Journal of Molecular Graphics and Modelling 75 (2017) 330–339

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Structural and functional effects of nucleotide variation on the human TB drug metabolizing enzyme arylamine *N*-acetyltransferase 1

Ruben Cloete[a], Wisdom A. Akurugu[a], Cedric J. Werely[b], Paul D. van Helden[b], Alan Christoffels[a,*]

[a] South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Private Bag X17, Bellville, Cape Town 7535, South Africa
[b] SAMRC Centre for Molecular and Cellular Biology, and DST-NRF Centre of Excellence for Biomedical TB Research. Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, P.O. Box 241, Cape Town 8000, South Africa

ARTICLE INFO

Article history:
Received 23 November 2016
Received in revised form 11 April 2017
Accepted 12 April 2017
Available online 10 June 2017

ABSTRACT

The human arylamine *N*-acetyltransferase 1 (NAT1) enzyme plays a vital role in determining the duration of action of amine-containing drugs such as para-aminobenzoic acid (PABA) by influencing the balance between detoxification and metabolic activation of these drugs. Recently, four novel single nucleotide polymorphisms (SNPs) were identified within a South African mixed ancestry population. Modeling the effects of these SNPs within the structural protein was done to assess possible structure and function changes in the enzyme. The use of molecular dynamics simulations and stability predictions indicated less thermodynamically stable protein structures containing E264K and V231G, while the N245I change showed a stabilizing effect. Coincidently the N245I change displayed a similar free energy landscape profile to the known R64W amino acid substitution (slow acetylator), while the R242M displayed a similar profile to the published variant, I263V (proposed fast acetylator), and the wild type protein structure. Similarly, principal component analysis indicated that two amino acid substitutions (E264K and V231G) occupied less conformational clusters of folded states as compared to the WT and were found to be destabilizing (may affect protein function). However, two of the four novel SNPs that result in amino acid changes: (V231G and N245I) were predicted by both SIFT and POLYPHEN-2 algorithms to affect NAT1 protein function, while two other SNPs that result in R242M and E264K substitutions showed contradictory results based on SIFT and POLYPHEN-2 analysis. In conclusion, the structural methods were able to verify that two non-synonymous substitutions (E264K and V231G) can destabilize the protein structure, and are in agreement with mCSM predictions, and should therefore be experimentally tested for NAT1 activity. These findings could inform a strategy of incorporating genotypic data (i.e., functional SNP alleles) with phenotypic information (slow or fast acetylator) to better prescribe effective treatment using drugs metabolized by NAT1.

1. Introduction

Arylamine *N*-acetyltransferases (NATs) belongs to a group of phase II enzymes that metabolize various aromatic and heterocyclic amines, hydroxylamines, arylhydrazines and arylhydrazides. These enzymes are found in both prokaryotes and eukaryotes [1,2]. In humans there are two functional enzymes – NAT1 and NAT2 (33 kD and 31 kD proteins respectively) – encoded by polymorphic genes. However, these 290 amino acid containing proteins are derived from a single protein-coding exon of 870-bp [3–5]. The coding regions of *NAT1* and *NAT2* exhibit an 87% nucleotide sequence identity and an 81% identity at the amino acid level, with 55 amino acid differences. NATs catalyze either the acetyl-CoA dependent *N*-acetylation of primary aromatic amines and hydrazines (usually deactivating) or the *O*-acetylation of their *N*-hydroxylated metabolites (usually activating). This reaction is believed to occur via a two-step "ping-pong" Bi–Bi mechanism in which the acetyl group is initially transferred to an enzyme-thio-ester intermediate (acryl-NAT) before being transferred to primary amine substrate [6,7]. Human NAT1 is known to catalyze the metabolism of para-aminosalicylic acid [8] used in the treatment of some drug resistant

* Corresponding author.
E-mail addresses: ruben@sanbi.ac.za (R. Cloete), waakurugu.akurugu@gmail.com (W.A. Akurugu), cjw1@sun.ac.za (C.J. Werely), pvh@sun.ac.za (P.D. van Helden), alan@sanbi.ac.za (A. Christoffels).

http://dx.doi.org/10.1016/j.jmgm.2017.04.026

tuberculosis cases [9]. Since human NAT1 plays a vital role in detoxification and potential metabolic activation of xenobiotic compounds, it is important in clinical pharmacology and toxicology.

Previous research has shown non-synonymous single nucleotide polymorphisms (nsSNPs) in the genomic DNA sequence of human *NAT1* that imply changes in its amino acid sequence [10–14]. This has the potential to destabilize the enzyme structure and negatively affect the function of the NAT1 enzyme. The high-resolution crystal structures for both human NAT1 and NAT2 were resolved by Sinclair et al. (2000), which contributed immensely to the understanding of substrate specificities of these enzymes. Additionally, the determination of the homologous X-ray crystal structure of NAT from *Salmonella typhimurium* identified a highly conserved catalytically essential triad of residues: Cysteine-Histidine-Aspartate, which is present in all functional NAT enzymes [15]. In addition, other important contact residues have also been identified near the active site [16]. Some nsSNPs in human NAT1 have been found to affect its structure and function [17]. The study by Walraven and colleagues (2008) reported that amino acid substitution resulted in the loss of crucial hydrogen bond interactions, which resulted in reduced protein thermal stability. Thereafter, the effects of a number of human NAT1 nsSNPs in the coding region and 3′-untranslated region have been explored experimentally [14,18–21]. Relative to the human reference allele *NAT1*4*, nsSNPs were found to increase or decrease human NAT1 activity, defining rapid or slow acetylation phenotypes.

An investigation of a South African mixed ancestry population in the Western Cape region observed four novel nsSNPs in human *NAT1* [C Werely, data] in addition to the two well-known *NAT1* SNPs (R64W and I263 V) [http://nat.mbg.duth.gr/]. Previous studies have shown that functional effects of nsSNPs can be predicted computationally [17,22]. Therefore, this study utilized the crystal structure of human NAT1 (PDBID: 2IJA) to computationally test the effects of four novel and two published nsSNPs on the structure and function of NAT1 using sequence and structure based methods namely; protein stability calculations using mCSM, molecular dynamic simulations and the routinely used SIFT and POLYPHEN-2 algorithms. Our results will allow classification of nsSNPs within specific phenotype classes and also inform which of the nsSNPs can be prioritized for experimental validation based on destabilizing effects on NAT1.

2. Materials and methods

We interrogated the functional effects of these four SNPs by SIFT, POLYPHEN-2 and mCSM analyses [16,23,24]. We introduced each of the four nsSNPs into the structure of NAT1 (PDBID: 2IJA) to perform stability calculations using mCSM (24). We further evaluated the influence of these SNPs and two known SNPs on the dynamic behavior of the NAT1-COA-PABA complex using Molecular dynamic simulations.

2.1. SIFT and POLYPHEN analyses

Two on-line servers were used to predict the functional effect of the four novel nsSNPs on the protein structure using the SIFT [25] and POLYPHEN-2 [26]. The SIFT program calculates a tolerance index score ranging from 0 (deleterious) to 1 (neutral) of a particular amino acid substitution based on multiple sequence alignments. While, Polyphen-2 computationally calculates the position specific independent count score of amino acid substitutions and categorizes them according to a probabilistic score ranging from 0 (neutral) to 1 (deleterious) and functional significance of benign (0.00-0.14), possibly damaging (0.15–0.84) and probably damaging (0.85-1). Both SIFT and Polyphen-2 has been shown to be accurate at predicting loss-of-function mutations with both programs obtaining a sensitivity of approximately 70% based on a study by Flanagan et al., 2010 [27]. Briefly, the FASTA NAT1 protein sequence and a file containing the wild type and SNP residues were uploaded to the SIFT server. The input NAT1 protein was searched against the UNIREF release April (2011) database with default parameter for median conservation of sequence. Sequences in UNIREF that had a similarity more than 90% to NAT1 were excluded from the SIFT predictions to obtain acceptable median conservation levels (2.75-3.25). The FASTA formatted NAT1 protein sequence was submitted to the POLYPHEN-2 server. The position of the WT residue and each of the six *NAT1* nsSNPs were specified.

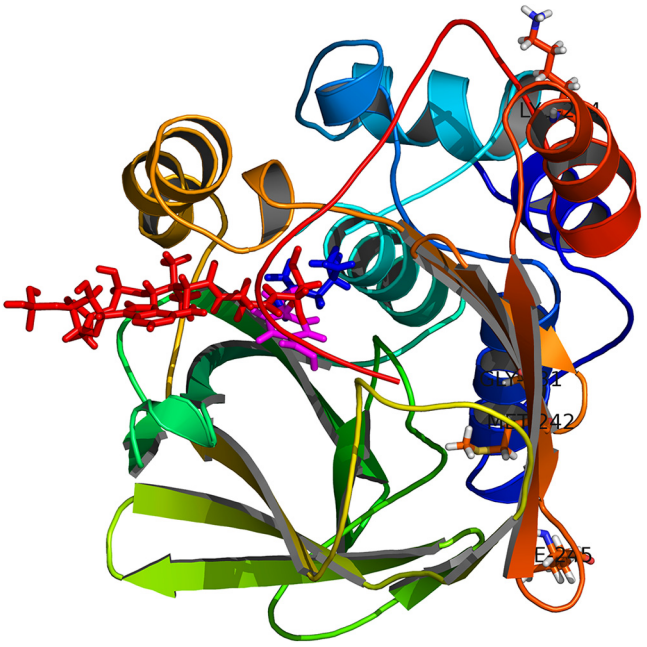

Fig. 1. The human NAT1 mutant crystal structure (2IJA) in complex with COA and PABA.
The four novel mutated residues identified in the South African mixed cohort are labeled and shown as sticks: acetylated cysteine CYS68 is shown in blue. Ligands are shown as sticks COA in red and PABA in magenta.

2.2. Protein structure stability analysis

The crystal structure of NAT1 (PDBID: 2IJA) was downloaded from PDB and the cysteine 68 residue was modified to be acetylated in complex with coenzyme A (COA) and TB drug para-aminobenzoic acid (PABA) [28] (Fig. 1). The wild type structure was mutated using the Pymol Mutagenesis Wizard and the resulting structures were submitted to the mCSM web server, which relies on graph-based signatures [24] (Fig. 1). These signatures encode distance patterns between atoms of the surrounding environment and physicochemical properties. In comparison to other predictive methods such as AUTOMUTE, CUPSAT, Dmutant, Eris, I-mutant-2.0, PoPMuSic-1.0, PoPMuSic-2.0 and SDM, mCSM outperformed all these methods achieving a correlation coefficient p-value of 0.824 and a standard error of 1.026 (kcal/mol) [24]. Therefore, mCSM web server was used to calculate the stability change ($\Delta\Delta G$ values) upon introduction of a mutation between the wild type and mutant structures [24]. The wild type and mutant structure were uploaded to the server and the position of the mutation was specified. Usually mutational changes with a negative stability score are classified as destabilizing (impact on protein structure) whereas a positive score