Arylamine *N*-acetyltransferase in Balb/c mice: identification of a novel mouse isoenzyme by cloning and expression *in vitro*

Susan L. KELLY and Edith SIM*

Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, U.K.

Three genes encoding arylamine *N*-acetyltransferase were identified in Balb/c mice. All three genes were cloned from genomic DNA, sequenced and expressed in a bacterial expression system. Two of the genes corresponded to *Nat*-1 and *Nat*-2 which have been previously identified in A/J and C57Bl/6 strains of mice (Martell et al., 1991). The new gene, designated *Nat*-3, can be distinguished from the other mouse *Nat* genes both by specific amplification using PCR and by restriction-endonuclease digestion. The products of all three genes are demonstrated to catalyse acetylation of aminofluorene and anisidine following expression in *Escherichia coli*.

INTRODUCTION

N-Acetyltransferase (NAT) catalyses the acetylation of arylamine and hydrazine xenobiotics, including carcinogens such as aminofluorene and benzidine. There is wide inter-individual variation in the capacity to acetylate a number of these substrates in both human and animal populations.

It was demonstrated over 30 years ago that the slow-acetylator phenotype in humans was inherited as a recessive Mendelian trait (Evans et al., 1960) and recently the molecular basis of this genetic polymorphism has been partially identified (see Sim et al., 1992; Deguchi, 1993; Grant, 1993 for reviews). There are two distinct loci on human chromosome 8 (AAC-1 and AAC-2) (Hickman et al., 1994) which encode active *N*-acetyltransferase isoenzymes (Blum et al., 1990a). These have been designated monomorphic *N*-acetyltransferase (Ohsako and Deguchi, 1990) or NAT-1 (Blum et al., 1990a) and polymorphic *N*-acetyltransferase (Ohsako and Deguchi, 1990) or NAT-2 (Blum et al., 1990a). The AAC-2 locus encoding NAT-2 in humans is multiallelic with at least five different alleles giving rise to the slow-acetylator phenotype (Vatsis et al., 1991; Hickman et al., 1992; Bell et al., 1993). They differ from the fast allele at up to three point mutations, resulting in either defective translation or a less stable enzyme (Blum et al., 1991). The AAC-1 locus encoding NAT-1 also appears to be multi-allelic (Vatsis and Weber, 1993).

Animal models of fast- and slow-*N*-acetylation have been identified (see Weber and Hein, 1985 for review) and the genetic basis for the polymorphism in a number of species has been elucidated. In rabbits there are two loci encoding two distinct NAT proteins but in slow acetylators, one of the *Nat* genes is completely missing (Blum et al., 1989a; Sasaki et al., 1991). Two *Nat* genes, both of which encode a functional NAT protein (Martell et al., 1992), have also been identified previously in mouse DNA and named *Nat*-1 and *Nat*-2 (Martell et al., 1991). A point mutation in *Nat*-2 in a slow-acetylating strain of mice results in the production of an enzyme which is transcribed/translated less efficiently and is much more unstable than that from a fast-acetylating strain (Martell et al., 1992). However, it appears that there may be two independently inherited acetylator polymorphisms in mice (Tannen and Weber, 1980). We report here that there is a third mouse gene, *Nat*-3 (Kelly and Sim, 1992), which encodes active arylamine NAT.

MATERIALS AND METHODS

Southern-blot analysis

Genomic DNA was prepared from the livers of A/J and C57Bl/6J mice, according to the method of Porteous (1985) and aliquots (15 μg) were incubated (5 h) with 45 units of each restriction enzyme (Boehringer–Mannheim) in the appropriate 'SuRE Cut' buffer at either 37 °C or 50 °C. The DNA was separated on an agarose (0.7%, w/v) gel, and 1 kb 'ladder' DNA (Gibco) and λ-*Hin*dIII molecular-mass markers were run in separate lanes. The DNA was then transferred to a poly(vinylidene difluoride)-based membrane (Immobilon-N, Millipore) using a standard Southern-blotting technique (Sambrook et al., 1989). The Southern blot was hybridized for 16 h at 64 °C with DNA probes labelled with ^{32}P (Feinberg and Vogelstein, 1983) to specific activities greater than 1×10^8 c.p.m./μg of DNA in 5 × SSPE (1 × SSPE: 0.15 M NaCl/10 mM NaH_2PO_4/1 mM EDTA, pH 7.4), 5 × Denhardt's solution [0.1% (w/v) Ficoll, 0.1% (w/v) polyvinylpyrrolidine, 0.1% (w/v) BSA] and 0.5% (w/v) SDS containing 100 μg/ml sonicated salmon sperm DNA. The membranes were washed twice at 64 °C for 15 min, in 2 × SSPE containing 0.1% (w/v) SDS and then twice at 64 °C for 30 min in 1 × SSPE containing 0.1% (w/v) SDS.

In order to restriction map clones, identified as containing *Nat* sequences, Southern blots of phage DNA (1 μg), digested with *Sac*I or of isolated *Sac*I fragments digested with *Acc*I, *Dra*I, *Eco*RV, *Hin*dIII, *Pst*I, *Sma*I or a combination of two of these enzymes, were hybridized with a mixture of *Nat*-1 and *Nat*-3 as probe. The hybridization and washing conditions were as described for genomic DNA.

The DNA probes used for the hybridizations encoded either mouse NAT-1 or NAT-3. Genomic DNA from A/J mice was amplified using the PCR with the primer pair Nat-5 and Nat-6 (Kelly and Sim, 1991). The products of this amplification were ligated into pUC18 cleaved with *Sma*I and the resulting clones were analysed by restriction digestion and partial sequencing.

Abbreviations used: IPTG, isopropyl-β-D-thiogalactopyranoside; DDT, dithiothreitol; AN, anisidine; 2-AF, 2-aminofluorene; PABA, *p*-aminobenzoic acid; SMZ, sulphamethazine.

* To whom correspondence should be addressed.

Table 1 Sequence of oligonucleotide primers used in the PCR and for sequencing the mouse *Nat* genes

Restriction endonuclease recognition sequences are underlined. The primers were used as follows: Mus1 and Mus9 were used to sequence all three *Nat* genes. Mus2, Mus3, Mus7, Mus8, Mus10 and Mus11 were used to sequence *Nat*-3. Mus12 paired with either Mus13, Mus14 or Mus15 was used to amplify the open reading frames of *Nat*-1, *Nat*-2 or *Nat*-3 respectively. Pet1 and Pet2 were used to sequence inserts within the expression vector pET-5a.

Primer	Sequence	Features
Mus1	5′-GAAGATGGCAGGAACCTGAGG	Antisense
Mus2	5′-GTTGCTGGACTTCTGGTAACC	Antisense
Mus3	5′-ATGGCTTAATGGGCACCATTC	Sense
Mus7	5′-CCTTGCCTTTCGCGACCATGGACATT	Sense, *Nru*I site
Mus8	5′-GATCATTTCCCCTGCAGACTCTAAATAGT	Antisense, *Pst*I site
Mus9	5′-CTACAGGTGACCATCAGTG	Sense
Mus10	5′-GCCTCATCATCTAATCTAGAT	Sense, *Xba*I site
Mus11	5′-AATCAAGACAACTATATCATC	Antisense
Mus12	5′-TGCCTTAGGGACATATGGACAT	Sense, *Nde*I site
Mus13	5′-AGAAGAATTCTGCTCCTTACCC	Antisense, *Eco*RI site
Mus14	5′-GTGTCCTTCCATGAATTCGGAA	Antisense, *Eco*RI site
Mus15	5′-AGATCGGATCCCCTTATTACTC	Antisense, *Bam*HI site
Pet1	5′-ACCACAACGGTTTCCCTCTAG	Sense
Pet2	5′-AAATAGGCGTATCACGAGGCC	Antisense

They were found to be of two types: one group of clones encoded 769 bp of mouse *Nat*-1 (Martell et al., 1991) and the other group encoded 769 bp of a unique *Nat* sequence (Kelly and Sim, 1992) which could be cleaved with either *Eco*RI or *Hae*III.

Isolation of genomic DNA encoding NAT

A library of Balb/c mouse genomic DNA in EMBL-3 Sp6/T7 was purchased from Clontech Lab. and grown in *Escherichia coli* NM538. Duplicate replicas of plaques were prepared on nitrocellulose filters (Sambrook et al., 1989) and screened with a mixture of ^{32}P-labelled probe encoding for NAT-1 and NAT-3. The hybridization conditions were as described for Southern-blot analysis. Positive plaques were rescreened twice after regrowth and following the tertiary screen the presence of *Nat*-1, -2 or -3 was confirmed by using an aliquot of each plaque as template DNA in the PCR primed off Nat-5 and Nat-6. The amplified DNA was incubated with either *Eco*RI (cleaves *Nat*-3), *Bgl*II (cleaves *Nat*-2) or *Hin*cII (cleaves *Nat*-1 once and *Nat*-3, twice). One of each of the three types of clone was cleaved with *Sac*I and the fragment containing the *Nat* gene was subcloned into pUC18 (Norrander et al., 1983) for restriction analysis and DNA sequencing.

PCR amplification

For the amplification of mouse *Nat*, 1 μg of genomic DNA or 10 ng of plasmid DNA was added to 50 pmol of each of a pair of oligonucleotide primers (Table 1) and either 2.5 units of *Taq* polymerase (Boehringer–Mannheim) or 2.5 units of native *Pfu* polymerase (Stratagene) in the appropriate buffer. A series of 30–40 thermal cycles was carried out as follows: denaturation, 0.5 min at 94 °C; annealing, 0.5 min at between 45 °C and 65 °C according to the primers; and extension, 1 min at 72 °C using *Taq* polymerase and 2 min at 75 °C using *Pfu* polymerase. DNA products for subcloning into plasmid vectors were separated by electrophoresis on agarose (1 %, w/v) gels and isolated using a silica matrix (Geneclean II, Bio 101, USA).

Apart from Nat-5 and Nat-6 which were described previously (Kelly and Sim, 1991), the primers used for the PCR are listed in Table 1. Mus12 used in conjunction with Mus13, Mus14 or Mus15 specifically amplifies mouse *Nat*-1, *Nat*-2 or *Nat*-3 respectively. Pet1 paired with Pet2 can be used to amplify any sequence inserted into the expression vector pET-5a (Studier et al., 1990).

DNA sequencing

The full sequence of *Nat*-3 was obtained in both directions by subcloning overlapping restriction fragments of DNA into the plasmid vector pUC18. Mini-plasmid preparations of these subclones were prepared from 5–10 ml of 16 h cultures of the recombinant plasmid in *E. coli* JM83 using the boiling method (Sambrook et al., 1989). The plasmid preparations were further purified by sequential extractions in phenol and chloroform and were precipitated with ethanol. Plasmid DNA (3–5 μg) was denatured in 0.2 M NaOH, neutralized with 0.1 vol. of 5 M ammonium acetate, pH 7.5, and finally precipitated with 4 vol. of ethanol. The DNA pellet was redissolved in 7 μl of water and sequenced using Sequenase DNA polymerase (United States Biochemical, Cleveland, OH, U.S.A.) in the dideoxy chain termination method. The oligonucleotides used to prime the sequencing reactions were either M13 Universal primer supplied by United States Biochemical, M13 Reverse primer supplied by Boehringer–Mannheim or the specific oligonucleotide primers described in Table 1.

The nucleotide sequences of the entire open reading frame of *Nat*-1 and *Nat*-2 were obtained in both directions following insertion into the plasmid pET-5a. The primers used to obtain the sequences of overlapping nucleotide fragments were Pet1, Pet2, Mus1, Mus9 (Table 1) and Nat-6 (Kelly and Sim, 1991).

Expression of *Nat*-1, *Nat*-2 and *Nat*-3

The coding regions of each of the three mouse *Nat* genes were inserted into the T7 transcription/expression region of the plasmid vector pET-5a [Novagen, AMS Biotechnology (U.K.) Ltd., Witney, Oxon, U.K.]. The cloning strategy was as illustrated in Figure 1. The recombinant plasmid was first established in *E. coli* strain NM544 for amplification and sequencing, and then transformed into the *E. coli* B. strain, BL21 (DE3), which contains the T7 RNA polymerase gene under *lac* control (Studier et al., 1990). Addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to growing cultures induces the T7 RNA polymerase, which then transcribes the genes in the plasmid vector, giving recombinant protein. *E. coli* BL21(DE3) containing recombinant plasmids were either grown for 16 h at 20 °C without induction or were induced with IPTG for 4 h at 37 °C by which time the A_{600} value had reached 0.6. The bacteria were harvested by centrifugation and the cells washed in one-fifth of the culture volume of cold 50 mM Tris/HCl, pH 8.0/2 mM EDTA. Following removal of the supernatant the cells were either lysed immediately or flash frozen in liquid N_2 and stored at −70 °C for up to 4 weeks.

Enzymic assays

in order to assay for NAT activity, the bacterial pellet was resuspended in one-twentieth of the culture volume of cold 20 mM Tris/HCl, pH 7.5, 1 mM EDTA, 1 mM dithiothreitol (DTT) containing 1 mM PEFABLOC (S. Black, Cheshunt, Herts., U.K.). Aliquots (0.5 ml) were sonicated twice for 1 min and the bacterial debris pelleted by centrifugation at 10000 *g* for

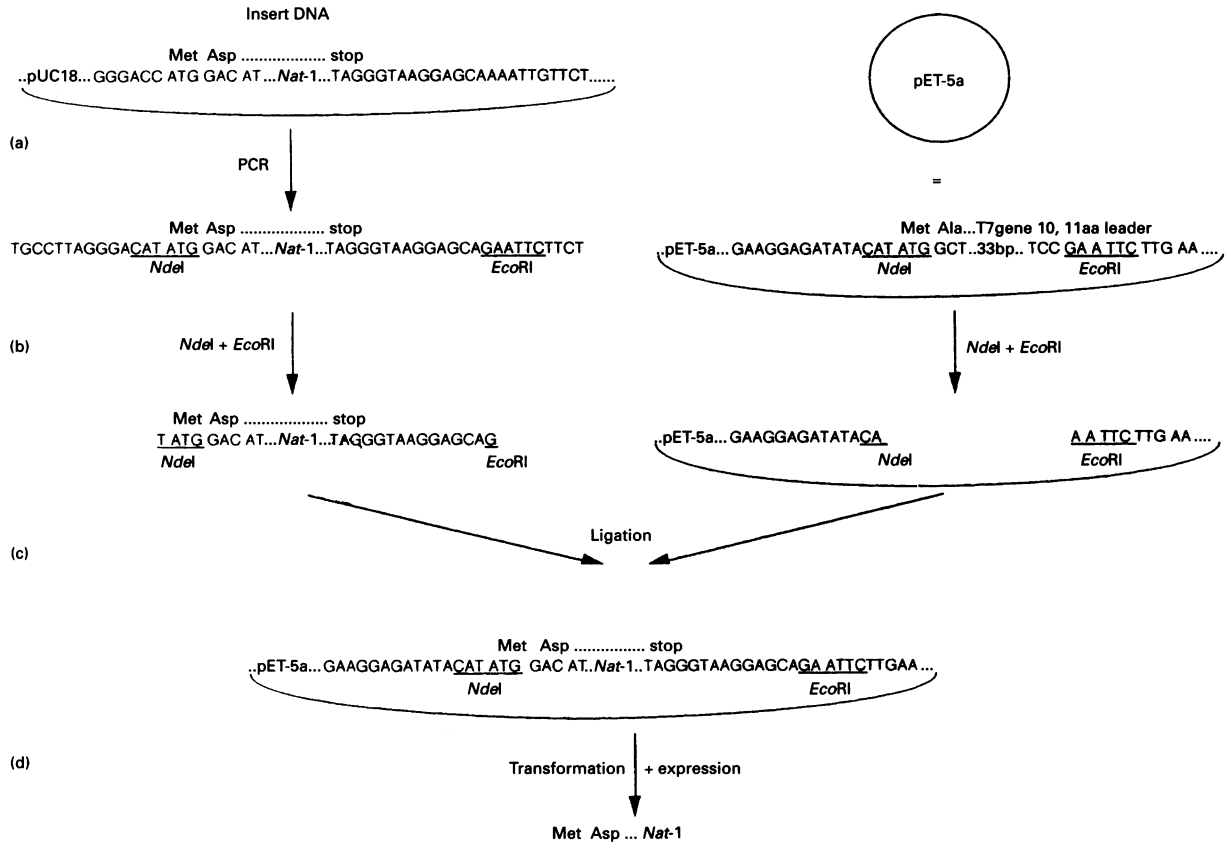

Figure 1 Scheme for expression of *Nat*-1, *Nat*-2 or *Nat*-3 using pET-5a

(**a**) Insert DNA was amplified from the clones in pUC18 using the PCR and oligonucleotide primers Mus12 paired with either Mus13 (*Nat*-1), Mus14 (*Nat*-2) or Mus15 (*Nat*-3) (see Table 1). The amplified DNA contained modified ends which allowed cleavage with the restriction endonucleases *Nde*I and either *Eco*RI (*Nat*-1 and *Nat*-2) or *Bam*HI (*Nat*-3). (**b**) The PCR products and the pET-5a vector DNA were incubated with *Nde*I and either *Eco*RI or *Bam*HI. (**c**) Following removal of the excised fragments of DNA, the insert containing the open reading frame of the *Nat* gene was ligated into the plasmid vector. (**d**) The recombinant plasmid was first established in *E. coli* strain NM544 for amplification and sequencing, and then transformed into *E. coli* strain BL21(DE3) which contains the T7 RNA polymerase gene under *lac* control. When induced with IPTG the cell was able to produce NAT protein.

5 min at 4 °C. *N*-Acetyltransferase activity in the supernatants from bacterial cell lysates containing pET-5a alone or with recombinant pET-5a carrying the gene for NAT1, NAT2 or NAT3 was determined as described previously (Ward et al., 1992) with sulphamethazine (SMZ; 1 mM), anisidine (AN; 1 mM), 2-aminofluorene (2-AF; 0.4 mM) or *p*-aminobenzoic acid (PABA; 1 mM) as substrate. The supernatants were diluted in 20 mM Tris/HCl, pH 7.5, 1 mM EDTA, 1 mM DTT so that the initial rate of acetylation of the arylamine substrate was linear. The protein content of the lysates was determined using the method of Bradford (1976) (Bio-Rad Laboratories Ltd., Hemel Hempstead, Herts, U.K.) calibrated with bovine γ-globulin.

SDS/PAGE

Analysis of expressed proteins was by electrophoresis on 10% (w/v) polyacrylamide gels (Schägger and von Jagow, 1987). Bacterial cells containing recombinant plasmids were harvested and resuspended in 0.1 vol. of lysis buffer [50 mM Tris/HCl, pH 8.0, 25% (w/v) sucrose, 1 mM EDTA] containing 1 mg/ml lysozyme. Following lysis by the addition of Triton X-100 to 0.6%, the suspension was treated with DNAase I (10 μg/ml) to reduce the viscosity (Harlow and Lane, 1988). The inclusion bodies were removed from the suspension by centrifugation at 10000 ***g*** for 10 min. The supernatant was retained for analysis of soluble protein and the pellet washed three times with 0.5% Triton X-100 in 1 mM EDTA. Finally the inclusion bodies were resuspended in 0.15 M NaCl/20 mM sodium phosphate, pH 7.5. Samples of each supernatant (15 μl), or of resuspended pellets from the inclusion body preparation (1.5 μl), were incubated at 37 °C for 30 min in buffer containing 2% (v/v) mercaptoethanol and 4% (w/v) SDS before being applied to the gel in a total volume of 30 μl. The molecular-mass markers were as follows: BSA, 66 kDa; egg albumin, 45 kDa; glyceraldehyde-3-phosphate dehydrogenase, 36 kDa; bovine carbonic anhydrase, 29 kDa; bovine pancreas trypsinogen, 24 kDa; soybean trypsin inhibitor, 20 kDa and bovine milk α-lactalbumin, 14.2 kDa (Sigma, SDS-7). The markers were treated as described above and applied in the same final volume as the samples.

RESULTS AND DISCUSSION

A third *Nat* gene in A/J and C57Bl/6 mouse strains

Preliminary evidence for the likelihood of three genes encoding NAT in the fast (C57Bl/6J) and slow (A/J) *N*-acetylating strains of mice was presented previously (Kelly and Sim, 1992). In summary, it was found that each of three restriction fragments of genomic DNA which hybridized to a full-length cDNA encoding rabbit NAT (Blum et al., 1989b) could be independently amplified

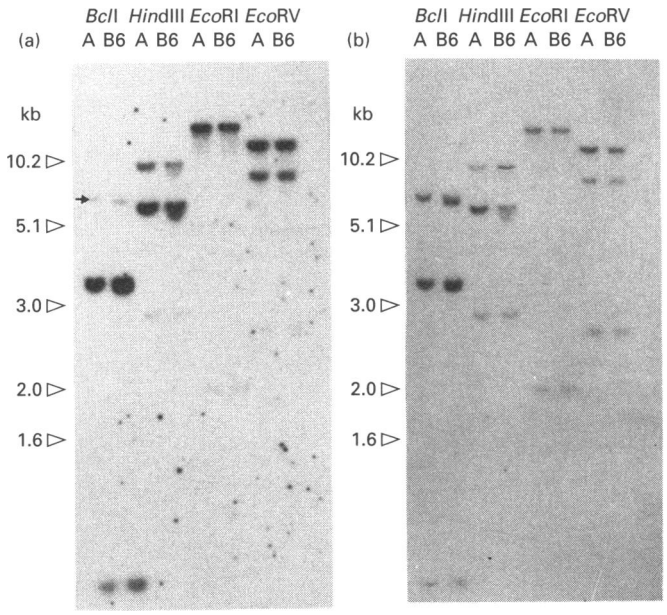

Figure 2 Southern blot analysis of mouse genomic DNA using (a) *Nat*-1 and (b) *Nat*-3 as probes

Genomic DNA from A/J and C57Bl/6 mice was prepared and digested with restriction enzymes as described in the Materials and methods section. The arrow (a) indicates the position of a possible RFLP in the *Bcl*I digests. Molecular-mass markers are indicated by open arrow heads.

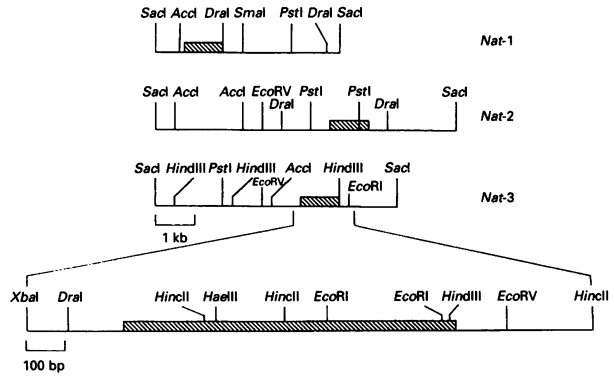

Figure 3 Restriction maps of three distinct genomic DNA clones

The three clones were isolated as described in the Materials and methods section and a partial restriction map for each is shown. The hatched area in each case represents an open reading frame. The expanded region of *Nat*-3 is the region which was sequenced.

using the PCR and primers specific for *Nat* following digestion with *Hin*dIII and size fractionation on an agarose gel. The product from the smallest of these *Hin*dIII-cleaved DNA fragments (2.7 kb) could be cut following incubation with either *Eco*RI or *Hae*III restriction enzymes, neither of which would be predicted to cut the DNA in published sequences of mouse *Nat*-1 or *Nat*-2 (Martell et al., 1991).

Southern-blot analysis of genomic DNA from A/J and C57Bl/6 mice after digestion with the restriction enzymes *Bcl*I, *Eco*RI, *Eco*RV or *Hin*dIII and using *Nat*-1 as probe revealed that there were likely to be at least two distinct arylamine *N*-acetyltransferase genes in A/J and C57Bl/6 mice (Figure 2a). These corresponded to bands which had previously been

```
                                                                    50
NAT-3 MDIEAYFERIGYQKSSNKLDLQTLTEILQHQIRAIPFENLNIHCGKTMEL
NAT-1 ............KN.V.....A....V....M..V......M...EA.H.
NAT-2 ............STRS....K.......................ES...
                                                                   100
NAT-3 SLEDTFHQIVRKKRGGWCLQVNHLLYWALAMIGFETTMLGGCVYVPSACK
NAT-1 D.Q.I.DH.....................TKM.........Y..ITPVS.
NAT-2 ...AI.D.....................TKL.........Y.FNTP.N.
                                                                   150
NAT-3 YSNTMIHLLLQVTISGKTYIVDSAFPFSCQLWEPLELTSGKDQPQVPAIF
NAT-1 ..SE.V...V.....DRK......YGG.Y.M...................
NAT-2 ..SG.....V.......D....AG.GR.Y.M...................
                                                                   200
NAT-3 HLREENGTWYLEQTKRQEYVSNQEFIDSNFLEKNTHRKIYSFTLEPRTIE
NAT-1 L.T........D.IR.EQ..P.E..VN.DL....KY...........V..
NAT-2 R.T........D.IR.EQ..P.....N.DL....KY..............
                                                                   250
NAT-3 DFWSISTYYQVSRTSVMTNTSLCSLHTKDGVHGLMGTILAYKKFNYKDNI
NAT-1 ..EYVNS.L.T.PA..FVS..F...Q.SE...C.V.STFTSRR.S...DV
NAT-2 ..E.MN..L.T.PA..F.SK.F...Q.PE...C.V.ST.T.RR.S....V
                                                                   290
NAT-3 DLVEFKTLKEEEIEEVLKSVFGIHLETKLVPKCGNVFFTI
NAT-1 ......YVN.....D...TA...S..R.F...H.ELV...
NAT-2 ......S.T.....D..RTI..VS..R.....H.DR....
```

Figure 4 Comparison of deduced amino acid sequences of *Nat*-1, *Nat*-2 and *Nat*-3 from Balb/c mice

The amino acid residues in *Nat*-1 and *Nat*-2 which differ from those in the amino acid sequence of the complete coding region of *Nat*-3 are shown using the one-letter code. Residues which are identical with those in *Nat*-3 are indicated by (.). The three conserved cysteine residues are underlined.

identified using a cDNA for rabbit NAT (Blum et al., 1989b) as a probe (Kelly and Sim, 1992). When the same Southern blot was stripped and hybridized to the putative *Nat*-3 DNA probe (Figure 2b), the same pattern of bands was obtained, but they were of different relative intensities.

No restriction-fragment-length polymorphisms were obtained between the fast- and slow-acetylator strains in either the *Hin*dIII or *Eco*RI digests in contrast with the published report of Martell et al. (1991). Bands of 2.7 kb were present in the *Hin*dIII digests of both strains, whereas Martell et al. (1991) obtained a band at 2.8 kb in their A/J strain and the smallest band they observed in the C57Bl/6J strain was 4.4 kb. Also, while the *Eco*RI digest illustrated in Figure 2 shows two bands present for the C57Bl/6J strain, Martell et al. (1991) obtained only one. The reason for these differences in the Southern blot of the C57Bl/6J DNA is uncertain but could be due to differences in the source of the animals.

Comparison of three distinct arylamine *Nat* clones

The three distinct *Nat* clones identified from 13 positive isolates obtained by screening the genomic DNA library are illustrated in Figure 3.

The sequence of the open reading frame of *Nat*-1 from Balb/c mice matches that of the published A/J and C57Bl/6 sequences except for a 'TC' to 'CT' inversion in Balb/c at nucleotide positions 74 and 75. This results in an amino acid substitution such that isoleucine in the published sequence (Martell et al., 1991) is replaced by threonine at amino acid residue 25 (Figure 4). However, DNA sequencing over this region in both directions in the A/J strain (results not shown) also gave the same 'CT' inversion in the nucleotide sequence. A comparison of the

```
   1  TCTAGATTTCTAGTTTCTGATACTTGGATGATAGAATAAATGAATAAACTTTCTCTCTA
  60  GGTTCACAGACACTGTTAAGCAGTTTTAAGTAATTATTATGACAGGTATTTAAAGCTAAT
 120  TTTGTATTTGTGTGTTTACATTTTCTCTTAAATAACTTTAAGAGATAAATAACCTTTATA
 180  AGGAACTCCAAGTGCAGATAGTAACATTTGACTGTTTTGTTTCCCTTGCCTTAGGGGACC

        M  D  I  E  A  Y  F  E  R  I  G  Y  Q  K  S  S  N  K  L  D
 240  ATGGACATTGAAGCATACTTTGAAAGAATTGGTTACCAGAAGTCCAGCAACAAACTGGAC

        L  Q  T  L  T  E  I  L  Q  H  Q  I  R  A  I  P  F  E  N  L
 300  TTGCAGACATTAACTGAAATCCTTCAGCATCAGATACGGGCTATTCCTTTTGAGAACTTG

        N  I  H  C  G  K  T  M  E  L  S  L  E  D  T  F  H  Q  I  V
 360  AACATCCATTGTGGGAAAACCATGGAACTGAGCTTAGAGGACACCTTTCATCAAATTGTG

        R  K  K  R  G  G  W  C  L  Q  V  N  H  L  L  Y  W  A  L  A
 420  AGGAAGAAGCGGGGAGGGTGGTGTCTTCAAGTCAACCATCTTCTCTACTGGGCTCTGGCC

        M  I  G  F  E  T  T  M  L  G  G  C  V  Y  V  P  S  A  C  K
 480  ATGATAGGGTTTGAGACCACAATGTTAGGAGGATGTGTTTATGTCCCTTCAGCCTGCAAG

        Y  S  N  T  M  I  H  L  L  L  Q  V  T  I  S  G  K  T  Y  I
 540  TATAGTAACACTATGATACACCTTCTACTACAGGTGACCATCAGTGGCAAAACATATATT

        V  D  S  A  F  P  F  S  C  Q  L  W  E  P  L  E  L  T  S  G
 600  GTAGATAGTGCATTCCCATTTTCCTGCCAGCTATGGGAGCCTCTGGAGTTGACATCTGGG

        K  D  Q  P  Q  V  P  A  I  F  H  L  R  E  E  N  G  T  W  Y
 660  AAGGATCAGCCTCAGGTTCCTGCCATCTTCCACCTGAGAGAAGAGAATGGAACCTGGTAC

        L  E  Q  T  K  R  Q  E  Y  V  S  N  Q  E  F  I  D  S  N  F
 720  CTGGAACAAACTAAAAGACAAGAATATGTTTCAAACCAAGAATTCATTGATTCTAATTTT

        L  E  K  N  T  H  R  K  I  Y  S  F  T  L  E  P  R  T  I  E
 780  CTTGAGAAGAACACACATCGAAAAATATATTCTTTTACTCTTGAACCACGAACGATTGAA

        D  F  W  S  I  S  T  Y  Y  Q  V  S  R  T  S  V  M  T  N  T
 840  GATTTCTGGAGTATAAGTACATACTACCAGGTATCTCGAACATCTGTGATGACAAACACA

        S  L  C  S  L  H  T  K  D  G  V  H  G  L  M  G  T  I  L  A
 900  TCACTTTGTTCCCTACATACCAAAGACGGAGTCCATGGCTTAATGGGCACCATTCTTGCC

        Y  K  K  F  N  Y  K  D  N  I  D  L  V  E  F  K  T  L  K  E
 960  TATAAGAAGTTCAATTATAAGGACAATATAGATCTGGTAGAGTTTAAGACTCTGAAGGAA

        E  E  I  E  E  V  L  K  S  V  F  G  I  H  L  E  T  K  L  V
1020  GAAGAAATAGAAGAAGTCCTGAAGAGTGTTTTTGGAATTCACTTGGAGACAAAGCTTGTG

        P  K  C  G  N  V  F  F  T  I  *
1080  CCCAAATGTGGCAATGTCTTTTTTACTATTTAGAGTAATAAGGGGAAATGATCTTTAATG
1140  TTTCTATATGCACTTATTCTTTCATGAAAAACAATCAAATTATGCAAAATTAGGTGACAG
1200  ACTGAGAAGCCTTGAGATATCATCACAAAGTTCATCAATGATTGACTGGTATATCAACAT
1260  CTGTCCCCTACATTATGGTAAGAAAATCATAGACATCAAATAGCTCTTTCTGCAAAATTA
1320  ACCATTTATGTAACTAAATGACTTTTGAAGAAATGTGCCATTAAAAATGTAATCGCAATA
1380  AAAGTGCTTTAATGATGATATAGTTGTCTTGATTTATACTATAAAACCCTCCATGTT
```

Figure 5 Nucleotide sequence of *Nat*-3 from Balb/c mice

The deduced amino acid sequence of the open reading frame is shown using the one-letter code. A possible polyadenylation signal is underlined in the 3′ non-coding region.

published amino acid sequences for chicken (Ohsako et al., 1988; Ohtomi et al., 1989), human (Blum et al., 1990a) and rabbit (Blum et al., 1990b,c) NATs, as well as those from mouse (Figure 4) shows a highly conserved threonine at residue 25. It therefore seems likely that there is also a threonine residue at this position in mouse *Nat*-1.

The open reading frame of *Nat*-2 in Balb/c mice is identical with that of the fast-acetylating C57Bl/6 mice. There are two allelic forms of *Nat*-2 which differ by a single point mutation at nucleotide position 296. This nucleotide is 'A' in *Nat*-2 from Balb/c and C57Bl/6 mice but is 'T' in A/J mice. This nucleotide change leads to substitution of asparagine at position 99 in Balb/c and C57Bl/6 mice for isoleucine in A/J mice. Expression of *Nat*-2 in COS-1 cells (Martell et al., 1992; Levy et al., 1992), has demonstrated that this single amino acid substitution is sufficient to change the activity of the expressed enzyme so that with asparagine at position 99 the $V_{max.}$ for *p*-aminobenzoic acid is 1.5-fold greater than that with isoleucine at position 99. Furthermore, the expressed enzyme from C57Bl/6 mice is 15-fold more stable at 37 °C than the enzyme from A/J mice. On this basis it was suggested that the single point mutation was responsible for determining the acetylator phenotype of the mouse strain. Balb/c mice would therefore be expected to be fast acetylators and this is in agreement with previous pharmacokinetic studies with Balb/c mice (Glowinski and Weber, 1982).

A more detailed restriction map of the fragment of the clone encoding NAT-3 from Balb/c mice is also illustrated in Figure 3 and the nucleotide sequence of this region (available from the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X72959) is shown in Figure 5. The nucleotide sequence of *Nat*-3 contains an open reading frame of 870 bp (from nucleotide 240 to 1109) and a putative polyadenylation signal has been identified in the 3′ non-coding region. The open reading frame shows 74.4% identity with that of mouse *Nat*-1 and 77.8% identity with *Nat*-2. The sequence similarity between *Nat*-3 and both *Nat*-1 and *Nat*-2 is less than *Nat*-1 compared with *Nat*-2 which show 83.9% identity over the sequence corresponding to the coding region. The deduced amino acid sequence of the open reading frame of the third *Nat* gene from Balb/c mice has been compared with other mammalian NAT sequences (Table 2). The two rabbit NAT proteins show only 13 amino acid differences, but the mouse proteins are less similar (Figure 4). The deduced amino acid sequences of the Balb/c *Nat*-1 and *Nat*-2 genes are 82% identical, NAT-3 is only 68% identical with NAT-1 and 74% identical with NAT-2. However, the cysteine residues, which are highly conserved in other NAT proteins from rabbit (Blum et al., 1990b,c), human (Blum et al., 1990a), mouse (Martell et al., 1991) and chick (Ohsako et al., 1988; Ohtomi et al., 1989), at positions 44, 68 and 223 are present in NAT-3. The region around the active-site cysteine at residue 68 (Dupret and Grant, 1992) is particularly well conserved within a region of 21 identical amino acid residues among the three mouse proteins (Figure 4).

Independent amplification of *Nat*-1, *Nat*-2 and *Nat*-3

The coding region of each of the three mouse *Nat* genes can be amplified by the PCR using distinct oligonucleotide primers (see Table 1). The primer pair Mus (12 and 13) amplifies *Nat*-1, Mus (12 and 14) amplifies *Nat*-2 and Mus (12 and 15) amplifies *Nat*-3. Within the region amplified, *Nat*-1 is cleaved by the restriction endonuclease *Ava*I but *Nat*-2 and *Nat*-3 are not. Only *Nat*-2 contains a restriction endonuclease site for *Pst*I and *Nat*-3 is cleaved specifically by *Hae*III. The specificity of the PCR and the cleavage of the amplified DNA is illustrated in Figure 6. This method of amplifying the DNA has been used to show that all three mouse *Nat* genes are present in the A/J, Balb/c and C57Bl/6 mice (results not shown) and can be used to identify the *Nat* genes in various other strains of mice and mouse cell lines. The method is also being used to amplify cDNA from specific tissues and cell lines in order to determine the expression of the *Nat* genes.

Expression of mouse *Nat* genes in *E. coli*

The enzymic activity of the expressed protein in the *E. coli* strain BL21(DE3) was assessed with four arylamine substrates: AN, 2-AF, PABA and SMZ. A control culture of *E. coli* BL21(DE3) containing the vector alone had no detectable arylamine activity.

Cultures of *E. coli* containing pET-5a with either *Nat*-1 or *Nat*-2 inserted into the T7 transcription site were induced with IPTG as described in the Materials and methods section. The best substrate for the NAT-1 activity was 2-AF, followed by AN; but no activity was detected even after 1 h incubation with either PABA or SMZ (Figure 7a).

The NAT-2 protein had detectable levels of acetylating activity with all four substrates. The catalytic activity was particularly high with 2-AF, 2.5-fold less active with PABA and 13-fold less active with AN as substrate than with 2-AF (Figure 7b). The NAT-2 protein also had low but detectable enzymic activity with

Table 2 Percentage similarity of mouse NAT-3 in a 290 amino acid overlap with other mammalian NAT proteins

Type of NAT		Similarity (%)	Reference
Mouse	NAT-1	67.6	Martell et al. (1991)
	NAT-2	73.1	Martell et al. (1991)
Human	NAT-1	68.3	Blum et al. (1990a)
	NAT-2	65.9	Kelly and Sim (1991)
Rabbit	NAT-1	62.8	Blum et al. (1990b)
	NAT-2	61.4	Blum et al. (1990c)

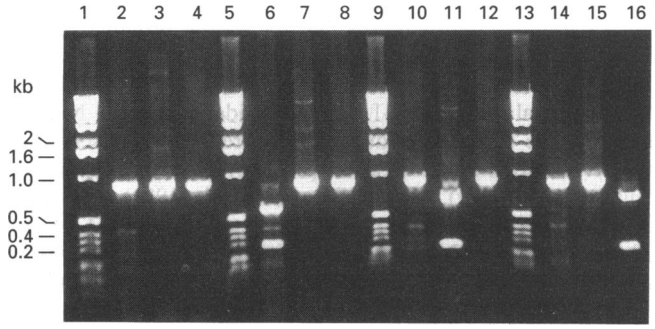

Figure 6 Amplification of *Nat*-1, *Nat*-2 and *Nat*-3 using PCR

Genomic DNA from Balb/c mice was used as template and primers were as follows: tracks 2, 6, 10 and 14, primers 12 and 13; tracks 3, 7, 11 and 15, primers 12 and 14; tracks 4, 8, 12 and 16, primers 12 and 15. The amplification products are shown before (tracks 2–4) and after digestion with *Ava*I (tracks 6–8), *Pst*I (tracks 10–12) or *Hae*III(tracks 14–16). The primers are as described in Table 1. Molecular-mass markers are in tracks 1, 5, 9 and 13.

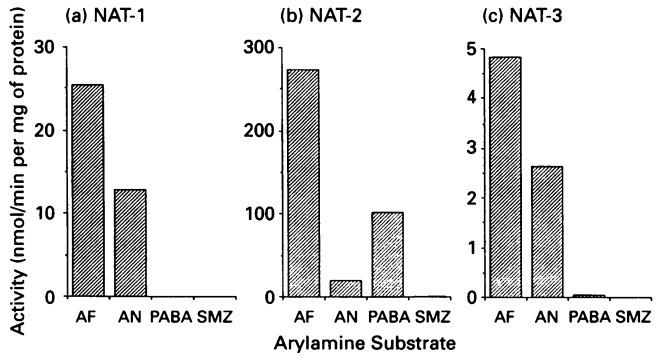

Figure 7 Arylamine NAT activity associated with (a) NAT-1, (b) NAT-2 and (c) NAT-3

The active protein was present in the *E. coli* lysate either after induction for 4 h with IPTG (**a** and **b**) or without induction (**c**). Enzymic activity was determined as described in the Materials and methods section. The substrates used are AN (anisidine), 2-AF (2-aminofluorene), PABA (*p*-aminobenzoic acid) and SMZ (sulphamethazine).

SMZ as substrate (approx. 1000-fold less than with 2-AF). Martell et al. (1992) could not detect any NAT-2 catalytic activity with SMZ as substrate following expression of *Nat*-2 in COS-1 cells. It is possible that NAT-2 is produced more efficiently in the *E. coli* expression system than in COS-1 cells.

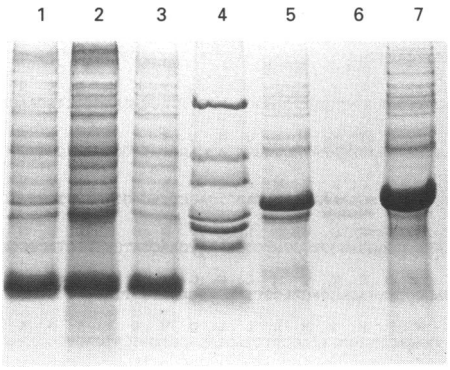

Figure 8 SDS/PAGE of proteins expressed in *E. coli*

The tracks contain soluble (tracks 1–3) and insoluble (tracks 5–7) protein from lysates of *E. coli* transformed with pET-5a into which *Nat*-1 (tracks 1 and 5), *Nat*-2 (tracks 2 and 6) or *Nat*-3 (tracks 3 and 7) has been inserted as illustrated in Figure 1. The growing cultures were induced with IPTG for 4 h. Details of the preparation of the supernatant and of the inclusion body preparation are given in the Materials and methods section. The molecular-mass markers (track 4) are 66, 45, 36, 29, 24, 20 and 14.2 kDa (Dalton Mark VII-L, Sigma).

High expression of recombinant proteins in *E. coli* frequently leads to the formation of insoluble aggregates of protein (or inclusions) which can be separated from crude cell lysates by centrifugation. No NAT activity with any of the substrates was detected after induction with IPTG of the culture containing recombinant *Nat*-3, and it was found that although the NAT-3 protein was highly expressed, it was present within inclusion bodies (Figure 8) as was much of the protein expressed from *Nat*-1. In contrast, very little of the expressed NAT-2 protein was within inclusion bodies and a faint band of protein at 31 kDa, likely to be due to NAT-2, was detectable on the gel within the supernatant of the *E. coli* lysate. All the expressed proteins gave products of approx. 31 kDa compared with the expected size of 33.7 kDa determined from each of the deduced amino acid sequences. However, because of the variation in the proportion of the expressed protein that was soluble and therefore active, it is not possible to compare quantitively the specific activities of the different NATs for any one substrate, but the substrate specificities of the individually expressed proteins can be examined.

The basal level of T7 RNA polymerase activity is sufficient to promote transcription of the target gene in the uninduced cell. Therefore, when a culture of *E. coli* BL21(DE3) containing the *Nat*-3 gene in pET-5a was grown at 20 °C for 16 h without induction, enough soluble NAT3 protein was obtained in the cell lysate to demonstrate that *Nat*-3 encodes a protein with NAT catalytic activity. No enzymic activity was detected with SMZ, but it is not possible to say that SMZ is not a substrate for the enzymic activity of NAT-3 because there was so little soluble protein. However, NAT-3 did have catalytic activity with 2-AF as substrate, 2-fold less activity with AN as substrate and barely detectable activity with PABA (Figure 7c).

Although SMZ is acetylated polymorphically in mice (Tannen and Weber, 1980), studies have shown that it is a poor substrate for liver NAT (Glowinski and Weber, 1982). Northern-blot analysis has demonstrated that mRNA for both NAT-1 and NAT-2 is present in the liver of mice (Martell et al., 1992; Levy et al., 1992) but whether the low activity of NAT-2 with SMZ is sufficient to account for the acetylation of this substrate in mice or whether NAT-3 is the enzyme responsible awaits further investigation.

We thank the Wellcome Trust for financial support, Ms A. Ward for assaying the NAT proteins with 2-AF, Dr. D, Hickman for help with Figure 1, Dr. J. R. Bright for submitting the nucleotide data to the database and Mrs A. Watson for technical assistance.

REFERENCES

Bell, D. A., Taylor, J. A., Butler, M. A., Stephens, E. A., Wiest, J., Brubaker, L. H., Kadlubar, F. F. and Lucier, G. W. (1993) Carcinogenesis **14**, 1689–1692

Blum, M., Grant, D. M., Demierre, A. and Meyer, U. A. (1989a) Proc. Natl. Acad. Sci. U.S.A. **86**, 9554–9557

Blum, M., Grant, D. M., Demierre, A. and Meyer, U. A. (1989b) Nucleic Acids Res. **17**, 3589

Blum, M., Grant, D. M., McBride, W., Hein, M. and Meyer, U. A. (1990a) DNA Cell Biol. **9**, 193–203

Blum, M., Heim, M. and Meyer, U. A. (1990b) Nucleic Acids Res. **18**, 5287

Blum, M., Heim, M. and Meyer, U. A. (1990c) Nucleic Acids Res. **18**, 5295

Blum, M., Demierre, A., Grant, D. M., Heim, M. and Meyer, U. A. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 5237–5241

Bradford, M. (1976) Anal. Biochem. **72**, 248–254

Deguchi, T. (1993) Biomed. Res. **13**, 231–242

Dupret, J.-M. and Grant, D. M. (1992) J. Biol. Chem. **267**, 7381–7385

Evans, D. A. P., Manley, K. A. and McKusick, V. A. (1960) Br. Med. J. **2**, 485–491

Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6–13

Glowinski, I. and Weber, W. W. (1982) J. Biol. Chem. **257**, 1424–1430

Grant, D. M. (1993) Pharmacogenetics **3**, 45–50

Harlow, E. and Lane, D. (1988) Antibodies: A Laboratory Manual, pp. 90–91, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Received 26 November 1993/22 March 1994; accepted 5 April 1994

Hickman, D., Risch, A., Buckle, V., Spurr, N. K., Jeremiah, S. J., McCarthy, A. and Sim, E. (1994) Biochem. J. **297**, 441–445

Hickman, D., Risch, A., Camilleri, J. P. and Sim, E. (1992) Pharmacogenetics **2**, 217–226

Kelly, S. L. and Sim, E. (1991) Human Exp. Toxicol. **10**, 33–38

Kelly, S. L. and Sim, E. (1992) Biochem. Soc. Trans. **20**, 106S

Levy, G. N., Martell, K. J., DeLean, J. H and Weber, W. W. (1992) Pharmacogenetics **2**, 197–206

Martell, K. J., Vatsis, K. P. and Weber, W. W. (1991) Mol. Pharmacol. **40**, 218–227

Martell, K. J., Levy, G. N. and Weber, W. W. (1992) Mol. Pharmacol. **42**, 265–272

Noriander, J., Kempe, T. and Messing, J. (1983) Gene **26**, 101–106

Ohsako, S. and Deguchi, T. (1990) J. Biol. Chem. **265**, 4630–4634

Ohsako, S., Ohtomi, M., Sakamoto, Y., Uyemura, K. and Deguchi, T. (1988) J. Biol. Chem. **263**, 7534–7538

Ohtomi, M., Sasaki, M. and Deguchi, T. (1989) Eur. J. Biochem. **185**, 253–261

Porteous, D. J. (1985) Somatic Cell Mol. Gen. **11**, 445–454

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sasaki, Y., Ohsako, S. and Deguchi, T. (1991) J. Biol. Chem. **266**, 13243–13250

Schägger, H. and von Jagow, G. (1987) Anal. Biochem. **166**, 368–379

Sim, E., Hickman, D., Coroneos, E. and Kelly, S. L. (1992) Biochem. Soc. Trans. **20**, 304–309

Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990) Methods Enzymol. **185**, 60–69

Tannen, R. H. and Weber, W. W. (1980) J. Pharmacol. Exp. Ther. **213**, 480–484

Vatsis, K. P. and Weber, W. W. (1993) Arch. Biochem. Biophys. **301**, 71–76

Vatsis, K. P., Martell, K. J. and Weber, W. W. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 6333–6337

Ward, A., Hickman, D., Gordon, J. W. and Sim, E. (1992) Biochem. Pharmacol. **44**, 1099–1104

Weber, W. W. and Hein, D. W. (1985) Pharmacol. Rev. **37**, 25–79