Investigation of the catalytic triad of arylamine N-acetyltransferases: essential residues required for acetyl transfer to arylamines

James SANDY^{1,2}, Adeel MUSHTAQ^{1,3}, Simon J. HOLTON^{1,4}, Pamela SCHARTAU⁵, Martin E. M. NOBLE⁶ and Edith SIM

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

The NATs (arylamine N-acetyltransferases) are a well documented family of enzymes found in both prokaryotes and eukaryotes. NATs are responsible for the acetylation of a range of arylamine, arylhydrazine and hydrazine compounds. We present here an investigation into the catalytic triad of residues (Cys-His-Asp) and other structural features of NATs using a variety of methods, including site-directed mutagenesis, X-ray crystallography and bioinformatics analysis, in order to investigate whether each of the residues of the catalytic triad is essential for catalytic activity. The catalytic triad of residues, Cys-His-Asp, is a well defined motif present in several families of enzymes. We mutated each of the catalytic residues in turn to investigate the role they play in catalysis. We also mutated a key residue, Gly¹²⁶, implicated in acetyl-CoA binding, to examine the effects on acetylation activity. In addition, we have solved the structure of a C70Q

mutant of *Mycobacterium smegmatis* NAT to a resolution of 1.45 Å (where 1 Å = 0.1 nm). This structure confirms that the mutated protein is correctly folded, and provides a structural model for an acetylated NAT intermediate. Our bioinformatics investigation analysed the extent of sequence conservation between all eukaryotic and prokaryotic NAT enzymes for which sequence data are available. This revealed several new sequences, not yet reported, of NAT paralogues. Together, these studies have provided insight into the fundamental core of NAT enzymes, and the regions where sequence differences account for the functional diversity of this family. We have confirmed that each of the three residues of the triad is essential for acetylation activity.

Key words: arylamine N-acetyltransferase (NAT), bioinformatics, crystallography, site-directed mutagenesis.

INTRODUCTION

The NATs (arylamine N-acetyltransferases) are a family of enzymes initially identified as being responsible for the metabolism of arylamines and arylhydrazines by transfer of an acetyl group from acetyl-CoA to the terminal nitrogen group of the substrate [1]. As a result of genome sequencing, a range of NAT homologues have now been identified, and a pattern has emerged [2]. The proteins are found in a range of organisms, and were proposed historically through enzymic studies to have an active-site cysteine residue [1]. Kinetic analyses using tissue homogenates [3], and more recently pure enzyme [4,5], are compatible with a Ping Pong mechanism, in which the active-site cysteine is acetylated initially by the acetyl donor. For such reaction mechanisms, a non-productive interaction with the acetyl acceptor in advance of cysteine acetylation can occur, and this has been specifically observed for NAT family members [4,6,7]. Studies on the NAT enzyme from *Salmonella typhimurium*, which was initially identified as also carrying out the O-acetylation of hydroxylamines, were the first to indicate that the active site of NATs contains a cysteine residue (in *S. typhimurium* NAT, this is residue 69) [8]. The active-site cysteine was also found to be essential by site-directed mutagenesis of the human NAT2 isoform [9,10], and by chemical modification of the cysteine residue in NATs from rabbit [11], *S. typhimurium* [12] and hamster [4].

Subsequently, crystallographic analysis of the enzyme from *S. typhimurium* indicated that the cysteine residue in the active site is located adjacent to a histidine residue and an aspartate

residue, which together form a proposed catalytic triad [12]. These residues forming the catalytic triad are highly conserved in NAT homologues for which sequence data are available. In addition, crystal structures are now available for NATs from other prokaryotes, including *Mycobacterium smegmatis* [13], *Pseudomonas aeruginosa* [14] and *Mesorhizobium loti* NAT1 [15]. The NAT enzymes all share a common fold, composed of three domains of approximately the same length. The first two domains, an α -helical bundle and a β -barrel (amino acids 1–85 and 86–174 respectively), are joined by a linker helix (amino acids 175–200) to the third domain, an α/β lid (residues 201–280). In each case the tertiary structures of the native enzymes are very similar, and the residues of the proposed catalytic triad can be superimposed. While it has been demonstrated that these NAT enzymes each contain the residues forming the catalytic triad, it has not been demonstrated whether each of the three residues, apart from cysteine, is required for enzymic activity [8,9,16]. In order to answer this question, we have generated NATs from *M. smegmatis* and *S. typhimurium* in which each of the residues of the catalytic triad has been mutated. We have investigated the effects of these mutations on the catalytic activity of the enzyme, and relate this to *in silico* studies of the sequences of all NAT homologues for which sequence information is currently available. In addition, because a glutamine residue is an approximate isostere of an acetylated cysteine, we have mutated the active-site cysteine of *M. smegmatis* NAT (Cys⁷⁰; equivalent to Cys⁶⁹ of *S. typhimurium* NAT) to a glutamine (mutant C70Q). This mutant provides a model for the structure of the

Abbreviations used: NAT, arylamine N-acetyltransferase; pNPA, *p*-nitrophenyl acetate.

¹ These authors contributed equally to this work.

² To whom correspondence should be addressed (email james.sandy@pharm.ox.ac.uk).

³ Present address: University of Washington, Seattle, WA, U.S.A.

⁴ Present address: EMBL, Hamburg, Germany.

⁵ Present address: Department of Chemistry, University of Cambridge, Downing Street, Cambridge, U.K.

⁶ Present address: Laboratory of Molecular Biophysics, University of Oxford, Oxford OX1 3QU, U.K.

Table 1 Oligonucleotides used for generating mutants of NATs from *M. smegmatis* and *S. typhimurium*

For NAT from *M. smegmatis*, two PCR products were ligated to produce the mutated open reading frame of *msnat*, since the high GC content precluded the use of the system employed for *S. typhimurium* NAT. *M. smegmatis* NAT cloned into pGEMT-EZ was used as template DNA. Each sense primer (S) was used with the cloning vector SP6 and each antisense primer (AS) was used with the cloning vector T7 (<http://www.promega.com/vectors>). The two products were then blunt-end-ligated before amplification and subsequent cloning into either pET28b or pBAD/gIII. For NAT from *S. typhimurium*, mutations in the open reading frame of *stnat* cloned into pBADgIII [5] were created using the GeneEditor™ kit (Promega Inc.) using the mutagenic oligonucleotide sense primers shown. Numbers correspond to nucleotides in relation to the first nucleotide of the open reading frame of the corresponding gene. Nucleotide mismatches compared with the wild-type template are underlined. All mutated DNA fragments were sequenced. Activity assays were carried out as described in the Materials and methods section. '—' indicates that no activity was detected with 4-aminoveratrol, *p*-anisidine, isoniazid, 5-aminosalicylate or *p*-aminobenzoic acid as substrate with either acetyl-CoA or pNPA as acetyl donor after 1 h at 37 °C with up to 1 mg of enzyme and 2 mM substrate or acetyl donor; '+' indicates that activity was detected (see Table 3).

Mutation	Activity	Oligonucleotide (5' → 3')	Sense/antisense
<i>M. smegmatis</i> NAT			
C ⁷⁰ → X	—	C ²²⁶ TGCTCGGCTACGTACTG ²⁴³	S
H ¹¹⁰ → X	—	G ³⁴⁶ CGGTGCCGGGGCGCC ³⁶⁰	S
D ¹²⁷ → X	—	C ³⁹⁷ AGACGCTGACCTCGCC ⁴¹³	S
Both enzymes			
C ⁷⁰ → A	—	G ²²⁵ CCGTTGTGTTCTAGGCGTACCCGCCACGGCG ¹⁹³	AS
C ⁷⁰ → Q	—	G ²²⁵ CCGTTGTGTTCTAGTACCCGCCACGGCG ¹⁹³	AS
C ⁷⁰ → S	—	G ²²⁵ CCGTTGTGTTCTAGCTGACCCGCCACGGCG ¹⁹³	AS
H ¹¹⁰ → R	—	G ³⁴⁵ ACCGACAGCAGTTGCGGGCTGGGCGGGCAG ³¹³	AS
H ¹¹⁰ → W	—	G ³⁴⁵ ACCGACAGCAGTTCCAGGTCTGGGCGGGCAG ³¹³	AS
H ¹¹⁰ → A	—	G ³⁴⁵ ACCGACAGCAGTTGGCGGTCTGGGCGGGCAG ³¹³	AS
D ¹²⁷ → W	—	T ³⁹⁶ CCGCCGAAGCCACCTTCAGCAGGTACCGGCC ³⁶⁴	AS
D ¹²⁷ → N	—	T ³⁹⁶ CCGCCGAAGCCACGTTGACCAAGTACCGGCC ³⁶⁴	AS
D ¹²⁷ → A	—	T ³⁹⁶ CCGCCGAAGCCACGGCCAGCAGGTACCGGCC ³⁶⁴	AS
<i>S. typhimurium</i> NAT			
C ⁶⁹ → A	—	C ¹⁹⁰ GACGCGGTGGATACGCTTTTGAAGTGAATGGC ²²³	S
C ⁶⁹ → Q	—	C ¹⁹⁰ GACGCGGTGGATACCAATTGAAGTGAATGGC ²²³	S
C ⁶⁹ → S	—	C ¹⁹⁰ GACGCGGTGGATACAGTTTGAAGTGAATGGC ²²³	S
G ¹²⁶ → P	+	T ³⁹⁸ GGGGTTTCCCGGCCAAACGCTAACCG ⁴²⁴	S
G ¹²⁶ → A	+	T ³⁹⁸ GGGGTTTGCCGGCCAAACGCTAACCG ⁴²⁴	S

acetylated intermediate that offers some insight into the structural stabilization of this species, and we have determined its crystal structure to a resolution of 1.45 Å (where 1 Å = 0.1 nm).

MATERIALS AND METHODS

Site-directed mutagenesis

A series of mutants was created (Table 1) using the gene sequences of NATs from *M. smegmatis* and *S. typhimurium* as templates, and using the method described previously [17]. The final mutant products were cloned into pET28b and pBAD/gIII vectors as described previously [5], with the mutated residues being confirmed by sequencing. The pET28b vectors were propagated in *Escherichia coli* strain BL21(DE3)pLysS, as described previously [13], in LB (Luria–Bertani) medium supplemented with final concentrations of 1 M sorbitol and 0.25 mM betaine, with induction of recombinant protein being carried out at 27 °C in the presence of 0.1 mM isopropyl β-D-thiogalactoside. The pBAD/gIII vector was propagated in *E. coli* TOP10 cells, as described previously [5], except that induction was carried out at 27 °C in the presence of 0.02 % (v/v) arabinose for *M. smegmatis* NAT and the derived mutants.

The recombinant NAT enzymes, all of which had an N-terminal hexahistidine affinity tag, were prepared from the soluble fraction of whole-cell lysates, which includes the periplasmic space. The NAT enzymes were purified using a Ni²⁺-nitrilotriacetate affinity resin, with the *M. smegmatis* NAT variants eluting in 50 mM imidazole [13] and the *S. typhimurium* NAT variants eluting in 250 mM imidazole [5]. The hexahistidine affinity tag was removed by digestion with thrombin (5 units/mg of protein); the protein was then dialysed into 20 mM Tris/HCl, pH 8.0, 1 mM EDTA and 1 mM dithiothreitol, filtered through a 0.22 µm filter, stored at 4 °C, and either used for assay of activity or concentrated

to 15 mg/ml for crystallization trials, as described previously [12,13].

Crystallization of C70Q mutant *M. smegmatis* NAT

The NAT protein was incubated with Molecular Dimensions Structure Screen I and II, and crystallization trials were set up using a Tecan Genesis ProTeam 150 liquid handling robot into 96-well plates. A drop size of 1 µl (protein/mother liquor 1:1, v/v) was used in all cases. Plates were sealed and incubated at 20 °C. Crystals generally appeared between 1 and 3 days.

Data collection and structure determination

X-ray diffraction data were collected at the ESRF (European Synchrotron Radiation Facility, Grenoble, France) on beamline 14.1 under cryogenic conditions. Diffraction images were analysed and integrated using MOSFLM [17]. Molecular replacement was carried out using Amore [17], with a modified 'A' chain from the native *M. smegmatis* NAT (PDB ID 1gx3) as a search model to give an initial *R*-factor after the 'FITING' stage of 0.42. Rigid body refinement was applied to the molecular replacement solution using REFMAC5 [17a], followed by several rounds of restrained refinement (also REFMAC5). Model rebuilding was carried out using O [18]. Solvent molecules were located using ARP/wARP [19]. Co-ordinates and structure factors have been deposited in the Protein Data Bank (PDB ID 1W5R).

Determination of enzymic activity

Protein samples were analysed by SDS/PAGE for purity, and the concentration of protein was determined either by measuring the absorbance at 280 nm using molar absorption coefficients of 1.15 litre · mol⁻¹ · cm⁻¹ for *S. typhimurium* NAT and 0.95 litre · mol⁻¹ · cm⁻¹ for *M. smegmatis* NAT (<http://www.basic.nwu.edu/>)

biotoools/proteincalc.html), or by determining the concentration of protein using the Bradford colorimetric method [20] with BSA as a standard. Determination of protein precipitation was by measurement of protein concentration during storage at 4°C in the supernatant following centrifugation at 14000 g for 10 min at 4°C.

Enzymic activity to quantify arylamine substrate remaining was determined spectrophotometrically (on a Cecil 5500 spectrophotometer) by analysis of acetylation of the arylamines *p*-aminobenzoic acid, *p*-anisidine, 4-aminoveratrole and 5-amino-salicylate as substrates at 37°C, and using acetyl-CoA as an acetyl donor. In experiments where pNPA (*p*-nitrophenyl acetate) was used as an acetyl donor, quantification was carried out by measurement of the hydrolysis of pNPA in the presence of arylamine substrate, as described previously [5].

Sequence analysis

Similar NAT sequences were obtained by a BLAST search [21] using the amino acid sequences from *M. smegmatis* and *S. typhimurium* NATs as search models. Sequences were then aligned using the CLUSTALW (<http://www.ebi.ac.uk/clustalw/>) [22] program (version 1.82) using the BLOSUM matrix, and redundant sequences were removed using the EXPASY reduce redundancy program (<http://us.expasy.org/tools/redundancy/>). To find consensus sequences, ESPRIPT [23] was used to identify different levels of sequence conservation (100%, 90%, 80% and 70%). Prokaryotic sequences were analysed separately from eukaryotic sequences in order to determine conservation among each group, and then all sequences were analysed together to find a global consensus sequence. Sequence conservation was also analysed using the AMAS server (http://barton.ebi.ac.uk/servers/amas_server.html). The alignment produced from CLUSTALW [22] was used with NAT sequences grouped as suggested in the AMAS manual. All numbering of residues in the text and Figures is based on the NAT sequence from *S. typhimurium* for ease of comparison. Although *S. typhimurium* NAT has only 280 amino acid residues, sequence numbering is from 1 to 310 because of the amino acid sequence of *Neurospora crassa* NAT, which contains an insertion in the sequence not found in any other NAT homologue.

RESULTS AND DISCUSSION

Expression of active-site cysteine mutants

Recombinant mutant *M. smegmatis* NAT enzymes in which the active-site cysteine was mutated to alanine, serine or glutamine were expressed as recombinant proteins using both the pBADgIII expression system and the pET28b expression vector (Table 1). Using the pBADgIII system, each of the NAT mutants was expressed as a soluble protein, such that the yield of purified protein was approx. 3 mg/litre of culture for each mutant. This is comparable with the yield of wild-type *M. smegmatis* NAT protein produced with the same expression vector under similar induction conditions [0.02% (v/v) arabinose for 4 h at 27°C]. The recombinant proteins were equally as soluble as the native protein, as observed using SDS/PAGE (Figure 1). Although other bands were present on the gel, there were no differences in these minor contaminants or degradation products between the wild-type enzyme and the site-directed mutants. This suggests that the mutant recombinant proteins were folded in solution, such that they were not susceptible to proteolysis, as has been observed with other, unfolded mutant NAT enzymes [10,24–26].

Figure 1 Site-directed mutants of *M. smegmatis* NAT are expressed as soluble proteins

An SDS/12%–PAGE gel stained with Coomassie Blue is shown. Each mutant was purified as described in the Materials and methods section, and then the hexahistidine tag was cleaved with thrombin. A sample of 10 µg of purified NAT enzyme was loaded in each lane.

Table 2 Crystallographic data for C70Q *M. smegmatis* NAT

Values in parentheses indicate the specific values in the highest-resolution shell. R_{merge} is defined as:

$$R_{\text{merge}} = \sum_h \sum_j |I_{h,j} - \langle I_h \rangle| / \sum_h \sum_j I_{h,j}$$

where $I_{h,j}$ is the intensity of the j th observation of unique reflection h . R_{conv} is defined as:

$$R_{\text{conv}} = \sum_h ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_h |F_{\text{obs}}|$$

where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes for reflection h . R_{free} is equivalent to R_{conv} , but is calculated using a 5% disjoint set of reflections excluded from the maximum likelihood refinement stages. RMSD, root mean square deviation. 1 Å = 0.1 nm.

Parameter	Value
Space group	P4 ₁ 2 ₁ 2
Cell dimensions (Å)	98.9, 98.9, 130.9
Maximum resolution (Å)	1.45
Highest-resolution shell (Å)	1.53–1.45
Observed reflections	330 604
Unique reflections	103 538
Completeness (%)	90.5 (84.7)
$I/\sigma I$	6.8 (1.8)
Mosaicity (°)	0.32
Redundancy	3.2 (2.7)
R_{merge} (%)	6.0 (41.4)
V_m (Å ³ /Da)	2.5
Refinement	
Protein atoms	1 molecule of C70Q NAT chain A residue 0–276*
Number of atoms	9326
Other atoms	300 water molecules
Resolution range (Å)	31.00–1.45
R_{conv} (%)	18.8
R_{free} (%)	21.5
RMSD bond length (Å)	0.012
RMSD bond angle (°)	1.482

* Residue 0 is the last residue of the N-terminal cloning artefact.

The C70Q *M. smegmatis* NAT mutant recombinant protein was expressed using the pET28b expression system in order to generate sufficient protein for crystallization. The protein was found to crystallize under a range of conditions, and the best crystals were obtained with 0.2 M MgCl₂, 0.1 M Tris/HCl, pH 8.5, and 30% poly(ethylene glycol) 4000. Upon X-ray analysis, the mutant enzyme was found to have crystallized in the space group P4₁2₁2 and diffracted to a resolution of 1.45 Å (Table 2). The X-ray data were processed as described in the Materials and methods section, and molecular replacement was carried out using an 'A' chain from wild-type *M. smegmatis* NAT as a template. Once the molecular replacement solution was found, rigid-body refinement was carried out, followed by model building using O, and further rounds of restrained refinement.

Figure 2 Structure of C70Q *M. smegmatis* NAT at 1.45 Å resolution

(A) Comparison of the peptide backbone of wild-type *M. smegmatis* NAT (blue) and the C70Q mutant (green), showing the catalytic triad as a ball and stick representation. (B) Superposition in the model of acetylcysteine and glutamine. (C) Close-up view of the modified catalytic triad in the C70Q *M. smegmatis* NAT. Maps are contoured at 1.5 σ . Panels (A) and (C) were produced using AESOP (M. E. M. Noble, unpublished work; details available from M. E. M. N. on application; email martin.noble@biop.ox.ac.uk); (B) was produced using the Deep-View Swiss pdb viewer [51] (<http://www.expasy.org/spdbv>).

The C α structure of the C70Q NAT recombinant protein is illustrated in Figure 2(A). The structure of the C70Q mutant is isomorphous with that of the wild-type enzyme from *M. smegmatis*, and superimposition of the C70Q mutant with the wild-type enzyme using Deep View (SwissPDB Viewer) gives a root mean square difference of 0.35 Å over 1080 backbone atoms. The regions that are different between the two structures are in solvent-exposed loops that are remote from sites of crystal contacts. Since glutamine represents a structural isostere to an acetylated cysteine residue (Figure 2B), the C70Q NAT mutant therefore represents a model of the acetylated intermediate state of the NAT enzyme. The fact that the cysteine-to-glutamine mutation is accommodated within the active site without any structural perturbation of the wild-type fold suggests that an acetylated cysteine group could also be accommodated (Figure 2C). These data do not preclude a conformational change of the protein before and during the acetylation process, as has been proposed for *S. typhimurium* NAT [5] and has been inferred for the human NAT1 enzyme from the effects of acetylation on protein stability [10]. The C70Q mutant adopts a conformation whereby the amide oxygen of the glutamine forms a 2.85 Å hydrogen bond with its own peptide nitrogen. In doing this, it displaces a tightly bound water molecule observed in the wild-type structure that has been suggested previously to define the centre of the oxyanion hole that facilitates formation of a tetrahedral intermediate during nucleophilic attack [12]. Observation of this interaction in this structure indicates that the oxyanion hole may play an additional role in stabilizing and orienting the acetyl moiety for subsequent

attack by the acetyl group acceptor of the NAT reaction. Other minor rearrangements are seen in the active site, notably in the orientation of the imidazole ring of the catalytic histidine, although we cannot exclude the possibility that this results from the different chemical nature of a glutamine compared with the true acetylated intermediate.

The acetylated intermediate of hamster NAT2 has been estimated to have a half-life of 88 s [27] using pure enzyme, which is longer than a previous estimate of less than 1 min for pigeon NAT in a liver homogenate [3]. Although the half-life of the acetylated intermediate is likely to be dependent on the specific NAT isoenzyme, these values, suggesting a transient acetylated intermediate, are compatible with the calculation that approx. 10% of active human NAT1 becomes labelled when reacted with radioactive acetyl-CoA [11,28]. It has been found, using an assay to detect the hydrolysis of acetyl-CoA directly, that the presence of arylamine substrate greatly increases the rate of hydrolysis of acetyl-CoA by NAT from *M. smegmatis* [29].

In view of the difficulty in producing sufficient quantities of the acetylated intermediate of a NAT enzyme for crystallization, the glutamine mutant therefore represents the best model currently available for the active-site-acetylated intermediate.

Activities of site-directed mutants

Open reading frames corresponding to the mutants of both *M. smegmatis* NAT and *S. typhimurium* NAT, in which Cys⁶⁹, His¹⁰⁷ and Asp¹²² (*S. typhimurium* NAT numbering) were mutated,

were created (Table 1). The mutated NAT enzymes were expressed as soluble recombinant proteins. The rate of precipitation of each of these proteins was found, on storage in the presence of 1 mM dithiothreitol, to be comparable with that of the wild-type enzyme. In each case the mutants were expressed at similar levels of soluble protein as the recombinant wild-type enzyme [30]. Recombinant NAT enzymes were each used (for enzyme assays) within the first 20 days of storage (at 4 °C) for comparison with the wild-type enzymes. None of the mutant proteins exhibited any acetylation activity against *p*-aminobenzoic acid, 4-aminoveratrole, 5-aminosalicylate or *p*-anisidine (Table 1). The latter three substrates are effectively acetylated by the wild-type enzyme at pH 7.5. A range of conditions, including modification of pH across the range 5.0–8.5, concentrations of substrates or acetyl donors of up to 2 mM, and the presence of up to 1 mg of protein per assay, equally did not result in any acetylation activity by the mutant enzymes after 1 h at 37 °C (Table 1). These data suggest that the residues Cys⁶⁹, His¹⁰⁷ and Asp¹²², which form the catalytic triad, are each individually essential for acetylation activity with acetyl-CoA as the acetyl donor.

Each of the catalytic triad residues was mutated to an alanine residue. None of these mutant enzymes exhibited acetylation activity. Enzymes in which the cysteine residue was mutated to either an alanine or a glutamine would not be expected to catalyse acetyl transfer. However, substitution of a cysteine with a serine residue, such as is involved in acetyl transfer in acetylcholinesterase [31], was expected to produce an enzyme with residual activity. However, no activity could be detected, suggesting that the serine residue is insufficiently nucleophilic to support NAT activity (Table 1). In an artificial cysteine protease, thiosubtilisin, in which serine has been replaced by cysteine, it has been demonstrated that the enzyme, as well as having protease activity, can catalyse an acetylation reaction when ethyl acetate is used as the acetyl donor, albeit at a very low rate [32].

The His¹⁰⁷ residue was mutated to an arginine residue to determine the role of the potential proton relay created by the histidine within the catalytic triad. Substitution of histidine with a tryptophan residue was also carried out to mimic the aromatic properties of histidine. Neither the His¹⁰⁷ → Arg nor the His¹⁰⁷ → Trp mutant supported catalytic activity in *M. smegmatis* NAT (Table 1).

In the acetylcholinesterases, the third residue of the catalytic triad is glutamate [31]. This is also the case in several species of *Bacillus* NATs (see Supplementary Figures 1 and 2 at <http://www.BiochemJ.org/bj/390/bj3900115add.htm>); however, it is not known whether these NAT homologues are active in catalysis. The replacement of Asp¹²² with Asn in *M. smegmatis* NAT resulted in an inactive protein (Table 1). This substitution is found in one isoform of human NAT2 (NAT 2*12D) which produces an inactive enzyme [33]. The studies presented here complement refolding studies with the insoluble recombinant hamster NAT2 enzyme, in which it was reported that enzyme in which the active-site aspartate was replaced by either alanine or asparagine could not be refolded in an active form, unlike the refolded native hamster NAT2 [4]. A catalytic aspartate has been demonstrated to be important in prokaryotic and eukaryotic members of the transglutaminase family [34,35] and in the serine proteases, such as subtilisin, whereas in the cysteine proteases asparagine is often found as the third residue of the triad [36]. Whether the NAT enzyme can act as a protease has not been reported.

Modification of the putative acetyl-CoA binding site

Unlike acetyltransferases of the unrelated GNAT (GCN5-related NAT) family, for which structures of the enzyme complexed

Table 3 Effect of substitution of Gly¹²⁶ of *S. typhimurium* NAT on acetyl transfer activity

Activity assays were carried out as described in the Materials and methods section using wild-type enzyme and the G126A and G126P mutants. nd, no activity detected.

Enzyme	Acetyl-CoA		pNPA	
	K_m (μ M)	k_{cat} (s^{-1})	K_m (μ M)	k_{cat} (s^{-1})
Wild-type	< 20	1.27	159	353
G126A	80	8.73×10^{-3}	1100	0.16
G126P	nd	nd	2350	0.89

with acetyl-CoA have been obtained for several family members [37,38], a structure of NAT with acetyl-CoA bound has not been forthcoming. However, there are regions of NAT that have been suggested, on the basis of sequence conservation [2,12], to be involved in acetyl-CoA binding, particularly in comparison with the RifF enzyme, a NAT homologue involved in amide bond formation in the synthesis of rifamycin which does not use acetyl-CoA as a cofactor [39]. It has been proposed that Gly¹²⁶ in NAT from *S. typhimurium* forms part of a structural P-loop involved in acetyl-CoA binding [12]. The Gly¹²⁶ residue of NAT from *S. typhimurium* was mutated to either an alanine or a proline. Proline was chosen for substitution because glycine is replaced in RifF by a proline [39]. When Gly¹²⁶ of *S. typhimurium* NAT was converted into a proline residue, the enzyme lost the ability to catalyse the transfer of an acetyl group from acetyl-CoA; however, the enzyme could still catalyse the acetylation of an arylamine using the acetyl donor pNPA (Table 3). The RifF enzyme, in which Gly¹²⁶ is a proline residue, is unable to catalyse acetylation of substrate using either acetyl-CoA or pNPA as an acetyl donor [39]. When Gly¹²⁶ in *S. typhimurium* NAT was replaced by alanine, the protein was less active (Table 3), but still supported the transfer of an acetyl group from acetyl-CoA with either isoniazid or *p*-anisidine as an acetyl acceptor. These recombinant mutant enzymes were found to be as soluble as the wild-type *S. typhimurium* NAT enzyme.

Gly¹²⁶ therefore appears to be involved in acetyl-CoA binding. It is likely that there are residues other than Gly¹²⁶ contributing to acetyl-CoA binding; however, these have not yet been identified.

It has also been demonstrated using random [40] and site-directed [41] mutagenesis that a range of amino acid substitutions can result in a decrease in activity despite a lack of effect on protein stability. These point mutations, resulting in substitution of a particular amino acid, are likely to be specific effects, rather than due to non-specific structural modifications resulting in poorly folded, and hence unstable, proteins.

G126P/A substitutions in naturally occurring NATs

A Gly¹²⁶ → Ala substitution is found in NAT enzymes from two *Bacillus* species, namely *B. cereus* and *B. anthracis* [42] (see Supplementary Figure 2 at <http://www.BiochemJ.org/bj/390/bj3900115add.htm>). Interestingly, a Gly¹²⁶ → Pro substitution is found in the NAT3 enzymes from mice and rats [43–45] as well as in the rifamycin amide synthase enzyme RifF [39]. This substitution in mouse and rat NAT3 may explain why these isoenzymes exhibit very low rates of acetylation activity.

In silico analysis of NAT sequences

The analysis of 29 prokaryotic NAT sequences provides a clearer view of which residues are conserved (Figure 3). There are only ten fully conserved residues among the prokaryotic NATs, with 17 highly conserved residues (90% level). This is in

Figure 3 Sequence conservation among prokaryotic NAT enzymes

A total of 29 prokaryotic NAT sequences were found using a BLAST search and aligned using the CLUSTALW facility (www.ebi.ac.uk/clustalw/). Once aligned, the sequences were subjected to analysis using the ESPRIPT [23] server. AESOP (M. E. M. Noble, unpublished work; details available from M. E. M. N. on application) was used to highlight the α -backbone by indicating 100% (red), 90% (orange), 80% (yellow) and 70% (green) conservation across all prokaryotic sequences aligned (**A** and **B**). The corresponding ribbon diagrams (**C** and **D**) are shaded ramped from N-terminus (blue) to C-terminus (red). The catalytic triad residues are shown in stick format. (**A**) and (**C**) are in the same orientation as each other, and are orthologous to (**B**) and (**D**).

contrast with the eukaryotic NATs, where 32 residues are completely conserved among 30 sequences, with 56 residues being highly conserved (90% level). There are two main clusters of conservation within the prokaryotic sequences, P*EN (residues 37–40) and RGG*CYE (residues 65–71) (where * refers to non-conserved residues); however, the eukaryotic sequences show a much higher degree of identity throughout the entire sequence, as might be expected from relative evolutionary timescales (Figure 4).

Among all prokaryotic and eukaryotic NAT sequences analysed here, only two do not have an arginine residue at position 65: *Bacillus subtilis* NAT [2] and the *Amycolatopsis mediterranei* NAT-like protein Riff [39] have a histidine and glutamine residue respectively. There is an allele of the human NAT (human

NAT2*5) where a glutamine occupies this position [46]. Unlike the Riff enzyme, the human NAT isoform still exhibits acetylation activity, albeit at a low rate [46].

Multiple NAT sequences have been found in several prokaryotic species, including *Mesorhizobium loti* [47], for which structural information is available [15]. As a result of our bioinformatic study (Figures 4 and 5, and Supplementary Figures 2 and 3 at <http://www.BiochemJ.org/bj/390/bj3900115add.htm>), we have identified several other prokaryotic organisms that have more than one NAT homologue, i.e. *B. cereus* and *B. anthracis*, each of which has three NAT homologues (Supplementary Figures 2 and 3).

Only seven invariant residues throughout the entire sequence have been identified; these are Leu²⁵, Gly⁶⁶, Gly⁶⁷, Cys⁶⁹, Asn⁷³,

Figure 4 Sequence conservation among NAT enzymes

All NAT sequences found from BLAST searches were aligned using the CLUSTALW facility (www.ebi.ac.uk/clustalw), and the alignment files were used for analysis using the AMAS server [52]. Prokaryote (A) and eukaryote (B) NATs were analysed separately, with the inclusion of the putative NAT sequence from *Neurospora crassa* in both analyses. The *N. crassa* NAT enzyme has 310 amino acids, and appears to have an extended loop region not present in any other NAT homologues, creating an insertion in the sequence alignment corresponding to residues 254–268 (A) and 240–249 (B). Mid-grey shading indicates the level of conservation between all species, and black shading indicates similarity between species. The light-grey shading indicates differences in amino acids between the sequences at each point along the alignment. The top bar represents the boundaries between domains based upon the *S. typhimurium* NAT enzyme domains: residues 1–85, α -helical bundle; 86–174, β -barrel domain; 175–200, linker helix; 201–280, α/β lid domain.

Figure 5 Fully conserved residues among NAT sequences

A total of 59 NAT sequences were aligned using the CLUSTALW server to analyse sequence conservation. Amino acids which are conserved across all NAT sequences are shaded in black. Amino acids conserved among eukaryotes only are shaded in light grey; those conserved in prokaryotes only are shaded in mid-grey. The sequence is shown in three sections approximating to the three domains, as indicated in the legend to Figure 4. Numbering is relative to that of the NAT sequence from *S. typhimurium*. Although some NAT enzymes are longer than 280 amino acids (e.g. *Neurospora crassa* with 310 amino acids), no sequence conservation is seen in the final C-terminal region. Amino acids are labelled in single-letter format. The amino acid sequence of *N. crassa* NAT was omitted from this analysis, as it appears to differ from both the prokaryotic-only and the eukaryotic-only alignments.

Gly⁸⁴ and His¹⁰⁷ (Figure 5). The third member of the catalytic triad (Asp¹²²) is not completely conserved, and is found as an asparagine in one human NAT2 isoform [33] which is inactive in acetyl-transfer activity. Glutamate has been reported at position 122 in four *Bacillus* species [*B. thuringiensis* (subspecies konkukian), *B. cereus* (subspecies A.T.C.C. 10987), *B. cereus* (subspecies A.T.C.C. 14579/DSM31) and *B. anthracis*] replacing the catalytically important Asp¹²², but it is not known whether these enzymes exhibit N-acetylation activity. These enzymes may play another, as yet unknown, role in their host organism.

Of the seven invariant residues observed in all NAT enzymes (Figure 5), Leu²⁵ is likely to be involved in structural stability, as it forms van der Waals forces with Leu⁷⁹ and a highly conserved Val¹¹². Leu⁷⁹ is completely conserved in all NAT enzymes analysed here, with the exception of NAT from *Streptomyces murayamaensis*, where an equally hydrophobic valine occupies the equivalent position (see Supplementary Figure 2 at <http://www.BiochemJ.org/bj/390/bj3900115add.htm>). Val¹¹² is also highly

conserved (> 70 % conservation among all NATs), being replaced by cysteine (4/59), alanine (9/59), threonine (4/59) or isoleucine (9/59) (Supplementary Figure 4). These interactions are likely to contribute towards maintaining the β -barrel domain position, hence contributing towards the structural integrity of the NAT enzymes. The Gly⁶⁶ and Gly⁶⁷ residues also appear to play a role in stabilizing the structure of the NAT enzymes, as they are involved in forming hydrogen bonds with the well conserved Pro³⁷ and Glu³⁹ (part of the well conserved PFENL region) [2] (Figure 4 and Supplementary Figures 2 and 4). The amide backbone nitrogen of Gly⁶⁶ forms hydrogen bonds with the carboxy side chain of Glu³⁹, while the amide backbone nitrogen of Gly⁶⁷ forms hydrogen bonds with the backbone carbonyl oxygen of Pro³⁷.

Cys⁶⁹ is the essential active-site cysteine [8,9,11,12,16]. Previous mutagenesis experiments with human NAT1 suggested that this cysteine residue may exist in an acetylated form, protecting the enzyme from proteolysis [10], although no evidence has been found for this as yet in the prokaryotic NAT enzymes.

Asn⁷³ is likely to play a role in stabilizing the catalytic cysteine, keeping it in the correct orientation, as it forms hydrogen bonds from the amide nitrogen of the backbone with the backbone carbonyl oxygen of Cys⁶⁹. Gly⁸⁴ is situated at the end of helix 4 prior to the β -barrel domain (Figure 4), and is likely to be important in allowing the adoption of the appropriate relative disposition of the α -helical and β -barrel domains of NAT. Finally, His¹⁰⁷ is the last completely conserved residue in the consensus sequence, and is important in forming the catalytic triad of residues. There are no fully conserved residues C-terminal of His¹⁰⁷, and this may relate to a divergent role for NAT enzymes. The C-terminus has been demonstrated to have a controlling effect on substrate specificity and enzymic activity [5]. It is possible that the wide range of NAT homologues may have widely divergent endogenous functions controlled by the C-terminal sequence.

In prokaryotes, only one additional completely conserved residue exists, Gly¹²⁴ (Figure 5). This residue both stabilizes the catalytically important Asp¹²² by forming a hydrogen bond and forms part of the loop containing Phe¹²⁵, implicated in many substrate/inhibitor studies of NAT enzymes [5,6,29,48]. However, a consensus sequence for eukaryotic enzymes only extends further into the C-terminal domain and may reflect a greater similarity of function in this subgroup of NAT enzymes. Discussion of the implications of the conservation of these residues with respect to structure is difficult due to the lack of a eukaryotic NAT structure. Previous reports of modelling of eukaryotic NAT enzymes have

only managed to model the first 131 amino acids due to the very low identity between the eukaryotic and prokaryotic NAT enzymes at the C-terminus [46,49].

Two arginine residues have been implicated as being essential for NAT activity (Arg⁹ and Arg⁶⁵) [41]; however, they are not completely conserved throughout the sequences highlighted in the present study. At position 9 there are examples of glutamine (*Staphylococcus aureus*), lysine (*Bacillus subtilis*, *Vibrio parahaemolyticus*, *Staphylococcus epidermidis*) and histidine (*Pseudomonas aeruginosa*, *Neurospora crassa*) residues (Supplementary Figure 2 at <http://www.BiochemJ.org/bj/390/bj3900115add.htm>). The NAT enzyme from *Pseudomonas aeruginosa* is extremely active with 5-aminosalicylate [50] and hydralazine substrates [14]. It is one of the most active prokaryotic NAT enzymes, suggesting that this arginine residue is not essential. The arginine residue at position 65 is conserved in all NAT enzymes apart from that found in *Bacillus subtilis*, where a histidine is present; however, there is evidence suggesting that this enzyme is active [2].

Conclusions

The results reported here demonstrate that each of the residues of the catalytic triad are essential for the activity of NATs from *S. typhimurium* and *M. smegmatis*. A C70Q mutant of the NAT enzyme from *M. smegmatis* is a good model for an acetylated enzyme intermediate. Site-directed mutagenesis and activity studies are compatible with Gly¹²⁶ of *S. typhimurium* NAT being involved in acetyl-CoA binding. *In silico* analysis of 59 NAT-like sequences has revealed that only seven amino acids are conserved among all species, and that some residues which were thought previously to be invariant amongst NATs may be replaced by similar residues without loss of activity, e.g. Arg⁹ to His in *P. aeruginosa* NAT. The active-site cysteine and histidine residues are, however, fully conserved among all NAT sequences analysed here. The lack of sequence identity in the C-terminal domain may reflect a difference in the endogenous roles of the different NAT homologues.

We are grateful to the Wellcome Trust for financial support and to the Medical Research Council (UK) for a studentship (A.M.). We thank the staff of the ESRF for their kind assistance and time. We are also grateful to Mark Payton, Annette Dreisbach and Katalin Pinter for their help in the early stages of the work presented here.

REFERENCES

- Weber, W. W. and Hein, D. W. (1985) *N*-Acetylation pharmacogenetics. *Pharmacol. Rev.* **37**, 25–79.
- Payton, M., Mushtaq, A., Yu, T.-W., Wu, L.-J., Sinclair, J. and Sim, E. (2001) Eubacterial arylamine *N*-acetyltransferases – identification and comparison of 18 members of the protein family with conserved active site cysteine, histidine and aspartate residues. *Microbiology* **147**, 1137–1147.
- Riddle, B. and Jencks, W. P. (1971) Acetyl-coenzyme A:arylamine *N*-acetyltransferase. Role of the acetyl-enzyme intermediate and the effects of substituents on the rate. *J. Biol. Chem.* **246**, 3250–3258.
- Wang, H., Guo, Z., Vath, G. M., Wagner, C. R. and Hanna, P. E. (2004) Chemical modification of hamster arylamine *N*-acetyltransferase 2 with isozyme-selective and nonselective *N*-aryl bromoacetamide reagents. *Protein J.* **23**, 153–166.
- Mushtaq, A., Payton, M. and Sim, E. (2002) The COOH terminus of arylamine *N*-acetyltransferase from *Salmonella typhimurium* controls enzymic activity. *J. Biol. Chem.* **277**, 12175–12181.
- Sandy, J., Fullam, E., Holton, S. J., Sim, E. and Noble, M. E. M. (2005) Binding of the anti-tubercular drug isoniazid to the arylamine *N*-acetyltransferase protein from *Mycobacterium smegmatis*. *Protein Sci.* **14**, 775–782.
- Delgoda, R., Lian, L. Y., Sandy, J. and Sim, E. (2003) NMR investigation of the catalytic mechanism of arylamine *N*-acetyltransferase from *Salmonella typhimurium*. *Biochim. Biophys. Acta* **1620**, 8–14.
- Watanabe, M., Sofuni, T. and Nohmi, T. (1992) Involvement of Cys69 residue in the catalytic mechanism of *N*-hydroxyarylamines *O*-acetyltransferase of *Salmonella typhimurium*. Sequence similarity at the amino acid level suggests a common catalytic mechanism of acetyltransferase for *S. typhimurium* and higher organisms. *J. Biol. Chem.* **267**, 8429–8436.
- Dupret, J. M. and Grant, D. M. (1992) Site-directed mutagenesis of recombinant human arylamine *N*-acetyltransferase expressed in *Escherichia coli*. Evidence for direct involvement of Cys68 in the catalytic mechanism of polymorphic human NAT2. *J. Biol. Chem.* **267**, 7381–7385.
- Butcher, N. J., Arulpragasam, A. and Minchin, R. F. (2004) Proteasomal degradation of *N*-acetyltransferase 1 is prevented by acetylation of the active site cysteine: a mechanism for the slow acetylator phenotype and substrate-dependent down-regulation. *J. Biol. Chem.* **279**, 22131–22137.
- Andres, H. H., Klem, A. J., Schopfer, L. M., Harrison, J. K. and Weber, W. W. (1988) On the active site of liver acetyl-CoA:arylamine *N*-acetyltransferase from rapid acetylators rabbits (III/IV). *J. Biol. Chem.* **263**, 7521–7527.
- Sinclair, J. C., Sandy, J., Delgoda, R., Sim, E. and Noble, M. E. (2000) Structure of arylamine *N*-acetyltransferase reveals a catalytic triad. *Nat. Struct. Biol.* **7**, 560–564.
- Sandy, J., Mushtaq, A., Kawamura, A., Sinclair, J., Sim, E. and Noble, M. (2002) The structure of arylamine *N*-acetyltransferase from *Mycobacterium smegmatis* – an enzyme which inactivates the anti-tubercular drug, Isoniazid. *J. Mol. Biol.* **318**, 1071–1083.
- Westwood, I. M., Holton, S. J., Rodrigues-Lima, F., Dupret, J. M., Bhakta, S., Noble, M. E. and Sim, E. (2005) Expression, purification, characterization and structure of *Pseudomonas aeruginosa* arylamine *N*-acetyltransferase. *Biochem. J.* **385**, 605–612.
- Holton, S. J., Dairou, J., Sandy, J., Rodrigues-Lima, F., Dupret, J.-M., Noble, M. E. M. and Sim, E. (2005) Structure of *Mesorhizobium loti* arylamine *N*-acetyltransferase 1. *Acta Crystallogr. F Struct. Biol. Crystallogr. Commun. Online* **61**, 14–16.
- Cheon, H. G., Boteju, L. W. and Hanna, P. E. (1992) Affinity alkylation of hamster hepatic arylamine *N*-acetyltransferases: isolation of a modified cysteine residue. *Mol. Pharmacol.* **42**, 82–93.
- Kawamura, A., Sandy, J., Upton, A., Noble, M. and Sim, E. (2003) Structural investigation of mutant *Mycobacterium smegmatis* arylamine *N*-acetyltransferase: a model for a naturally occurring functional polymorphism in *Mycobacterium tuberculosis* arylamine *N*-acetyltransferase. *Protein Expression Purif.* **27**, 75–84.
- Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*
- Jones, T. A., Cowan, S., Zou, J.-Y. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
- Perrakis, A. (1997) wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 448–55.
- Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.
- Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
- Gouet, P., Courcelle, E., Stuart, D. I. and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308.
- Martell, K., Levy, G. and Weber, W. (1992) Cloned mouse *N*-acetyltransferases: enzymatic properties of expressed *Nat-1* and *Nat-2* gene products. *Mol. Pharmacol.* **42**, 265–272.
- Fretland, A. J., Leff, M. A., Doll, M. A. and Hein, D. W. (2001) Functional characterization of human *N*-acetyltransferase 2 (NAT2) single nucleotide polymorphisms. *Pharmacogenetics* **11**, 207–215.
- Leff, M. A., Fretland, A. J., Doll, M. A. and Hein, D. W. (1999) Novel human *N*-acetyltransferase 2 alleles that differ in mechanism for slow acetylator phenotype. *J. Biol. Chem.* **274**, 34519–34522.
- Wang, H., Vath, G. M., Gleason, K. J., Hanna, P. E. and Wagner, C. R. (2004) Probing the mechanism of hamster arylamine *N*-acetyltransferase 2 acetylation by active site modification, site-directed mutagenesis, and pre-steady state and steady state kinetic studies. *Biochemistry* **43**, 8234–8246.
- Jencks, W. P., Gresser, M., Valenzuela, M. S. and Huneeus, F. C. (1972) Acetyl coenzyme A: arylamine acetyltransferase. Measurement of the steady state concentration of the acetyl-enzyme intermediate. *J. Biol. Chem.* **247**, 3756–3760.

- 29 Brooke, E., Davies, S. G., Mulvaney, A. W., Pompeo, F., Sim, E. and Vickers, R. J. (2003) An approach to identifying novel substrates of bacterial arylamine N-acetyltransferases. *Bioorg. Med. Chem.* **11**, 1227–1234
- 30 Sinclair, J., Delgoda, R., Noble, M., Jarmin, S., Goh, N. and Sim, E. (1998) Purification, characterization and crystallization of an N-hydroxyarylamine O-acetyltransferase from *Salmonella typhimurium*. *Protein Expression Purif.* **12**, 371–380
- 31 Massiah, M. A., Viragh, C., Reddy, P. M., Kovach, I. M., Johnson, J., Rosenberry, T. L. and Mildvan, A. S. (2001) Short, strong hydrogen bonds at the active site of human acetylcholinesterase: proton NMR studies. *Biochemistry* **40**, 5682–5690
- 32 Tai, D. F. and Liaw, W. C. (2002) Thiolsubtilisin acts as an acetyltransferase in organic solvents. *FEBS Lett.* **517**, 24–26
- 33 Zang, Y., Zhao, S., States, J. C. and Hein, D. W. (2004) Functional effects of the G364A (D122N) genetic polymorphism in human N-acetyltransferase2 (NAT2): disruption of catalytic core accelerates protein degradation? In Proceedings of the Third International Workshop on Arylamine N-acetyltransferases, 27–28 August 2004, Vancouver, Canada, abstract #20
- 34 Noguchi, K., Ishikawa, K., Yokoyama, K., Ohtsuka, T., Nio, N. and Suzuki, E. (2001) Crystal structure of red sea bream transglutaminase. *J. Biol. Chem.* **276**, 12055–12059
- 35 Krejci, E., Duval, N., Chatonnet, A., Vincens, P. and Massoulie, J. (1991) Cholinesterase-like domains in enzymes and structural proteins: functional and evolutionary relationships and identification of a catalytically essential aspartic acid. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6647–6651
- 36 Vernet, T., Tessier, D. C., Chatellier, J., Plouffe, C., Lee, T. S., Thomas, D. Y., Storer, A. C. and Menard, R. (1995) Structural and functional roles of asparagine 175 in the cysteine protease papain. *J. Biol. Chem.* **270**, 16645–16652
- 37 Scheibner, K. A., De Angelis, J., Burley, S. K. and Cole, P. A. (2002) Investigation of the roles of catalytic residues in serotonin N-acetyltransferase. *J. Biol. Chem.* **277**, 18118–18126
- 38 Vetting, M. W., Roderick, S. L., Yu, M. and Blanchard, J. S. (2003) Crystal structure of mycothiol synthase (Rv0819) from *Mycobacterium tuberculosis* shows structural homology to the GNAT family of N-acetyltransferases. *Protein Sci.* **12**, 1954–1959
- 39 Pompeo, F., Mushtaq, A. and Sim, E. (2002) Expression and purification of the rifamycin amide synthase, Riff, an enzyme homologous to the prokaryotic arylamine N-acetyltransferases. *Protein Expression Purif.* **24**, 138–151
- 40 Summerscales, J. E. and Josephy, P. D. (2004) Human acetyl-CoA:arylamine N-acetyltransferase variants generated by random mutagenesis. *Mol. Pharmacol.* **65**, 220–226
- 41 Delomenie, C., Goodfellow, G. H., Krishnamoorthy, R., Grant, D. M. and Dupret, J. M. (1997) Study of the role of the highly conserved residues Arg9 and Arg64 in the catalytic function of human N-acetyltransferases NAT1 and NAT2 by site-directed mutagenesis. *Biochem. J.* **323**, 207–215
- 42 Hasmann, M. J., Seidl, P. H., Engelhardt, G. and Schleifer, K. H. (1986) Acetyl-coenzyme A:arylamine N-acetyltransferases in microorganisms: screening and isolation of an enzyme from *Bacillus cereus*. *Arch. Microbiol.* **146**, 275–279
- 43 Kelly, S. L. and Sim, E. (1994) Arylamine NAT in Balb/c mice: identification of a novel mouse isoenzyme by cloning and expression *in vitro*. *Biochem. J.* **302**, 347–353
- 44 Fretland, A. J., Doll, M. A., Gray, K., Feng, Y. and Hein, D. W. (1997) Cloning, sequencing, and recombinant expression of NAT1, NAT2, and NAT3 derived from the C3H/HeJ (rapid) and A/HeJ (slow) acetylator inbred mouse: functional characterization of the activation and deactivation of aromatic amine carcinogens. *Toxicol. Appl. Pharmacol.* **142**, 360–366
- 45 Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., Scott, G., Steffen, D., Worley, K. C., Burch, P. E. et al. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature (London)* **428**, 493–521
- 46 Rodrigues-Lima, F. and Dupret, J. M. (2002) 3D model of human arylamine N-acetyltransferase 2: structural basis of the slow acetylator phenotype of the R64Q variant and analysis of the active-site loop. *Biochem. Biophys. Res. Commun.* **291**, 116–123
- 47 Rodrigues-Lima, F. and Dupret, J. M. (2002) *In silico* sequence analysis of arylamine N-acetyltransferases: evidence for an absence of lateral gene transfer from bacteria to vertebrates and first description of paralogs in bacteria. *Biochem. Biophys. Res. Commun.* **293**, 783–792
- 48 Brooke, E., Davies, S. G., Mulvaney, A. W., Okada, M., Pompeo, F., Sim, E., Vickers, R. J. and Westwood, I. M. (2003) Synthesis and *in vitro* evaluation of novel small molecule inhibitors of bacterial arylamine N-acetyltransferases (NATs). *Bioorg. Med. Chem. Lett.* **13**, 2527–2530
- 49 Rodrigues-Lima, F., Delomenie, C., Goodfellow, G. H., Grant, D. M. and Dupret, J. M. (2001) Homology modelling and structural analysis of human arylamine N-acetyltransferase NAT1: evidence for the conservation of a cysteine protease catalytic domain and an active-site loop. *Biochem. J.* **356**, 327–334
- 50 Delomenie, C., Fouix, S., Longuemaux, S., Brahimi, N., Bizet, C., Picard, B., Denamur, E. and Dupret, J.-M. (2001) Identification and functional characterization of arylamine N-acetyltransferases in eubacteria: evidence for highly selective acetylation of 5-aminosalicylic acid. *J. Bacteriol.* **183**, 3417–3427
- 51 Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723
- 52 Livingstone, C. D. and Barton, G. J. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput. Appl. Biosci.* **9**, 745–756
- 53 Clamp, M., Cuff, J., Searle, S. M. and Barton, G. J. (2004) The Jalview Java Alignment Editor. *Bioinformatics* **20**, 426–427

Received 14 February 2005/22 April 2005; accepted 4 May 2005

Published as BJ Immediate Publication 4 May 2005, doi:10.1042/BJ20050277