Screening Of Zinc Database Against Streptococcal Cysteine Protease Enzyme For Identification Of Novel Group A Streptococcus Inhibitors

Ratul Bhowmik[1*], Ranajit Nath[2], Ratna Roy[3]

[1*]*Department of Pharmaceutical Chemistry, SPER, Jamia Hamdard, New Delhi, India.*

[2]*Department of Pharmaceutics, NSHM Knowledge Campus, Kolkata-Group of Institutions, Kolkata, West Bengal, India*

[3]*Department of Pharmacology, NSHM Knowledge Campus, Kolkata-Group of Institutions, Kolkata, West Bengal, India*

*Corresponding author: Ratul Bhowmik, Department of Pharmaceutical Chemistry, SPER, Jamia Hamdard, New Delhi, India. Email: ss713724@gmail.com

ABBREVIATIONS

ADMET: Adsorption, Distribution, Metabolism, Excretion, Toxicity

GAS: Group A Streptococcus

MD: Molecular Dynamics

NMR: Nuclear Magnetic Reasonance

PAINS: Pan-assay interference compounds

PDB: Protein Data Bank

SBVS: Structure-based virtual screening

ABSTRACT

Inhibition of streptococcal cysteine protease has recently emerged as quite a promising target to treat severe cases of Group A Streptococcus infections. For the identification of streptococcal cysteine protease inhibitors, structure-based virtual screening (SBVS) of ZINC Database was performed. The docking protocol was performed with the help of AutoDock Tools and AutoDock Vina software. Based on binding affinity and similarity of interactions with our target receptor streptococcal cysteine protease, 4 hit compounds were identified, which were further subjected to ADMET (Adsorption, Distribution, Metabolism, Excretion, Toxicity) and Drug-likeness to identify the best hit compound. The most potent compound showed binding of -7.7 KJ/mol with receptor streptococcal cysteine protease. It also showed 6 similar amino acid interactions with the receptor's native ligand along with good ADME and Drug-likeness properties. Furthermore, the molecular dynamics simulation analysis revealed that the complex formed between the protein streptococcal cysteine protease and the hit compound ZINC000205429716 had good structural stability. The current study reveals the successful use of in silico SBVS methods for the identification of novel and possible streptococcal cysteine protease inhibitors, with compound ZINC000205429716 serving as a potential lead for the creation of Group A Streptococcus inhibitors.

Keywords: Group A Streptococcus, streptococcal cysteine protease, structure-based virtual screening, ADMET, molecular dynamics simulation

INTRODUCTION

Streptococcus pyogenes, commonly known as GAS (Group A Streptococcus) cause a wide range of cutaneous infections, from superficial streptococcal pyoderma to moderately severe cellulitis and even life-threatening necrotizing fasciitis. The annual prevalence of streptococcus pyoderma is estimated to be more than 111 million cases worldwide, with economically disadvantaged children living in tropical and subtropical areas being the most frequently affected. Epidemiological data show a link between streptococcal pyoderma and the development of post-infectious glomerulonephritis and invasive diseases in patients, which has serious implications for their prognosis. Following repeated GAS exposure, serious postinfectious immune sequelae such as rheumatic fever and acute glomerulonephritis can develop. Despite the availability of sequence information for several GAS genomes and detailed characterization of their virulence factors, a commercial GAS vaccine that is both safe and effective has yet to be developed (1,2).

During an infection, GAS produces a large number of secreted and cell-associated proteins, such as toxins, superantigens, and proteases. The most important among them is the streptococcal cysteine protease, which is mostly responsible for severe cases of GAS infection. Streptococcal cysteine protease is a virulence factor that cleaves human fibronectin and degrades vitronectin. It also cleaves human IL1B precursor, resulting in biologically active IL1B. Furthermore, it induces apoptosis in human monocytes and epithelial cells in vitro, as well as decreasing phagocytic activity in monocytic cells. Streptococcus cysteine protease thus plays a role in bacterial colonization, invasion, and wound healing inhibition (3,4). Thus inhibition of the streptococcal cysteine protease would help in preventing the severe cases of GAS infections.

The goal of this work is to structure-based virtual screening (SBVS) to find tiny powerful molecules that can target our target receptor streptococcal cysteine protease. The study's findings indicated that the discovered hit compounds might be possible lead molecules in

partially streptococcal cysteine protease, which can then be utilized to treat severe cases of GAS infection.

MATERIALS AND METHODS

Computer Environment

VSDK (Virtual Screening by Docking) may be done and executed on any version of Microsoft Windows or the LINUX platform. A high-speed computer machine with multiple operating systems was utilized to do virtual screening (windows, Linux). It also had a Java environment, a strong internet connection, and a stable power supply.

Structure-Based Virtual Screening

When it comes to medication activity, molecular recognition is regarded as the most important factor. The phrase "drug action" refers to the pharmacological activity displayed by drug molecules when binding to the targeted protein and creating a stable protein-ligand complex. SBVS, also known as structure-based virtual screening, aims to exploit and explore the molecular recognition between the target protein and the chosen ligand molecules to select specific molecules that show good binding affinity with the active sites of the targeted biological receptor, allowing 3D structures to be inferred. The docking approach is based on identifying the optimal conformation or pose of the ligand with the receptor's specific active region. The dock score binding affinity indicates the binding relationship between the ligand molecules and the targeted protein. Based on calculated binding interactions, docking scores may also be used to predict the biological activity of ligand molecules. In general, the initial stage in SBVS is the selection of the targeted protein as well as the 3D database of the ligand. The next stage is to utilize virtual filters to dock and score compounds to discover and choose compounds for future investigation.

Steps used in Virtual Screening

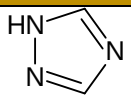
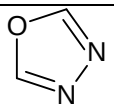
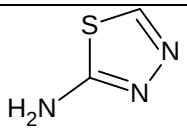
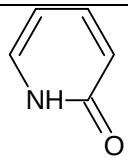
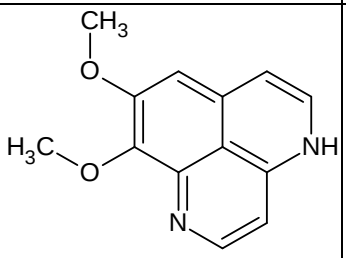
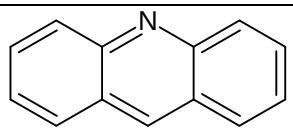
The initial stage was to construct a database of tiny molecules by selecting a library, removing counter ions, adding hydrogen, resolving valency issues, protonation at physiological pH, calculating 2D characteristics, converting 2D molecules to 3D, and minimizing energy. The second stage was to identify our target receptor using NMR or X-ray, with a resolution value of less than 2 Armstrong. The final step was to locate the binding location of our desired receptor. The fourth step was to run the docking technique to estimate the optimal ligand conformation at the selected receptor's binding region. The docking score or binding affinity

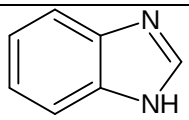
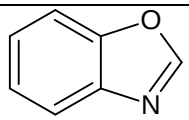
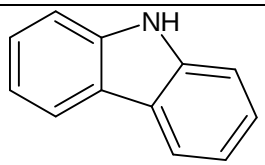
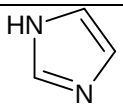
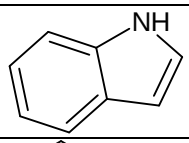
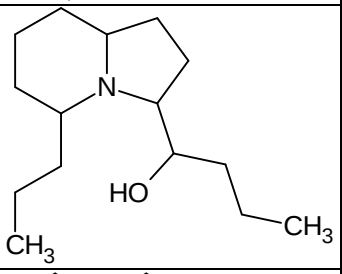
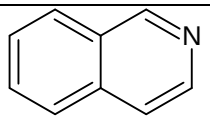
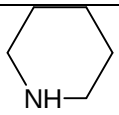
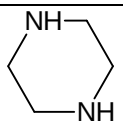
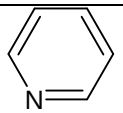
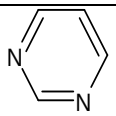
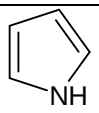
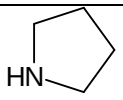
was utilized to estimate and assess the interaction energy between our ligand and the target receptor. The sixth and last stage was to filter the docked molecules further based on ADMET characteristics.

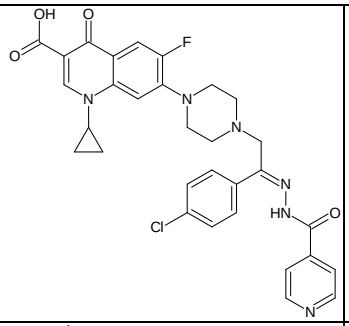
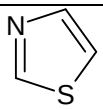
Selection and Preparation of Ligand Library

A literature review was used to choose several antibacterial medicinally significant scaffolds (Table 1). The selected scaffolds' structures were drawn on the ZINC database website, and their substructures were generated (5). The created substructures were then downloaded in SDF format. Using the python software prepare ligand4.py, the substructures were then translated to pdbqt format. These substructures were utilized to build a virtual library that would later be used for molecular docking and ADMET evaluation.

Table 1: This table demonstrates the names and structures of scaffolds chosen for this study

Sl. No.	Name	Scaffolds	Smiles
1.	1,2,4-Triazole		C1=NC=NN1
2.	1,3,4-Oxadiazole		C1=NN=CO1
3.	2-Amino-1,3,4-thiadiazole		C1=NN=C(S1)N
4.	2-Hydroxypyridine		C1=CC(=O)NC=C1
5.	Aaptamine		COC1=C(C2=NC=CC3=C2C(=C1)C=CN3)OC
6.	Acridine		C1=CC=C2C(=C1)C=C3C=CC=CC3=N2

7.	Benzimidazole		C1=CC=C2C(=C1)NC=N2
8.	Benzoxazole		C1=CC=C2C(=C1)N=CO2
9.	Carbazole		C1=CC=C2C(=C1)C3=CC=CC=C3N2
10.	Imidazole		C1=CN=CN1
11.	Indole		C1=CC=C2C(=C1)C=CN2
12.	Indolizidine		CCCC1CCCC2N1C(CC2)C(CCC)O
13.	Isoquinoline		C1=CC=C2C=NC=CC2=C1
14.	Piperidine		C1CCNCC1
15.	Piperazine		C1CNCCN1
16.	Pyridine		C1=CC=NC=C1
17.	Pyrimidine		C1=CN=CN=C1
18.	Pyrrole		C1=CNC=C1
19.	Pyrrolidine		C1CCNC1

20.	Quinolone		C1CC1N2C=C(C(=O)C3=CC(=C(C=C32)N4CCN(CC4)CC(=NNC(=O)C5=CC=NC=C5)C6=CC=C(C=C6)Cl)F)C(=O)O
21.	Thiazole		C1=CSC=N1

Selection and Preparation of Receptor

The chosen receptor molecule streptococcal cysteine protease (PDB-Id: 2UZJ) was downloaded in PDB format from the protein data bank database (6). After that, the protein molecule was imported into the AutoDock Tools program. To begin, the co-crystallized ligand was extracted to verify the protein. Following that, the protein was prepared by removing water molecules, eliminating unnecessary chains or heteroatoms, mending missing atoms, adding hydrogen atoms, computing charges (Kollman charges), and lastly converting it to pdbqt format. Finally, a grid box was generated with the co-crystallized ligand in the middle. The grid box dimensions were stored as config.txt file for docking with Autodock Vina (7). The generated protein pdbqt file was then used to extract the co-crystallized ligand.

Molecular Docking using Autodock Vina

The prepared protein was then docked against our prepared library set of ligands using AutoDock Vina (7). Perl Script was used for docking of our multiple ligands (8). The results were displayed in terms of binding affinity. The binding affinity represents the binding energy. The binding energy exhibits the extent of binding of the ligand molecule. Furthermore, the best type of configuration would be the one that would bind with its target. The docking results were analyzed using Discovery Studio Biovia 2021 and Pymol software (9–11).

ADMET analysis

The initially screened ligands obtained with the help of dock scores were then subjected to ADMET analysis. SwissADME and Pre-ADMET web servers were used to predict drug-likeness and ADMET properties of our ligand molecules (12,13). Lipinski's rule was used to check whether the initially screened ligand molecules were suitable for docking. According to

Lipinski's rule of five, a compound, to qualify as a ligand, should have less than 500 Da molecular weight, high lipophilicity i.e. value of Log P less than five, hydrogen bond donors less than 5, and hydrogen bond acceptors less than 10. Ligand violating any two rules of Lipinski's was considered unsuitable for further screening. Other than Lipinski's rule, physicochemical analysis, as well as Drug-likeliness properties of all the ligand molecules, were also taken into consideration for the drug screening process. Parameters such as PAINS (Pan-assay interference compounds) and synthetic accessibility of the compounds were also taken into consideration. Moreover, Pre-ADMET was used for predicting the toxicity profiles of our initially screened ligand molecules. Molinsipration web server was also used to predict the bioavailability parameter of our screened ligand molecules.

Boiled-Egg Analysis

For predicting blood-brain barrier permeability as well as gastrointestinal absorption of our selected phytochemicals, BOILED EGG was used (14). According to BOILED-Egg plot analysis, compounds found in the yellow region were considered to be having higher blood-brain barrier permeability, whereas compounds found in the white region of the plot were considered to be having higher gastrointestinal absorption properties. The BOILED-Egg plot analysis was performed using the SwissADME web server.

Molecular Dynamics Simulations Study

The molecule with the best binding affinity along with satisfactory ADMET properties was further subjected to a molecular dynamics simulation study. Molecular Dynamics (MD) Simulation is a computer-based simulation approach used to analyze the physical motions of atoms or molecules. MD simulations can identify a few critical hydrogen bond interactions. MD simulations assist in protein docking and virtual screening advances. The iMODS server was utilized in this work to simulate molecular dynamics. The iMODS service aids in the exploration of normal mode analysis and generates accessible information about routes that may involve macromolecules or homologous structures (15).

RESULTS

Molecular Docking Results

A total of 2200 compounds were downloaded from the ZINC database webserver in SDF format for docking against our targeted receptor streptococcal cysteine protease. The native ligand of the protein streptococcal cysteine protease was also separately docked with the

targeted protein. The docking score of our native ligand with our target protein was found to be -6.2 KJ/mol. Now, among the 2100 compounds, the ones demonstrating dock scores of more than -6.2 KJ/mol were selected for further analysis. A total of four compounds were initially screened which showed docking scores of -7 KJ/mol and above (Figure 1, Table 2). Moreover, the amino acid interactions of our native ligand with target protein were also compared with that of amino acids interactions of the initially screened four ligands with the target protein.

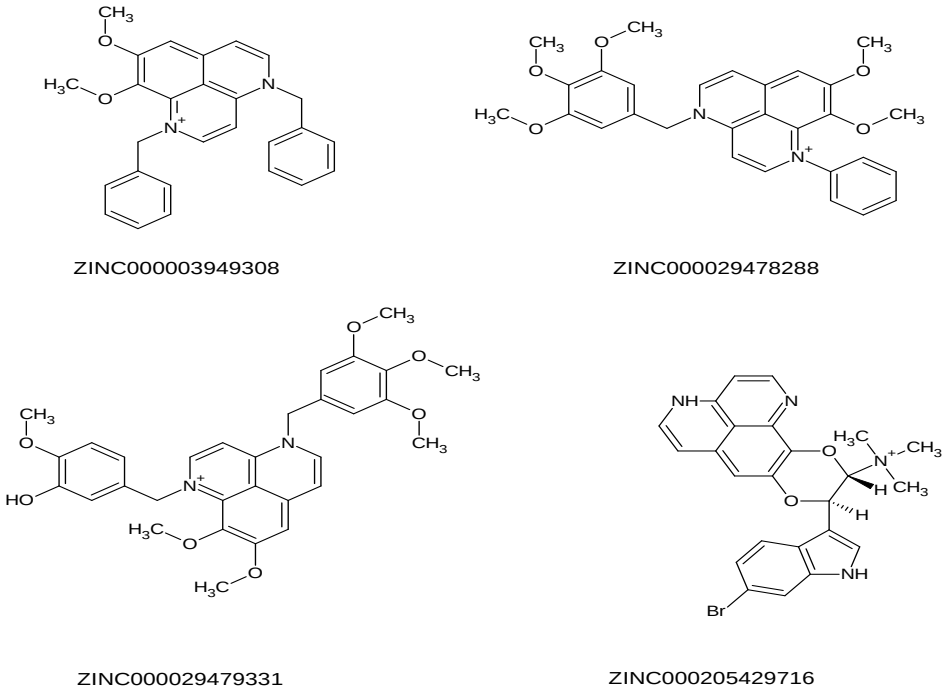


Figure 1: This figure represents the initially screened hit compounds obtained from virtual screening

Compound ZINC000205429716 showed the best binding affinity of about -7.7 KJ/mol, as well as showed 6 common amino acid interactions as that of the native ligand. The common amino acid interactions of the compound ZINC000205429716 and the native ligand were found at Asn161, Tyr244, Gln245, Ser163, Asn242, Arg162. Compound ZINC000205429716 showed overall decent amino acid interactions. It demonstrated hydrogen bonding interactions at Ser163, show 2 Pi-Alkyl interactions at ARG162 and 1 Pi-Alkyl interaction at Tyr244. Furthermore, it also showed Amide-Pi Stacked interaction and Pi-Sigma at Tyr244. Van der Waals interactions were observed at Gln245, Ser246, Asn161, Gly243, Asn242, Asp225, Leu224 (Figure 2). The structural analysis of the compound was done using Discovery Studio Biovia 2021 and Pymol software (Figure 3, Figure 4).

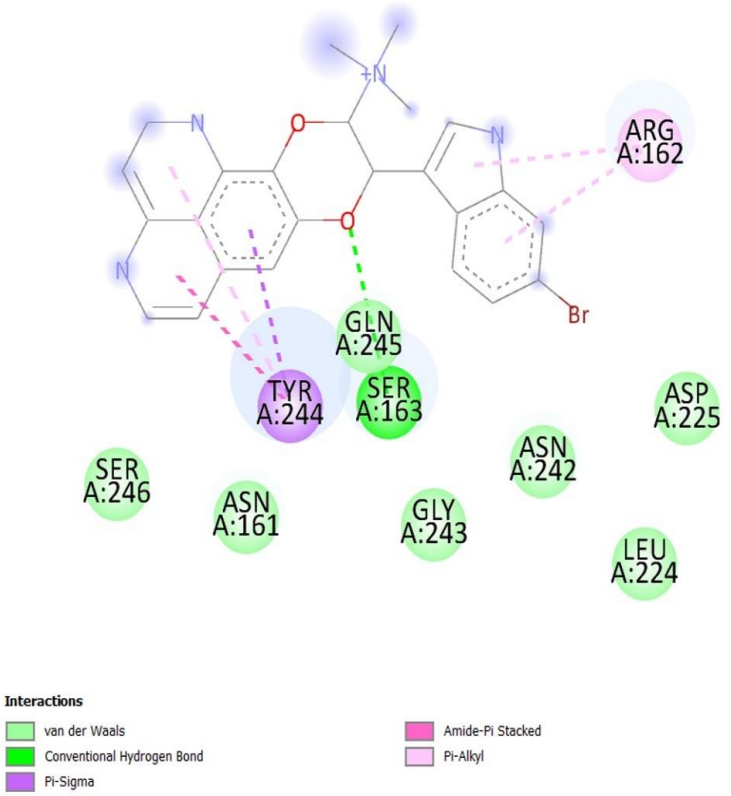


Figure 2: This figure represents 2D amino acid interactions of the hit compound ZINC000205429716 with Streptococcus cysteine protease

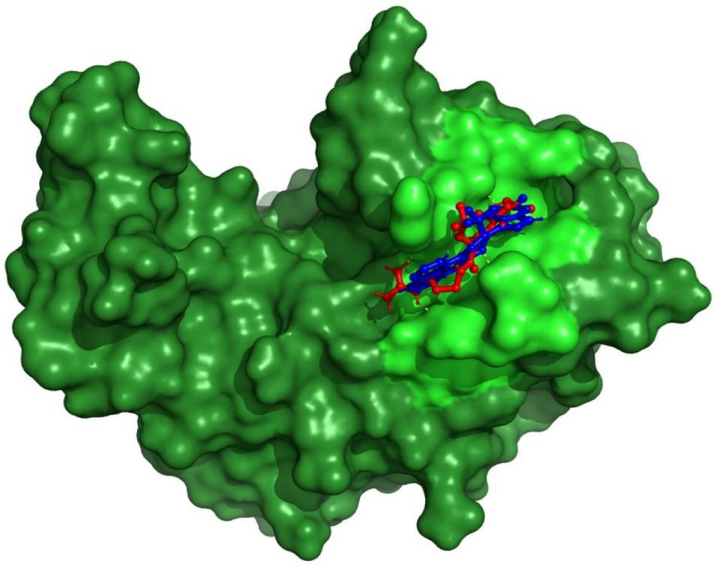


Figure 3: Figure represents superimposed structures of native ligand (red color) with hit compound ZINC000205429716 (blue color) with protein Streptococcal cysteine protease (deep green). The active site is represented in light green color.

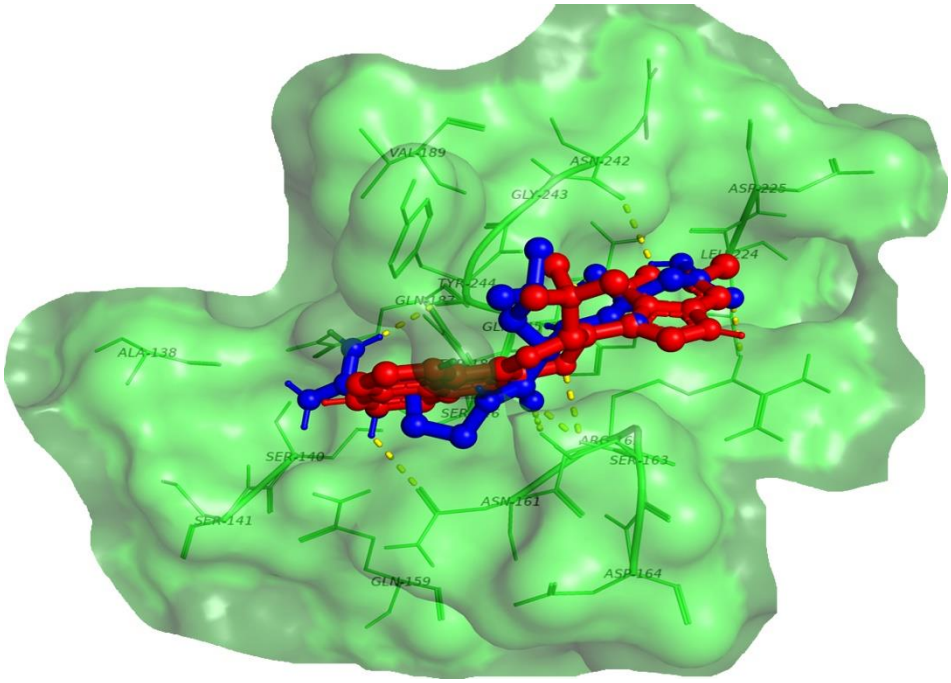


Figure 4: Figure represents detailed amino acid interactions with superimposed structures of native ligand (red color) with hit compound ZINC000205429716 (blue color) with protein Streptococcal cysteine protease (deep green). The active site is represented in a light green color.

ADMET analysis

The ADME analysis of the initially screened compounds was done using SwissADME and Pre-ADMET webserver. Using Lipinski's rule of five, the ADME analysis of the previously screened molecules was performed. Physiochemical parameters and drug-likeness properties of the initially screened four ligand molecules were all noted down. The ADME results demonstrated that all of the four ligands passed Lipinski's rule of five analysis (Table 2). The toxicity analysis using the Pre-ADMET demonstrated that the ligands molecules showed moderate to very low toxicity (Table 3). Moreover, the Molinspiration webserver's bioavailability results demonstrated that each of the initially screened four ligands showed good specificity towards enzyme inhibition activity (Table 4). Furthermore, the PAINS (pan-assay interference compounds) analysis showed zero alerts for all four ligands. The overall synthetic accessibility score for the four ligands was in the range of 2.92-4.23.

Table 2: Table represents Drug-Likeness properties and Dock score of the initially screened hit compounds

Sr. No.	Ligands	BBB	GI absorption	Permeability glycoprotein substrate	Log S (SILICOS-IT) (scale insoluble<-10<poorly<-6<moderately<-4<soluble<-2very<0	Lipinski Rule	Dock Score
1	ZINC000003949308	Yes	High	Yes	-8.95	Passed	-7.0
2	ZINC000029478288	Yes	High	Yes	-8.86	Passed	-7.0
3	ZINC000029479331	No	High	Yes	-8.74	Passed	-7.6
4	ZINC000205429716	Yes	High	Yes	-9.47	Passed	-7.7

Table 3: Detailed toxicity analysis of the initially screened hit compounds

ID	ZINC000003949308	ZINC000029478288	ZINC000029479331	ZINC000205429716
algae_at	0.0330949	0.0134343	0.0061521	0.0144541
Ames_test	mutagen	mutagen	mutagen	non-mutagen
Carcino_Mouse	negative	negative	negative	negative
Carcino_Rat	negative	negative	negative	positive
daphnia_at	0.0276493	0.0218115	0.0241821	0.0200488
hERG_inhibition	medium_risk	medium_risk	medium_risk	medium_risk
medaka_at	0.00199505	0.00136381	0.00168615	0.00132119
minnow_at	0.010438	0.00505731	0.00601564	0.0113621
TA100_10RLI	negative	negative	negative	negative
TA100_NA	positive	negative	negative	negative
TA1535_10RLI	negative	negative	negative	negative
TA1535_NA	negative	negative	negative	negative

Table 4: Bioavailability properties of the initially screened hit compounds.

Bioavailability parameters	ZINC000003949308	ZINC000029478288	ZINC000029479331	ZINC000205429716
GPCR ligand	0.16	-0.16	0.13	**0.21**
Ion channel modulator	**0.35**	0.17	0.06	**0.23**
Kinase inhibitor	-0.3	-0.27	-0.28	**0.29**
Nuclear receptor ligand	-0.7	-0.53	-0.58	-0.3
Protease inhibitor	-0.21	-0.19	-0.18	-0.24
Enzyme inhibitor	**0.39**	**0.24**	**0.26**	**0.2**

Boiled-Egg Analysis

The figure 5 shows the Boiled-Egg plot of our initially screened four ligands. Molecule 1 represents ZINC000003949308, molecule 2 represents ZINC000029478288, molecule 3 represents ZINC000029479331, molecule 4 represents ZINC000205429716. Molecule 1, molecule 2, and molecule 4 showed good blood-brain barrier permeability as well as high gastrointestinal retention properties. On the other hand, molecule 3 demonstrated good gastric retention properties but with poor blood-brain barrier permeability property.

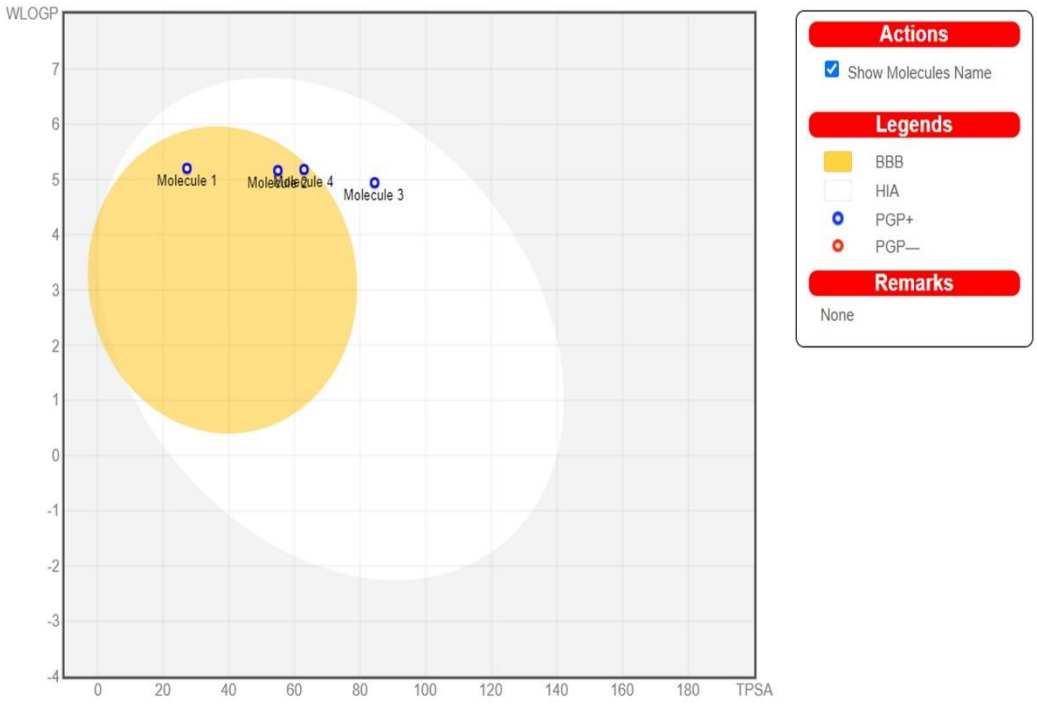


Figure 5: The white part of the BOILED-EGG Model represents the physicochemical space of molecules with the highest probability of being absorbed by the GI (gastrointestinal tract), while the yellow part represents the physicochemical space of molecules with the highest probability of permeating to the brain.

Molecular Dynamics Simulation Study Analysis

The Molecular Dynamics simulation results are shown in Figure 6. Compound ZINC000205429716 was selected as the hit compound as it showed the best binding affinity along with good ADME properties. Here the docked complex of our ligand ZINC000205429716 with receptor streptococcal cysteine protease was considered for MD simulation (Figure 6). Normal mode analysis mobility allows us to analyze the large-scale B-factor and mobility as well as the stability of the molecules. The IMOD server exposed the internal coordinates analysis depending on the protein-ligand structural interactions. IMODs also measure the B-factor and structural deformity and calculate the eigenvalue.

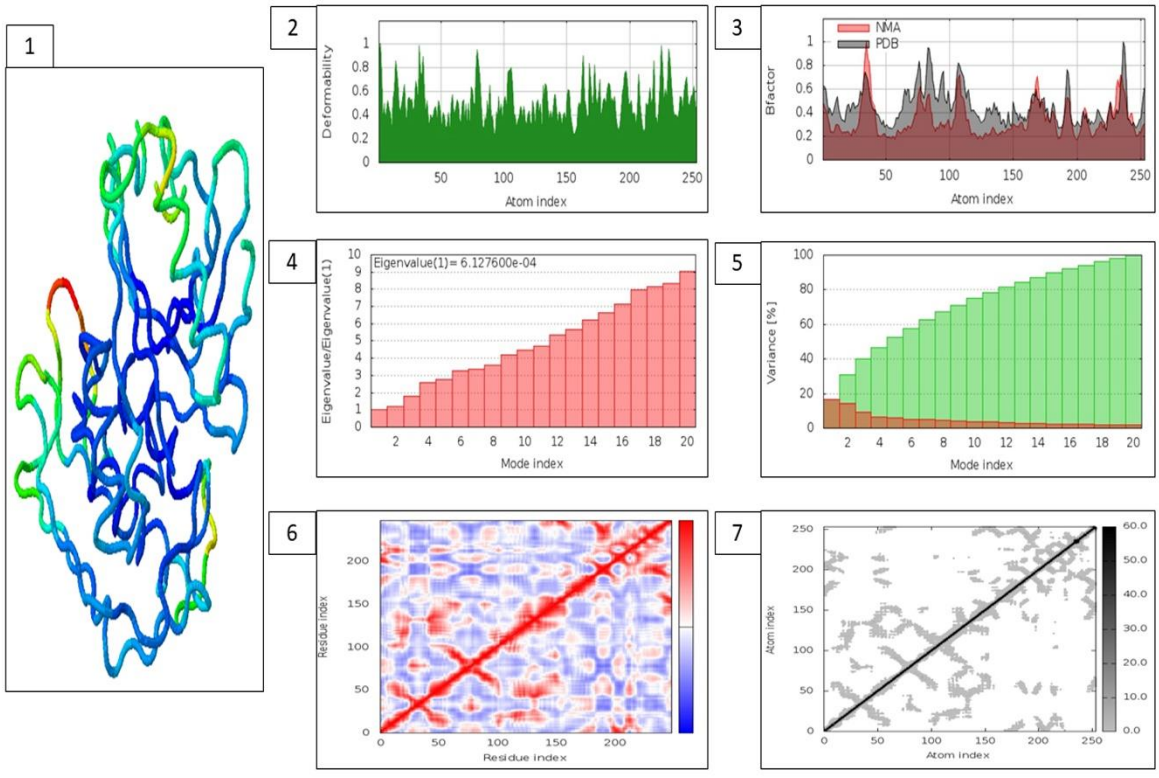


Figure 6: Figure 6 represents molecular dynamic simulation analysis of the final hit compound ZINC000205429716 with receptor Streptococcal cysteine protease.

Image 1 of Figure 6 represents the docked complex of our protein and ligand. Image 2 of Figure 6 represents the deformability graph. The deformity graph illustrated peaks in the graph which represent regions in the protein with deformability. Image 3 represents the B-Factor graph. The main-chain deformability, also known as the B-Factor, is a measure of a molecule's ability to deform at each of its residues. Image 4 represents the eigenvalue of the complex. The motion stiffness is represented by the eigenvalue associated with each normal mode. Its value is proportional to the amount of energy required to distort the structure. The simpler the deformation, the lower the eigenvalue. Our docked complex demonstrated an eigenvalue of 6.127600e-04. Image 5 represents the variance plot. The variance plot demonstrates individual variances in red color whereas cumulative variance in green color. Image 6 represents the covariance map. This map demonstrates the correlation motion between a pair of residues in red color, uncorrelated motion in white color, and anti-correlated motion in blue color. Image 7 represents the elastic map of our docked complex. Each dot in the graph represents one spring inside the atoms' pair. The dots are colored dependent on stiffness, with darker grey dots indicating stiffer springs and lighter grey dots indicating softer springs.

From the molecular dynamics study, it was evident that our complex showed a good amount of deformability. Furthermore, it also showed a moderate eigenvalue, suggesting that it could be deformed. The variance map exhibited a higher degree of cumulative variances than an individual variance. The elastic network map also produced satisfactory results.

CONCLUSION

Streptococcal cysteine protease is primarily responsible for cleaving transmembrane proteins associated with the epithelial barrier, allowing streptococcus bacteria to pass through. Hence, it is considered an ideal target for developing lead candidates against severe cases of GAS infections. The target receptor for this study was streptococcal cysteine protease. Approximately 2100 substructures were generated using several scaffolds and the ZINC Database webserver. These substructures were then subjected to screening, docking, ADMET analysis, and molecular dynamics simulations. We discovered a hit molecule, ZINC000205429716, that has a high affinity for streptococcal cysteine protease. The hit molecule identified in this study met all the ADMET requirements. The structure of ZINC000205429716, like the native ligand of the target receptor, revealed some similar interactions with streptococcal cysteine protease, making it an appropriate molecule to block the target receptor. According to the molecular dynamics simulation study, the complex formed

between our protein and the hit compound ZINC000205429716 exhibited good structural stability. To summarize, the ZINC000205429716 molecule is a leading candidate for binding with streptococcal cysteine protease, which could lead to the treatment of severe GAS infections. This study's findings will be useful in pre-clinical and subsequent in vivo trials.

ACKNOWLEDGMENT

Nil

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

1. Sumitomo T, Nakata M, Higashino M, Terao Y, Kawabata S. Group A streptococcal cysteine protease cleaves epithelial junctions and contributes to bacterial translocation. J Biol Chem. 2013;288(19):13317–24.
2. Nelson, DC, Garbe, J, Collin, M. The cysteine proteinase SpeB from Streptococcus pyogenes – a potent modifier of immunologically important host and bacterial proteins. Biol Chem. 2011;392(December):1077–88.
3. Sumitomo T, Mori Y, Nakamura Y, Honda-Ogawa M, Nakagawa S, Yamaguchi M, et al. Streptococcal cysteine protease-mediated cleavage of desmogleins is involved in the pathogenesis of cutaneous infection. Front Cell Infect Microbiol. 2018;8(JAN):1–10.
4. Olsen JG, Dagil R, Niclasen LM, Sørensen OE, Kragelund BB. Structure of the Mature Streptococcal Cysteine Protease Exotoxin mSpeB in Its Active Dimeric Form. J Mol Biol [Internet]. 2009;393(3):693–703. Available from: http://dx.doi.org/10.1016/j.jmb.2009.08.046
5. Irwin JJ, Shoichet BK. ZINC - A free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005;45(1):177–82.
6. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Vol. 28, Nucleic Acids Research. 2000. p. 235–42.
7. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2009;31(2):455–61.

8. Sharma V, Pattanaik KK, Jayprakash V, Basu A, Mishra N. A utility script for automating and integrating AutoDock and other associated programs for virtual screening. Bioinformation. 2009;4(2):84–6.

9. DeLano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsl protein Crystallogr. 2002;40.

10. Seeliger D, De Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. J Comput Aided Mol Des. 2010;24(5):417–22.

11. Visualizer DS. v4. 0.100. 13345. In: Accelrys Software Inc. 2005.

12. 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. 2017;7(March):1–13.

13. Lee SK, Lee IH, Kim HJ, Chang GS, Chung JE, No KT. The PreADME Approach: Web-based program for rapid prediction of physico-chemical, drug absorption and drug-like properties. EuroQSAR 2002 Des Drugs Crop Prot Process Probl Solut. 2002;418–20.

14. Daina A, Zoete V. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. ChemMedChem. 2016;11(May):1117–21.

15. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. IMODS: Internal coordinates normal mode analysis server. Nucleic Acids Res. 2014;42(W1):271–6.