Published in final edited form as:

Pharmacogenet Genomics. 2006 July ; 16(7): 515–525.

Functional properties of an alternative, tissue-specific promoter for human arylamine N-acetyltransferase 1

David F. Barker, Anwar Husain, Jason R. Neale, Benjamin D. Martini, Xiaoyan Zhang, Mark A. Doll, J. Christopher States, and David W. Hein

Department of Pharmacology and Toxicology and James Graham Brown Cancer Center, University of Louisville School of Medicine, Louisville, KY

Abstract

Variable expression of human arylamine N-acetyltransferase 1 (NAT1) due to genetic polymorphism, gene regulation or environmental influences is associated with individual susceptibility to various cancers. Recent studies of *NAT1* transcription showed that most mRNAs originate at a promoter, P1, located 11.8 kb upstream of the single open reading frame (ORF) exon. We have now characterized an alternative *NAT1* promoter lying 51.5 kb upstream of the *NAT1* ORF. In the present study, analysis of human RNAs representing 27 tissue types by RT-PCR and quantitative RT-PCR showed the upstream 51.5 kb promoter, designated P3, to be most active in specific tissues, including kidney, liver, lung, and trachea. All *NAT1* P3 mRNAs included 5'-untranslated region (5'-UTR) internal exons of 61 and 175 nucleotides in addition to the 79 nucleotide 5'-UTR exon present in P1 mRNA. CAP-dependent amplification of 5' P3 mRNA termini defined an 84 bp transcription start region in which most start sites are centrally clustered. The hepatoma-derived HepG2 cell line expressed a high level of P3 mRNA with the same spliced structure and start site pattern as found in normal tissues. A 435 bp minimal promoter was defined by transfection of HepG2 with luciferase expression constructs containing genomic fragments from the P3 start region. These findings imply a fundamental role for P3 in *NAT1* regulation and define additional regions for genetic polymorphisms associated with enhanced cancer risk.

Keywords

tissue-specific transcription; arylamine N-acetyltransferase; promoter regions; alternative splicing; carcinogens; tumor cells; neoplastic transformation

INTRODUCTION

Human arylamine N-acetyltransferase 1 (NAT1) and 2 (NAT2), encoded by adjacent genes on 8p22, are isozymes with important roles in the biotransformation of pharmaceuticals and environmental carcinogens [1]. NAT1 and NAT2 proteins share 80% amino acid identity, but exhibit distinct substrate selectivity for acetylation of exocyclic arylamines [2]. *NAT2* coding region single nucleotide polymorphisms (SNPs) that greatly reduce enzymatic activity are common in most human populations and have been shown to underlie variable susceptibility to urinary bladder cancer in smokers [3] and in workers exposed to aromatic amines [1,4]. *NAT2* polymorphism may also be involved in variable susceptibility to other cancers caused by exposure to heterocyclic amines in tobacco smoke and well-done meats [5-7]. In contrast to *NAT2*, *NAT1* coding region polymorphisms that reduce enzyme activity are rare on a population basis [8-10]. The most common known *NAT1* polymorphisms are located in the 3'-

UTR and although the location of a functional polyadenylation sequence is modified, there is limited direct evidence for any effect on acetylation phenotype [11-14]. Associations of *NAT1* 3' UTR polymorphisms with cancer susceptibility have been reported [15-19], but not all studies are concordant [9,20,21], perhaps because NAT1 activity is also subject to modulation by intracellular conditions [22-24]. Although epidemiological studies with known *NAT1* polymorphisms are controversial, it is likely that NAT1 activity has an important role in the metabolism of environmental carcinogens, because *NAT1* mRNA and catalytic activity are expressed in nearly all human tissues and NAT1 metabolizes several aromatic and heterocyclic amine carcinogens classified as known or suspected carcinogens [1,25]. Humans are exposed to many of these carcinogens via tobacco products, well-done meats and industrial processes. Hence, investigations of genetic and other factors influencing *NAT1* expression are expected to reveal important aspects of cancer susceptibility associated with environmental exposures.

Recent studies, drawing on information present in cDNA sequence databases as well as experimental approaches including 5'-RACE, RT-PCR and transfection of promoter reporter constructs, showed that transcription of *NAT1* is considerably more complex than previously recognized. [26-29]. For both *NAT1* and *NAT2*, a single exon includes a complete 870 nucleotide (nt) open reading frame (ORF). However most *NAT2* mRNAs are spliced and originate at a promoter located 8.6 kb 5' to the *NAT2* ORF exon [28]. Most *NAT1* mRNAs originate 11.8 kb 5' to the *NAT1* ORF exon, at a promoter designated P1, and contain a 5'-UTR non-coding 79 nt exon [26,28]. Less frequently, variant P1 mRNAs also include one or more additional 5'-UTR exons resulting in altered translational efficiency [27]. The P1 promoter is active in a wide variety of tissue types and likely accounts for most of the apparently ubiquitous expression of NAT1 enzyme in human tissues and cell lines.

A minor class of *NAT1* mRNAs is represented by three similar *NAT1* cDNA sequences in Genbank, CD702544, BM926650 and BM924372. The 5' ends of these three cDNAs all lie within a 34 bp segment, defining a possible alternative promoter, designated P3, located approximately 51.5 kb upstream of the *NAT1* ORF exon. The CD702544 cDNA is from nasopharynx. The other two cDNAs were isolated from libraries prepared with RNA representing more than one organ type and are thus of uncertain tissue origin. Alternative promoters have potential roles in mediating developmental and tissue-specific gene expression as well as responses to internal signals and the external environment [30]. Thus, we have investigated the functional properties of the *NAT1* P3 promoter to determine the pattern of tissue-specific expression, the structure of the mRNA product, the transcription start site region and the minimal functional promoter sequence. Results of these analyses indicate that the P3 promoter is highly expressed in liver, lung, kidney and trachea. Altered expression of this promoter due to transcriptional control responses or genetic variation within the regulatory region could play an important role in determining individual acetylation phenotypes and susceptibility to environmental carcinogens.

METHODS

Human RNA samples

Purified total RNAs from human tissues were obtained from two separate and distinct commercial sources, Ambion (Ambion Inc., Austin, TX) and BD (BD Biosciences Clontech, Palo Alto, CA). The human test panel, from Ambion, included adipose tissue, bladder, brain, cervix, colon, esophagus, heart, kidney, liver, lung, ovary, placenta, prostate, skeletal muscle, small intestine, spleen, testes, thymus, thyroid and trachea. Each of the Ambion panel RNAs was a pool derived from three individuals. The human test panel obtained from BD included RNAs from brain, cerebellum, colon, fetal brain, fetal liver, heart, kidney, lung, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis,

thymus, trachea and uterus. Additional total RNAs obtained from BD represented human bladder and liver. For 5'-RACE analysis, kidney and lung RNA samples of individuals different from those included in the test panel were also obtained from BD. The BD RNAs generally were pooled from a variable number of individuals, ranging from three to 63, however the brain, heart, liver, placenta, and stomach RNAs and the kidney RNA used for 5'-RACE were derived from single individuals.

Cell lines and RNA extraction

HepG2 human hepatocellular carcinoma cells were obtained from the ATCC (#HB-8065) and cultured in a humidified 5% CO₂ atmosphere in 1:1 DMEM/F12 (Gibco) supplemented with 10% fetal bovine serum (Hyclone), penicillin (100 U/ml) and streptomycin (100 µg/ml) (BioWhittaker). MCF-7 cells were grown as described [26], and samples of additional cell lines, grown in standard conditions, were obtained from University of Louisville investigators. RNAs were extracted with the RNeasy kit (Qiagen Inc., Valencia, CA). RNA from the HeLa human cervical carcinoma cell line was obtained from Invitrogen (Carlsbad, CA).

RT-PCR and determination of product exon compositions

Superscript III (Invitrogen, Carlsbad, CA) reverse transcriptase (RT) was used to prepare cDNA according to the manufacturer's protocol in a 20 µl reaction with random hexamer primers and 1 µg of total RNA. For RT-PCR of the complete 5'-UTR of P1 or P3 mRNAs, the cDNA was diluted 5-fold with water and 4 µl was used as template for PCR with a reverse primer, CDRT1, located in the *NAT1* protein coding region and either a forward primer specific for P1 mRNA (P1For) or one specific for P3 mRNA (P3For, Table 1). PCR was carried out at 94°C for 3 minutes followed by 35 cycles of 94°C for 30 seconds, 62°C for 60 seconds, 72°C for 75 seconds, and final extension at 72°C for 3 minutes. Control PCRs prepared with all components except RT were also performed. Products were analyzed on 1.2% agarose gels in sodium boric acid [31]. For determination of exon structures, RT-PCR product fragments were excised from agarose, gel-purified with QiaexII (Qiagen) and cloned with the pcDNA3.1/V5-His-TOPO® vector (Invitrogen). Clone colonies were assessed for the presence and length of inserted fragments by colony PCR with the T7 promoter and BGH reverse primers (Invitrogen). For DNA sequencing, plasmid DNAs were prepared and sequenced as described [26]. For EcoRI digest analysis, 5 µl of the colony PCR product was incubated with 10 units of EcoRI (New England Biolabs) as described by the manufacturer and electrophoresed in 1.5% agarose.

Quantitative real-time RT-PCR

For quantitative real-time RT-PCR, TaqMan® analysis was performed using the 7700 sequence detection system from ABI (Applied Biosystems, Foster City, CA). The 20 µl PCR reactions were 1X TaqMan® Universal Master Mix, with 300 nM of the primers P3qF and P3qR and 100 nM probe (P3pr, Table 1). Quantitation of the endogenous control 18S rRNA was performed using TaqMan® Ribosomal RNA Control Reagents for 18S rRNA (Applied Biosystems). Four µl of diluted cDNA, equivalent to 40 ng of the initial RNA template, was used in each PCR. Two independent RT reactions were prepared for each RNA, and for each cDNA product, real-time PCRs were carried out in duplicate, for a total of four readings. For PCR an initial incubation at 50°C for 2 minutes and 94°C for 10 minutes was followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Each assay plate included reaction mix controls with no added template. As an additional negative control, TaqMan® quantitation of reactions prepared in parallel without addition of RT were also run at least once for each set of replicate RT reactions. Baselines and threshold levels were selected as recommended by ABI and values of Ct, the cycle number at which the measured fluorescence reached the set threshold value, were recorded. For each real-time run, a Δ Ct (*NAT1* Ct – 18S rRNA Ct) was calculated for each sample and a $\Delta\Delta$ Ct value was then derived by subtracting the smallest

Δ Ct from all Δ Ct values. The formula $2^{-\Delta\Delta Ct}$ was used to calculate an initial relative concentration value for all samples, except that samples with a *NAT1* Ct of 40 were assigned as zero. The two initial values obtained with one RT were then averaged and normalized with respect to the average of all samples in the same RT group and the average of the measures obtained with each separate RT was plotted.

5' RLM-RACE analysis of P3 transcription start sites

The First-Choice™ kit (Ambion, Austin, TX) was used for RLM-RACE analysis of total RNA from kidney, lung, trachea, liver, HepG2 and fetal liver as previously described [26]. Following ligation of the 5' RACE adapter oligomer, reverse transcription with Superscript III was performed with random decamer primers or a *NAT1* specific primer (NAT1RT, Table 1), in a 20 μ l reaction. Primers utilized for nested PCR were primers to the 5' adapter (First-Choice™, Ambion) and *NAT1* specific outer and *NAT1* P3 specific inner primers, (P3outr, P3innr respectively, Table 1, Figure 1). For each nested PCR, 1 μ l of the RT reaction was used as template for a 25 μ l outer PCR reaction and 2 μ l of the outer PCR product was used as template for a 50 μ l inner PCR reaction. All PCRs with Amplitaq DNA polymerase (ABI), were started in a preheated 94°C block followed by 94°C for 3 minutes and then 35 cycles of 30 seconds at 94°C, 30 seconds at 60°C, 90 seconds at 72°C, and final extension at 72°C for 7 minutes. The major product band(s) from independent nested PCRs were gel-purified, cloned into the pcDNA3.1/V5-His-TOPO® vector (Invitrogen) and sequenced to determine transcription start site locations.

Functional analysis of promoter segments in luciferase expression vectors

DNA from BAC clone CTD2547L16 was used as a template for PCR using the high fidelity Phusion polymerase (New England Biolabs, Beverly, MA) and primers with artificial HindIII restriction sites included at their 5' ends (Table 1). The products were purified with the Qiaquick PCR purification kit (Qiagen), digested with HindIII, gel purified and ligated into the HindIII site of pGL3-Basic (Promega, Madison, WI). Clones with an appropriately oriented single insert were identified by colony PCR using orientation specific primers (P3ori and pGL3ori, Table 1) and by agarose gel analysis of plasmid DNAs with and without HindIII digestion. Duplicate clones of independently amplified products were prepared for each segment and assayed for promoter strength by transfection of the HepG2 cell line. For these assays, 24 well plates were seeded with 2.5×10^5 HepG2 cells per well and grown for 24 hours to 80-90% confluence. Each transfection was performed in triplicate wells with purified plasmid DNAs (Qiagen Midi-Kit) and Transfast Transfection Reagent (Promega), following the manufacturer's recommendations using 250 ng of DNA and a 1:1 transfection reagent:DNA ratio. Ten ng of the phRL-TK plasmid carrying the Renilla luciferase gene was included as a control for transfection and expression efficiency. Cells were incubated with the transfection mixture for 2 hours, followed by addition of complete growth medium and incubation for 48 hours prior to luciferase assay. Each experimental plate included control transfections with no DNA, pGL3 Basic, and pGL3-Control. Three replicate transfections were performed with each promoter construct. Dual assay of firefly and Renilla luciferase was carried out and promoter strengths calculated as described previously [26] by calculating the ratio of firefly to Renilla luciferase activity and normalizing with respect to the baseline promoter strength of pGL3-Basic.

RESULTS

RT-PCR detection of *NAT1* P3 mRNAs in specific human tissues

Forward primers specific for the amplification of *NAT1* mRNAs originating from the major promoter P1 and the alternative P3 promoter were designed based on unique sequences present in the most 5' portion of representative cDNAs (Fig. 1). RT-PCR reactions with each specific

primer and a common reverse primer located within the *NAT1* ORF exon (Fig. 1) were performed with total RNAs from 27 different human tissues. Fifteen of the tissues were represented by two independent RNA samples from distinct commercial sources. A similar P1-specific RT-PCR pattern was detected with RNAs from every tested tissue (Fig. 2). The most prominent product of approximately 425 bp corresponds to the expected size for the typical mRNA originating at P1 (Fig. 1). The presence of additional fainter bands is consistent with previous reports describing minor P1 transcripts with additional, alternatively spliced exons within the 5'-UTR [26-28]. In contrast to the ubiquitous P1-specific RT-PCR products, strong P3-specific products detectable in ethidium bromide stained gels were only observed with RNAs from fetal liver, kidney, liver, lung, ovary and trachea. The typical P3 RT-PCR for these tissues resulted in a relatively strong product of approximately 650 bp, and variable, larger, faint product(s), most prominently at ~750 bp (Fig. 2). A similar P3-specific RT-PCR pattern was observed for both of the independent kidney, liver, lung and trachea samples tested. Weak bands were occasionally detected with P3-specific RT-PCR of RNAs from colon, cervix, fetal brain, placenta, prostate and testis, but these were typically very faint, variable in size and/or less consistently detected. No P1 or P3 RT-PCR products were detected in any of the controls prepared without RT enzyme (data not shown).

Exon composition of *NAT1* P3 mRNAs

P3-specific RT-PCR products from fetal liver, kidney, liver, lung, prostate and trachea (Fig. 2) were extracted from agarose gels and candidate plasmid clones were prepared and characterized for insert size by colony PCR. The structures of thirteen of these clones, representing each tissue tested and each distinct insert size detected, were determined by DNA sequencing. Subsequently, additional RT-PCR product clones from the same tissues as well as ovary were classified by comparing their lengths and *EcoRI* digest patterns with the previously sequenced clones. For fetal liver, kidney, liver, lung, ovary and trachea, characterization of clones of the most prominent P3 RT-PCR product revealed the 663 bp fragment shown as the major P3 mRNA in Fig. 3. The larger, distinct, minor RT-PCR product bands from fetal liver, kidney, liver, lung ovary and trachea RNAs (Fig. 2) were shown to correspond to the minor P3 mRNA m1 (Fig. 3) including the additional 118 nt exon. Other RT-PCR products (Fig. 3), detected as rare clones with unusual insert sizes, included either a 46 nt exon segment, corresponding to mRNA m2, found in liver and trachea, or a 151 nt exon segment present in mRNA m3, found in fetal liver and prostate.

Quantitative analysis of *NAT1* P3 expression in human tissues

Based on the structure of the major P3 mRNA, (Fig. 3), a P3-specific TaqMan® (Applied Biosystems) quantitative RT-PCR (q-RT-PCR) probe and intron-spanning primer set was designed, with a forward primer, P3qF, located near the 3' terminus of the 175 bp exon, a reverse primer, P3qR, within the *NAT1* ORF-containing exon and a probe, P3pr, from within the 79 bp exon segment (Table 1, Figure 1). The q-RT-PCR measurements of P3 mRNA in human tissues (Fig. 4) were consistent with the strengths of RT-PCR product bands observed for the same RNA samples (Fig. 2). Among the 15 tissues represented by two independent samples, kidney, liver, lung and trachea showed consistently high expression, and prostate and testis showed low but detectable P3 expression with the TaqMan® assay. No P3 mRNA was detected in any of the bladder, brain, colon, heart, placenta, skeletal muscle, small intestine, spleen or thymus samples. Among the tissues with only one representative sample, ovary and fetal liver had substantial expression of P3 and there was low detectable expression in fetal brain. No P3 mRNA was detected by the TaqMan® analysis of the single samples of adipose, cerebellum, cervix, esophagus, salivary gland, spinal cord, stomach, thyroid, or uterus. Although kidney, liver, lung and trachea showed the highest P3 expression in both of the sample panels tested, there was no clear hierarchy of P3 mRNA expression among them. In panel A,

for example, trachea had the highest P3 expression and kidney was low, but this relation was reversed in test panel B.

Identification and characterization of a human cell line expressing the P3 promoter

Eight human cell lines were assessed for the presence of P3-specific expression by RT-PCR and q-RT-PCR as described for RNAs from human tissues (Fig. 5). The relative level of P3 expression was highest for the hepatoma-derived line HepG2, weak for the colon adenocarcinoma line Caco-2, near the lower limit of detection for HEK293 and undetectable for the five other cell lines tested. The pattern of P3 RT-PCR products for HepG2 (data not shown) was similar to that for the human RNAs (Fig. 2). Cloning and sequencing of representative RT-PCR fragments confirmed that the major P3 mRNA in HepG2 corresponded to the major P3 mRNA found in human tissues (Fig. 3). One distinct minor P3 mRNA with an additional 150 bp exon, m4 (Figure 3), was detected as a rare HepG2 RT-PCR product.

Identification of P3 transcription start sites with 5'-RLM-RACE

To further define the location of P3 transcription start sites, 5'-RLM RACE was performed with RNAs from the HepG2 cell line and the four human tissues with the highest P3 expression. For each RNA source, 18 to 22 independent 5'-RACE products containing P3-specific transcription start sites (TSSs) were cloned and sequenced. All of the P3 TSSs identified in kidney, liver, lung and trachea occurred within the 84 nts upstream of the splice recognition site at the 3' terminus of the P3 exon (Fig. 6). Among the 84 P3 TSSs from these four human tissues, 48 (57%) occurred within the six nt region from -46 to -41 with respect to the splice junction. Most of the remainder, 38/84 (45%), occurred downstream of the -41 position and relatively few, 6/84 (7%), occurred upstream of the -46 site. The HepG2 cell line had a very similar pattern of P3 TSSs, including 14/18 TSSs located in the -46 to -41 segment and two others at -72 and -84. Two additional HepG2 P3 TSSs occurred outside of the 84 nt segment, at -348 and at +62 with respect to the coordinates shown in Figure 6. Of the 102 total TSSs observed, 85 occurred at an adenine, 12 at guanine, 3 at cytosine and 2 at thymine. Many of the transcription start sites, notably the two most commonly used sites at positions -41 and -43 fall in the context of an initiator-type promoter element sequence [32], which is consistent with the absence of any adjacent upstream TATA-box in this region.

The PCR strategy for P3 5'-RACE analysis employed an outer primer from the *NAT1* 79 nt 5'-UTR exon and an inner primer from the 175 nt exon segment. Thus, the 5'-RACE results showed that all mRNAs which include the 175 and 79 nt exons are initiated at P3. The initial RT-PCR with a forward primer located in the P3 start region and a reverse primer in the *NAT1* ORF also showed that all *NAT1* mRNAs initiated at P3 include both of the 175 nt and 79 nt exons. Taken together, these findings support the scheme presented (Fig. 3) as an accurate and complete description of *NAT1* mRNAs produced from the P3 promoter, and validate the use of a forward primer from within the 175 base exon for P3-specific q-RT-PCR. The correlation between the strength of P3-specific RT-PCR products (Fig. 2) with the quantitative data (Fig. 4) also indicates that both of these distinct primer pairs are specific for the same mRNA.

Results of the 5'-RACE analysis also showed that the pattern of transcription initiation within the P3 promoter segment is similar in all of the high-expressing tissues, with a strong tendency for initiation to occur at adenines between -41 and -46 or within 10 to 15 nts 3' to that position. The 5'-RACE analysis of the HepG2 P3 mRNAs also largely reflected the pattern defined by the human tissue samples, with 14/18 TSSs occurring within the -41 to -46 region (Fig. 6). Although the presence of 2/18 P3 mRNAs in HepG2 with TSSs not found in any normal tissue is a minor concern, the characteristics of P3 mRNAs in HepG2 are generally identical to what is found in normal tissues with respect to TSS location and splicing pattern. Thus, the relatively

high level of P3 mRNA in HepG2 detected by q-RT-PCR (Fig. 5) primarily reflected the transcription pattern of normal cells, indicating that HepG2 is a suitable cell line for identifying and characterizing the biologically relevant P3 promoter and associated transcription factors.

Definition of the functional boundaries of the P3 promoter

Luciferase reporter constructs containing genomic segments spanning the P3 TSS region were transfected into the HepG2 cell line to identify the functional boundaries of the P3 promoter (Fig. 7). A 2372 bp fragment, including 1926 bp upstream of the P3 TSS region and also including 130 bp 3' to the 61 bp exon (Fig. 7) has a promoter strength of 7, and 5'-shortened fragments of 893 bp, 728 bp and 565 bp with the same 3' terminus exhibit promoter strengths of 35, 52 and 53 respectively. Removal of 130 bp 3' to exon 61 from the 565 fragment further increased promoter strength to 76. However, deletion of the remaining 232 nts lying 3' to the P3 start exon including the entire intron and the 61 bp exon caused a 5-fold reduction in promoter strength to 15. (Fig. 7). Hence, this region may contain important binding sites for transcription factors or proteins of the spliceosome that support mRNA stabilization. These results show that the 435 bp segment from 119 bp upstream of the 84 bp P3 start site region to 232 bp downstream includes elements necessary and sufficient for P3 promoter expression in the HepG2 cell line.

DISCUSSION

Our results confirm functional expression of an alternative *NAT1* promoter located 51.5 kb upstream of the single *NAT1* ORF exon [26]. In contrast to ubiquitous expression of the *NAT1* P1 promoter, high level P3 expression was found in a few specific tissues, including kidney, liver, lung and trachea and possibly ovary and fetal liver. Evidence for P3 tissue-specificity was confirmed in two sets of human RNA samples obtained from independent sources and by the use of distinct sets of P3-specific primers for RT-PCR, quantitative RT-PCR and also 5'-RACE. Experimental conditions for the ready detection and quantitation of *NAT1* P3-type transcripts were established, in contrast with two earlier negative reports [27, 28], possibly due to technical differences or interindividual variability of P3 expression.

The typical P3 mRNA (Fig. 3) includes internal 5'-UTR exons of 61, 175 and 79 nts as well as the variable 5' exon, which may be as large as 78 nt but is usually less than 46 nts long. Since the typical P1 mRNA 5' UTR includes only the 79 nt exon and a variable start exon of less than 89 nt, the 5'-UTR of a P3 mRNA is usually about 225 nt longer. The 175 nt exon sequence includes two potential AUG start sites that have adequate and strong initiation potential [33] and their presence could sharply reduce translation of the *NAT1* ORF but might also provide control points for translation regulation [33,34]. A minor P3 mRNA detected in many tissues also includes an 118 nt exon between the 175 and 79 nt exons and the occasional presence of this 118 nt exon was previously observed for mRNA products from the P1 promoter [26-28]. The 150 nt exon found in the rare P3 mRNA m4 (Fig. 3) was also found in one *NAT1* P1 mRNA from blood (Genbank accession BG941635). The other exons represented in m2, m3 and m5 forms of P3 mRNA are all located 5' to the *NAT1* P1 promoter and have not been reported previously. Their presence could be due to rare alternative or improper splicing of the 39,039 nt intron. Further investigation will be required to determine the relative proportions of mRNAs initiated at the P1 and P3 promoters in different cell types as well as the composition of differentially spliced products. The presence of the additional exons that are unique to P3 mRNAs could affect translation initiation, translational regulation or mRNA nuclear export, stability, or cytoplasmic localization [35], resulting in important differences in expression of NAT1 enzyme activity.

Determination of the biological relevance of the novel P3 promoter and its role in *NAT1* expression will be an important area of future investigation. Many of the tissues in which P3

is expressed have potentially high exposure to the external environment, including trachea and lung through respiration, liver through digestion and kidney through high volume blood flow. Regulated transcription at the P3 promoter may modulate NAT1 activity in response to environmental changes or constitutive P3 expression may supplement NAT1 enzyme production at these key organ sites. Prior to pursuit of these hypotheses, further confirmation and characterization of the pattern of tissue-specific expression and sublocalization of areas of high P3 expression within the architecture of organs from patients with a known medical history will be important.

Investigation of the role of the P3 promoter in a variety of circumstances where NAT1 transcription is controlled or altered will also be of interest. *NAT1* is known to be expressed in placenta from very early times of development [36], but we detected only P1 type mRNAs in two placental RNA samples. The murine *Nat2* gene, the likely ortholog of human *NAT1* [37, 38], is expressed in the neural tube at the time of closure and is also temporally and spatially regulated during cardiogenesis [39]. However, *Nat2* is not known to have any alternative promoter and we detected no homology to the human P3 segment in the murine genome. Thus, current data do not support a role of the *NAT1* P3 promoter in regulated expression during development. *NAT1* mRNA has also been shown to be one of the most frequently upregulated mRNAs in estrogen receptor positive breast tumors [40-42] and high NAT1 expression may play a direct role in tumor cell hyperproliferation and resistance to apoptosis [43]. The mechanism of altered *NAT1* mRNA regulation in breast tumors is not currently understood, but could involve the activation of an alternative promoter. Further analysis of the expression pattern and control elements of the *NAT1* P3 promoter as well as of the more widely expressed P1 promoter will be needed to attain a complete understanding of *NAT1* transcriptional regulation and its relation to the physiological roles of NAT1.

Finally, our characterization of the P3 promoter and the typical P3 mRNA structure confirms that functional elements of the *NAT1* gene are encompassed within a genomic segment of at least 53 kb, extending the target region that must be considered for the presence of functionally relevant genetic polymorphisms. SNPs within the P3 promoter, the 5'-UTR exons or within the long intron could all cause changes in NAT1 activity or regulation. There is substantial disequilibrium across this ~50 kb segment in several populations [44]. Thus, some epidemiological associations with high frequency polymorphisms in the *NAT1* 3'-UTR that are difficult to explain in terms of any direct functional effect may be due to linkage disequilibrium with as yet undetected functional SNPs located in the 5' portion of the gene. Identification of such a SNP and its functional mechanism could help to clarify the role of *NAT1* in human susceptibility to carcinogenesis.

Acknowledgements

Grant support: This work was partially supported by USPHS grants CA34627 (D.W.H.) and ES12557 (A.H.) and a grant from the University of Louisville Center for Genetics and Molecular Medicine (D.W.H.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

1. Hein DW, Doll MA, Fretland AJ, Leff MA, Webb SJ, Xiao GH, et al. Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol Biomarkers Prev* 2000;9:29-42. [PubMed: 10667461]
2. Kawamura A, Graham J, Mushtaq A, Tsiftoglou SA, Vath GM, Hanna PE, et al. Eukaryotic arylamine N-acetyltransferase. Investigation of substrate specificity by high-throughput screening. *Biochem Pharmacol* 2005;69:347-359. [PubMed: 15627487]
3. Garcia-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet* 2005;366:649-659. [PubMed: 16112301]

4. Carreon T, Ruder AM, Schulte PA, Hayes RB, Rothman N, Waters M, et al. NAT2 slow acetylation and bladder cancer in workers exposed to benzidine. *Int J Cancer* 2006;118:161–168. [PubMed: 16003747]
5. Deitz AC, Zheng W, Leff MA, Gross M, Wen WQ, Doll MA, et al. N-Acetyltransferase-2 genetic polymorphism, well-done meat intake, and breast cancer risk among postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 2000;9:905–910. [PubMed: 11008907]
6. Hamasaki T, Inatomi H, Katoh T, Aono H, Ikuyama T, Muratani T, et al. N-acetyltransferase-2 gene polymorphism as a possible biomarker for prostate cancer in Japanese men. *Int J Urol* 2003;10:167–173. [PubMed: 12622714]
7. Chan AT, Tranah GJ, Giovannucci EL, Willett WC, Hunter DJ, Fuchs CS. Prospective study of N-acetyltransferase-2 genotypes, meat intake, smoking and risk of colorectal cancer. *Int J Cancer* 2005;115:648–652. [PubMed: 15700302]
8. Hughes NC, Janezic SA, McQueen KL, Jewett MA, Castranio T, Bell DA, et al. Identification and characterization of variant alleles of human acetyltransferase NAT1 with defective function using p-aminosalicylate as an in-vivo and in-vitro probe. *Pharmacogenetics* 1998;8:55–66. [PubMed: 9511182]
9. Lin HJ, Probst-Hensch NM, Hughes NC, Sakamoto GT, Louie AD, Kau IH, et al. Variants of N-acetyltransferase NAT1 and a case-control study of colorectal adenomas. *Pharmacogenetics* 1998;8:269–281. [PubMed: 9682272]
10. Lo-Guidice JM, Allorge D, Chevalier D, Debuysere H, Fazio F, Lafitte LJ, et al. Molecular analysis of the N-acetyltransferase 1 gene (NAT1*) using polymerase chain reaction-restriction fragment-single strand conformation polymorphism assay. *Pharmacogenetics* 2000;10:293–300. [PubMed: 10862520]
11. Zenser TV, Lakshmi VM, Rustan TD, Doll MA, Deitz AC, Davis BB, et al. Human N-acetylation of benzidine: role of NAT1 and NAT2. *Cancer Res* 1996;56:3941–3947. [PubMed: 8752161]
12. Bruhn C, Brockmoller J, Cascorbi I, Roots I, Borchert HH. Correlation between genotype and phenotype of the human arylamine N-acetyltransferase type 1 (NAT1). *Biochem Pharmacol* 1999;58:1759–1764. [PubMed: 10571250]
13. de Leon JH, Vatsis KP, Weber WW. Characterization of naturally occurring and recombinant human N-acetyltransferase variants encoded by NAT1. *Mol Pharmacol* 2000;58:288–299. [PubMed: 10908296]
14. Bell DA, Badawi AF, Lang NP, Ilett KF, Kadlubar FF, Hirvonen A. Polymorphism in the N-acetyltransferase 1 (NAT1) polyadenylation signal: association of NAT1*10 allele with higher N-acetylation activity in bladder and colon tissue. *Cancer Res* 1995;55:5226–5229. [PubMed: 7585580]
15. Zheng W, Deitz AC, Campbell DR, Wen WQ, Cerhan JR, Sellers TA, et al. N-acetyltransferase 1 genetic polymorphism, cigarette smoking, well-done meat intake, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 1999;8:233–239. [PubMed: 10090301]
16. Hein DW, Leff MA, Ishibe N, Sinha R, Frazier HA, Doll MA, et al. Association of prostate cancer with rapid N-acetyltransferase 1 (NAT1*10) in combination with slow N-acetyltransferase 2 acetylase genotypes in a pilot case-control study. *Environ Mol Mutagen* 2002;40:161–167. [PubMed: 12355549]
17. Cascorbi I, Roots I, Brockmoller J. Association of NAT1 and NAT2 polymorphisms to urinary bladder cancer: significantly reduced risk in subjects with NAT1*10. *Cancer Res* 2001;61:5051–5056. [PubMed: 11431340]
18. Ishibe N, Sinha R, Hein DW, Kulldorff M, Strickland P, Fretland AJ, et al. Genetic polymorphisms in heterocyclic amine metabolism and risk of colorectal adenomas. *Pharmacogenetics* 2002;12:145–150. [PubMed: 11875368]
19. Li D, Jiao L, Li Y, Doll MA, Hein DW, Bondy ML, et al. Polymorphisms of cytochrome P4501A2 and N-acetyltransferase genes, smoking, and risk of pancreatic cancer. *Carcinogenesis* 2006;27:103–111. [PubMed: 15987714]
20. Millikan RC. NAT1*10 and NAT1*11 polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2000;9:217–219. [PubMed: 10698485]
21. Gu J, Liang D, Wang Y, Lu C, Wu X. Effects of N-acetyl transferase 1 and 2 polymorphisms on bladder cancer risk in Caucasians. *Mutat Res* 2005;581:97–104. [PubMed: 15725609]

22. Butcher NJ, Ilett KF, Minchin RF. Substrate-dependent regulation of human arylamine N-acetyltransferase-1 in cultured cells. *Mol Pharmacol* 2000;57:468–473. [PubMed: 10692486]
23. Rodrigues-Lima F, Dupret JM. Regulation of the activity of the human drug metabolizing enzyme arylamine N-acetyltransferase 1: role of genetic and non genetic factors. *Curr Pharm Des* 2004;10:2519–2524. [PubMed: 15320760]
24. Dairou J, Malecaze F, Dupret JM, Rodrigues-Lima F. The xenobiotic-metabolizing enzymes arylamine N-acetyltransferases in human lens epithelial cells: inactivation by cellular oxidants and UVB-induced oxidative stress. *Mol Pharmacol* 2005;67:1299–1306. [PubMed: 15644493]
25. Anonymous. Report on Carcinogens. Eleventh Edition. National Toxicology Program, Department of Health and Human Services, Public Health Service; 2005.
26. Husain A, Barker DF, States JC, Doll MA, Hein DW. Identification of the major promoter and non-coding exons of the human arylamine N-acetyltransferase 1 gene (NAT1). *Pharmacogenetics* 2004;14:397–406. [PubMed: 15226672]
27. Butcher NJ, Arulpragasam A, Goh HL, Davey T, Minchin RF. Genomic organization of human arylamine N-acetyltransferase Type I reveals alternative promoters that generate different 5'-UTR splice variants with altered translational activities. *Biochem J* 2005;387:119–127. [PubMed: 15487985]
28. Boukouvala S, Sim E. Structural analysis of the genes for human arylamine N-acetyltransferases and characterisation of alternative transcripts. *Basic Clin Pharmacol Toxicol* 2005;96:343–351. [PubMed: 15853926]
29. Grant DM. Pharmacogenetics and the regulation of gene transcription. *Pharmacogenetics* 2004;14:391–393. [PubMed: 15226670]
30. Ayoubi TA, Van De Ven WJ. Regulation of gene expression by alternative promoters. *FASEB J* 1996;10:453–460. [PubMed: 8647344]
31. Brody JR, Kern SE. Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *Biotechniques* 2004;36:214–216. [PubMed: 14989083]
32. Javahery R, Khachi A, Lo K, Zenzie-Gregory B, Smale ST. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol* 1994;14:116–127. [PubMed: 8264580]
33. Meijer HA, Thomas AA. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem J* 2002;367:1–11. [PubMed: 12117416]
34. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 2002;299:1–34. [PubMed: 12459250]
35. Moore MJ. From birth to death: the complex lives of eukaryotic mRNAs. *Science* 2005;309:1514–1518. [PubMed: 16141059]
36. Smelt VA, Upton A, Adjaye J, Payton MA, Boukouvala S, Johnson N, et al. Expression of arylamine N-acetyltransferases in pre-term placentas and in human pre-implantation embryos. *Hum Mol Genet* 2000;9:1101–1107. [PubMed: 10767335]
37. Estrada-Rodgers L, Levy GN, Weber WW. Characterization of a hormone response element in the mouse N-acetyltransferase 2 (Nat2*) promoter. *Gene Expr* 1998;7:13–24. [PubMed: 9572394]
38. Boukouvala S, Price N, Plant KE, Sim E. Structure and transcriptional regulation of the Nat2 gene encoding for the drug-metabolizing enzyme arylamine N-acetyltransferase type 2 in mice. *Biochem J* 2003;375:593–602. [PubMed: 12904181]
39. Wakefield L, Cornish V, Broackes-Carter F, Sim E. Arylamine N-acetyltransferase 2 expression in the developing heart. *J Histochem Cytochem* 2005;53:583–592. [PubMed: 15872051]
40. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–752. [PubMed: 10963602]
41. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100:8418–8423. [PubMed: 12829800]
42. Bieche I, Girault I, Urbain E, Tozlu S, Lidereau R. Relationship between intratumoral expression of genes coding for xenobiotic-metabolizing enzymes and benefit from adjuvant tamoxifen in estrogen receptor alpha-positive postmenopausal breast carcinoma. *Breast Cancer Res* 2004;6:R252–263. [PubMed: 15084249]

43. Adam PJ, Berry J, Loader JA, Tyson KL, Craggs G, Smith P, et al. Arylamine N-acetyltransferase-1 is highly expressed in breast cancers and conveys enhanced growth and resistance to etoposide in vitro. *Mol Cancer Res* 2003;1:826–835. [PubMed: 14517345]
44. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P, et al. A haplotype map of the human genome. *Nature* 2005;437:1299–1320. [PubMed: 16255080]

Figure 1. Structure and genomic alignment of distinct *NAT1* mRNAs

The genomic diagram at the top shows the locations of the *NAT1* promoter P1, the newly identified P3 promoter, the typical 5'-UTR exon elements and the single ORF-containing exon. The sizes of the 5'-UTR exons and the genomic distances between them and the promoter locations are indicated in bp. The alignments of representative *NAT1* cDNAs from GenBank (<http://www.ncbi.nlm.nih.gov/>), BG655073 from pancreatic islet and CD702544 from nasopharynx, originating respectively at P1 and P3, are shown. The locations of primers P1For, P3For and CDRT1, used for specific RT-PCR amplification of P1 and P3 are shown, with expected product sizes of 423 bp and 663 bp respectively. The locations of primers P3qF, P3qR and P3pr, used for P3-specific quantitative RT-PCR are shown in relation to the corresponding exon sequences and genomic locations. The positions of primers NAT1RT, P3outr and P3innr, used for P3-specific 5'RACE are also shown.

Figure 2. RT-PCR analysis of *NAT1* P1 and P3 mRNAs in human tissues

RNAs from various human tissues obtained from two independent commercial sources were analyzed for expression of *NAT1* P1 and P3 mRNAs by RT-PCR using specific primers (Fig. 1) and agarose gel electrophoresis. Panels A and B respectively show results for sets of tissue RNAs obtained from Ambion (Austin, TX) and from BD (BD Biosciences Clontech, Palo Alto, CA). Marker lanes with bands ranging from 300 to 1000 bp are included in each gel portion shown. Tissues labeled in bold are represented in both panels.

Figure 3. Structures of NAT1 P3 mRNAs from human tissues and the HepG2 cell line

The top line shows the genomic locations of *NAT1* P3 promoter and exon elements, with the location of the P1 promoter shown for reference. Coordinates shown for the indicated exon termini are from the May 2004 genome build (<http://genome.ucsc.edu/>). The structures of individual mRNAs and the samples in which they occurred are shown. Abbreviations are: fetal liver (FL), kidney (K), Liver (L), lung (Lu), Ovary (O), trachea (T), prostate (P), HepG2 (H). Exon elements shown in black are present in all P3 mRNAs and are the only exons present in the major P3 mRNA. The 118 base exon shown in dark outline is present in a minor fraction of P3 mRNAs from several human tissues. Lightly outlined exons occur in rare P3 RT-PCR products detected as atypical cDNA clones or were found as rare variants in the course of characterizing the RT-PCR products in the HepG2 cell line or 5'-RACE analyses (see text).

Figure 4. Analysis of *NAT1* P3 expression in human tissues by q-RT-PCR

The human tissue RNAs tested by RT-PCR (Fig. 2) were analyzed for expression of *NAT1* P3 mRNA by quantitative real-time RT-PCR as described in Methods. Panels A and B respectively show the results for tissues obtained from Ambion (Austin, TX) and BD (BD Biosciences Clontech, Palo Alto, CA). The relative mRNA content shown is on an arbitrary scale which is distinct in each panel. The BD liver sample, marked Liver (BD) was also analyzed together with the samples in Panel A to provide a point of reference with panel B. Single underlining of labels indicates tissues represented in both panels A and B. The ranges of measured values in panel A were: Kidney (0.9, 1.3), Liver (2.1, 5.2), Liver[BD](3.9, 4.1), Lung (2.6, 2.9), Ovary (0.5, 0.4), Trachea (13.3, 10.4) and in panel B: Fetal Liver (0.3, 0.4), Kidney (8.2, 9.0), Liver (5.5, 8.3), Lung (3.5, 4.2), Trachea (0.7, 3.4).

Figure 5. Analysis of NAT1 P3 expression in human cell lines by q-RT-PCR

RNAs isolated from human cell lines were analyzed for expression of NAT1 P3 mRNA by quantitative real-time RT-PCR as described in Methods. The cell lines are A549 from a lung adenocarcinoma, Caco-2, SW480, and HT-29 from colon adenocarcinomas, HEK293 from embryonic kidney, HeLa from a cervical carcinoma, HepG2 from an hepatocellular carcinoma and MCF-7 from a mammary adenocarcinoma. The BD liver sample, marked Liver (BD) was analyzed together with the cell line samples to provide a point of reference. The ranges of measured values were: HepG2#1 (0.6, 1.1), HepG2#2 (0.46, 0.51) and Liver[BD] (5.5, 8.3).

Figure 6. *NAT1* P3 promoter transcription start sites in human tissues and the HepG2 cell line
 The transcription start sites identified by P3-specific 5' RLM-RACE are shown for the four human tissues with highest P3 expression and the HepG2 cell line. Nucleotide positions are numbered with respect to the first nucleotide of the underlined donor splice sequence. Each TSS from an independently derived and cloned 5' RACE PCR product is represented by one ♦. The RNAs from kidney (1 female), liver (1 male), lung (2 males 1 female) and trachea (≥ 1 male ≥ 1 female), were all obtained from BD and the number and gender of the donor individual(s) are indicated in parentheses. The liver and trachea RNAs were from the same lot as analyzed in panel B of Figures 2 and 4, but kidney and lung samples were from different individuals.

Figure 7. Definition of the functional P3 promoter in HepG2

Four different genomic forward primers and three different genomic reverse primers (Table 1) located as shown, were used to amplify the indicated segments from the CTD2547L16 BAC and inserted into the pGL3-Basic luciferase reporter vector in the forward orientation as described in Methods. The assessed promoter strengths of each segment, relative to pGL3-Basic, are shown in the box at right. The promoter strength of pGL3-Control, which includes the SV40 promoter and enhancer, was 493.

Table 1

Names, sequences and utilities of oligonucleotides

Primer	Sequence	Utility
P3For	5'-CCTAGGCCAACTGCACAAATC-3'	P3-specific RT-PCR Forward
P1For	5'-TTGCCGGCTGAAATAACCTG-3'	P1 specific RT-PCR Forward
CDRT1	5'-AATCATGCCAGTGCTGTATTTTGG-3'	NAT1 RT-PCR reverse
P3qF	5'-GGGGATAATCGGACAATAACAATC-3'	P3 q-RT-PCR Forward
P3qR	5' TCCAAGTCCAATTTGTTCTAGACT 3'	NAT1 q-RT-PCR reverse
P3pr	6FAM 5'-CAATCTGTCTTCTGGATTAA-3' MGBNFQ	NAT1 TaqMan [®] probe
NAT1RT	5'-AAAATCTTCAATTGTTTCGAGG-3'	Specific RT for 5'-RACE
P3outr	5'-GTAGGTTGATCTTCAGTTTTAATCCAGA-3'	NAT1 P3 5'-RACE outer
P3innr	5'-TAAAAAAGGGTGGAGAGTTGTATTG-3'	NAT1 P3-specific 5'RACE inner
2115F	5'-CCCCCAAGCTTGAAGGAAAAGCCCAGAAATG-3'	P3 2372 PCR with reverse primer R1
636F	5'-CCCCCAAGCTTAGGATGGTGATCAAAAGATGCT-3'	P3 893 PCR with reverse primer R1
471F	5'-CCCCCAAGCTTCATAGCTAACTGACTGCAGTGTT-3'	P3 728 PCR with reverse primer R1
308F	5'-CCCCCAAGCTTAAGTTTGGCTCAAAGTGTCC-3'	P3 565 PCR with reverse primer R1
R1	5'-CCCCCAAGCTTCGCTGAGATAAACGTGGAG-3'	P3 reverse primer R1
Rss	5'-CCCCCAAGCTTCTGATTGTGTCAGTTTGGCCTA-3'	P3 203 PCR with 308F primer
Rls	5'-CCCCCAAGCTTACAGTTGCGTTCTGTTCTCCTGG-3'	P3 435 PCR with 308F primer
P3ori	5'-TGGGGTAATACGGTTTGTCC-3'	P3 primer for orientation PCR
pGL3ori	5'-GTTCCATCTTCCAGCGGATA-3'	pGL3-Basic primer for orientation PCR