# SCIENTIFIC REPORTS

**OPEN**

# Comparative analysis of xenobiotic metabolising *N*-acetyltransferases from ten non-human primates as *in vitro* models of human homologues

Theodora Tsirka[1], Maria Konstantopoulou[1], Audrey Sabbagh[2,3], Brigitte Crouau-Roy[4,5], Ali Ryan [6,7], Edith Sim[6,7], Sotiria Boukouvala[1] & Giannoulis Fakis[1]

Xenobiotic metabolising *N*-acetyltransferases (NATs) perform biotransformation of drugs and carcinogens. Human NAT1 is associated with endogenous metabolic pathways of cells and is a candidate drug target for cancer. Human NAT2 is a well-characterised polymorphic xenobiotic metabolising enzyme, modulating susceptibility to drug-induced toxicity. Human NATs are difficult to express to high purification yields, complicating large-scale production for high-throughput screens or use in sophisticated enzymology assays and crystallography. We undertake comparative functional investigation of the NAT homologues of ten non-human primates, to characterise their properties and evaluate their suitability as models of human NATs. Considering the amount of generated recombinant protein, the enzymatic activity and thermal stability, the NAT homologues of non-human primates are demonstrated to be a much more effective resource for *in vitro* studies compared with human NATs. Certain NAT homologues are proposed as better models, such as the NAT1 of macaques *Macaca mulatta* and *M. sylvanus*, the NAT2 of *Erythrocebus patas*, and both NAT proteins of the gibbon *Nomascus gabriellae* which show highest homology to human NATs. This comparative investigation will facilitate *in vitro* screens towards discovery and optimisation of candidate pharmaceutical compounds for human NAT isoenzymes, while enabling better understanding of NAT function and evolution in primates.

The xenobiotic metabolising *N*-acetyltransferase (NAT) enzymes are involved in the biotransformation of drugs and carcinogens. NATs are conjugation enzymes typically employing acetyl-coenzyme A (CoA) as donor substrate to perform *N*-acetylation of an arylamine or arylhydrazine acceptor substrate during a two-step ping-pong Bi-Bi reaction catalysed by a triad of cysteine, histidine and aspartate residues[1,2]. The diversity of NAT homologues is extensive, as evidenced through previous comparative genomic surveys and phylogenetic analyses across a broad taxonomic range of prokaryotes and eukaryotes[3–6]. The phylogeny of *NAT* homologues in both simple (fungi) and complex (vertebrates) eukaryotes suggests a dynamic birth-and-death evolutionary pattern, where multiple lineage-specific gene duplications and losses, combined with selection-driven sequence divergence, have promoted functional diversification of NAT enzymes, presumably allowing rapid adaptation to variable xenobiotic environments[5,6]. Enzymatic and structure-function investigations, particularly with microbial NATs, support this model, as they reveal variability even in key catalytic components in certain organisms[7–10].

In primates, our phylogenetic analysis[6] predicted an initial duplication of the ancestral *NAT* gene in the common ancestor of Simiiformes, generating *NAT1* and a second *NAT* gene which subsequently duplicated in the common ancestor of Catarrhini (Old World monkeys and apes, including human) to give rise to *NAT2* and the pseudogene *NATP*. Platyrrhini (New World monkeys) appear to lack the third *NAT* locus, while other independent duplication events may have occurred in the lower primate lineages of Strepsirrhini and Tarsiiformes. NAT1

[1]Democritus University of Thrace, Department of Molecular Biology and Genetics, Alexandroupolis, Greece. [2]IRD UMR216, Mère et enfant face aux infections tropicales, Paris, France. [3]PRES Sorbonne Paris Cité, Université Paris Descartes, Faculté de Pharmacie, Paris, France. [4]CNRS, Université Paul Sabatier, ENFA, UMR5174EDB (Laboratoire Évolution & Diversité Biologique), Toulouse, France. [5]Université de Toulouse 3, UMR5174EDB, Toulouse, France. [6]University of Oxford, Department of Pharmacology, Oxford, UK. [7]Kingston University London, Faculty of Science, Computing and Engineering, Kingston upon Thames, UK. Sotiria Boukouvala and Giannoulis Fakis contributed equally to this work. Correspondence and requests for materials should be addressed to G.F. (email: gfakis@mbg.duth.gr)

evolution in primates appears to be primarily subject to purifying selection, potentially maintaining some endogenous metabolic function. NAT2 evolution is more likely driven by diversifying selection favouring variability in amino acid sequence, possibly as adaptation to changing xenobiotic exposures. These observations are supported by population genetic studies of human *NAT* genes, predicting that changing selective pressures on NAT2, rather than NAT1, may have favoured expansion of the slow acetylator genotype/phenotype after the agricultural revolution of the Neolithic period[11–16]. Human *NAT1* gene expression begins very early in embryogenesis[17] and is evident in a range of tissues, driven by a housekeeping promoter[18,19]. By contrast, human *NAT2* shows a more restricted expression pattern consistent with its well-established role as hepatic xenobiotic metabolising enzyme[20]. Polymorphisms are also more common and their effects on xenobiotic acetylation more apparent for human *NAT2* compared with *NAT1* gene[21].

Expanding previous *in silico* phylogenetic analyses of primate NATs[6], we undertook a comparative functional investigation of the NAT1 and NAT2 homologues of eleven primate species, including human. The recombinant NAT proteins of non-human primates are proposed as models of their human counterparts, particularly for *in vitro* screens towards the discovery and optimisation of candidate pharmaceutical compounds.

## Results and Discussion

### Primate *NAT* genes and proteins.
We have previously reported[6] sequencing of *NAT* genes from the non-human primates *Sapajus apella* (a New World monkey from South America), *Allenopithecus nigroviridis*, *Cercopithecus diana*, *Chlorocebus tantalus*, *Erythrocebus patas*, *Macaca sylvanus* and *Mandrillus sphinx* (Old World monkeys from across Africa), *Macaca mulatta* and *Trachypithecus cristatus* (Old World monkeys from South Asia), and *Nomascus gabriellae* (an ape from Southeast Asia). Of those ten species, the gibbon *N. gabriellae* is taxonomically more closely related to *Homo sapiens* (human). The corresponding consensus classification of these primates is shown in Supplementary Fig. S1, with additional information provided in Supplementary Table S1.

Suitable species-specific mnemonics were assigned to symbols of annotated *NAT* genes (Table 1), according to current consensus nomenclature rules[22], and complete coding sequences of 873 bp were cloned in expression vector. The translated 290 amino acid NAT sequences were highly conserved; identity ranged from 94.16 to 100% for NAT1 and from 85.57 to 99.66% for NAT2 sequences, with higher variability observed among NAT2 than among NAT1 homologues (Supplementary Fig. S2 and Supplementary Table S2). The phylogeny of protein sequences was in line with the species taxonomy of Supplementary Fig. S1, but the NAT1 and NAT2 homologues clustered as separate groups, supporting two distinct lineages of *NAT* orthologues in the simians (Supplementary Fig. S3). These observations are consistent with our earlier evolutionary studies of primate NATs[6].

Nine NAT1 homologues were expressed in recombinant form, representing all eleven primate species in this study, including human. NAT1 protein sequences were identical between *C. tantalus* and *E. patas*, as well as between *M. mulatta* and *M. sylvanus*. Expression of nine NAT2 homologues was also performed, representing all species except *A. nigroviridis* and *S. apella* (Table 1). Using a standard expression and affinity chromatography purification method, the 18 recombinant proteins were recovered at variable levels of yield and purity (Fig. 1). Expression of NAT homologues of non-human primates was consistently higher compared with (HUMAN) NAT1 and (HUMAN)NAT2 proteins, which were generated in very low amounts. Overall, the expression of non-human NAT1 homologues was considerably higher compared with non-human NAT2 homologues, enabling more effective purification of the former.

Unlike the (HUMAN)NAT proteins, further chromatographic purification was possible for the gibbon (NOMGA)NAT1 and (NOMGA)NAT2, as well as for other readily expressed NAT homologues of non-human primates. Under the gel filtration conditions used, the NAT proteins were recovered as monomers with the expected molecular mass (Supplementary Fig. S4). Purified proteins (ALLNI)NAT1, (NOMGA)NAT1 and (TRACR)NAT1 (Supplementary Fig. S5) were further selected for crystallisation screens with commercial panels (Morpheus®, JCSG plus™, Midas™, Structure Screen 1 & 2 HT-96 and Wizard, all from Molecular Dimensions), using the facilities of the Department of Biochemistry at Oxford University. Although these initial crystallisation attempts did not provide any useful crystals, the production of purified NAT proteins (3–10 mg per litre of culture) from various non-human primates is encouraging, laying the groundwork for future screens.

Previous studies have shown that human NATs are very difficult to express to high purification yields in *Escherichia coli*, complicating their large-scale production for high-throughput screens of chemical libraries, or for use in sophisticated crystallographic and enzymological investigations. For crystallisation, Wu and colleagues[23] generated recombinant (HUMAN)NAT1 and (HUMAN)NAT2 proteins to purification yields of 1.6 and 0.5 mg per litre of bacterial culture, respectively (http://www.thesgc.org/structures/2IJA; http://www.thesgc.org/structures/2PQT; http://www.thesgc.org/structures/2PFR). Another group has also been successful in producing fully purified (HUMAN)NAT1 and (HUMAN)NAT2 proteins to high yields (4 and 1.4 mg/L, respectively), and these preparations have been used to advance knowledge of NAT enzymatic catalysis and inhibition[24–27]. However, the bulk of studies measuring recombinant activity of human NAT isoenzymes have employed either crude lysates (of bacterial, yeast or mammalian cells) or partially purified proteins in low yields. In many instances, enzymatic activity of human NATs and their polymorphic variants has been measured in preparations where recombinant protein is detectable only by Western blot[28–40].

The recombinant NAT proteins of laboratory rodents have been popular substitutes of human NATs in studies of catalytic function or *in vitro* screens for pharmaceutical compounds[39–47]. However, compared with the NAT proteins of non-human primates, the rodent homologues are less suitable models, given their considerably lower identity to human NATs (Supplementary Table S2). Moreover, although human *NAT1* and rodent *Nat2* genes are orthologous (i.e. they share common phylogenetic ancestry and function), this is not true for human *NAT2* and rodent *Nat1* genes which have derived through independent genomic duplications occuring after the divergence of rodents and primates[6,44,47].

| Scientific name (synonym)[a] | Common name[a] | Taxon mnemonic[a] | Taxon ID[a] | Gene[b] | GenBank ID[b] | Amplification primers[c] |
|---|---|---|---|---|---|---|
| *Allenopithecus nigroviridis* | Allen's swamp monkey | ALLNI | 54135 | *NAT1* | KU640972 | #1 and #2 |
| *Cercopithecus diana* | Diana monkey | CERDI | 36224 | *NAT1* | KU640973 | #1 and #2 |
| | | | | *NAT2* | KU640982 | #1 and #3 |
| *Chlorocebus tantalus* (*Cercopithecus tantalus*) | Tantalus monkey | CHLTN | 60712 | *NAT1* | KU640974 | #6 and #7 |
| | | | | *NAT2* | KU640983 | #1 and #3 |
| *Erythrocebus patas* (*Cercopithecus patas*) | Red guenon | ERYPA | 9538 | *NAT1* | KU640975 | #6 and #7 |
| | | | | *NAT2* | KU640984 | #1 and #3 |
| *Homo sapiens* | Human | HUMAN | 9606 | *NAT1* | X17059 (allele *NAT1*4*) | #1 and #2 |
| | | | | *NAT2* | X14672 (allele *NAT2*4*) | Clone available |
| *Macaca mulatta* | Rhesus macaque | MACMU | 9544 | *NAT1* | KU640969 | #1 and #2 |
| | | | | *NAT2* | AJ504440 (allele *NAT2*1*) | Clone available |
| *Macaca sylvanus* | Barbary macaque | MACSY | 9546 | *NAT1* | KU640970 | #1 and #2 |
| | | | | *NAT2* | KU640978 | #1 and #3 |
| *Mandrillus sphinx* (*Papio sphinx*) | Mandrill | MANSP | 9561 | *NAT1* | KU640971 | #1 and #2 |
| | | | | *NAT2* | KU640980 | #1 and #3 |
| *Nomascus gabriellae* | Red-cheeked gibbon | NOMGA | 61852 | *NAT1* | KU640967 | #1 and #2 |
| | | | | *NAT2* | KU640985 | #1 and #4 |
| *Sapajus paella* (*Cebus apella*) | Brown-capped capuchin | SAPAP | 9515 | *NAT1* | KU640966 | #1 and #2 |
| *Trachypithecus cristatus* (*Presbytis cristata*) | Silvered leaf-monkey | TRACR | 122765 | *NAT1* | KU640968 | #1 and #2 |
| | | | | *NAT2* | KU640977 | #1 and #5 |

**Table 1.** Description of *NAT* gene constructs and their primate species of origin. [a]The current scientific names of species with synonymous names in parentheses (first column) and a common name (second column), are from the UniProt Taxonomy database (http://www.uniprot.org/taxonomy/). Official taxon mnemonics (third column) and taxon identification numbers (fourth column) are from the same source. According to current consensus nomenclature (http://nat.mbg.duth.gr/), taxon mnemonics are attached to the symbols of *NAT* genes to identify their specific organism of origin[22]. [b]All *NAT* gene annotations (fifth column) are as first described by[6], except for the *NAT2* gene of *M. mulatta*[60], and the *NAT1* and *NAT2* genes of *H. sapiens*[66]. The recombinant proteins expressed in this study were as predicted by translation of the nucleotide sequences for which GenBank identification numbers are provided (sixth column). [c]The complete coding sequence of each *NAT* gene was amplified for cloning, using combinations of the following primers, as indicated (seventh column): Primer #1 (forward), GCGGCAGCC<u>CATATG</u>GACATTGAAGCATA; Primer #2 (reverse), TCGAGT<u>GCGGCCGC</u>CTAAATAGTAAAAAATCTATCACC; Primer #3 (reverse), GTGCTCGAGT<u>GCGGCCGC</u>CTAAATAGTGAAGGATC; Primer #4 (reverse), TCGAGT<u>GCGGCCGC</u>CTAAATAGTAAGGGATCCATC; Primer #5 (reverse), TCGAGT<u>GCGGCCGC</u>CTAAATAGTAAAGAATCCATC. The recognition sites for restriction endonucleases *Nde*I (forward primer) and *Not*I (reverse primers) are underlined. Clones for the *NAT2* genes of *M. mulatta* and *H. sapiens* were available in our laboratories from previous studies[61]. The *NAT1* constructs of *C. tantalus* and *E. patas* were generated from the *NAT1* construct of *C. diana* via site-directed mutagenesis with Primer #6 (forward), CCATGGACTTAG**G**CTTAGAGGCCAT, and Primer #7 (reverse), ATGGCCTCTAAG**C**CTAAGTCCATGG (in bold, the mutagenised nucleotide representing the only non-synonymous SNP differentiating the *NAT1* gene of *C. diana* from the other two homologues). Amplification of the *NAT2* genes of *A. nigroviridis* (GenBank ID: KU640981) and *S. apella* (GeneBank ID: KU640976) was not possible due to limited availability of genomic DNA.

The advantages of expressing recombinant NAT1 and NAT2 proteins of non-human primates, for use as *in vitro* models of their human homologues, are obvious. In particular, the gibbon *N. gabriellae* could serve as a convenient model, since it was possible to efficiently express and purify its NAT isoenzymes (Supplementary Fig. S4 and Supplementary Fig. S5). Sequence identity is 95.88% between (HUMAN)NAT1 and (NOMGA)NAT1, as well as between (HUMAN)NAT2 and (NOMGA)NAT2, which corresponds to a difference of only 12 amino acids per each pair of homologues (Supplementary Fig. S2 and Supplementary Table S2).

**Enzymatic activities of primate NAT proteins.** All recombinant NAT proteins were assayed for enzymatic activity against the (HUMAN)NAT1 selective substrate *p*-aminobenzoate (PABA) and the (HUMAN)NAT2 selective substrates procainamide (PA, an antiarrhythmic arylamine) and isoniazid (INH, an anti-tubercular arylhydrazine). Assays were also performed with the less selective arylamine substrates 5-aminosalicylate (5AS) and *p*-anisidine (PANS)[35]. Overall, the NAT isoenzymes of non-human primates demonstrated the expected substrate specificity pattern, but enzyme activities varied considerably among species (Fig. 2).

The specific activities measured for (HUMAN)NAT1 and (HUMAN)NAT2 were lowest, even with selective substrates, and this was possibly due to the poor quality of recombinant preparations generated for those two proteins under the standard expression/purification conditions used (Fig. 1). However, the gibbon (NOMGA)NAT1 and (NOMGA)NAT2 isoenzymes provided relatively low specific activities too, despite their more robust recombinant expression and purification. Compared with other primates, NAT enzymatic activities may be lower
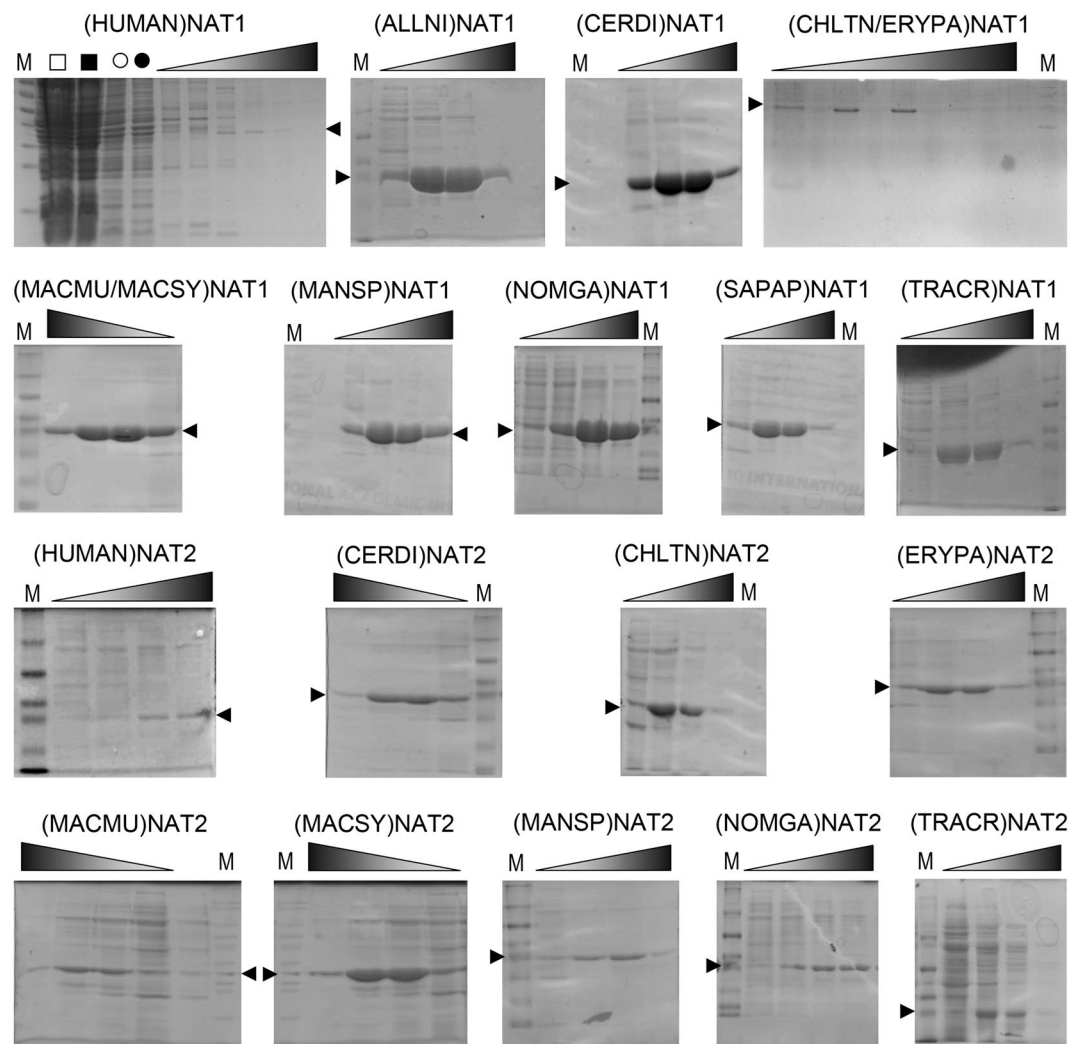
**Figure 1.** Recombinant NAT proteins of primates. The NAT proteins of *Homo sapiens* (HUMAN), *Allenopithecus nigroviridis* (ALLNI), *Cercopithecus diana* (CERDI), *Chlorocebus tantalus* (CHLTN), *Erythrocebus patas* (ERYPA), *Macaca mulatta* (MACMU), *Macaca sylvanus* (MACSY), *Mandrillus sphinx* (MANSP), *Nomascus gabriellae* (NOMGA), *Sapajus apella* (SAPAP) and *Trachypithecus cristatus* (TRACR) were expressed in *E. coli* and isolated by affinity chromatography. *C. tantalus* and *E. patas* have identical NAT1 amino acid sequences, thus only one recombinant protein was expressed, designated (CHLTN/ERYPA)NAT1. This was also the case for the identical NAT1 proteins of *M. mulatta* and *M. sylvanus*, expressed as (MACMU/MACSY)NAT1. SDS-PAGE gels are shown with chromatographic fractions eluted with a concentration gradient of up to 250 mM of imidazole, indicated by shaded triangles at the top of each image. Each lane was loaded with 30 µl of eluate. The symbols in the first image are whole cell lysate without IPTG induction (□), whole cell lysate after IPTG induction (■), soluble cell extract before the affinity chromatography (○), and initial flowthrough of soluble cell extract through the affinity chromatography column (●). The small black arrowheads indicate bands of recombinant NAT proteins, the molecular weight of which (~31 kDa) was estimated relative to protein markers (lanes M). Full-length gels are presented in Expanded Data Supplementary Fig. 1 at the end of the Supplementary Information file.

in human and other apes. This similarity, further supports the utility of (NOMGA)NAT1 and (NOMGA)NAT2 as *in vitro* models of the more poorly expressed human NATs.

The non-selective substrate PANS was efficient with both the NAT1 and the NAT2 isoenzymes of non-human primates (Fig. 2). The (MACMU/MACSY)NAT1 homologue of the two macaque species generated excessive activities with both PANS and PABA, which were at least 8-fold higher than the activities measured for all other NAT1 homologues (Fig. 2a). Similarly, (ERYPA)NAT2 was the most active NAT2 isoenzyme with PANS and its selective substrates PA and INH (Fig. 2b).

The highly active (MACMU/MACSY)NAT1 and (ERYPA)NAT2 homologues were further used to test selectivity against a series of additional NAT substrates, namely the toxic arylamines 2-aminophenol (2AP), 4-chloroaniline (CLA), 3,4-dichloroaniline (3,4DCA) and 4-phenoxyaniline (POA), and the pharmaceutical aryl-hydrazine hydralazine (HDZ)[35]. Both NAT homologues provided activity with all five substrates, but the activities measured with (ERYPA)NAT2 were substantially higher (>10,000 nmol/min/mg) compared with those measured
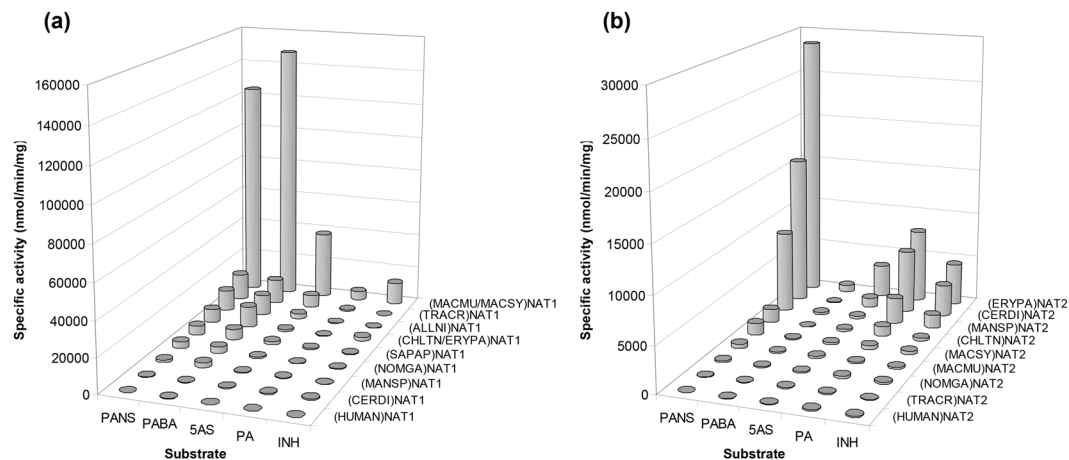
**Figure 2.** Enzyme activity screens of primate NAT homologues. Overview of the specific activities measured with the recombinant NAT1 (**a**) and NAT2 (**b**) isoenzymes of primate species *Allenopithecus nigroviridis* (ALLNI), *Cercopithecus diana* (CERDI), *Chlorocebus tantalus* (CHLTN), *Erythrocebus patas* (ERYPA), *Macaca mulatta* (MACMU), *Macaca sylvanus* (MACSY), *Mandrillus sphinx* (MANSP), *Nomascus gabriellae* (NOMGA), *Sapajus apella* (SAPAP), *Trachypithecus cristatus* (TRACR) and *Homo sapiens* (HUMAN). Assays were performed with acetyl-CoA as acyl-group donor substrate and a series of acceptor substrates, namely *p*-anisidine (PANS), *p*-aminobenzoate (PABA), 5-aminosalicylate (5AS), procainamide (PA) and isoniazid (INH). Enzyme activity was detected with all NAT homologues, but varied considerably per protein and substrate used. Assays were performed in duplicate and variability between replicates was within 10% from the average specific activity shown. A presentation of enzyme activity data is provided in Expanded Data Supplementary Fig. 2 at the end of the Supplementary Information file.

with (MACMU/MACSY)NAT1 (<3500 nmol/min/mg) (Fig. 3). The results suggest that 2AP, CLA, 3,4DCA, POA, and HDZ are more NAT2 selective, particularly taking into account that (MACMU/MACSY)NAT1 was about 4.5-fold more active than (ERYPA)NAT2 with the non-selective substrate PANS (Fig. 2). The pattern agrees with previous investigations with human NATs, generated in low yields detectable by Western blot[35].

Previous studies have demonstrated that NATs can utilise acyl-group donor substrates other than acetyl-CoA[10,35,48]. We thus tested (MACMU/MACSY)NAT1 and (ERYPA)NAT2 with acetyl-, *n*-propionyl-, malonyl-, succinyl- and hexanoyl-CoA as donor substrates. As demonstrated (Fig. 4), (MACMU/MACSY)NAT1 provided activities with all five acyl-CoA compounds, in the following declining order: acetyl-CoA (100%) > *n*-propionyl-CoA (50%) > malonyl-CoA (25%) > succinyl-CoA (15%) and hexanoyl-CoA (8%). Activities with these compounds have also been reported with hamster NAT2 recombinant protein, which is functionally equivalent to human NAT1[35]. The (ERYPA)NAT2 protein demonstrated more restricted acyl-CoA selectivity, utilising acetyl-CoA (100%) and *n*-propionyl-CoA (88%), but providing only marginal (<3%) activities with the remaining three acyl-CoA substrates tested. This pattern is more similar to that observed with certain microbial NATs, which are effective utilisers of acetyl- and *n*-propionyl-CoA, but show virtually no preference for other acyl-CoA compounds[10,48].

**Thermal stability of primate NAT proteins.** Differential Scanning Fluorimetry (DSF) is widely employed as a quick, sensitive, low-cost, high-throughput technique for initial screens of pharmaceutical targets against candidate inhibitors or other ligands[49,50]. We found that application of the technique with (HUMAN)NAT1 and (HUMAN)NAT2 is complicated by the poor quality of recombinant protein preparations generated with our expression/purification scheme, which is commonly used by many laboratories. In contrast, DSF was successful with the recombinant NAT proteins of non-human primates, allowing accurate assessment of their thermal stability with or without substrates (Figs 5 and S6). The use of recombinant NATs from non-human primates would, therefore, facilitate initial DSF-based screens of chemical libraries for identification of leads that may be further evaluated as potential inhibitors of the human homologues. For example, the gibbon (NOMGA)NAT1 and (NOMGA)NAT2 isoenzymes appeared quite robust in DSF assays, providing Tm values of 46.34 ± 0.11 and 51.42 ± 0.15 °C, respectively. As mentioned above, the gibbon NAT proteins are highly similar in amino acid sequence with their human homologues (Supplementary Table S2) and they are readily expressed and purified in recombinant form (Supplementary Fig. S4). Both (NOMGA)NAT isoenzymes provided Tm values in the higher range (Fig. 5), generating optimal DSF curves. The recombinant NAT homologues of the gibbon are, therefore, proposed as appropriate models of human NATs in DSF-based screens for candidate pharmaceutical compounds.

The determined Tm values for NAT homologues of non-human primates ranged from 34.01 ± 1.34 to 46.34 ± 0.11 °C (mean value 40.07 ± 3.78 °C) for NAT1, and from 47.18 ± 0.4 to 53.23 ± 1.77 °C (mean value 50.45 ± 2.12 °C) for NAT2, supporting that the NAT2 homologues are considerably more resilient to thermal denaturation than the NAT1 homologues, as has been reported previously for the human NAT isoenzymes[31]. Among NAT1 proteins, (NOMGA)NAT1 and (TRACR)NAT1 had the highest and lowest Tm values, respectively, translating into a difference of about 12 °C (Fig. 5a and c). Variability in Tm was more restricted among NAT2
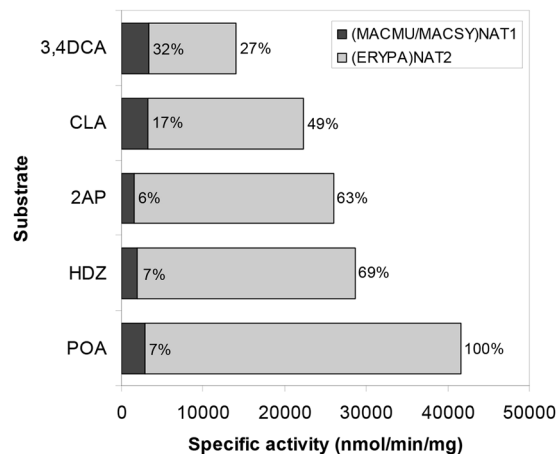
**Figure 3.** Additional acceptor substrate preferences of primate NAT1 and NAT2 homologues. The arylamines 4-phenoxyaniline (POA), 2-aminophenol (2AP), 4-chloroaniline (CLA) and 3,4-dichloroaniline (3,4DCA), and the arylhydrazine hydralazine (HDZ) were used in assays with either (MACMU/MACSY)NAT1 of the two macaques or (ERYPA)NAT2 of *Erythrocebus patas*. These were the recombinant NAT proteins providing highest activities in the screens of Fig. 2. In the plot, the light grey bars are the specific activities of (ERYPA) NAT2 with each one of the five substrates tested, also provided as percentages (%) relative to the specific activity measured with POA (100%). The dark grey bars are the specific activities of (MACMU/MACSY)NAT1, and each measurement is also provided relative (%) to the corresponding specific activity of (ERYPA)NAT2 per each substrate tested. Assays were performed in duplicate and variability between replicates was within 10% from the average specific activity shown. A presentation of enzyme activity data is provided in Expanded Data Supplementary Fig. 2 at the end of the Supplementary Information file.
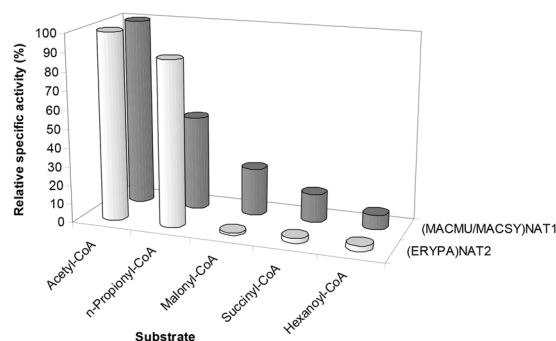


**Figure 4.** Donor substrate preferences of primate NAT1 and NAT2 homologues. Two parallel sets of assays were performed with acetyl-, *n*-propionyl-, malonyl-, succinyl- or hexanoyl-CoA as donor substrates. The first set employed (MACMU/MACSY)NAT1 protein with *p*-aminobenzoate (PABA) as acceptor substrate, while the second utilised (ERYPA)NAT2 protein with *p*-anisidine (PANS). On the graph, the bars of each series are the relative specific activities measured with each one of the five acyl-group donor substrates tested, presented as percentage (%) of the corresponding specific activity measured with acetyl-CoA (100%). Assays were performed in triplicate and variability between replicates was within 10% from the calculated average specific activity. A presentation of enzyme activity data is provided in Expanded Data Supplementary Fig. 2 at the end of the Supplementary Information file.

proteins (maximum difference of about 6 °C), with (MACSY)NAT2 and (MACMU)NAT2 providing the highest and lowest Tm values, respectively (Fig. 5b and d). This is noteworthy, since those two proteins are 99.31% identical and belong to species of the same genus (Supplementary Fig. S2 and Supplementary Table S2).

**Other comparative aspects of primate NAT1 homologues.** Human NAT1 has been the focus of investigations seeking to unravel its postulated role in endogenous metabolism and disease, particularly cancer[51–53]. Human NAT1 and its murine homologue have been implicated in folate metabolism via the acetylation of *p*-aminobenzoylglutamate (pABGlu)[51], although recent studies suggest a more complex association between folate and NAT[40,54,55]. We show here that NAT1 is capable of pABGlu acetylation in non-human primates too, as expected (Supplementary Fig. S7).

Moreover, human NAT1 has been investigated as a candidate target for small-molecule inhibition[56]. A naphthoquinone inhibitor developed for (HUMAN)NAT1[46], also inhibited (MACMU/MACSY)NAT1 at 10 μM
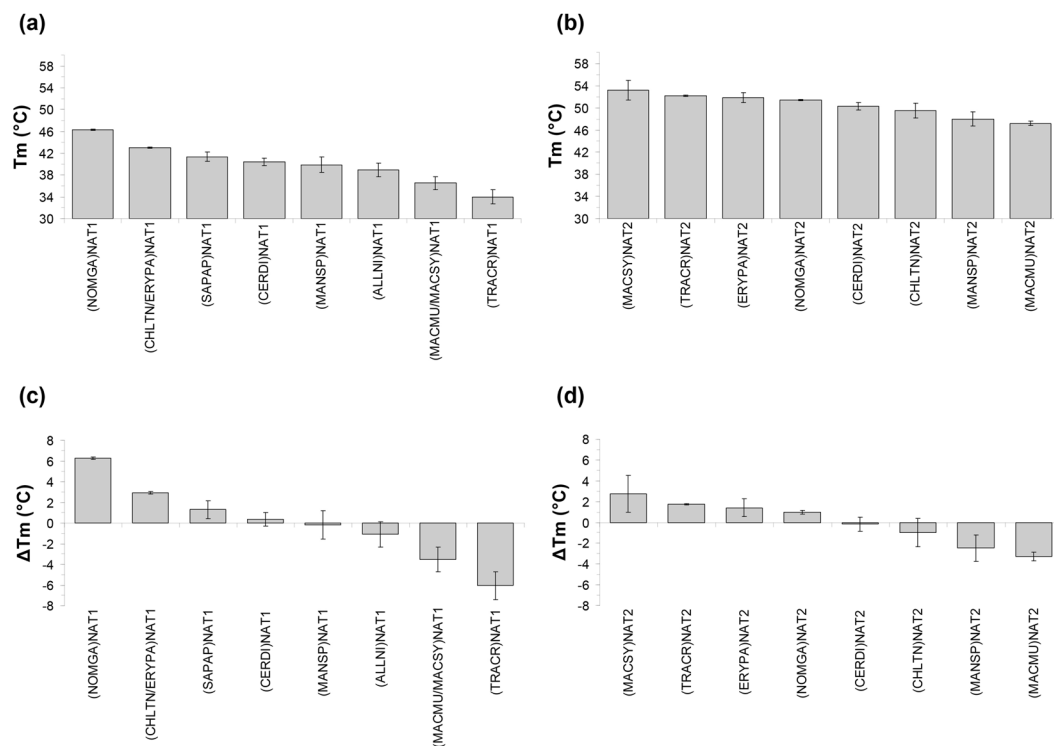
**Figure 5.** Thermal stability of primate NAT proteins. Overview of Tm values determined by DSF analysis of recombinant NAT1 (**a**) and NAT2 (**b**) homologues of non-human primates *Allenopithecus nigroviridis* (ALLNI), *Cercopithecus diana* (CERDI), *Chlorocebus tantalus* (CHLTN), *Erythrocebus patas* (ERYPA), *Macaca mulatta* (MACMU), *Macaca sylvanus* (MACSY), *Mandrillus sphinx* (MANSP), *Nomascus gabriellae* (NOMGA), *Sapajus apella* (SAPAP) and *Trachypithecus cristatus* (TRACR). The difference of each NAT1 (**c**) or NAT2 (**d**) Tm value from the corresponding calculated mean temperature is also plotted (the *x*-axis of graphs c and d corresponds to mean values of 40.07 °C and 50.45 °C for NAT1 and NAT2 proteins, respectively). Duplicate experiments were performed generating sigmoid curves that were fitted to the Boltzmann equation (Adj. R$^2$ ≥ 0.99). The screens were typically repeated multiple times with different protein preparations (particularly for low-Tm NAT homologues), in order to ensure accuracy and reproducibility of results.

concentration (Fig. 6a). However, inhibition was lower (50% decrease in enzyme activity with PABA) for (MACMU/MACSY)NAT1, compared with the reported 92% with 5AS for (HUMAN)NAT1[46]. Our enzymatic assays (Fig. 2a) show that (MACMU/MACSY)NAT1 is much more active than (HUMAN)NAT1, indicating that full inhibition would likely require a higher concentration of the naphthoquinone in the reaction. The *in silico* docking of the inhibitor to the modelled structure of (MACMU/MACSY)NAT1 protein indicated only a minor shift to the expected positioning within the active site. Docking of the inhibitor was the same between (HUMAN) NAT1 and the modelled structure of (NOMGA)NAT1 (Fig. 6b,c). The limited availability of the compound at the time of this study precluded more sophisticated assays using a concentration range of the inhibitor with NAT1 proteins from different non-human primates. Since those efficiently expressed homologues may constitute a useful resource for future high-throughput chemical screens, further optimisation of inhibition assays is warranted. DSF would also be useful as a more affordable high-throughput method to screen for NAT inhibitors.

**Structure-function comparisons of primate NAT proteins.** The comparative approach undertaken in this study places the phylogeny of primate NAT homologues in the context of enzymatic function. The evolution of NAT enzymes has presumably been influenced by xenobiotic challenges conferred by the various habitats of primate species, particularly their diet. Although primate NAT proteins are highly conserved in amino acid sequence, they differ considerably in enzymatic activity (Fig. 2). Associating genetic variability with enzymatic function allows the identification of amino acid residues that may have been influenced by environmental factors during primate NAT evolution. Such comparisons may also reveal apparently subtle differences between NAT homologues that may still have a significant phenotypic impact, further advancing our knowledge of NAT catalytic mechanism. This approach is focusing only on functional variability developed through the slow action of natural selection, so it is complementary to artificial mutagenesis of recombinant proteins which usually aims to alter specific residues with presupposed functional consequences[31,57–59]. To this end, some interesting comparisons are discussed below.

*(MACMU/MACSY)NAT1 vs. (MANSP)NAT1.* As mentioned above, the macaque (MACMU/MACSY)NAT1 was demonstrated to be considerably more active compared with other primate NAT1 homologues (Fig. 2a). It shares highest phylogenetic and primary sequence similarity (98.63%) with (MANSP)NAT1 which, however,
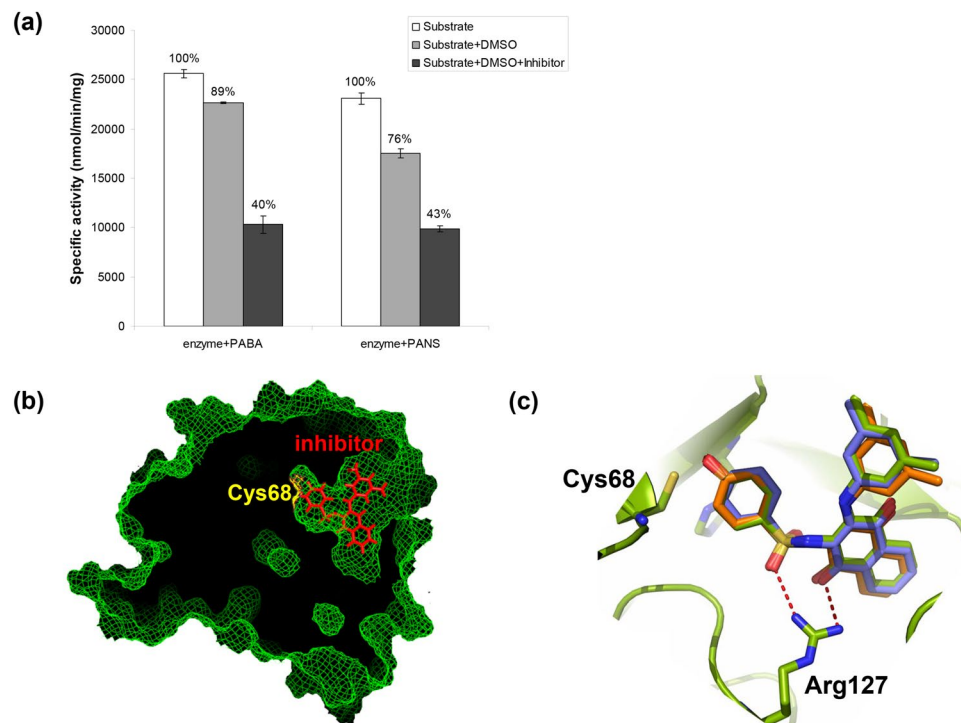
**Figure 6.** Interaction of primate NAT1 homologues with a small-molecule inhibitor. A synthetic naphthoquinone inhibitor of (HUMAN)NAT1 and (MOUSE)NAT2[46] was used. (**a**) Inhibition of (MACMU/MACSY)NAT1 homologue by 10 μM of the inhibitor in 5% (v/v) DMSO. Two replicate experiments were performed and the results are provided as average specific activity ± standard deviation. The enzymatic substrates used were *p*-aminobenzoate (PABA) and *p*-anisidine (PANS). (**b**) Overview of inhibitor binding to the active site of (MACMU/MACSY)NAT1 protein, modelled against the crystal structure of (HUMAN) NAT1 (PDB ID: 2PQT). (**c**) Detailed view of inhibitor binding to the superimposed models of (HUMAN) NAT1, (NOMGA)NAT1 and (MACMU/MACSY)NAT1 proteins. A slight shift in positioning of the inhibitor (particularly its *m*-xylene ring) within the active site of (MACMU/MACSY)NAT1 was observed, relative to the other two NAT1 homologues where binding showed complete overlap.

provided 55-fold lower specific activity with the selective substrate PABA. Both proteins were very effective during recombinant expression/purification (Fig. 1) and (MANSP)NAT1 had a higher Tm by 3.4 °C (Fig. 5a and c). Therefore, inherent instability of (MANSP)NAT1 protein is unlikely to explain its lower enzymatic activity relative to (MACMU/MACSY)NAT1. At the primary sequence level, the two homologues vary only at amino acid positions 29, 95, 273 and 280 (Supplementary Fig. S2). (MACMU/MACSY)NAT1 was the only NAT1 homologue in the study with glutamate at position 29, instead of glutamine. It was also the only NAT1 homologue with asparagine, instead of serine, at position 95. Conversely, (MANSP)NAT1 was the only homologue with valine, instead of isoleucine, at position 273. Position 280 is less conserved among the compared NAT1 sequences, bearing a methionine in (MACMU/MACSY)NAT1, an isoleucine in (MANSP)NAT1 and a valine or methionine in other primates.

Variant residue 29 is located at the α2-helix, on the surface of the protein molecule and away from the active site. In contrast, residue 95 is positioned within the catalytic pocket, next to the highly conserved Val93 which has been identified as one of the key hydrophobic residues involved in binding of NAT1 selective substrates[23] (Fig. 7a). More sophisticated structure-function investigations may reveal if the bulkier Asn95 of (MACMU/MACSY)NAT1 might confer increased enzymatic activity, relative to (MANSP)NAT1 and other primate homologues with serine at this position. Residues 273 and 280, on the other hand, are unlikely to directly interact with the active site, although they are located at the beginning of the carboxy-terminal tail, which in human NATs is folding into the catalytic pocket where it interacts with CoA[23].

*(ERYPA)NAT2 vs. (CERDI)NAT2 and (CHLTN)NAT2.* Among the primate NAT2 homologues examined, (ERYPA)NAT2 combined highest enzymatic activity with increased thermal stability (Figs 2b and 5b,d). It shares 99.66% and 99.31% primary sequence identity with its phylogenetically closest homologues (CERDI)NAT2 and (CHLTN)NAT2, respectively. Compared with other NAT2 homologues, those three proteins demonstrated robust recombinant expression and purification (Fig. 1). (ERYPA)NAT2 and (CERDI)NAT2 produced comparable levels of enzymatic activity which were considerably higher (10- to 20-fold with PANS) relative to (CHLTN)NAT2 (Fig. 2b). The latter protein also had the lowest Tm of the three, suggesting decreased stability (Fig. 5b and d).

(ERYPA)NAT2 differs from (CERDI)NAT2 only at amino acid position 276. Unlike other NAT2 sequences in this study, which have a glycine at position 276, (CERDI)NAT2 has a glutamate (Supplementary Fig. S2). This
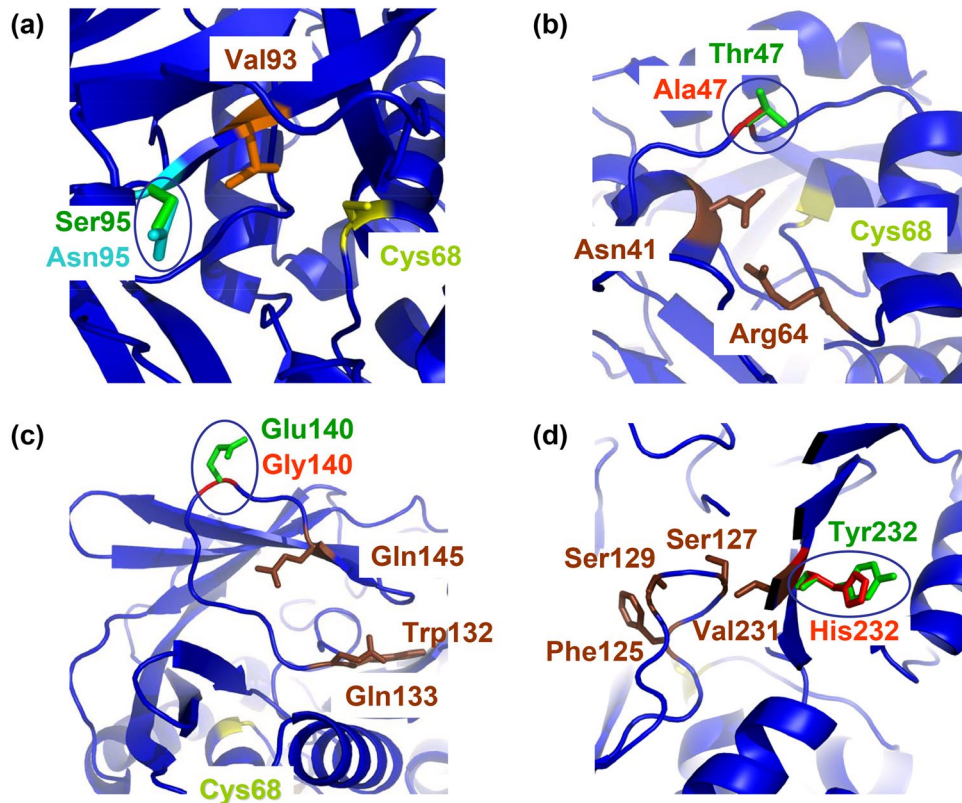
**Figure 7.** Examples of amino acid residues potentially differentiating the enzymatic function of highly homologous NAT proteins of non-human primates. (**a**) Detailed partial view of (MACMU/MACSY) NAT1, showing residue 95 (circled) and other proximal amino acids of relevance. Residue 95 is predicted to functionally differentiate (MACMU/MACSY)NAT1 (Asn95) from its 98.63% homologous (MANSP)NAT1 (Ser95). (**b,c**) Detailed partial view of (ERYPA)NAT2, showing circled residues 47 (**b**) and 140 (**c**), together with other proximal amino acids of relevance. Residues 47 and 140 are predicted to functionally differentiate (ERYPA)NAT2 (Ala47, Gly140) from its 99.31% homologous (CHLTN)NAT2 (Thr47, Glu140). (**d**) Detailed partial view of (MACMU)NAT2, showing residue 232 (circled) and other proximal amino acids of relevance. Residue 232 is predicted to functionally differentiate (MACMU)NAT2 (His232) from its 99.31% homologous (MACSY)NAT2 (Tyr232). The NAT1 and NAT2 homologues of non-human primates were modelled against the crystal structure of (HUMAN)NAT1 (PDB ID: 2PQT) and (HUMAN)NAT2 (PDB ID: 2PFR), respectively.

residue is located in the carboxy-terminal tail of NAT2, on the surface of the protein molecule, and has no apparent interaction with other amino acids. This single difference between (ERYPA)NAT2 and (CERDI)NAT2 is, therefore, unlikely to have any significant consequence on enzymatic function, consistent with the experimental findings.

(CHLTN)NAT2 is different from (ERYPA)NAT2 by two amino acids at positions 47 and 140. (CHLTN)NAT2 has a threonine and a glutamate, unlike other investigated NAT2 homologues which have an alanine and a glycine at these positions, respectively (Supplementary Fig. S2). The residue at position 47 is located behind the α-helix of the catalytic cysteine, and is also within spatial proximity to Asn41 and Arg64 which interact via hydrogen bonding to stabilise the active site[23]. It is possible that the bulkier polar threonine of (CHLTN)NAT2 destabilises this interaction, as opposed to the smaller hydrophobic alanine of (ERYPA)NAT2 and other NAT2 homologues (Fig. 7b). On the opposite side of the catalytic core, residue 140 is located right at the turn of an important loop stabilised through hydrogen bonding between Trp132, Gln133 and Gln145[23]. It is possible that the bulky negatively charged Glu140 of (CHLTN)NAT2 decreases the stability of this loop, in contrast to the small glycine of (ERYPA)NAT2 and other NAT2 homologues (Fig. 7c).

Polymorphisms affecting amino acid residues 64 (*NAT2*14* and *NAT2*19* alleles) and 145 (*NAT2*17* allele) have been reported for human NAT2 and they confer the slow acetylator phenotype (http://nat.mbg.duth.gr/). Slow acetylator variants are very common in human populations[16] and this should be expected to be the norm for populations of other primates too. It is, therefore, possible that the low activity of certain NAT homologues in this study is due to slow acetylator variants present in the single individual sampled per each primate species. To date, only two NAT polymorphic variants have been reported in the literature for (MACMU)NAT2 and these are functionally differentiated by a valine to isoleucine substitution at position 231[60,61]. Genotyping of a larger number of individuals would allow insight into *NAT* gene polymorphism in primates other than human.

*(MACSY)NAT2 vs. (MACMU)NAT2.* Another interesting observation arises from comparison of (MACSY) NAT2 and (MACMU)NAT2 belonging to the two macaque species (genus *Macaca*). As discussed above, the two homologues have 99.31% amino acid sequence identity, but the enzymatic activity of (MACSY)NAT2 was up

to 3-fold higher relative to (MACMU)NAT2 (Fig. 2b). Among all investigated NAT2 proteins, (MACSY)NAT2 provided highest and (MACMU)NAT2 lowest Tm values which were different by about 6 °C (Fig. 5b and d). Moreover, recombinant expression of (MACSY)NAT2 was considerably more efficient compared with (MACMU)NAT2 (Fig. 1). The two homologues are different by only two amino acids at positions 155 and 232, where (MACSY)NAT2 has glutamate and tyrosine, and (MACMU)NAT2 has glutamine and histidine, respectively (Supplementary Fig. S2).

Among all primate NAT proteins investigated, glutamine is found at position 155 only in (MACMU)NAT2 (Supplementary Fig. S2). However, this residue is located on the surface of the protein molecule and so it is predicted to have limited impact. Position 232, on the other hand, is located in the middle of β13-sheet, one of the four anti-parallel β-sheets of domain III. Sheets β12–β15 form a wall just behind loop 125–129 comprising important amino acids that determine substrate selectivity of human NAT isoenzymes[23,31]. The intermediate space is also accessible to the tip of the carboxy-terminal tail. Residue 232 is located right next to the above-mentioned (MACMU)NAT2 polymorphic site p.Val231Ile, which sterically interacts with loop 125–129[61]. Although the side chain of residue 232 is oriented to the opposite direction from loop 125–129, it is possible that the size and polarity of amino acid 232 may influence the structural integrity of β13-sheet, with impact on loop 125–129 and the carboxy-terminal chain (Fig. 7d).

## Conclusions

This study undertook comparative functional investigation of NAT homologues from 10 non-human primates, in order to characterise their properties and evaluate their suitability as models of human NATs, particularly for *in vitro* screens requiring large quantities of pure recombinant protein. Considering factors, such as the overall amount of expressed recombinant protein, the enzymatic activity, the thermal stability and the percent identity to human NATs, certain homologues are proposed as more appropriate, i.e. (MACMU/MACSY)NAT1 and (ERYPA)NAT2, as well as the two NAT homologues of the gibbon *N. gabriellae* which show higher homology to human NATs. The practicality of using such homologues (instead of hard-to-express human NAT proteins) for standard *in vitro* screens lies in that the NAT proteins of non-human primates demonstrate higher sequence identity to human NATs (compared with previously used rodent homologues), and in that recombinant protein production is readily achievable once the *NAT* genes of non-human primates have been amplified and cloned from tiny amounts of genomic DNA. However, currently available rodent models for NAT[20] should remain relevant, if suitable compounds become validated as candidates for *in vivo* preclinical tests in the future. Moreover, the comparative investigation of *NAT* genes and their enzyme products in non-human primates allows better understanding of their role in xenobiotic metabolism and facilitates the identification of functional residues that may have been subject to evolutionary pressures. To that end, genotyping of multiple individuals from different non-human primate species may help identify genetic variability similar to that observed for *NAT1* and *NAT2* genes in human populations.

## Methods

**Generation of gene constructs.**    Polymerase chain reaction (PCR) was performed with *Pfu*-DNA polymerase, using 10 ng of genomic DNA template. Primers (VBC Biotech) incorporating restriction sites at the ends of each PCR product were used according to the scheme of Table 1, and the primer annealing temperature was 56 °C. Subsequent restriction cloning into the pET28b(+) plasmid was carried out as described[61]. When necessary, site-directed mutagenesis was performed with the QuikChange II kit (Agilent) and appropriate mutagenic primers. The insert of each construct was sequenced (GATC Biotech) with vector-specific primers to verify in-frame translation into the expected N-terminal hexa-histidine tagged NAT amino acid sequence. Computational analyses of sequences (e.g. alignments, calculation of percent identities and phylogeny) were carried out with BioEdit Sequence Alignment Editor and MEGA[62]. All *NAT* gene constructs were maintained as clones in *E. coli* strain BL21(DE3)pLysS (Promega).

**Expression and chromatographic purification of recombinant proteins.**    Expression of recombinant proteins was performed in 250 ml of Terrific Broth medium with 50 μg/ml kanamycin, during overnight incubation at 15 °C (120 rpm), after induction with 1 mM isopropyl-1-thio-D-galactopyranoside (IPTG) at culture optical density (600 nm) of approximately 1. Soluble fractions of bacterial lysates were recovered in 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 0.1% (w/v) 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 5 mM imidazole, 5% (v/v) glycerol, 2 mM β-mercaptoethanol and a commercial cocktail of protease inhibitors. Affinity chromatography took place through Ni-NTA resin (Qiagen) and recombinant proteins were eluted with increasing amounts of imidazole (25, 50, 100 and 250 mM). Procedures were performed as described[10].

For proteins further purified by ion exchange and gel filtration chromatography, recombinant expression took place in 500 ml cultures and affinity chromatography followed according to the procedures above. Ion exchange chromatography was performed with an ÄKTA Purifier Core System carrying a positively charged Mono Q 5/50 GL column (GE Healthcare). This column was appropriate, as the isoelectric point (pI) of NAT proteins was calculated to range between 5 and 6, and the protein molecules were thus expected to be negatively charged in the chromatographic buffer used (pH 7.4). A concentration gradient (50 mM to 1 M) of NaCl was achieved by gradual mixing of buffer A (20 mM Tris-HCl pH 7.4, 50 mM ε-aminocaproic acid, 5 mM EDTA) with buffer B (buffer A plus 1 M NaCl), and the procedure was monitored by measuring absorbance at 280 nm.

Gel filtration chromatography was performed using a Superdex 75 10/300 GL column attached to the ÄKTA Purifier Core System (GE Healthcare). The expected molecular weight of mammalian NAT proteins is 30–35 kDa. Protein separation took place in buffer containing 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, and the procedure was monitored at 280 nm. Molecular weight standards (Sigma-Aldrich) comprised proteins (0.5 mg each)

bovine serum albumin (66.8 kDa), ovalbumin (44.3 kDa), soybean trypsin inhibitor (21.5 kDa) and bovine heart cytochrome c (12.4 kDa), which were eluted at volumes of 3.5, 4.5, 6 and 7 ml, respectively.

Buffer change of protein preparations was performed between different chromatographic procedures by overnight (4 °C) dialysis through 10 kDa molecular weight cut-off pore size membrane (Sigma-Aldrich). Concentration of protein solutions was carried out by centrifugation (4 °C) through filter devices (Ultracel YM-3, Microcon) with molecular weight cut-off pore size of 3 kDa. Removal of the N-terminal hexa-histidine tag of recombinant proteins was achieved via overnight (4 °C) cleavage by 1 U of human plasma thrombin (Sigma-Aldrich) per milligram of protein in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl. Protein concentration was determined spectrophotometrically at 280 nm, and preparations were inspected by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) after each purification process. Yields of incompletely purified recombinant proteins were estimated by densitometric analysis of gels, as described[61].

**Enzyme activity and inhibition assays.** The colorimetric method with Ellman's reagent (5,5′-dithiobis-2-nitrobenzoic acid) was used to measure NAT enzyme activity of recombinant proteins[63], following our standardised in-house protocol[10,61]. Replicate reactions were performed with acetyl-CoA (CID: 444493), *n*-propionyl-CoA (CID: 92753), malonyl-CoA (CID: 644066), succinyl-CoA (CID: 92133) or hexanoyl-CoA (CID: 449118) as donor substrates, and 2-aminophenol (2AP; CID: 5801), 4-chloroaniline (CLA; CID: 7812), *p*-anisidine (PANS; CID: 7732), 4-phenoxyaniline (POA; CID: 8764), *p*-aminobenzoate (PABA; CID: 4876), 5-aminosalicylate (5AS; CID: 4075), 3,4-dichloroaniline (3,4DCA; CID: 7257), procainamide (PA; CID: 4913), *p*-aminobenzoylglutamate (pABGlu; CID: 196473), isoniazid (INH; CID: 3767) or hydralazine (HDZ; CID: 3637) as acceptor substrates. The amount of dimethylsulphoxide (DMSO) per reaction was 0.5% (v/v). All compounds were from Sigma-Aldrich.

Inhibition assays with a previously characterised human NAT1 selective naphthoquinone inhibitor[46] were carried out using 1 μg of affinity chromatography purified recombinant NAT1 protein mixed with 10 μM of inhibitor in 20 mM Tris-HCl (pH 7.5), 5% (v/v) DMSO, prior to addition of 0.5 mM of substrate (PANS or PABA). Reactions were initiated with 0.5 mM acetyl-CoA and stopped after 30 s for colorimetric measurement of released CoA. Control reactions with 5% (v/v) DMSO but no inhibitor were also performed.

**Differential scanning fluorimetry.** Differential scanning fluorimetry (DSF) was performed exactly as described[10,61] to determine the denaturation midpoint temperatures ($T_m$) of different NAT proteins, as well as to examine the effects of compound binding.

**Molecular modelling of protein structures.** The published crystal structures[23] of human NAT1 (PDB ID: 2PQT) and NAT2 (PDB ID: 2PFR) proteins were used to model the secondary and tertiary structures of NAT homologues from other primates. Alignments of non-human NATs to the secondary structure of human NATs were performed with T-COFFEE Expresso (http://tcoffee.crg.cat/apps/tcoffee/do:expresso)[64] and graphically visualised with ESPript3.0 (http://espript.ibcp.fr)[65]. The tertiary structure of non-human NATs was modelled to the three-dimensional structure of human NATs using SWISS-MODEL (http://swissmodel.expasy.org/) and graphically visualised on PyMOL (Schrödinger, LLC). Docking analysis was performed as described[39]. In the available structure of human NAT1, the catalytic Cys68 appears modified as S-(2-anilino-2-oxoethyl)-L-cysteine and this modification was removed to restore the functional residue.

## References

1. Weber, W. W., Cohen, S. N. & Steinberg, M. S. Purification and properties of N-acetyltransferase from mammalian liver. *Ann N Y Acad Sci* **151**, 734–741 (1968).
2. Sinclair, J. C., Sandy, J., Delgoda, R., Sim, E. & Noble, M. E. Structure of arylamine N-acetyltransferase reveals a catalytic triad. *Nat Struct Biol* **7**, 560–564, https://doi.org/10.1038/76783 (2000).
3. Vagena, E., Fakis, G. & Boukouvala, S. Arylamine N-acetyltransferases in prokaryotic and eukaryotic genomes: a survey of public databases. *Curr Drug Metab* **9**, 628–660 (2008).
4. Glenn, A. E., Karagianni, E. P., Ulndreaj, A. & Boukouvala, S. Comparative genomic and phylogenetic investigation of the xenobiotic metabolizing arylamine N-acetyltransferase enzyme family. *FEBS Lett* **584**, 3158–3164, https://doi.org/10.1016/j.febslet.2010.05.063 (2010).
5. Martins, M., Dairou, J., Rodrigues-Lima, F., Dupret, J. M. & Silar, P. Insights into the phylogeny or arylamine N-acetyltransferases in fungi. *J Mol Evol* **71**, 141–152, https://doi.org/10.1007/s00239-010-9371-x (2010).
6. Sabbagh, A. *et al.* Rapid birth-and-death evolution of the xenobiotic metabolizing NAT gene family in vertebrates with evidence of adaptive selection. *BMC Evol Biol* **13**, 62, https://doi.org/10.1186/1471-2148-13-62 (2013).
7. Yu, T. W. *et al.* Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proc Natl Acad Sci USA* **96**, 9051–9056 (1999).
8. Pluvinage, B. *et al.* The Bacillus anthracis arylamine N-acetyltransferase ((BACAN)NAT1) that inactivates sulfamethoxazole, reveals unusual structural features compared with the other NAT isoenzymes. *FEBS Lett* **585**, 3947–3952, https://doi.org/10.1016/j.febslet.2011.10.041 (2011).
9. Kubiak, X. *et al.* Structural and biochemical characterization of an active arylamine N-acetyltransferase possessing a non-canonical Cys-His-Glu catalytic triad. *J Biol Chem* **288**, 22493–22505, https://doi.org/10.1074/jbc.M113.468595 (2013).
10. Karagianni, E. P. *et al.* Homologues of xenobiotic metabolizing N-acetyltransferases in plant-associated fungi: Novel functions for an old enzyme family. *Sci Rep* **5**, 12900, https://doi.org/10.1038/srep12900 (2015).
11. Patin, E. *et al.* Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. *Am J Hum Genet* **78**, 423–436, https://doi.org/10.1086/500614 (2006).
12. Patin, E. *et al.* Sub-Saharan African coding sequence variation and haplotype diversity at the NAT2 gene. *Hum Mutat* **27**, 720, https://doi.org/10.1002/humu.9438 (2006).
13. Magalon, H. *et al.* Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. *Eur J Hum Genet* **16**, 243–251, https://doi.org/10.1038/sj.ejhg.5201963 (2008).
14. Luca, F. *et al.* Multiple advantageous amino acid variants in the NAT2 gene in human populations. *Plos One* **3**, e3136, https://doi.org/10.1371/journal.pone.0003136 (2008).

15. Mortensen, H. M. *et al.* Characterization of genetic variation and natural selection at the arylamine N-acetyltransferase genes in global human populations. *Pharmacogenomics* **12**, 1545–1558, https://doi.org/10.2217/pgs.11.88 (2011).

16. Sabbagh, A., Darlu, P., Crouau-Roy, B. & Poloni, E. S. Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey. *Plos One* **6**, e18507, https://doi.org/10.1371/journal.pone.0018507 (2011).

17. Smelt, V. A. *et al.* Expression of arylamine N-acetyltransferases in pre-term placentas and in human pre-implantation embryos. *Hum Mol Genet* **9**, 1101–1107 (2000).

18. Boukouvala, S. & Sim, E. Structural analysis of the genes for human arylamine N-acetyltransferases and characterisation of alternative transcripts. *Basic Clin Pharmacol Toxicol* **96**, 343–351, https://doi.org/10.1111/j.1742-7843.2005.pto_02.x (2005).

19. Husain, A. *et al.* Functional analysis of the human N-acetyltransferase 1 major promoter: quantitation of tissue expression and identification of critical sequence elements. *Drug Metab Dispos* **35**, 1649–1656, https://doi.org/10.1124/dmd.107.016485 (2007).

20. Boukouvala, S. & Fakis, G. Arylamine N-acetyltransferases: what we learn from genes and genomes. *Drug Metab Rev* **37**, 511–564, https://doi.org/10.1080/03602530500251204 (2005).

21. Hein, D. W. N-acetyltransferase SNPs: emerging concepts serve as a paradigm for understanding complexities of personalized medicine. *Expert Opin Drug Metab Toxicol* **5**, 353–366, https://doi.org/10.1517/17425250902877698 (2009).

22. Hein, D. W., Boukouvala, S., Grant, D. M., Minchin, R. F. & Sim, E. Changes in consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenet Genomics* **18**, 367–368, https://doi.org/10.1097/FPC.0b013e3282f60db0 (2008).

23. Wu, H. *et al.* Structural basis of substrate-binding specificity of human arylamine N-acetyltransferases. *J Biol Chem* **282**, 30189–30197, https://doi.org/10.1074/jbc.M704138200 (2007).

24. Wang, H. *et al.* Over-expression, purification, and characterization of recombinant human arylamine N-acetyltransferase 1. *Protein J* **24**, 65–77 (2005).

25. Liu, L. *et al.* Arylamine N-acetyltransferases: characterization of the substrate specificities and molecular interactions of environmental arylamines with human NAT1 and NAT2. *Chem Res Toxicol* **20**, 1300–1308, https://doi.org/10.1021/tx7001614 (2007).

26. Liu, L., Wagner, C. R. & Hanna, P. E. Human arylamine N-acetyltransferase 1: *in vitro* and intracellular inactivation by nitrosoarene metabolites of toxic and carcinogenic arylamines. *Chem Res Toxicol* **21**, 2005–2016, https://doi.org/10.1021/tx800215h (2008).

27. Liu, L., Wagner, C. R. & Hanna, P. E. Isoform-selective inactivation of human arylamine N-acetyltransferases by reactive metabolites of carcinogenic arylamines. *Chem Res Toxicol* **22**, 1962–1974, https://doi.org/10.1021/tx9002676 (2009).

28. Hein, D. W., Ferguson, R. J., Doll, M. A., Rustan, T. D. & Gray, K. Molecular genetics of human polymorphic N-acetyltransferase: enzymatic analysis of 15 recombinant wild-type, mutant, and chimeric NAT2 allozymes. *Hum Mol Genet* **3**, 729–734 (1994).

29. Doll, M. A., Jiang, W., Deitz, A. C., Rustan, T. D. & Hein, D. W. Identification of a novel allele at the human NAT1 acetyltransferase locus. *Biochem Biophys Res Commun* **233**, 584–591, https://doi.org/10.1006/bbrc.1997.6501 (1997).

30. de Leon, J. H., Vatsis, K. P. & Weber, W. W. Characterization of naturally occurring and recombinant human N-acetyltransferase variants encoded by NAT1. *Mol Pharmacol* **58**, 288–299 (2000).

31. Goodfellow, G. H., Dupret, J. M. & Grant, D. M. Identification of amino acids imparting acceptor substrate selectivity to human arylamine acetyltransferases NAT1 and NAT2. *Biochem J* **348**(Pt 1), 159–166 (2000).

32. Fretland, A. J., Doll, M. A., Leff, M. A. & Hein, D. W. Functional characterization of nucleotide polymorphisms in the coding region of N-acetyltransferase 1. *Pharmacogenetics* **11**, 511–520 (2001).

33. Fretland, A. J., Leff, M. A., Doll, M. A. & Hein, D. W. Functional characterization of human N-acetyltransferase 2 (NAT2) single nucleotide polymorphisms. *Pharmacogenetics* **11**, 207–215 (2001).

34. Zang, Y., Zhao, S., Doll, M. A., States, J. C. & Hein, D. W. The T341C (Ile114Thr) polymorphism of N-acetyltransferase 2 yields slow acetylator phenotype by enhanced protein degradation. *Pharmacogenetics* **14**, 717–723 (2004).

35. Kawamura, A. *et al.* Eukaryotic arylamine N-acetyltransferase. Investigation of substrate specificity by high-throughput screening. *Biochem Pharmacol* **69**, 347–359, https://doi.org/10.1016/j.bcp.2004.09.014 (2005).

36. Zang, Y., Doll, M. A., Zhao, S., States, J. C. & Hein, D. W. Functional characterization of single-nucleotide polymorphisms and haplotypes of human N-acetyltransferase 2. *Carcinogenesis* **28**, 1665–1671, https://doi.org/10.1093/carcin/bgm085 (2007).

37. Zang, Y., Zhao, S., Doll, M. A., Christopher States, J. & Hein, D. W. Functional characterization of the A411T (L137F) and G364A (D122N) genetic polymorphisms in human N-acetyltransferase 2. *Pharmacogenet Genomics* **17**, 37–45, https://doi.org/10.1097/01.fpc.0000236325.73186.2c (2007).

38. Zhu, Y. & Hein, D. W. Functional effects of single nucleotide polymorphisms in the coding region of human N-acetyltransferase 1. *Pharmacogenomics J* **8**, 339–348, https://doi.org/10.1038/sj.tpj.6500483 (2008).

39. Laurieri, N. *et al.* A novel color change mechanism for breast cancer biomarker detection: naphthoquinones as specific ligands of human arylamine N-acetyltransferase 1. *Plos One* **8**, e70600, https://doi.org/10.1371/journal.pone.0070600 (2013).

40. Stepp, M. W., Mamaliga, G., Doll, M. A., States, J. C. & Hein, D. W. Folate-Dependent Hydrolysis of Acetyl-Coenzyme A by Recombinant Human and Rodent Arylamine N-Acetyltransferases. *Biochem Biophys Rep* **3**, 45–50, https://doi.org/10.1016/j.bbrep.2015.07.011 (2015).

41. Sticha, K. R., Sieg, C. A., Bergstrom, C. P., Hanna, P. E. & Wagner, C. R. Overexpression and large-scale purification of recombinant hamster polymorphic arylamine N-acetyltransferase as a dihydrofolate reductase fusion protein. *Protein Expr Purif* **10**, 141–153 (1997).

42. Wang, H., Vath, G. M., Gleason, K. J., Hanna, P. E. & Wagner, C. R. 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, https://doi.org/10.1021/bi0497244 (2004).

43. Wang, H., Wagner, C. R. & Hanna, P. E. Irreversible inactivation of arylamine N-acetyltransferases in the presence of N-hydroxy-4-acetylaminobiphenyl: a comparison of human and hamster enzymes. *Chem Res Toxicol* **18**, 183–197, https://doi.org/10.1021/tx049801w (2005).

44. Kawamura, A. *et al.* Mouse N-acetyltransferase type 2, the homologue of human N-acetyltransferase type 1. *Biochem Pharmacol* **75**, 1550–1560, https://doi.org/10.1016/j.bcp.2007.12.012 (2008).

45. Russell, A. J. *et al.* Selective small molecule inhibitors of the potential breast cancer marker, human arylamine N-acetyltransferase 1, and its murine homologue, mouse arylamine N-acetyltransferase 2. *Bioorg Med Chem* **17**, 905–918, https://doi.org/10.1016/j.bmc.2008.11.032 (2009).

46. Laurieri, N. *et al.* Small molecule colorimetric probes for specific detection of human arylamine N-acetyltransferase 1, a potential breast cancer biomarker. *J Am Chem Soc* **132**, 3238–3239, https://doi.org/10.1021/ja909165u (2010).

47. Laurieri, N. *et al.* Differences between murine arylamine N-acetyltransferase type 1 and human arylamine N-acetyltransferase type 2 defined by substrate specificity and inhibitor binding. *BMC Pharmacol Toxicol* **15**, 68, https://doi.org/10.1186/2050-6511-15-68 (2014).

48. Lack, N. A. *et al.* Temperature stability of proteins essential for the intracellular survival of Mycobacterium tuberculosis. *Biochem J* **418**, 369–378, https://doi.org/10.1042/bj20082011 (2009).

49. DeSantis, K., Reed, A., Rahhal, R. & Reinking, J. Use of differential scanning fluorimetry as a high-throughput assay to identify nuclear receptor ligands. *Nucl Recept Signal* **10**, e002, https://doi.org/10.1621/nrs.10002 (2012).

50. Velikova, N. *et al.* Putative histidine kinase inhibitors with antibacterial effect against multi-drug resistant clinical isolates identified by *in vitro* and in silico screens. *Sci Rep* **6**, 26085, https://doi.org/10.1038/srep26085 (2016).

51. Butcher, N. J. & Minchin, R. F. Arylamine N-acetyltransferase 1: a novel drug target in cancer development. *Pharmacol Rev* **64**, 147–165, https://doi.org/10.1124/pr.110.004275 (2012).

52. Stepp, M. W., Doll, M. A., Carlisle, S. M., States, J. C. & Hein, D. W. Genetic and small molecule inhibition of arylamine N-acetyltransferase 1 reduces anchorage-independent growth in human breast cancer cell line MDA-MB-231. *Mol Carcinog* **57**, 549–558, https://doi.org/10.1002/mc.22779 (2018).
53. Wang, L., Minchin, R. F. & Butcher, N. J. Arylamine N-acetyltransferase 1 protects against reactive oxygen species during glucose starvation: Role in the regulation of p53 stability. *Plos One* **13**, e0193560, https://doi.org/10.1371/journal.pone.0193560 (2018).
54. Laurieri, N. *et al*. From arylamine N-acetyltransferase to folate-dependent acetyl CoA hydrolase: impact of folic acid on the activity of (HUMAN)NAT1 and its homologue (MOUSE)NAT2. *Plos One* **9**, e96370, https://doi.org/10.1371/journal.pone.0096370 (2014).
55. Witham, K. L., Minchin, R. F. & Butcher, N. J. Role for human arylamine N-acetyltransferase 1 in the methionine salvage pathway. *Biochem Pharmacol* **125**, 93–100, https://doi.org/10.1016/j.bcp.2016.11.015 (2017).
56. Sim, E., Fakis, G., Laurieri, N. & Boukouvala, S. Arylamine N-acetyltransferases–from drug metabolism and pharmacogenetics to identification of novel targets for pharmacological intervention. *Adv Pharmacol* **63**, 169–205, https://doi.org/10.1016/b978-0-12-398339-8.00005-7 (2012).
57. Dupret, J. M. & Grant, D. M. 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 (1992).
58. Delomenie, C., Goodfellow, G. H., Krishnamoorthy, R., Grant, D. M. & Dupret, J. M. 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**(Pt 1), 207–215 (1997).
59. Minchin, R. F. & Butcher, N. J. The role of lysine(100) in the binding of acetylcoenzyme A to human arylamine N-acetyltransferase 1: implications for other acetyltransferases. *Biochem Pharmacol* **94**, 195–202, https://doi.org/10.1016/j.bcp.2015.01.015 (2015).
60. Fakis, G., Boukouvala, S., Kawamura, A. & Kennedy, S. Description of a novel polymorphic gene encoding for arylamine N-acetyltransferase in the rhesus macaque (Macaca mulatta), a model animal for endometriosis. *Pharmacogenet Genomics* **17**, 181–188, https://doi.org/10.1097/FPC.0b013e328011e3ad (2007).
61. Tsirka, T., Boukouvala, S., Agianian, B. & Fakis, G. Polymorphism p.Val231Ile alters substrate selectivity of drug-metabolizing arylamine N-acetyltransferase 2 (NAT2) isoenzyme of rhesus macaque and human. *Gene* **536**, 65–73, https://doi.org/10.1016/j.gene.2013.11.085 (2014).
62. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599, https://doi.org/10.1093/molbev/msm092 (2007).
63. Brooke, E. W. *et al*. An approach to identifying novel substrates of bacterial arylamine N-acetyltransferases. *Bioorg Med Chem* **11**, 1227–1234 (2003).
64. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205–217, https://doi.org/10.1006/jmbi.2000.4042 (2000).
65. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320–324, https://doi.org/10.1093/nar/gku316 (2014).
66. Blum, M., Grant, D. M., McBride, W., Heim, M. & Meyer, U. A. Human arylamine N-acetyltransferase genes: isolation, chromosomal localization, and functional expression. *DNA Cell Biol* **9**, 193–203 (1990).

## Acknowledgements

## Author Contributions

T.T. performed the experimental procedures; M.K. contributed to cloning; A.S. and B.C.R. provided expert scientific input and resources relevant to genomic DNA and *NAT* gene sequences of primates; A.R. and E.S. supported the purification and crystallographic trials of recombinant NAT proteins, and provided expert technical and scientific input; S.B and G.F. jointly conceived of the project, supervised the work and wrote the paper. The materials and datasets generated during the study are available from G.F. on reasonable request. None of the funding bodies had any involvement in the design and conduct of this research or preparation of the manuscript. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at https://doi.org/10.1038/s41598-018-28094-6.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.