tardive dyskinesia (Schulze et al., 2001). A possible explanation for the discrepant findings could be incomplete determination of the different haplotypes existing in the populations. Thus, as shown here, the −164 SNP can be located in at least four different haplotypes having different functional impact, and it is thus important to take the complete haplotypes into consideration when investigating associations of phenotype rather than focusing on single SNPs.

Although CYP1A2 is only involved in the metabolism of

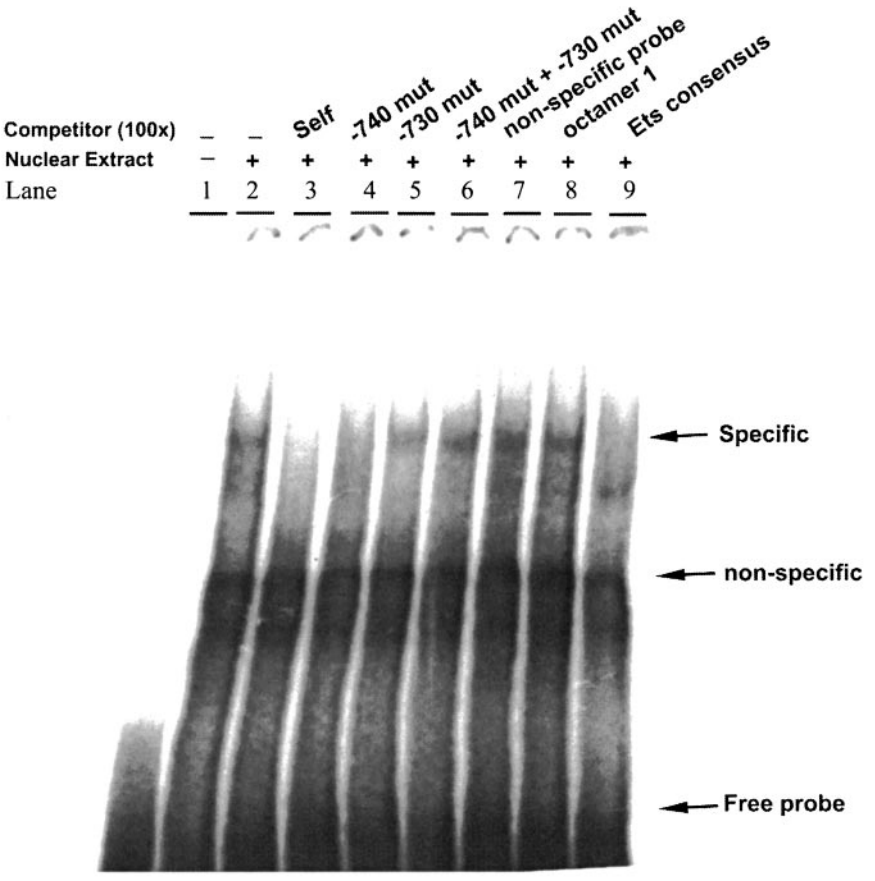


**Fig. 4.** EMSA of B16A2 hepatoma cell line nuclear protein binding to the 41-bp wt oligonucleotide probe 1 spanning the −730 and −740 regions of 1A2 intron 1. Radiolabeled oligonucleotide probe 1 wt was incubated without nuclear protein (lane 1) or with 7 $\mu$g of nuclear extracts from the B16A2 cells (lane 2). Competition assay was performed in the presence of 100 M excess of the respective unlabeled oligonucleotide as indicated. DNA protein complexes are indicated with arrows. Two DNA-protein complexes were observed (lane 2), one of which corresponded to a specific binding of nuclear proteins as revealed by efficient competition in the presence of 100 M excess cold wt probe (lane 3) in the reaction, but not by a nonspecific oligonucleotide probe (lane 7) or consensus sequence for octamer 1 (lane 8), demonstrating binding specificity. The formation of protein-DNA complex was not affected by a probe carrying the −740T>G SNP since a 100-fold molar excess of the mutated probe competed efficiently with the wt probe for the binding of nuclear protein (lane 4, −740 mut). However, a probe carrying the −730 C>T SNP alone (lane 5, −730 mut) or in combination with the −740T>G SNP (lane 6, −740 mut + −730 mut) failed to inhibit binding of the nuclear protein and did not compete with the wt probe for protein binding. Complexes were resolved by electrophoreses in 4% polyacrylamide, 0.5× Tris borate-EDTA gel. Sequences of oligonucleotides used are listed in Table 2.

about 5% of commonly prescribed drugs, it apparently participates in the metabolism of 75% of drugs associated with adverse drug reactions metabolized by enzymes having variant alleles (Phillips et al., 2001). Interindividual differences in its activity might thus be of substantial importance for the determination of the outcome of drug treatment, and knowledge about the basis for such interindividual differences, both genetic and environmental, might be useful to avoid adverse drug reactions. In combination with the well established role of CYP1A2 for the metabolic activation of procarcinogens, the polymorphism here described in intron 1 might, thus, also be of critical importance for determination of the individual's susceptibility to liver cancer risk following long-term exposure to several dietary procarcinogens.

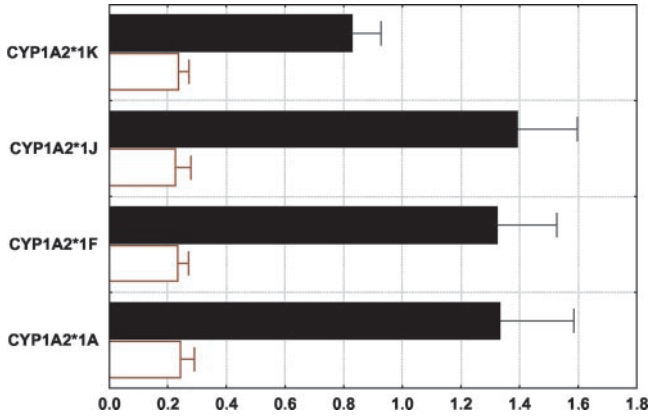


**Fig. 5.** Functional characterization of the CYP1A2 intron 1 haplotypes on constitutive expression (white bars) and enzyme inducibility (black bars) using a reporter gene assay. A series of *CYP1A2* reporter constructs was made that contained the different intron 1 haplotypes (see Fig. 1 for sequence of the insert). These constructs were transfected into B16A2 cells, and luciferase activities were assayed 48 h after transfection with or without TCDD. The *CYP1A2*1A* is wt, *CYP1A2*1F* contains −164C>A, *CYP1A2*1J* contains −164C>A and −740T>G, and the *CYP1A2*1K* contains −164C>A, −740T>G, and −730C>T. The relative activity refers to the ratio of normalized luciferase activity of each construct to that of the promoterless pGL3-Basic plasmid. Values are the average of three independent experiments + S.D., performed in quadruplicate. There was no significant difference in constitutive expression between the different intron 1 haplotypes. However, upon TCDD treatment, cells transfected with *CYP1A2*1K* plasmid had significantly lower luciferase activity compared with other intron 1 haplotypes ($p < 0.01$, independent $t$ test)


### Acknowledgments

We owe special thanks to the Ethiopian volunteers for participating in the study.


### References


Aklillu E, Herrlin K, Gustafsson LL, Bertilsson L, and Ingelman-Sundberg M (2002a) Evidence for environmental influence on CYP2D6-catalysed debrisoquine hydroxylation as demonstrated by phenotyping and genotyping of Ethiopians living in Ethiopia or in Sweden. *Pharmacogenetics* **12:**375–383.

Aklillu E, Oscarson M, Hidestrand M, Leidvik B, Otter C, and Ingelman-Sundberg M (2002b) Functional analysis of six different polymorphic CYP1B1 enzyme variants found in an Ethiopian population. *Mol Pharmacol* **61:**586–594.

Basile VS, Ozdemir V, Masellis M, Walker ML, Meltzer HY, Lieberman JA, Potkin SG, Alva G, Kalow W, Macciardi FM, and Kennedy JL (2000) A functional polymorphism of the cytochrome P450 1A2 (CYP1A2) gene: association with tardive dyskinesia in schizophrenia. *Mol Psychiatry* **5:**410–417.

Bertilsson L, Carrillo JA, Dahl ML, Llerena A, Alm C, Bondesson U, Lindstrom L, Rodriguez de la Rubia I, Ramos S, and Benitez J (1994) Clozapine disposition covaries with CYP1A2 activity determined by a caffeine test. *Br J Clin Pharmacol* **38:**471–473.

Carrillo JA, Christensen M, Ramos SI, Alm C, Dahl ML, Benitez J, and Bertilsson L (2000) Evaluation of caffeine as an in vivo probe for CYP1A2 using measurements in plasma, saliva and urine. *Ther Drug Monit* **22:**409–417.

Chida M, Yokoi T, Fukui T, Kinoshita M, Yokota J, and Kamataki T (1999) Detection

of three genetic polymorphisms in the 5′-flanking region and intron 1 of human CYP1A2 in the Japanese population. *Jpn J Cancer Res* **90:**899–902.
Chung I and Bresnick E (1995) Regulation of the constitutive expression of the human CYP1A2 gene: cis elements and their interactions with proteins. *Mol Pharmacol* **47:**677–685.
Chung I and Bresnick E (1997) Identification of positive and negative regulatory elements of the human cytochrome P4501A2 (CYP1A2) gene. *Arch Biochem Biophys* **338:**220–226.
Dittmer J and Nordheim A (1998) Ets transcription factors and human disease. *Biochim Biophys Acta* **1377:**F1–F11.
Dwivedi PP, Omdahl JL, Kola I, Hume DA, and May BK (2000) Regulation of rat cytochrome P450C24 (CYP24) gene expression. Evidence for functional cooperation of Ras-activated Ets transcription factors with the vitamin D receptor in 1,25-dihydroxyvitamin D(3)-mediated induction. *J Biol Chem* **275:**47–55.
Eaton DL, Gallagher EP, Bammler TK, and Kunze KL (1995) Role of cytochrome P4501A2 in chemical carcinogenesis: implications for human variability in expression and enzyme activity. *Pharmacogenetics* **5:**259–274.
Gallagher EP, Kunze KL, Stapleton PL, and Eaton DL (1996) The kinetics of aflatoxin B1 oxidation by human cDNA-expressed and human liver microsomal cytochromes P450 1A2 and 3A4. *Toxicol Appl Pharmacol* **141:**595–606.
Ghadersohi A and Sood AK (2001) Prostate epithelium-derived Ets transcription factor mRNA is over-expressed in human breast tumors and is a candidate breast tumor marker and a breast tumor antigen. *Clin Cancer Res* **7:**2731–2738.
Ikeya K, Jaiswal AK, Owens RA, Jones JE, Nebert DW, and Kimura S (1989) Human CYP1A2: sequence, gene structure, comparison with the mouse and rat orthologous gene and differences in liver 1A2 mRNA expression. *Mol Endocrinol* **3:**1399–1408.
Laumonnier Y, Nadaud S, Agrapart M, and Soubrier F (2000) Characterization of an upstream enhancer region in the promoter of the human endothelial nitric-oxide synthase gene. *J Biol Chem* **275:**40732–40741.
Le Jossic C, Glaise D, Corcos L, Diot C, Dezier JF, Fautrel A, and Guguen-Guillouzo C (1996) trans-Acting factors, detoxication enzymes and hepatitis B virus replication in a novel set of human hepatoma cell lines. *Eur J Biochem* **238:**400–409.
McLellan RA, Oscarson M, Seidegard J, Evans DA, and Ingelman-Sundberg M (1997) Frequent occurrence of CYP2D6 gene duplication in Saudi Arabians. *Pharmacogenetics* **7:**187–191.
McManus ME, Burgess WM, Veronese ME, Huggett A, Quattrochi LC, and Tukey RH (1990) Metabolism of 2-acetylaminofluorene and benzo(a)pyrene and activation of food-derived heterocyclic amine mutagens by human cytochromes P-450. *Cancer Res* **50:**3367–3376.
Minchin RF, McManus ME, Boobis AR, Davies DS, and Thorgeirsson SS (1985) Polymorphic metabolism of the carcinogen 2-acetylaminofluorene in human liver microsomes. *Carcinogenesis* **6:**1721–1724.
Mullick J, Anandatheerthavarada HK, Amuthan G, Bhagwat SV, Biswas G, Camasamudram V, Bhat NK, Reddy SE, Rao V, and Avadhani NG (2001) Physical interaction and functional synergy between glucocorticoid receptor and Ets2 proteins for transcription activation of the rat cytochrome P-450c27 promoter. *J Biol Chem* **276:**18007–18017.
Nakajima M, Yokoi T, Mizutani M, Kinoshita M, Funayama M, and Kamataki T (1999) Genetic polymorphism in the 5′-flanking region of human CYP1A2 gene: effect on the CYP1A2 inducibility in humans. *J Biochem (Tokyo)* **125:**803–808.
Nakajima M, Yokoi T, Mizutani M, Shin S, Kadlubar FF, and Kamataki T (1994) Phenotyping of CYP1A2 in Japanese population by analysis of caffeine urinary metabolites: absence of mutation prescribing the phenotype in the CYP1A2 gene. *Cancer Epidemiol Biomarkers Prev* **3:**413–421.
Nordmark A, Lundgren S, Ask B, Granath F, and Rane A (2002) The effect of the CYP1A2 *1F mutation on CYP1A2 inducibility in pregnant women. *Br J Clin Pharmacol* **54:**504–510.
Oscarson M, Gullsten H, Rautio A, Bernal ML, Sinues B, Dahl ML, Stengard JH, Pelkonen O, Raunio H, and Ingelman-Sundberg M (1998) Genotyping of human cytochrome P450 2A6 (CYP2A6), a nicotine C-oxidase. *FEBS Lett* **438:**201–205.
Ozaki I, Mizuta T, Zhao G, Yotsumoto H, Hara T, Kajihara S, Hisatomi A, Sakai T, and Yamamoto K (2000) Involvement of the Ets-1 gene in overexpression of matrilysin in human hepatocellular carcinoma. *Cancer Res* **60:**6519–6525.
Phillips KA, Veenstra DL, Oren E, Lee JK, and Sadee W (2001) Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. *JAMA (J Am Med Assoc)* **286:**2270–2279.
Quattrochi LC, Shih H, and Pickwell GV (1998) Induction of the human CYP1A2 enhancer by phorbol ester. *Arch Biochem Biophys* **350:**41–48.
Quattrochi LC, Vu T, and Tukey RH (1994) The human CYP1A2 gene and induction by 3-methylcholanthrene. A region of DNA that supports AH-receptor binding and promoter-specific induction. *J Biol Chem* **269:**6949–6954.
Rasmussen BB, Brix TH, Kyvik KO, and Brosen K (2002) The interindividual differences in the 3-demethylation of caffeine alias CYP1A2 is determined by both genetic and environmental factors. *Pharmacogenetics* **12:**473–478.
Sachse C, Brockmoller J, Bauer S, and Roots I (1999) Functional significance of a C→A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br J Clin Pharmacol* **47:**445–449.
Schmidt JV, Su GH, Reddy JK, Simon MC, and Bradfield CA (1996) Characterization of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development. *Proc Natl Acad Sci USA* **93:**6731–6736.
Schrenk D, Brockmeier D, Morike K, Bock KW, and Eichelbaum M (1998) A distribution study of CYP1A2 phenotypes among smokers and non-smokers in a cohort of healthy Caucasian volunteers. *Eur J Clin Pharmacol* **53:**361–367.
Schulze TG, Schumacher J, Muller DJ, Krauss H, Alfter D, Maroldt A, Ahle G, Maroldt AO, Novo y Fernandez A, Weber T, et al. (2001) Lack of association between a functional polymorphism of the cytochrome P450 1A2 (CYP1A2) gene and tardive dyskinesia in schizophrenia. *Am J Med Genet* **105:**498–501.
Shimoda K, Someya T, Morita S, Hirokane G, Yokono A, Takahashi S, and Okawa M (2002) Lack of impact of CYP1A2 genetic polymorphism (C/A polymorphism at position 734 in intron 1 and G/A polymorphism at position −2964 in the 5′-flanking region of CYP1A2) on the plasma concentration of haloperidol in smoking male Japanese with schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* **26:**261–265.
Sofowora GG, Choo EF, Mayo G, Shyr Y, and Wilkinson GR (2001) In vivo inhibition of human CYP1A2 activity by oltipraz. *Cancer Chemother Pharmacol* **47:**505–510.
Sogawa K, Nakano R, Kobayashi A, Kikuchi Y, Ohe N, Matsushita N, and Fujii-Kuriyama Y (1995) Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation. *Proc Natl Acad Sci USA* **92:**1936–1940.
Uwaifo AO and Bababunmi EA (1984) Liver carcinogenesis in tropical Africa. *IARC Sci Publ* **63:**59–88.
Wooding SP, Watkins WS, Bamshad MJ, Dunn DM, Weiss RB, and Jorde LB (2002) DNA sequence variation in a 3.7-kb noncoding sequence 5′ of the CYP1A2 gene: implications for human population history and natural selection. *Am J Hum Genet* **71:**528–542.
Yokomori N, Kobayashi R, Moore R, Sueyoshi T, and Negishi M (1995) A DNA methylation site in the male-specific P450 (Cyp 2d-9) promoter and binding of the heteromeric transcription factor GABP. *Mol Cell Biol* **15:**5355–5362.
Zhu Y and Pfaff DW (1994) Protein-DNA-binding assay for analysis of steroid-sensitive neurons in mammalian brain. *Methods Neurosci* **22:**245–264.

**Address correspondence to:** Dr. Magnus Ingelman-Sundberg, Division of Molecular Toxicology, Institute of Environmental Medicine, Karolinska Institutet, SE-171 77 Stockholm, Sweden. E-mail: magnus.ingelman-sundberg@imm.ki.se