Eubacterial arylamine \(N\)-acyetyltransferases – identification and comparison of 18 members of the protein family with conserved active site cysteine, histidine and aspartate residues

Mark Payton,¹ Adeel Mushtaq,¹ Tin-Wein Yu,² Ling-Juan Wu,³ John Sinclair⁴ and Edith Sim¹

Author for correspondence: Edith Sim. Tel: +44 1865 271596. Fax: +44 1865 271853.
e-mail: esim@molbiol.ox.ac.uk

¹ Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK
² Department of Chemistry, Box 351700, University of Washington, Seattle, WA 98195-1700, USA
³ Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK
⁴ The Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Arylamine \(N\)-acyetyltransferases (NATs) are enzymes involved in the detoxification of a range of arylamine and hydrazine-based xenobiotics. NATs have been implicated in the endogenous metabolism of p-aminobenzoyl glutamate in eukaryotes, although very little is known about the distribution and function of NAT in the prokaryotic kingdom. Using DNA library screening techniques and the analysis of data from whole-genome sequencing projects, we have identified 18 nat-like sequences from the Proteobacteria and Firmicutes. Recently, the three-dimensional structure of NAT derived from the bacterium Salmonella typhimurium (PDB accession code 1E2T) was resolved and revealed an active site catalytic triad composed of Cys\(^{69}\)-His\(^{107}\)-Asp\(^{122}\). These residues have been shown to be conserved in all prokaryotic and eukaryotic NAT homologues together with three highly conserved regions which are found proximal to the active site triad. The characterization of prokaryotic NATs and NAT-like enzymes is reported. It is also predicted that prokaryotic NATs, based on gene cluster composition and distribution amongst genomes, participate in the metabolism of xenobiotics derived from decomposition of organic materials.

Keywords: gene cluster, NAT, prokaryotic, endogenous function

INTRODUCTION

Arylamine \(N\)-acyetyltransferases (NATs; EC 2.3.1.5) are cytosolic enzymes of approximately 30 kDa found in almost all multicellular eukaryotes (Sim et al., 2000). NAT was first identified in humans in the 1960s by its ability to inactivate the anti-tubercular drug isoniazid (INH) (Evans et al., 1960). This inactivation differs amongst individuals due to point mutations within the ORF which result in non-conservative amino acid changes and distinct phenotypic variants (Deguchi et al., 1990; Blum et al., 1992; Hickman et al., 1995). NAT can exist as multiple isoenzymes and in humans there are two isoforms which have distinct, but overlapping substrate specificities (Blum et al., 1990; Ohsako & Deguchi, 1990; Vatsis et al., 1995). NAT is known to inactivate arylamine and hydrazine-based xenobiotics by the transfer of the acetyl group from acetyl-CoA to the terminal nitrogen atom (Weber & Hein, 1985). Carcinogens, such as 2-aminofluorene and benzidine, are also activated by NAT through the O-acetylation of hydroxylamine metabolites to form reactive acetoxy esters (Hanna, 1996; King et al., 1997). As well as xenobiotic metabolism, NAT (human NAT1 and its functional homologues) has been implicated in the endogenous metabolism of the folate breakdown product p-aminobenzoyl glutamate (Minchin, 1995; Payton et al., 1999; Ward et al., 1995).

The Ames tester strains of Salmonella typhimurium (TA98 and XG1024) are used to predict the mutagenic properties of aromatic amines and nitroarenes. Sensitivity to these compounds is through \(N\)-hydroxylamine O-acetyltransferase activity and was the first
identification of bacterial NAT (McCoy et al., 1983; Watanabe et al., 1992). Since then nat, and associated acetylation activity, has been found in the Mycobacterium smegmatis and Mycobacterium tuberculosis (Payton et al., 1999). The pathogenic mycobacterium M. tuberculosis is the main cause of tuberculosis in man (http://www.who.int/gtb/index.htm) and the choice of therapy is often INH. An increased production of NAT in mycobacteria results in an increased tolerance to INH (Payton et al., 1999). The study of NAT has once again become associated with the anti-tubercular agent INH. NAT activity has also been reported in Escherichia coli (Chang & Chung, 1998).

NAT is highly conserved in the eukaryotic kingdom and this study was undertaken to determine how widespread NAT is in the prokaryotic kingdom and to what extent conserved residues are maintained. Genes transcribed in bacteria often appear as related operons. Characterization of these bacterial operons may lead to a better understanding of the endogenous function of NATs in prokaryotes.

METHODS

Bacterial strains and culture. The bacterial strains Escherichia coli BL21(DE3) pLysS and JM109 (Promega), Salmonella typhimurium LT2 and SGSC1412 (http://www.tigr.org/db/mdb/mdb.html) and Bacillus subtilis 168 (Kunst et al., 1997) and 6633 (Difco) were grown in standard LB medium (Sambrook et al., 1989). Mycobacterium smegmatis was grown in 7H9 plus ADC (Parish & Stoker, 1998) containing the appropriate antibiotics for plasmid selection. To test NAT induction by substrate, bacterial and mycobacterial cultures were grown until mid-exponential phase before diluting 1:10 into fresh medium containing the appropriate concentration of INH.

Isolation and cloning of prokaryotic nat. The N-terminal 175 aa of Sal. typhimurium NAT was used to identify NAT-like sequences from forward and reverse translations of bacterial genomes with electronic database searches (http://www.tigr.org/db/mdb/mdb.html) using BLAST and FASTA3 programs (http://www.ebi.ac.uk/ fasta3/). Sequences representative of the mycobacteria M. smegmatis and M. tuberculosis were determined as described previously (Payton et al., 1999). DNA representing the ORF of nat from all described prokaryotes was amplified by PCR using pfu DNA polymerase (Stratagene) supplemented with 6% DMSO for the GC-rich mycobacterial genomes. Products were ligated into PCR-Script (Stratagene) for sequence analysis and were subcloned to yield cleavable histidine tags in the vectors pET28b (2.3 kDa tag) or pRSET (3-1 kDa tag) (Invitrogen) for protein expression. The mycobacterial nat sequences were also subcloned into the mycobacterial expression vector pACE-1 (Parish et al., 1997).

The endogenous copies of M. smegmatis and B. subtilis nat were disrupted using either kanamycin or β-galactosidase cassettes inserted 5’ to the first region of conservation (Fig. 1) using technology described by Hinds et al. (1999) and Sambrook et al. (1989).

Recombinant protein production. Soluble recombinant NAT derived from E. coli, B. subtilis and Amycolatopsis mediterranei was produced containing a cleavable N-terminal hexahistidine tag. Recombinant protein was obtained using the expression host E. coli BL21(DE3) pLysS in LB medium containing 1 M sorbitol and 2.5 mM betaine. A 50 ml culture was inoculated at 1:10 dilution from an overnight culture and grown at 27 °C with shaking until the optical density at 600 nm reached 0.6. Protein expression was induced with 0.1 mM IPTG and the bacteria were grown for a further 4 h with shaking. In all cases cells were pelleted at 3000 g for 10 min, resuspended in 1/10 of the culture volume of 50 mM Tris/HCl (pH 8.0), 4 mM dL-DTT and 1 mM Pefabloc SC (protease inhibitor; Pentapharm) and snap frozen in liquid nitrogen. The cell suspension was thawed at 30 °C and subjected to sonication on wet ice for 2 min, with 1 min intervals, using a microprobe (Soniprep 150; MSE) set to 7 microns. Cell debris was pelleted at 20000 g for 30 min and the soluble fraction collected. Pure recombinant NAT was isolated from soluble bacterial lysates as described by Sinclair et al. (1998). Protein purity was determined by SDS-PAGE and protein was quantified using the Bradford assay (Sigma). The heterologous expression of nat from Sal. typhimurium was as described by Sinclair et al. (1998).

Recombinant protein encoded by M. smegmatis and M. tuberculosis nat was obtained using the mycobacterial expression vector pACE-1 and the host M. smegmatis as described previously (Payton et al., 1999).

Western blot analysis. Protein samples were separated by SDS-PAGE and transferred to a nylon support (Roche) for Western blot analysis. Antisera were raised in rabbits and generated against purified recombinant NAT proteins isolated in SDS-PAGE gel slices representative of Sal. typhimurium NAT. Unless otherwise stated, antisera were used at a dilution of 1:100000. Bound rabbit immunoglobulin was detected by chemiluminescence using mouse anti-rabbit IgG mAbs (used at a dilution of 1:12000) conjugated to horseradish peroxidase (Sigma). Equal standardized amounts of soluble protein were added in all cases. The protein loaded onto gels was prepared from the same number of cells as determined by optical density at 600 nm prior to lysis to ensure equal loading.

Enzymic assays. All enzymic reactions were carried out under conditions where the initial rate was linear. Michaelis constants and specific activities for N-acetyltransferase were measured from soluble cell fractions and pure recombinant protein preparations as described previously (Sinclair et al., 1998). Acetyl-CoA (440 μM) with up to five times the $K_m$ concentration of arylamine or hydrazine substrate were used. Loss of substrate was indicative of enzyme activity and was determined either colorimetrically (Sinclair et al., 1998) or, in the case with INH, by fluorometric detection following appropriate dilutions of the reaction mixtures (Ellard & Gammon, 1976).

Computer analysis. Database searches were performed with the programs FASTA and BLAST using a cut-off value of greater than $1 \times 10^{-4}$ (http://www.ebi.ac.uk/ fasta3/; http://www.hgsc.bcm.tmc.edu/ SearchLauncher/). ORF prediction and manipulation was executed using DNA Strider (DNA Strider 1.2.1, CEA France) and NIX (http://www.hgmp.mrc.ac.uk/) in combination with homology searches (http://www.hgsc.bcm.tmc.edu/ SearchLauncher/). Alignment of predicted amino acid sequences and homology shading was performed using the European Bioinformatics Institute server (http://www.ebi.ac.uk/fasta3/) and MacBoxshade (http://www.netaxis.com/~jayfar/mops.html).

RAP isolation and Northern analysis. Total RNA was prepared from M. smegmatis, E. coli and B. subtilis using
Table 1. Oligonucleotide primers used to amplify coding regions of prokaryotic and human NATs

<table>
<thead>
<tr>
<th>Source of nat (coding region)</th>
<th>5' sequence (5'-3')</th>
<th>3' sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human NAT2 (873 bp)</td>
<td>TACGAGTGCATATGCATGGCATGTAAGGGATCCCATCACC</td>
<td>ATGGCCTTTACTGGCCGCTTAAATAGTAAGGGATCCATCACC</td>
</tr>
<tr>
<td>Human NAT1 (873 bp)</td>
<td>TACGAGTGCATATGCATGGCATGTAAGGGATCCATCCT</td>
<td>ATGGCCTTTACTGGCCGCTTAAATAGTAAGGGATCCATCCT</td>
</tr>
<tr>
<td>E. coli (846 bp)</td>
<td>TACGAGTGCATATGCATGGCATGTAAGGGATCCATCCT</td>
<td>ATGGCCTTTACTGGCCGCTTAAATAGTAAGGGATCCATCCT</td>
</tr>
<tr>
<td>B. subtilis (756 bp)</td>
<td>TACGAGTGCATATGCATGGCATGTAAGGGATCCATCCT</td>
<td>ATGGCCTTTACTGGCCGCTTAAATAGTAAGGGATCCATCCT</td>
</tr>
<tr>
<td>M. smegmatis (828 bp)</td>
<td>TACGAGTGCATATGCATGGCATGTAAGGGATCCATCCT</td>
<td>ATGGCCTTTACTGGCCGCTTAAATAGTAAGGGATCCATCCT</td>
</tr>
<tr>
<td>M. tuberculosis (852 bp)</td>
<td>TACGAGTGCATATGCATGGCATGTAAGGGATCCATCCT</td>
<td>ATGGCCTTTACTGGCCGCTTAAATAGTAAGGGATCCATCCT</td>
</tr>
<tr>
<td>Amy. mediterranei (787 bp)</td>
<td>TACGAGTGCATATGCATGGCATGTAAGGGATCCATCCT</td>
<td>ATGGCCTTTACTGGCCGCTTAAATAGTAAGGGATCCATCCT</td>
</tr>
</tbody>
</table>

Catrimox-14 (Iowa Biotechnology) as described previously (Payton & Pinter, 1999). RNA (20 μg) was size-fractionated by electrophoresis on a formaldehyde denaturing 0.9% agarose gel (Sambrook et al., 1989) followed by pre-treatment with 0.05 M NaOH to hydrolyse RNA and ensure effective transfer of large transcripts to the nylon filter (Roche). RNA markers (Promega) were treated in an identical manner to the RNA samples. Probes representing the corresponding bacterial nat ORFs (see Table 1 for primers used) were generated by PCR amplification using a molar ratio of 1:1 dTTP/DIG-labelled dUTP (Roche) to ensure effective detection of rare transcripts. Nylon filters were probed for 16 h at 50 °C in DIG Easy Hyb buffer (Roche), then washed twice for 5 min in 2× SSC/0.1% SDS (room temperature) and twice for 15 min in 0.1× SSC/0.1% SDS (50 °C). Signal detection was according to the manufacturer’s recommendations and filters were exposed to autoradiographic film (Sigma) for between 20 min and 2 h.

RESULTS

Identification of prokaryotic nat

Eighteen prokaryotic NAT-like sequences, including the partial sequence representative of Legionella pneumophila, were identified (Fig. 1). Whilst NAT is retained across the animal kingdom (Delomene et al., 1997; http://athena.louisville.edu/medschool/pharmacology/NAT.html), the distribution of prokaryotic nat is so far restricted to the Proteobacteria and Firmicutes, although representative members of all phyla have already been sequenced (Fig. 2) (http://www.tigr.org/tdb/mbd/mbd.html).

Homologous regions within prokaryotic NAT

Sequence comparison of eukaryotic and prokaryotic NAT demonstrates a higher degree of conservation towards the N termini of these proteins compared to the C-terminal 100 aa (Fig. 1). Of the previously described key residues, Arg9, Arg56 and Cys48 (Delomene et al., 1997; Watanabe et al., 1992) in the human NAT isoenzymes, Cys48 is the only acid to be completely retained in the prokaryotic homologues. Structural predictions and crystallographic analysis of Sal. typhimurium NAT previously revealed the existence of three domains (Fig. 1) and an active site catalytic triad (asterisks; Fig. 1) consisting of Cys89, His107, Asp122 (Hubbard et al., 1996; Sinclair et al., 2000). A comparison of the amino acid sequences representative of NATs from the Proteobacteria and Firmicutes (Table 2) reveals high conservation in domains 1 and 2 (38–52%), with a marked decrease in domain 3 (<20–42%).

An additional loop is present in all eukaryotic NATs on comparison with the prokaryotic isoenzymes (dashed line; Fig. 1). Alignment of the amino acid sequences representative of prokaryotic NAT-like enzymes reveals three highly conserved regions (boxed; Fig. 1). The regions of conservation consist of (i) I/VPFENLxx, (ii) RGGfCdrxxxf and (iii) DaGdjj (using the class 1 alphabet), corresponding to residues 36–43, 65–75 and 122–127 of Sal. typhimurium NAT, respectively. Crystallographic analysis demonstrates that these regions are found within the putative substrate cleft in close association with the active site Cys (Fig. 3). Nine residues are conserved throughout all species, including the eukaryotes (Leu2, Gly46-47, Cys48, Leu57, Gly58, His59, Asp122 and Gly123; numbering with respect to the Sal. typhimurium NAT sequence) and are predominately clustered around the three regions of conservation defined as (i)–(iii) (Fig. 1). A highly conserved Gly126 is absent in the putative nat sequences of Actinosynnema pretiosum, Amy. mediterranei, Streptomycetes achromogenes and Bacillus anthracis (†; Fig. 1).

NAT sequences representative of the Bacillus/ Clostridium group (Firmicutes; Fig. 2) are highly similar to those determined for the Proteobacteria (Table 2). There appear to be two distinct prokaryotic isoforms of the NAT enzyme: one isoenzyme is maintained in many of the members of the Proteobacteria for which sequence data are available (Figs 1 and 2), but in the Actinomycetales of the Firmicutes, the NAT-like sequences appear to fall into two subcategories. Corynebacterineae (mycobacteria) NATs retain all three regions associated with the well characterized eukaryotic isoform, whereas Pseudonocardineae (Amy. mediterranei and Act. pretiosum) and Streptomycineae (Str. achromogenes) NATs display a marked decrease in similarity in the first region of conservation (Fig. 1;
M. PAYTON and OTHERS

Fig. 1. Alignment of the primary amino acid sequence of 18 predicted and characterized prokaryotic NATs compared to the human NAT1 and NAT2 isoforms. The three NAT enzyme domains revealed by structural analysis are shown (domains 1–3) together with the three regions of conservation (boxed, i–iii). Highly conserved residues of the catalytic triad (Cys69-His107-Asp122) and Gly126 are also illustrated using * and †, respectively: the numbers refer to NAT from Sal. typhimurium.

A dashed line represents a loop only associated with eukaryotic NAT and a solid line represents the difference in the NAT-like enzymes which lack NAT activity. Sequences are grouped as illustrated in Fig. 2.

region i, I/VPFENLxx), although retaining the catalytic triad. An insertion of up to 8 aa is also seen in the NAT sequences of these three organisms (Fig. 1; bold line) compared to other prokaryotic NAT sequences.

Prokaryotic NAT enzymic activities

nat-like sequences were cloned from bacteria representative of the Proteobacteria (E. coli and Sal. typhimurium) and the Firmicutes [Act. mediterranei (Pseudocardineae), M. smegmatis and M. tuberculosis (Corynebacterineae) and B. subtilis (Bacillus/Clostridium group)] for recombinant protein production. Prokaryotic recombinant NAT isoenzymes were purified and the enzymic activities were determined using the well characterized eukaryotic NAT substrates p-aminobenzoic acid (p-ABA), which is a specific substrate for human NAT1, and INH, which is a specific substrate for human NAT2 (Fig. 4a; Andres et al., 1983; Hickman et al., 1995). With the substrates tested, the bacterial NATs are more similar to the human NAT2. The Km for p-ABA with all of the recombinant, purified bacterial NAT proteins from Sal. typhimurium, E. coli and M. tuberculosis was greater than 5 mM, whereas activities could readily be detected with INH with each of these enzymes. The Km for the enzyme from B. subtilis was 39 µM and so is very similar to the previously reported value for the mycobacterial NAT enzymes (Payton et al., 1999). The E. coli enzyme had a Km of 1-2 mM for INH.
The expressed and purified NAT-like sequence isolated from *Amy. mediterranei* lacked any measurable acetylation activity with either INH or *p*-ABA (Fig. 4a).

### Detection of endogenous and recombinant prokaryotic NAT

Endogenous bacterial NAT enzymic activities were found in the soluble lysates of *E. coli*, *Sal. typhimurium*, *B. subtilis* and *M. smegmatis* using INH as substrate, albeit at a low level (Table 3). No detectable acetylation of INH was observed using soluble lysates of the NAT knockout mutants of *M. smegmatis* and *B. subtilis* (Table 3). The level of endogenous activity in *E. coli* is much less than that measured by Chang & Chung (1998) and no activity was detected with *p*-ABA. The lack of endogenous activity with *p*-ABA matched the observations obtained with the pure, recombinant enzymes (see above).

Recombinant and endogenous NAT-like proteins were studied by Western blot analysis using an antiserum generated against pure recombinant *Sal. typhimurium*

---

**Fig. 2.** Distribution of NAT-like sequences within Eubacteria (excluding the Archaea) described in this report. Numbers in parentheses represent whole genomes sequenced to date.

**Table 2.** Amino acid similarities (%) between prokaryotic NAT domains 1 and 2, and domain 3

<table>
<thead>
<tr>
<th>Domain 3</th>
<th><em>S. typhimurium</em></th>
<th><em>C. crescentus</em></th>
<th><em>M. tuberculosis</em></th>
<th><em>B. subtilis</em></th>
<th><em>A. mediterranei</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Salmonella typhimurium</em></td>
<td></td>
<td>52</td>
<td>52</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td><em>Caulobacter crescentus</em></td>
<td></td>
<td>42</td>
<td>51</td>
<td>39</td>
<td>41</td>
</tr>
<tr>
<td><em>Mycobacterium tuberculosis</em></td>
<td>29</td>
<td>36</td>
<td>38</td>
<td>37</td>
<td>41</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td></td>
<td>41</td>
<td>34</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td><em>Amycolatopsis mediterranei</em></td>
<td>&lt; 20*</td>
<td>&lt; 20*</td>
<td>&lt; 20*</td>
<td>31</td>
<td></td>
</tr>
</tbody>
</table>
NAT (Payton et al., 1999). Expression of NAT in Sal. typhimurium (strain LT2) was shown to increase with time and reached a maximum by late exponential/stationary phase (Fig. 5a), which may indicate that the expression of the nat gene is regulated throughout the growth cycle. All subsequent cultures were grown to late exponential phase prior to analysis. Endogenous NAT was detected in whole-cell preparations of Sal. typhimurium (Fig. 5a) and soluble lysates of Sal. typhimurium (LT2), E. coli (JM109) and M. smegmatis (strain Mc²155) (Fig. 5b, c). Immunoblot analysis revealed a doublet with soluble and whole-cell lysates of Sal. typhimurium and E. coli, and a single band with soluble lysates of M. smegmatis (Fig. 5a–c). It is possible that the antibody could be identifying another protein in addition to NAT in lysates of Sal. typhimurium and E. coli, although it may be that there is some proteolysis of NAT. Cultures of M. smegmatis, E. coli or Sal. typhimurium were grown in increasing amounts of substrate (INH) and showed no corresponding increase in NAT production (Fig. 5b, c). Analysis of equal amounts of recombinant NAT-like proteins representative of the Proteobacteria and Firmicutes demonstrated cross-reactivity with the antiserum raised to Sal. typhimurium NAT in a manner that reflected sequence homology (Fig. 5d; Table 2). To demonstrate the weakly cross-reacting forms, a long exposure was necessary and an additional smaller molecular mass band is observed in the pure recombinant NAT preparation from M. smegmatis. This lower molecular mass band (which

![Fig. 3. A space-fill model of the crystal structure of Sal. typhimurium NAT illustrates the size of the substrate cleft and position of the C terminus cap-like domain 3 (red). The three regions of conservation line the substrate-binding pocket (turquoise) and the active site cysteine is shown (blue). The N-terminal domains 1 (purple) and 2 (yellow) are linked via an inter-domain helix (green) to the third C-terminal domain (red).](image)

![Fig. 4. Characterized reactions catalysed by NAT and NAT-like enzymes. (a) Known NAT substrates and (b) predicted ansamycin precursors cyclized by NAT-like enzymes to yield proansamycin X (Amy. mediterranei) and rubranasol (Str. achenomogenous).](image)

Table 3. A comparison of INH acetylation using whole-cell lysates

| Species       | Rate [pmol min⁻¹ (mg protein)⁻¹] ⁺  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sal. typhimurium</td>
<td>7.9 ± 0.1</td>
</tr>
<tr>
<td>E. coli</td>
<td>6.8 ± 0.1</td>
</tr>
<tr>
<td>B. subtilis</td>
<td>6.5 ± 0.1</td>
</tr>
<tr>
<td>B. subtilis (NAT KO)</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>M. smegmatis</td>
<td>10.8 ± 0.4</td>
</tr>
<tr>
<td>M. smegmatis (NAT KO)</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

⁺ NAT KO refers to strains in which the nat gene has been knocked out as described in Methods.

† Results are shown as the mean of four independent determinations ± sd.
Fig. 5. Western blot analysis of endogenous (a–c) and pure recombinant protein (d) probed with antiserum raised to Sal. typhimurium NAT. (a) Time course in which cell pellets of Sal. typhimurium were analysed from cultures at 1, 2, 3 and 4 h (lanes 1–4, respectively) after a 1:10 inoculation. The NAT band is indicated by an arrow. (b) Lanes: 1, pure recombinant Sal. typhimurium NAT; 2, soluble lysate of M. smegmatis; 3–5, representative lysates of E. coli; 6–8, Sal. typhimurium grown in the presence of 0, 0.5 and 10 mg INH ml⁻¹, respectively. (c) Pure recombinant M. smegmatis NAT, 32.4 kDa (lane 1), and Sal. typhimurium NAT, 34.4 kDa (lane 2). Lanes 3–6 are representative lysates of M. smegmatis grown in 0, 0.5, 1 and 4 µg INH ml⁻¹, respectively. (d) Pure recombinant prokaryotic NATs. Lanes: 1, E. coli (34.5 kDa); 2, Sal. typhimurium (34.4 kDa); 3, B. subtilis (30.7 kDa); 4, Amy. mediterranei (32.6 kDa); 5, M. tuberculosis (33.3 kDa); 6, M. smegmatis (32.4 kDa). All recombinant proteins include a 2.2 kDa hexahistidine purification tag. There has been some proteolytic degradation of the histidine tag giving rise to the second lower molecular mass band in lane 6.

Fig. 6. Northern blot analysis using 20 µg total RNA isolated from (a) M. smegmatis, (b) E. coli and (c) B. subtilis. RNA markers (Promega) were used to determine transcript size.

accounts for less than 10% of the total intensity in lane 6, Fig. 5d) appears on storage of the pure recombinant M. smegmatis NAT protein and is likely to be a proteolytic fragment as a result of proteolytic digestion at the thrombin cleavage site following the hexahistidine tag introduced into the recombinant protein (Sinclair et al., 1998).

Prokaryotic NATs reveal polymorphisms

Of the prokaryotic nat genes sequenced (following amplification using proof reading pfu DNA polymerase and cloning) two strains of B. subtilis (strains 168 and 6633) and Sal. typhimurium (LT2 and SGSC1412) were compared. Polymorphisms were found in both species which resulted in non-conservative mutations (Fig. 1). B. subtilis (strains 6633 → 168), Thr⁴¹ → Met, His⁶⁹ → Tyr, Arg⁶⁵ → Gly and Leu²⁵⁹ → Phe; Sal. typhimurium (strains SGSC1412 → LT2), Asn³⁰ → His, Gly⁴⁹ → Val, Ile¹⁵² → Met, Ile⁴³⁶ → Val and His⁸⁵⁹ → Ser. Since these differences have been identified in the DNA sequences of cloned variants, they are not PCR artefacts.

nat operon analysis

Transcript sizes of approximately 7, 9 and 4 kb were associated with RNA isolated from M. smegmatis, E. coli and B. subtilis, respectively (lanes a–c, Fig. 6) and corresponded well to predicted transcript sizes from whole genome analysis (Fig. 7). Transcript sizes for the Amy. mediterranei and Str. achromogenes nat-like gene clusters have been reported (Floss & Yu, 1999; Sohng et al., 1997).

The genes found within the putative nat operons of the pathogenic mycobacteria M. tuberculosis, Mycobacterium bovis and Mycobacterium avium encode highly similar enzymes. Homology searches of these predicted proteins reveal a cluster of enzymes responsible for the metabolism of aromatic and biphenyl-based compounds. These enzymes include a possible oxidoreductase (Rv3570c), 2-hydroxy-6-oxo-6-phenylhexa-2,4-dione hydrolase (biphenyl metabolism, Rv3569c), biphenyl-2,3-diol-1,2-dioxygenase (extradiol ring cleavage and biphenyl metabolism, Rv3568c), nitrolo(tri)acetate monoxygenase (Rv3567c) and NAT (Rv3566c). The codes in parentheses indicate the identification number of putative proteins in the M. tuberculosis genome (Cole et al., 1998). The genes of other bacterial operons which encode NAT (Figs 6 and 7) predominantly encode novel or hypothetical proteins with no functional database homologues.
**DISCUSSION**

Initial searches of databases for prokaryotic NATs suggest that the distribution of nat may be limited to the *Proteobacteria* and *Firmicutes* in the prokaryotic kingdom (Fig. 2). NAT consists of three domains (designated 1, 2 and 3) which are of approximately equal length (Fig. 1). The first two N-terminal domains are highly conserved in NATs throughout both the eukaryotic and prokaryotic kingdoms. The C-terminal domain (3) shows an increased sequence divergence (Fig. 1, Table 2).

Alignment of the primary amino acid sequence reveals three highly conserved regions (designated i–iii) found in all NATs (Fig. 1). In domains 1 and 2 the three highly conserved regions (i–iii) line the putative substrate pocket (Fig. 3). The third C-terminal domain in *Sal. typhimurium* NAT forms a distinct α/β lid to the molecule to create an active site cleft (Fig. 3). It was previously shown using human NAT1 that the first two domains alone were necessary to generate an acetylated intermediate and that the third domain was essential in the transfer of the acetyl group to the acceptor amine (Sinclair & Sim, 1997). Although these studies were performed with human NAT, they provide further evidence that the C-terminal domain is the key to arylamine substrate specificity.

Eukaryotic organisms have up to three isoenzymes of NAT encoded by distinct genes (Hein et al., 2000) and these multiple genes have probably arisen as a result of gene duplication. If there is only one isoform in mammals, it is the equivalent to human NAT1 which has been shown previously to acetylate p-ABA and p-aminobenzoyleglutamate (p-ABGlu) (Cribb et al., 1991; Minchin, 1995; Ward et al., 1995). The widespread distribution of this isoenzyme amongst eukaryotes may be related to its putative endogenous role in the N-acetylation of p-ABGlu (Minchin, 1995; Ward et al., 1995), a folate catabolite. Prokaryotic organisms use p-ABA to synthesize folate *de novo* and therefore the N-acetylation of p-ABA would be detrimental to the survival of a bacterium.

Acetyl-CoA is an important co-factor for numerous processes in the cell, including the synthesis of fatty acids, cholesterol, and coenzyme A (CoA). It is also a key intermediate in several metabolic pathways, such as the Krebs cycle and the citric acid cycle. In this context, the presence of NAT in prokaryotes may serve to balance the levels of p-ABA, preventing its accumulation and potential toxic effects. This balance can be crucial for the survival and metabolic efficiency of these organisms.
enzymes and binds to a variety of motifs (Engel & Wierenga, 1996). A 4 aa motif commencing with a Gly has been associated with phosphate binding to acetyl-CoA (Kinoshita et al., 1999). There is an identifiable motif in Sal. typhimurium NAT in which Gly\textsuperscript{126} is situated adjacent to the active site Cys\textsuperscript{69} (Sinclair et al., 2000). This Gly\textsuperscript{126} is highly conserved in all prokaryotic NAT enzymes identified in this report with the exception of the NAT-like enzymes of Amy. mediterranei (G\textsuperscript{126}P), Act. pretiosum (G\textsuperscript{126}A) and Str. acromogenes (G\textsuperscript{126}A).

If Gly\textsuperscript{126} is required for acetyl-CoA binding, the absence of this residue would suggest a loss in acetyl-CoA binding and thus explain the lack of acetylation activity found with the NAT-like enzyme of Amy. mediterranei (Table 3). The highly conserved regions (i)–(iii) (Fig. 1) are found to be closely associated with the active site triad within the crystal structure of Sal. typhimurium NAT (Fig. 3). In the NAT-like sequences identified in the Pseudonocardiae and Streptomycineae (Fig. 2), an addition of up to 8 aa is seen in combination with partial disruption of region (i). This may be a key factor in the recognition of much larger substrates (Fig. 4b) known for the NAT-like enzymes of Amy. mediterranei, Act. pretiosum and Str. acromogenes.

Eukaryotic NAT is known to activate potential carcinogens and polymorphic NAT in eukaryotes often results in the generation of variants in which acetylation rates are altered (Blum et al., 1992). These NAT isoenzymes have been shown to metabolize xenobiotics at different rates and have been linked to disease susceptibility (e.g. Risch et al., 1995). Polymorphic NAT isoenzymes in the Enterobacteriaceae are of clinical interest as different strains within these genera may contribute to the activation of potential carcinogens in the human gut. Further kinetic studies will determine the effect of the mutations identified in this report on enzymatic activity.

The NAT-like proteins isolated from Amy. mediterranei (Fig. 7), Act. pretiosum and Str. acromogenes (Sohng et al., 1997) have been shown to be transcribed as part of an antibiotic synthesis gene cluster (Floss & Yu, 1999). It has been deduced that these particular NATs may be key to the cyclization of the ansamycin precursors to yield proansamycin X (Amy. mediterranei; Yu et al., 1999) and rubranasol (Str. acromogenes; Sohng et al., 1997) (Fig. 4b). This cyclization is via amide formation and the enzyme responsible for the production of proansamycin X has been accordingly termed rifamycin amide synthase (RAS; Yu et al., 1999). RAS forms part of an immense 90 kb cluster (Fig. 7) which encodes the proteins responsible for the complete biosynthesis of the antibiotic rifamycin B in Amy. mediterranei (Floss & Yu, 1999; Schupp et al., 1998; Tang et al., 1998).

Bacterial genomes lack much of the non-coding sequence associated with the eukaryotic genome. Often eubacterial transcripts consist of related genes in an operon, as found with the rifamycin cluster of Amy. mediterranei. Whole-genome analysis of the pathogenic mycobacteria within the M. tuberculosis and M. avium complexes are revealing unexpected cellular processes such as an ability to survive anaerobically (Cole et al., 1998). We have found transcription of nat in M. tuberculosis by RT-PCR (not shown). It is therefore likely, as with the other predicted gene clusters in this report, that the slow-growing mycobacteria also encode a nat operon (Fig. 7). The putative nat operon in pathogenic mycobacteria (Fig. 7) encodes enzymes capable of metabolizing aromatic and biphenyl-based compounds. As well as generating an additional source of carbon and nitrogen, this cluster contributes to the ability of the mycobacterium to survive in a highly toxic environment. Such a finding is supported by the genome of M. tuberculosis encoding over 20 separate cytochrome P450 enzymes (Cole et al., 1998), more than any other bacterium studied to date.

Unlike the bacterial and mycobacterial cytochrome P450 enzymes (Fulco, 1991; Poupin et al., 1999), NAT production in Sal. typhimurium would appear not to be regulated by the presence of substrate, although higher concentrations of substrate may be required (Fig. 5b, c). It has been reported that xenobiotic enzymes are often involved in co-metabolism and therefore produced in response to a mixture of substrates or a compound involved earlier in the degradation pathway (French et al., 1999). This may also be the case with the nat operon, implying regulation by aromatic and/or biphenyl-based compounds other than NAT ligands.

To date, approximately 33 complete prokaryotic genomes (including 5 archaeal sequences) have been determined (Fig. 2) and in excess of 90 are ongoing (http://www-fp.mcs.anl.gov/~gaasterland/genres.html; http://www.tigr.org/tdb/mdb/mdb.html). However, only a fraction of the genomes sequenced appear to possess nat. Alignment of the primary amino acid sequences of both NAT and RAS enzymes reveals a high homology, particularly in the N-terminal domains 1 and 2 (Table 2) and is indicative of divergent evolution from a common ancestor. The microbes shown in this report which possess nat inhabit environments such as soil and faecal material, in which there are high concentrations of aromatic compounds due to the decomposition of organic material. It is therefore possible that selection pressures on these microbes have resulted in the acquisition of genes encoding xenobiotic-metabolizing enzymes. This is supported by the gene content of the nat operons of the M. tuberculosis and M. avium complexes described in this report.

**ACKNOWLEDGEMENTS**

We are grateful to the Wellcome Trust for providing support and the Medical Research Council for a graduate studentship (A. Mushag). We thank Dr R. Minchin, Laboratory for Cancer Medicine, Royal Perth Hospital, Australia, Dr S. Clifton, Genome Sequencing Center, Washington University School of Medicine, USA, Dr T. Victor, Stellenbosch University, South Africa and Dr J. M. Duprat and colleagues, Hôpital Robert Debré, Paris, France, for their gifts of DNA and helpful discussions.
REFERENCES


Received 26 September 2000; revised 18 December 2000; accepted 24 January 2001.