

Research Article

Ensemble-Based Virtual Screening of *Mycobacterium tuberculosis* ClpC1 Inhibitors

Hafsa Tayyab, Nor FW Ridzwan, Saharuddin B Mohamad*

Institute of Biological Sciences, Faculty of Science, University of Malaya, Jalan Profesor Diraja Ungku Aziz, 50603 Kuala Lumpur, Malaysia

***Corresponding author:** Saharuddin B Mohamad, Institute of Biological Sciences, Faculty of Science, University of Malaya, Jalan Profesor Diraja Ungku Aziz, 50603 Kuala Lumpur, Malaysia.

Received: 17 May 2022; **Accepted:** 31 May 2022; **Published:** 13 June 2022

Citation: Hafsa Tayyab, Nor FW Ridzwan, Saharuddin B Mohamad. Ensemble-Based Virtual Screening of *Mycobacterium tuberculosis* ClpC1 Inhibitors. Archives of Clinical and Biomedical Research 6 (2022): 503-516.

Abstract

Tuberculosis (TB) is an airborne transmissible disease caused by *Mycobacterium tuberculosis* (*Mtb*), responsible for 1.3 million deaths annually. Due to increased cases of drug resistance to *Mtb*, a new treatment regime for TB needs to be discovered. ClpC1 plays a crucial role in the protein homeostasis of *Mtb* thus, presents as a promising target in controlling TB infection. The present study aimed to identify potential inhibitors for the ClpC1 N-terminal domain (ClpC1-NTD) by applying the relaxed complex scheme in virtual screening that accounts for the target and ligand flexibility. A filtered library of natural product compounds was virtually screened against each of the selected ClpC1-NTD dominant conformations from the ensemble generated using molecular dynamics simulation. The promising compounds with the strong binding affinity to ClpC1 protein were then further analysed for their molecular interactions. The stability of the most potent compound was examined through a complex MD simulation while the pharmacokinetics properties were gathered using SwissADME and pkCSM. The results showed that ligand NP132 formed a strong and stable complex with good pharmacokinetics and toxicological profile.

Keywords: Caseinolytic protein C1; Molecular docking; Molecular dynamics simulation; Natural product compounds; Tuberculosis

1. Introduction

Tuberculosis (TB) is among the deadliest diseases in the world, and as stated in the World Health Organization (WHO) TB report, it accounts for 10 million new cases annually [1]. The current TB treatments may need to be replaced with novel approaches to tackle the increasing cases of drug resistance to the etiological agent for TB, *Mycobacterium tuberculosis* (*Mtb*) [2]. A promising target for anti-mycobacterial drug development is the caseinolytic protein C1 (ClpC1), a protein chaperone encoded by the *Mtb clpC1* gene [3]. It plays a pivotal role in *Mtb* survival and homeostasis by refolding damaged proteins during infection, thus preventing their degradation [4]. The N-terminal domain (NTD) of ClpC1 contains the binding domain that would trigger a proteolytic complex formation, and based on experimental evidence, ClpC1-NTD is essential for the growth and survival of *Mtb* [5]. Inhibition of ClpC1-NTD would disrupt the formation of the proteolytic complex, resulting in increased protein toxicity levels and causing mycobactericidal effects, significantly reducing the number of *Mtb* [6]. As ClpC1-NTD has been structurally and functionally characterised, this makes it an attractive target for drug design and discovery.

Through experimental means, several natural product compounds have been identified as potential inhibitors for ClpC1-NTD such as cyclomarin A, ecumicin, and rufomycin I. However, cyclomarin A and ecumicin displayed poor pharmacokinetic properties in mice models due to their large size and high flexibility [7]. Experimental identification of ClpC1-NTD inhibitors is expensive nonetheless, recent technological advances have allowed using computational means to ease the drug discovery process. Finding potential inhibitors using *in silico* analysis comes under Computer-Aided Drug Design (CADD) and has proven to be an efficient, cost-, and time-effective method [8, 9]. Among the CADD techniques, the relaxed complex scheme allows to computationally visualise the actual dynamics of a protein [10].

This study employed an ensemble-based virtual screening approach for the identification of potential inhibitors for ClpC1-NTD [11]. The protein is not a rigid structure as often treated by most molecular docking tools but is flexible [8]. Therefore, Molecular Dynamics (MD) simulation was utilised to ensure that the *in silico* analysis closely mimicked the natural condition where the protein and ligand are flexible when interacting with each other [12]. Investigation on the promising compound stability with the protein was carried out by a subsequent protein-ligand (complex) MD simulation to strengthen the findings.

2. Materials and Methods

2.1. Data Collection and Optimisation

A crystallographic structure (PDB ID: 6CN8) was selected from Protein Data Bank based on its high resolution. This structure was pre-processed using UCSF Chimera 1.13.1 (<https://www.cgl.ucsf.edu/chimera/>) [13] where the crystallised water, ligand, and ions were removed. Then the missing residues in the loop region were refined using Chimera MODELLER 9.21 [14] graphical interface, and the model with the lowest Normalised Discrete Optimised Protein Energy (zDOPE) score was chosen from the seven models that were generated. This model was validated through PROCHECK server (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) [15] to evaluate the stereochemical properties of the modelled protein. Furthermore, the ligands were obtained from Super Natural II (http://bioinf-applied.charite.de/supernatural_new/index.php?site=home), a natural product

compounds database. The Lipinski's rule of five criteria which constitutes that the molecular weight of compounds is less than 500 Dalton, lipophilicity (logP) is not greater than 5, there are not more than 10 hydrogen bonds (H-bonds) acceptors, and 5 H-bond donors [16] was applied to extract the ligands used in this study. The library also included a positive control (PubChem CID: 76871757) extracted from PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>).

2.2. Protein Molecular Dynamics (MD) Simulation

The modelled protein structure was subjected to an MD simulation using GROMACS 5.0.4 (<http://www.gromacs.org/>) [17] on an Intel® Core™ i7-4930K machine with 96 Gib memory (RAM) equipped with GeForce GTX780 Ti as the Graphics Processing Unit (GPU) and running on Ubuntu Linux package. This stage included the computational generation of protein ensemble and the selection of dominant conformations in the protein trajectory [11].

2.2.1. Conformational Ensemble Generation

According to a study by Nguyen et al. [18], the TIP3P water model is the best fit for the CHARMM force field. Therefore, during protein MD simulation the CHARMM36 force field [19] was used for topology generation. Then, the system was solvated using the TIP3P water model [20] in a cubic box with a minimum distance of 1 nm. Next, two Na⁺ ions were added to the system to neutralise the total net charge of -2 and followed by energy minimisation (steepest descent method for 50,000 steps) to remove steric clashes and inappropriate geometry [21]. The minimised system was subjected to the Number of Particles, Volume, and Temperature (NVT) equilibration run for 200 ps under constant temperature condition, and the Number of Particles, Pressure, and Temperature (NPT) equilibration run for 500 ps under constant pressure condition. Finally, the production run was conducted for a 100,000 ps time scale under constant temperature (300 K) and pressure (1 bar) using the Langevin thermostat and piston methods [22] respectively. The time step for the MD simulation was 2 fs, and the Particle Mesh Ewald (PME) method [23] was used to analyse the long-range electrostatic interactions. The 100,000 ps production run was carried out to monitor the conformational changes of protein and generated snapshots every 10 ps interval, which resulted in 10,001 structures, including the initial structure. The graphs for MD simulation results were generated using modules in GROMACS and GRACE 5.1.22 program (<https://plasma-gate.weizmann.ac.il/Grace/>) [24].

2.2.2. RMSD-Based Clustering

The protein trajectory was clustered based on RMSD using the GROMOS algorithm in GROMACS. A set of general guidelines from literature (<https://ctlee.github.io/BioChemCoRe-2018/clustering/>) were followed for determining the best RMSD cut-off value such as the number of total clusters should be less than 40, there should not be many clusters with only one member, and more than 90% of the trajectory should be represented in less than seven clusters. The RMSD cut-off value 1.70 Å, which fulfilled the above criteria was chosen and the representative conformations were separated for subsequent analyses.

2.3. Ensemble-Based Virtual Screening

The ligand library was screened against each of the selected protein conformations using AutoDock 4.2 (<http://autodock.scripps.edu/>) [25]. The respective charges and hydrogens were added to prepare the protein

conformations and ligands individually for molecular docking using AutoDockTools [25]. For each conformation, the grid box with dimensions 70×70×70 with spacing 0.375 Å, was set up on the binding site. Then grid box was centred on coordinates $x = -2.442$, $y = 6.599$, and $z = -6.208$. The dockings were carried out using a Lamarckian Genetic Algorithm (LGA) [26] with 10 search runs for initial virtual screening and 100 search runs for the re-scoring analysis. Parameters including population size, number of energy evaluations and generations, and mutation and crossover rates were left at default. The ligand with the strongest binding affinity with the protein was selected to further study the interactions using Chimera.

2.4. Complex MD Simulation

To determine the stability of the selected protein-ligand complex, another MD simulation was carried out. The complex MD simulation protocol was similar to the protein MD simulation carried out previously. The exception in the protocol is on the ligand topology, which was generated using an external tool SwissParam (<https://www.swissparam.ch/>) [27] and later combined with protein topology. Then the system was subjected to solvation, neutralisation with three Na⁺ ions, and energy minimisation. The NVT equilibration run duration was the same however, the NPT equilibration run was conducted for 1,000 ps under constant pressure condition instead of 500 ps. Production run and analysis of complex MD trajectory involved similar procedures.

2.5. ADMET Profile Generation

To investigate the behaviour of predicted potent compound and the positive control inside the human host, the *in silico* Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) profile was generated using SwissADME (<http://www.swissadme.ch/>) [28] and pkCSM (<http://biosig.unimelb.edu.au/pkcsml/>) [29] web-based applications. The Simplified Molecular-Input Line-Entry System (SMILES) was used as the input for SwissADME and pkCSM. These quick, accurate, and easy-to-use web servers provided information regarding the gastrointestinal absorption rate, blood-brain barrier permeability, inhibition of CYP3A4 (an important Cytochrome P450 [CYP] isoform), drug clearance from the body, and mutagenicity of the uploaded compounds.

3. Results and Discussion

3.1. ClpC1 Structure Selection and Ligand Library Construction

The crystal structure of *Mtb* ClpC1-NTD bound to rufomycin I (PDB ID: 6CN8) with a high resolution (1.40 Å) was selected for this *in silico* study. Subsequently, loop refinement was carried out for the six missing residues and the model (Model 1) with the lowest zDOPE score (-2.24) was chosen (Supplementary Table S1). As stated by Gruswitz et al. [30], a zDOPE score value of less than -1 would produce a reliable model. Furthermore, a Ramachandran plot (Supplementary Figure S1) was constructed for Model 1 and indicated that 100% residues were residing in the most favoured regions, thus making it a valid model. Simultaneously the ligand library was constructed based on Lipinski's rule of five criteria [16] for small oral drugs (Supplementary Table S2), and 150 Natural Product (NP) compounds were filtered. Additionally, rufomycin I (PubChem CID: 76871757) was selected as the positive control because it was reported to have a strong binding affinity to ClpC1 protein [7].

3.2. Protein Ensemble Generation MD Simulation and Clustering

An ensemble of protein conformations was generated over a 100,000 ps MD simulation run instead of using a rigid structure to avoid bias as ligands are known to favour some protein conformations over others and form stable complexes [12, 31]. The energy analysis (Supplementary Figure S2) showed that energy minimisation of the system was successful, as indicated by the steady convergence of the potential energy curve and the large negative value (-603,660 kJ/mol) obtained. The computational temperature (300 K) and pressure (1 bar) of the system respectively, were maintained over the entire protein MD simulation. The low total energy value (-466,582 kJ/mol) for the MD simulation run was also an indication that the system had reached equilibrium.

For the protein trajectory analysis, there were slightly high fluctuations observed at the beginning of the simulation until 20,000 ps in both RMSD and Radius of Gyration (Rg) plots (Figure. 1A and B respectively). However, the protein structure was compact before 20,000 ps, and the entire trajectory reached an equilibration state after 20,000 ps. As shown in Figure 1D, the visualized structures at specific time intervals indicated that the protein remained stable and compact. The RMSF plot (Figure 1C) provides information on the protein flexibility based on each of the residues [32]. As expected, the terminal loop regions displayed the highest flexibility. In addition, ClpC1 belongs to the AAA+ protein superfamily which is known to be highly dynamic and contains very flexible loop regions [33]. Contrarily, the RMSF values for the alpha-helices remained low throughout the protein MD simulation indicating well-structured regions [21].

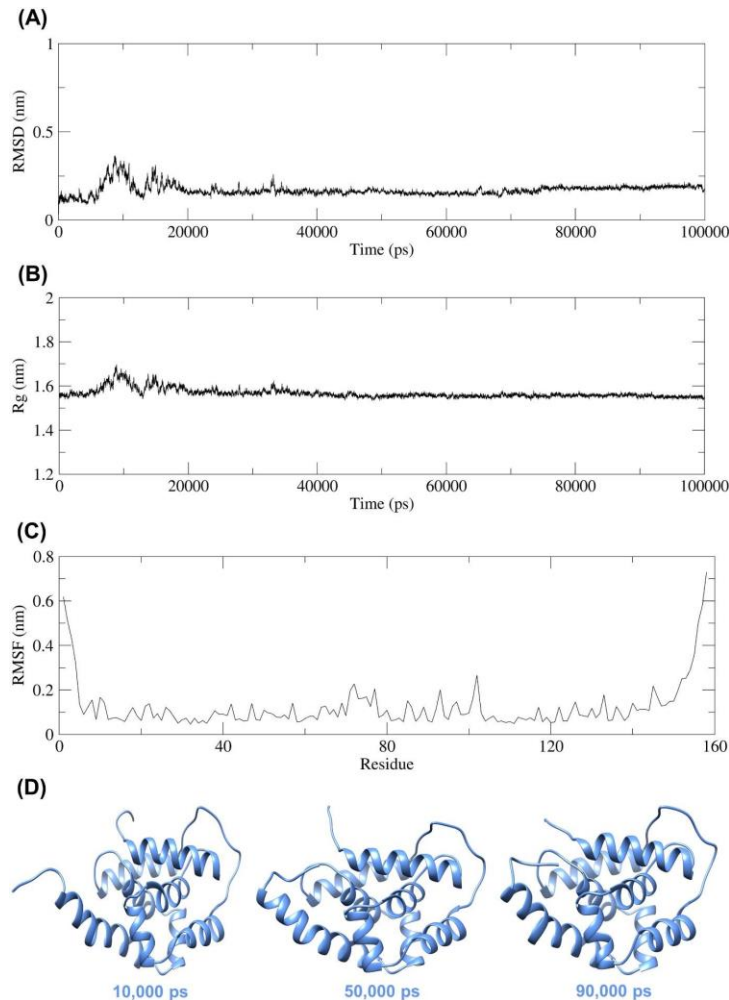


Figure 1: Trajectory analysis carried out to investigate stability and fluctuation during 100,000 ps MD simulation. (A) Protein C α atoms RMSD plot. (B) Protein Rg plot. (C) RMSF plot for protein residues. (D) Protein showing compact and stable structures during MD simulation at specific time intervals, generated using UCSF Chimera.

Finally, the protein trajectory was clustered using the RMSD-based (GROMOS) method since it would be computationally unfeasible to utilise each of the 10,001 protein conformations (including the starting structure) generated via MD simulation for the molecular docking analysis. The GROMOS method clusters similar conformations together based on the selected RMSD cut-off value [34]. Table 1 shows the 11 clusters produced from choosing the 1.70 Å RMSD cut-off value where the top three clusters represented 98% of the total protein conformations. Thus, the average (dominant) conformations from each of the three clusters were utilised in the subsequent virtual screening (VS) analysis.

Number of clusters	Total members	Identifier for the representative conformation in each cluster
Cluster 1	9,094	5,349 (conformation 1) ^a
Cluster 2	499	1,051 (conformation 2) ^b

Cluster 3	209	917 (conformation 3) ^c
Cluster 4	70	1,711
Cluster 5	62	876
Cluster 6	30	1,254
Cluster 7	17	3,307
Cluster 8	10	1,501
Cluster 9	6	1,401
Cluster 10	3	761
Cluster 11	1	886
^{a,b,c} the respective selected dominant conformations (1,2,3) to be utilised in the subsequent analyses.		

Table 1: Protein trajectory clustering analysis using the GROMOS method with 1.70 Å RMSD cut-off.

3.3. Ensemble-Based Virtual Screening Analysis and Interactions

The ligand library containing 150 NP compounds was virtually screened against each of the three dominant protein conformations independently using the molecular docking protocol for AutoDock 4.2 [25], resulting in 450 initial dockings with 10 search runs each. From the initial VS analysis, the protein-ligand complexes that had lowest binding energy values were separated (Supplementary Table S3) and then the top 10 ligands were selected based on their lowest binding energy values (Supplementary Table S4 and Figure 2), as that indicated the optimal binding pose [31, 35]. Furthermore, a re-scoring analysis (Table 2) was carried out to validate the results from the initial VS analysis, following the same protocol. Subsequently, the positive control (rufomycin I) and the top 10 ligands were docked against protein conformation 1, 2, and 3 respectively, resulting in 33 dockings with 100 search runs each. Out of the 10 selected ligands, compound NP132 (ID: SN00055391) with the strongest binding affinity (-9.20 kcal/mol) was identified as a potential inhibitor for ClpC1-NTD. As expected, rufomycin I also displayed a strong binding affinity (-14.59 kcal/mol) with the protein.

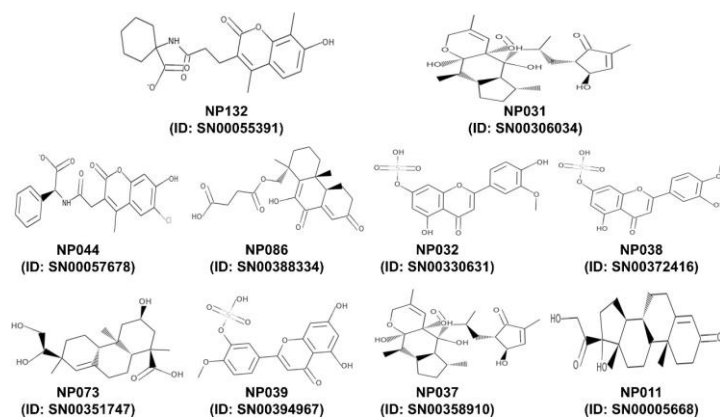


Figure 2: Top 10 ligands from initial virtual screening analysis identified based on lowest binding energy with their corresponding Super Natural II database IDs.

Ligands	Lowest binding energy (kcal/mol)			Average binding energy (kcal/mol)
	Conformation	Conformation	Conformation	
	1	2	3	
Rufomycin I	-10.58	-17.64	-15.54	-14.59 ^a
NP132	-8.15	-8.76	-10.68	-9.20 ^b
NP031	-8.38	-9.00	-9.71	-9.03
NP044	-8.08	-9.39	-9.48	-8.98
NP086	-7.96	-8.82	-10.04	-8.94
NP032	-8.88	-8.56	-9.31	-8.92
NP038	-8.84	-8.35	-9.50	-8.90
NP073	-8.44	-8.68	-9.55	-8.89
NP039	-8.33	-8.83	-9.12	-8.76
NP037	-7.88	-8.54	-9.51	-8.64
NP011	-7.83	-8.13	-8.86	-8.27
^{a,b} the lowest binding energy values indicating the most stable complexes.				

Table 2: The re-scoring results for the positive control and the top 10 ligands with their respective lowest binding energy values.

Further analysis was carried out to determine the interactions between the selected protein (conformation 1) and ligand (NP132) complex with the lowest binding energy. The lowest binding energy (more negative) value indicates that the protein and ligand complex is in the lowest energy state and thus closest to the native configuration [35]. Conformation 1 was selected, as it represented 9,094 of the total 10,001 conformations produced. The two protein residues (Arg-10 and Lys-85) were connected to the respective O atoms of NP132 with three H-bonds as represented in Table 3 and Figure 3 respectively. Experimental evidence suggested that cyclomarin A and ecumicin bind to Lys-85 [7], and ClpS adaptor protein which interacts with ClpC1-NTD, binds to Arg-10 [36]. This demonstrated that compound NP132 bounded to the same side of ClpC1-NTD as the known inhibitors cyclomarin A and ecumicin. Furthermore, the presence of H-bonds between ligand and the active site of protein could be an indication of inhibitory activity [37]. Moreover, Arg-10 and Lys-85 are within a highly conserved region in *Mycobacterium* species [38]. Conservation of Arg-10 and Lys-85 across bacterial species and H-bond interactions with the same residues suggests that they may be essential for survival and play a role in inhibitory behaviour [39, 40]. The bond lengths formed were also significant as shorter H-bond lengths meant stronger bonds [41]. Additionally, H-bonds formation indicated that the ligand remained stable in the binding site of protein [42].

Protein Residue Name-Position & Atom	Ligand Atom	H-bond Distance (Å)
Arg-10:1HH1	O	2.024
Lys-85:HZ3	O	2.152
Lys-85:HZ3	O	1.922

Table 3: The protein (conformation 1) and ligand (NP132) atoms involved in H-bonds formation and the respective distance between them.

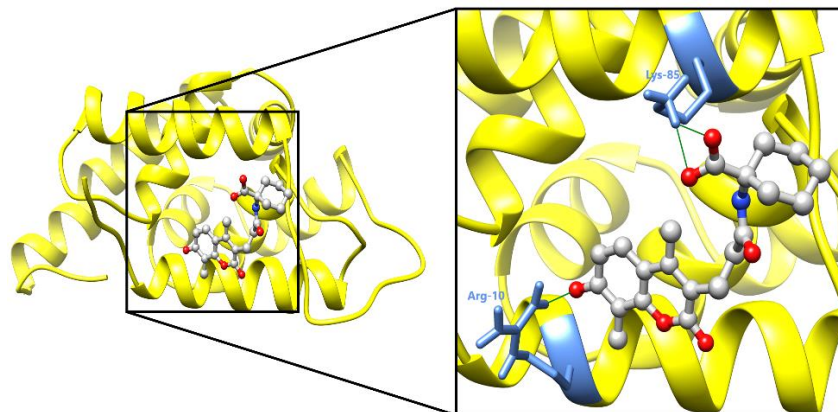


Figure 3: The protein (conformation 1) and ligand (NP132) complex with the lowest binding energy. The H-bonds (green lines) formation between conformation 1 residues (blue) and NP132 (ball and stick model), generated using UCSF Chimera.

3.4. Protein and Ligand Complex MD Simulation

The stability of the protein conformation 1 and ligand NP132 complex was further evaluated during a 100,000 ps MD simulation. The energy and trajectory analyses were analogous to the protein MD simulation, whereby the protein-ligand complex was successfully minimised (-745,200 kJ/mol), the temperature (300 K) and pressure (1 bar) remained constant, and the total energy (-532,200 kJ/mol) indicated that the system had attained equilibrium (Supplementary Figure S3). Furthermore, the MD trajectory showed an overall stable trend for the calculated RMSD value (Figure 4A) for the ligand atoms (purple). However, for the Carbon-alpha ($C\alpha$) backbone of the protein (black) and the protein-ligand complex (orange), RMSD values were ~0.50 nm at 70,000 ps. Since a smaller deviation indicates a more stable structure, the larger RMSD values could be due to the presence of NP132 disrupting the protein structure. Moreover, the Rg value for the protein (Figure 1B) was ~1.55 nm and for the complex (Figure 4C) was ~1.61 nm. A higher Rg value for the complex could indicate a lack of compactness [43] which may also be due to the presence of NP132. However, the Solvent-Accessible Surface Area (SASA) plot (Figure 4D) had a value of 91.06 for the complex, which indicated that the structure had remained intact. Additionally, the ligand did not dislodge from the protein but remained bounded with H-bonds (Figure 4E). Also, the farthest distance between the ligand and protein was recorded to be 0.25 nm. To conclude, the complex remained compact and stable throughout the MD simulation.

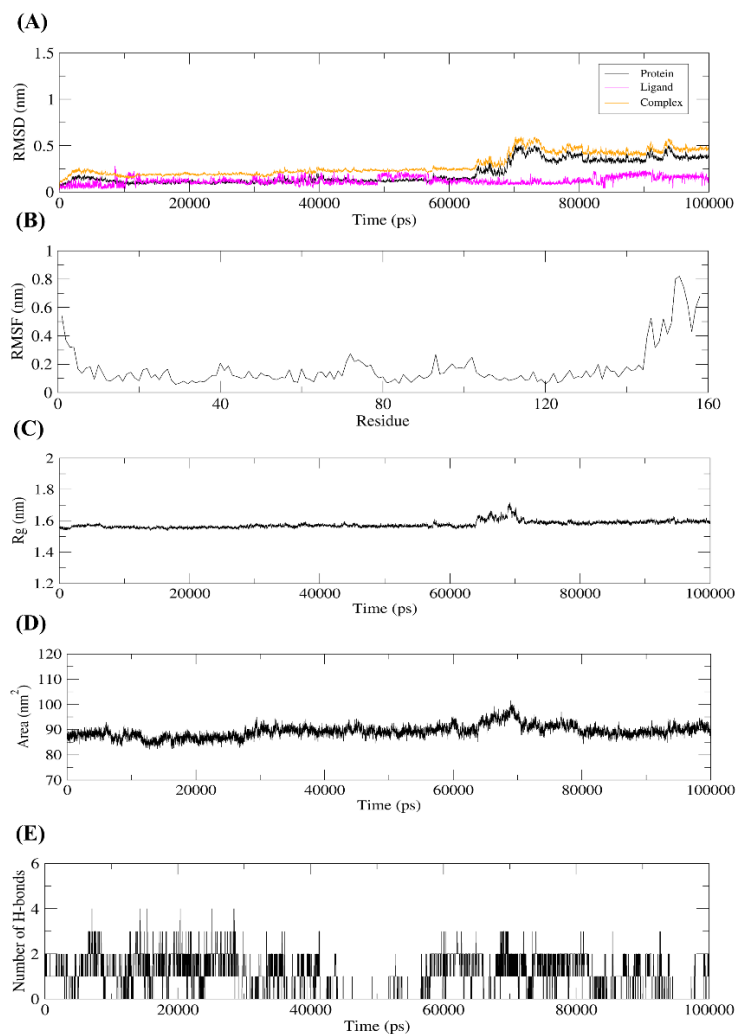


Figure 4: Trajectory analysis carried out to investigate stability and fluctuation during 100,000 ps MD complex simulation. (A) Protein, ligand, and complex RMSD plot. (B) Protein RMSF plot. (C) Protein-ligand complex Rg plot. (D) Protein-ligand complex SASA plot. (E) Protein-ligand complex H-bond pattern plot.

3.5. Physicochemical and Pharmacokinetic Properties of NP132 and Rufomycin I

Lipinski's rule of five constitutes the physicochemical properties of small drug-like compounds (Supplementary Table S5). Since NP132 followed Lipinski's rule of five, this indicates that the compound will have high solubility, permeability, and bioavailability [44]. Contrarily, rufomycin I violated three of Lipinski's rule of five as the molecular weight was greater than 500 Dalton, logP was also greater than 5, and there were more than 5 H-bond donors. In addition, NP132 was non-toxic, and rufomycin I exhibited low toxicity in humans. For a compound to have good oral bioavailability, following at least two Lipinski's rule of five is required [45]. Furthermore, for an oral drug, the uptake of a compound through the intestines is essential [46]. The ADMET profile (Table 4) indicated that the gastrointestinal absorption rate for the potent compound NP132 was high as opposed to rufomycin I, which could be attributed to its larger size. As stated by Wang and Craik [47], the larger the size (molecular weight) of a compound, the lower the absorption rate will be. The results further showed that NP132 and rufomycin I would not

permeate the blood-brain barrier. This physiological barrier prevents the therapeutic compounds in blood from crossing into the brain, leading to lower brain toxicity [48].

Parameters:	Compound	
	NP132	Rufomycin I
Gastrointestinal absorption	High	Low
Blood-brain barrier permeant	No	No
CYP3A4 inhibition	Non-inhibitor	Inhibitor
Total clearance (log ml/min/kg)	0.773	-0.081
AMES toxicity	Non-toxic	Non-toxic

Table 4: The *in silico* ADMET profile for compounds NP132 and rufomycin I using SwissADME and pkCSM.

The CYP3A4 inhibition was also studied as this is the most abundant of the CYP enzymes and is involved in the metabolism of many drugs [49]. Since NP132 was a non-inhibitor, it could be metabolised by the liver [50]. On the contrary, rufomycin I inhibited CYP3A4 probably due to its high logP value (5.1774) and large molecular weight (1012.20 Dalton) (Supplementary Table S5), which may cause severe side effects [51]. Additionally, the excretion rates of the compounds from the body were determined. Where the 0.773 log ml/min/kg and -0.081 log ml/min/kg values for a total clearance of NP132 and rufomycin I respectively meant that NP132 would be excreted quickly by the kidneys, while rufomycin I would have a slower excretion rate. Lastly, AMES toxicity was used to test for mutagenicity, and both NP132 and rufomycin I were non-mutagenic. Therefore, they might have a high probability of passing the toxicity tests *in vitro* and *in vivo* [52].

4. Conclusion

Drug resistance to TB has made most of the available treatments inefficient, which has led to a search for novel treatment options. One of the targets being studied currently is the chaperone protein ClpC1 involved in protein homeostasis and is essential for *Mtb* survival and proliferation in the host. In this *in silico* study, the ensemble-based virtual screening approach was utilised to search for potential inhibitors by targeting the N-terminal domain (binding domain) of the ClpC1 protein. Taking into consideration protein and ligand flexibility, several potential inhibitors were identified based on their strong binding affinities. However, the compound NP132 (ID: SN00055391) with the strongest binding affinity and significant H-bonds formation was tested further for its stability with the protein. Observing the complex MD simulation data, the protein structure seemed slightly disrupted, probably due to the presence of the ligand NP132 nonetheless, the complex remained compact and stable. Furthermore, comparing the potent compound NP132 against the positive control rufomycin I indicated that NP132 could be a more suitable inhibitor, as determined by the physicochemical, pharmacokinetic, and toxicological properties studied. The small drug-like compound NP132 followed Lipinski's rule of five and was non-toxic and non-mutagenic. It could also be absorbed rapidly into the bloodstream and metabolised by the CYP3A4 enzyme for quick removal from the body. The results from these *in silico* analyses should be further studied by *in vitro* and *in vivo* assays for determining the inhibitory action of promising compound NP132 on ClpC1-NTD.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. World Health Organization. Global tuberculosis report (2019).
2. Singh S, Khare G, Bahal RK, et al. Identification of *Mycobacterium tuberculosis* BioA inhibitors by using structure-based virtual screening. *Drug Des Dev Ther* 12 (2018): 1065-1079.
3. Vasudevan D, Rao SP, Noble CG. Structural basis of mycobacterial inhibition by cyclomarin A. *J Biol Chem* 288 (2013): 30883-30891.
4. Garg M, Sidhu NK, Kakarala KK. *In silico* analysis of Clp protease catalytic subunit-2 of *Mycobacterium tuberculosis*: modeling and docking analysis. *Int J Adv Res* 2 (2014): 11170-11203.
5. Choules MP, Wolf NM, Lee H, et al. Rufomycin targets ClpC1 proteolysis in *Mycobacterium tuberculosis* and *M. abscessus*. *Antimicrob Agents Chemother* 63 (2019): e02204-e02218.
6. Lupoli TJ, Vaubourgeix J, Burns-Huang K, et al. Targeting the proteostasis network for mycobacterial drug discovery. *ACS Infect Dis* 4 (2018): 478-498.
7. Wolf NM, Lee H, Choules MP, et al. High-resolution structure of ClpC1-rufomycin and ligand binding studies provide a framework to design and optimize anti-tuberculosis leads. *ACS Infect Dis* 5 (2019): 829-840.
8. Leelananda SP, Lindert S. Computational methods in drug discovery. *Beilstein J Org Chem* 12 (2016): 2694-2718.
9. Khan SU, Ahemad N, Chuah LH, et al. Sequential ligand- and structure-based virtual screening approach for the identification of potential G protein-coupled estrogen receptor-1 (GPER-1) modulators. *RSC Adv* 9 (2019): 2525-2538.
10. Barakat K, Mane J, Friesen D, et al. Ensemble-based virtual screening reveals dual-inhibitors for the p53-MDM2/MDMX interactions. *J Mol Graph Model* 28 (2010): 555-568.
11. Amaro RE, Li WW. Emerging methods for ensemble-based virtual screening. *Curr Top Med Chem* 10 (2010): 3-13.
12. Lionta E, Spyrou G, Vassilatis DK, et al. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem* 14 (2014): 1923-1938.
13. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera - a visualization system for exploratory research and analysis. *J Comput Chem* 25 (2004): 1605-1612.
14. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234 (1993): 779-815.
15. Laskowski RA, MacArthur MW, Moss DS, et al. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26 (1993): 283-291.
16. Lipinski CA, Lombardo F, Dominy BW, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46 (2001): 3-26.

17. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: high performance molecular simulations through multi level parallelism from laptops to supercomputers. *SoftwareX* 1 (2015): 19-25.
18. Nguyen TT, Viet MH, Li MS. Effects of water models on binding affinity: evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase. *Sci World J* 2014 (2014): 536084.
19. Klauda JB, Venable RM, Freites JA, et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* 114 (2010): 7830-7843.
20. Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79 (1983): 926-935.
21. Shukla R, Shukla H, Sonkar A, et al. Structure-based screening and molecular dynamics simulations offer novel natural compounds as potential inhibitors of *Mycobacterium tuberculosis* isocitrate lyase. *J Biomol Struct Dyn* 36 (2018): 2045-2057.
22. Feller SE, Zhang Y, Pastor RW, et al. Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys* 103 (1995): 4613-4621.
23. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 98 (1993): 10089-10092.
24. Turner P. GRACE (5.1.22) [Computer software]. Center for coastal and land-margin research. Oregon graduate institute of science and technology, Beaverton, OR (2005).
25. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30 (2009): 2785-2791.
26. Fuhrmann J, Rurainski A, Lenhof HP, et al. A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *J Comput Chem* 31 (2010): 1911-1918.
27. Zoete V, Cuendet MA, Grosdidier A, et al. SwissParam: a fast force field generation tool for small organic molecules. *J Comput Chem* 32 (2011): 2359-2368.
28. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 7 (2017): 42717.
29. Pires DE, Blundell TL, Ascher DB. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem* 58 (2015): 4066-4072.
30. Gruswitz F, Chaudhary S, Ho JD, et al. Function of human Rh based on structure of RhCG at 2.1 Å. *Proc Natl Acad Sci* 107 (2010): 9638-9643.
31. Sliwoski G, Kothiwale S, Meiler J, et al. Computational methods in drug discovery. *Pharmacol Rev* 66 (2014): 334-395.
32. Bornot A, Etchebest C, De Brevern AG. Predicting protein flexibility through the prediction of local structures. *Proteins* 79 (2011): 839-852.
33. Gavrish E, Sit CS, Cao S, et al. Lassomycin, a ribosomally synthesized cyclic peptide, kills *Mycobacterium tuberculosis* by targeting the ATP-dependent protease ClpC1P1P2. *Chem Biol* 21 (2014): 509-518.
34. Morcos F, Jana B, Hwa T, et al. Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proc Natl Acad Sci* 110 (2013): 20533-20538.
35. Yan Z, Wang J. Specificity quantification of biomolecular recognition and its implication for drug discovery. *Sci Rep* 2 (2012): 1-7.
36. Marsee JD, Ridings A, Yu T, et al. *Mycobacterium tuberculosis* ClpC1 N-terminal domain is dispensable

- for adaptor protein-dependent allosteric regulation. *Int J Mol Sci* 19 (2018): 3651.
37. Itoh Y, Nakashima Y, Tsukamoto S, et al. N⁺-CH...O Hydrogen bonds in protein-ligand complexes. *Sci Rep* 9 (2019): 1-12.
 38. Jung IP, Ha NR, Kim AR, et al. Mutation analysis of the interactions between *Mycobacterium tuberculosis* caseinolytic protease C1 (ClpC1) and ecumicin. *IntJ Biol Macromol* 101 (2017): 348-357.
 39. Mamputha S, Lu ZL, Roeske RW, et al. Conserved amino acid residues that are important for ligand binding in the type I gonadotropin-releasing hormone (GnRH) receptor are required for high potency of GnRH II at the type II GnRH receptor. *Mol Endocrinol* 21 (2007): 281-292.
 40. Mok CK, Chen GW, Shih KC, et al. Evolutionarily conserved residues at an oligomerization interface of the influenza A virus neuraminidase are essential for viral survival. *Virology* 447 (2013): 32-44.
 41. Remer LC, Jensen JH. Toward a general theory of hydrogen bonding: the short, strong hydrogen bond [HOH...OH]. *J Phys Chem A* 104 (2000): 9266-9275.
 42. Williams MA, Ladbury JE. Hydrogen bonds in protein-ligand complexes. In: Böhm H-J, Schneider G (eds) *Protein-ligand interactions: from molecular recognition to drug design*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim (2003): 137-161.
 43. Arnittali M, Rissanou AN, Harmandaris V. Structure of biomolecules through molecular dynamics simulations. *Procedia Comput Sci* 156 (2019): 69-78.
 44. Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* 1 (2004): 337-341.
 45. Turner JV, Agatonovic-Kustrin S. *In silico* prediction of oral bioavailability. In: Taylor JB, Triggler DJ (eds) *Comprehensive medicinal chemistry II*. Elsevier, Oxford (2007): 699-724.
 46. Doogue MP, Polasek TM. The ABCD of clinical pharmacokinetics. *Ther Adv Drug Saf* 4 (2013): 5-7.
 47. Wang CK, Craik DJ. Cyclic peptide oral bioavailability: lessons from the past. *PeptSci* 106 (2016): 901-909.
 48. Durán-Iturbide NA, Díaz-Eufracio BI, Medina-Franco JL. *In silico* ADME/Tox profiling of natural products: a focus on Biofacquim. *ACS Omega* 5 (2020): 16076-16084.
 49. Sevrioukova IF, Poulos TL. Understanding the mechanism of cytochrome P450 3A4: recent advances and remaining problems. *Dalton Trans* 42 (2013): 3116-3126.
 50. Basheer L, Kerem Z. Interactions between CYP3A4 and dietary polyphenols. *OxidMed Cell Longev* 2015 (2015): 854015.
 51. Lynch T, Price AL. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *Am Fam Physician* 76 (2007): 391-396.
 52. McCarren P, Springer C, Whitehead L. An investigation into pharmaceutically relevant mutagenicity data and the influence on Ames predictive potential. *J Cheminformatics* 3 (2011): 1-20.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC-BY\) license 4.0](https://creativecommons.org/licenses/by/4.0/)