Arylamine N-acetyltransferases: a pharmacogenomic approach to drug metabolism and endogenous function

E. Sim*[1], K. Pinter*, A. Mushtaq*, A. Upton*, J. Sandy*, S. Bhakta* and M. Noble†

*Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, U.K., and †Laboratory of Molecular Biophysics, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, U.K.

Abstract

The arylamine N-acetyltransferases (NATs) are a unique family of enzymes that catalyse the transfer of an acetyl group from acetyl-CoA to the terminal nitrogen of hydrazine and arylamine drugs and carcinogens. The NATs have been shown to be important in drug detoxification and carcinogen activation, with humans possessing two isoenzymes encoded by polymorphic genes. This polymorphism has pharmacogenetic implications, leading to different rates of inactivation of drugs, including the anti-tubercular agent isoniazid and the anti-hypertensive drug hydralazine. Mice provide a good model for human NAT, allowing genetic manipulation of expression to explore possible endogenous roles of these enzymes. The first three-dimensional NAT structure was resolved for NAT from *Salmonella typhimurium*, and subsequently the structure of NAT from *Mycobacterium smegmatis* has been elucidated. These identified a 'Cys-His-Asp' catalytic triad (conserved in all NATs), which is believed to be responsible for the activation of the active site cysteine residue. As more genomic data become available, NAT homologues continue to be found in prokaryotic species, many of which are pathogenic, including *Mycobacterium tuberculosis*. The discovery of NAT in *M. tuberculosis* is particularly significant, since this enzyme participates in inactivation of isoniazid in the bacterium, with implications for isoniazid resistance. Structural studies on NAT proteins and phenotypic analyses of organisms (both mice and prokaryotes) following genetic modifications of the *nat* genes are leading to an understanding of the potentially diverse roles of NAT in endogenous and xenobiotic metabolism. These studies have indicated that NAT, particularly in Mycobacteria, has the potential to be a drug target. Combinatorial chemical approaches, together with *in silico* structural studies, will allow for advances in the identification of NAT substrates and inhibitors, both as experimental tools and as potential drugs.

Introduction

Arylamine N-acetyltransferases (NATs) were initially identified as the enzymes responsible for the inactivation of the anti-tubercular drug isoniazid. This finding was an early example of pharmacogenetic polymorphism, since it was demonstrated that NAT activity was under genetic control. Individuals were either assigned as fast or slow acetylators (see [1] for an early review). NAT enzymes are now known to acetylate arylamine, arylhydroxylamines and arylhydrazines by transfer of an acetyl moiety from acetyl-CoA. In humans, there are two NAT isoenzymes encoded at two polymorphic loci. Animal models for acetylator polymorphism have indicated that NAT in other species exists as multiple isoenzymes, and these isoenzymes are encoded at polymorphic loci. Species investigated include birds [2], hamsters [3], rats [4] and rabbits [5]. Recent research has focused on mice [6].

Key words: catalytic triad, mycobacteria, *nat* (arylamine N-acetyltransferase gene) knockout mice, xenobiotic.

Abbreviations used: NAT, arylamine N-acetyltransferase; *p*-abaglu, *p*-aminobenzoylglutamate; CMV, cytomegalovirus.

[1]To whom correspondence should be addressed (e-mail edith.sim@pharmacology.oxford.ac.uk).

In addition to the identification of NAT in eukaryotes, there has been considerable interest in the identification of homologous NAT sequences in prokaryotes [7,8]. The existence of *nat* genes in prokaryotes has proved to be of particular interest for several reasons: (1) Many of NAT substrates are anti-bacterial compounds. These include the sulphonamides and also the hydrazine, isoniazid, which is a well-known anti-tubercular drug. (2) Prokaryotic NATs have yielded a three-dimensional crystal structure. (3) NAT homologues have been identified for which an enzymic function other than acetylation of xenobiotics has been identified.

One of the recurring themes of research into xenobiotic-metabolizing enzymes, particularly those which are expressed constitutively, is whether these enzymes have an endogenous role. The study of prokaryotic NAT homologues that are involved in the synthesis of antibiotics, e.g. rifamycin [9,10] and rubradirin [11], suggests that NAT is capable of carrying out another function: intramolecular amide formation.

Our team in Oxford in collaboration with many other scientists has set out to address three inter-related questions: (i) what are the structures of the NAT enzymes, and how are

these related to function?; (ii) do the NAT enzymes have endogenous roles?; and (iii) can NAT enzymes carry out different functions in the range of organisms in which they are found?

The overall strategies to answer these questions have used a range of techniques, including X-ray crystallography of recombinantly expressed proteins, coupled with identification of novel ligands to answer question (i). A pharmacogenomic approach, i.e. identification of NAT homologues by homology searching coupled with transgenic technology with mice and bacteria, has been used to answer questions (ii) and (iii).

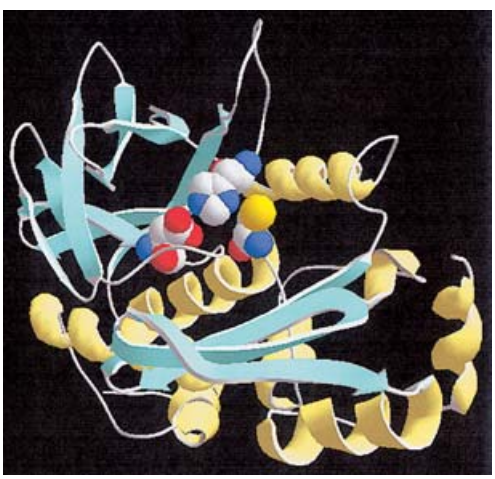
The remainder of this review will address mainly a cross-section of our recent studies on NAT that have opened doors into fundamental topics, ranging from mammalian embryonic development to understanding mycobacterial cell-wall synthesis. Earlier, more comprehensive, reviews are also available [12–16]. The answer does indeed appear to be that, using a pharmacogenomic approach, NAT can be identified in a range of species and can have different roles, depending on the organisms in which it is found. This has been most clearly identified through identification of NAT homologues that can be deduced to catalyse arylamide bond formation, but where no acetyl transfer from acetyl-CoA can be detected [10]. The extent to which diversity of activity is underpinned by structural understanding of both the different isoenzymes that have been identified and their pharmacogenetic variants encoded by multiple alleles is still being investigated, and will also be addressed.

Structural studies

To date, the three-dimensional crystal structure of two NAT isoenzymes has been determined. These enzymes, from *Salmonella typhimurium* [17] and *Mycobacterium smegmatis* [18], show remarkable similarity, and their structural determination identified the key to the reaction mechanism of the NATs. In the active site, there is a catalytic triad of an essential cysteine residue juxtaposed with both a histidine and an aspartate residue (Figure 1). Recently, using site-directed mutagenesis studies of the *S. typhimurium* enzyme, it has been shown that each of these residues is essential for the catalytic mechanism (A. Mushtaq, P. Schartau, M. Payton and E. Sim, unpublished work). The cysteine, histidine and aspartate residues act in concert, and result in an acetyl group being transferred from acetyl-CoA to NAT. Precise information on the lifetime of the acetyl intermediate is lacking at present, and it is possible that the NATs from diverse organisms (and, indeed, different polymorphic variants) might differ. Acetylated intermediates have been detected using [^{14}C]acetyl-CoA and NAT from rabbit [20], and the NAT1 human isoenzyme [21]. However, it has proved extremely difficult to detect an acetylated NAT from the organism *S. typhimurium* using [^{13}C]acetyl-CoA and ^{13}C-NMR studies, despite having a stable concentrated solution of active recombinant protein [22]. It may be that these data suggest a genuine difference in the lifetime of the

Figure 1 | Structure of arylamine N-acetyltransferase has an active-site catalytic triad

The ribbon diagram shows the cysteine, histidine and aspartate residues of the active site in space-filling representation. The molecule consists of three domains, with the C-terminal domain poised over the active site in this orientation. The diagram shows *S. typhimurium* NAT [17], but the NAT from *M. smegmatis* [18] has a similar overall organization at the active site.



acetylated intermediate of these species of NAT, or it may be due to the different sensitivities of the two techniques. It has, however, been demonstrated in both human and *S. typhimurium* NAT that the C-terminus appears to control the rate at which the acetyl-CoA is hydrolysed [21,23]. In neither of these species of NAT is the acetyl-CoA hydrolysed in the absence of an arylamine, unless the C-terminus is missing (Figure 2). In the case of *S. typhimurium* NAT, only the terminal 11 residues are sufficient to suppress the NAT-catalysed hydrolysis of acetyl-CoA in the absence of arylamine substrate. These studies suggest that the C-terminus is directly responsible for arylamine substrate specificity in both human and *S. typhimurium* NAT, as had been predicted from a comparison of sequence and specificity of different members of the NAT family [24,25].

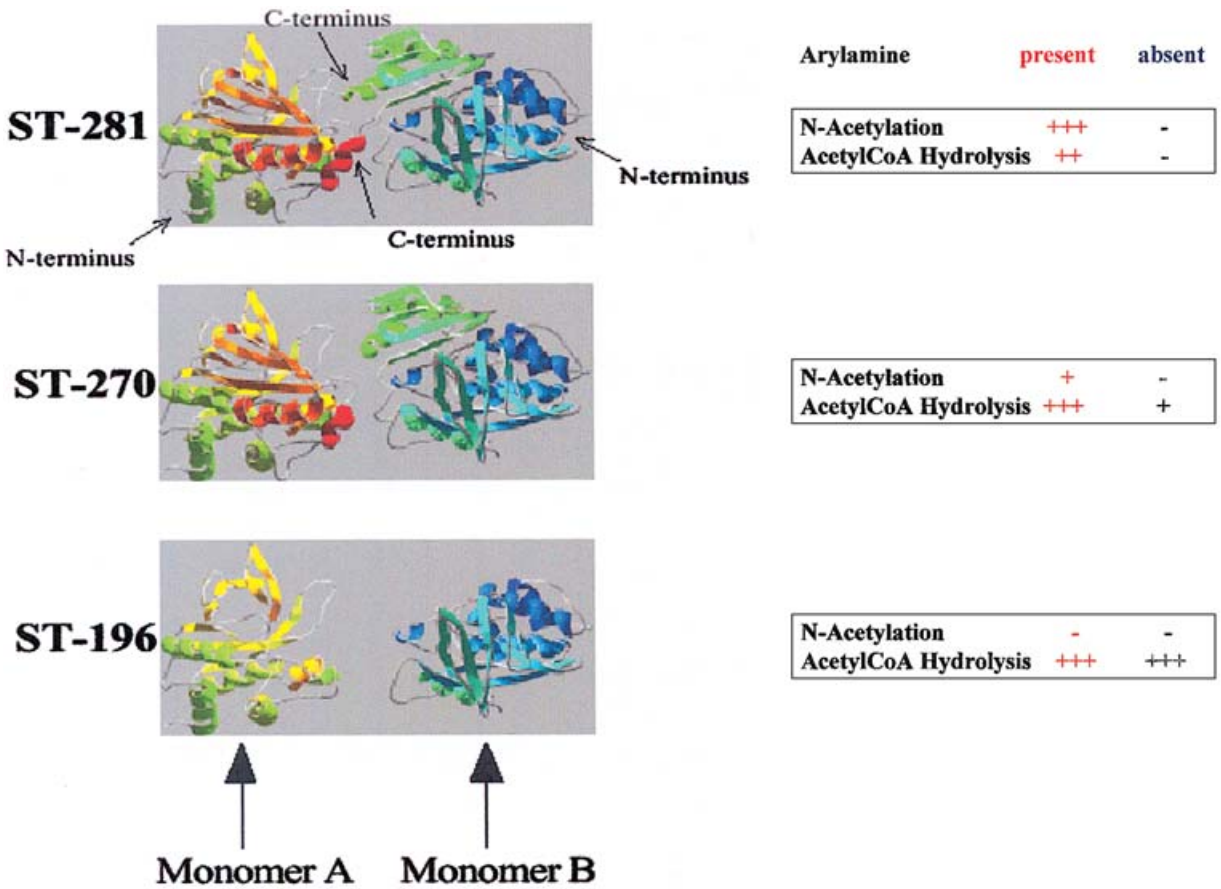
In sequence comparisons of all of the known NAT sequences, therefore, it is important to focus on the C-terminal region, which is likely to be important in determining whether the enzymes might be capable of carrying out different functions.

Genetic modifications of mycobacterial NAT

In order to determine the possible endogenous functions of NATs [26], we have adopted two strategies. Initially, we have generated organisms in which the *nat* gene has been knocked out, or where the expression of *nat* has been increased. We have used this genetic strategy initially with a eukaryote (the mouse) and prokaryotes (*M. smegmatis* and *Mycobacterium bovis* BCG) as models for *Mycobacterium tuberculosis*, in

Figure 2 | Role of C-terminus of NAT from *S. typhimurium* in controlling activity

C-terminal deletion mutants were generated and assayed for different enzymic activities: acetylation of the arylamine anisidine using acetyl-CoA as acetyl donor; hydrolysis of acetyl-CoA when the arylamine anisidine was present; or hydrolysis of acetyl-CoA when the arylamine was not present. Full-length NAT only hydrolyses acetyl-CoA, concomitantly with acetylation of anisidine. Removing the C-terminus results in an enzyme that does not acetylate anisidine, and that hydrolyses acetyl-CoA in the absence of an arylamine substrate. See [23] for further details.



which the *nat* gene sequence is identical with that of *M. bovis* BCG [27,28]. In the case of mycobacteria, we have been particularly interested in NAT enzymic activity, since NAT can inactivate the anti-tubercular drug isoniazid, and appears to do so within mycobacterial cells [29,30]. This may have important consequences for treatment of tuberculosis. Pharmacogenetic variation within NAT in different clinical isolates of *M. tuberculosis* [30], which leads to a lowering of NAT activity, still needs to be investigated fully in relation to isoniazid resistance of these strains. The relationship between a mutation in *nat* in *M. tuberculosis* and isoniazid resistance is not simple, but it appears that NAT is a contributory factor controlling isoniazid sensitivity, together with other genetic factors, such as mutation in the isoniazid-activating enzyme katG [31].

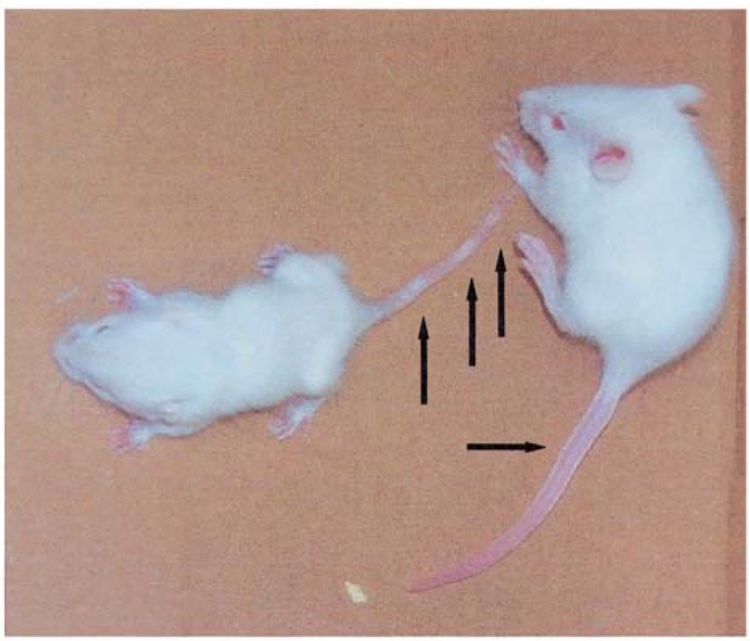
In parallel with the naturally occurring pharmacogenetic variation in mycobacteria, a series of genetic manipulation experiments in *M. smegmatis* has demonstrated that knocking out the *nat* gene has two effects: first, during growth in the presence of isoniazid, altering the level of *nat* expression modifies the growth of the organism in the presence of the anti-mycobacterial agent. Isoniazid is more effective, as would be expected since the enzyme that inactivates it is missing [32]. The corollary is true when the expression of *nat* is enhanced: isoniazid becomes less effective [29]. Secondly, the growth of *M. smegmatis* is altered even in the absence of isoniazid, and this suggests that NAT also has an endogenous role. The growth of *M. smegmatis* is altered such that the lag phase is prolonged, even in the absence of isoniazid [32]. We have observed this effect in a range of mycobacteria in which we have knocked out the *nat* gene, and we have also observed that the lag phase is shorter when the expression of the *nat* gene has been increased [33].

Genetic modifications of murine NAT

In mice, as well as in humans, *nat* genes exist as isoenzymes that show polymorphism, and mice are considered to represent a good model for human NATs. Both mice and humans have three loci encoding NAT homologues located

Figure 3 | Chimaeric mice overexpressing human NAT1 show a 'kinky tail' phenotype

The human *NAT1* gene under control of the CMV promoter was inserted into the murine *nat2* gene of the Ola129 strain of mice by homologous recombination. Resulting transgenic Ola129 embryonic stem cells were implanted into blastocysts of the PO strain of mice. Chimaeric embryos (identified by PCR analysis) recovered by Caesarean section were grossly abnormal. Of the few chimaeric pups born naturally, detected by brown patches (from the Ola129 contribution) on the white coat (of the PO strain) and PCR analysis, only a very limited chimaerism was detected (small brown spot on nose of mouse on left) and no germ-line transmission was obtained. Two of the chimaeric pups are shown, and the arrows indicate the kinks in the tail. Mice in which the *nat2* gene has been interrupted by a neomycin cassette containing the *lacZ* gene are apparently phenotypically normal [49].



within less than 200 kb at isologous regions in their respective genomes [34,35]. In humans, two of the loci are functional and one is a pseudogene [36]. In mice, all three appear to generate functional transcripts, although no apparent specific substrate for the product of the third locus (*nat3*) has been identified. NAT3 will poorly acetylate substrates specific for gene products of each of the other two loci (NAT1 and NAT2) [37–39]. The distribution of NAT isoenzymes in humans and mice demonstrates that there is one isoenzyme that acetylates isoniazid (and sulphamethazine in humans). This enzyme (NAT2 in humans; perversely named NAT1 in mice) is found predominantly in the liver and gut, although traces have been detected in other tissues, particularly as cDNA (see [15] for a review). Another isoenzyme which specifically acetylates *p*-aminobenzoic acid and *p*-aminobenzoylglutamate (*p*-abaglu) has also been identified in both mice (murine NAT2; [40]) and humans (human NAT1; [41]). It is this latter finding that has been particularly intriguing, since *p*-abaglu is a folate catabolite that is acetylated to its urinary metabolite in both mice and humans [12]. It is possible that a NAT isoenzyme carries out this function. Since all human and murine [6] NAT isoenzymes are subject to polymorphisms, this feature was intriguing in relation to possible consequences for folate metabolism. Another feature of the *p*-abaglu-acetylating NAT isoenzymes in humans and mice is that the enzyme is very widespread amongst different tissues, including neuronal tissue, and is expressed very early in development prior to implantation and to neurulation [40,42]. Therefore, since folate has a protective effect in neural tube development, it seemed sensible to alter the expression of the *nat* gene responsible for *p*-abaglu acetylation in mice. Whereas the *nat* knock-out shows very little phenotype [43], when we overexpressed the human *NAT1* gene from humans in mice we found that the number of transgenic animals obtained was extremely small, with many resorbed and grossly abnormal embryos following both random and targeted integration of the *NAT* sequence under control of a strong promoter [(cytomegalovirus (CMV)] into the murine *nat2* gene. The few chimaeras that we did obtain from such targeted integration had a defect, resulting in a kinky tail (Figure 3). Abnormalities of this type have been described as models for neural tube defects [44]. Unfortunately, we were unable to breed from any of the chimaeras, since no germ-line transmission of the overexpressed *NAT* transgene was obtained. The overall view of this time-consuming study was that overexpression of a *nat* gene to produce too much of the NAT isoenzyme that metabolizes *p*-abaglu in mice is detrimental in early development, and

results either in resorption of embryos or in defective pups.

A similar overexpression study has been performed by Erickson and co-workers, and they have been able, using random integration, to generate mice from which they can breed (cf. [45] and [43]). The level of overexpression of NAT activity proved to be less than they would have expected [45], and it may be that the difference between the studies is that the Erickson and co-workers used a different strain of mice.

Conclusions

The structural and transgenic studies which we have carried out have begun to answer the questions we wished to address. (1) Having prokaryotic NAT structures has allowed modelling of other NAT structures, and this has resulted in proposals for how a common mutation in human NAT2 may result in the slow acetylator phenotype [46]. (2) The transgenic studies with NAT enzymes from both mycobacteria and mice indicate that NAT does have an endogenous role(s), as yet unidentified. (3) The transgenic studies on mycobacteria and mice do indicate that NAT enzymes carry out different functions in the range of organisms in which they are found, and this is also true of the transgenic studies with *Amycolatopsis mediteraneii* [47], in which the NAT homologue, rifamycin amide synthetase, is found. The extent to which the NATs that have been described in a wide range of other bacteria [48] have unique endogenous functions remains to be established.

The authors thank all of their colleagues in the NAT field, particularly those who have worked in the Oxford lab for ideas and discussion. We are grateful to Dr F. Brook and Professor R. Gardner (Department of Zoology, University of Oxford, Oxford, U.K.) for assistance with transgenic studies. The Wellcome Trust has provided generous financial support and we are also grateful to Action Research, and to the Biotechnology and Biological Sciences Research Council and GlaxoSmithKline (A.U.), and the Medical Research Council (A.M.) for studentships. We also thank Rebecca Weaver for her help in preparing the manuscript.

References

1 Weber, W.W. and Hein, D.W. (1985) Pharmacol. Rev. **37**, 25–79
2 Ohsako, S., Ohtomi, M., Sakamoto, Y. and Deguchi, T. (1988) J. Biol. Chem. **263**, 7534–7538
3 Sticha, K.R.K., Sieg, C.A., Bergstrom, C.P., Hanna, P.E. and Wagner, C.R. (1997) Protein Expression Purif. **10**, 141–153
4 Doll, M.A. and Hein, D.W. (1995) Pharmacogenetics **5**, 247–251
5 Andres, H.H., Vogel, R.S., Tarr, G.E. and Weber, W.E. (1987) Mol. Pharmacol. **31**, 446–456
6 Boukouvala, S., Price, N. and Sim, E. (2002) Pharmacogenetics **12**, 385–394
7 Payton, M., Mushtaq, A., Yu, T., Wu, L., Sinclair, J. and Sim, E. (2001) Microbiology **147**, 1137–1147
8 Rodrigues-Lima, F. and Dupret, J.M. (2002) Biochem. Biophys. Res. Commun. **293**, 783–792
9 August, P.R., Tang, L., Yoon, Y.J., Ning, S., Muller, R., Yu, T.W., Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X. et al. (1998) Chem. Biol. **5**, 69–79
10 Pompeo, F., Mushtaq, A. and Sim, E. (2002) Protein Expression Purif. **24**, 138–151
11 Sohng, J.K., Oh, T.J., Lee, J.J. and Kim, C.G. (1997) Mol. Cell **7**, 674–681
12 Sim, E., Payton, M., Noble, M.E.M. and Minchin, R. (2000) Hum. Mol. Genet. **9**, 2435–2441
13 Upton, A., Johnson, N., Sandy, J. and Sim, E. (2001) Trends Pharmacol. Sci. **22**, 140–146
14 Pompeo, F., Brooke, E., Kawamura, A., Mushtaq, A. and Sim, E. (2002) Pharmacogenetics **3**, 19–30
15 Butcher, N.J., Boukouvala, S., Sim, E. and Minchin, R.F. (2002) Pharmacogenomics J. **2**, 30–42
16 Hein, D.W. (2000) Toxicol. Lett. **112–113**, 349–356
17 Sinclair, J.C., Sandy, J., Delgoda, R., Sim, E. and Noble, M.E.M. (2000) Nat. Struct. Biol. **7**, 560–564
18 Sandy, J., Mushtaq, A., Kawamura, A., Sinclair, J., Sim, E. and Noble, M. (2002) J. Mol. Biol. **318**, 1071–1083
19 Reference deleted
20 Andres, H.H., Klem, A.J., Schopfer, L.M., Harrison, J.K. and Weber, W.W. (1988) J. Biol. Chem. **16**, 7521–7527
21 Sinclair, J. and Sim, E. (1996) Biochem. Pharmacol. **53**, 11–16
22 Delgoda, R., Lian, L-Y., Sandy, J. and Sim, E. (2003) Biochim. Biophys. Acta **1620**, 8–14
23 Mushtaq, A., Payton, M. and Sim, E. (2002) J. Biol. Chem. **277**, 12175–12181
24 Goodfellow, G.H., Dupret, J.M. and Grant, D.M. (2000) Biochem. J. **348**,159–166
25 Sasaki, Y., Ohsako, S. and Deguchi, T. (1991) J. Biol. Chem. **266**, 13243–13250
26 Butcher, N.J., Ilett, K.F. and Minchin, R.F. (2000) Mol. Pharmacol. **57**, 468–473
27 Cole, S.T., Brisch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, III, C.E. et al. (1998) Nature (London) **393**, 537–544
28 www.pasteur.fr/recherche/unites/Lgmb/mycogenomics.html
29 Payton, M., Auty, R., Delgoda, R., Everett, M. and Sim, E. (1999) J. Bacteriol. **181**, 1343–1347
30 Upton, A., Everett, M., Mushtaq, A., van Helden, P., Victor, T., Wagner, R. and Sim, E. (2001) Mol. Microbiol. **42**, 309–319
31 Slayden, R.A. and Barry, III, C.E. (2000) Microbes Infect. **2**, 659–669
32 Payton, M., Gifford, C., Schartau, P., Hagemeier, C., Mushtaq, A., Lucas, S., Pinter, K. and Sim, E. (2001) Microbiology **147**, 3295–3302
33 Bhakta, S., Upton, A., Payton, M., Pompeo, F., DaSilva, R., Besra, G.D., Knutton, S., Busby, S., Parish, T. and Sim, E. (2003) Keystone Meeting on Tuberculosis, New Mexico, January
34 Fakis, G., Boukouvala, S., Buckle, V., Payton, M., Denning, C. and Sim, E. (2000) Cytogenet. Cell Genet. **90**, 134–138
35 Matas, N., Thygesen, P., Stacey, M., Risch, A. and Sim, E. (1997) Cytogenet. Cell Genet. **77**, 290–295
36 Blum, M., Grant, D.M., McBride, W. and Meyer, U.A. (1990) DNA Cell Biol. **9**, 193–203
37 Kelly, S.L. and Sim, E. (1994) Biochem. J. **302**, 347–353
38 Fretland, A.J., Doll, M.A., Gray, K., Feng, Y. and Hein, D.W. (1997) Toxicol. Appl. Pharmacol. **142**, 360–366
39 Estrada-Rodgers, L., Levy, G.N. and Weber, W.W. (1998) Drug Metab. Dispos. **26**, 502–505
40 Upton, A., Smelt, V., Mushtaq, A., Aplin, R., Johnson, N., Mardon, H. and Sim, E. (2000) Biochim. Biophys. Acta **1524**, 143–148
41 Minchin, R.F. (1995) Biochem. J. **307**, 1–3
42 Smelt, V.A., Upton, A., Adjaye, J., Payton, M.A., Boukouvala, S., Johnson, N., Mardon, H.J. and Sim, E. (2000) Hum. Mol. Genet. **9**, 1101–1107
43 Rodrigues-Lima, F., Blomeke, B., Sim, E. and Dupret, J.M. (2002) Pharmacogenomics J. **2**, 152–155
44 Murdoch, J.N., Henderson, D.J., Doudney, K., Gaston-Massuet, C., Phillips, H.M., Paternotte, C., Arkell, R., Stanier, P. and Copp, A.J. (2003) Hum. Mol. Genet. **12**, 87–98
45 Erickson, R., Morgan, C. and McQueen, C.A. (2000) 2nd International NAT workshop, Eynsham, Oxford, abstract no. 1
46 Rodrigues-Lima, F. and Dupret, J.M. (2002) Biochem. Biophys. Res. Commun. **291**, 116–123
47 Floss, H.G. and Yu, T.W. (1999) Curr. Opin. Chem. Biol. **3**, 25–79
48 Delomenie, C., Fouix, S., Longuemaux, S., Brahimi, N., Bizet, C., Picard, B., Denamur, E. and Dupret, J.M. (2001) J. Bacteriol. **183**, 3417–3427
49 Cornish, V., Pinter, K., Boukouvala, S., Johnson, N., Labrousse, C., Payton, M., Priddle, H. and Sim, E. (2003) Pharmacogenomics J., in the press

Received 3 February 2003