Article

In Silico Screening and Analysis of Potential Inhibitors of Arylamine N-Acetyltransferases (NATs) from the Traditional Chinese Medicine: A Study Using Free Available Tools

Francisco Honeidy Carvalho Azevedo[1,2], Joveline Richardt Lange[1], Mayrla Letícia Alves de Oliveira[3], Inês Maria de Souza Araujo[4], Ricardo Ferreira de Oliveira[1,5], Luis Andre Baptista[6] and Hermes Luís Neubauer de Amorim[1,*]

[1] Laboratory of Structural Bioinformatics, Lutheran University of Brazil (ULBRA), Canoas 92425-900, Rio Grande do Sul, Brazil
[2] Saint Augustine College (FSA), Teresina 64019-480, Piauí, Brazil
[3] State University of Piauí (UESPI), Teresina 64002-150, Piauí, Brazil
[4] Federal University of Piauí (UFPI), Teresina 64049-550, Piauí, Brazil
[5] Institute of Exact and Technological Sciences, Feevale University, Novo Hamburgo, 93525-075, Rio Grande do Sul, Brazil
[6] Institute of Chemistry, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre 91501-970, Rio Grande do Sul, Brazil
* Correspondence: hermes.amorim@ulbra.br; Tel.: +55-051-3477-4000 ext 2774

Abstract: Arylamine N-acetyltransferases (NATs) are cytosolic enzymes, highly polymorphic, present in both eukaryotes and prokaryotes. These enzymes play an important role in the detoxification and activation of xenobiotics as well as in the synthesis of endogenous compounds. Specific NATs have been pointed out in the literature as possible therapeutic targets. In particular, the human NAT1, for the treatment of certain cancers, and the NAT from *M. tuberculosis* (TBNAT), for the treatment of tuberculosis. This paper describes an *in silico* approach to prospect and select potentially inhibitors of NAT1 and TBNAT from the Traditional Chinese Medicine (TCM) using free available tools. A library with ligands from TCM was previously screened in order to select only compounds with optimal pharmacological properties. The affinity of the selected ligands with respect to NAT enzymes was then evaluated by virtual screening (VS). Subsequently, the complexes with the best ligands were submitted to molecular dynamics (MD) simulations aiming to obtain better quality information on affinity and selectivity. The results for one specific ligand, ZINC14690579, indicated its potential for affinity and selectivity. ZINC14690579 structure may represent the discovery of a new scaffold for future development of NAT inhibitors.

Keywords: Arylamine N-acetyltransferases; cancer; tuberculosis; drug discovery; Traditional Chinese Medicine; virtual screening; molecular dynamics simulation; MM-PBSA.

1. Introduction

Arylamine N-acetyltransferases (NATs) are cytosolic enzymes that catalyze the transfer of the acetyl group from acetyl coenzyme A (AcCoA) to the free amino group of arylamines and hydrazines[1–4]. NATs are highly polymorphic and present in both eukaryotes and prokaryotes[5,6]. These enzymes play an important role in the detoxification and metabolic activation of xenobiotics[4,7,8] as well as in the synthesis of endogenous compounds[9,10].

NATs are found in the human proteome in two forms, NAT1 and NAT2[4,11]. Several studies have revealed an association between NAT1 activity and the risk of developing certain types of cancer (for review see [12]). Consequently, human NAT1-specific inhibitors has been suggested as

potential biomarkers as well as potential drugs for cancer treatment [12–15]. In another therapeutic perspective, studies showed that NAT gene of *Mycobacterium tuberculosis* (*M. tuberculosis*) is part of a cluster essential for the survival of the mycobacteria in the environment inside macrophage [10,16,17]. This points to NAT from *M. tuberculosis* (TBNAT) as a potential drug target for antimycobacterial therapy[18,19].

The challenge for the design of selective inhibitors for eukaryotic and prokaryotic NATs begins with the structural similarity shared by these enzymes. In fact, the fold of human NAT1 and NAT2 closely resembles the overall structure of TBNAT[11,20,21]. NATs overall fold is composed of three independent domains of approximately the same length. The first two domains, an α-helical bundle and a β-barrel are joined by a linker helix to the third domain, an α/β lid. For native enzymes, the residues of the catalytic triad (Cys-His-Asp) can be easily superimposed. The modular structure of NATs makes these enzymes highly flexible, adding another challenge for the *in silico* discovery of inhibitors. For instance, previous studies have suggested that there may be multiple conformational states in the region that defines the active site of NAT enzymes[22], which in itself determines that the interpretation of data obtained from computational techniques necessarily take into account the conformational dynamics in the ligand-receptor interaction.

There is currently a wide range of free resources[23–28] and free software packages[29–35] dedicated to drug discovery and design. The availability of free resources and tools allows scientists from all over the world, independent of the financial support, to plan and test hypotheses through *in silico* experiments. This paper describes an *in silico* approach to prospect and select potentially selective inhibitors of NAT1 and TBNAT from the Traditional Chinese Medicine (TCM) using free available tools and resources. A library with 36,043 ligands from TCM was previously screened in order to select only compounds with optimal ADMET (absorption, distribution, metabolism and excretion and toxicological properties) pharmacological properties. The affinity of the 1,902 selected ligands with respect to NAT1, NAT2 and TBNAT enzymes was then evaluated by virtual screening (VS). Subsequently, the complexes with the best ligands were submitted to molecular dynamics (MD) simulations aiming to obtain better quality information on the affinity of the ligands. In particular, the MM-PBSA method combined with MD simulations was used to incorporate conformational fluctuations and entropic contributions to the affinity estimation [36–38]. One potential selective inhibitor for NAT1 and TBNAT was chosen as a promising compound for future development.

2. Results and Discussion

2.1. Virtual screening and docking results

Figure 1 shows the 2D chemical structures of the ten compounds selected from the virtual screening. Since of none of the ten best ligands found for NAT1 presented better binding score when compared to NAT2 (Table S1 in Supplementary Materials), the selection criterion was based on the best binding score with TBNAT and worse binding score with NAT2. Thus, only the ten best compounds for TBNAT were locally re-evaluated for the purpose of verifying the results generated by the virtual screening with AutoDock Vina[33] (Vina) software implemented in the MTiOpenScreen[28] web server (Table 1). Table 1 also shows the binding score values for the same

ligands with the AutoDock[32] software. Autodock uses a completely different methodological approach for score function calculation in comparison to AutoDock Vina. This allows, taking into account the standard error of each program, a preliminary covalidation of the VS. From VS and docking data it is possible to observe that both AutoDock Vina and AutoDock 4.2 tended to have worse binding energy scores for NAT1 when compared to NAT2 and TBNAT. So far we do not have an explanation for this since the binding residues in the homology model of human NAT1 are spatially conserved in relation to the two crystallographic structures available up to the time of the conclusion of this study. Specifically, are deposited in the Protein Data Bank (PDB)[39,40] structures with ID codes 2IJA[41], which refers to the structure of the mutant NAT1 F125S, and 2PQT, which refers to a structure with the cysteine residue at position 68 substituted by a non-standard S-(2-anilino-2-oxoethyl)-l-cysteine (Tyx68) residue. After superimposition it is possible to see that the only conformational exception are the residues His107 and Phe125 in the crystallographic structure 2PQT, which are influenced by the Tyx68 side-chain (Figure S1 in Supplementary Materials). As consequence, in the 2PQT structure the side-chain of His107 shows a torsion about 25° in chi2 angle and a displacement of 1.9 Å of Phe125 side-chain. Even though these differences may not have substantial influence in the docking results since there is no way to anticipate with which residues a non-substrate inhibitor will bind at the enzyme recognition site. Furthermore, as shown in section 2.2., it was possible to observe a spatial repositioning of the ligands during molecular dynamics simulations due to rearrangements in segments of the main chain around the NAT1 recognition site.
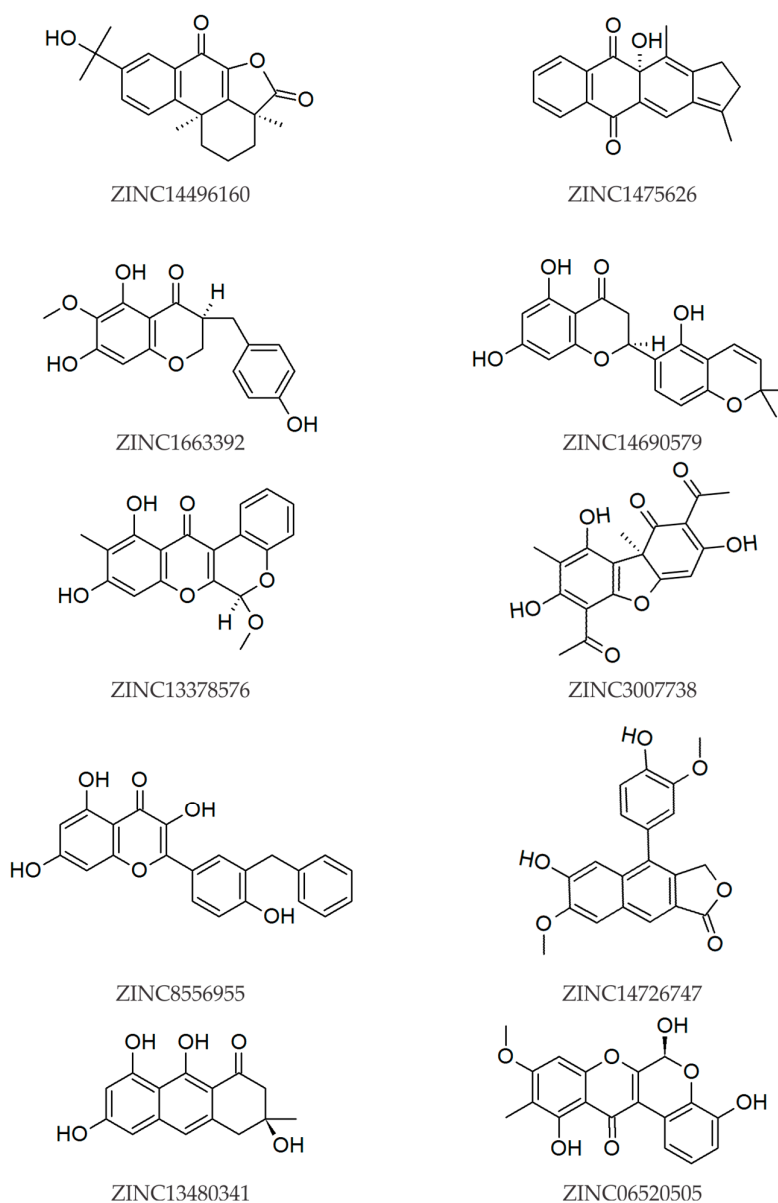
Figure 1. 2D chemical structures of the selected compounds from virtual screening with AutoDock Vina[33] implemented in the MTiOpenSreen[28] web server. Each compound is identified according to the registration number in the ZINC15[27] database.

In Table 1 are highlighted in bold the compounds that presented better binding energy for TBNAT than for NAT2 in both Vina and AutoDock: ZINC14496160, ZINC14715626, ZINC14690579, ZINC53007738 and ZINC13480341. The complexes with these compounds were then selected for analysis by molecular dynamics simulation.

Table 1. Comparison between the results of virtual screening with AudoDock Vina and the docking with AutoDock 4.2. The energies are in kcal/mol.

Ligand	Target enzyme, docking software and binding score energies					
	NAT1		NAT2		TBNAT	
	Vina	AutoDock	Vina	AutoDock	Vina	AutoDock
ZINC14496160	**-2.6**	**-2.6**	**-7.4**	**-8.3**	**-8.6**	**-9.3**
ZINC14715626	**5.8**	**3.7**	**-7.3**	**-7.5**	**-8.9**	**-8.1**
ZINC1663392	-6.0	-3.6	-8.7	-7.3	-8.1	-9.4
ZINC14690579	**-3.1**	**-1.5**	**-3.5**	**9.1**	**-9.0**	**-7.5**
ZINC13378576	-2.4	-0.6	-9.2	-7.9	-8.5	-7.5
ZINC53007738	**8.9**	**8.3**	**-6.0**	**-6.9**	**-8.3**	**-7.6**
ZINC85569505	-5.1	-6.4	-9.3	-8.3	-8.7	-8.1
ZINC14726747	0.7	1.3	-7.6	-7.8	-6.6	-7.3
ZINC13480341	**5.8**	**-3.6**	**-7.3**	**-6.5**	**-8.9**	**-7.3**
ZINC06520505	-2.2	6.3	-8.6	-7.1	-8.5	-7.1

2.2. Molecular dynamics simulations results

Aiming to increase the sampling all MD simulations were performed in triplicate. The stability of each replicate was evaluated from atom-positional root mean-square deviation (RMSD) analysis (Figure S1 in Supplementary Materials). The last 10 ns of each replicate were extracted and subsequently concatenated forming a single trajectory of 30 ns for further analysis. The conformational stability of the concatenated trajectories for each complex was estimated from the analysis of the mean value and standard deviation of the RMSD (Table S2 in Supplementary Materials). Except for TBNAT complexes, protein RMSD average values for concatenated trajectories were found between 0.17 to 0.20 nm with a maximum fluctuation of ± 0.02 nm, suggesting stability within the ensemble of conformations considered. Similarly, the RMSD average values for the ligands were found between 0.03 to 0.12 nm with a maximum fluctuation of ± 0.03 nm. Protein TBNAT-complexes presented greater variations of RMSD but with stabilization around 0.3 nm in the last 10 ns of the MD simulations (Figure S2 in Supplementary Materials). Ligand RMSDs for the TBNAT complexes ranged from 0.03 to 0.10 nm with a maximum fluctuation of ± 0.01 nm, except for TBNAT-ZINC14496160 which showed a RMSD of 0.78 ± 0.03 nm. The higher mean value of RMSD and the greater amplitude of the conformational fluctuations observed for the TBNAT complexes can be attributed to a greater flexibility and to more significant conformational changes in the enzyme during the simulations. In fact, it is possible that the greater flexibility observed for the MD simulations of TBNAT-ZINC14690579 compared to those observed for NAT1 and NAT2 (Figure S2 and Table S2 in Supplementary Materials) are related to the existence of an additional seventeen amino acid insertion spanning residues 167–183 in human NATs[21]. This region, absent in prokaryotic NATs, is reported to increase NAT protein stability[42].

The average number of hydrogen bonds (Table S2 in Supplementary Materials) ranged from a minimal of 0.44 (for NAT1-ZINC14715626) to a maximum of 2.09 (for NAT2-ZINC14496160), indicating that the affinity of the ligands occurred mainly through hydrophobic contacts.

Binding score energies of NAT1 and TBNAT inhibitor complexes were analyzed from the concatenated production phase of MD simulations by MM-PBSA method . Table 2 shows MM-PBSA binding score energies for the five compounds select from VS and docking. It is possible to observe that the compound with the greatest binding score difference between TBNAT and NAT2 was ZINC14690579 ($\Delta\Delta G_{bind}$ = -12.7 kcal/mol). This data is in agreement to Table 1, where only ZINC14690579 presented difference in binding score ($\Delta\Delta G_{bind(Vina)}$ = -5.5 kcal/mol; $\Delta\Delta G_{bind(AutoDock)}$ = -16.6 kcal/mol) greater than the standard error of each docking program (which is of about 2–3 kcal/mol in prediction of binding free energy in cross-validation studies)[32,33].

In an unexpected manner, the difference in binding score for compound ZINC14690579 ($\Delta\Delta G_{bind}$ = -15.7 kcal/mol) also demonstrating a potential selectivity factor between NAT1 and NAT2. Thus, the interaction of the respective ligand with TBNAT and NAT1 was subjected to a more detailed analysis.

Table 2. Binding score energies for the five compounds with more selective-inhibitory potential from VS revaluated using MM-PBSA method. The energies are in kcal/mol.

Ligand	Target enzyme and binding score energies		
	NAT1	NAT2	TBNAT
ZINC14715626	-20.4	-24.4	-19.5
ZINC14496160	-19.3	-19.6	-22.1
ZINC14690579	**-20.5**	**-4.8**	**-17.5**
ZINC53007738	-15.0	-13.5	-1.8
ZINC13480341	-18.6	-12.9	-18.7

From Table 3 it is possible to observe that the component that most contributes to the interaction of ZINC14690579 with TBNAT and NAT1 is related to the non-polar/hydrophobic contributions (ΔE_{vdw}). In this sense, the sum of the energetic terms allows to conclude that the estimated selectivity profile (predicted in function of $\Delta\Delta G_{binding}$) for NAT1/NAT2 and TBNAT/NAT2 is mainly related to the smaller contribution of the non-polar/hydrophobic interactions of NAT2 with ZINC14690579. Specifically for NAT1, a greater contribution of the electrostatic contributions (ΔE_{elec}) is observed when compared to NAT2 and TBNAT. As will be seen below, this must be related to the contribution of basic residues around the NAT1 binding site.

Table 3. MM-PBSA calculation for NAT-ZINC14690579 complexes. The energies are in kcal/mol.

NAT	Energy components[a]				$\Delta G_{binding}$	$\Delta\Delta G_{binding}$[b]
	ΔE_{elec}	ΔE_{vdw}	$\Delta E_{polar\text{-}solv}$	$\Delta E_{non\text{-}polar}$		
NAT1	-15.2	-38.1	37.0	-4.2	-20.5	-15.7
NAT2	-6.0	-13.5	16.1	-1.4	-4.8	0.0
TBNAT	-5.7	-29.1	20.6	-3.1	-17.5	-12.7

[a]ΔE_{elec} = electrostatic component; ΔE_{vdw} = noncovalent van der Waals component;

$\Delta E_{polar\text{-}solv}$ = polar salvation component; $\Delta E_{non\text{-}polar}$ = non-polar salvation component

[b]$\Delta\Delta G_{binding}$ = ΔG_{NAT} - ΔG_{NAT2}

2.2.1. Molecular dynamics simulation of NAT1-ZINC14690579 complex

In order to determine the contribution of NAT1 residues involved in the interaction with the ligand, free energy decomposition was performed. The number of hydrophobic contacts and hydrogen bonds between NAT1 and ZINC14690579 was also calculated (Table 4). All analysis and calculations were realized considering the 30 ns concatenated trajectory. From Table 4 it is possible to observe that the residues that contributed the most to the interaction are non-polar (Val93, Ile106, Phe125, Leu209, Val216 and Phe217), which is in agreement with the value of ΔE_{vdw} (-38.1 kcal/mol) in Table 3. These residues account for 54.95% of the total hydrophobic contacts with the ligand (Table 4). From Table 3 it can be observed that the value of the term relative to the electrostatic interactions (ΔE_{elec}) between NAT1 and ZINC14690579 (-15.2 kcal/mol) is more than double compared to NAT2 (-6.0 kcal/mol) and TBNAT (-5.7 kcal/mol). From free energy decomposition data in Table 4 it is possible to suggest that the magnitude of the electrostatic contribution between protein and ligand in the NAT1-ZINC14690579 complex derives from the contribution of a group of basic residues (Lys99, Lys100, Arg165, Arg167, Lys188, Lys220 and Arg286). It is interesting to note that five of these residues (Lys99, Lys100, Arg165, Arg167 and Lys188) are distant from the ligand (Figure 2, Panels (a) and (b)), but together they form a basic cluster that contributes significantly to the interaction. Considered together, the fact that the main contributions to the ligand-receptor interaction in the NAT1-ZINC14690579 complex occur through non-polar contacts and long-distance electrostatic interactions suggests that the driving force that maintains the ligand in the active site is mainly the result of a sum of non-directional interactions. In fact, the number of ligand-receptor hydrogen bonds calculated for the NAT1-ZINC14690579 concatenated trajectory was found small, 1.74 ± 0.83 (Table S2 in Supplementary Materials). From Table 4 it can be seen that the most important hydrogen bond interaction was with the backbone of Phe217 (42.2% of the total hydrogen bonds). The other three amino acids with which hydrogen bonds were detected (Phe125, Ser215, Val216) accounted for 34% of the total of these interactions.

Redocking of ZINC14690579 in a structure of NAT1-ZINC14690579 complex collected at 20 ns from MD simulation resulted in a binding energy score of -8.1 kcal/mol. Compared with the value obtained for the docking of ZINC14690579 on the NAT1 homology modeling on Table 1 (-3.1

kcal/mol) this results in a difference of 5 kcal/mol, which is greater than the AutoDock Vina standard error. This finding may indicate that the crystallographic structures of NAT1 may not be the most suitable for docking studies with homology models built from these.

Comparison of the homology modeling structure of NAT1 and a 20ns structure from MD simulation shows important conformational differences in the region of NAT1 binding site (Figure 2, Panel(c)). In particular, the main chain of the segment formed by the residues Ser95 to Lys99 moves 3.45 Å towards the ligand. Following this movement, the backbone and side chain of Val93, the residue that most contributes individually to the protein-ligand interaction, moves around 1.50 closer to the ZINC14690579. In the same sense, a shift of about 1.7 Å in the backbone of the α-helical interdomain (helices α8-α10) allows the side chains of Leu209, Ser215, Val216 and Phe217 to maximize the interaction with the ligand.

Table 4. Relative contribution of residues to binding between NAT1 and ZINC14690579

Residue	Complex NAT1-ZINC14690579		
	Contribution Energy (kJ/mol)[1]	Hydrophobic contacts[2] (%)	Hydrogen bonding (%)[2,3]
Cys68	-2.02	8.53	n.d.
Val93	-6.65	9.73	n.d.
Pro97	-1.97	2.70	n.d.
Lys99	-1.51	n.d.	n.d.
Lys100	-2.30	0.52	n.d.
Ser102	-1.73	2.30	n.d.
Ile106	-4.22	7.08	n.d.
Phe125	-4.70	8.37	16.7
Arg165	-1.24	n.d.	n.d.
Arg187	-2.13	n.d.	n.d.
Lys188	-1.30	n.d.	n.d.
Leu209	-4.63	6.07	n.d.
Ser215	-1.33	1.62	12.8
Val216	-5.95	10.5	4.51
Phe217	-5.52	13.2	42.2
Lys220	-1.66	0.52	n.d.
Phe222	- 1.15	1.12	n.d.
Arg286	-2.66	3.00	n.d.
Phe288	-2.70	4.73	n.d.

[1] Energy contribution estimated by the MM-PBSA method[38].

[2] Percentage of hydrophobic contacts and hydrogen bonds calculated with the LIGPLOT[43] program.

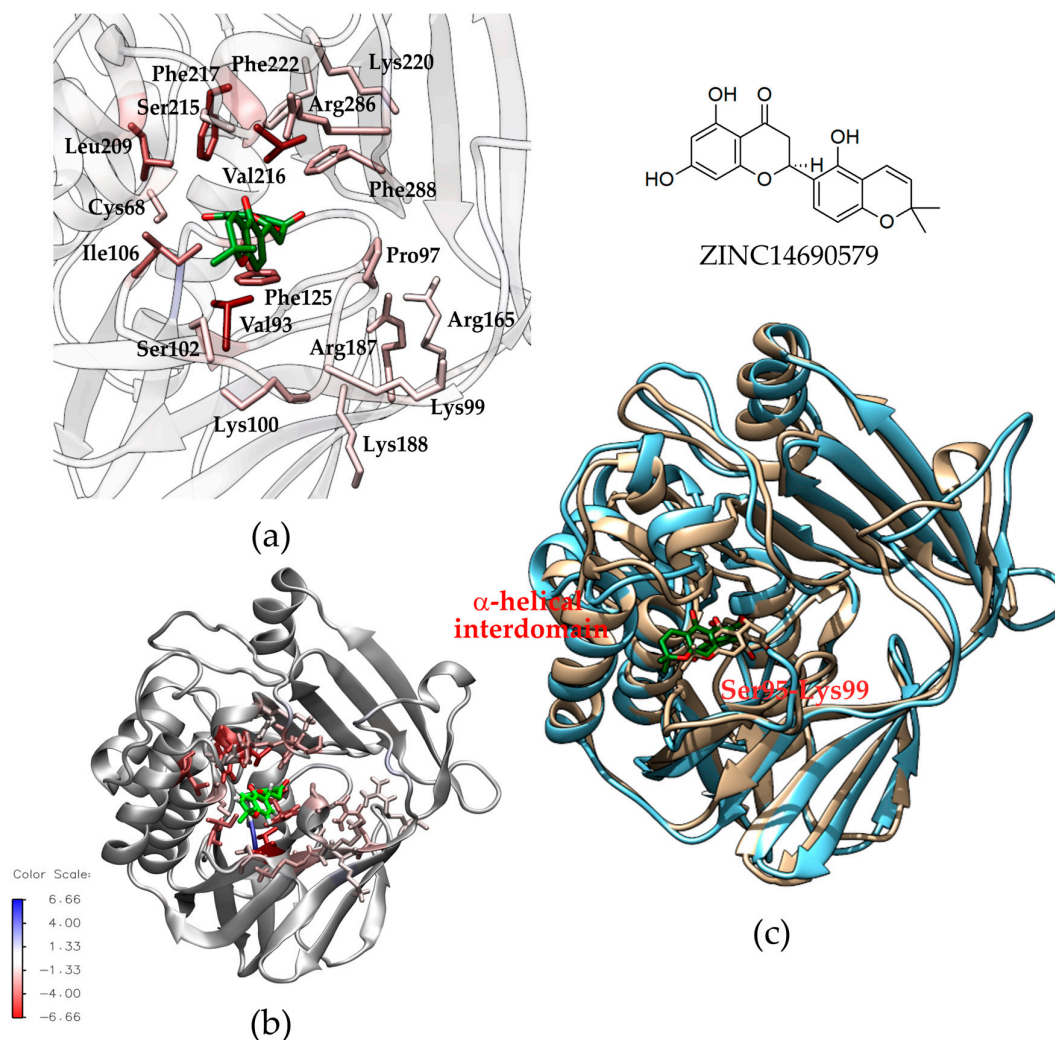[3] Hydrogen bonds considered only for residues that had more than 2% of the total interactions.

Figure 2. Schematic representation of the interaction of NAT1 with ZINC14690579. (a) Representation of main interacting residues of NAT1 with ZINC14690579. NAT1 is represented as transparent ribbon and sticks with shades of red. The ligand is represented as green sticks. The contribution of each residue to the interaction with the ligand is proportional to the increase in red tonality (see Table 4). (b) Representation of NAT1- ZINC14690579 in silico complex. NAT1 is represented as gray ribbon and sticks with shades of red. The color bar indicates the relative interaction energy of each residue with the ligand. (c). Structural changes induced by ligand binding at the recognition site of NAT1 depicted by the superimposition of the crystallographic structure of NAT1 (represented gray ribbon) and the 20ns structure from the MD simulation (represented as blue light ribbon). Ligand structure is represented as gray sticks (from docking) and green stick (from DM simulation). Structure of ZINC14690579 on top of panel (c).

2.2.2. Molecular dynamics simulations of TBNAT-ZINC14690579 complex

Decomposition of free energy and the counting of hydrophobic contacts and hydrogen bonds were performed in the same way as for the NAT1-ZINC14690579 complex. From Table 5 it is possible to observe that all the ten residues that contributed to the interaction are non-polar. This is in agreement with the value of ΔE_{vdw} (-29.1 kcal/mol) in Table 3. Already, the lack of polar residues and hydrogen bonds involved in the interaction with the ligand is in agreement with the low contribution of the electrostatic interactions ($\Delta E_{elec} =$ -5.7 kcal/mol). These observations suggest that the protein-ligand interactions in the TBNAT-ZINC14690579 complex are mainly non-directional. For instance, one of the main elements involved in the interaction between TBNAT and ZINC14690579 derives from a cluster of non-polar residues located between Leu99-Ala100 and Ala103-Pro107 (Table 5; Figure 3, Panels (a) and (b)). It is interesting to note that the interaction with this cluster of non-polar residues is a result of a backbone shift of the Leu99-Pro107 segment of approximately 6.2 Å toward the ligand (Figure 3, Panel (c)). Other conformational changes that contribute to maximize the interaction with the ligand involve displacement of the side chains of Phe130, Ala201 and Phe204 in 2.47, 8.31 and 2.97 Å respectively. In particular, changes in the position of residues Ala201 and Phe204 is accompanied by a movement of the backbone of the third domain of approximately 7.5 Å compared to the crystallographic structure of the TBNAT. The obtained results point to a series of concerted rearrangements that result in the maximization of protein-ligand interactions. These observations can be interpreted in the way that the conformational adaptability of the NATs allows these enzymes to perform various functions[21,22].

Table 5. Relative contribution of residues to binding between TBNAT and ZINC14690579

Residue	Complex TBNAT-ZINC14690579		
	Contribution Energy (kJ/mol)[1]	Hydrophobic contacts[2] (%)	Hydrogen bonding[2,3] (%)
Leu99	-3.16	13.45	n.d.
Ala100	-1.52	6.51	n.d.
Ala103	-1.95	3.98	n.d.
Pro104	-1.56	0.17	n.d.
Leu105	-1.93	0.34	n.d.
Pro106	-3.53	9.16	n.d.
Pro107	-1.78	9.59	n.d.
Phe130	-1.75	8.78	n.d.
Ala201	-2.04	1.76	n.d.
Phe204	-1.93	6.85	n.d.

[1] Energy contribution estimated by the MM-PBSA method[38].
[2] Percentage of hydrophobic contacts and hydrogen bonds calculated with the LIGPLOT[43] program.
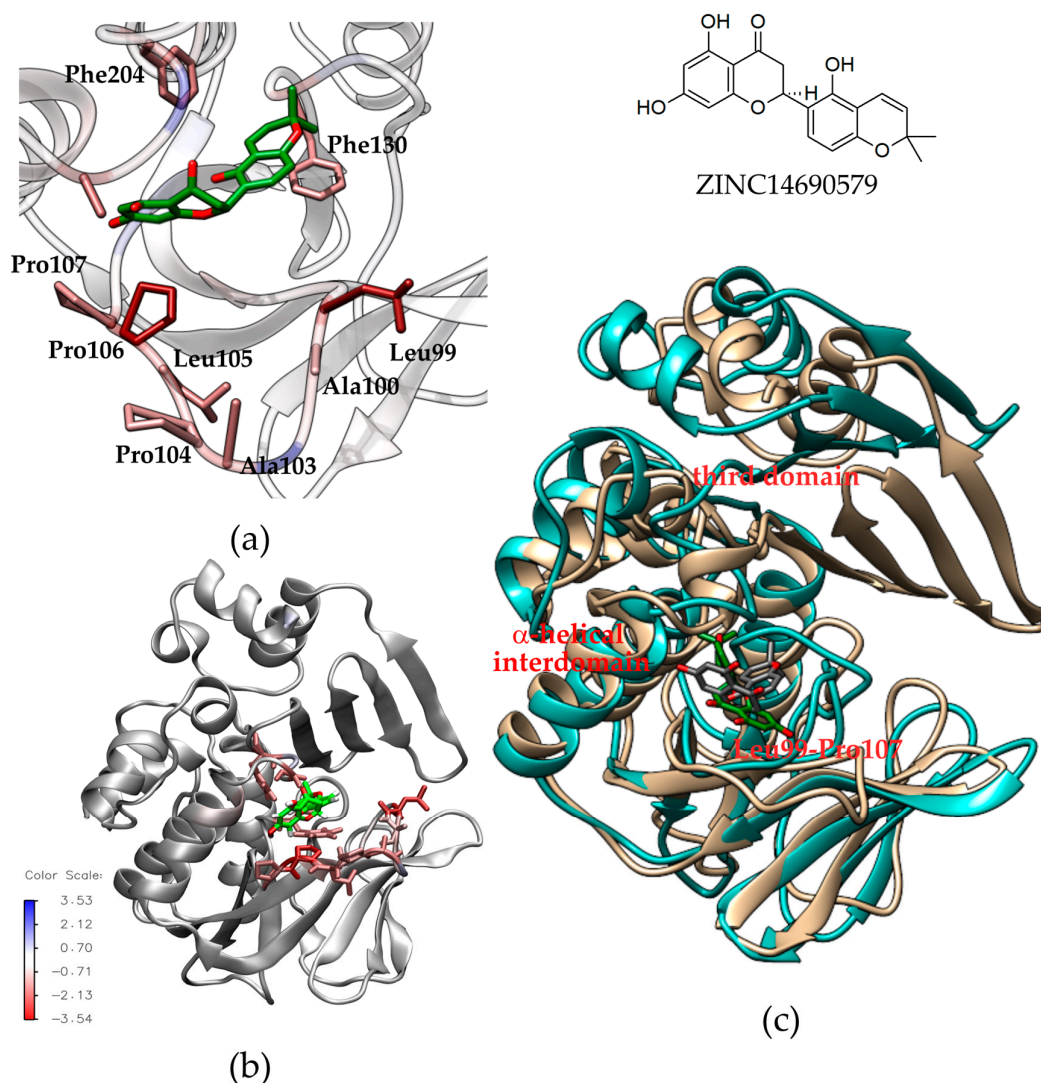[3]Hydrogen bonds considered only for residues that had more than 2% of the total interactions.

Figure 3. Schematic representation of the interaction of TBNAT with ZINC14690579. (a) Representation of main interacting residues of TBNAT with ZINC14690579. TBNAT is represented as transparent ribbon and sticks with shades of red. The ligand is represented as green sticks. The contribution of each residue to the interaction with the ligand is proportional to the increase in red tonality (see Table 5). (b) Representation of TBNAT-ZINC14690579 in silico complex. TBNAT is represented as gray ribbon and sticks with shades of red. The color bar indicates the relative interaction energy of each residue with the ligand. (c). Structural changes induced by ligand binding at the recognition site of TBNAT depicted by the superposition of the crystallographic structure of TBNAT (represented gray ribbon) and the 20ns structure from the MD simulation (represented as blue light ribbon). Ligand structure is represented as gray sticks (from docking) and green stick (from DM simulation). Red arrows point to the main conformational changes in the recognition site. Structure of the ligand on top of panel (c).

A potential anti-tuberculosis activity of ZINC14690579 can be extrapolated from evidence of inhibitory activity of the extract from *Premna odorata* Blanco (Lamiaceae), a medicinal plant traditionally used in Albay Province, in south-eastern Luzon, Philippines, against *M. tuberculosis*[44]. It was observed that the most active fraction of the extract of *Premna odorata* Blanco has a ZINC1469057 analog, the diosmetin (Figure 4). Although not conclusive, this data combined with the

in silico observations in this work may indicate that ZINC14690579 scaffold may be potentially effective for inhibition of TBNAT and, consequently, treatment of tuberculosis.
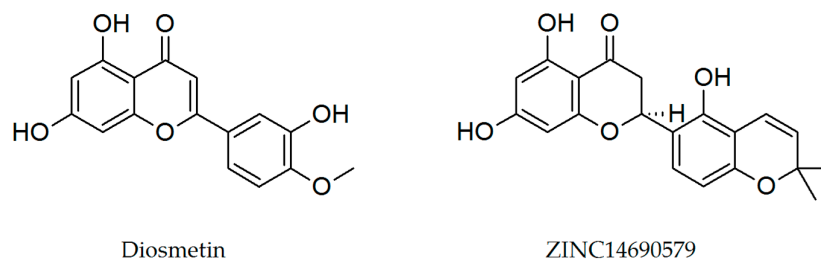

Figure 4. Comparison between 2D structures of Diosmetin and ZINC14690579.

3. Materials and Methods

3.1. Library preparation

The library containing 36,043 compounds from the Traditional Chinese Medicine (TCM) was downloaded in the SDF (Structures Data File) format from the ZINC database[27], catalog TCM Database@Taiwan[45]. Before the use, the TCM library was previously formatted and cleaned (to remove inorganics, large compounds, mixtures, empty structures, salts and duplicates) with Bank Formatter and Bank Cleaner tools, respectively, both implemented in the Mobyle[46] portal. This procedure was followed by the addition of hydrogens and attribution of the 3D structure for each compound with the Standardizer[47] software. Geometries of the ligands were then optimized with the AMMOS[48] software implemented in the Mobyle portal. Afterwards, the library was submitted to the FAF-Drugs3[5,6] server. Filtering rules were chosen in order to detect and to remove potentially non-bioavailable compounds, substructures with potential toxic groups, aggregators and Pan-Assay interference compounds[26]. At the end of this stage, 1,902 compounds were left in the library.

3.2. Target preparation

NAT2 and TBNAT structural models were obtained from crystallographic structures deposited in the Protein Data Bank[39,40] (PDB) (www.rcsb.org): NAT2, accession code 2PFR (resolution 1.9 Å)[50] and TBNAT, accession code 4BGF (resolution 2.1 Å)[51]. Since the two human NAT1 structures deposited in PDB have missing atoms and modified amino acid residues, the structural model of NAT1 was obtained by homology modelling. The 3D coordinates of wide-type NAT1 (amino-acid sequence access code P18440-1 of UniProtKB[52] database) was generated using the SWISS-MODEL Workspace[53]. As template, the crystallographic structure of mutant (F125S) NAT1 (PDB accession code 2IJA, resolution 1.7 Å)[41] was used. Solvent, ions and ligands were removed from the target structures for the subsequent steps.

3.3. Virtual screening

Virtual screening with the filtered library was performed with the AudoDock Vina[33], version 1.1.2, software implemented in the MTiOpenScreen[28] web server. All screenings were performed using NAT1, NAT2 and TBNAT as targets. Each enzyme was treated as rigid and each ligand in the

library was treated as flexible. Gasteiger[54] partial charges were calculated after addition of all hydrogens. Nonpolar hydrogens of enzyme and ligand were subsequently merged. A box of approximately 14×14×14 Å was generated for the each target. The grid boxes were centralized on the peptidic O of Ile106 (for NAT1), peptidic O of Val106 (for NAT2) and peptidic O of Thr109 (for TBNAT).

MTiOpenScreen returns only the 1000 best ligands for each target. As for NAT1 no compound had more favorable binding score than for NAT2, the ten best compounds were selected considering only TBNAT, i.e., the ten ligands with most favourable binding score for TBNAT and the worst binding score for NAT2. In order to confirm the results of the virtual screening, a second round of docking simulations were locally performed with the ten best compounds selected from MTiOpenScreen. In this round the dockings were performed with AutoDock Vina[33] implemented in the PyRx[35] graphic interface. Except where otherwise indicated, the standard parameters for the programs were used.

3.4. Docking with AutoDock 4.2

Autodock[32], version 4.2, a docking program that uses a different approach to the scoring function in comparison to AutoDock Vina, was used to compare the results obtained from the virtual screening. The docking protocol was the same as described for AutoDock Vina, except for the dimensions of the grid box and the use of the Lamarckian Genetic Algorithm. Targets and ligands were prepared for docking simulations with the AutoDockTools (ADT)[55] interface, version 1.5.6. A cubic box of $40 \times 40 \times 40$ points with a spacing of 0.35 Å between the grid points was generated for each protein target. The grid boxes were centered as for virtual screening protocol. Global search Lamarckian genetic algorithm (LGA)[56] and local search (LS) pseudo-Solis and Wets[57] methods were applied in the docking search. Each single docking simulation consisted of 10 independent runs. The initial population was 150, the maximum number of generations was 27,000 and the maximum number of energy evaluations was 2.5×10^6. Default values were selected for other parameters.

3.5. Molecular dynamics simulations

3.5.1. Ligand setup

The topology for each ligand was generated from the Automated Topology Builder (ATB)[58][59] web server. The conformation with better energy score from the docking simulation with Autodock Vina was taken as input structure. Missing hydrogens were added and the ligand structure was uploaded to the ATB server preserving the original geometry.

3.5.2. Molecular dynamics protocol

Molecular dynamics simulations were carried out using GROMOS96 53a6[60] force field implemented in GROMACS package[34], version 4.5.6. The protonation states of histidines were determined from H++[61] web server. All systems were simulated in NPT ensemble and periodic boundary conditions. The dimensions of the central box were chosen so that the minimum distance between of any protein atom to the closest box wall was 12 Å. The simulations were carried out using explicit solvent water molecules described by the simple point charge (SPC)[62] model. Initially, the protein structure in each system was submitted to a maximum of 500 steps of steepest

descent energy minimization. To relax strong solvent-solvent and solvent-protein non-bonded interactions, 100 ps of MD simulation was performed restraining the protein structure. Initial velocities were assigned according to Maxwell distribution. Counter-ions were added to neutralize each system. Table 1 summarizes the characteristics of each NAT simulated.

Table 6. Characteristics of each NAT simulated.

NAT enzyme	Organism	Number of residues	Number of atoms	Liquid charge
NAT1	Homo sapiens	289	3098	-3
NAT2	Homo sapiens	290	3044	-5
TBNAT	*M. tuberculosis*	270	2684	-6

All simulations were performed in triplicate, using different initial random velocity distributions, for 20 ns with an integration time step of 2 fs. Each system was heated with gradual increments in the following temperatures: 150 K (100 ps), 200 K (100 ps), and 250 K (100 ps). After, the temperature of each system was adjusted to 309.15 K. The first 10 ns of MD simulations was considered as part of the heating (0.30 ns) and the equilibration (9.70 ns) steps and had not been used in the data analysis. The temperature of each group (protein, ligand, solvent and ions) was independently coupled to a thermal bath with a relaxation time of 0.1 ps using the v-rescale thermostat[63]. The pressure in each system was weakly coupled to a pressure bath of 1 atm applying an isotropic scaling and 0.5 ps of relaxation time using the Parrinello-Rahman barostat[64,65]. Bond lengths were constrained using the LINCS algorithm[66] with 4th order expansion. Electrostatic interactions among non-ligand atoms were evaluated by the smooth particle mesh Ewald (SPME)[67] method with a charge grid spacing of approximately 1.0 Å. The charge grid was interpolated on a cubic grid with the direct sum tolerance set. Lennard-Jones interactions were evaluated using a 14 Å atom-based cutoff. The pair list was updated every 10 steps.

3.6.3. Analysis of the MD trajectories

The last 10ns of each triplicate was concatenated in a 30 ns trajectory for further analysis. As default, the trjconv module of GROMACS was used to restore, if necessary, protein and ligand inside the simulation box, as well as to make the fit of the conformations of each trajectory on the respective reference crystallographic structure. g_rms and g_hbond modules of GROMACS were used, respectively, to calculate the root-mean-square deviation (RMSD) of protein and ligand and the number of hydrogen bonds between protein and ligand. Biding score energies were estimated using the Molecular Mechanics – Poisson Boltzmann Surface Area (MM-PBSA) method, implemented through the module g_mmpbsa in GROMACS by Kumari and co-workers[38]. MM-PBSA based interaction energies were computed on snapshots collected every 2 ps from the DM trajectories. All calculations with g_mmpbsa were performed using default values. Protein-ligand hydrophobic contacts and hydrogen bonds were computed on snapshots collected every 100 ps from the DM trajectory using the LIGPLOT[43] program.

Visual inspection of the systems was performed using the Visual Molecular Dynamics (VMD)[68] and Chimera[69] software.

4. Conclusions

The use of free tools and resources allowed us to conclude the objectives of this study with a potential inhibitor identified. ZINC14690579 showed agreement in all methodologies used, indicating a potential for affinity and selectivity for NAT1 and TBNAT in relation to NAT2. This selectivity is mainly related to the smaller contribution of the non-polar/hydrophobic interactions of ZINC14690579 with NAT2 . The results indicate that the binding of ZINC14690579 to both NAT1 and TBNAT involves a set of concerted conformational changes resulting in accommodation of the ligand at the binding site. These initial observations on the conformational behaviour of NATs are in accordance with the theory that the conformational adaptability of these enzymes allows them to perform various functions, and should be taken into account in the *in silico* study of new inhibitors. In this sense, ZINC14690579 may represent the discovery of a new scaffold for future development of NAT inhibitors.

References

1.   Weber, W. W.; Cohen, S. N. N-acetylation of drugs: isolation and properties of an N-acetyltransferase from rabbit liver. *Mol. Pharmacol.* **1967**, *3*, 266–73.

2.   Riddle, B.; Jencks, W. P. Acetyl-coenzyme A: arylamine N-acetyltransferase. Role of the acetyl-enzyme intermediate and the effects of substituents on the rate. *J. Biol. Chem.* **1971**, *246*, 3250–8.

3.   Drummond, G. S.; Kelker, H. C.; Weber, W. W. N-acetylation of drugs. Observations on the properties of partially purified N-acetyltransferase from peripheral blood of rabbit. *Biochem. J.* **1980**, *187*, 157–62.

4.   Weber, W. W.; Hein, D. W. N-acetylation pharmacogenetics. *Pharmacol. Rev.* **1985**, *37*, 25–79.

5.   Upton, A.; Johnson, N.; Sandy, J.; Sim, E. Arylamine N-acetyltransferases - of mice, men and microorganisms. *Trends Pharmacol. Sci.* **2001**, *22*, 140–6.

6.   Boukouvala, S.; Fakis, G. Arylamine N -Acetyltransferases: What We Learn from Genes and Genomes. *Drug Metab. Rev.* **2005**, *37*, 511–564, doi:10.1080/03602530500251204.

7.   Hein, D. W.; Doll, M. A.; Rustan, T. D.; Gray, K.; Feng, Y.; Ferguson, R. J.; Grant, D. M. Metabolic activation and deactivation of arylamine carcinogens by recombinant human NAT1 and polymorphic NAT2 acetyltransferases. *Carcinogenesis* **1993**, *14*, 1633–8.

8.   Hein, D. W. Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis. *Mutat. Res.* **2002**, *506–507*, 65–77.

9.   Sim, E.; Pinter, K.; Mushtaq, A.; Upton, A.; Sandy, J.; Bhakta, S.; Noble, M. Arylamine N-acetyltransferases: a pharmacogenomic approach to drug metabolism and endogenous function. *Biochem. Soc. Trans.* **2003**, *31*, 615–619, doi:10.1042/bst0310615.

10.   Bhakta, S.; Besra, G. S.; Upton, A. M.; Parish, T.; Sholto-Douglas-Vernon, C.; Gibson, K. J. C.; Knutton, S.; Gordon, S.; daSilva, R. P.; Anderton, M. C.; Sim, E. Arylamine *N* -Acetyltransferase Is Required for Synthesis of Mycolic Acids and Complex Lipids in *Mycobacterium bovis* BCG and Represents a Novel Drug Target. *J. Exp. Med.* **2004**, *199*, 1191–1199, doi:10.1084/jem.20031956.

11.   Wu, H.; Dombrovsky, L.; Tempel, W.; Martin, F.; Loppnau, P.; Goodfellow, G. H.; Grant, D. M.; Plotnikov, A. N. Structural Basis of Substrate-binding Specificity of Human Arylamine N-Acetyltransferases. *J. Biol. Chem.* **2007**, *282*, 30189–30197, doi:10.1074/jbc.M704138200.

12.   Butcher, N. J.; Minchin, R. F. Arylamine N-Acetyltransferase 1: A Novel Drug Target in Cancer Development. *Pharmacol. Rev.* **2012**, *64*, 147–165, doi:10.1124/pr.110.004275.

13.   Sim, E.; Abuhammad, A.; Ryan, A. Arylamine N-acetyltransferases: from drug metabolism and pharmacogenetics to drug discovery. *Br. J. Pharmacol.* **2014**, *171*, 2705–2725, doi:10.1111/bph.12598.

14.   Duval, R.; Xu, X.; Bui, L.-C.; Mathieu, C.; Petit, E.; Cariou, K.; Dodd, R. H.; Dupret, J.-M.; Fernando, R.-L. Identification of cancer chemopreventive isothiocyanates as direct inhibitors of the arylamine N-acetyltransferase-dependent acetylation and bioactivation of aromatic amine carcinogens. *Oncotarget* **2016**, doi:10.18632/oncotarget.7086.

15.   Francis, S.; Laurieri, N.; Nwokocha, C.; Delgoda, R. Treatment of Rats with Apocynin Has Considerable Inhibitory Effects on Arylamine N-Acetyltransferase Activity in the Liver. *Sci. Rep.* **2016**, *6*, 26906, doi:10.1038/srep26906.

16.   Anderton, M. C.; Bhakta, S.; Besra, G. S.; Jeavons, P.; Eltis, L. D.; Sim, E. Characterization of the putative operon containing arylamine N-acetyltransferase (nat) in Mycobacterium bovis BCG. *Mol. Microbiol.* **2006**, *59*, 181–192, doi:10.1111/j.1365-2958.2005.04945.x.

17.   Yam, K. C.; D'Angelo, I.; Kalscheuer, R.; Zhu, H.; Wang, J.-X.; Snieckus, V.; Ly, L. H.; Converse, P. J.; Jacobs, W. R.; Strynadka, N.; Eltis, L. D. Studies of a Ring-Cleaving Dioxygenase Illuminate the Role of Cholesterol Metabolism in the Pathogenesis of Mycobacterium tuberculosis. *PLoS Pathog.* **2009**, *5*, e1000344, doi:10.1371/journal.ppat.1000344.

18.   Abuhammad, A.; Fullam, E.; Lowe, E. D.; Staunton, D.; Kawamura, A.; Westwood, I. M.; Bhakta, S.; Garner, A. C.; Wilson, D. L.; Seden, P. T.; Davies, S. G.; Russell, A. J.; Garman, E. F.; Sim, E. Piperidinols That Show Anti-Tubercular Activity as Inhibitors of Arylamine N-Acetyltransferase: An Essential Enzyme for Mycobacterial Survival Inside Macrophages. *PLoS One* **2012**, *7*, e52790, doi:10.1371/journal.pone.0052790.

19.   Abuhammad, A.; Fullam, E.; Bhakta, S.; Russell, A.; Morris, G.; Finn, P.; Sim, E. Exploration of Piperidinols as Potential Antitubercular Agents. *Molecules* **2014**, *19*, 16274–16290, doi:10.3390/molecules191016274.

20.   Sandy, J.; Mushtaq, A.; Holton, S. J.; Schartau, P.; Noble, M. E. M.; Sim, E. Investigation of the catalytic triad of arylamine N-acetyltransferases: essential residues required for acetyl transfer to arylamines. *Biochem. J.* **2005**, *390*, 115–123, doi:10.1042/BJ20050277.

21.   Sim, E.; Walters, K.; Boukouvala, S. Arylamine N-acetyltransferases: From Structure to Function. *Drug Metab. Rev.* **2008**, *40*, 479–510, doi:10.1080/03602530802186603.

22.   Zhang, N.; Walters, K. J. Insights into how protein dynamics affects arylamine N-acetyltransferase catalysis. *Biochem. Biophys. Res. Commun.* **2009**, *385*, 395–401, doi:10.1016/j.bbrc.2009.05.065.

23.   Frey, J. G.; Bird, C. L. Web-based services for drug design and discovery. *Expert Opin. Drug Discov.* **2011**, *6*, 885–895, doi:10.1517/17460441.2011.598924.

24.   Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.;

Fortier, A.; Gasteiger, E.; Grosdidier, A.; Hernandez, C.; Ioannidis, V.; Kuznetsov, D.; Liechti, R.; Moretti, S.; Mostaguir, K.; Redaschi, N.; Rossier, G.; Xenarios, I.; Stockinger, H. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603, doi:10.1093/nar/gks400.

25.   Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B.; Zhou, Y.; Wishart, D. S. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **2014**, *42*, D1091–D1097, doi:10.1093/nar/gkt1068.

26.   Lagorce, D.; Sperandio, O.; Baell, J. B.; Miteva, M. A.; Villoutreix, B. O. FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res.* **2015**, *43*, W200–W207, doi:10.1093/nar/gkv353.

27.   Sterling, T.; Irwin, J. J. ZINC 15: Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337, doi:10.1021/acs.jcim.5b00559.

28.   Labbé, C. M.; Rey, J.; Lagorce, D.; Vavruša, M.; Becot, J.; Sperandio, O.; Villoutreix, B. O.; Tufféry, P.; Miteva, M. A. MTiOpenScreen: A web server for structure-based virtual screening. *Nucleic Acids Res.* **2015**, *43*, W448–W454, doi:10.1093/nar/gkv306.

29.   DeLano, W. L. The case for open-source software in drug discovery. *Drug Discov. Today* **2005**, *10*, 213–217, doi:10.1016/S1359-6446(04)03363-X.

30.   Geldenhuys, W. J.; Gaasch, K. E.; Watson, M.; Allen, D. D.; Van der Schyf, C. J. Optimizing the use of open-source software applications in drug discovery. *Drug Discov. Today* **2006**, *11*, 127–132, doi:10.1016/S1359-6446(05)03692-5.

31.   Villoutreix, B. O.; Renault, N.; Lagorce, D.; Sperandio, O.; Montes, M.; Miteva, M. A. Free resources to assist structure-based virtual ligand screening experiments. *Curr. Protein Pept. Sci.* **2007**, *8*, 381–411.

32.   Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791, doi:10.1002/jcc.21256.

33.   Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31*, 455–461, doi:10.1002/jcc.21334.

34.   Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854, doi:10.1093/bioinformatics/btt055.

35.   Dallakyan, S.; Olson, A. J. Small-Molecule Library Screening by Docking with PyRx. In; 2015; pp. 243–250.

36.   Jayashree, S.; Thomas, C.; Piotr, C.; Peter, K.; Case, D. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409, doi:10.1021/ja981844+.

37.   Homeyer, N.; Gohlke, H. Free energy calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area method. *Mol. Inform.* **2012**, *31*, 114–122, doi:10.1002/minf.201100135.

38.   Kumari, R.; Kumar, R.; Lynn, A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951–1962, doi:10.1021/ci500020m.

39.   Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242, doi:10.1093/nar/28.1.235.

40.   Rose, P. W.; Prli´c, A.; Altunkaya, A.; Bi, C.; Bradley, A. R.; Christie, C. H.; Di Costanzo, L.; Duarte, J. M.;

Dutta, S.; Feng, Z.; Green, R. K.; Goodsell, D. S.; Hudson, B.; Kalro, T.; Lowe, R.; Peisach, E.; Randle, C.; Rose, A. S.; Shao, C.; Tao, Y. P.; Valasatava, Y.; Voigt, M.; Westbrook, J. D.; Woo, J.; Yang, H.; Young, J. Y.; Zardecki, C.; Berman, H. M.; Burley, S. K. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2017**, *45*, D271–D281, doi:10.1093/nar/gkw1000.

41. Tempel, W., Wu, H., Dombrovski, L., Loppnau, P., Weigelt, J., Sundstrom, M., Arrowsmith, C.H., Edwards, A.M., Grant, D.M., Bochkarev, A., Plotnikov, A. N. The Crystal Structure of Human N-acetyltransferase 1 mutant, F125S Available online: http://dx.doi.org/10.2210/pdb2ija/pdb.

42. Walraven, J. M.; Trent, J. O.; Hein, D. W. Computational and Experimental Analyses of Mammalian Arylamine N-Acetyltransferase Structure and Function. *Drug Metab. Dispos.* **2007**, *35*, 1001–1007, doi:10.1124/dmd.107.015040.

43. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127–34, doi:10.1093/nar/25.24.4940.

44. Lirio, S. B.; Macabeo, A. P. G.; Paragas, E. M.; Knorn, M.; Kohls, P.; Franzblau, S. G.; Wang, Y.; Aguinaldo, M. A. M. Antitubercular constituents from Premna odorata Blanco. *J. Ethnopharmacol.* **2014**, *154*, 471–474, doi:10.1016/j.jep.2014.04.015.

45. Chen, C. Y.-C. TCM Database@Taiwan: The World's Largest Traditional Chinese Medicine Database for Drug Screening In Silico. *PLoS One* **2011**, *6*, e15939, doi:10.1371/journal.pone.0015939.

46. Neron, B.; Menager, H.; Maufrais, C.; Joly, N.; Maupetit, J.; Letort, S.; Carrere, S.; Tuffery, P.; Letondal, C. Mobyle: a new full web bioinformatics framework. *Bioinformatics* **2009**, *25*, 3005–3011, doi:10.1093/bioinformatics/btp493.

47. Standardizer 2016.

48. Pencheva, T.; Lagorce, D.; Pajeva, I.; Villoutreix, B. O.; Miteva, M. A. AMMOS software: Method and application. In *Methods in Molecular Biology*; 2012; Vol. 819, pp. 127–141 ISBN 9781617794643.

49. Miteva, M. A.; Violas, S.; Montes, M.; Gomez, D.; Tuffery, P.; Villoutreix, B. O. FAF-Drugs: free ADME/tox filtering of compound collections. *Nucleic Acids Res.* **2006**, *34*, W738–W744, doi:10.1093/nar/gkl065.

50. Tempel, W., Wu, H., Dombrovski, L., Loppnau, P., Weigelt, J., Sundstrom, M., Arrowsmith, C.H., Edwards, A.M., Grant, D.M., Bochkarev, A., Plotnikov, A. N. The Crystal Structure of Human N-acetyltransferase 2 in complex with CoA Available online: http://dx.doi.org/10.2210/pdb2pfr/pdb.

51. Abuhammad, A.; Lowe, E. D.; Mcdonough, M. A.; Shaw Stewart, P. D.; Kolek, S. A.; Sim, E.; Garman, E. F. Structure of arylamine N-acetyltransferase from Mycobacterium tuberculosis determined by cross-seeding with the homologous protein from M. marinum: Triumph over adversity. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 1433–1446, doi:10.1107/S0907444913015126.

52. Uniprot UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169, doi:10.1093/nar/gkw1099.

53. Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **2014**, *42*, W252–W258, doi:10.1093/nar/gku340.

54. Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity?a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228, doi:10.1016/0040-4020(80)80168-2.

55. Sanner, M. F. Python: a programming language for software integration and development. *J. Mol. Graph. Model.* **1999**, *17*, 57–61.

56.   Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662, doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

57.   Solis, F. J.; Wets, R. J.-B. Minimization by Random Search Techniques. *Math. Oper. Res.* **1981**, *6*, 19–30, doi:10.1287/moor.6.1.19.

58.   Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037, doi:10.1021/ct200196m.

59.   Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *J. Comput. Aided. Mol. Des.* **2014**, *28*, 221–233, doi:10.1007/s10822-014-9713-7.

60.   Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676, doi:10.1002/jcc.20090.

61.   Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537–W541, doi:10.1093/nar/gks375.

62.   Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271, doi:10.1021/j100308a038.

63.   Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 14101, doi:10.1063/1.2408420.

64.   Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190, doi:10.1063/1.328693.

65.   Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076, doi:10.1080/00268978300102851.

66.   Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472, doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

67.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593, doi:10.1063/1.470117.

68.   Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, doi:10.1016/0263-7855(96)00018-5.

69.   Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612, doi:10.1002/jcc.20084.