Identification of amino acids imparting acceptor substrate selectivity to human arylamine acetyltransferases NAT1 and NAT2

Geoffrey H. GOODFELLOW*, Jean-Marie DUPRET† and Denis M. GRANT*¹

*Genetics and Genomic Biology Program, Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8, and

†Institut National de la Santé et de la Recherche Médicale U458, Hôpital Robert Debré, 48 boulevard Serurier, 75019 Paris, France

The human arylamine *N*-acetyltransferases NAT1 and NAT2 catalyse the acetyl-CoA-dependent N- and O-acetylation of primary arylamine and hydrazine xenobiotics and their N-hydroxylated metabolites. We previously used a panel of recombinant NAT1/NAT2 chimaeric proteins to identify linear amino acid segments that have roles in imparting the distinct catalytic specificities to these proteins [Dupret, Goodfellow, Janezic and Grant (1994) *J. Biol. Chem.* **269**, 26830–26835]. These studies indicated that a conserved central region (residues 112–210) distinct from that containing the active-site cysteine residue Cys⁶⁸ was important in determining NAT substrate selectivity. In the present study we have refined our analysis through further chimaera generation of this conserved region and by subsequent site-directed mutagenesis of individual amino acids. Enzyme-kinetic analysis of these mutant proteins with the NAT1-selective and NAT2-selective substrates *p*-aminosalicylic acid (PAS) and

sulphamethazine (SMZ) respectively suggests that residues 125, 127 and 129 are important determinants of NAT1-type and NAT2-type substrate selectivity. Modification of Arg¹²⁷ had the greatest effect on specificity for PAS, whereas changing Phe¹²⁵ had the greatest effect on specificity for SMZ. Selected NAT mutants exhibited *K_m* values for acetyl-CoA that were comparable with those of the wild-type NATs, implying that the mutations affected acceptor substrate specificity rather than cofactor binding affinity. Taken together with previous observations, these results suggest that residues 125, 127 and 129 might contribute to the formation of the active-site pocket surrounding Cys⁶⁸ and function as important determinants of NAT substrate selectivity.

Key words: enzyme kinetics, site-directed mutagenesis, structure–function.

INTRODUCTION

Primary arylamine and hydrazine xenobiotics and their oxidized metabolites can be acetylated by either of two isoenzymes of cytosolic acetyl-CoA (Ac-CoA): arylamine *N*-acetyltransferase (NAT; EC 2.3.1.5), which have been designated NAT1 and NAT2. Human NAT1 and NAT2 are the products of distinct gene loci separated by roughly 170–360 kb on chromosome 8 [1,2]. The NAT2 gene locus is the site of the human acetylation polymorphism, because variant alleles at this locus [3] are associated with decreased levels of immunoreactive NAT2 protein in human liver cytosol [4] and a decrease in the elimination *in vivo* of drugs such as isoniazid, sulphamethazine (SMZ) and procainamide in individuals of the ‘slow acetylator’ phenotype. The expression of NAT1 is independent of the isoniazid acetylation polymorphism, although observed variation in the acetylation of NAT1-selective substrates such as *p*-aminosalicylic acid (PAS) and *p*-aminobenzoic acid [5] has also been shown to have a genetic basis [6]. Both human NAT1 and NAT2 are capable not only of the N-acetylation of arylamines to innocuous acetamides but also the O-acetylation and, to a limited extent, N,O-acetyl transfer of their hydroxylamine and hydroxamic acid metabolites to reactive, protein- and DNA-binding acetoxy esters [7]. Human and animal studies have clearly shown the significant role of the NATs in carcinogen bioactivation and the formation of DNA adducts that together influence an individual’s susceptibility to numerous cancers [8].

The human wild-type NAT1*4 and NAT2*4 genes share 87% nucleotide sequence identity; their 290-residue products are 81% identical [9]. Moreover, of the 55 amino acid differences between wild-type NAT1 and NAT2 (NAT1 4 and NAT2 4 respectively),

only 28 (10%) are non-conservative. Nevertheless these proteins exhibit marked differences in their patterns of arylamine acceptor substrate specificity. We have been using recombinant expression systems and protein modification techniques to investigate the structural features of NAT1 and NAT2 that impart their distinct catalytic properties [10,11]. Using site-directed mutagenesis, we identified Cys⁶⁸ as the cysteine residue likely to be directly involved in the acetyl transfer reaction [10]. Independently, it was demonstrated that the Cys⁶⁹ residue of *N*-hydroxyarylamine O-acetyltransferase of *Salmonella typhimurium* was responsible for mediating Ac-CoA binding [12]. Cys⁶⁹, the only conserved cysteine residue between O-acetyltransferase of *S. typhimurium* and NATs of higher organisms, resides within a highly conserved region. In contrast, site-directed substitution of Arg⁹ and Arg⁶⁴ had a greater effect on the function of NAT1 than on that of NAT2 [13], suggesting that the catalytic mechanisms of these two enzymes might have distinctive properties.

To localize linear regions of amino acid sequence associated with substrate specificity and protein stability, we constructed a panel of NAT1/NAT2 chimaeric genes and determined the enzyme-kinetic characteristics of their expressed products [11]. We observed that regions distinct from those adjacent to catalytic residue Cys⁶⁸ made significant contributions to differences in both the *K_m* and *V_{max}* of these proteins. Specifically, a central region (residues 112–210) from NAT1 4 or NAT2 4 seemed to correlate with NAT1-type or NAT2-type behaviour respectively. Examination of this central region reveals a highly conserved 42-residue segment (residues 107–148) in which the sequences of NAT1 4 and NAT2 4 differ at only three positions (residues 125, 127 and 129) (Figure 1). In the present study we used chimaera construction and site-directed mutagenesis to evaluate the con-

Abbreviations used: Ac-CoA, acetyl-CoA; NAT, arylamine *N*-acetyltransferase; NAT1 4 and N1, wild-type NAT1; NAT2 4 and N2, wild-type NAT2; PAS, *p*-aminosalicylic acid; SMZ, sulphamethazine.

¹ To whom correspondence should be addressed (e-mail grant@sickkids.on.ca).

Figure 1 Highly conserved 25-residue segment within the conserved central region of NAT1 and NAT2

The open box represents the protein-coding region of *NAT* genes and the dark grey boxes represent the wild-type NAT1 and NAT2 proteins. The solid vertical lines show the three amino acid differences between NAT1 and NAT2 in this highly conserved region flanked by the restriction sites *Pst*I and *Bbv*I at the gene level; within this conserved region the light grey boxes enclose the three amino acids that differ at positions 125, 127 and 129 between NAT1 (Phe¹²⁵, Arg¹²⁷ and Tyr¹²⁹) and NAT2 (Ser¹²⁵, Ser¹²⁷, Ser¹²⁹).

tribution of these three amino acids to the substrate affinity, maximal velocity, molecular activity and stability of human NAT1 and NAT2.

EXPERIMENTAL

Materials

The NAT1-selective arylamine acceptor substrate PAS, the NAT2-selective arylamine acceptor substrate SMZ, the acetyl donor Ac-CoA and the components of the cofactor-regenerating system [acetyl-D,L-carnitine and carnitine acetyltransferase (EC 2.3.1.7)] were obtained from Sigma. AcPAS was produced enzymically in definable quantities from PAS by using wild-type recombinant NAT1 as a catalyst [11]; AcSMZ was synthesized and purified as described previously [14]. The expression vector pKEN2 and the *Escherichia coli* host strain XA90 (F'*lacI*^{q1}) were kindly provided by Dr G. L. Verdine (Department of Chemistry, Harvard University, Cambridge, MA, U.S.A.). Enzymes for cloning and sequencing were obtained from New England Biolabs and Pharmacia. Amplitaq DNA polymerase and other related PCR reagents were from Perkin-Elmer (Applied Biosystems Canada). Oligonucleotides for site-directed mutagenesis were synthesized by General Synthesis and Diagnostics (Toronto, Ontario, Canada). All other chemicals and reagents were of analytical grade from standard suppliers.

Construction of NAT triple mutants

The region of the *NAT* genes encoding residues 125, 127 and 129 is flanked by the *Pst*I and *Bbv*I restriction endonuclease sites at nt 332 and 408 respectively in both genes (Figure 1). Mutants containing three amino acid substitutions were therefore constructed by creating novel chimaeric genes from existing chimaeras that had been characterized previously [11]. These chimaeras, 1-2 and 2-2, had been generated by an exchange of corresponding fragments at the *Pst*I restriction site at position 332 in the wild-type NAT1 and NAT2 genes. Phagemid pKEN2-based expression vectors containing the coding regions for 1-2 and 2-2 were digested with *Eco*RI and *Sal*I. The *Eco*RI/*Sal*I expression cassettes were gel-purified and then digested with *Bbv*I to yield 408 bp *Eco*RI/*Bbv*I and 462 bp *Bbv*I/*Sal*I DNA fragments.

Appropriate gel-purified 408 bp *Eco*RI/*Bbv*I DNA fragments from 1-2 and 2-2 were exchanged, fragments were re-ligated and subcloned into pKEN2, then transformed into the *E. coli* strain DH5 α . Diagnostic digestion of the resultant plasmids with the restriction enzymes *Alw*NI (present in the 332–408 region of only the NAT1 gene), *Hinc*II (present in the 1–332 region of only the NAT2 gene) and *Hpa*I (present in the 1–332 region of only the NAT2 gene) verified the identities of the constructs.

Construction of NAT single and double mutants

PCR-mediated site-directed mutagenesis was used to construct mutant NAT proteins containing one or two amino acid substitutions. Expression vectors pNAT1 and pNAT2, containing the human wild-type NAT1*4 and NAT2*4 coding regions respectively [10] were used as PCR templates to generate mutagenic insert DNA fragments [15,16] corresponding to nt 1–408 of the 870 bp *NAT* coding region. Amplifications used 5' primers specific for either NAT1 (N1PC5) or NAT2 (N2PC5), and 3' mutagenic primers containing a *Bbv*I restriction site (Table 1). Amplifications were performed in a Perkin-Elmer DNA Thermal Cycler 480 under the following conditions: denaturation, 30 s at 95 °C; annealing, 30 s at 45 °C for the NAT1 mutants or 58 °C for the NAT2 mutants; extension, 45 s at 72 °C, for a total of 30 cycles, followed by a final extension at 72 °C for 7 min. Amplified products were gel-purified, digested with *Eco*RI and *Bbv*I, then precipitated with ethanol to remove end fragments and buffer salts. *Bbv*I/*Sal*I fragments (nt 409–870) were isolated from wild-type pNAT1 and pNAT2, ligated to appropriate PCR-generated *Eco*RI/*Bbv*I mutagenic fragments (1–408 bp) and subcloned into *Eco*RI/*Sal*I-digested pKEN2. Modified vectors were transformed into *E. coli* strain DH5 α and protein-coding regions were completely sequenced to verify the presence of only the desired mutation. Constructs with the desired mutation(s) were then transformed into the *E. coli* expression strain XA90 for kinetic analysis of the recombinant proteins.

Nomenclature of NAT1 and NAT2 mutant proteins

Amino acid substitutions were introduced at positions 125, 127 and 129 of NAT1 4 and NAT2 4. At these three positions, the

Table 1 Oligonucleotide primers used for mutagenesis and sequencing

Bold type identifies the location of start and stop codons; italics represent the nucleotides recognized by the restriction enzymes noted. The amino acids encoded by the mutagenic primers at positions 125, 127 and 129 are underlined and labelled; the location of each oligonucleotide in the NAT1 or NAT2 coding sequence is indicated relative to the ATG start codon (+1). The 5' primers used for expression cassette PCR (ECPCR) of the original NAT1 and NAT2 expression cassettes are identical with those used for PCR-mediated site-directed mutagenesis. The ECPCR primers N1PC5 and N1PC3 denote NAT1-specific primers and were used to amplify the NAT1 expression cassettes, whereas N2PC5 and N2PC3 denote NAT2-specific primers and were used to amplify the NAT2 expression cassettes [10].

<u>ECPCR primers</u>		
<u>5' primers</u>		
N1PC5	5'-CGCGCGAATTCAGGAGGAATTCAT ATG GACATTGAAGCATATTCTGAAAGAA-3'	+1 (start) +28
N2PC5	5'-CGCGCGAATTCAGGAGGAATTCAT ATG GACATTGAAGCATATTTTGAAGAA-3'	+1 (start) +28
<u>3' primers</u>		
N1PC3	5'-GCGCATATGTCGA CTA AATAGTAAAAATCTATCACCATG-3'	+873 (stop) +847
N2PC3	5'-GCGCATATGTCGA CTA AATAGTAAGGGATCCATCACCAGG-3'	+873 (stop) +847
<u>3' mutagenic primers</u>		
S125F	5'-CCAGAAATTAATTCTAGAGGCTGCCACATCTGGGAGGAGCTT CAA ACCAG-3'	+419 (BbvI) (Ser ¹²⁹) (Ser ¹²⁷) (Phe ¹²⁵) +368
S127R	5'-CCAGAAATTAATTCTAGAGGCTGCCACATCTGGGAGGATCTT CCAG ACCAG-3'	(BbvI) (Ser ¹²⁹) (Arg ¹²⁷) (Ser ¹²⁵)
S129Y	5'-CCAGAAATTAATTCTAGAGGCTGCCACATCTGGT AGG ATCTT CCAG ACCAG-3'	(BbvI) (Tyr ¹²⁹) (Ser ¹²⁷) (Ser ¹²⁵)
S125	5'-CCAGAAATTAATTCTAGAGGCTGCCACATCTGGT AGG ATCTT CCAG ACCAG-3'	(BbvI) (Tyr ¹²⁹) (Arg ¹²⁷) (Ser ¹²⁵)
S127	5'-CCAGAAATTAATTCTAGAGGCTGCCACATCTGGT AGG ATCTT CAA ACCAG-3'	(BbvI) (Tyr ¹²⁹) (Ser ¹²⁷) (Phe ¹²⁵)
S129	5'-CCAGAAATTAATTCTAGAGGCTGCCACATCTGGGAGGATCTT CAA ACCAG-3'	(BbvI) (Ser ¹²⁹) (Arg ¹²⁷) (Phe ¹²⁵)

*NAT1*4* gene encodes Phe¹²⁵, Arg¹²⁷ and Tyr¹²⁹, whereas the *NAT2*4* gene encodes Ser¹²⁵, Ser¹²⁷ and Ser¹²⁹. NAT1 4 and NAT2 4 are hereafter abbreviated to N1 and N2 respectively. Because different amino acid residues occur at each of these three positions in NAT1, the one-letter amino acid code letters F, R and Y are used to refer to positions 125, 127 and 129 respectively, removing the need to include the amino acid position. For instance, the modified recombinant protein N1-FS/YS is a NAT1 mutant with residue Phe¹²⁵ replaced with Ser and residue Tyr¹²⁹ replaced with Ser.

Expression, enzyme assay, purification and immunodetection of mutant NAT proteins

Conditions for bacterial growth, lysate preparation and NAT enzyme assays were as described previously [10]. Pilot studies were performed to determine both the appropriate dilution factor for each bacterial lysate (to ensure no more than 10–15 % conversion of substrate into product) and the optimum range of arylamine acceptor substrate concentrations. The acetylated metabolites AcSMZ and AcPAS were quantified by HPLC [17] for 10–16 different substrate concentrations in the ranges 1–2000 μ M for SMZ (limited by substrate solubility) and 2.5–

40000 μ M for PAS. Kinetic constants (apparent K_m and V_{max}) were estimated by non-linear regression of the standard Michaelis–Menten kinetic equation with the program ULTRA FIT (Biosoft, Ferguson, MO, U.S.A.), with the Marquardt algorithm for least-squares error minimization during curve fitting. All kinetic parameters were determined in at least two independent experiments.

To determine whether the alterations in enzyme kinetics were associated with altered binding of cofactor, true K_m and V_{max} values for Ac-CoA were obtained by varying the concentrations of both the acceptor amine and the cofactor.

Recombinant human NAT1 and NAT2 were purified as glutathione S-transferase fusion proteins by standard protocols with glutathione 4B–Sepharose and thrombin protease from Pharmacia. Purified NAT1 and NAT2 recombinant proteins were subjected to PAGE [12 % (w/v) gel], transferred to nitrocellulose and probed with anti-NAT antisera to create standard curves for the quantification of immunoreactive NAT protein in bacterial lysate samples. Aliquots of the purified NAT were included on each blot (internal standard) to quantify the molecules of NAT present in the bacterial lysate sample.

SDS/PAGE and immunoblotting of bacterial lysates containing wild-type and mutant NAT proteins were performed as

described previously [10], as were experiments to estimate protein stability. Selected NAT2 mutants (N2-SF, N2-SR, N2-SY, N2-SF/SR and N2-SF/SY) had 20 μ g of bacterial lysate protein loaded instead of 10 μ g. Protein contents of lysates were determined with a dye-binding method [18].

RESULTS

Expression and immunodetection of mutant NAT proteins

All the NAT1 mutants except N1-YS and N1-FS/YS had levels of soluble immunoreactive protein that were comparable with that of wild-type NAT1. N1-YS and N1-FS/YS displayed 3-fold and 1.5-fold decreased amounts of immunoreactive protein respectively, relative to wild-type NAT1 (Figures 2a and 2b). Expression levels of all NAT1 proteins were similar in the soluble (supernatant) and insoluble (pellet) fractions (Figures 2b–2d). Thus the lower levels of soluble N1-YS and N1-FS/YS were a result of decreased overall efficiency of expression rather than an alteration in the proportion compartmentalized into insoluble

Figure 2 Immunodetection of recombinant wild-type and mutant NAT1 proteins

Bacterial lysates were subjected to PAGE [12% (w/v) gel], transferred to nitrocellulose membrane and hybridized with a rabbit polyclonal antiserum raised against a NAT1 fusion protein. Internal standard (purified NAT1) (100 ng) was run on lane 1 of gels (a)–(d) and 10 μ g of bacterial lysate proteins was loaded in all other lanes. (a) Lanes 2–8, supernatant (s) fractions of N1-FS, N1-RS, N1-YS, N1-FS/RS, N1-FS/YS, N1-RS/YS and N1-FS/RS/YS respectively. (b) Lanes 2–6, wild-type NAT1 (s), wild-type NAT1 (p), wild-type NAT1 pellet (p), N1-FS (s) and N1-FS (p). (c) Supernatant (s) (lanes 2, 4 and 6) and pellet (p) (lanes 3, 5 and 7) fractions of N1-RS, N1-YS and N1-FS/RS. (d) Supernatant (s) (lanes 2, 4 and 6) and pellet (p) (lanes 3, 5 and 7) fractions of N1-FS/YS, N1-RS/YS and N1-FS/RS/YS. Lanes C, 10 μ g of the vector alone (pKEN2/XA90). The positions of molecular mass markers are indicated (in kDa) at the left.

Figure 3 Immunodetection of recombinant wild-type and mutant NAT2 proteins

Bacterial lysates were subjected to PAGE [12% (w/v) gel], transferred to nitrocellulose membrane and hybridized with a rabbit polyclonal antiserum raised against recombinant human NAT2. (a) Lanes 2–9, supernatant (s) fractions of N2-SF, N2-SR, N2-SY, N2-SF/SR, N2-SF/SY, N2-SR/SY, N2-SF/SR/SY and wild-type NAT2 respectively. (b) Supernatant (s) (lanes 1, 3, 5, 7 and 9) and pellet (p) (lanes 2, 4, 6, 8 and 10) fractions of N2-SF, N2-SR, N2-SY, N2-SF/SR and N2-SF/SY. (c) Supernatant (s) (lanes 2, 4 and 6) and pellet (p) (lanes 3, 5 and 7) fractions of N2-SR/SY, N2-SF/SR/SY and wild-type NAT2. Internal standard (purified NAT1) (500 ng) was run in lane 1 of (a) and (c) and in lane 11 of (b). Bacterial lysate protein (20 μ g) was loaded in lanes 2–6 of (a) and lanes 1, 3, 5, 7 and 9 of (b); 4 μ g was loaded in lanes 2, 4, 6, 8 and 10 of (b); 2 μ g was loaded in lanes 3, 5 and 7 of (c); in all other lanes 10 μ g of bacterial lysate protein was loaded. Lanes C, vector alone (pKEN2/XA90) (10 μ g). The arrows indicate the location of immunoreactive NAT2 protein. The position of a 29 kDa molecular mass marker is indicated at the left of each panel.

inclusion bodies (Figures 2c and 2d). N2-SY, N2-SF/SR/SY and wild-type NAT2 had 2-fold increased levels of soluble immunoreactive protein (Figure 3a). N2-SR and N2-SY were expressed equally in soluble and insoluble fractions (Figure 3), whereas the remaining NAT2 mutants and wild-type NAT2 had more than 60% of the NAT protein expression in the insoluble phase (Figures 3b and 3c).

Kinetic analyses of mutant NAT1 and NAT2 proteins

Kinetic constants varied widely between the modified proteins. Apparent K_m (an estimate of the affinity of an enzyme for its substrate) values for PAS (K_m^{PAS}) and SMZ (K_m^{SMZ}) varied 415-fold and 370-fold respectively; maximal velocities (V_{max}) varied 14000-fold and 155-fold. The molecular catalytic activity (k_{cat}) of the modified NAT1 and NAT2 proteins varied 1685-fold and 631-fold respectively. Specific observations are summarized below.

Figure 4 Substrate affinities of wild-type and mutant NAT1 (top panels) and NAT2 (bottom panels) proteins for PAS (left panels) and SMZ (right panels)

Results are representative of at least two independent experiments.

Alterations at position 127 had the greatest effect on enzyme affinity for PAS, especially for the NAT1 mutant proteins. A single Arg → Ser substitution at this position in NAT1 produced a 42-fold increase in K_m^{PAS} (Figure 4). Each of the NAT1 double mutants containing Arg → Ser substitutions (N1-FS/RS and N1-RS/YS) also displayed markedly increased K_m^{PAS} values (146-fold and 41-fold respectively) (Figure 4). We calculated the ratio of $k_{\text{cat}}^{\text{SMZ}}/K_m^{\text{SMZ}}$ to $k_{\text{cat}}^{\text{PAS}}/K_m^{\text{PAS}}$ for each protein as a measure of NAT1-type (low value) or NAT2-type (high value) kinetic behaviour. For N1-RS this ratio was close to unity, which was more than 28000-fold greater than that for wild-type NAT1 and indicated an almost equivalent kinetic selectivity for PAS and SMZ (Figure 5). N1-RS lacked the catalytic specificity of wild-type NAT1 for PAS, as exhibited by a 7300-fold decrease in $k_{\text{cat}}^{\text{PAS}}/K_m^{\text{PAS}}$. Amino acid 127 had a more modest but still significant role in determining the affinity for PAS in the NAT2 mutants. Replacement of Ser¹²⁷ with Arg in NAT2 yielded a protein, N2-SR, with a 2.3-fold lower K_m^{PAS} than wild-type NAT2 (Figure 4).

Modifications at position 125 had the largest effect on K_m^{SMZ} . Replacement of Ser¹²⁵ with Phe in NAT2 produced the greatest increase in K_m^{SMZ} (2-fold) of any of the singly modified NAT2 proteins (Figure 4). Furthermore, N2-SR/SY had a K_m^{SMZ} comparable with that of wild-type NAT2; on the addition of a Ser¹²⁵ → Phe substitution to generate the NAT2 triple mutant, a 65-fold increase in K_m^{SMZ} resulted (Figure 4). Indeed, the K_m^{SMZ} displayed by the NAT2 triple mutant was even higher than that of wild-type NAT1. Concomitant with the large increase in K_m^{SMZ} , there also was a 800-fold increase in NAT1-type behaviour ($k_{\text{cat}}^{\text{PAS}}/K_m^{\text{PAS}}$); these factors together were reflected in the specificity-constant ratio of 0.043 for the NAT2 triple mutant (almost 1500-fold smaller than for wild-type NAT2). The greatest single

change in K_m that we observed involved the effect of amino acid substitutions at position 125 in NAT1 with the substrate SMZ. Replacement of Phe¹²⁵ with Ser¹²⁵ in NAT1 (producing N1-FS) caused a 220-fold decrease in K_m^{SMZ} (Figure 4). N1-FS had an apparent K_m^{SMZ} of 20 μM , more than 5-fold smaller than even that displayed by wild-type NAT2. The NAT1 double mutant N1-FS/RS, with a $k_{\text{cat}}^{\text{SMZ}}/K_m^{\text{SMZ}}$ -to- $k_{\text{cat}}^{\text{PAS}}/K_m^{\text{PAS}}$ ratio of more than 135 (Figure 5), combined the increase in NAT2-type characteristics of N1-FS and the decrease in NAT1-type behaviour of N1-RS.

All the modified NAT1 proteins except N1-YS had a greater than 36-fold decrease in the rate of PAS turnover (Table 2). N1-YS had the highest molecular activity of all the NAT1 mutant proteins for PAS, yet it had the lowest molecular activity for SMZ (Table 2). This suggests that this residue is permissive to substitutions with respect to the acetylation of PAS, but requires Tyr¹²⁹ for optimal SMZ acetylation. The NAT2 mutant proteins, excluding the NAT2 triple mutant, had k_{cat} values comparable with that of wild-type NAT2 (Table 3). N2-SF/SR/SY not only had a molecular activity for PAS acetylation that was more than 100-fold that of wild-type NAT2, this mutant also had the highest rate of SMZ turnover (Table 3). No significant correlation was observed between the molecular activities of the NAT mutant proteins for PAS and SMZ (results not shown), which suggests that these mutations are eliciting different effects on enzyme catalysis that are dependent on the particular acceptor substrate.

Stabilities of mutant proteins

Stabilities of the wild-type and mutant NAT proteins are listed in Table 4. Wild-type NAT1 and NAT2 had $t_{1/2}$ values at 37 °C of

Figure 5 Ratios of the specificity constants k_{cat}^{SMZ}/K_m^{SMZ} to k_{cat}^{PAS}/K_m^{PAS} in wild-type and mutant NAT proteins

Table 2 k_{cat} values for wild-type and mutant NAT1 proteins

Bacterial lysates containing expressed recombinant wild-type and mutant NAT1 proteins were used to quantify enzyme activity by HPLC as described in the Materials and methods section. Recombinant NAT1 was purified as a fusion protein with glutathione S-transferase and then cleaved with thrombin as described in the Materials and methods section. Molecular activity (k_{cat}) is expressed as the number of molecules of product (Ac-PAS or Ac-SMZ) formed per molecule of enzyme (NAT1) in a given unit of time (min). PAS and SMZ were used as the NAT1-selective and NAT2-selective probe drugs respectively.

NAT1 mutant	k_{cat} (mol/min per mol)	
	PAS	SMZ
N1-FS	221	147
	247	133
		127
N1-RS	76.0	153
	75.8	104
		102
N1-YS	1577	5.30
	1575	5.45
		5.84
N1-FS/RS	84.6	125
	88.2	120
N1-FS/YS	53.3	27.4
	47.2	33.1
		30.2
N1-RS/YS	147	73.6
	164	82.1
N1-FS/RS/YS	238	163
	275	148
Wild-type	10954	252
	10758	247
	8007	
	7560	

Table 3 k_{cat} values for wild-type and mutant NAT2 proteins

Bacterial lysates containing expressed recombinant wild-type and mutant NAT2 proteins were used to quantify enzyme activity by HPLC as described in the Materials and methods section. Recombinant NAT2 was purified as a fusion protein with glutathione S-transferase and then cleaved with thrombin as described in the Materials and methods section. Molecular activity (k_{cat}) is expressed as the number of molecules of product (Ac-PAS or Ac-SMZ) formed per molecule of enzyme (NAT2) in a given unit of time (min). PAS and SMZ were used as the NAT1-selective and NAT2-selective probe drugs respectively.

NAT2 mutant	k_{cat} (mol/min per mol)	
	PAS	SMZ
N2-SF	0.37	3.82
	0.34	4.24
	0.40	4.26
N2-SR	1.04	0.84
	1.26	0.89
N2-SY	0.26	0.76
	0.23	0.75
		0.89
N2-SF/SR	0.37	1.19
	0.39	1.24
	0.38	1.34
N2-SF/SY	0.70	5.63
	0.66	5.00
N2-SR/SY	1.64	1.59
	1.26	1.71
	1.44	
N2-SF/SR/SY	144	3.12
	128	8.24
	174	6.79
Wild-type NAT2	160	
	1.41	5.47
	1.46	2.54
		2.93
		2.26

4.7 and 37 h respectively. In general, the NAT2 mutants were more stable than the corresponding NAT1 modified proteins. N1-RS and N1-YS were relatively unstable, with 6-fold and 5-fold decreases respectively in their stabilities *in vitro* compared with N1 (Table 4). In contrast, the NAT1 mutants N1-FS, N1-FS/RS and N1-FS/RS/YS showed 4–5-fold increases in stability (Table 4). These results suggest that the Phe¹²⁵ → Ser substitution imparts stability to the modified NAT1 proteins and that the removal of either Arg¹²⁷ or Tyr¹²⁹ decreases stability. N2-SF, N2-SF/SR and N2-SF/SY had 2–3-fold decreases in stability (Table 4). The 8-fold decrease in stability of the NAT2 triple mutant was the largest for all the NAT2 modified proteins (Table 4). Overall, the removal of Ser¹²⁵ had the most significant effect on enzyme stability of the modified NAT2 proteins (Table 4). We observed no loss of immunoreactive protein accompanying the loss of enzyme function, suggesting that the observed effects were not due to proteolytic degradation.

Binding of cofactor to wild-type and modified NAT proteins

Double-reciprocal plots of 1/velocity against 1/[substrate] produce parallel lines for enzymes possessing a Ping Pong kinetic mechanism. Each of the recombinant NAT proteins except the NAT2 triple mutant displayed a pattern representative of Ping Pong kinetics (results not shown). The true K_m values for wild-type NAT1 (Figure 6a) and the NAT1 triple mutant (Figure 6b)

Table 4 Stabilities of wild-type and mutant NAT1 and NAT2 proteins *in vitro*

The $t_{1/2}$ is calculated as the time at which 50% of the initial activity remained after preincubation of the bacterial lysate at 37 °C; it was used as an estimate of the thermal stability of the wild-type and mutant proteins. Concentrations of PAS and SMZ at one-half maximal velocity were used respectively for the NAT1 and NAT2 mutants.

NAT variant	$t_{1/2}$ (h)
Wild-type NAT1	4.7
N1-FS	24
N1-RS	0.8
N1-YS	0.9
N1-FS/RS	25
N1-FS/YS	1.5
N1-RS/YS	2
N1-FS/RS/YS	18
Wild-type NAT2	37
N2-SF	21
N2-SR	27
N2-SY	27
N2-SF/SR	14
N2-SF/SY	13
N2-SR/SY	36
N2-SF/SR/SY	4.8

were 495 and 690 μM respectively. Although wild-type NAT2 had an affinity for Ac-CoA (675 μM) comparable with that of the NAT1 proteins (Figure 6c), the NAT2 triple mutant exhibited a decrease in enzyme velocity as the Ac-CoA concentration increased above 100 μM , which prevented a calculation of the true K_m for Ac-CoA. However, an analysis of the kinetic data at fixed Ac-CoA concentrations of 50 and 100 μM yielded a true K_m for Ac-CoA of 275 μM (Figure 6c). These modest variations in Michaelis–Menten kinetic constants among the mutant proteins were due largely to alterations in acceptor amine affinity.

DISCUSSION

The major finding of our study was that residues 125, 127 and 129 are important in determining NAT substrate selectivity. Enzyme–substrate interactions are dependent on charge interactions, a number of weak attractions (hydrogen bonds, hydrophobic interactions and van der Waals forces), and a complementary three-dimensional shape to confer specificity on the enzymic reaction. The active site comprises only a small portion of a protein, usually occupying a cleft or crevice in the molecule, and includes residues that can lie far apart in the linear amino acid sequence. The ionic interactions of charged residues are important determinants of biological specificity in enzyme–substrate [19] interactions, receptor–drug [20] complexes and intramolecular [21] and intermolecular [22] protein associations. In the present study, all of the modified NAT1 proteins with an Arg $\rightarrow$ Ser substitution had significant decreases in NAT1-type behaviour (Figure 5). The marked decrease in catalytic specificity for PAS by the NAT1 mutants lacking Arg¹²⁷ could therefore be due to the absence of an interaction between the negatively charged CO_2^- group of PAS and the positively charged Arg¹²⁷. Interestingly, mouse NAT2, which has the highest amino acid sequence identity with human NAT1 (82%) [3] and the same kinetic profile [high affinity for PAS and low affinity for SMZ (D. M. Grant, unpublished work)], also has an Arg residue at position 127. Future kinetic studies with panels of aniline derivatives differing in the nature of their substituents at the *para*

Figure 6 Secondary plots of $1/V_{\max}$ against $1/[\text{Ac-CoA}]$.

Plots of (a) wild-type NAT1 (■), (b) the NAT1 triple mutant (○) and (c) wild-type NAT2 (■) and the NAT2 triple mutant (○). The abscissa intercept and the ordinate intercept represent the $-1/K_m$ and $1/V_{\max}$ values respectively used to generate the true kinetic parameters (K_m and $V_{\max}$) for the cofactor Ac-CoA. True K_m values were calculated to be: wild-type NAT1, 495 μM ; NAT1 triple mutant, 690 μM ; wild-type NAT2, 675 μM ; and NAT2 triple mutant, 275 μM .

position might determine whether a charge interaction between a *p*-carboxylate group on the acceptor substrate contributes to high substrate affinity. Cheon and Hanna [23] showed that the arginine-selective modifying reagent phenylglyoxal inhibited the N-acetylation of the NAT1-selective substrate *p*-aminobenzoic acid (which has a *p*-carboxylate group) by both hamster NAT1 and NAT2. However, the addition of Ac-CoA prevented the inhibition of only NAT1 by phenylglyoxal, and suggested that an Arg residue was located near the active site of NAT1. In view of our results, it is possible that the Arg residue proposed to be near the active site is Arg¹²⁷.

Kyte–Doolittle hydropathy calculations predict that the region between residues 125 and 129 in both NAT1 and NAT2 is hydrophilic. However, surface probability predictions from the linear amino acid sequence suggest that this region is located in the interior of the NAT protein. To satisfy both of these conditions this region would have to remain exposed to the solvent but not be on the surface of the protein. Two of the possible locations for this amino acid region are along a passageway to the active site or within the active site itself. Confirmation of these speculations awaits the identification of a crystal structure of the NAT proteins. One of the contributing

elements to the large decreases in K_m^{SMZ} of N1-FS and N1-FS/RS might simply be steric factors in the vicinity of residue 125. The phenyl ring of Phe takes up considerably more volume than the hydroxymethyl side chain of Ser. Thus the amino acid substitutions in N1-FS and N1-FS/RS presumably increase the volume of unoccupied space in the interior of these proteins. In view of evidence that the substrates selectively metabolized by NAT1 are generally smaller than NAT2-selective substrates, perhaps larger substrates such as SMZ that were previously excluded by the presence of the phenyl ring in Phe¹²⁵ can now gain better access to the substrate recognition atoms in the active site. Differences in protein flexibility complement the above-mentioned steric restrictions. In fact, NAT2-type catalytic specificity might be determined by a combination of these factors. Wild-type NAT2 is predicted to be much more flexible than wild-type NAT1 in the region spanning residues 125, 127 and 129. The introduction of either a single NAT2 residue into NAT1 at position 125 (N1-FS) or a pair of NAT2 residues at positions 125 and 127 (N1-FS/RS) causes significant increases in predicted flexibility. Our results suggest that flexibility might be one of the important factors at positions 125 and 127 in the determination of NAT2-type behaviour.

Another possible explanation for the large increases in NAT2-type catalytic specificity of N1-FS and N1-FS/RS might arise from the hydroxy group of Ser¹²⁵. An unpaired hydrogen-bond donor or acceptor is rare in a protein [24], suggesting that the hydroxy group of Ser might be imparting additional substrate specificity for SMZ. In this regard, it is interesting to note that human NAT2, the only acetyltransferase from higher organisms that has a Ser residue at position 125, has the lowest reported K_m^{SMZ} . Thus the residue at position 125 might be a key determinant of SMZ substrate selectivity in all NATs. Further site-directed mutagenesis studies could determine the contribution of both the size and the hydrogen-bonding ability of the residue at position 125 to acceptor substrate selectivity.

Protein folding is directly influenced by the nature of the amino acid sequence and, in particular, by side-chain flexibility. Single amino acid substitutions can produce alterations in enzyme-kinetic parameters without altering the secondary structure [22,25]. In contrast, point mutations can also cause changes in local secondary structure with no effect on protein function [26]. In our studies we were unable to ascertain the extent of the structural perturbations that our mutagenesis produces, owing to a lack of three-dimensional structural information. However, enzyme activity is dependent on the protein's folding correctly into its native structure. Thus variations in enzyme activity resulting from our site-directed mutagenesis are representative of the effect of the modified residues and three-dimensional shape. The enzyme activity of N2-SF/SR/SY is inhibited when the final Ac-CoA concentration is more than 100 μ M. Given that Ac-CoA is a rather large molecule, with a molecular mass of 809.6 Da, perhaps the observed enzyme inhibition is a result of a small alteration in the local secondary structure that prevents the long β -mercaptoethylamine and pantothenate units of Ac-CoA from obtaining efficient entry into and exit from the enzyme in the presence of higher Ac-CoA concentrations.

Both enzyme stability and recombinant protein expression efficiency can influence the levels of enzyme activity observed in a system *in vitro*. However, these are not the only factors responsible for the molecular activity of the NAT1 mutant proteins. For instance, N1-RS, with a soluble NAT expression level comparable with that of N1, had a low molecular activity

for PAS that might be attributed to decreased stability *in vitro*. In contrast, N1-YS displayed marked decreases in both intrinsic stability and soluble immunoreactive NAT protein, yet still had the highest molecular activity for PAS.

In conclusion, we have demonstrated that selected single point mutations in the human NAT proteins can produce drastic effects on their kinetic parameters. Specifically, residues Arg¹²⁷ and Ser¹²⁵ make significant contributions to NAT1-type and NAT2-type behaviour respectively. In view of previous observations that discrete regions of the primary structure of NATs contribute to enzyme specificity [11], our results suggest that residues 125, 127 and 129 might contribute to forming the active-site region surrounding the catalytic residue Cys⁶⁸.

We are grateful to Dr N. Hughes and Dr A. Gaedigk for their technical assistance. This work was supported by Grant MT10890 and a Scientist Award (to D. M. G.) from the Medical Research Council of Canada, and a Studentship (to G. H. G.) from the Research Institute, Hospital for Sick Children.

REFERENCES

- Blum, M., Grant, D. M., McBride, O. W., Heim, M. and Meyer, U. A. (1990) DNA Cell Biol. **9**, 193–203
- Matas, N., Thygesen, P., Stacey, M., Risch, A. and Sim, E. (1997) Cytogenet. Cell. Genet. **77**, 290–295
- Vatsis, K. P., Weber, W. W., Bell, D. A., Dupret, J.-M., Price Evans, D. A., Grant, D. M., Hein, D. W., Lin, H. J., Meyer, U. A., Relling, M. V. et al. (1995) Pharmacogenetics **5**, 1–17
- Grant, D. M., Mörike, K., Eichelbaum, M. and Meyer, U. A. (1990) J. Clin. Invest. **85**, 968–972
- Grant, D. M., Blum, M., Beer, M. and Meyer, U. A. (1991) Mol. Pharmacol. **39**, 184–191
- Hughes, N. C., Janezic, S. A., McQueen, K. L., Jewett, M. A., Castranio, T., Bell, D. A. and Grant, D. M. (1998) Pharmacogenetics **8**, 55–66
- Hein, D. W., Doll, M. A., Rustan, T. D., Gray, K., Feng, Y., Ferguson, R. J. and Grant, D. M. (1993) Carcinogenesis **14**, 1633–1638
- Badawi, A. F., Stern, S. J., Lang, N. P. and Kadlubar, F. F. (1996) Prog. Clin. Biol. Res. **395**, 109–140
- Grant, D. M. and Meyer, U. A. (1993) in Human Protein Data, vol. 2 (Haeberli, A., ed.), VCH, Weinheim
- Dupret, J.-M. and Grant, D. M. (1992) J. Biol. Chem. **267**, 7381–7385
- Dupret, J.-M., Goodfellow, G. H., Janezic, S. A. and Grant, D. M. (1994) J. Biol. Chem. **269**, 26830–26835
- Watanabe, M., Sofuni, T. and Nohmi, T. (1992) J. Biol. Chem. **267**, 8429–8436
- Delomélie, C., Goodfellow, G. H., Krishnamoorthy, R., Grant, D. M. and Dupret, J.-M. (1997) Biochem. J. **323**, 207–215
- du Souich, P., McLean, A. J., Stoeckel, K., Ohlendorf, D. and Gibaldi, M. (1979) Clin. Pharmacol. Ther. **26**, 757–765
- Vallette, F., Mege, E., Reiss, A. and Adesnik, M. (1989) Nucleic Acids Res. **17**, 723–733
- Kadowaki, H., Kadowaki, T., Wondisford, F. E. and Taylor, S. I. (1989) Gene **76**, 161–166
- Grant, D. M., Lottspeich, F. and Meyer, U. A. (1989) FEBS Lett. **244**, 203–207
- Bradford, M. M. (1976) Anal. Biochem. **72**, 248–254
- Barski, O. A., Gabbay, K. H. and Bohren, K. M. (1996) Biochemistry **35**, 14276–14280
- Wuytens, G., Verschueren, K., de Winter, J. P., Gajendran, N., Beek, L., Devos, K., Bosman, F., de Waele, P., Andries, M., van den Eijnden-van Raaij, A. J. et al. (1999) J. Biol. Chem. **274**, 9821–9827
- Merickel, A., Kaback, H. R. and Edwards, R. H. (1997) J. Biol. Chem. **272**, 5403–5408
- Vaughan, C. K., Buckle, A. M. and Fersht, A. R. (1999) J. Mol. Biol. **286**, 1487–1506
- Cheon, H. G. and Hanna, P. E. (1992) Biochem. Pharmacol. **43**, 2255–2268
- Baker, E. N. and Hubbard, R. E. (1984) Prog. Biophys. Mol. Biol. **44**, 235–238
- Zhang, X. J., Baase, W. A., Shoichet, B. K., Wilson, K. P. and Matthews, B. W. (1995) Protein Eng. **8**, 1017–1022
- Cordes, M. H., Walsh, N. P., McKnight, C. J. and Sauer, R. T. (1999) Science **284**, 325–328