

Homology modelling and structural analysis of human arylamine *N*-acetyltransferase NAT1: evidence for the conservation of a cysteine protease catalytic domain and an active-site loop

Fernando RODRIGUES-LIMA*, Claudine DELOMÉNIÉ†, Geoffrey H. GOODFELLOW‡, Denis M. GRANT‡ and Jean-Marie DUPRET*¹

*CNRS-UMR7000, Faculté de Médecine Pitié-Salpêtrière, 105 bd de l'Hôpital, 75013 Paris, France, †INSERM U458, Hôpital Robert Debré, 48 bd Sérurier, 75019 Paris, France, and ‡Genetics and Genomic Biology Program, Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G1X8

Arylamine *N*-acetyltransferases (EC 2.3.1.5) (NATs) catalyse the biotransformation of many primary arylamines, hydrazines and their *N*-hydroxylated metabolites, thereby playing an important role in both the detoxification and metabolic activation of numerous xenobiotics. The recently published crystal structure of the *Salmonella typhimurium* NAT (*S*tNAT) revealed the existence of a cysteine protease-like (Cys-His-Asp) catalytic triad. In the present study, a three-dimensional homology model of human NAT1, based upon the crystal structure of *S*tNAT [Sinclair, Sandy, Delgoda, Sim and Noble (2000) *Nat. Struct. Biol.* 7, 560–564], is demonstrated. Alignment of *S*tNAT and NAT1, together with secondary structure predictions, have defined a consensus region (residues 29–131) in which 37% of the residues are conserved. Homology modelling provided a good quality model of the corresponding region in human

NAT1. The location of the catalytic triad was found to be identical in *S*tNAT and NAT1. Comparison of active-site structural elements revealed that a similar length loop is conserved in both species (residues 122–131 in NAT1 model and residues 122–133 in *S*tNAT). This observation may explain the involvement of residues 125, 127 and 129 in human NAT substrate selectivity. Our model, and the fact that cysteine protease inhibitors do not affect the activity of NAT1, suggests that human NATs may have adapted a common catalytic mechanism from cysteine proteases to accommodate it for acetyl-transfer reactions.

Key words: acetyl-CoA, flexibility, steric hindrance, structure prediction, substrate specificity.

INTRODUCTION

The acetyl coenzyme A (Ac-CoA):arylamine *N*-acetyltransferases (NATs; EC 2.3.1.5), catalyse the transfer of an acetyl group from Ac-CoA to the nitrogen or oxygen atom of primary arylamines, hydrazines and their *N*-hydroxylated metabolites. They, therefore, play an important role in the detoxification and potential metabolic activation of numerous xenobiotics. Human NAT1 and NAT2 are encoded by two distinct loci located on chromosome 8 [1]. NAT2 protein content is the basis of the well-known isoniazid acetylation polymorphism, resulting in significant toxicological implications. Although *NAT1* expression is independent of the classically defined acetylation polymorphism, it also displays marked inter-individual variations which may be a source of pharmacological susceptibility [2,3].

NAT isoforms have been detected in several species, including mammals [4–8], chickens [9], and prokaryotes [10,11]. The crystallographic structure of *Salmonella typhimurium* NAT (*S*tNAT) was recently resolved [12]. It allowed the identification of three critical residues, which presumably act together as a protease-like catalytic triad (residues Cys⁶⁹-His¹⁰⁷-Asp¹²²). Experimental evidence is consistent with the participation of a cysteine residue in the catalytic activity of NATs. Site-directed mutagenesis has shown that the Cys⁶⁹ residue in *S*tNAT [13] or the homologous human Cys⁶⁸ [14] mediate Ac-CoA binding and are thus essential for enzyme activity. It has also been proposed

that a highly conserved basic residue, corresponding to position Arg⁶⁴ in the human NATs, contributes to the conformational stability of NATs [15]. The *S*tNAT structural data are consistent with the proposed roles of the human Cys⁶⁸ and Arg⁶⁴ residues [12].

Although human NAT1 and NAT2 share 81% amino-acid-sequence identity [4,16], their kinetic selectivity for amine-containing acceptor substrates differ markedly [17]. Functional analysis of recombinant NAT1/NAT2 chimaeras [18] and site-directed mutagenesis [19] have suggested that amino acids 125 and 127 are important determinants of NAT1-type and NAT2-type acceptor substrate selectivity. These residues have been proposed to be located either along a passageway to the active-site region or within the active site itself [19]. Elucidation of the structure of human NATs may confirm this functional hypothesis.

We have used a computational approach to determine the structure of human NAT1, which circumvented the absence of a crystallographic structure of human NATs. This approach produces valid structural models for protein sequences with available related templates (defined as > 30–35% amino-acid-sequence identity) [20]. Comparison of these homology models with known structures may also reveal similarities which allow biochemical and biological functions to be inferred [21]. Due to the similarity between *S*tNAT and human NAT1, we used *S*tNAT as a structural template for homology modelling. In the

Abbreviations used: Ac-CoA, acetyl-CoA; NAT, arylamine *N*-acetyltransferase; *S*tNAT, *Salmonella typhimurium* NAT; r.m.s.d., root-mean-square deviation.

¹ To whom correspondence should be addressed (e-mail jmdupret@infobiogen.fr).

present paper, we describe a homology model encompassing the N-terminal region of human NAT1 (residues 29–131). We show that *S.typhimurium* NAT and human NAT1 have a similar three-dimensional structure within this N-terminal region. Furthermore, structural analyses of our NAT1 model and previous studies are consistent with the hypothesis that human NATs share a common partial mechanistic strategy with cysteine proteases and possess the required groups for acetyl transfer from Ac-CoA to the acceptor substrates. We also suggest that there are structural determinants which control substrate specificity.

EXPERIMENTAL

Bacterial strain and reagents

The phagemid expression vector, pKEN2, and the *Escherichia coli* host strain, XA90, (*F'lacI^{q1}*) were a gift from Dr G. L. Verdine (Department of Chemistry, Harvard University, Cambridge, MA, U.S.A.). The arylamine acceptor substrate, 2-aminofluorene, the acetyl donor, Ac-CoA, and the cysteine protease inhibitors E64 and leupeptin were purchased from Sigma.

Sequence alignments, secondary structure prediction, and protein fold recognition

Sequences of human NAT1 (SwissProt P18440), NAT2 (SwissProt P11245) and *S. typhimurium* NAT (SwissProt Q00267) were used. The BLAST 2 (BLOSUM 62 matrix) search engine [22] was used to create sequence alignments, whereas Clustal W1.8 [23] enabled the alignment of multiple sequences. The neural networking PHD method [24] was used to predict secondary structure elements. Protein-fold recognition was carried out by the 3D-PSSM [25] and FRUSDP [26] threading algorithms.

Homology modelling of the NAT1 protein structure and comparison of structures

The three-dimensional structure of the N-terminal domain of the human NAT1 enzyme was modelled by comparative protein modelling methods using the program SWISS-MODEL [27] in the optimized mode. The structure of the enzyme domain was modelled on the basis of its structural similarity with the N-terminal domain of the *S.typhimurium* NAT (Protein Data Bank entry 1e2t). The degree of identity between the template (residues 29–131) and the NAT1 sequence (residues 29–131) was 37%, which enabled a preliminary model to be generated by SWISS-MODEL. The sequence alignment was then improved manually; Swiss-PdbViewer 3.5 [27] was used to produce a structure-based alignment and SWISS-MODEL was used in the optimized mode to minimize energy. The final model was evaluated with PROCHECK [28], and Swiss-PdbViewer 3.5 was then used to analyse and visualize the structures. The incremental combinatorial extension [29] and the TOP [30] algorithms were used to align and compare the three-dimensional structures. The TOP algorithm was used to search the Protein Data Bank and the SCOP database [31] for structural similarities.

NAT1 activity assay in the presence of protease inhibitors

Recombinant human NAT1 was expressed from p-NAT1 as described previously [14]. Undiluted recombinant *E. coli* lysate was pre-incubated at 37 °C with various concentrations of the cysteine protease inhibitors E64 or leupeptin. Aliquots were then removed at selected times for subsequent enzyme assays. The

acetylation of 2-aminofluorene was assessed in 5 min assays with 100 μ M Ac-CoA in the absence of the Ac-CoA regenerating system (D,L-acetylcarnitine and carnitine acetyltransferase). Reverse-phase HPLC [14] was used to quantify the acetylated products of NAT activity.

RESULTS

Amino acid sequence alignment and fold analysis of *S. typhimurium* NAT, human NAT1 and NAT2

We used BLAST 2 and Clustal W1.8 to compare the amino acid sequences of *S. typhimurium* NAT (*S.typhimurium* NAT) and human NAT1 and NAT2. The amino acid sequences of the two human NATs were 81% identical (Figure 1), whereas *S.typhimurium* NAT was less similar to the human NATs (27% with NAT1, 29% with NAT2). The N-terminal region of the NAT sequences (residues 1–150) showed a higher degree of amino-acid identity (35% between *S.typhimurium* NAT and NAT1, 31% between *S.typhimurium* NAT and NAT2), and contained three highly conserved sequences (corresponding to the NAT1 residues 34–42, 64–85, and 107–131 in Figure 1). Three strictly conserved residues (Cys⁶⁹, His¹⁰⁷, and Asp¹²²), which form the catalytic triad of *S.typhimurium* NAT, are located within these conserved sequences. The overall identity between *S.typhimurium* NAT and NAT2 is only modestly higher than that between *S.typhimurium* NAT and NAT1 (29% and 27% identity respectively); however, the catalytic core of *S.typhimurium* NAT (residues 63–131) showed a markedly higher identity with the NAT1 catalytic core (residues 63–131) (40% identity with NAT1, 31% identity with NAT2).

3D-PSSM and FRUSDP algorithms were used for fold recognition analyses by threading and gave similar results. The predicted three-dimensional fold of the two human NATs is consistent with the structure of *S.typhimurium* NAT. Furthermore, the best sequence/structure alignment of human NATs and *S.typhimurium* NAT structure was within these three N-terminal conserved regions. Overall, this suggests that the N-terminal region of human NAT proteins, especially their catalytic core, adopt the *S.typhimurium* NAT fold (Figure 1).

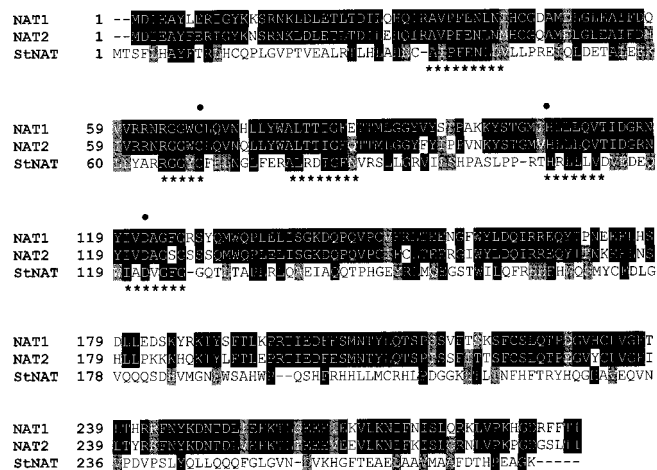


Figure 1 Amino acid sequence alignment of human NATs and *S.typhimurium* NAT

The following sequences (with accession numbers in parentheses) were used for the alignments: human NAT1 (P18440), human NAT2 (P11245) and *S. typhimurium* NAT (Q00267). Alignments were made by Clustal W1.8 analysis. Conserved or similar amino acids are shaded. Catalytic residues are indicated by a dot. Conserved regions of human NATs threaded with 3D-PSSM and matching the 3-dimensional structure of *S.typhimurium* NAT are indicated by asterisks.

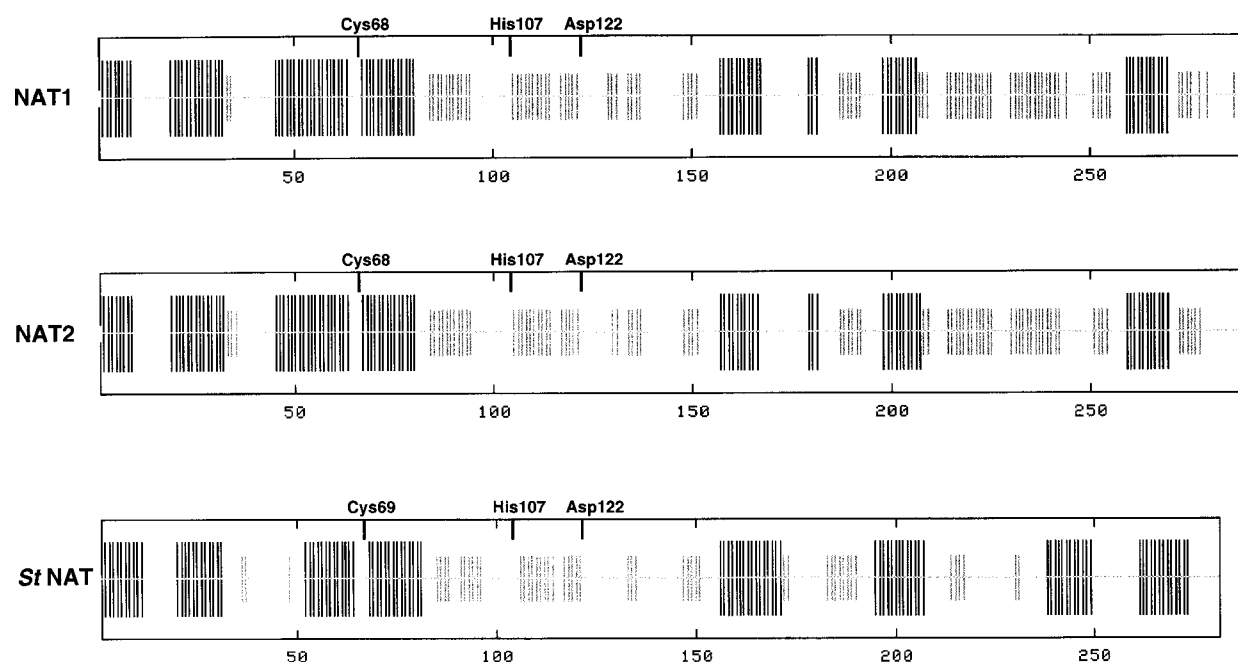


Figure 2 Comparison of secondary structure prediction of human NATs and *S*/NAT

Secondary structure elements were predicted by PHDsec. Helix, strand and coil regions are represented by black bars, grey bars (smaller) and grey lines respectively. Catalytic triad residues are indicated above the bars.

Prediction and comparison of secondary structures

The PHD algorithm [24] was used to predict the secondary structures of *S*/NAT, NAT1 and NAT2. The predicted structural elements of the N-terminal region (residues 1–150) were strongly conserved and similarly arranged for all three enzymes (Figure 2). Moreover, the residues of the catalytic triad were identified in the same structural elements for each enzyme. It was predicted that Cys, His, and Asp residues would be located at the beginning of an α -helix, at the beginning of a β -strand and at the end of a β -strand respectively (Figure 2). These predictions of the structural elements, in particular the catalytic triad, are consistent with the crystal structure of *S*/NAT [12], and suggest that the N-terminal region of the three NAT enzymes may have similar three-dimensional structures.

Molecular modelling of human NAT1 enzyme

Sequence, secondary structure and fold recognition analyses suggested that the *S*/NAT structure [12] could be used as a template for homology modelling of the NAT1 enzyme. We used the SWISS-MODEL program to produce a homology-based model of the human NAT1 enzyme. A region of the N-terminal domain of NAT1 (residues 29–131), which encompasses the catalytic core (residues 63–131), was selected because it had the highest sequence identity (37%) to the N-terminal domain of *S*/NAT (residues 29–131). The percentage of sequence identity was above the 30% limit that is generally considered to be the threshold limit for accurate homology modelling [32]. This indicated that the homology modelling of the human NAT1 enzyme (residues 29–131), based on the *S*/NAT crystal structure (residues 29–131), would presumably yield a reliable structure [33].

The best model obtained for the N-terminal domain of NAT1 (residues 29–131) is shown as a Swiss-PdbViewer representation

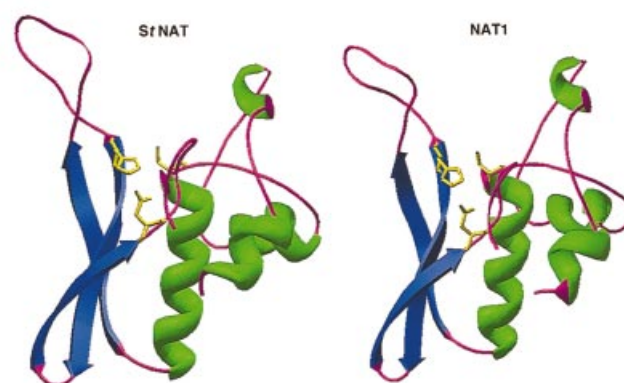


Figure 3 Structural model of human NAT1 (residues 29–131)

Swiss-PdbViewer (ribbon diagram) representation of the homology-modelled human NAT1 (residues 29–131) and the three-dimensional structure of *S*/NAT (residues 29–131; Protein Data Bank entry 1e2t). Catalytic triad residues are shown as stick models (yellow).

(Figure 3). The quality of this model was evaluated using PROCHECK. No residue was found in the disallowed regions of the Ramachandran diagram. The total quality G-factor was -0.1 , which is indicative of a good quality model (acceptable values of the G-factor in PROCHECK are between 0 and -0.5 , with the best models displaying values close to zero). As expected, the backbone of the predicted structure was extremely similar to the crystal structure of *S*/NAT, given the identity of the sequence with the template ([27] and Figure 3). In fact, the structural alignment of both structures using combinatorial extension and TOP algorithms showed that the α -carbon co-

				RMSD (C α)	C α aligned			
				Å				
NAT1	66	GWCLQVNHLLYWA [.]	105	MIHL [.] LLQV	118	NYLIVDAG [.]	-	-
FXIII	312	GQCWVFAGYFNTFL	371	NYHCWNEA	392	WQAVDET	1.1	52
CATX	29	GSCWAHASTSAMAD	178	LNHV [.] LSVA	196	WIVR [.] NEW	1.1	38
STAPH	22	GWCAGYTMSALLNA	117	AGHAWAVV	137	IILLW [.] LPW	1.5	44
BLEOH	71	PRCWIFSCINVMRI	369	MTHT [.] MTFT	392	WRV [.] ENCW	1.7	44

Figure 4 Alignment of NAT1 with structural homologues

The TOP algorithm was used to search the Protein Data Bank and SCOP databases for NAT1 structural homologues. The four proteins with the best alignments are shown (FXIII, human factor XIII; CATX, human cathepsin X; STAPH, bacterial staphopain; BLEOH, human bleomycin hydrolase). For each protein, the number of C α atoms aligned with NAT1 structure are indicated, as is the r.m.s.d. value obtained for these atoms. Only structurally aligned regions containing a catalytic triad residue are shown. Catalytic triad residues are indicated by a dot. Conserved or similar residues are shaded.

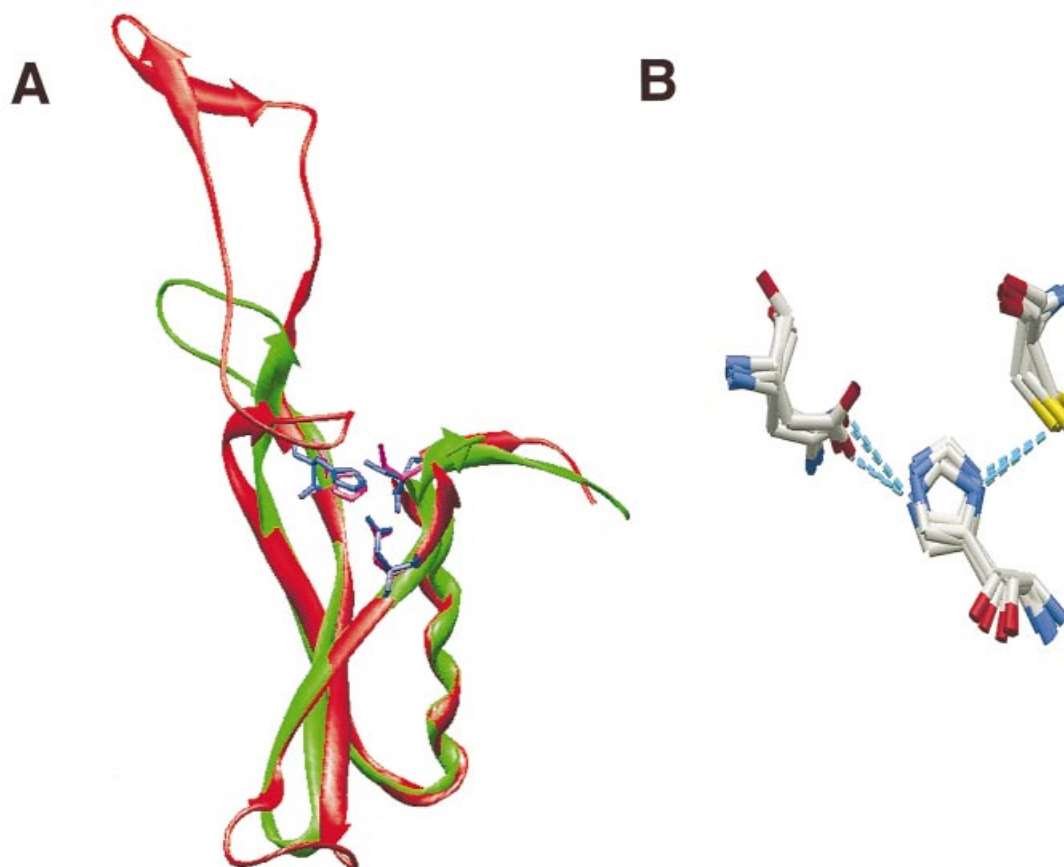


Figure 5 Structural analysis of human NAT1 catalytic core (residues 63–131)

(A) Structural alignment of NAT1 model and of human factor XIII structure. The TOP algorithm (least-squares alignment of C α atoms) was used to align the NAT1 model and human factor XIII structure (Protein Data Bank entry 1 evu). The Swiss-PdbViewer (ribbon diagram) representation was used to show the best aligned regions of NAT1 (residues 63–131, green) and factor XIII (residues 309–400, red). NAT1 and factor XIII catalytic triad are in magenta and blue respectively. (B) Comparison of human NAT1 catalytic triad with members of the cysteine protease family. Catalytic triads of NAT1, human factor XIII, human cathepsin X, human bleomycin hydrolase and staphopain are shown following least-squares overlap on the triad residue C α atoms using Swiss-PdbViewer. Hydrogen bonds are shown as broken lines.

ordinates differed by only 0.5 Å (1 Å = 10⁻¹⁰ m). Both structures contained a domain (residues 68–131 in NAT1) consisting of a segregated α -helix and three strands of an antiparallel β -sheet, which is similar to those in alpha and beta (a + b) proteins (Figure 3). One of the most interesting features of the *Sr*NAT crystal structure was the presence of a cysteine protease-like catalytic triad ([33] and Figure 3). The NAT1 homology model

demonstrated that the location of the catalytic triad (residues Cys⁶⁸, His¹⁰⁷ and Asp¹²²) in this alpha and beta (a + b) fold (Figure 3) was conserved from *Sr*NAT [12] to NAT1. Overlap of the NAT1 model catalytic triad (Cys⁶⁸, His¹⁰⁷, Asp¹²²) with the *Sr*NAT equivalent yielded a root-mean-square deviation (r.m.s.d.) value on the three C α atoms of only 0.2 Å (results not shown).

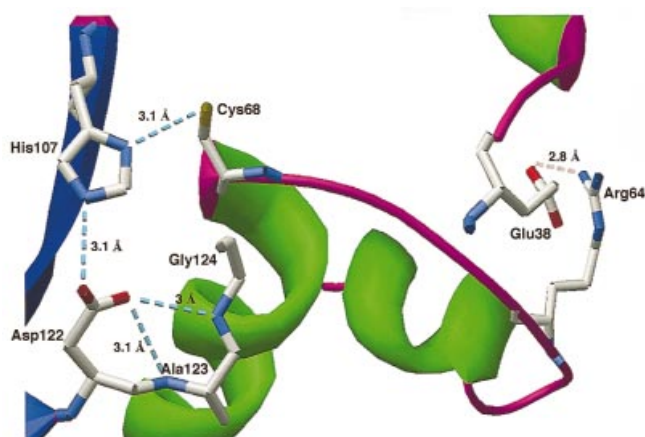


Figure 6 Structural model of the human NAT1 catalytic triad

NAT1 catalytic triad residues are shown. The hydrogen bond network is shown by turquoise broken lines. Salt bridge interactions between the highly conserved residues, Glu³⁸ and Arg⁶⁴, are shown by red hatched lines.

We employed the same approach to model the N-terminal domain of NAT2 (residues 29–131), which shares 88% and 31% identity with the equivalent domains of NAT1 and *St*NAT respectively. Interestingly, despite the high sequence identity between the N-terminal regions of NAT1 and NAT2, the PROCHECK algorithm found that the NAT2 model was not sufficiently reliable with a PROCHECK G-factor of -0.48 (results not shown). Sequence alignment of the human NAT proteins (Figure 1) showed that the N-terminal regions used for modelling differed by thirteen non-conservative amino acids and four conservative amino acids.

Structural analysis of the NAT1 catalytic core

Our modelled NAT1 structure (residues 29–131) was used to search for similar protein structures and protein domain structures in the Protein Data Bank and SCOP database using the TOP algorithm. The optimal structural alignments (Figure 4) of C α atoms were found with the catalytic cores from two members of the cysteine protease family: human factor XIII transglutaminase (Protein Data Bank entry 1evu) and human procathepsin X (Protein Data Bank entry 1deu). Indeed, the α -carbon co-ordinates of the structurally aligned residues (52 equivalent C α atoms for factor XIII and 38 for procathepsin X) differed from the NAT1 model by just 1.1 Å (Figure 4). Good structural alignments (Figure 4) were also obtained with the catalytic core of bacterial staphopain (Protein Data Bank entry 1cv8) and human bleomycin hydrolase (Protein Data Bank entry 1cb5), which had C α atoms r.m.s.d. values of 1.5 Å and 1.7 Å respectively. The domains of these four proteases, which match the NAT1 catalytic core structure, contain the cysteine protease family catalytic triad (Cys, His, Asp or Asn) (Figure 4). In NAT1, *St*NAT [12] and the cysteine protease family [34], the catalytic triad is clustered in an alpha and beta (a+b) motif which consists of an α -helix and three β -sheets (Figure 5A and [12,34]). Furthermore, the structural alignment of the NAT1 model catalytic core with that of human factor XIII clearly showed that the catalytic triad of each enzyme had a similar spatial position within the alpha and beta (a+b) fold (Figure 5A). In fact, the overlap of the NAT1 model active site triad

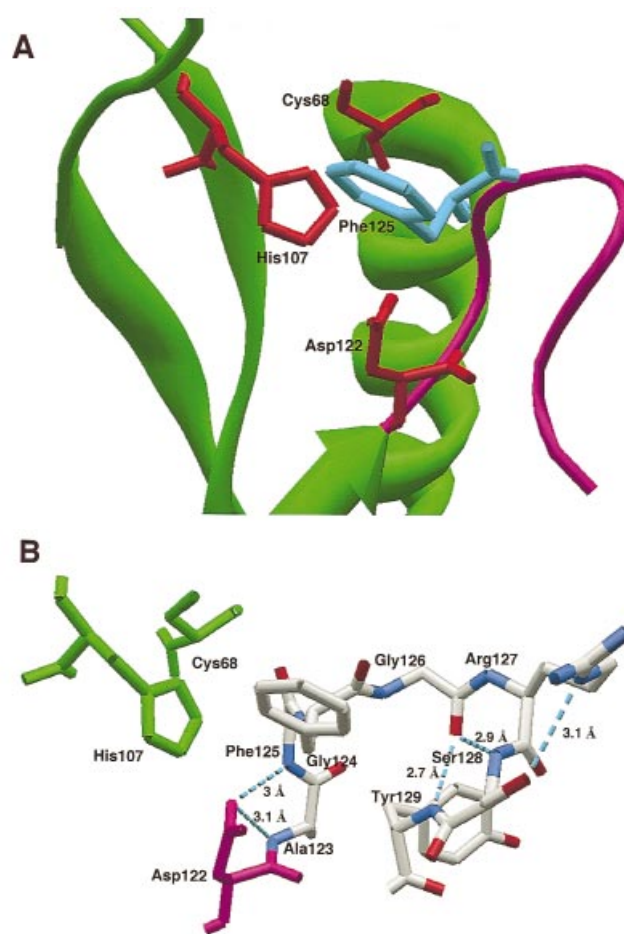


Figure 7 Structural model of the human NAT1 active site loop

(A) NAT1 catalytic core and active site loop. NAT1 catalytic triad residues are shown in red. The Phe¹²⁵ residue is shown in turquoise. The loop is shown in magenta. (B) NAT1 active site loop hydrogen bond network (residues 122–129). NAT1 catalytic triad residues are shown in green, except for Asp¹²² which is in magenta. Loop residues are in grey (residues 123–129). Backbone amine and carboxy groups are in blue and red respectively. Hydrogen bonds are shown by turquoise broken lines.

(Cys⁶⁸, His¹⁰⁷, Asp¹²²) with the equivalent catalytic triad of human factor XIII yielded a r.m.s.d. value on the three C α atoms of 0.4 Å (Figure 5). Similar results were obtained with the catalytic triad of the other cysteine protease structures (Figures 4 and 5B).

Overall, this structural evidence suggests that NAT1 and members of the cysteine protease family indeed share a common catalytic mechanism in which an ‘activated’ cysteine residue acts as a catalytic nucleophilic group. The kinetic mechanism of cysteine proteases involves the activation of the catalytic triad cysteine residue by a hydrogen bond network in which the Asp (or Asn) and His residues of the triad increases the nucleophilicity of the thiol group by a ‘charge-relay system’ [34,35]. The predicted hydrogen bonds between the carboxylate group of Asp¹²² and the imidazole base of His¹⁰⁷, and between the thiol group of Cys⁶⁸ and the imidazole base of His¹⁰⁷ are consistent and support the existence of such a hydrogen bond network in NAT1 (Figure 6). In addition, we found that the Asp¹²² residue could also be stabilized by hydrogen bonds with the backbone amine groups of Ala¹²³ and Gly¹²⁴. The highly

conserved Arg⁶⁴ and Glu³⁸ residues may provide stabilization of the catalytic Cys⁶⁸ residue through a salt-bridge interaction (salt-bridge length of 2.8 Å in NAT1) akin to that for *Sr*NAT [12]. To determine the specificity of the NAT1 active site, we assessed whether NAT1 enzyme activity could be inhibited by two cysteine protease inhibitors. NAT1 activity was not inhibited, even in the presence of high concentrations (1 mM final) of the specific cysteine protease inhibitor, E64, or the less specific inhibitor, leupeptin (results not shown).

Our NAT1 model revealed that a loop (coloured magenta in Figure 7A.), originating at the catalytic residue Asp¹²², is part of the catalytic pocket. It has been suggested that the residues in this loop may be involved in the binding of Ac-CoA [12] and are also perhaps important determinants of the substrate selectivity for human NATs [19]. Site-directed mutagenesis suggested that the NAT1 Phe¹²⁵ residue (corresponding to Ser¹²⁵ in NAT2) is involved in substrate recognition [19]. Our NAT1 model (Figure 7A) clearly showed that the phenyl ring of the Phe¹²⁵ residue is proximal to the catalytic triad, with an average distance of 3.4 Å to the His¹⁰⁷ and Asp¹²² residues. Since this Phe¹²⁵ residue takes up considerably more space than the hydroxymethyl side chain of Ser¹²⁵ in NAT2 (200 Å³ and 100 Å³ respectively), then it probably restricts access of the large substrates to the active site. Thus this may explain why NAT1-selective substrates are generally smaller than NAT2-selective substrates. It has also been suggested that a difference in protein flexibility, in particular at position 127 (Arg¹²⁷ for NAT1, Ser¹²⁷ for NAT2), is involved in the catalytic specificity of the human NATs [19].

When we examined the hydrogen-bond network of the NAT1 model in more detail, we found that the backbone oxygen atom of Gly¹²⁶ was linked by hydrogen bonds to the backbone nitrogen atoms of Ser¹²⁸ and Tyr¹²⁹ (Figure 7B). Substitution of Tyr¹²⁹ with a serine residue (as in NAT2) in the NAT1 model still allowed hydrogen bonds to form (results not shown). In addition, as mentioned above, the lateral chain of the catalytic Asp¹²² residue is paired by hydrogen bonds to the backbone amine groups of Ala¹²³ and Gly¹²⁴. Interestingly, the Gly¹²⁴ residue is strictly conserved in all NAT species. We also found, for the NAT1 model, that the side chain epsilon nitrogen atom of Arg¹²⁷ (which is replaced by Ser¹²⁷ in NAT2) and the side chain oxygen atom of Ser¹²⁸ are paired by a hydrogen bond. When the Arg¹²⁷ residue was replaced by a serine residue (as in NAT2), this hydrogen bond no longer existed (results not shown). This suggests that this hydrogen bond may only exist in NAT1. Thus the loop starting at residue Asp¹²² may contribute to the formation of the active-site region and be stabilized by hydrogen bonds (Figure 7B).

DISCUSSION

The recent determination of the three-dimensional structure of *Sr*NAT has provided a better understanding of the molecular mechanisms that underlie the biological function of the arylamine *N*-acetyltransferases [12]. It has also provided the information required to build computer-generated structural models for other NATs. This approach has been proved to be more reliable than approaches based solely on sequence similarities [21,36].

We used the three-dimensional structure of *Sr*NAT to build a molecular model of the N-terminal region (residues 29–131) of human NAT1 (Figure 3). The *Sr*NAT and human NATs share a greater percentage of amino-acid identity within this region (Figure 1), and are predicted to have comparable secondary structure elements adopting very similar arrangements (Figure 2). Fold-recognition analysis supported that the structure of

*Sr*NAT could be used as a template for homology modelling of the NAT1 N-terminal region (residues 29–131). Furthermore, it has been demonstrated that the first 150 N-terminal amino acids of NAT1 are sufficient to bind Ac-CoA, and that a fragment containing the first 204 N-terminal amino acids catalyses the formation of an acetylated enzyme intermediate [37]. A good quality homology model was obtained for NAT1 and demonstrated that the modelled N-terminal region *Sr*NAT and NAT1 adopt a similar three-dimensional structure (Figure 3). As expected, our homology model shows that both *Sr*NAT and NAT1 possess a catalytic triad (Cys⁶⁸-His¹⁰⁷-Asp¹²² in NAT1) in a similar spatial pattern, suggesting that these triads are a common feature of all members of the NATs family. Aspartate and histidine residues probably participate in the withdrawal of a proton from the active thiol group [35]. The resulting thiol nucleophile may act as an acetyl acceptor, allowing the generation of an acetylated enzyme intermediate [38]. Site-directed mutagenesis has also provided evidence for the direct involvement of the triad cysteine residue in the catalytic mechanism of both human NATs [14] and *Sr*NAT [13]. The *Sr*NAT structure [12] and our NAT1 model also support previous findings in which the highly conserved human Arg⁶⁴ has a structural role [15] rather than being involved in general base catalysis [39].

Cysteine-protease-class enzymes also possess a catalytic triad [34,40], either Cys-His-Asp or Cys-His-Asn. The observation that NATs and members of the cysteine protease family share a common catalytic core fold, which contains a triad in the same spatial position (Figure 5), probably reflects the significant contribution of these specific residues towards proton transfer during the catalytic process. Thus both enzyme families probably incorporate a common mechanistic pathway to lower the free energy for a part of their rate-limiting transition states. Similar structural features have been reported for other enzymes that catalyse different overall reactions, which undoubtedly decreases the number of unique protein folds required to support metabolic reactions [41]. The lack of NAT1 inhibition by cysteine protease inhibitors suggests that NAT1 and cysteine proteases probably have distinct substrate specificities, and therefore different functions. Thus the NAT family may be a member of a growing class of enzymes recruited from proteases that have adapted a common enzymic mechanism and accommodated it to a group of transfer reactions involving Ac-CoA and arylamines as substrates [42].

We have previously identified three amino acids which may contribute to the acceptor substrate selectivity of human NAT1 and NAT2 [19]. Among these residues, Phe¹²⁵ of NAT1 was proposed to be involved in steric control of substrate recognition. The phenyl ring in Phe¹²⁵ might facilitate the access of NAT1-selective substrates, such as *p*-aminosalicylic acid, and restrict the access of larger NAT2-selective substrates, such as sulphamethazine. Our present data support this hypothesis, since we predict that Phe¹²⁵ is part of a loop (spanning residues Asp¹²² to Met¹³¹) which is involved in the formation of the catalytic region. The NAT1 model further shows that the bulky side-chain of Phe¹²⁵ (average volume approx. 200 Å³) is located proximal to the catalytic His¹⁰⁷ residue and faces a passageway to the catalytic core (Figure 7A). The distance between Phe¹²⁵ and the residues of the catalytic triad was estimated to be between 3.3 Å and 5.7 Å (Figure 7A). The presence of a smaller amino acid residue, serine (average volume approx. 100 Å³), at position 125 in human NAT2 supports the hypothesis that the amino acid at this position may be an important determinant of acceptor-substrate selectivity in NATs. Thus, a phenylalanine residue at position 125 may cause steric hindrance and limit substrate access, whereas serine allows the access of a broader range of substrates. It is

noteworthy that most known bacterial NATs contain a conserved phenylalanine residue aligning with Phe¹²⁵ in the NAT1 sequence. Their low *N*-acetylation activity with sulphamethazine and high catalytic efficiency with 5-aminosalicylic acid demonstrates that they exhibit a similar reaction spectrum to human NAT1 [42a]. It has been shown that substrate specificity recognition by protein-tyrosine phosphatases was influenced by the steric hindrance of an arginine residue located in the vicinity of the active site [43]. Similar observations have been also reported for pancreatic lipases [44].

Although human NAT1 and NAT2 have markedly different kinetic selectivities, they also present a certain degree of cross-reactivity. This is shown by the existence of an amine substrate, aminofluorene, which has similar Michaelis constants for both enzymes [14]. This suggests that, in addition to steric hindrance, the flexibility of the regions surrounding the active site is also important for the acceptor-substrate recognition of NATs [19]. Our present results are consistent with such a hypothesis, since the NAT1 model predicts that Phe¹²⁵ and two other amino acids, previously shown to be involved in the acceptor-substrate specificity of human NATs (Arg¹²⁷ and Tyr¹²⁹ in NAT1), are located in a flexible loop within the active site region (Figure 7). Interestingly, an analogous flexible loop, present in the cysteine-protease-like catalytic region of a human ubiquitin hydrolase, may be involved in substrate specificity through structural flexibility [45]. In many enzymes loops often contain functional residues, since they tend to be more flexible than helical and sheet structures during conformational changes [46]. The three-dimensional architecture of these protein loops is not solely determined by their primary amino-acid sequences, but also involves interactions such as hydrogen bonding [46].

Our NAT1 model predicts that this above-mentioned loop is stabilized by five hydrogen bonds, two of which involve the lateral chain of the Asp¹²² catalytic residue and backbone atoms (Figure 7B). One of those backbone atoms belongs to the strictly conserved residue Gly¹²⁴. The presence of such a conserved glycine residue at the beginning of the loop may confer a high degree of local flexibility. Moreover, the presence of a residue lacking a side chain at position 124 may allow the adjacent polypeptides, namely the helix and a β -strand of the catalytic-site core, to pack closely together. Further site-directed mutagenesis of this strictly conserved residue will confirm this hypothesis. Interestingly, it has been shown that a conserved glycine residue contributes to active-site loop flexibility and activity of pepsin [47].

It is also predicted that there is a single hydrogen bond between the side chains of Arg¹²⁷ and Ser¹²⁸ of NAT1. Since this loop is located within a central region conserved between NAT1 and NAT2 (which only differ at positions 125, 127 and 129), it can be assumed that the same loop is also present in NAT2. In this example, the presence of Ser¹²⁷ in place of Arg¹²⁷ impairs hydrogen-bond formation with Ser¹²⁸. The reduced number of predicted hydrogen bonds in NAT2 (four hydrogen bonds) with respect to NAT1 (five hydrogen bonds) further supports the hypothesis that NAT2 is more flexible than NAT1 in the acceptor substrate recognition region spanning residues 125–129 [19].

The corresponding region of *Sr*NAT contains a putative Ac-CoA-binding P-loop motif (Gly¹²⁶-Gly¹²⁷-Gln¹²⁸-Thr¹²⁹) [12] that is probably present in human NAT1 (Gly¹²⁶-Arg¹²⁷-Ser¹²⁸-Tyr¹²⁹). Although the sequences of these putative motifs differ markedly, they may be functionally related. Indeed, the nucleotide-binding capability of P-loops is dependent on their unique three-dimensional structure rather than on their primary sequences [48]. Interestingly, although the putative P-loops sequences of NAT1 and NAT2 differ at positions 127 and 129,

both enzymes have similar true Michaelis constants for Ac-CoA [19].

In conclusion, our homology model of human NAT1 provides an important step towards gaining a better understanding of the contribution that specific residues impart in the conformational flexibility and substrate selectivity of human NATs. In light of the presence of a cysteine-protease-like catalytic triad in *Sr*NAT, this approach may also provide new insights into the identification of structurally related proteins.

We thank Dr R. Krishnamoorthy for helpful advice. We thank the 'Ligue Nationale contre le Cancer' for funding the equipment.

REFERENCES

- 1 Matas, N., Thygesen, P., Stacey, M., Risch, A. and Sim, E. (1997) Mapping AAC1, AAC2 and AACP, the genes for arylamine *N*-acetyltransferases, carcinogen metabolising enzymes on human chromosome 8p22, a region frequently deleted in tumours. *Cytogenet. Cell Genet.* **77**, 290–295
- 2 Grant, D. M., Hughes, N. C., Janezic, S. A., Goodfellow, G. H., Chen, H. J., Gaedigk, A., Yu, V. L. and Grewal, R. (1997) Human acetyltransferase polymorphisms. *Mutat. Res.* **376**, 61–70
- 3 Hein, D. W., Doll, M. A., Fretland, A. J., Leff, M. A., Webb, S. J., Xiao, G. H., Devanaboyina, U. S., Nangju, N. A. and Feng, Y. (2000) Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol. Biomarkers Prev.* **9**, 29–42
- 4 Blum, M., Grant, D. M., McBride, W., Heim, M. and Meyer, U. A. (1990) Human arylamine *N*-acetyltransferase genes: isolation, chromosomal localization, and functional expression. *DNA Cell Biol.* **9**, 193–203
- 5 Sasaki, Y., Ohsako, S. and Deguchi, T. (1991) Molecular and genetic analyses of arylamine *N*-acetyltransferase polymorphism of rabbit liver. *J. Biol. Chem.* **266**, 13243–13250
- 6 Doll, M. A. and Hein, D. W. (1995) Cloning, sequencing and expression of NAT1 and NAT2 encoding genes from rapid and slow acetylator inbred rats. *Pharmacogenetics* **5**, 247–251
- 7 Kelly, S. L. and Sim, E. (1994) Arylamine *N*-acetyltransferase in Balb/c mice: identification of a novel mouse isoenzyme by cloning and expression *in vitro*. *Biochem. J.* **302**, 347–353
- 8 Ferguson, R. J., Doll, M. A., Rustan, T. D. and Hein, D. W. (1996) Cloning, expression, and functional characterization of rapid and slow acetylator polymorphic *N*-acetyltransferase encoding genes of the Syrian hamster. *Pharmacogenetics* **6**, 55–66
- 9 Ohsako, S., Ohtomi, M., Sakamoto, Y., Uyemura, K. and Deguchi, T. (1988) Arylamine *N*-acetyltransferase from chicken liver II. Cloning of cDNA and expression in Chinese hamster ovary cells. *J. Biol. Chem.* **263**, 7534–7538
- 10 Sinclair, J. C., Delgoda, R., Noble, M. E., Jarmin, S., Goh, N. K. and Sim, E. (1998) Purification, characterization, and crystallization of an *N*-hydroxyarylamine *O*-acetyltransferase from *Salmonella typhimurium*. *Protein Expr. Purif.* **12**, 371–380
- 11 Payton, M., Auty, R., Delgoda, R., Everett, M. and Sim, E. (1999) Cloning and characterization of arylamine *N*-acetyltransferase genes from *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*: increased expression results in isoniazid resistance. *J. Bacteriol.* **181**, 1343–1347
- 12 Sinclair, J. C., Sandy, J., Delgoda, R., Sim, E. and Noble, M. E. (2000) Structure of arylamine *N*-acetyltransferase reveals a catalytic triad. *Nat. Struct. Biol.* **7**, 560–564
- 13 Watanabe, M., Sofuni, T. and Nohmi, T. (1992) Involvement of Cys69 residue in the catalytic mechanism of *N*-hydroxyarylamine *O*-acetyltransferase of *Salmonella typhimurium*. Sequence similarity at the amino acid level suggests a common catalytic mechanism of acetyltransferase for *S. typhimurium* and higher organisms. *J. Biol. Chem.* **267**, 8429–8436
- 14 Dupret, J. M. and Grant, D. M. (1992) Site-directed mutagenesis of recombinant human arylamine *N*-acetyltransferase expressed in *Escherichia coli*. Evidence for direct involvement of Cys68 in the catalytic mechanism of polymorphic human NAT2. *J. Biol. Chem.* **267**, 7381–7385
- 15 Deloménie, C., Goodfellow, G. H., Krishnamoorthy, R., Grant, D. M. and Dupret, J. M. (1997) Study of the role of the highly conserved residues Arg-9 and Arg-64 in the catalytic function of human *N*-acetyltransferases NAT1 and NAT2 by site-directed mutagenesis. *Biochem. J.* **323**, 207–215
- 16 Ohsako, S. and Deguchi, T. (1990) Cloning and expression of cDNAs for polymorphic and monomorphic arylamine *N*-acetyltransferases from human liver. *J. Biol. Chem.* **265**, 4630–4634

- 17 Grant, D. M., Blum, M., Beer, M. and Meyer, U. A. (1991) Monomorphic and polymorphic human arylamine *N*-acetyltransferases: a comparison of liver isozymes and expressed products of two cloned genes. *Mol. Pharmacol.* **39**, 184–191
- 18 Dupret, J. M., Goodfellow, G. H., Janezic, S. A. and Grant, D. M. (1994) Structure-function studies of human arylamine *N*-acetyltransferases NAT1 and NAT2. Functional analysis of recombinant NAT1/ NAT2 chimeras expressed in *Escherichia coli*. *J. Biol. Chem.* **269**, 26830–26835
- 19 Goodfellow, G. H., Dupret, J. M. and Grant, D. M. (2000) Identification of amino acids imparting acceptor substrate selectivity to human arylamine acetyltransferases NAT1 and NAT2. *Biochem. J.* **348 Pt 1**, 159–166
- 20 Burley, S. K. (2000) An overview of structural genomics. *Nat. Struct. Biol.* **7** (Suppl.), 932–934
- 21 Thornton, J. M., Todd, A. E., Milburn, D., Borkakoti, N. and Orengo, C. A. (2000) From structure to function: approaches and limitations. *Nat. Struct. Biol.* **7** (Suppl.), 991–994
- 22 Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
- 23 Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402
- 24 Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584–599
- 25 Fischer, D., Barret, C., Bryson, K., Elofsson, A., Godzik, A., Jones, D., Karplus, K. J., Kelley, L. A., MacCallum, R. M. and Pawowski, K. et al. (1999) CAFASP-1: critical assessment of fully automated structure prediction methods. *Proteins Suppl.* **3**, 209–217
- 26 Fischer, D. and Eisenberg, D. (1996) Protein fold recognition using sequence-derived predictions. *Protein Sci.* **5**, 947–955
- 27 Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723
- 28 Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486
- 29 Shindyalov, I. N. and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739–747
- 30 Guoguang, L. (2000) TOP: A new method for protein structure comparisons and similarity searches. *J. Appl. Crystallog.* **33**, 179–183
- 31 Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540
- 32 Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325
- 33 Sim, E., Payton, M., Noble, M. and Minchin, R. (2000) An update on genetic, structural and functional studies of arylamine *N*-acetyltransferases in eucaryotes and procaryotes. *Hum. Mol. Genet.* **9**, 2435–2441
- 34 Bromme, D., Bonneau, P. R., Purisima, E., Lachance, P., Hajnik, S., Thomas, D. Y. and Storer, A. C. (1996) Contribution to activity of histidine-aromatic, amide-aromatic, and aromatic-aromatic interactions in the extended catalytic site of cysteine proteinases. *Biochemistry* **35**, 3970–3979
- 35 Fersht, A. (1999) Structure and mechanism in protein science. A guide to enzyme catalysis and protein folding. W. H. Freeman, New York
- 36 Lee, R. H. (1992) Protein model building using structural homology. *Nature (London)* **356**, 543–544
- 37 Sinclair, J. and Sim, E. (1997) A fragment consisting of the first 204 amino-terminal amino acids of human arylamine *N*-acetyltransferase one (NAT1) and the first transacetylation step of catalysis. *Biochem. Pharmacol.* **53**, 11–16
- 38 Andres, H. H., Klem, A. J., Schopfer, L. M., Harrison, J. K. and Weber, W. W. (1988) On the active site of liver acetyl-CoA. Arylamine *N*-acetyltransferase from rapid acetylator rabbits. *J. Biol. Chem.* **263**, 7521–7527
- 39 Watanabe, M., Igarashi, T., Kaminuma, T., Sofuni, T. and Nohmi, T. (1994) *N*-hydroxyarylamine *O*-acetyltransferase of *Salmonella typhimurium*: proposal for a common catalytic mechanism of arylamine acetyltransferase enzymes. *Environ. Health Perspect.* **102** (Suppl. 6), 83–89
- 40 Rawlings, N. D. and Barrett, A. J. (1994) Families of cysteine peptidases. *Methods Enzymol.* **244**, 461–486
- 41 Babbitt, P. C. and Gerlt, J. A. (1997) Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272**, 30591–30594
- 42 Li, A. X. and Steffens, J. C. (2000) An acyltransferase catalyzing the formation of diacylglycerol is a serine carboxypeptidase-like protein. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6902–6907
- 42a Deloménie, C., Fouix, S., Longuemaux, S., Brahimi, N., Bizet, C., Picard, B., Denamur, E. and Dupret, J. M. (2001) Identification and functional characterization of arylamine *N*-acetyltransferases in eubacteria: evidence for highly selective acetylation of 5-aminosalicylic acid. *J. Bacteriol.*, in the press
- 43 Peters, G. H., Iversen, L. F., Branner, S., Andersen, H. S., Mortensen, S. B., Olsen, O. H., Moller, K. B. and Moller, N. P. (2000) Residue 259 is a key determinant of substrate specificity of protein- tyrosine phosphatases 1B and alpha. *J. Biol. Chem.* **275**, 18201–18209
- 44 Carriere, F., Withers-Martinez, C., van Tilbeurgh, H., Roussel, A., Cambillau, C. and Verger, R. (1998) Structural basis for the substrate selectivity of pancreatic lipases and some related proteins. *Biochim. Biophys. Acta* **1376**, 417–432
- 45 Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D. and Hill, C. P. (1997) Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 Å resolution. *EMBO J.* **16**, 3787–3796
- 46 Lesk, A. M. (1991) Protein architecture. A practical approach, Oxford University Press, Oxford
- 47 Okoniewska, M., Tanaka, T. and Yada, R. Y. (2000) The pepsin residue glycine-76 contributes to active-site loop flexibility and participates in catalysis. *Biochem. J.* **349**, 169–177
- 48 Via, A., Ferre, F., Brannetti, B., Valencia, A. and Helmer-Citterich, M. (2000) Three-dimensional view of the surface motif associated with the P-loop structure: cis and trans cases of convergent evolution. *J. Mol. Biol.* **303**, 455–465

Received 19 December 2000/7 February 2001; accepted 8 March 2001