

DMD #15040

Computational and Experimental Analyses of Mammalian Arylamine *N*-acetyltransferase Structure and Function

Jason M. Walraven, John O. Trent, and David W. Hein

Departments of Pharmacology & Toxicology and Medicine, James Graham Brown Cancer
Center, University of Louisville School of Medicine, Louisville, KY

DMD #15040

Running Title Page

a) Mammalian NAT Structure and Function

b) Corresponding author: David W. Hein, Ph.D., Department of Pharmacology and Toxicology,
University of Louisville School of Medicine, Louisville, KY 40292, Phone: (502) 852-5141;
Fax: (502) 852-7868; Email: d.hein@louisville.edu

c) The number of text pages: 27
 The number of tables: 1
 The number of figures: 5
 The number of references: 40
 The number of words in the Abstract: 220
 The number of words in the Introduction: 750
 The number of words in the Discussion: 1095

d) Non-standard Abbreviations Used in this Paper:

 Arylamine *N*-acetyltransferase (NAT); single nucleotide polymorphism (SNP);
sulfamethazine (SMZ); polymerase chain reaction (PCR)

DMD #15040

ABSTRACT

Arylamine *N*-acetyltransferases (NATs) play an important role in the metabolism of arylamine and hydrazine drugs and many arylamine pro-carcinogens. The two human *N*-acetyltransferases, *NAT1* and *NAT2*, are widely distributed in human tissues and highly polymorphic. While many xenobiotic procarcinogens and drugs are known mammalian NAT substrates, it is unclear what physiological roles these enzymes might play, what endogenous substrates they primarily act upon, or the mechanisms underlying the functional effects of specific NAT gene coding region SNPs. Analyses of mammalian NAT protein structures can greatly help to answer these questions. Homology modeling techniques can be used to approximate mammalian NAT structures using known bacterial NAT crystal structures as templates. In comparison to the bacterial template NATs used for homology modeling, mammalian NATs have a 17 residue insert of unknown structure and function. Homology modeling analyses yielded two different alignments (Modeller 8v1 or 3DCoffee algorithms) that placed this insert in two likely alternative locations. Secondary structure prediction techniques and experimental analyses of a series of human *NAT2* mutants with artificial deletions/replacements of the insert region distinguished one of these alternatives as the most likely insert location and provided a better understanding of its structure and function. This study demonstrates both the utility and limitations of computational structural modeling with proteins that differ as much as the mammalian and bacterial NATs.

DMD #15040

INTRODUCTION

Arylamine *N*-acetyltransferases (NATs; EC 2.3.1.5) catalyze the *N*-acetylation of arylamines and hydrazines and *O*-acetylation of *N*-hydroxy-arylamines and heterocyclic amines. These reactions are important for the activation and deactivation of exocyclic amine-containing pro-carcinogens, and for the metabolism of some pharmaceutical drugs (Weber and Hein, 1985). In a ping-pong bi-bi reaction mechanism, the enzyme first acetylates the active site cysteine using acetyl-coenzyme A, and then transfers the acetyl group to the substrate's exocyclic nitrogen (*N*-acetylation) or the oxygen of its oxidized exocyclic nitrogen (*O*-acetylation) (Hein, 1988). While *N*-acetylation is typically considered a deactivation step, *O*-acetylation is an activation step, resulting in the formation of reactive arylnitrenium species that can react with DNA to form adducts (Hanna, 1996). Since human NAT genes are highly polymorphic, it is important to understand how different NAT genotypes alter *N*-acetylation phenotypes, thereby influencing cancer susceptibilities and pharmaceutical drug toxicities (Hein et al., 2000).

Expression of human NAT1 and NAT2 is widely distributed throughout body tissues (Windmill et al., 2000). While many xenobiotic NAT substrates have been discovered, only one potential endogenous substrate, *p*-aminobenzoylglutamate, has been discovered (Minchin, 1995). Therefore, it is highly probable that other yet unknown endogenous NAT substrates and physiological roles exist. Discovery of potential substrates and useful inhibitors is limited by our knowledge of mammalian NAT enzyme structure.

To-date no publication has described the crystal structure of mammalian NATs. For this reason, investigators have utilized homology modeling techniques to gain clues about the key structural characteristics of mammalian NATs, including the location and potential

DMD #15040

mechanism(s) of SNP-induced mutations, and the shape of the active site for computational docking of NAT substrates (Rodrigues-Lima et al., 2001; Rodrigues-Lima and Dupret, 2002; Kawamura et al., 2005; Lau et al., 2006). Quality mammalian NAT structures could be used to rationalize experimental data and screen for potential substrates and inhibitors, potentially pointing to an endogenous physiological role. There are limitations to homology modeling techniques, however, since they sometimes rely on incomplete structural template information.

Bacterial NAT crystal structures were used as templates to create the aforementioned mammalian NAT homology models. Bacterial NAT crystal structures have been solved for *Salmonella typhimurium*, *Pseudomonas aeruginosa*, *Mycobacterium smegmatis*, and *Mesorhizobium loti*, and can be found with Protein Data Bank accession numbers 1E2T, 1W4T, 1GX3, and 2BSZ, respectively (Sinclair et al., 2000; Sandy et al., 2002; Westwood et al., 2005; Holton et al., 2005). Alignment scores between bacterial and mammalian NAT protein sequences, and among the bacterial NAT sequences, are relatively low (~ 30%). Aligning mammalian and bacterial NAT protein sequences reveals important differences between the bacterial and mammalian NAT protein structures that could limit the utility of the mammalian NAT homology models.

Several studies describe what appears to be an inserted sequence in the second domain of mammalian NATs when aligned with bacterial NATs (Payton et al., 2001; Kawamura et al., 2005; Lau et al., 2006). Although this insert is sometimes referred to as a loop or coil because of its unstructured appearance in homology models, the lack of a template structure for this region is immediately responsible for these designations. Homology modeling can only generate structure for a target protein sequence if a corresponding template structure sequence is provided. Previously reported studies modeled mammalian NATs with this insert adjacent to the active site

DMD #15040

pocket (Kawamura et al., 2005; Savulescu et al., 2005; Lau et al., 2006), while other studies place it in a different location (Payton et al., 2001) or do not recognize its existence at all (Rodrigues-Lima et al., 2001).

It is unknown whether this insert has defined secondary structure and is integrated into the second domain beta barrel, or if it takes the form of an unstructured loop or coil. In the latter case, a loop adjacent to the active site pocket could affect the size, shape, and/or accessibility of the active site pocket. In addition, since the location of the loop in a homology model is determined by the placement of the insert in the alignment, as is demonstrated by the differences in previously reported alignments (Payton et al., 2001; Rodrigues-Lima et al., 2001; Savulescu et al., 2005; Lau et al., 2006), it is possible that a slightly different or more optimal alignment could place the insert in a different location.

Our homology modeling analyses yielded two different, and apparently equally valid, alignments that placed this insert in two different locations. We used secondary structure prediction programs and designed a series of human NAT2 deletion mutants to positively define the insert location and better understand its structure and function.

DMD #15040

METHODS

Homology Modeling

Human NAT2 and bacterial NAT sequences from *Salmonella typhimurium*, *Pseudomonas aeruginosa*, and *Mesorhizobium loti* were aligned using the Modeller 8v1 auto align function and the 3DCoffee alignment algorithm (Sali and Blundell, 1993; O'Sullivan et al., 2004). The known crystal structures for *Salmonella typhimurium*, *Pseudomonas aeruginosa*, and *Mesorhizobium loti* NATs (PDB accession numbers 1E2T, 1W4T, 2BSZ) were used as template structures for creating homology models of human NAT2. The Modeller and 3DCoffee alignments, with bacterial NAT template structural information, were used to generate homology models by satisfaction of spatial restraints using the program Modeller 8v1 (Sali and Blundell, 1993).

Secondary Structure Prediction

N-acetyltransferase protein sequences from *Salmonella typhimurium*, *Pseudomonas aeruginosa*, *Mesorhizobium loti*, and human NAT2 were evaluated using twelve secondary structure prediction programs: APSSP2, PROF king, PROF sec, PORTER, JPRED, JUFO, SCRATCH, SABLE1, SABLE2, SAM-T99, PSIPred, and YASPIN (McClelland and Rumelhart, 1988; Kneller et al., 1990; Rost, 1996; Karplus et al., 1998; Cuff and Barton, 2000; McGuffin et al., 2000; Ouali and King, 2000; Meiler et al., 2001; Raghava, 2002; Adamczak et al., 2004; Adamczak et al., 2005; Lin et al., 2005; Pollastri and McLysaght, 2005). Secondary structure predictions were then aligned based on Modeller or 3DCoffee protein sequence alignments.

DMD #15040

Human NAT2 Mutation

A 17 residue segment was deleted from human NAT2, at 157-173 or 167-183 based on Modeller and 3DCoffee alignment results, respectively. Residues immediately surrounding the deletion site were replaced by beta turn-facilitating residues in order to prevent disruption of the beta barrel secondary structure. Eight different deletion mutants were created. For the Modeller-based deletion, turns QEGST, PHAG, ANDH, and SDGSD were inserted on either side of the deletion in four different deletion mutants. For the 3DCoffee-based deletion, turns HHEH, GRSG, GG, and GGSG were inserted on either side of the deletion in four different deletion mutants. In both sets of mutants, the first three turns match turn residues found in the same location in structures 1E2T, 1W4T, 2BSZ, respectively (Figure 1), and the fourth turn was made entirely of “turn promoting” residues (Paul and Rosenbusch, 1985). Codons chosen for the turn residues were based on the codon usage table for *Homo sapiens* (www.kazusa.or.jp/codon).

The human NAT2 nucleotide sequence was mutated by overlap extension (Horton et al., 1989). Human NAT2 coding region sequences on either side of the targeted deletion were amplified independently by polymerase chain reaction (PCR) from a prokaryotic expression vector pKK223-3 containing human NAT2 (Ferguson et al., 1994), using plasmid-specific and human NAT2-specific primers (Table 1). Each set of NAT2-specific primers contained complementary nucleotide sequences coding for the replacement turn residues. For each mutant, the two separate PCR products were combined, denatured at 98°C, and the complementary sequences allowed to anneal at 30°C, followed by overlap extension using nested forward and reverse human NAT2-specific PCR primers. The forward and reverse nested primers introduced restriction sites EcoR1 and HindIII, respectively. Independent duplicates of each mutant were cloned and expressed in parallel as cloning and expression controls.

DMD #15040

A restriction fragment length analysis using restriction enzymes KpnI and ApoI was used to verify the successful removal of the 17 residue segment in the Modeller-based and 3DCoffee-based deletion mutants, respectively, prior to cloning. Successful deletion of the Modeller-based insert changed the KpnI digestion pattern from one to zero cuts. Successful deletion of the 3DCoffee-based insert changed the ApoI digestion pattern from two to one cut. Digestion products were resolved on a 1% agarose gel.

Cloning, Sequencing, and Expression

Independent duplicates of the reference human NAT2 coding region were cloned and expressed in parallel with the independent duplicates of each mutant. All were directionally cloned into pKK223-3 bacterial expression vector (Pharmacia-LKB Biotechnology, Piscataway, NJ) using EcoRI and HindIII restriction sites as described previously (Doll and Hein, 1995).

The expected nucleotide sequence was confirmed by automated sequencing of sense and antisense strands as previously described (Walraven et al., 2006) using human NAT2 specific sequencing primers (Table 1), prior to transformation and expression in JM105 strain *Escherichia coli* as previously described (Chung and Miller, 1988; Chung et al., 1989; Doll and Hein, 1995). Bacteria were lysed by sonication in homogenization buffer as previously described (Doll and Hein 1995), and total bacterial lysate protein concentration was determined for each expression (Bradford, 1976).

Enzyme Activity and Stability

As a measure of NAT2 enzyme activity, sulfamethazine (SMZ) N-acetylation assays were performed by incubating bacterial lysate, 1 mM acetyl-coenzyme A, and 300 μ M SMZ at

DMD #15040

37°C for 10 minutes under linear rate conditions, and the reaction was stopped by adding 1/10 volume of 1 M acetic acid. Control reactions substituted water for acetyl-coenzyme A. Acetylated SMZ product was separated and measured using a Beckman System Gold high-performance liquid chromatography system as described previously (Leff et al, 1999). Sulfamethazine is a NAT2-selective substrate (Grant et al., 1991).

Thermostability was measured by incubating the bacterial lysate (1 mg/ml) at 30°-50°C for 10 minutes, immediately followed by assays for SMZ N-acetylation as described above. Results were normalized to activity of untreated lysate to give the percentage of activity remaining.

DMD #15040

RESULTS

Homology Modeling

Alignments between bacterial template NATs and human NAT2 revealed a 17 residue segment, or insert, in human NAT2 that has no corresponding bacterial template sequence. Modeller and 3DCoffee alignment algorithms placed this 17 residue insert in different locations, 157-173 (GIWYLDQIRREQYITNK) and 167-183 (EQYITNKEFLNSHLLPK), respectively.

The NAT2 homology models were superimposable with the bacterial template structures, with the Cys68-His107-Asp122 catalytic triad positioned as expected in the active site pocket. Due to a lack of structural template information, the 17 residue insert appeared as loops in the second domain of human NAT2. In the models, the Modeller alignment-based loop was located one beta strand away from the 3DCoffee-based loop and away from the active site, whereas the 3DCoffee alignment-based loop was adjacent to the active site pocket (Figure 2).

Since these homology models were not used for detailed evaluation of specific residues or for substrate docking, an analysis of the structural quality was not necessary. Such detailed evaluations would require molecular dynamics simulations for structural optimization. The Modeller program generates models of sufficient quality for the immediate purposes of this study.

Secondary Structure Prediction

Bacterial NAT secondary structure predictions were very similar to their actual secondary structures as defined on the RCSB Protein Data Bank website (www.rcsb.org), thus providing confidence that the human NAT2 predictions also were accurate. The predicted bacterial NAT

DMD #15040

and human NAT secondary structures were strikingly similar, with nearly all major secondary structures shared (Figure 3). The individual prediction programs gave slightly different variations for each sequence, but the prediction programs were overall very consistent.

When the bacterial NAT and human NAT2 predictions were aligned based on the Modeller alignment, the 17 residue human NAT2 insert was located in the same place as a predicted beta strand. In addition, a bacterial NAT beta strand on the c-terminal side of the insert did not match up with a human NAT2 beta strand. This apparent mismatch was corrected when the prediction results were aligned based on the 3DCoffee alignment. In this case, the 17 residue insert did not have any predicted secondary structure and the surrounding predicted secondary structures matched with the bacterial NAT secondary structures.

Another important observation was made regarding the C-terminal tail. The bacterial NAT C-terminal helix found in all four known bacterial NAT crystal structures was not predicted for the human NAT2 sequence by any of the prediction programs.

Human NAT2 Deletion Mutants

Based on restriction fragment length analyses and complete sequencing results, the eight different human NAT2 deletion mutants were successfully constructed and cloned into bacterial expression vector pKK223-3. Duplicate clones of the reference and deletion mutant human NAT2 provided highly reproducible results. This validates the cloning process and expression system used in this study as highly consistent, as demonstrated previously (Walraven et al., 2006).

Enzyme activity results (Figure 4) clearly demonstrated that deletion of the Modeller alignment insert resulted in complete loss of enzyme activity, regardless of the turn used to

DMD #15040

replace the residues surrounding the deletion. On the other hand, all of the 3DCoffee deletion mutants had measurable N-acetylation activity. The activities for these mutants were highly variable, with one turn variant exhibiting SMZ NAT activity nearly equivalent to the reference activity.

Thermostability measurements for the reference human NAT2 and the highest NAT2-activity 3DCoffee deletion mutant revealed that the deletion mutant was much less stable than reference human NAT2 (Figure 5). The reference NAT2 enzyme had only slightly reduced activity at the higher temperatures, whereas the deletion mutant began losing activity at lower temperatures and completely lost activity at the higher temperatures. One of the deletion mutants with lower activity was also tested for thermostability, but no SMZ NAT2 activity was detectable after incubation at 32°C (data not shown).

Deleting the 17 residue insert resulted in decreased activity due at least in part to reduced stability. The choice of turn residues was apparently also critical to the mutant enzyme's stability. The mutant with the highest activity was also the mutant with the shortest turn substituted in for the segment, which was composed entirely of glycines (GG).

DMD #15040

DISCUSSION

Previous studies have used bacterial template structures to create mammalian NAT homology models (Rodrigues-Lima et al., 2001; Rodrigues-Lima and Dupret, 2002; Kawamura et al., 2005; Lau et al., 2006). However, since the generation of a homology model is not the primary focus of this study, many of the structural details discussed in those studies will not be revisited here. This study focused on the undefined and not well understood mammalian NAT insert. Since homology models will conform to the template structure(s) provided, the regions that lack a template require more attention. This is especially true when the regions in question could influence the active site pocket's shape and substrate accessibility.

Previous studies have generated alignments that placed the insert in different locations. Lau et al. (2006) placed the insert residues at either 172-188 or 170-186 (in the alignment, QP is repeated on both sides of the insert) which is modeled as a loop in the second domain, adjacent to the active site pocket. The alignment published by Savulescu et al. (2005) places the insert in the same location, and Kawamura et al. (2005) reported models with a loop also in the second domain, adjacent to the active site. These three studies placed the loop close to the 3DCoffee-based insert (167-183) described in our study. Payton et al. (2001) published an alignment with the insert located at 156-172, which is nearly identical to the Modeller-based insert (157-173) described in our study.

The results of the present study clearly identify the correct insert location and confirm that the insert is a loop. The loop is not integrated into the secondary structure of the protein and has no defined secondary structure. Since bacterial NATs function well without an insert, it is conceivable that deleting the loop in a way that facilitated the folding of the beta barrel at the

DMD #15040

deletion site could result in active protein, probably with slightly altered characteristics. On the other hand, it is much less likely that deleting the wrong insert location would yield an active protein, since it would remove a significant portion of the surrounding beta barrel secondary structure in the second domain. If the insert were not a loop, but were integrated into the secondary structure of the protein, both deletions would significantly alter the fold of the protein or prevent it from folding altogether.

When the 3DCoffee-assigned insert (167-183) was deleted, the enzyme remained active, though with altered characteristics. On the other hand, deletion of the Modeller-assigned insert (157-173) resulted in undetectable activity. These results support the conclusion that residues 167-183 are not only the correct insert location, but they form a loop that does not play a critical role in the integrity of the beta barrel secondary structure. Deleting residues 157-173 probably removed a portion of the second domain beta barrel and altered the folding and/or function of the protein. For the purposes of this study, it is not necessary to know whether the loss of activity was due to misfolding of the protein, altered protein function, or a combination of both.

Secondary structure predictions supported and confirmed these conclusions. According to the alignments in Figure 3, deletion of the Modeller-assigned insert resulted also in the deletion of a beta strand in the second domain beta barrel. This is consistent with the loss of activity seen for the Modeller alignment-based mutants. In addition, no secondary structure was predicted for the 3DCoffee-assigned insert, which is consistent with the conclusion that the insert must be a loop or coil with no defined secondary structure, and not integrated into the surrounding beta barrel. These results are also consistent with our hypothesis that the mammalian NAT proteins can function without the loop.

The secondary structure prediction that the bacterial NAT C-terminal helix is not shared

DMD #15040

by human NAT2 suggests that the C-terminal tail is free to interact with the active site. The C-terminal helix is likewise not predicted for human NAT1 or any of the rat Nats (unpublished data). This would be consistent with previous reports that the 10 C-terminal residues are critical for control of acetyl-coenzyme A hydrolysis (Mushtaq et al., 2002). Since acetyl-coenzyme A hydrolysis takes place in the active site, it is reasonable to assume that the C-terminal tail must interact with the active site pocket.

Although the loop is not required for NAT catalytic activity, the reduced stability of loop-less human NAT2 suggests that this loop may serve a role in stabilizing the protein's structure. How the loop might accomplish this task is beyond the scope of homology modeling. Our models suggest that the loop could also play a role in active site access since it is adjacent to the active site and flexible enough to extend over the bacterial NAT-like active site cavity. Combined with the possibility that the C-terminal tail interacts with the active site, the active site pocket could look much different than its representation in current homology models.

After our experiments were completed, the crystal structure for a mutant human NAT1 protein was deposited into the RCSB Protein Data Bank under the code 2IJA. Although the publication describing this structure has not yet been released, it substantiates the findings of our human NAT2 study. The overall structure of our NAT2 model compares favorably to 2IJA, with a root mean square deviation of 1.4 Å between the two structures. In addition, the NAT2 loop location as determined by our study (167-183) is nearly identical to the 2IJA loop residues (165-185). Two structural differences between 2IJA and our model are the association of the loop with a beta strand in domain three, and the insertion of the C-terminal tail adjacent to the active site pocket. These interesting structural features have a dramatic impact on the size and shape of the active site pocket and rationalize the influence of the C-terminal tail on acetyl-coenzyme A

DMD #15040

hydrolysis. These two major differences also support our study's finding that the loop may have a stabilizing effect on the protein's structure, and make sense of the predicted lack of secondary structure in the C-terminal NAT2 sequence. More information is needed about the flexibility of these important structural features in solution.

This study demonstrates the strength of combining computational and experimental approaches to determining protein structure and function. Based on the initial success of mammalian crystallization attempts (2IIA), it might be expected that the results of this study will be confirmed by the eventual crystallization of human NAT2. This study both identified the location and potential function of the mammalian NAT loop, and contributes to our understanding of arylamine N-acetyltransferase structure and function.

DMD #15040

ACKNOWLEDGMENTS

We thank Dr. David Barker for his technical assistance and helpful advice.

DMD #15040

REFERENCES

- Adamczak R, Porollo A, and Meller J (2005) Combining Prediction of Secondary Structure and Solvent Accessibility in Proteins. *Proteins* 59:467-75.
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248-254.
- Cheng J, Randall A, Sweredoski M, and Baldi P (2005) SCRATCH: a Protein Structure and Structural Feature Prediction Server. *Nucleic Acids Research* 33:72-76.
- Chung CT and Miller RH (1988) A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Res* 16:3580.
- Chung CT, Niemela SL, and Miller RH (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Sci USA* 86:2172-2175.
- Cuff JA and Barton GJ (2000) Application of Multiple Sequence Alignment Profiles to Improve Protein Secondary Structure Prediction. *Proteins* 40:502-511.
- Doll MA and Hein DW (1995) Cloning, sequencing and expression of NAT1 and NAT2 encoding genes from rapid and slow acetylators inbred rats. *Pharmacogenetics* 5:502-505.
- Ferguson RJ, Doll MA, Rustan TD, Gray K, Hein DW (1994) Cloning, expression, and functional characterization of two mutant (NAT2(191) and NAT2(341/803)) and wild-type human polymorphic *N*-acetyltransferase (NAT2) alleles. *Drug Metab Dispos* 22(3):371-376.
- Grant DM, Blum M, Beer M, Meyer UA (1991) Monomorphic and polymorphic human arylamine *N*-acetyltransferases: a comparison of liver isozymes and expressed products

DMD #15040

- of two cloned genes. *Mol Pharmacol* 39(2):184-191.
- Hanna PE (1996) Metabolic activation and detoxification of arylamines. *Curr Med Chem* 3:195-210.
- Hein DW (1988) Acetylator genotype and arylamine-induced carcinogenesis. *Biochim Biophys Acta* 948:37-66.
- Hein DW, Doll MA, Fretland AJ, Leff MA, Webb SJ, Xiao GH, Devanaboyina US, Nangju NA, and Feng Y (2000) Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. *Cancer Epidemiol Biomarkers Prev* 9(1):29-42.
- Holton SJ, Dairou J, Sandy J, Rodrigues-Lima F, Dupret J-M, Noble ME, and Sim E (2005) Structure of Mesorhizobium Loti Arylamine N-Acetyltransferase 1. *Acta Crystallograph Sect F Struct Biol Cryst Commun* 61:14-16.
- Horton RM, Hunt HD, Ho SN Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* 77(1):61-68.
- Karplus K, Barret C, and Hughey R (1998) Hidden (Markov) models for detecting remote protein homologies. *Bioinformatics* 14(10):846-856.
- Kawamura A, Graham J, Mushtaq A, Tsiftoglou SA, Vath GM, Hanna PE, Wagner CR, and Sim E (2005) Eukaryotic arylamine N-acetyltransferase investigation of substrate specificity by high-throughput screening. *Biochem Pharmacol* 69: 347-359.
- Kneller DG, Cohen FE and Langridge R (1990) Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network. *J Mol Biol* 214:171-182.
- Leff MA, Epstein PN, Doll MA, Fretland AJ, Devanaboyina US, Rustan TD, and Hein DW (1999) Prostate-specific human N-acetyltransferase 2 (NAT2) expression in the mouse. *J Pharmacol Exp Ther* 290:182-187.

DMD #15040

- Lin K, Simossis VA, Taylor WR, and Heringa J (2005) A Simple and Fast Secondary Structure Prediction Algorithm using Hidden Neural Networks. *Bioinformatics* 21(2):152-9.
- McClelland JL and Rumelhart DE (1988) *Explorations in Parallel Distributed Processing*. 3:318-362. MIT Press, Cambridge MA.
- McGuffin LJ, Bryson K, and Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics*. 16:404-405.
- Meiler J, Muller M, Zeidler A, and Schmaschke F (2001) Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks. *J Mol Model* 7(9):360-369.
- Minchin RF (1995) Acetylation of *para*-aminobenzoylglutamate, a folate catabolite, by recombinant human NAT and U937 cells. *Biochem J* 307:1-3.
- Mushtaq A, Payton M, Sim E (2002) The COOH terminus of arylamine *N*-acetyltransferase from *Salmonella typhimurium* controls enzymatic activity. *J Biol Chem* 277:12175-12181.
- O'Sullivan O, Suhre K, Abergel C, Higgins DG, and Notredame C (2004) 3DCoffee: Combining protein sequences and structures within multiple sequence alignments. *J Mol Biol* 340:385-395.
- Ouali M, and King RD (2000) Cascaded multiple classifiers for secondary structure prediction. *Prot Sci* 9:1162-1176.
- Paul C, Rosenbusch JP (1985) Folding patterns of porin and bacteriorhodopsin. *EMBO J* 4(6):1593-1597.
- Payton M, Mushtaq A, Yu T-W, Wu L-J, Sinclair J, and Sim E (2001) Eubacterial arylamine *N*-acetyltransferase identification and comparison of 18 members of the protein family with

DMD #15040

- conserved active site cysteine, histidine, and aspartate residues. *Microbiology* 147:1137-1147.
- Pollastri G and McLysaght A (2005) Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 21(8):1719-20.
- Raghava GPS (2002) APSSP2 : A combination method for protein secondary structure prediction based on neural network and example based learning. CASP5. A-132.
- Rodrigues-Lima F, Delomenie C, Goodfellow GH, Grant DM, and Dupret J-M (2001) Homology modelling and structural analysis of human arylamine *N*-acetyltransferase NAT1: Evidence of the conservation of a cysteine protease catalytic domain and an active site loop. *Biochem J* 356:327-334.
- Rodrigues-Lima F and Dupret J-M (2002) 3D model of human arylamine *N*-acetyltransferase 2: Structural basis of the slow acetylator phenotype of the R64Q variant and analysis of the active-site loop. *Biochem Biophys Res Commun* 291:116-123.
- Rost B (1996) PHD: predicting one-dimensional protein structure by profile based neural networks. *Meth Enzymol* 266:525-539.
- Sali A and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779-815.
- Sandy J, Mushtaq A, Kawamura A, Sinclair J, Sim E, and Noble M (2002) The structure of arylamine *N*-acetyltransferase from *Mycobacterium smegmatis* – an enzyme which inactivates the anti-tubercular drug, isoniazid. *J Mol Biol* 318:1071-1083.
- Savulescu MR, Mushtaq A, and Josephy PD (2005) Screening and characterizing human NAT2 variants. *Methods Enzymol* 400:192-215.
- Sinclair JC, Sandy J, Delgoda R, Sim E, and Noble ME (2000) Structure of arylamine *N*-

DMD #15040

acetyltransferase reveals a catalytic triad. *Nat Struct Biol* 7:560-564.

Weber WW and Hein DW (1985) N-acetylation pharmacogenetics. *Pharmacol Rev* 37:25-27.

Westwood IM, Holton SJ, Rodrigues-Lima F, Dupret J-M, Bhakta S, Noble ME, and Sim E
(2005) Expression, purification, characterization and structure of *Pseudomonas aeruginosa* arylamine N-Acetyltransferase. *Biochem J* 385:605-612.

Windmill KF, Gaedigk A, Hall PM, Samaratunga H, Grant DM, and McManus ME (2000)
Localization of N-acetyltransferases NAT1 and NAT2 in human tissue. *Toxicol Sci*
54:19-29.

DMD #15040

FOOTNOTES

This work was partially supported by U.S. Public Health Service grant CA34627 from the National Cancer Institute and training grant ES011564 from the National Institute of Environmental Health Sciences. This work constitutes partial fulfillment by Jason Walraven for the Ph.D. in Pharmacology and Toxicology at the University of Louisville School of Medicine.

Send reprint requests to: David W. Hein, Ph.D., Department of Pharmacology and Toxicology, University of Louisville School of Medicine, Louisville, KY 40292. E-mail: d.hein@louisville.edu

DMD #15040

FIGURE LEGENDS

Figure 1. Bacterial template structure beta-barrel turn residues. Residues 146-179 are shown for all three bacterial NATs, with beta sheet residues indicated by arrows and coil/turn residues by straight lines. The first underlined segment on each line corresponds to the turn used for a Modeller-based mutant, and the second for a 3DCoffee-based mutant. This figure was adapted from information on the Protein Data Bank website (www.rcsb.org; accession numbers 1E2T, 1W4T, and 2BSZ).

Figure 2. Homology models of human NAT2, and the *Mesorhizobium loti* NAT1 bacterial template structure (shown for reference). The Modeller alignment produces a second domain loop away from the active site, whereas the 3DCoffee alignment produces a second domain loop adjacent to the active site pocket. Asterisks identify the active site loop, and the residue numbers of each insert are shown.

Figure 3. Secondary structure prediction results were aligned based on Modeller and 3DCoffee alignments. This figure is designed to give an *overall summary view* of how prediction programs assigned secondary structure and how the bacterial NAT and human NAT2 secondary structures align when the Modeller and 3DCoffee alignments are applied. Results from twelve secondary structure prediction programs are listed under each NAT sequence. Actual secondary structure is shown immediately beneath each bacterial NAT sequence. Blue “E” characters represent beta sheets, and red “H” characters represent helices. A green bar spans the insert residues according to each alignment. This figure is also provided as supplementary data in a more legible format.

DMD #15040

Figure 4. SMZ *N*-acetyltransferase activities of reference and mutant human NAT2. Mutants are labeled according to the turn sequence used. Duplicate clones are indicated with “1” and “2” above the sequence for each turn. Activity is nanomoles of product per minute reaction time per milligram of total protein. Activity was not detected for substrates marked ND (limit of detection = 0.06 nmoles/min/mg). Error bars indicate standard error (n = 3).

Figure 5. Thermostability of reference and mutant human NAT2. Percentage of SMZ activity remaining plotted on the ordinate versus 10 minute incubation temperature on the abscissa. Error bars indicate standard error (n = 3).

DMD #15040

Table 1

Primers Used For Mutant Construction and Cloning

Name	Sequence	Features
m-h2-1e2tF	5' - <u>CAGGAGGGCAGCACCC</u> TTAATTCTCATCTCCTGCC-3', ^b	QEGST peptide
m-h2-1e2tR	5' - <u>GGTGCTGCCCTCCTGTGT</u> CAAGCAGAAAATGCA-3', ^b	QEGST rc peptide
m-h2-1w4tF	5' - <u>AGACCCACGCCGGCTTTCT</u> TAATTCTCATCTCCTG-3', ^b	PHAG peptide
m-h2-1w4tR	5' - <u>AAAGCCGGCGTGGGGTCTCT</u> TCTGTCAAGCAGA-3', ^b	PHAG rc peptide
m-h2-2bszF	5' - <u>GACAGAAGCCAACGACC</u> CTTAATTCTCATCTCCTGCC-3', ^b	ANDH peptide
m-h2-2bszR	5' - <u>AAGGTGGTCGTTGGCTTCTGT</u> CAAGCAGAAAATGC-3', ^b	ANDH rc peptide
m-h2-CustF	5' - <u>ACAAGCGACGCCAGCGAC</u> CTTAATTCTCATCTCCTGCC-3', ^b	SDGSD peptide
m-h2-CustR	5' - <u>GTCGCTGGCGTCGCTTGT</u> CAAGCAGAAAATGCA-3', ^b	SDGSD rc peptide
3d-h2-1e2tF	5' - <u>ATCCACCACGAGCACCACCA</u> AAAAATATACTTATTTACGC-3', ^b	HHEH peptide
3d-h2-1e2tR	5' - <u>GTGGTGCTCGTGGTGGATT</u> TGGTCCAGGTACCAG-3', ^b	HHEH rc peptide
3d-h2-1w4tF	5' - <u>ATCGGCAGAAAGCGGCCACCA</u> AAAAATATACTTATTTACGC-3', ^b	GRSG peptide
3d-h2-1w4tR	5' - <u>GTGCCGCTTCTGCCGATT</u> TGGTCCAGGTACCAG-3', ^b	GRSG rc peptide
3d-h2-2bszF	5' - <u>ATCAGGGGCGGCAAAACCA</u> AAAAATATACTTATTTACG-3', ^b	GG peptide
3d-h2-2bszR	5' - <u>GTGTTTGCCGCCCTGATT</u> TGGTCCAGGTAC-3', ^b	GG rc peptide
3d-h2-CustF	5' - <u>ATCGGCGGCAGCGGCCACCA</u> AAAAATATACTTATTTACGC-3', ^b	GGSG peptide
3d-h2-CustR	5' - <u>GCCGCTGCCGCCGATT</u> TGGTCCAGGTACCA-3', ^b	GGSG rc peptide
hNat2EcoRI-F	5' - <u>ATGCGAGAATTTCAT</u> GGACATTGAAGCATATTTTGAAAGAATT-3', ^a	PCR, EcoRI
hNat2HindIII-R	5' - <u>ATGCGAAAGCTTAAGGGTTTAT</u> TTTGTTCCTTATTCTAAAT-3', ^a	PCR, HindIII
pKK-F	5' - TTATCAGACCGCTTCTGCGT-3'	PCR, Fwd-Seq
h2-Seq1F	5' - CCGGGGTGGGTGGTGTCT-3'	Fwd-Seq
h2-Seq2F	5' - CACCTTCTCCTGCAGGTGACCAT-3'	Fwd-Seq
h2-Seq3F	5' - TTTGGTGGGCTTCATCCTCA-3'	Fwd-Seq
pKK-R	5' - GCTGTTGACAATTAATCATCGG-3'	PCR, Rev-Seq
h2-Seq1R	5' - CAACCTCTTCCTCAGTGAGAG-3'	Rev-Seq
h2-Seq2R	5' - ACATCTGGGAGGAGCTTCCA-3'	Rev-Seq
h2-Seq3R	5' - TGACCTGGAGACACCACCA-3'	Rev-Seq

^a Underline indicates restriction site ^b Underline indicates turn residue codons (rc = reverse complement)




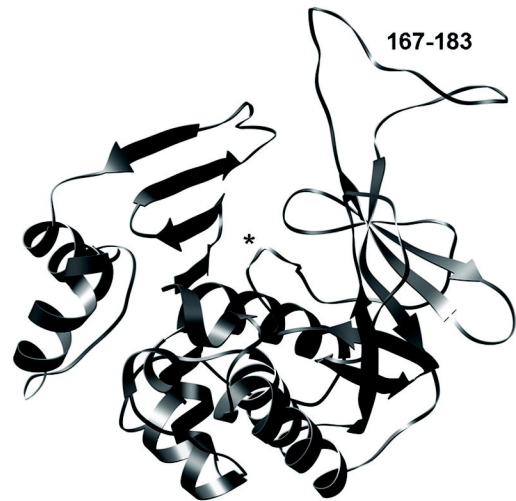
<i>Salmonella typhimurium</i>	 HGEYRLM <u>QEGS</u> TWILQFR <u>HHEH</u> WQSMYCFDLGVQ
<i>Pseudomonas aeruginosa</i>	 EADAGQVHCVRLVD <u>PHAGL</u> YESAVR <u>GRSG</u> WLPLY
<i>Mesorhizobium loti</i>	 EQKTPHEPFRIVE <u>ADDH</u> FRLQAAI <u>GGD</u> WRSLYRF

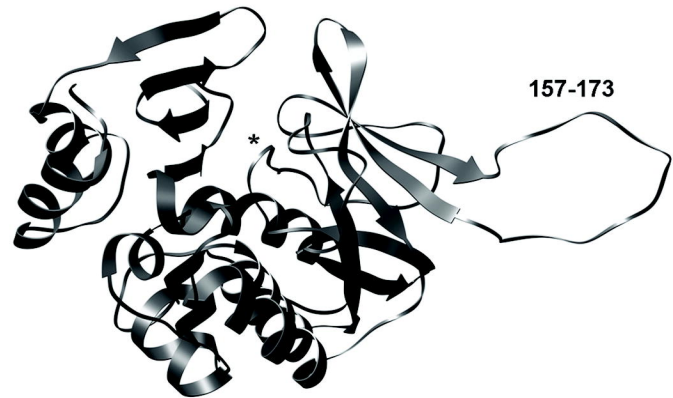
Figure 1



Human NAT2
3DCoffee Alignment



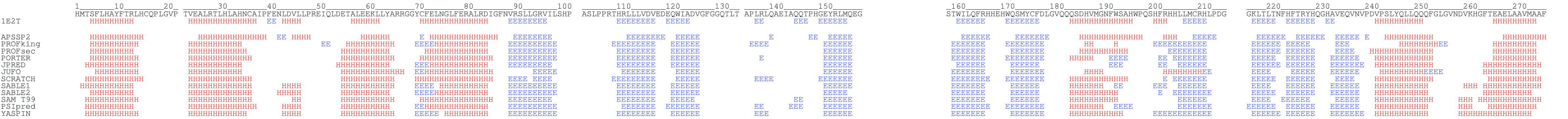
Mesorhizobium loti NAT1



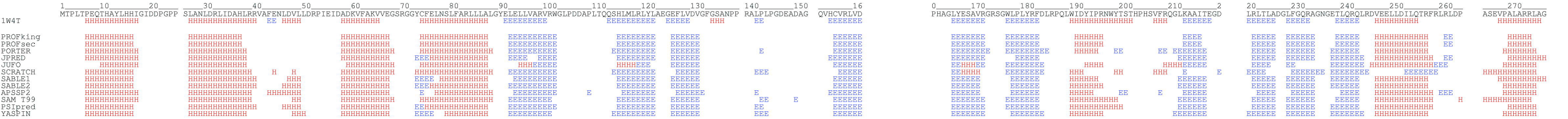
Human NAT2
Modeller Alignment

Modeller Alignment

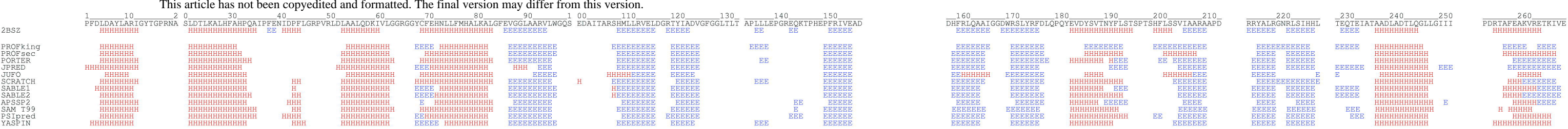
1E2T – Salmonella typhimurium NAT



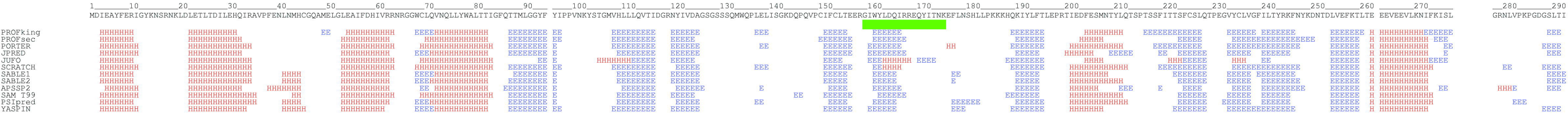
1W4T – Pseudomonas aeruginosa NAT



2BSZ – Mesorhizobium loti NAT

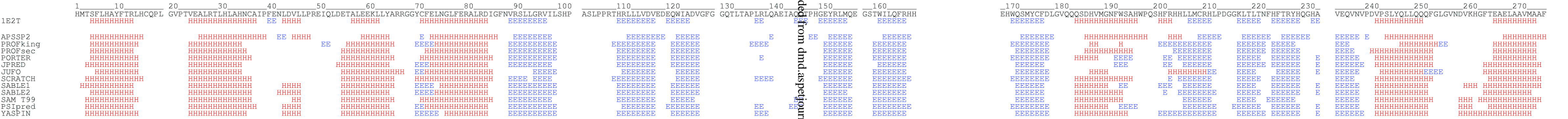


Human NAT2

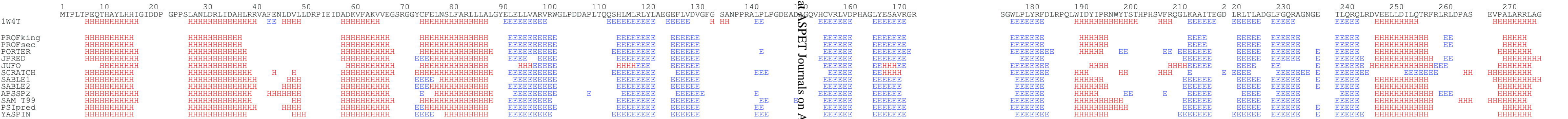


3DCoffee Alignment

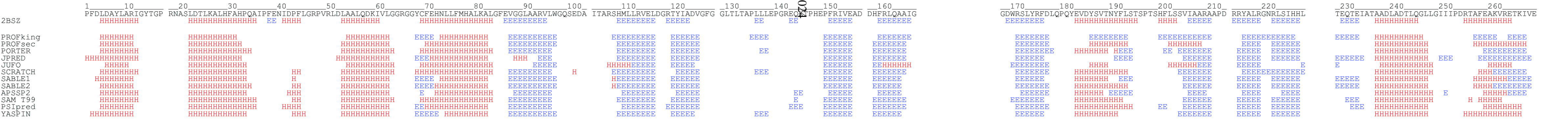
1E2T – Salmonella typhimurium NAT



1W4T – Pseudomonas aeruginosa NAT



2BSZ – Mesorhizobium loti NAT



Human NAT2

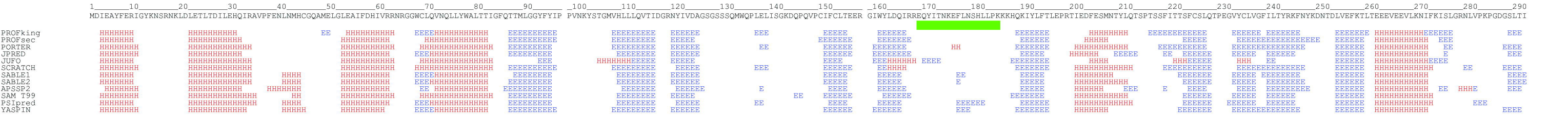


Figure 3

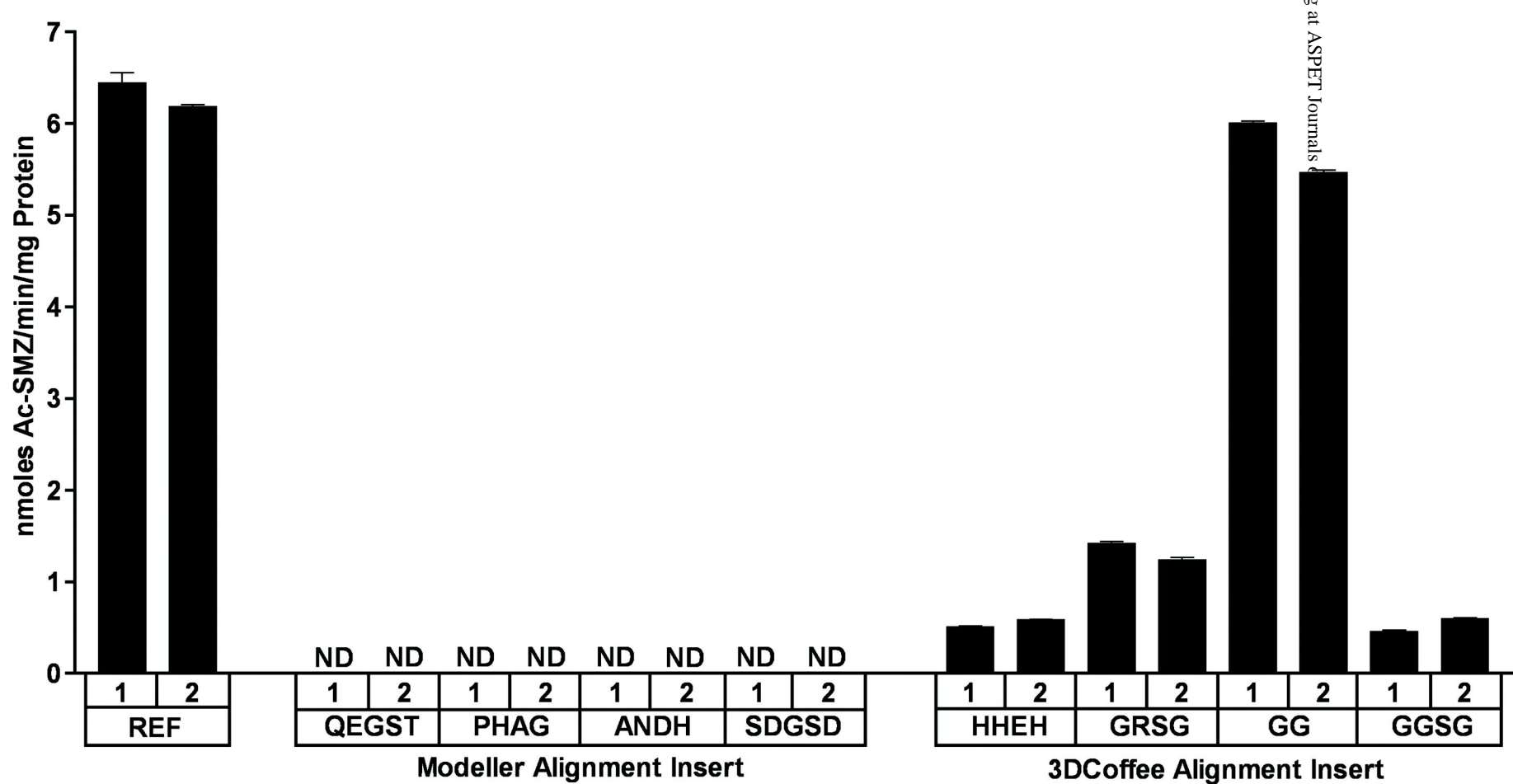


Figure 4

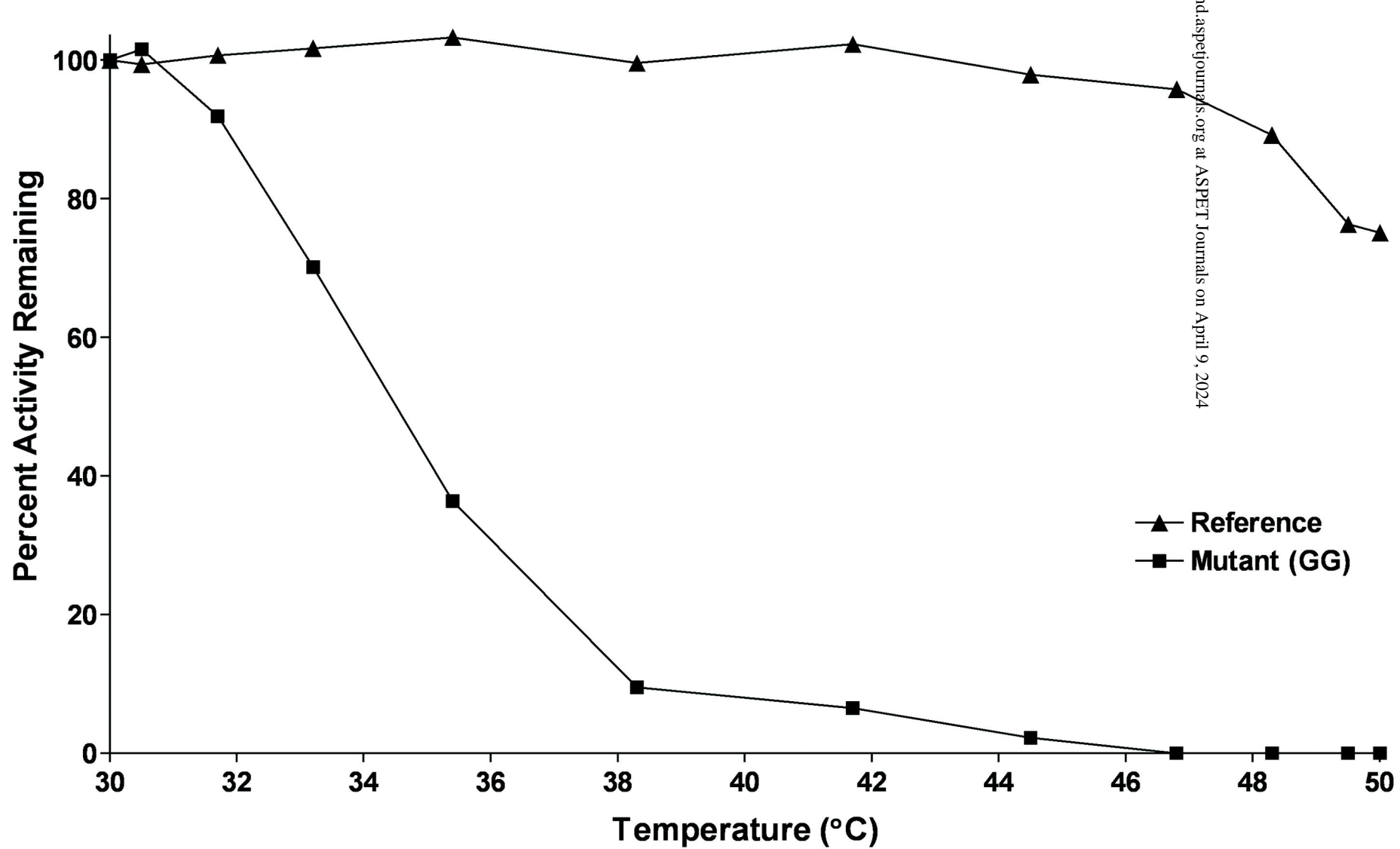


Figure 5