Structural Heterogeneity of Caucasian N-Acetyltransferase at the NAT1 Gene Locus

Kostas P. Vatsis and Wendell W. Weber
Department of Pharmacology, Medical School, The University of Michigan, Ann Arbor, Michigan 48109-0626

Received August 31, 1992, and in revised form October 19, 1992

The human N-acetylation polymorphism is a genetic trait phenotypically reflected by differences in N-acetyltransferase (NAT) activity with therapeutic agents (rapid and slow acetylation), but a genetic invariability in N-acetylation of some arylamine drugs is also known. There are two highly similar human NAT genes: NAT1 is thought to encode a genetically invariant protein, whereas NAT2 has conclusively been shown to represent a polymorphic locus. This study demonstrates the presence of discrete NAT1 structural variants among Caucasians. These were detected by direct sequencing of 1.6-kilobase NAT1 fragments generated by the polymerase chain reaction with liver and leukocyte DNA from 13 subjects of established acetylator phenotype and NAT2 genotype. A prominent alteration in one of the variants was obliteration of the consensus polyadenylation signal (AATAAA→AAAAA). Several mutations were discernible in all regions of the second variant allele, including silent (codon 153) and nonsilent (Ser→Ala) substitutions in the coding region and deletion of nine bases from an AT-rich segment in the 3' untranslated region. One-half of the unrelated subjects were either homozygous or heterozygous for the mutant NAT1 alleles, both of which obeyed a Mendelian inheritance pattern. These novel results unequivocally show that human NAT1, like NAT2, is a polymorphic locus. © 1993 Academic Press, Inc.

The N-acetylation polymorphism is an inherited trait demonstrable in human populations by polymodal frequency distribution histograms for the elimination of pharmacologically diverse therapeutic agents in vivo, as well as by catalytic activity differences in hepatic and extrahepatic cytosolic N-acetylation of these substances in vitro. Such differences distinguish phenotypically rapid from slow acetylators. On the other hand, unimodal frequency distribution histograms have been obtained for the metabolism of p-aminosalicylic acid and sulfanilamide by rapid and slow acetylators of isoniazid, a phenomenon referred to as monomorphic (genetically invariant) acetylation. Of toxicological interest is the additional involvement of cytosolic NATs (acetyl-CoA:arylamine N-acetyltransferase, EC 2.3.1.5) in the bioactivation of environmental carcinogens (1, 2).

Two human NAT cDNAs (3) and genes (4, 5) [NAT1 and NAT2 (4)] were shown to encode highly similar proteins with overlapping substrate specificities, and several lines of evidence suggested that NAT1 and NAT2 represented the monomorphic and polymorphic proteins, respectively (3, 4, 6). Structural variants of NAT2 with coding region and far downstream mutations were subsequently identified in Japanese and Caucasian slow acetylators of sulfamethazine, caffeine, dapsone, and isoniazid (3, 5, 7–9). Mammalian cell culture expression systems revealed that the mutant NAT2 alleles did not affect mRNA steady-state levels but caused parallel decreases in the cytosolic content of immunoreactive NAT2 and in N-acetylation activities (3, 9). These findings were in accord with those obtained with hepatic tissue from phenotypically slow acetylators with mutant NAT2 genotypes (9), thereby establishing that NAT2 is a polymorphic gene locus. Unlike these strides with NAT2, however, a sys-

---

1This work was supported by Grant GM-44965 from the National Institutes of Health.
2To whom correspondence should be addressed at Department of Pharmacology, M6322/0626 Medical Science I, Medical School, The University of Michigan, Ann Arbor, MI 48109-0626. Fax: (313) 743-4450.

3Abbreviations: N-acetyltransferase proteins and genes are depicted in regular type (NAT) and italics (NAT), respectively; PCR, polymerase chain reaction; WBC(s), white blood cell(s); UTR, untranslated region; IP(s), internal (sequencing) primer(s); 'wt', 'wild-type'; bp, base pair(s); kb, kilobase(s); nt, nucleotide(s); RFLP, restriction fragment length polymorphism; PABP, poly(A) binding protein.
FIG. 1. Structure of human NATI. The location of the 5' UTR has been defined by NAT cDNAs (3, 5), but the actual transcriptional start site(s) have not been mapped. The four cysteines conserved in human NATI and NAT2 (3–5) are shown in the coding exon (open bar). Stippled bar, 5' flank; hatched bars, 5' and 3' UTR. E, EcoRI.

Tematic survey of NATI genotypes has not been undertaken to date.

The present study is the first report on the existence of human NATI variant alleles, identified by direct sequencing of a 1.6-kb fragment of PCR-generated NATI with liver and leukocyte DNA from 13 individuals (three of whom were family members) of defined acetylator phenotype and NAT2 genotype. The most striking change in one of the NATI variants was obliteration of the highly conserved polyadenylation signal (AA-TAAA→AAAAAAA). Different mutational events took place in the second variant, including silent and nonsilent (Ser-214→Ala) point mutations in the coding region and a sizeable 9-bp deletion (10) from an AT-rich segment in the 3' UTR. The mutant NATI alleles were present in 50% of the subjects, and displayed Mendelian inheritance characteristics.

EXPERIMENTAL PROCEDURES

DNA isolation and determination of acetylator status. Genomic DNA was available from seven subjects of established acetylator phenotype.
Oligonucleotides. NAT1 amplification primers (see Fig. 2) were made with BamHI and SstI linkers and a GC pair attached at the linker ends (total length, 36 nt). Double-stranded sequencing of NAT1 was performed with a total of eight internal (sequencing) primers (IPs), which are not shown here since the gene sequence has been published (4). Mutation sites were verified with at least two IPs in opposite orientations: the sequence and gene location of five of these IPs are given in the appropriate figure legend. End-labeling of IPs was carried out with [γ-32P]ATP, and radiolabeled oligonucleotides were purified by gel filtration chromatography (8).

PCR amplifications and direct sequencing. The composition of PCR mixtures (with amplification primers at 0.5 μM for NAT1 and 0.25 μM for NAT2) and amplification conditions have been reported (8). Direct sequencing of PCR-generated NAT1 and NAT2 was as before (8), except that 5'-25P-labeled IPs were included at 106 cpm and the annealing temperature was at the Tm for the particular primer. Samples were submitted to electrophoresis on 6% polyacrylamide/8 M urea gels, except where indicated otherwise.

RESULTS

The protein-coding region and 5' UTR of NAT1 are uninterrupted by introns (Fig. 1), whereas only the coding exon of NAT2 is free of intervening introns (3–5). Amplification of NAT1 by PCR (Fig. 2, upper) led to formation of a single DNA band of the expected size (1.6 kb) (Fig. 2, lower, lanes 1–13); no bands were generated in the absence of DNA (lane 14). Both strands of the entire 1.6-kb NAT1 fragment from the 13 individuals were sequenced directly. The normal or 'wt' sequence was found to be identical to that determined by single-stranded sequencing of NAT1 from a heterozygote (4), except that a C instead of an A was invariably present at position 1095 in the 3' UTR. This is also true for the gene and corresponding cDNA from Japanese (3, 5).

The 'wt' NAT1 sequence is depicted in column a of Figs. 3–5. The upstream (Fig. 3) and coding region (Fig. 4) segments of the NAT1 samples in columns b and c were devoid of mutations, whereas the samples in columns d and e exhibited heterozygosity in the 5' flanking region (nt -1129–1144; Fig. 3, upper), 5' UTR (nt -1064 to -1091); the exact deletion site is unknown, but the mutant sequences are the same regardless of the location of the deleted nonanucleotide.
### Table 1
Caucasian NAT1 and NAT2 Genotypes Ascertained from Direct DNA Sequencing

<table>
<thead>
<tr>
<th>Sample/subject</th>
<th>N-acetylation phenotype</th>
<th>$5'$Flank/5'UTR nt</th>
<th>Protein-coding region</th>
<th>$3'$Untranslated region Positions</th>
<th>Polyadenylation signal</th>
<th>nt $1085^f$</th>
<th>nt $1095^f$</th>
<th>Genotype designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 3, 4</td>
<td>RR, Rr, rr, rr</td>
<td>C</td>
<td>ACG (Thr) TCA (Ser)</td>
<td>Codon 153 T Codon 214</td>
<td>Intact</td>
<td>Normal</td>
<td>A</td>
<td>AAT/AAA Mutated</td>
</tr>
<tr>
<td>5, 10, 11</td>
<td>RR, Rr, rr, rr</td>
<td>C</td>
<td>ACG (Thr) TCA (Ser)</td>
<td></td>
<td>Intact</td>
<td>AAAAAAA</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>RR, Rr, rr</td>
<td>C</td>
<td>ACG (Thr) TCA (Ser)</td>
<td></td>
<td>Intact</td>
<td>AAAAAAA</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>RR, Rr, rr</td>
<td>C</td>
<td>ACG (Thr) TCA (Ser)</td>
<td></td>
<td>Intact</td>
<td>AAAAAAA</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2, 12, 13</td>
<td>RR, Rr, rr</td>
<td>C</td>
<td>ACG (Thr) TCA (Ser)</td>
<td></td>
<td>Intact</td>
<td>AAAAAAA</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

*Subjects 2, 8, and 9 are members of a two-generation family.

$^f$Phenotyping methods and N-acetylation activities for individuals 1–7 have been reported (8). Urinary caffeine metabolite ratios of 1.1 and 0.9 were determined for persons 8 and 9, respectively [a ratio of 1.82 separates slow from rapid acetylators (12)]. The acetylator phenotype of individuals 10, 12, and 13 was assessed with sulfamethazine in vivo, and the amount of N-acetylated metabolite detected in the urine 4.5 h after drug administration (expressed as % of total) was 40.9, 37.1, and 57.0%, respectively [the antinode for this method is $\sim$72% (13)]. A ratio of 0.11 was obtained for N-acetylated to unmetabolized dapsone in serum from subjects 11 and 12 [the antinode is 0.26 (12)].

$^g$Codon 153 lies within an inverted repeat with termini at nt 446 and 464. In addition to the mutated $G^{469}$. NAT1 from the deletion variants contained comigrating bands at five of the eight juxtaposed positions on the palindromic stem but only in opposite orientations [nt 460–464 ($5' \leftrightarrow 3'$) and 450–446 ($3' \rightarrow 5'$)], suggesting that the duplicate bands at these positions may not be real heterozygosities; definitive proof will be required before this possibility can be ruled out.

$^h$The intact and deleted sequences were shown in Fig. 5 and are illustrated schematically in Fig. 6.

$^i$The normal or mutated polyadenylation signal occupies positions 1086–1091 in the 'wt' and nondeletion variant NAT1, but is shifted 9 bp upstream (nt 1077–1082) in NAT1 from the deletion variants.

$^j$When gaps are inserted in the deletion variant allele, the sequence from nt 1096 to 1104 is identical to that of the 'wt' and nondeletion variant NAT1 alleles, i.e., TTTAAGAA.

$^k$NAT1 genotype designations are informal; both letters and numbers were arbitrarily assigned to the 'wt' ($V_1$) and two allelic variants ($V_2$ and $V_3$).

$^l$The genotypes of subjects 1–7 have been published (8). PCR-generated NAT2 (8) with DNA from subjects 8–13 was sequenced directly from a position 100 nt 5' of the initiator ATG codon to the end of the 3' UTR (nt 1169); the sequences were determined with the IPs reported previously (8) and an additional IP located near the 3' end of the coding region (nt 844–860, 5'-AACCTGTTGATGGATGTC-3'). No mutations were seen in the 3' UTR of NAT2 from these slow acetylators.

### Discussion
The results of this investigation offer lucid evidence that the Caucasian NAT1 gene locus, hitherto described...
as monomorphic (3–5), is indeed polymorphic. The validity of the observed NAT1 mutations is documented by the heterozygosities in over 40% of the 12 unrelated individuals (Table 1, penultimate column), and further substantiated by the Mendelian inheritance of the variant alleles (subject 2 is the offspring of subjects 8 and 9; Table 1, last two columns). Moreover, acetylator phenotypes [assessed with ‘NAT2-type substrates’ (6)] and NAT2 genotypes were interspersed with NAT1 genotypes, strengthening the suggestion from catalytic activity experiments (6) that the two NAT genes are independently expressed.

None of the most commonly studied restriction sites was eliminated or created by the mutations in variants v2 and v3, indicating that RFLP analysis would have disclosed no banding pattern differences for NAT1. In fact, Southern blots of KpnI-, BamHI-, or EcoRI-digested DNA from three human livers showed no RFLPs when probed with NAT1 cDNA, whereas a KpnI RFLP was apparent in blots probed with NAT2 cDNA (3). The absence of RFLPs, therefore, by no means signifies structural invariability for a given gene.

The 3' UTR of 'wt' NAT1 (allele V1) contains multiple copies of AAT, ATA, and TAA motifs (nt 1065–1090; Fig. 6, line 1), which resemble the 'destabilizing' AT-rich elements in the 3' UTR of several eukaryotic mRNAs with high turnover rates (14). These 'destabilizing' units are known to facilitate mRNA degradation by disrupting association of the poly(A) tract with PABP (14–16), inferring that the AT-rich structure in the 3' UTR of allele V1 could similarly promote destruction of V1 mRNA. Variant v2 mRNA may also be short-lived as a consequence of the mutated polyadenylation signal in allele v2 (nt 1086–1091; Fig. 6, line 2), since (a) point mutations within the highly conserved AATAAA sequence prevent mRNA cleavage and polyadenylation (17), and (b) nonpolyadenylated mRNAs undergo accelerated degradation in the absence of a stabilizing poly(A)–PABP complex (16). Low levels of 'wt' (V1) mRNA or variant v2 mRNA would be accompanied by formation of scant but otherwise normal V1 or v2 protein. Two observations, made with tissue preparations of undetermined NAT1 genotype, lend support to this postulate: hepatic NAT1 mRNA is scarce (3); and immunoactive NAT1 is undetectable in crude liver cytosol (4) and only barely detectable on Western blots of column fractions with partially purified cytosol possessing
ample activity toward p-aminosalicylic acid, even when
loaded in an amount five-fold greater than that of NAT2-
containing fractions (6). In sharp contrast, the 3’ UTR of
NAT2 is devoid of AT-rich segments (3, 4), hepatic NAT2
mRNA is abundant (3, 9), and immunoreactive NAT2 is
readily seen in crude liver cytosol with moderately high
sulfamethazine N-acetylation activity (4, 9).

Removal of one-third of the ‘destabilizing’ AT-rich
segment from the 3’ UTR of mutant allele v3 (Fig. 6, line
3) could result in elevated steady-state levels of v3 mRNA.
On the other hand, upstream and coding region mutations
in v3 may adversely affect the translational efficiency of
v3 mRNA or lead to production of a labile protein, as has
been described for two Caucasian NAT2 variants (9). In
any event, deletions from 3’ UTR sequences are not un-
common. Two deletions are seen in the 3’ UTR of ‘wt’
NAT1 from Japanese relative to that from Caucasians
(Fig. 6, lines 1 and 4), and have also been found in this
region of eukaryotic genes and cDNAs encoding two P450s
and an array of proteins unrelated to drug metab-
olism (18).

Human NAT1 and NAT2 are expressed in liver (6),
whereas NAT1 is by far the predominant acetylase in
mononuclear leukocytes (19). NAT1 has a selectivity for
p-aminosalicylic and p-aminobenzoic acids (6), both of
which have traditionally been classified as ‘monomorphic
substrates’ (1–3, 6, 19–21). In this connection, some old
and new observations merit consideration. Dissimilarities
in p-aminosalicylic acid N-acetylation rates had been
noted almost 30 years ago by Jenne (21) with partially
purified human liver cytosolic NAT preparations, but the
significance of this finding was ignored because of the
constancy in rates of elimination of this compound by
humans in vivo (20, 21). A difference in the thermosta-
bility of p-aminobenzoate N-acetylase was later detected
in lymphocytes from 39 persons, and a host of criteria
favored the conclusion that the differential heat stability
most likely reflected a “structural gene difference” (22).
Moreover, a population study on p-aminobenzoate N-
acetylation by whole blood from 200 human subjects re-
vealed a frequency distribution with a tendency toward
bimodality (23). More recent investigations have corro-
borated and extended the in vitro findings of Jenne (21)
by demonstrating a dramatic variation (more than 25-
fold) in p-aminosalicylate and p-aminobenzoate N-acety-
lization by cytosol from 39 human autopsy liver specimens
(6), as well as a seven-fold variation in rates of p-ami-
obenzoate N-acetylation by the cytosolic fraction of
mononuclear leukocytes from 23 individuals of undefined
NAT1 genotype (19). Lastly, and contrary to the earlier
in vivo results (1, 2, 20, 21), a preliminary study presented
after completion of the present work indicates that p-
aminosalicylate elimination in 130 healthy volunteers,
assessed by HPLC measurements of N-acetylated and
parent drug in urine, is bimodal. It seems reasonable,
therefore, that the profound structural heterogeneity at
the human NAT1 locus described in the present paper
could be the genetic basis for the substantial catalytic
activity differences seen with p-aminosalicylate and p-
aminobenzoate in vitro and in vivo, thereby suggesting
that human NAT1 participates in an acetylation poly-
morphism distinct from that typified by isoniazid and
sulfamethazine.

REFERENCES

Oxford University Press, New York.
177, 1252–1257.
265, 12757–12760.
Acad. Sci. USA 88, 6333–6337.
112A.
Genet. 5, 391.
12. Kilbane, A. J., Silbart, L. K., Manis, M., Beitins, I. Z., and Weber,
467–473.
377.
245.
889–904.

of the 9th International Symposium on Microsomes and Drug Oxida-
tions, p. 244. Jerusalem, Israel, July 6–9.