Identification of Potent Inhibitors of ATP synthase subunit c (AtpE) from *Mycobacterium tuberculosis* using in silico Approach

Mustafa Alhaji Isa

Department of Microbiology, Faculty of Sciences, University of Maiduguri, P.M.B. 1069, Maiduguri, Nigeria
Corresponding author email: mustafaisa@unimaid.edu.ng

Abstract

ATP synthase subunit c (AtpE) is an enzyme that catalyzes the production of ATP from ADP in the presence of sodium or proton gradient from *Mycobacterium tuberculosis* (MTB). This enzyme considered an essential target for drug design and its shares the same pathway with the target of Isoniazid. Thus, this enzyme would serve as an alternative target of the Isoniazid. The 3D model structure of the AtpE was constructed based on the principle of the homology modeling using the Modeller9.16. The developed model was subjected to the energy minimization and refinement using molecular dynamic (MD) simulation. The minimized model structure was searched against Zinc and PubChem database to determine ligands that bind to the enzyme with minimum binding energy using RASPD and PyRx tool. A total of 4776 compounds capable of binding to AtpE with minimum binding energies were selected. These compounds further screened for physicochemical properties (Lipinski rule of five). All the compounds that possessed the desirable properties selected and used for molecular docking analysis. Five (5) compounds with minimum binding energies ranged between ─8.69, and ─8.44kcal/mol, less than the free binding energy of ATP were selected. These compound further screened for the absorption, distribution, metabolism, excretion, and toxicity (ADME and toxicity) properties. Of the five compounds, three (ZINC14732869, ZINC14742188, and ZINC12205447) fitted all the ADME and toxicity properties and subjected to MD simulation and Molecular Mechanics Generalized Born and Surface Area (MM-GBSA) analyses. The results indicated that the ligands formed relatively stable complexes and had free binding energies, less than the binding energy of the ATP. Therefore, these ligands considered as prospective inhibitors of MTB after successful experimental validation

Key Word: MTB, AtpE, ADME, Homology Modeling, and MD Simulation

Introduction

Tuberculosis (TB) is among the major infectious diseases that responsible for the mortality and morbidity worldwide. The disease has numerous available drugs for its treatment yet claims

the lives of countless individuals (Isa *et al.,* 2018; Jothieswari and Bhaskar Reddy, 2015). This is due to the endemic effects of multi-drug resistant TB (MDR-TB), extensively drug-resistant TB (XDR-TB), and total drug-resistant TB (TDR-TB), which pose a serious menace to the tuberculosis control program (WHO, 2015). India represents more than one-fourth of the world's TB cases and death (WHO, 2015). Therefore, it is essential to develop new antituberculosis drugs which can inhibit both actively multiplying bacilli and a non-growing persistent population of MTB to prevent reactivation of the infection. Isoniazid (INH) is the first-line antibiotic for the treatment of all types of TB caused by MTB. It was discovered 1912 and successfully used for the treatment of TB in 1951. The drug acts by inhibiting the cell wall synthesis of the organisms. Initially, the compound is a prodrug, but it would activate by the enzymes called Catalase-peroxidase encoded by the *katG* gene. The activated drug form complex with NADH, thus inhibit fatty acid synthesis of the bacterium, which is one of the significant components of the bacterium cell wall. Therefore, the drug has substantial early bactericidal action against quickly multiplying cells (Jindani *et al.,* 1980; Hafner *et al.,* 1997). Isoniazid, combined with rifampicin, has long been used for the treatment of TB. It is active against metabolically-active multiplying or replicating bacilli. INH resistance is the most frequently occurring in MTB at a rate of 10^{5-6} frequency in vitro as compared to other tuberculosis drugs. The two critical molecular mechanisms for the drug resistance in INH commonly attributed to a mutation in *katG* and *inhA* gene or its promoter region. Indeed, a lot of researchers have conducted which demonstrated that these two genes mainly associated with INH resistance. Among these genes, S315T in *katG* is the most commonly found and accounts for 50-90% of all the INH-resistant found in clinical isolates (Telenti, 1997; Hazbón *et al.,* 2006). This mutation is associated with higher levels of resistance at MIC > 1µg/mL to INH (Fenner *et al.,* 2012) and appeared consistently in MDR-TB strains (Hazbón *et al.,* 2006). This process followed by mutation in the promoter region of *inhA* which causes overexpression of *inhA* or a decrease in the binding affinity of the INH-NAD adduct (Rozwarski *et al.,* 1998). However, ATP synthase subunit c (AtpE) have the same pathway with the target of Isoniazid (Lee *et al.,* 2001; Johnson *et al.,* 2006; Cloete *et al.,* 2016). Thus, this enzyme would serve as an alternative target of the Isoniazid since it plays a vital role by providing ATP during the dormancy state of the MTB (Murphy and Brown, 2007). ATP synthase subunit c is an enzyme that catalyzes the production of ATP from ADP in the presence of sodium or proton gradient. The protein (F-type ATPases) has two structural domains F1 and F0 domains). F1 domain had the extramembranous catalytic core, while F0 domain had membrane proton channel joined together by the peripheral stalk and central stalk. The catalytic domain of F1 joins through a

rotary mechanism during the catalytic process to the central stalk sections of proton translocation. Residues between the 10-14 subunits serve as homomeric c-ring and form the central rotor element of the F1. F0 consists of the residues between 5-25 and 57 – 77 in its domain. However, ATP synthase subunit c is conserved in human beings, because of the subtle difference between human and bacterial make it an attractive target for drug design and development (Cloete *et al.,* 2016). Therefore, a drug inhibiting this enzyme must be specific and unique to bacteria, with less toxicity to human beings. These justify the reason for the selection of the enzyme mentioned above in this study.

Methods

Retrieval of Sequence and Search of Templates

The sequence of AtpE from MTB retrieved from the National Centre for Biotechnological Information (NCBI). The retrieved AtpE sequence was subjected to Basic Local Alignment Search Tool (BLASTP) (Altschul *et al.,* 1997) against the Protein Data Bank (PDB) (Bernstein *et al.,* 1977). This step was carried out to identify an appropriate template for the homology modeling. After identifying the suitable template, the AtpE sequence, and the template were aligned using ClustalW (Thompson *et al.,* 1994) to obtained sequence identity and similarity using default parameter.

Homology Modeling of AtpE

Homology modeling was used to build the three-dimensional structure (3D) of the AtpE based on the principle of spatial strain using Modeller9.16 (Pieper *et al.,* 2010). The process starts with the alignment of the target and the template sequence, then all the information of the template obtained during alignment was transferred to the target sequence. This information includes hydrogen bonds, main chain, side chain, and dihedral angle to build the 3D structure of AtpE. Ten (10) 3D model structure of the AtpE obtained and the one with least Discrete Optimized Protein Energy (DOPE) value was selected for energy minimization and evaluation.

Energy Minimization and Model Evaluation

The 3D model of AtpE with least Discrete Optimized Protein Energy (DOPE) value was selected and subjected to energy minimization and refinement using molecular dynamic simulation analysis for 10ns via AMBERTOOLS10 (Case *et al.,* 2015). The energy minimization and refinement was carried out to stabilize the model structure before to the molecular docking studies. The minimized model structure was superposed to the template to determine their deviation based on the root mean square deviation (RMSD) value. Finally. The

minimized model structure was assessed to determine the stereochemistry quality using the Ramachandran plot, ERRAT (Colovos and Yeates, 1993), and Verify_3D (Lüthy *et al.,* 1992).

Selection of Ligand for Docking Analysis

The model structure of AtpE complexed with ligand was submitted to RASPD tool (Mukherjee and Jayaram, 2013) to obtain ligands capable of binding to the protein. RASPD tool can identify ligands from Zinc database capable of binding to the protein complex with minimum binding energy. A total of four thousand seven hundred and seventy-six (4776) compounds were obtained and further validated their binding energies using PyRx program. The compounds with better binding energies were selected and screened for the Lipinski rule of five (Molecular weight (≤500 Da), Log P (≤5), hydrogen bond donor (≤5), and hydrogen bond acceptor (≤10)) to determine compounds with desirable physicochemical properties using DataWarrior program.

Molecular Docking Studies

Molecular docking studies were carried out to determine the interaction between the AtpE and the selected ligands using an AutoDock4.2 tool (Morris *et al.,* 1998). The process begins with the conversion of both AtpE and the ligands to the PDBQT type file, which contained all the information of the individual atoms such as atom type, partial charges, and the torsional degree of freedom. The gasteiger charges (Gasteiger and Marsili 1980) calculated and the free binding energy was determined using Lamarckian genetic algorithms. The x, y, and z coordinate of the AtpE set, and the grid map was fixed at 60 x 60 x 60 with a spacing of 0.375Å. Finally, the RMSD calculated, and the protein-ligand complex was analyzed using Pymol (Delano, 2002) and Ligplot+ program (Laskowski and Swindells 2011; Wallace *et al.,* 1996).

ADME and Toxicity Analysis

Absorption Distribution, Metabolism, Excretion, and Toxicity (ADME and Toxicity) analysis of all the selected ligands with good binding energies were carried out to determine compounds with desirable pharmacokinetic properties using ADME/TOX tool (Lipinski *et al.,* 2012; Veber *et al.,* 2002), AdmetSAR tool (Cheng *et al.,* 2012), and DataWarrior (Sander *et al.,* 2015). The pharmacokinetic properties predicted in this study include Human Intestinal Absorption (HIA), Blood-Brain Barrier (BBB) penetration, Cytochrome P450 (CYP450 2D6) Inhibitor, Plasma Protein Binding (PPB), Mutagenicity, Tumorigenicity, Irritation, and Reproduction. All the ligands with suitable pharmacokinetic properties selected for the MD simulation analysis.

Molecular Dynamic (MD) Simulation Analysis

The MD simulation analysis was carried out to determine the stability of the ligands at the binding pocket of the AtpE. In this study, all the ligands with the desirable ADME and toxicity properties were selected and subjected to MD simulation analysis using AMBERTOOLS10 package (Case *et al.,* 2015). During the process, the protonate 3D and the antechamber were used for the addition of explicit hydrogen and missing parameter to the ligands respectively. The force field of the protein assigned with GAFF, while the ligand force field assigned with ff12SB. The coordinate and topology file of the protein-ligand complex was constructed using the tleap component of the Amber tool. The complex system was neutralized using a buffer solution of 10Å within the octahedral box of TIP3P water. The system was minimized to remove structural artifact occurred during the model building. Also, further minimization of 2500 steps of steepest descent and 2500 steps of the conjugate gradient was carried out without any restrained. The initial and final temperature of 0 and 300k respectively were used for heating the system using Langevin dynamics temperature regulator. Finally, the production of the simulation performed at constant temperature and pressure of 300k and 1atm using the time step of 2fs. The analysis of the root mean square deviation (RMSD), root mean square fluctuation (RMSF), the radius of gyration were carried out using the PTRAJ component of the AMBERTOOL10. Also, Molecular Mechanics Generalized Born and Surface Area (MM-GBSA) analysis were carried out to determine the free binding energy of the protein-ligand complex using the MD simulation trajectory of the last 5ns.

Results and Discussion

Homology modeling

The sequence of AtpE was retrieved from NCBI with accession number NP_215821.1. The sequence was subjected to BLASTP against proteins with known 3D structures in the PDB. Six protein templates (4V1F, 4MJN, 3ZK1, 3V3C, 2WIE, and 2W5J) were selected based on the sequence identity, sequence similarity and high statistical significance (Less e-value). But, 4V1F further chosen due to high-resolution and used as a template for homology modeling. The result of pair sequence alignment between the template (4V1F) and the AtpE revealed that their sequences share 90.1% (73/81aa) sequence identity, 93.8% (76/81aa) sequence similarity, and 2.5% (2/81aa) gap (Figure 1). The presence of the high similarity between the AtpE and the template is a clear indication that their structures are highly conserved. It is therefore used

as a template to build the modeled structure of the AtpE. During the model building, all the coordinates of the template structure such as structurally variable regions (SVRs), structurally conserved regions (SCRs), and N, and C termini were all transferred to the AtpE via satisfaction of spatial restraints. All the main chain and side chain were set using rotamers in the modeled structure (Figure 2a). The process generated ten modeled structures of AtpE, and the ones with minimum Discrete Optimized Protein Energy (DOPE) were selected. The selected modeled structure was subjected to energy minimization (Van der Waals repulsion energy and steric clashes) using MD simulation with AMBERTOOL10. The final structure after minimization used for structural superimposition with the Cα trace of the template (4V1F). The RMSD between the Cα trace of the template and the modeled structure is 0.578Å. This RMSD value showed that the modeled structure was highly reliable resembling the template structure (Figure 2b). Analysis of the structural quality of the modeled structure, final minimized structure, and the template was carried out using PROCHECK. The program analyzes their stereochemistry quality based on phi and psi (Φ-ψ) plot, residue-by-residue geometry, overall structure geometry and G-factor. Also, Verify_3D and ERRAT Quality Factor for the template, modeled structure, and the final minimized structure were determined. ERRAT (Colovos and Yeates, 1993) was used to analyze and assess the statistics between different atom types based on non-bonded interaction. Verify_3D (Luethy *et al.,* 1992) was used to ascertain the compatibility of the 3D model with its amino acid sequence (ID) by giving a structural class to alpha, beta, loop, polar, nonpolar, etc. and comparing the output with proper structures. The result of the Ramachandran map of the modeled structure revealed that 91.8% of the entire residues were in the most favorable region [A, B, L]. Then followed by 6.6% in the additional allowed region [a, b, l, p], 1.6% and 0% in the generously allowed region [~a, ~b, ~l, ~p] and the disallowed region [XX] respectively. On the other hand, the minimized modeled structure and the template had 98.4% and 93.2% of all their residues in the most favorable regions [A, B, L]. However, 0% and 6.8% in additional allowed region [a, b, l, p] and 1.6% and 0% in the generously allowed region [~a, ~b, ~l, ~p] respectively (Table 4.1). Both the modeled and minimized modeled structures had excellent and acceptable quality with high structural integrity since a good quality model must have at least 90% in the most favorable region based on the analysis of 118 structures with 2Å resolution and ≤20 R factor. G-factor was used to measure how unusual the 3D structure was if the overall G-factor threshold was < -0.5, the structure was considered as unusual, and if it was < -1.0 regarded as highly unusual. Errat results indicated the percentage of non-bonded interactions between various atoms, in which the calculated error values fell below the rejection limit of 95%. The Errat score of the modeled

structure (before minimization) was 97.260%, while after minimization the score increased to 98.630%. It showed that the modeled structure became more stable after minimization (Table 1). Also, the number of steric clashes decreased from 62 to 49 after minimization. Similarly, the van der Waals repulsion energy also decreased from 63.8962kcal/mol to 30.4879kcal/mol. The decrease in the value of van der Waals repulsion energy in the minimized modeled structure accompanied by an increase in the stability (Table 1).

```
atpE      1  M--DPT--IAAGALIGGGLIMAGGAIGAGIGDGVAGNALISGVARQPEAQGRLFTPFFIT
4V1F_A    1  MELDPNALITAGALIGGGLIMGGGAIGAGIGDGIAGNALISGIARQPEAQGRLFTPFFIT


atpE     57  VGLVEAAYFINLAFMALFVFATPV-K
4V1F_A   61  VGLVEAAYFINLAFMALFVFATPGLQ
```

Figure 1: Pair sequence alignment between AtpE and the Mycobacterial ATP synthase rotor ring showing the conserved regions in the black color
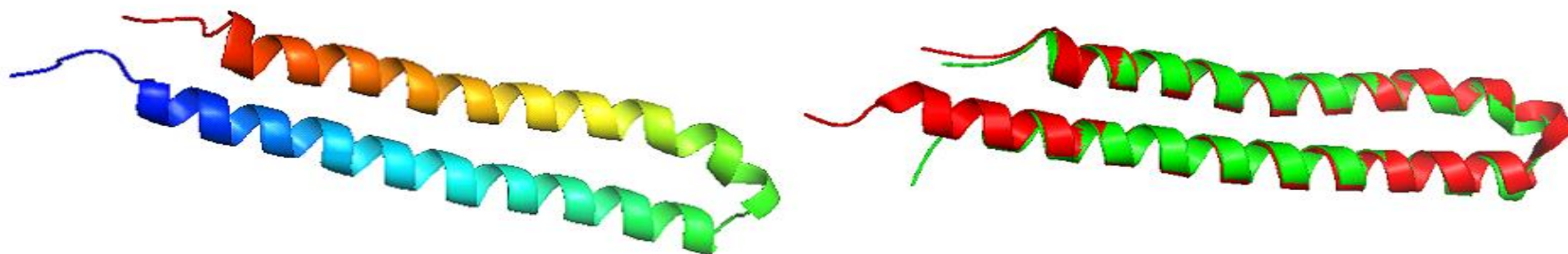
Figure 2: The 3D modeled structure of AtpE and the structural superimposition of the modeled structure and the template. **(a)** The Modeled structure of AtpE **(b)** Structural superimposition of the Cα traces of template (4V1F) (red) and the modeled structure (green) with the RMSD of 0.121Å

Table 1: Structural evaluation of the template (4V1F), modeled structure of AtpE and the minimized modeled structure of AtpE

Proteins	Procheck				ERRAT Quality Factor (%)	Verify_3D (%)	Total Number of steric clashes	VDW repulsion energy (kcal/mol)
	Most favored	Additional allowed	Generously allowed	Disallowed				
AtpE	91.80	6.60	1.60	0.0	97.26	33.33	62	63.90
Minimized AtpE	98.40	0.00	1.60	0.0	98.63	34.57	49	30.49
4V1F	93.20	6.80	0.00	0.0	99.14	23.26	49	30.49

Molecular Docking Analysis

Molecular docking is a technique used in biological testing of lead molecules. It helps in identifying compounds that have the potential to serve as drugs in the future. This process depends on the capability of compounds to bind to the macromolecule (mostly protein) with minimum binding energy and form a stable complex. RASPD and PyRx were used to screened compounds from the large compound database (Zinc and PubChem database). A total of 4776 compounds capable of binding to AtpE with minimum binding energies were selected. These compounds further screened for physicochemical properties: Molecular weight (≤500), number of hydrogen bond acceptor (≤10), number of hydrogen bond donor (≤5) and LogP (≤5)) and drug-likeness (Table 2). All the compounds that possessed the desirable properties were used for molecular docking analysis, to determine their binding energies with AtpE. Five (5) compounds found to yield minimum binding energies ranged between ─8.69, and ─8.44kcal/mol, less than the free binding energy of ATP (─2.19kcal/mol) (Figure 3). Based on the docking result, ZINC14732869 had the minimum binding energy of ─8.69kcal/mol and observed to interacted with AtpE via a hydrogen bond by accepting electrons from the hydroxyl group of Thr56 (distance=2.72Å). Also, it exhibited hydrophobic interactions with Ala34, Ile55, Leu35, Pro52, Phe53, Gly27, Ala31, Val30 and Leu49 (Table 3) (Figure 4a). All these residues mentioned above involved in both hydrogen and hydrophobic interactions formed the significant part of the catalytic domain (F0). Therefore, inhibition of these residues by the ligand would block the catalytic activity of the AtpE. This action would in turns affect the growth of the organism. Similarly, ZINC12079131 had the minimum binding energy of ─8.27kcal/mol and interacted and formed two hydrogen bonds with the hydroxyl group of Thr56 (distance$_1$ = 2.65Å and distance$_2$ = 2.90Å). It also presented hydrophobic interactions with Gly27, Val30, Ala31, Ala34, Leu35, Leu49, Pro52 and Phe53 (Figure 4b). ZINC14742188 exhibited minimum binding energy of ─8.63kcal/mol and interacted with AtpE and formed hydrogen bond by accepting electrons from the hydroxyl group of Thr56 (distance = 3.09Å). Also, it exhibited hydrophobic interactions with Gly27, Val30, Ala31, Ala34, Leu35, Leu49, Pro52, Ile55 and Phe53 (Figure 4c). Lastly, ZINC14531471 had a binding affinity of ─8.44kcal/mol and interacted and formed two hydrogen bonds with Asn33 (distance$_1$ = 3.19Å and distance$_2$ = 3.34Å). Besides, it presented hydrophobic interactions with Val30, Ala31, Ala34, Leu35, Leu49, Phe53, Ile55 and Thr56 (Figure 4e). All these residues mentioned above formed the catalytic domain (F1) of AtpE. Therefore, the interactions of the

9

ligands with these residues interfere with the growth of the MTB and subsequently led to its death (Table 3).

ADME and Toxicity Analyses

The result of the molecular docking analysis revealed that five compounds (ZINC14732869, ZINC12079131, ZINC14742188, ZINC12205447, and ZINC14531471) had good binding affinities. These compounds analyzed for ADMET properties such as Human Intestinal Absorption (HIA), Blood-Brain Barrier(BBB), Cytochrome P450 (CYP450 2D6) inhibitor, Aqueous Solubility, Plasma Protein Binding (PPB), Ames test, Carcinogens, Mutagenicity, Tumorigenic, Reproducibility, and Irritability (Table 4). Human Intestinal absorption and Blood-Brain Barrier of all the compounds were positive except ZINC14531471, which was contrary because it was unable to cross the Blood-Brain Barrier. Similarly, all the compounds were found to be non-inhibitors of CYP450 2D6 except ZINC14531471 which had an AC50 greater than 57μM (>57μM) based on the model calculation of Cheng *et al.* (2012). Concerning toxicity parameters (Mutagenicity, tumorigenic, reproducibility and irritability), the entire compound was predicted to be non-toxic except ZINC12079131 which was found to be highly irritating (Table 4). However, one compound (ZINC14531471) was found to be non-toxic, but was unable to cross the Blood-Brain Barrier and was also found to be a CYP450 2D6 inhibitor. Of the five compounds, only three (ZINC14732869, ZINC14742188, and ZINC12205447) fitted all the ADME and toxicity properties. Therefore, these compounds selected for the MD simulation analysis (Table 4).

Table 2: Molecular Properties and Drug-likeness of the selected ligands interacted with AtpE

| S/No.	Zinc Code	Molecular Weight	cLogP	H-bond Acceptors	H-bond Donors	Drug-likeness
1	ZINC14732869	499.68	0.85	7	1	7.40
2	ZINC12079131	490.63	2.11	8	1	-3.45
3	ZINC14742188	486.64	0.26	7	1	6.76
4	ZINC12205447	488.61	0.53	8	1	6.51
5	ZINC14531471	494.70	0.84	7	2	6.20
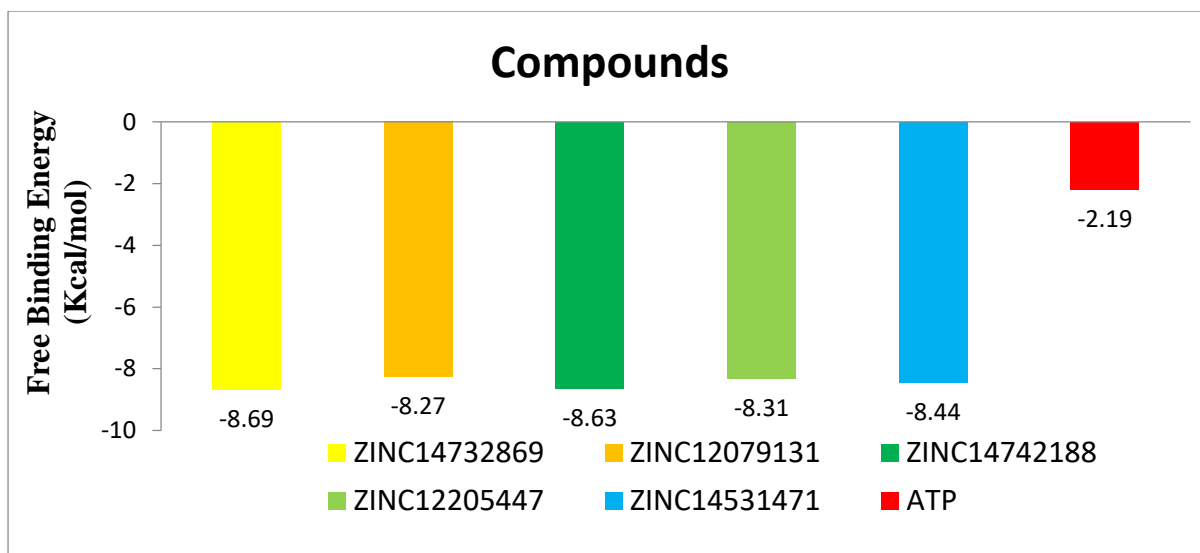
Figure 3: Distribution different free binding energies of the selected ligands Interacted with AtpE

Table 3: Docking score of ligands interacted with AtpE

| S/No.	Zinc Code	Minimum Free Energy of Binding (kcal/mol)	Interacting Residues	Distance (Å)	Residues involved in hydrophobic interaction
1.	ZINC14732869	−8.69	Thr56	2.72	Ala34, Ile55, Pro52, Phe53, Leu59, Val60, Gly23, Gly27, Ala31, Val130, Leu49, Leu35
2.	ZINC12079131	−8.27	Thr56 Thr56	2.65 2.90	Pro52, Ala34, Leu35, Leu49, Phe53, Leu59, Val60, Gly27, Val27, Val130, Ala31
3.	ZINC14742188	−8.63	Thr56	3.09	Leu35, Leu49, Phe53, Leu59, Val30, Gly23, Val60, Leu59, Ile55, Pro52, Ala31, Gly27
4.	ZINC12205447	−8.31	Thr56	2.73	Leu49, Ile55, Pro52, Ala31, Gly27, Leu59, Val160, Phe53, Leu35, Ala34
5.	ZINC14531471	−8.44	Asn33 Asn33	3.19 3.34	Leu59, Thr56, Val30, Phe53, Ala31, Ala34, Leu49, Leu35, Pro52, Ile55

(a) ZINC14732869
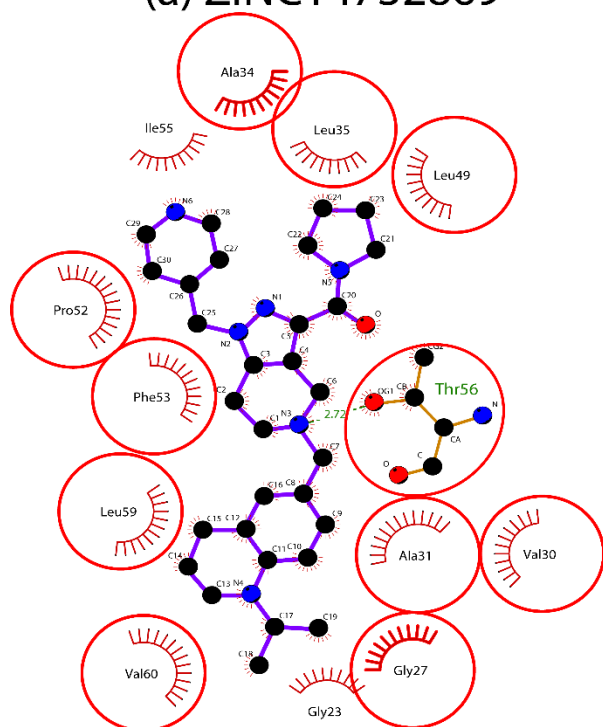
(b) ZINC12079131
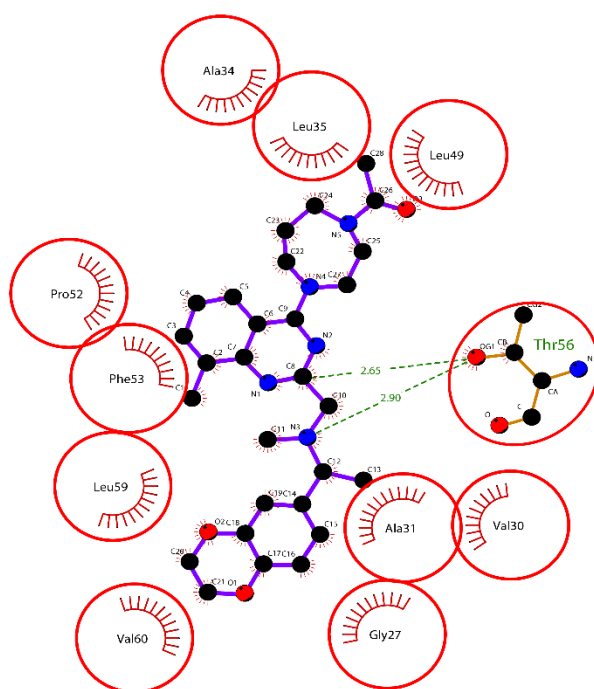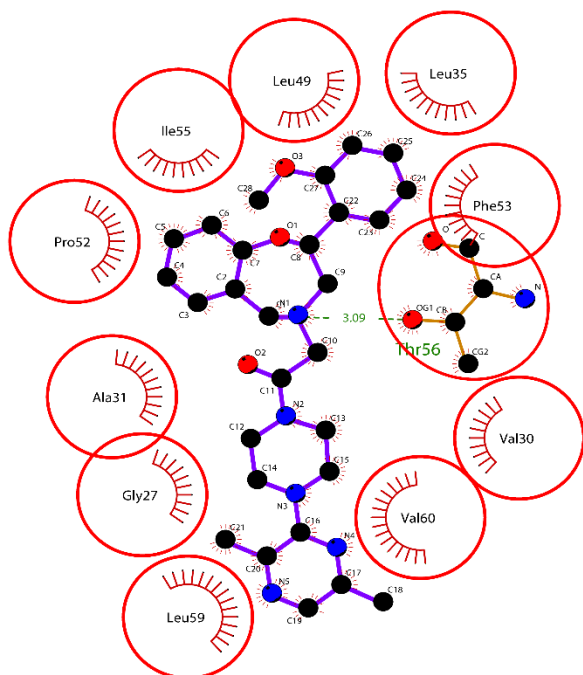
(c) ZINC14742188
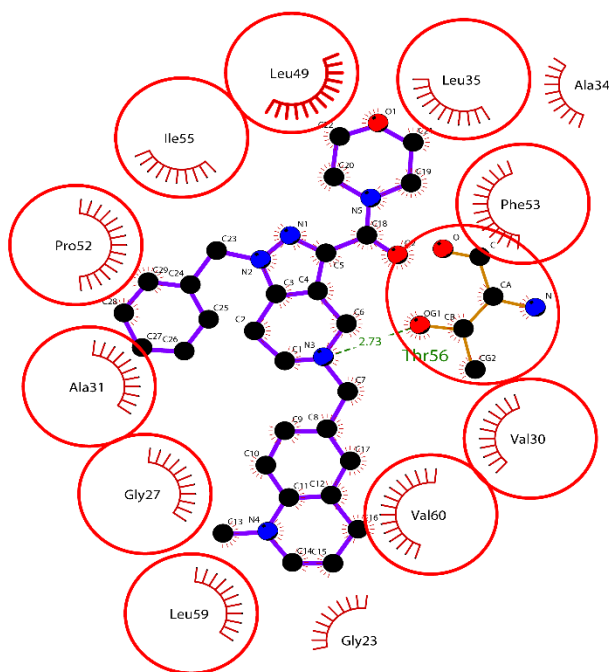
(d) ZINC12205447
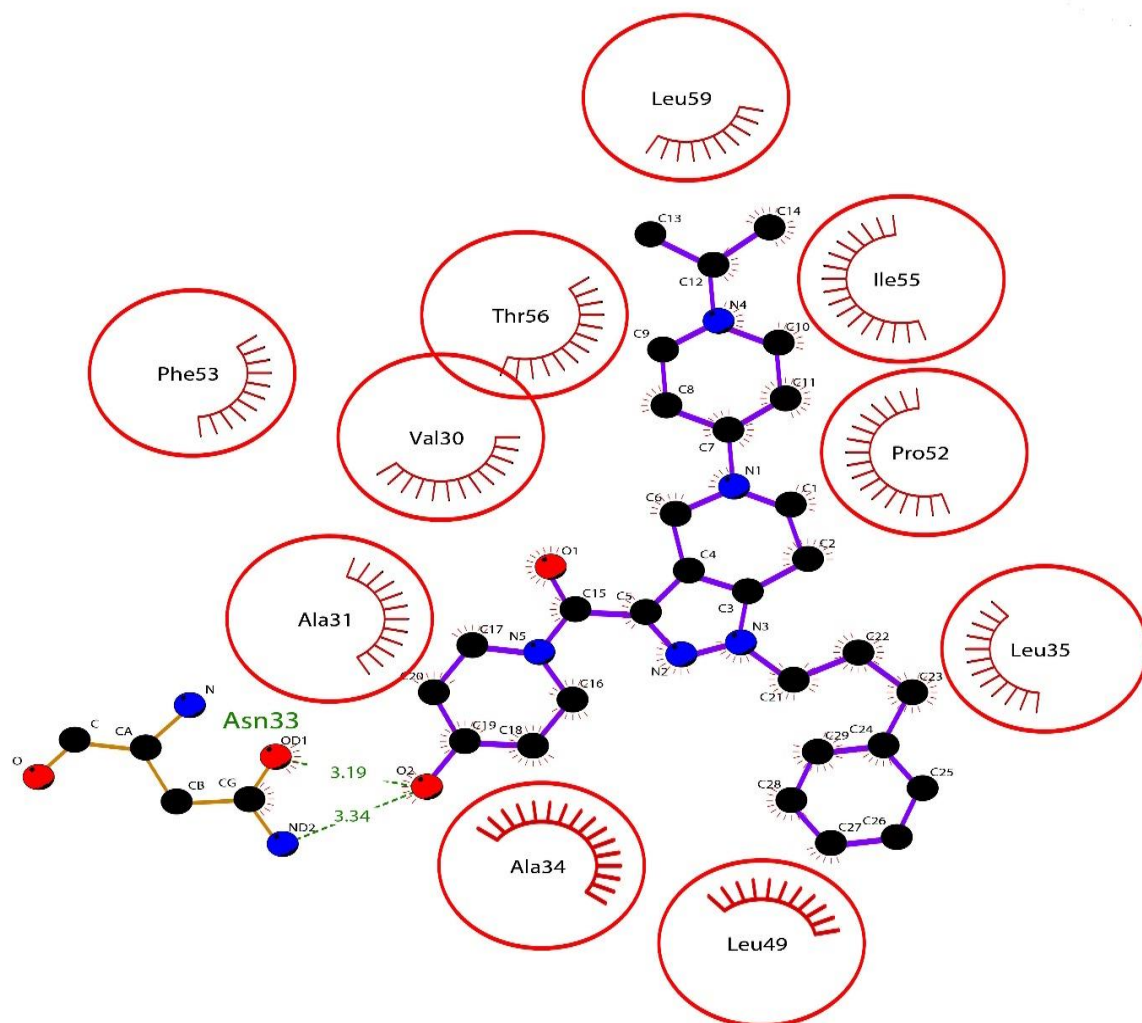
(e) ZINC14531471

Figure 4: Interactions of AtpE residues with the selected ligands (a) ZINC14732869 (b) ZINC12079131 (c) ZINC14742188 (d) ZINC12205447 (e) ZINC14531471

**Table 4:** ADME and Toxicity analyses of selected ligands interacted with AtpE

S/N	Compounds	HIA	BBB	CYP450 2D6 Inhibitor	PPB (%)	Aqueous Solubility	AMES Test	Carcinogens	Mutagenic	Tumorigenic	Reproducibility	Irritant
1	ZINC14732869	+	+	Non-inhibitor	71.07	-3.274	Non AMES toxic	Non-carcinogens	none	none	none	none
2	ZINC12079131	+	+	Non-inhibitor	44.78	-3.513	Non AMES toxic	Non-carcinogens	none	none	none	high
3	ZINC14742188	+	+	Non-inhibitor	65.45	-2.772	Non AMES toxic	Non-carcinogens	none	none	none	none
4	ZINC12205447	+	+	Non-inhibitor	44.60	-2.896	Non AMES toxic	Non-carcinogens	none	none	none	none
5	ZINC14531471	+	-	Inhibitor	11.87	-2.769	Non AMES toxic	Non-carcinogens	none	none	none	none

BBB= Blood-Brain Barrier, HIA = Human Intestinal Absorption, PPB = Plasma Protein Binding, Aqueous Solubility = Insoluble < -10 < Poorly soluble < -6 < Moderately soluble < -4 < Soluble < -2 < Very soluble < 0 < Highly soluble

**Molecular Dynamic Simulation Analysis**

From the results of the docking studies, five compounds with minimum binding energies were selected. These compounds further screened for ADME and toxicity properties. Three compounds (ZINC14732869, ZINC14742188, and ZINC12205447) were selected based on their pharmacokinetic properties. These compounds were subjected to the analysis of MD simulation to determine the structural stability of their complexes. Also, MD simulation of the AtpE complexed with ATP was carried out to compare its stability with the selected ligand. The stability of the four complexes (AtpeE─ZINC14732869, AtpE─ZINC14742188, AtpE─ZINC12205447, and AtpE─ATP) was ascertained by carefully examining the RMSD during the 50ns MD simulation. The deviations of the compound concerning their binding affinity as well as the motion of every residue within the complexes were determined based on the RMSF. The radius of gyration was also checked to determine the compactness of each complex based on the extent of how folded or unfolded the complex was. The AtpeE─ZINC14732869 complex equilibrated at 5ns and remained steady until fluctuated between 30 to 40ns, but later stabilized throughout the 50ns, with a mean value of 5.48±0.06414Å, maximum value of 9.77Å and minimum value of 1.28Å, higher than the mean RMSD value of AtpE─ATP complex (4.98±0.04894Å). The mean RMSD value of 5.48±0.06414Å suggested less flexibility probably due to interactions of the ligand with the flexible loop region of the AtpE, leading to high stability of the complex, although, it is less stable when compared to AtpE─ATP complex. Similarly, AtpE─ZINC14742188 and AtpE─ZINC12205447 complexes equilibrated and oscillated at 20ns with the average mean values of 7.6037±0.04346Å and 6.5452±0.06623, maximum values of 10.37Å and 10.22Å, and a minimum value of 1.72Å and 1.13Å respectively, higher than the mean RMSD value of AtpE─ATP complex (4.98±0.04894Å). The two (AtpE─ZINC14742188 and AtpE─ZINC12205447) complexes are less stable when compared with either AtpeE-ZINC14732869 or AtpE─ATP complex, although, it is not statistically significant ($p < 0.05$). However, both complexes had high RMSD values in most of their regions (Figure 5). The RMSF values of all the complexes (AtpeE─ZINC14732869, AtpE─ZINC14742188, and AtpE─ZINC12205447) shown in figure 6. These represent the extent of movement of the initial position of each residue and atoms in the AtpE and the ligands. In the AtpeE─ZINC14732869 complex, all the residues of N-terminal, ω-loop, and C-terminal had RMSF values less than or equal to 10Å (≤10Å), except residues between Met1─Ala6 and Val80─Lys81. The low RMSF values in all the regions of the complex occurred probably due

to the interaction of the ZINC14732869 to the flexible loop region of AtpE. This result further strengthened the RMSD result, where the complex had low RMSD values some of its regions. Similarly, the AtpE─ZINC14742188 complex had RMSF values of less than or equal to 10Å (≤10Å) in the N-terminal, C-terminal, and ω-loop regions except for residues between Met1─Ala7 and Ile36─Ser37, which had the RMSF value higher than 10Å (>10Å). These low RMSF values in the residues above were probably as a result of hydrophobic interactions between the residues and the ligand, which led to low RMSF values in the residues. Lastly, AtpE─ZINC12205447 complex residues located in N-terminal, C-terminal, and ω-loop regions had RMSF values of >10Å, except Met17, Ala18, Gly19, Gly20, Gly32, Asn33, Ala34, Leu35 and Gly38 which had value <10Å (Figure 6). The radius of gyration of the complexes was analyzed to determine whether they were stably folded after the 50ns MD simulation. All the complexes (AtpeE─ZINC14732869, AtpE─ZINC14742188, AtpE-ZINC12205447, and AtpE─ATP) fluctuated throughout the 50ns with a mean value of 18.57954 ± 0.03009Å, 17.7808 ± 0.03160Å, 17.4353 ± 0.03786Å, and 18.8310 ± 0.02508Å respectively. Therefore, all the complex structures had unfolded polypeptide structures (Figure 7).
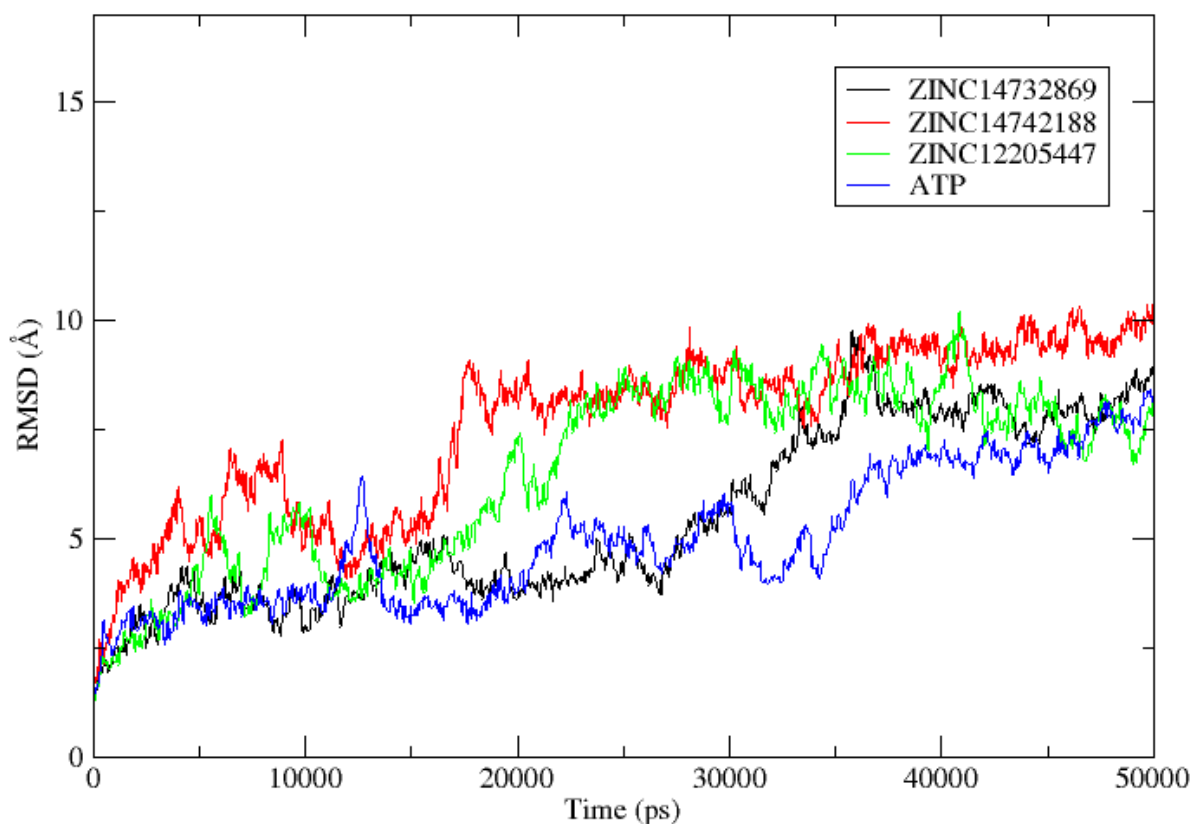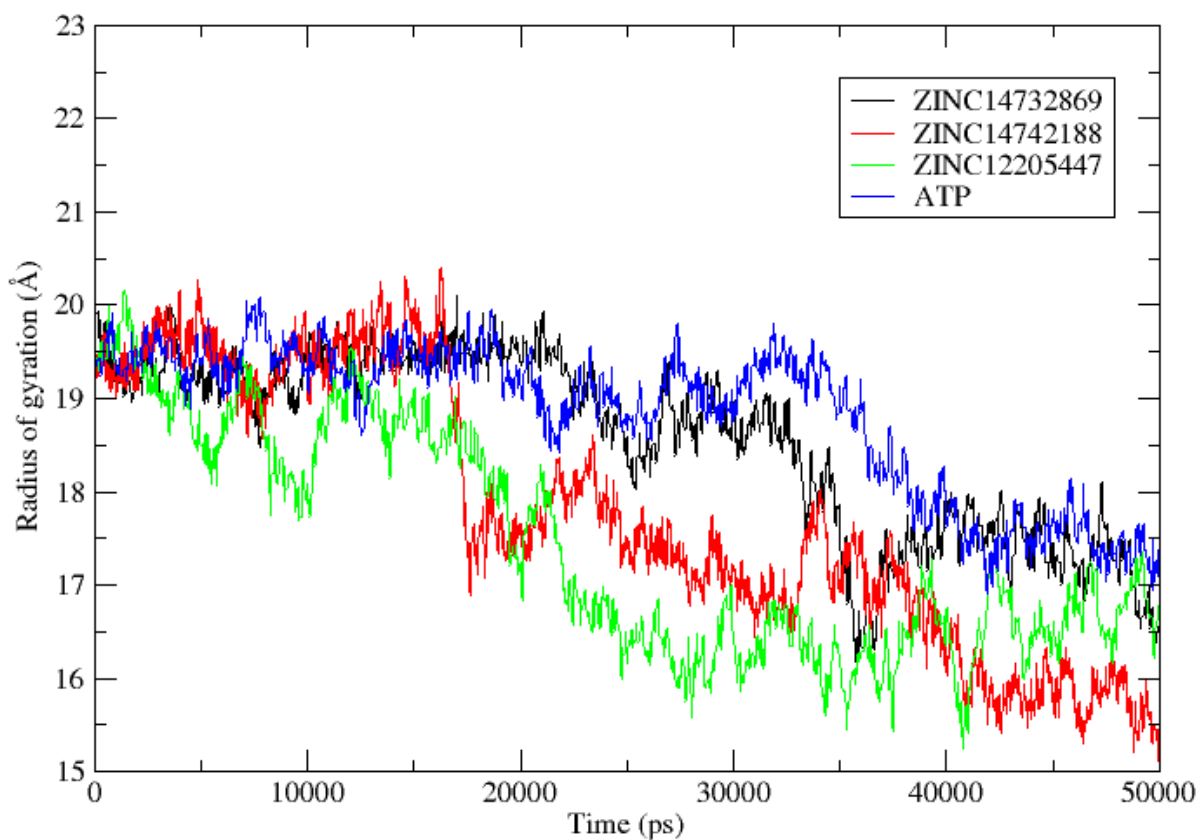

Figure 5: The MD simulation (RMSD analysis) of AtpeE─ZINC14732869, AtpE─ZINC14742188, AtpE─ZINC12205447, and AtpE─ATP complexes for 50ns
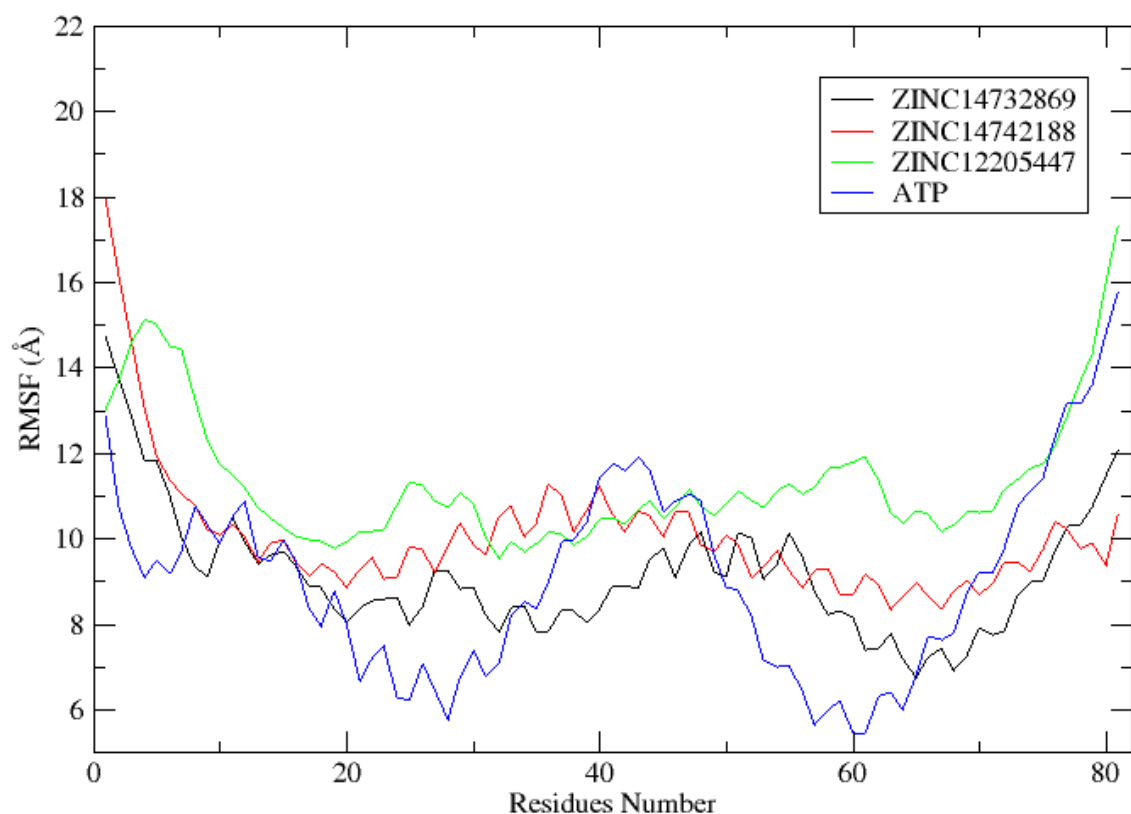
Figure 6: The MD simulation (RMSF analysis) of AtpeE─ZINC14732869, AtpE─ZINC14742188, AtpE─ZINC12205447, and AtpE─ATP complexes for 50ns

Figure 7: The MD simulation (Radius of gyration analysis) of AtpeE─ZINC14732869, AtpE─ZINC14742188, AtpE─ZINC12205447, and AtpE─ATP complexes for 50ns

17

MM-GBSA Analysis

The free binding energy of the protein-ligand complex was determine using MM-GBSA method implemented in Amber14. The analysis was carried out the using the average of 500 snapshots and at the interval of 10ns of last 5ns of the MD simulation trajectory. The energy used to determine the free binding energy include polar energy (G_{polar}), nonpolar ($G_{nonpolar}$) energy, van der Waals (E_{vdW}), and gas-phase electrostatic energy (E_{ele}), of the complexes. The result of the analysis showed that ZINC14742188 had less binding energy of ─26.13±0.3176kcal/mol, followed by ZINC14732869 (─24.47± 0.3135kca/mol) and ZINC12205447 (─22.76±0.3194kcal/mol), lower than the binding free energy of the ATP (─9.09±0.4646kcal/mol) (Table 5).

Table 5: MM-GBSA Analysis of the Selected Ligands

Compounds	ΔG$_{vdw}$	ΔG$_{ele}$	ΔG$_{polar}$	ΔG$_{nonpolar}$	ΔG$_{MM-GBSA}$
ZINC14732869	─31.59± 0.3970	─29.83± 0.7662	40.60± 0.8820	─3.65± 0.0209	**─24.47± 0.3135**
ZINC14742188	─31.47±0.3343	─3.37±1.3207	11.72±1.2036	─3.01±0.0240	**─26.13±0.3176**
ZINC12205447	─27.46±0.3254	─12.03±1.4011	19.55±1.3813	─2.82±0.0220	**─22.76±0.3194**
ATP	─26.79± 0.5005	─9.79±0.4811	32.73±0.4812	─5.24±0.0294	**─9.09±0.4646**

Conclusion

A total of four thousand seven hundred and seventy-six (4776) compounds were obtained and further validated their binding energies using PyRx program. The selected ligands were used for the molecular docking analysis to determine the binding energy between the AtpE and the ligands. Five (5) compounds with minimum binding energies ranged between ─8.69, and ─8.44kcal/mol, less than the free binding energy of ATP (─2.19kcal/mol) were obtained. The compounds were further filtered for the ADME and toxicity properties. Of the five compounds, only three (ZINC14732869, ZINC14742188, and ZINC12205447) fitted all the ADME and toxicity properties. These compounds were subjected to MD simulation and MM-GBSA analyses. The results of the analyses show that all the ligands formed relatively stable complexes and had free binding energies, less than the binding energy of the ATP. Therefore, these ligands considered as prospective inhibitors of MTB after successful experimental validation

Conflict of Interest

We have no conflict of interest

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, *25*(17), 3389-3402.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer Jr, E. F., Brice, M. D., Rodgers, J. R., ... & Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *European Journal of Biochemistry*, *80*(2), 319-324.

Case, D. A., J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke et al. "AMBER 2015." University of California, San Francisco (2015).

Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., ... & Tang, Y. (2012). admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties.

Cloete, R., Oppon, E., Murungi, E., Schubert, W. D., & Christoffels, A. (2016). Resistance-related metabolic pathways for drug target identification in MTB. *BMC Bioinformatics*, *17*(1), 75.

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*, *2*(9), 1511-1519.

DeLano, W. L. (2002). The PyMOL user's manual. *DeLano Scientific, San Carlos, CA*, *452*.

Fenner,L., Egger, M., Bodmer, T., Altpeter, E., Zwahlen, M., Jaton, K., ... & Siegrist, H. H. (2012). Effect of mutation and genetic background on drug resistance in MTB. *Antimicrobial agents and chemotherapy*, *56*(6), 3047-3053.

Hafner, R., Cohn, J. A., Wright, D. J., Dunlap, N. E., Egorin, M. J., Enama, M. E., ... & DATRI 008 Study Group. (1997). The early bactericidal activity of isoniazid in pulmonary tuberculosis: Optimization of methodology. *American journal of respiratory and critical care medicine*, *156*(3), 918-923.

Hazbón, M. H., Brimacombe, M., del Valle, M. B., Cavatore, M., Guerrero, M. I., Varma-Basil, M., ... & León, C. I. (2006). Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant MTB. *Antimicrobial agents and chemotherapy*, *50*(8), 2640-2649.

Gasteiger, Johann, and Mario Marsili. "Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges." Tetrahedron 36, no. 22 (1980): 3219-3228.

Isa, M. A., Majumdar, R. S., Haider, S., & Kandasamy, S. (2018). Molecular modelling and dynamic simulation of UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA) from Mycobacterium tuberculosis using in silico approach. *Informatics in Medicine Unlocked*, *12*, 56-66.

Jindani, A., Aber, V. R., Edwards, E. A., & Mitchison, D. A. (1980). The early bactericidal activity of drugs in patients with pulmonary tuberculosis. *American Review of Respiratory Disease*, *121*(6), 939-949.

Johnson, R., Streicher, E. M., Louw, G. E., Warren, R. M., Van Helden, P. D., & Victor, T. C. (2006). Drug resistance in MTB. *Current issues in molecular biology*, *8*(2), 97-112.

Jothieswari, D., & Bhaskar Reddy, K. (2015). Molecular Docking studies of potential chemical inhibitors on multi-drug resistance genes in Mycobacterium tuberculosis. *International journal of Innovative Drug Discovery*, *5*(1), 40-45.

Laskowski, R. A., Luscombe, N. M., Swindells, M. B., & Thornton, J. M. (1996). Protein clefts in molecular recognition and function. *Protein science: a publication of the Protein Society*, *5*(12), 2438.

Lee, A. S., Teo, A. S., & Wong, S. Y. (2001). Novel mutations in ndh in Isoniazid-resistant MTB Isolates. *Antimicrobial agents and chemotherapy*, *45*(7), 2157-2159.

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, *64*, 4-17.

Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*(6364), 83.

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, *19*(14), 1639-1662.

Mukherjee, G., & Jayaram, B. (2013). Rapid identification of hit molecules for target proteins via physicochemical descriptors. *Physical Chemistry Chemical Physics*, *15*(23), 9107-9116

Murphy, D. J., & Brown, J. R. (2007). Identification of gene targets against dormant phase MTB infections. *BMC infectious diseases*, *7*(1), 84.

Pieper, U., Webb, B. M., Barkan, D. T., Schneidman-Duhovny, D., Schlessinger, A., Braberg, H., ... & Datta, R. S. (2010). ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic acids research*, *39*(suppl_1), D465-D474.

Rozwarski, D. A., Grant, G. A., Barton, D. H., Jacobs, W. R., & Sacchettini, J. C. (1998). Modification of the NADH of the isoniazid target (InhA) from MTB. *Science*, *279*(5347), 98-102.

Sander, T., Freyss, J., von Korff, M., & Rufener, C. (2015). DataWarrior: an open-source program for chemistry aware data visualization and analysis. *Journal of chemical information and modeling*, *55*(2), 460-473.

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, *45*(12), 2615-2623.

Telenti, A., Philipp, W. J., Sreevatsan, S., Bernasconi, C., Stockbauer, K. E., Wieles, B., ... & Jacobs Jr, W. R. (1997). The emb operon, a gene cluster of MTB involved in resistance to ethambutol. *Nature medicine*, *3*(5), 567.

Thompson D. J., Higgins D. G. and Gibson T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (22): 4673–4680.

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1996). Derivation of 3D coordinate templates for searching structural databases: Application to Ser-His-Asp catalytic triads in the serine proteinases and lipases. *Protein Science*, *5*(6), 1001-1013.

World Health Organization. (2015). *Digital health for the End TB Strategy: an agenda for action* (No. WHO/HTM/TB/2015.21). World Health Organization (http://www.who.int/tb/areas-of-work/digital-health/Digital_health_EndTBstrategy.pdf, accessed 8 August 2016).