INTERNATIONAL JOURNAL
OF
PHARMACEUTICAL SCIENCES
AND
RESEARCH

MOLECULAR DOCKING STUDY OF HEMAGGLUTININ PROTEIN OF INFLUENZA A VIRUS TO DEVELOP A NOVEL ANTI-INFLUENZA DRUG

Anika Chebrolu

Texas Academy of Mathematics & Science, University of North Texas, Denton, TX, USA.

ABSTRACT: The Hemagglutinin (HA) protein of the Influenza virus plays a vital role in viral replication. In this study, numerous small molecules were screened systematically against the HA protein to find a compound that can bind to it selectively. The HA protein's three-dimensional crystal structure was obtained from RCSB PDB and inputted into the FTMap Web Tool to identify the best possible drug gable hotspot. Over 68 million compounds from the ZINC-15 database were filtered for drug-likeness properties and narrowed down to 250,000 ligands. These chosen molecules were then tested for their binding affinity to the selected hotspot on the HA protein using the CLC Drug Discovery Software. The top 100 molecules with the highest binding affinity were tested for ADMET properties using the admet SAR web tool. Six compounds that satisfied the given conditions were then subjected to a more intensive binding affinity test using the AutoDock Vina program of the PyRx software. The molecule with the highest binding affinity to HA, *6-(4-Methylpiperidin-1-yl) sulfonyl-1-azatricyclo [6.3.1.04, 12]dodeca-4,6,8(12)-trien-2-one*, was selected as the lead molecule that could be developed further as a potential drug against the Influenza A virus.

INTRODUCTION:

The Influenza Virus is Composed of 11 Proteins: Hemagglutinin (HA), Neuraminidase (NA), matrix protein 1 (M1), matrix-2 proton channel (M2), nucleoprotein (NP), non-structural protein 1 (NS1), nuclear export protein (NEP; formerly known as NS2), polymerase acid protein (PA), polymerase basic proteins (PB1 and PB2), and PB1-F2 [1]. The viral envelope around each virion consists of two glycoproteins, HA and NA [1]. Anti-Influenza drugs normally target the M2 Channel Proteins, NA Proteins, and Cap-Endonucleases due to their important roles in the virus life cycle [2].

The M2 blockers stabilize the virus budding site by balancing the pH levels across the virus and the host cell, while the NA & Cap-Endonucleases play an important role in viral replication [2]. Currently, the drugs being used to treat the Influenza virus in the US include Oseltamivir (Tamiflu), Zanamivir (Relenza) & Peramivir (Rapivab), which are NA inhibitors; Rimantadine & Amantadine, which are M2 Channel Blocker Inhibitors and finally, the recently approved Baloxavir, which is a Cap Endonuclease Inhibitor [1, 2].

Unfortunately, all these therapies have their limitations and side effects. Further, there could also be a growing resistance to these drugs due to the virus' ability to adapt to changes in the host environment because of the rapid mutations observed in the virus [3, 4]. Hence, there is an urgent need for the ongoing development of novel and more effective therapies for combating the Influenza virus.

The HA protein is an abundant glycoprotein located on the virus's outer shell. It facilitates the binding of the virion to sialic acid sugars located on the surface of skin cells before further interaction such as RNA replication occurs. The host cell then engulfs the virus and forms an endosome with the virus particles, which it later attempts to digest. During this process, the interior of the endosome is acidified. At a certain pH level, the HA protein undergoes extensive conformational changes and releases a fusion peptide; this allows the endosome to fuse with the virus particle membrane to continue with the replication process. Thus, the HA protein plays a vital role in the entrance and the replication processes of the Influenza virion. Through its inhibition, many key mechanisms of the Influenza virions will be terminated; hence, the HA protein is an ideal target for anti-Influenza drugs [5, 6, 7].

In-silico, or computational methods, have been widely utilized in pharmacology to rapidly discover and short-list compounds by calculating binding capacities, drug-likeness properties, and ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) [8, 9].

Molecular docking is a type of bioinformatic modelling that has increasingly been considered an important tool in *in-silico* drug discovery. Compared with traditional experimental high-throughput screening, molecular docking is a virtual, high throughput screening to score for lead molecules from large libraries of compounds. This approach has the advantage of low cost and preliminary effective screening against a chosen drug target [10, 11].

Molecular docking involves assessing and predicting the interactions between two molecules to determine if it is a stable adduct [12]. These include shape complementarity tests, based primarily on the shape of the drug target and how a ligand can best fit into a designated binding site on the target, and secondly, the use of interaction simulation tests to determine the optimized confirmation between a ligand and drug target so that the free energy of the overall system is minimized [13]. This study will aim to identify ligands that can selectively bind to the HA protein using *in-silico* molecular docking methods.

MATERIALS AND METHODS:

Retrieval and Preparation of the HA Protein: The Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) was used to retrieve the 3-dimensional crystal structure of HA from the Influenza virus A/Pennsylvania/14/2010 of the H3N2 strain. With a PDB code of 6X6P, this file was downloaded and saved in the Brookhaven protein data bank (pdb) file format [14]. The FTMap Computational Mapping Web Tool was used to determine the best possible binding site on the HA protein [15]. The FTMap web tool employs sixteen distinct small organic molecules, known as "probe" molecules, that revolve about the surface of a certain macromolecule and bind or interact with regions that contribute to the ligands' binding free energy. These regions, known as drug gable "hotspots," feature a concave structure with a pattern of protein-to-ligand interactions; hence, they possess a higher tendency to bind with organic compounds. The amino acid residues of the defined hotspots and the number of total clusters at each residue were analyzed based on interaction rates to locate a proper binding site [15, 16].

Retrieval and Preparation of the Ligands: The ZINC15 small molecule database was used to retrieve over 68 million compounds with their 3-dimensional structures [17]. Using the ZINC Tranche Browser to select subsets of compounds based on their physical properties, these molecules were filtered for standard reactivity, neutral net charges, and physiologically relevant pH of ~7.4. All the selected molecules were additionally subject to screening through Chris Lipinski's Rule of 5, a set of criteria used to identify small molecules that will likely be absorbed and permeated easily in *in-vivo* systems [18]. The criteria for Lipinski's Rule of 5 are:

❖ Molecular weight under 500 g/mol.

❖ Value of logP partitioning coefficient lower than 5.

❖ Fewer than 5 hydrogen bond donors.

❖ Fewer than 10 hydrogen bond acceptors.

None of these drug-likeness constraints can be violated, or the ligands will not have properties suitable for drug-like compounds [18].

Preliminary Molecular Docking: A preliminary docking test was performed using QIAGEN Bioinformatics' CLC Drug Discovery Workbench Software version 3.0.2. CLC Drug Discovery uses computational techniques to dock ligands and gives a docking score and access to atomic level interactions in the docking process [19]. CLC Drug Discovery Workbench was used to find the molecules with the highest binding affinity to the chosen binding site, found earlier using the FTMap web tool. A very negative score indicates a strong binding, while a less negative or positive score indicates a non-existent or weak binding [20]. The scoring formula is:

$$Score = S_{target\text{-}ligand} + S_{ligand}$$

The HA Protein was uploaded on the CLC Drug Discovery Software along with the previously chosen drug-like molecules. Each molecule was tested against the chosen binding site of the HA Protein.

Analysis of ADMET Properties: The top 100 molecules with the highest binding affinity based on the previous stage then moved on to the ADMET verification step, where each of these molecules was tested for ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties using the admet SAR 2.0 web tool [21].

The admet SAR online web tool was developed by the Shanghai Key Laboratory of New Drug Design and uses over 200,000 ADMET annotated data points from about 96,000 different compounds to virtually test the properties of small molecules [21]. The properties to be tested included blood-brain barrier (BBB+), human intestinal absorption (HIA+), CaCO-2 permeability (CaCO-2+),

carcinogenicity (carcinogenicity-) and human ether-a-go-go-related gene inhibition (hERG)[21]. The top 100 molecules in the. SDF format was converted to the SMILES format using the Open Babel GUI software to facilitate this testing process [22].

Final Docking: The molecules that passed the ADMET properties in the previous analyses were then put through a more comprehensive binding energy test using the AutoDock Vina program in the PyRX software, version 0.8 [23]. The PyRx software is a virtual screening software for computational drug discovery that can be used to screen large libraries of compounds against potential drug targets [23].

The AutoDock Vina program uses a sophisticated gradient optimization algorithm for computational modeling and considers many aspects of binding (the number of atoms, the number of torsions, the size of the search space and the exhaustiveness of the search, *etc.*) to determine the binding energy of a ligand to its corresponding protein. These results identified a lead molecule with the best binding properties to the HA protein.

RESULTS AND DISCUSSION: To prepare the HA Protein for molecular docking, the obtained crystal structure of the HA molecule was inputted into the FTMap Computational Mapping Web tool and its best binding spot was identified based on non-bonded and hydrogen-bonded interaction levels across the surface of the macromolecule. It was decided that the binding site would extend to twenty angstroms past the identified site in all directions to provide a suitable binding environment size for the ligands [15].
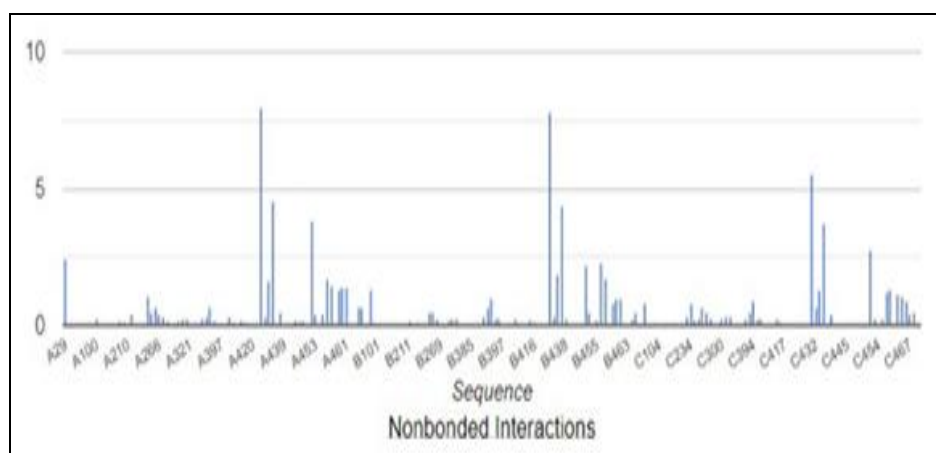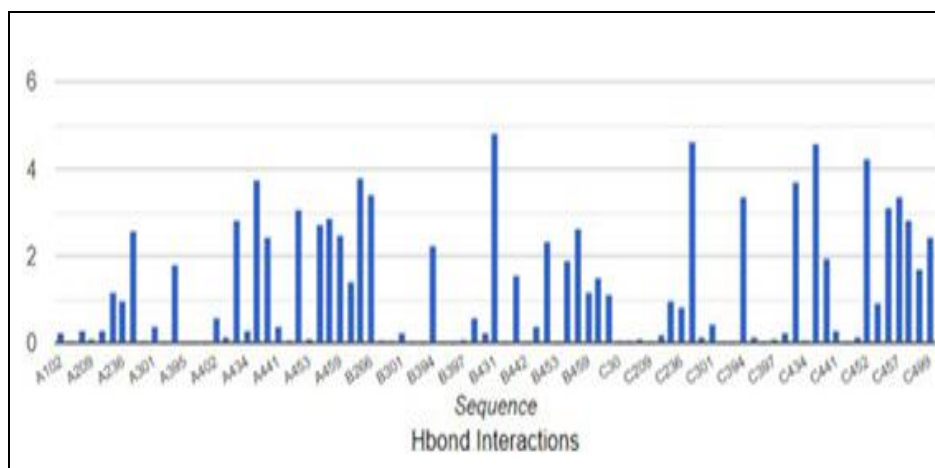
FIG. 1: THE GRAPHS ABOVE WERE GENERATED BY THE FTMAP COMPUTATIONAL MAPPING WEBTOOL AND DISPLAY THE INTERACTION LEVELS OF THE PROTEIN AT DIFFERENT SURFACE AREAS ON THE HA PROTEIN. ON THE GRAPHS, THE Y-AXIS DISPLAYS THE CONTACT RATE AS A PERCENT OF TOTAL CONTACTS. THIS SHOWS THE LEVEL OF INTERACTION AS A PERCENT OF THE TOTAL INTERACTIONS BETWEEN THE PROTEIN AND PROBE MOLECULES. THE X-AXIS SHOWS THE RESIDUE, OR AMINO ACID, ON THE PROTEIN AT WHICH THESE INTERACTIONS OCCUR AT. THE GRAPH AT THE TOP DEPICTS THE NON-BONDED INTERACTIONS SUCH AS ELECTROSTATIC AND VAN DER WAALS INTERACTIONS. THE GRAPH AT THE BOTTOM DEPICTS HYDROGEN-BONDED INTERACTIONS. CLUSTERS AND PEAKS ON THE GRAPH SHOW AREAS WITH HIGHER INTERACTION LEVELS, WHICH SIGNIFIES BETTER BINDING FOR ORGANIC COMPOUNDS [15]

Further, the binding site is located on the polypeptide chain of the HA protein that is associated with the attack of the virus on the cell, making it an extremely effective target site [6]. Binding to this site will alter the shape of this polypeptide chain, which will alter the protein's ability to enter the Influenza virion into the host cell. This specific polypeptide chain is likely common as a binding site in previous studies because of its high interaction levels compared to the other sites on the protein **Fig. 1.**

Over 68 million ligands from the ZINC15 small-molecule database were subjected to a drug-likeness filter which screened the molecules using Chris Lipinski's Rule of 5. The molecules were further filtered for other properties, including standard reactivity, neutral net charges, and a physiologically relevant pH of ~7.4, all of which narrowed the library of molecules down to around 250,000 compounds [17].

These 250,000 molecules were then tested against the HA protein for their binding affinity using the CLC Drug Discovery Workbench Software. Stronger binding to the protein is signified through a higher binding affinity [20]. The top 100 molecules with the highest binding affinity were identified as potential HA inhibitors. These top 100 molecules were then subjected to the ADMET verification stage, where they were each tested for ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties. Only six compounds passed the given ADMET properties and moved onto the lead molecule declaration phase **Table 1.**

TABLE 1: TABLE 1 DEPICTS A FEW OF THE IMPORTANT ADMET (ABSORPTION, DISTRIBUTION, METABOLISM, EXCRETION, AND TOXICITY) PROPERTIES THAT DETERMINE THE DISPOSITION OF A COMPOUND IN BIOLOGICAL SYSTEMS AND THEIR DESIRED OUTCOME IN TERMS OF THE SOFTWARE

Blood Brain Barrier	Human Intestinal Absorption	Caco-2 Permeability	Human Oral Bioavailability	Carcinogenicity	Human Ether a-go-go Related Gene Inhibition
Yes	Yes	Yes	Yes	No	No

While each of these six molecules has already surpassed the chosen ADMET and docking tests, to further narrow the results and choose the best possible lead molecule out of these given six ligands, a much slower binding energy test was performed on each of these molecules using the AUTODOCK Vina program in the PyRx software [23]. With a binding affinity of 8.2 kcal/mol, *6-(4-*

Methylpiperidin - 1 - yl) sulfonyl - 1 - azatricyclo [6.3.1.04, 12] dodeca - 4, 6, 8 (12) – trien – 2 - one

was declared the lead molecule in this study **Fig. 2, Fig. 3.**

FIG. 2: MOLECULAR STRUCTURE OF THE 6-(4-METHYLPIPERIDIN-1-YL) SULFONYL-1-AZATRICYCLO [6.3.1.04, 12] DODECA-4,6, 8(12)-TRIEN-2-ONE
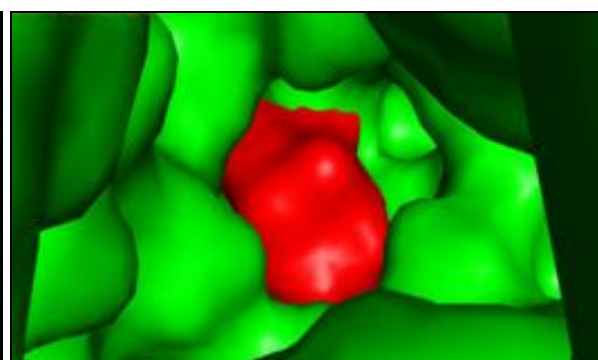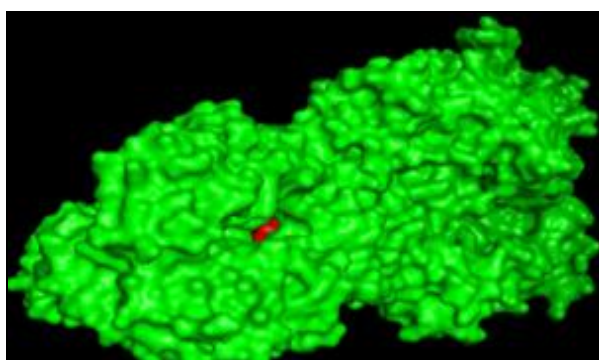

FIG. 3 AND 4: DOCKING POSE OF THE 6-(4-METHYLPIPERIDIN-1-YL) SULFONYL-1-AZATRICYCLO [6.3.1.04,12]DODECA-4,6,8(12)-TRIEN-2-ONE

A t-test was performed to ensure that the data is statistically significant to determine if the binding affinity towards HA protein and the lead molecule in this study were drawn from different populations with a significant difference within their population means. This would show that the HA values for the lead molecule in the study are statistically significant and not just random data. The null hypothesis in this experiment would be that there is no difference in the binding affinity. The alternative hypothesis would be that the lead molecule would have a binding affinity significantly poorer than the other molecules. There was over a 99.99% confidence that the data is not random.

All the studies to which we have compared our research to have determined different lead drug candidates that have the potential to inhibit the HA protein. Each ligand is different in its molecular conformation and size, contributing to a different binding free energy. Each study uses different methodologies and unique software packages, resulting in varied outcomes. Therefore, it is

unknown whether a drug can truly be effective or beneficial without pursuing in-vitro studies, and hence, it is hard to compare lead molecules in this stage. Therefore, we have compared our methodology to other studies instead of comparing lead molecules to validate our findings.

Aurélie Perrier *et al.,* found ligands to inhibit the membrane fusion process of the Influenza virion by targeting the HA protein [24]. However, their study primarily focused on interaction levels between the protein and ligands, while our study focuses on a broader interaction relationship while also considering other aspects of drug discovery such as ADMET and drug-likeness properties to identify a more well-rounded potential drug molecule.

Yao *et al.* used a different methodology involving a fluorescence polarization probe to achieve high-throughput screening and identify potential drug candidates against the HA protein [25]. The algorithmic software methodology used in our study was able to screen a much larger number of compounds and provided much quicker results. Jin

Il Kim *et al*. used a high-throughput screening method with a green fluorescent protein-tagged recombinant Influenza virus and discovered IY7640 as a Hemagglutinin targeting anti-Influenza drug and targeted the Influenza B virus specifically [26]. While the Influenza B virus has been more predominant in recent years, the Influenza A virus mutates almost three times faster than the B virus, making it an equally important target. Recently, attempts have been made to discover drugs that work against multiple proteins simultaneously, usually targeting the HA and NA proteins of the Influenza virus. In this direction, we can further evaluate the top drug candidates found in this current study and test them against NA and other Influenza virus proteins to determine their potential for efficacy against multiple targets [27].

As one can understand, biological systems are highly complex. Despite the technological advancements, these systems are algorithmically challenging to replicate due to inaccuracy related to protein flexibility, molecule conformational changes *in-vivo* and other host factors. Further, the promiscuity in these software-driven methods hinders accurate predictions. As a result, *in-silico* drug discovery always has a level of serendipity associated with it, contributing to the drawbacks of in-silico testing. However, recent advances in technology have allowed in-silico pharmacology to reduce the number of molecules that need to be made and tested and increase the speed of the drug discovery process, which ultimately results in large reductions in the cost of developing a drug and, therefore, is increasingly being used as an alternative to lab-like testing[9, 28].

CONCLUSION: In this study, an attempt was made to develop and find a lead compound that can successfully bind to the HA protein of the Influenza A virus from a large database of chemical ligands. A lead molecule was successfully declared with the elemental formula of *6-(4-Methylpiperidin-1-yl) sulfonyl-1-azatricyclo [6.3.1.04, 12] dodeca-4,6,8(12)-trien-2-one* and this compound demonstrated excellent binding affinity to HA protein along with excellent drug-likeness properties and ADMET properties. Our lead compound can further be tested *in-vivo* for HA inhibition and can be developed as a potential drug for effectively treating Influenza A infection.

REFERENCES:

1. Javanian M, Barary M, Ghebrehewet S, Koppolu V, Vasigala V and Ebrahimpour S: A brief review of influenza virus infection. Journal of Medical Virology 2021; 93(8): 4638-4646.
2. Principi N, Camilloni B, Alunno A, Polinori I, Argentiero A and Esposito S: Drugs for influenza treatment: Is there significant news. Front Med 2019; 6: 109.
3. Farrukee R and Hurt AC: Antiviral Drugs for the Treatment and Prevention of Influenza. Curr Treat Options Infect Dis 2017; 9: 318–332.
4. Lampejo T: Influenza and antiviral resistance: an overview. Eur J Clin Microbiol Infect Dis 2020; 39: 1202-1208.
5. Wi N and Wilson IA: Structural Biology of Influenza Hemagglutinin: An Amaranthine Adventure. Viruses 2020; 12(9): 1053.
6. Donald BJ: Structural transitions in Influenza Hemagglutinin at Membrane Fusion pH. Nature 2020; 583(7814): 150-153.
7. Pabis A, Rawle R and Kasson P: Influenza hemagglutinin drives viral entry *via* two sequential intramembrane mechanisms. Proc Nat Acad Sci USA 2020; 117(13): 7200.
8. Rognan D: The impact of in silico screening in the discovery of novel and safer drug candidates. Pharma and Therapeutics Elsevier 2017; 175: 47-66.
9. Clent BA, Wang Y, Britton HC, Otto F, Swain CJ, Todd MH, Wilden JD and Tabor AB: Molecular Docking with Open Access Software: Development of an Online Laboratory Handbook with Remote Workflow for Chemistry and Pharmacy Master's Students to Undertake Computer-Aided Drug Design. J Chem Educ 2021; 98(9): 2899-2905.
10. Pedro HMT: Key Topics in Molecular Docking for Drug Design. International Journal of Molecular Sciences 2019; 20(18): 4574.
11. Bhagat RT, Butle SR, Khobragade DS, Wankhede SB, Prasad CC, Mahure DS and Armarkar AV: Molecular docking in Drug Discovery. Journal of Pharmaceutical Research International 2021; 33(30): 46-58.
12. Dar AM and Mir S: Molecular Docking- Approaches, types, applications and basic challenges. J Anal Bioanal Tech 2017; 8: 356.
13. Spyrakis F, Benedetti P, Decherchi S, Rocchia W, Cavalli A, Alcaro S, Ortuso F, Baroni M and Cruciani G: A pipeline to enhance ligand virtual screening- Integrating molecular dynamics and fingerprints for ligand and proteins. J Chem Inf Model 2015; 55(10): 2256-74.
14. Burley SK: Protein Data Bank (PDB): The single global macromolecule structure archive. Methods in Molecular Biology 2017; 1607: 627-641.
15. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D and Vajda S: The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc 2015; 10(5): 733-755.
16. Beglov D, Hall Dr, Wakefield AE and Vajda S: Exoploring the structural origins of cryptic sites on proteins. Proc Natl Acad Sci USA 2018; 115(315): 3416-25.

17. Sterling T and Irwin JJ: ZINC 15 – Ligand discovery for everyone. J Chem Inf Model 2015; 55(11): 2324-37.
18. Benet LZ, Hosey CM, Ursu O and Oprea TI: BDDCS, The rule of 5 and druggability. Adv Drug Deliv Rev 2016; 101: 89-98.
19. Introduction to CLC Drug Discovery Workbench. Available online: http://resources.qiagenbioinformatics.com/manuals/clcdrugdiscoveryworkbench/current/index.php?manual=Introduction_CLC_Drug_Discovery_Workbench.html. (Accessed 06 June 2021).
20. Pagadala NS, Syed K and Tuszynski J: Software for molecular docking: a review. Biophys Rev 2017; 9(2): 91-102.
21. Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z, Li W, Liu G and Tang Y: admet SAR 2.0: Web-service for prediction and optimization of chemical ADMET properties. Nucleic Acids Res 2019; 35(6): 1067-9.
22. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T and Hutchison GR: Open Babel- An open chemical toolbox. J of Cheminformatics 2011; 3: 33.
23. Trott O and Olson AJ: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010; 31(2): 455-461.
24. Perrier A, Eluard M, Petitjean M and Vanet A: *In-silico* design of new inhibitors against hemagglutinin of Influenza. JPhys Chem B 2019; 123(3): 582-592.
25. YaoY, Kadam R, Lee C, WoehlJ, WuN, Zhu X, Kitamura S, Wilson I and Wolan D: An Influenza A hemagglutinin small-molecule fusion inhibitor identified by a new high-throughput fluorescence polarization screen. Proc Natl Acad Sci USA 2020; 117(31): 18431-18438.
26. Kim JI, Lee S, Lee GY, Park S, Bae JY, Heo J, Kim HY, Woo SH, Lee HU, Ahn CA, Bang HJ, Ju HS, Ok K, Byun Y, Cho DJ, Shin JS, Kim DY, Park MS and Park MS: Novel small molecule targeting the hemagglutinin stalk of influenza viruses. J Virol 2019; 93(17): e00878-19.
27. Xu L, Jiang W, Jia H, Zheng L, Xing J, Liu A and Du G: Discovery of multitarget-directed ligands against Influenza A virus from compound Yizhihaothrough a predictive system for compound-protein interactions. Front Cell Infect Microbiol 2020; 10: 16.
28. Chebrolu A and Madhavan S: Molecular docking study of ibuprofen derivatives as selective inhibitors of cyclooxygenase-2. IJPSR 2020; 11(12): 6526-31.