Life Science Informatics Publications
Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences
Journal Home page http://www.rjlbpcs.com/
RJLBPCS
ISSN 2454-6348

Original Research Article **DOI: 10.26479/2020.0602.05**

COMPUTATIONAL IDENTIFICATION OF NOVEL INHIBITORS OF THE MAIN PROTEASE OF THE NOVEL CORONAVIRUS (2019-nCoV)

Rithvik Ganesh*

Jasper High School, 6800 Archgate Dr, Plano, TX 75024, USA.

ABSTRACT: The novel Coronavirus (2019-nCoV), a closely related pathogen to the respiratory syndrome CoV (SARS-CoV), has caused thousands of worldwide death along with severe economic damage. The main protease of the coronavirus plays a crucial role in the viral gene expression through polyprotein processing, making it an appealing target linked to drug design. Inhibitors of the main protease of the coronavirus are hypothesized to help diminish the effects of the virus as a whole. Over 11 million compounds from the Mcule Database were screened against two potential anti-COVID compounds in a structure-based virtual screen. The top 233 compounds were then subject to molecular docking trial with AUTODOCK VINA. The top 74 compounds were then subject to a Compound-Protein Interaction (CPI) neural network algorithm. Finally, the top 30 compounds with positive classifications were subject to ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) property verification. After analysis, two top compounds were identified as potential 2019-nCoV inhibitors.

Keywords: 2019-nCoV, CPI Prediction, Molecular Docking, ADMET Verification, Mcule.

Article History: Received: March 30, 2020; Revised: April 09, 2020; Accepted: April 22, 2020.

Corresponding Author: Rithvik Ganesh*

Jasper High School, 6800 Archgate Dr, Plano, TX 75024, USA.

1. INTRODUCTION

Coronaviruses (CoVs) are a large family of viruses and severe pathogens that are involved in a plethora of diseases in numerous species [1, 2, 3]. Severe acute respiratory syndrome (SARS-CoV), was the cause of approximately 800 deaths in the year 2003. CoVs, which are a genus of the *Coronaviridae* family, have been cited as enveloped with a large plus-strand RNA genome [4, 5]. CoVs specifically can cause diseases including respiratory-related diseases [6, 7]. Unfortunately,

there are no available vaccines, antiviral, or other specific treatments for human and animal coronavirus, which drastically escalates the need for drug-like inhibitors of the 2019-nCoV pathogen. A handful of approved medicines have been cited as potential 2019-nCoV inhibitors, such as Remdesivir and Chloroquine [8, 9, 10]. Other recent scientific advances have pointed to phytochemical-based treatment approaches for vaccine and biopharmaceutical production [11, 12]. The development of novel and effective biopharmaceutical anti-COVID compounds is an urgent need of today's world. The main protease of the coronavirus plays a crucial role in the viral gene expression through polyprotein processing, making it an appealing target linked to drug design. Viral Replication in the virus has been seen as an extremely viable drug discovery path. The main protease of the virus could cut two replicase polyproteins, which mediate the function of the virus [13, 14, 15]. An inhibitor could interfere with this function, which would subsequently shut down the harmful virus replication [16]. *In silico* (computational) methods have been successfully employed in the pharmacological fields of science to identify inhibitors with substantial binding energies, drug-like and lead-like properties, stability with a given target, and ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) clarification. *In silico* methods employ either structural-based drug design methods or ligand-based drug design techniques to identify potential drug-like compounds [17, 18, 19, 20]. Computational screening methodologies have been used to identify various unique and diverse inhibitors for a wide spectrum of diseases [21, 22, 23]. Machine learning-based methods have been implemented in various portions of the *in silico* drug development space. Compound-Protein Interactions (CPIs) play an important role in the virtual screening of drug discoveries [24, 25]. Recently, deep learning methods have been used to identify potential patterns in the physics, mathematics, and chemistry between the binding mechanism of successful inhibitors of proteins. Convolutional Neural Networks and Graph Neural Networks have been successful architectures used in CPI algorithms [26, 27, 28]. In this study, an attempt has been made to identify inhibitors of the main protease of the coronavirus for an 2019-nCoV treatment. A robust *in silico* methodology has been implemented to identify these inhibitors in a quick, effective, and efficient screening process.

2. MATERIALS AND METHODS

Protein Target Preparation

The 3D structure of the crystal structure of COVID-19 main protease in complex with an inhibitor N3 was retrieved from the RCSB PDB database (ID: 6LU7) [29]. The experimental data snapshot was determine using X-Ray Diffraction and was resolved to 2.16 Å. The protein was refined, optimized, preprocessed, and modified with the Maestro Protein Preparation wizard [30].

Structural Similarity Virtual Screen

A database of precisely 11,201,268 compounds were retrieved from the Mcule Stock Compounds, with known synthesis routes. The 2-D Structures of the compounds Remdesivir and Chloroquine

were retrieved from the PubChem Chemical Database [31]. The 11,201,268 compounds were screened against Remdesivir and Chloroquine using a Tanimoto Coefficient, to identify compounds which may maintain key positive structural characteristics of the two while being diverse and unique. The Tanimoto (or Jaccard) coefficient T is the most popular similarity measure for comparing chemical structure through fingerprint-based analysis [32, 33].

Grid Generation and Molecular Docking

A receptor grid file was generated using the GLIDE module of the Schrodinger Biosuite. A co-crystallized ligand was then separated from the active site of the corresponding receptor chain. This format of the crystal structure of COVID-19 main protease was saved for docking with AUTODOCK VINA. The pure protein structure was uploaded as PDB file to the docking platform, using in-built AUTODOCK VINA Tools on the Mcule Interface [34, 35]. Using a combination of polar hydrogen atoms and Gasteiger charges, a binding site of the COVID-19 main protease was determined. Using an exhaustiveness of 2, the top 74 compounds with substantial binding energies were advanced to the next research phase.

Compound-Protein Interaction (CPI) Neural Network

The top 74 compounds were now subject to a Compound-protein interaction (CPI) prediction algorithm with end-to-end learning of neural networks for subsequent graphs and sequences by Tsubaki et. al [36, 37, 38]. For the CPI prediction algorithm, graphically-analyzed compounds and amino-acid protein sequences are analyzed using vector-clustering algorithms. A training database of known compound-protein interactions are used and analyzed using ROC curves with an AUC value of 0.99. The 74 compounds, paired with the refined protein target sequence, are then inputted into the CPI algorithm, which outputs compound likely to interact. The 30 compounds with positive interaction probabilities then advanced to an ADMET and drug-likeliness verification stage.

ADMET Property Verification

In this study, several drug-likeliness properties were analyzed for the ligands with the best docking and CPI results. ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) Verification is critical to the capacity of anti-COVID compounds [39]. After Rapid-Screening preliminary docking, the compounds were tested against Lipinski, Ghose, Veber, Egan, and Muegge filters, which are applied to classify drug-like compounds [40, 41, 42, 43, 44]. Lead-likeliness and drug-likeliness scores from SwissAdme were used as another threshold for analysis [45]. The prediction of these various pharmacokinetic features were utilized by tools that factor molecular similarity and predictive regression to analyze submitted compounds and ligands. The top 30 compounds were subject to various thresholds and analyses using a plethora of regression tools to validate key ADMET and pharmacokinetic properties.

3. RESULTS AND DISCUSSION

The 11,201,268 compounds in this study were subject to a robust filtration methodology, featuring structural similarity assessments, molecular docking, CPI Neural Networks, and ADMET Property Verification. Initially, the full Mcule was screened against two potential anti-COVID compounds using Tanimoto Coefficient Classification. The top 233 compounds that maintained key structural features but contained diverse scaffolds as well were then subject to a molecular docking trial with AUTODOCK VINA. Following protein and ligand processing, an exhaustive and flexible docking approach was used to identify compounds that had a substantial binding energy to the crystal structure of COVID-19 main protease. The top 74 compounds with substantial binding energy to the main protease were then subject to a Compound-Protein Interaction (CPI) neural network algorithm. A training database of compounds and proteins were analyzed with a ROC curve and AUC value of 0.99. A Graphical and Convolutional Neural Network, for the compound and protein respectively, were then used to identify potential inhibitors based on the trained database. The top 30 compounds with positive classifications were subject to ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) property verification criteria. After analysis with SwissADME tools, the top two compounds were identified as 2019-nCoV inhibitors.

Table 1: Structural Similarity and Molecular Docking Results

Ligand IUPAC Name	Ligand InChlKey	VINA Docking Score	Tanimoto Coefficient
[(3R,4R)-1-(6-Methyl-4-quinolinyl)-4-phenyl-3-piperidinyl]methanol	XCVCHJUVCOKJII-MOPGFXCFSA-N	-7.8	0.562
2-Benzyl-8-(4-quinolinyl)-2,8-diazaspiro[4.5]decan-3-one	PZCZBEBMWKYSQT-UHFFFAOYSA-N	-7.3	0.589

Table 2: CPI Neural Network Results

Ligand IUPAC Name	AUC Value	Interaction Probability	Non-Interaction Probability
[(3R,4R)-1-(6-Methyl-4-quinolinyl)-4-phenyl-3-piperidinyl]methanol	0.99	0.987	0.013
2-Benzyl-8-(4-quinolinyl)-2,8-diazaspiro[4.5]decan-3-one	0.99	0.981	0.019

Table 3: Drug-like properties of the lead ligands from SwissADME

Ligand IUPAC Name	Lipinski	Ghose	Veber	Egan	Muegge	Solubilty
[(3R,4R)-1-(6-Methyl-4-quinolinyl)-4-phenyl-3-piperidinyl]methanol	YES	YES	YES	YES	YES	YES
2-Benzyl-8-(4-quinolinyl)-2,8-diazaspiro[4.5]decan-3-one	YES	YES	YES	YES	YES	YES

Table 1 depicts the structural similarity and molecular docking results of the two lead ligands in the study. As seen here, the structural similarity coefficients show similarity two successful *in vitro* anti-COVID compounds, while maintaining individual scaffolds for variety and diversity. The AUTODOCK VINA docking score is substantial, showing high potential for chemical interaction

Table 2 depicts the Compound-Protein Interaction Neural Network algorithm results. The Area Under the Curve (AUC) value of 0.99 for the training set is excellent for careful discrimination of inhibitors. The High Positive Interaction Probability for both ligands is an excellent indication, consistent with ligand-based docking results.

Table 3 depicts the positive ADMET properties, consistent with Lipinski, Ghose, Veber, Egan, and Muegge drug classification rules. Among other positive indications from the SwissADME regression tool, there was a positive indication for solubility for both ligands. Solubility was measured using 3 generated criteria. Solubility is crucial to ADMET-based absorption for compounds. The first metric was an implementation of the ESOL model [46], the second was aqueous solubility model developed by Ali et al. [47], and the final metric was developed by SILICOS-IT [48].

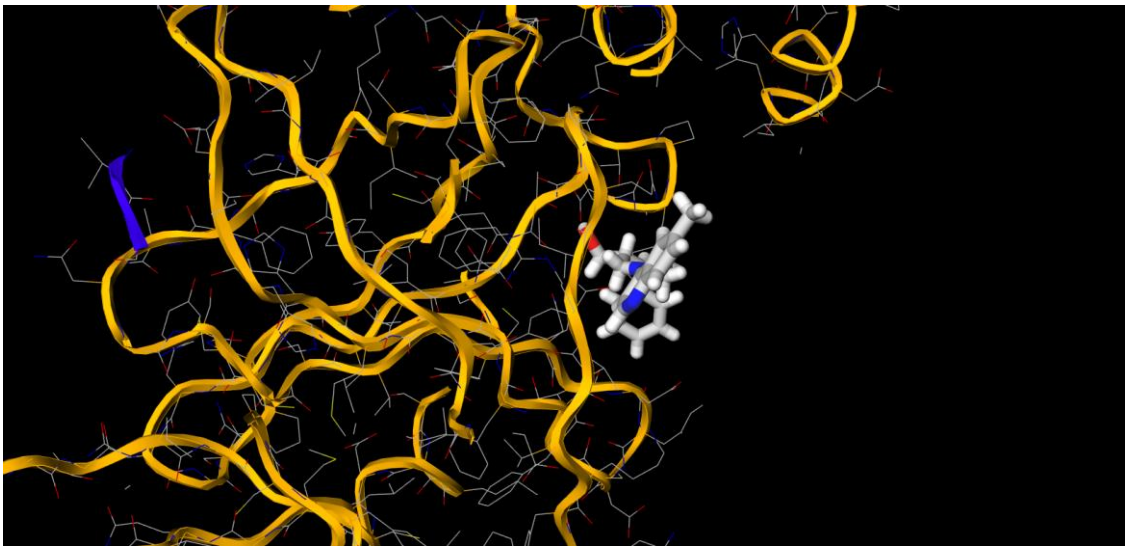

Figure 1a: Images of the docking poses of [(3R,4R)-1-(6-Methyl-4-quinolinyl)-4-phenyl-3-piperidinyl]methanol with the crystal structure of COVID-19 main protease

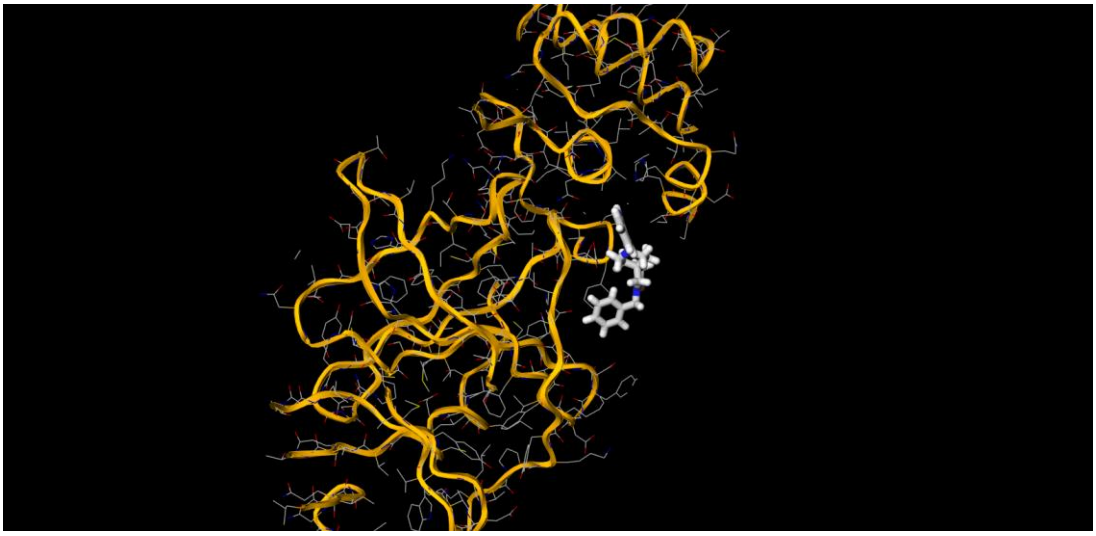

Figure 1b: Images of the docking poses of 2-Benzyl-8-(4-quinolinyl)-2,8-diazaspiro[4.5]decan-3-one with the crystal structure of COVID-19 main protease

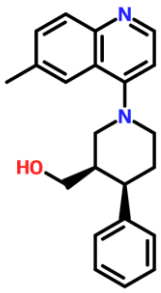

Figure 2a: Image of [(3R,4R)-1-(6-Methyl-4-quinolinyl)-4-phenyl-3-piperidinyl]methanol

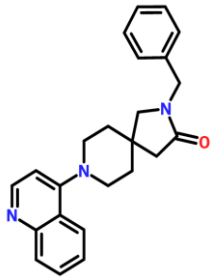

Figure 2a: Image of 2-Benzyl-8-(4-quinolinyl)-2,8-diazaspiro[4.5]decan-3-one

4. CONCLUSION

The main protease of COVID-19 facilitates the virus' key processes in patients. The inhibition of this protein can help control the progression of COVID-19 in patients, and could have potential in other similar viruses in the same family. In this in silico study, with the aid of neural networks, molecular docking, and ADMET verification, [(3R,4R)-1-(6-Methyl-4-quinolinyl)-4-phenyl-3-piperidinyl] methanol and 2-Benzyl-8-(4-quinolinyl)-2,8-diazaspiro[4.5]decan-3-one have been identified as potential COVID-19 main protease inhibitors. These inhibitors could serve as effective drug candidates to control the progression of COVID-19 in patients.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The authors confirm that the data supporting the findings of this research are available within the article.

FUNDING

None.

ACKNOWLEDGEMENT

The researcher would like to thank Dr. Gaurav Chopra (Assistant Professor of Analytical and Physical Chemistry, Dept. of Chemistry, Purdue University) and Jonathan Fine from the Chopra Lab at Purdue University for assistance as well as references to software used in this research.

CONFLICT OF INTEREST

Authors have no conflict of interest.

REFERENCES

1. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B. Identification of a new human coronavirus. Nature medicine. 2004 Apr;10(4):368-73.
2. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. InCoronaviruses 2015 (pp. 1-23). Humana Press, New York, NY.
3. Cavanagh D. Coronavirus avian infectious bronchitis virus. Veterinary research. 2007 Mar 1;38(2):281-97.
4. Siddell SG, Anderson R, Cavanagh D, Fujiwara K, Klenk HD, Macnaughton MR, Pensaert M, Stohlman SA, Sturman L, Van der Zeijst BA. Coronaviridae. Intervirology. 1983;20(4):181-9.
5. Gonzalez JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, Enjuanes L. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. Archives of virology. 2003 Nov 1;148(11):2207-35.
6. Tang JW, Tambyah PA, Hui DS. Emergence of a novel coronavirus causing respiratory illness from Wuhan, China. Journal of Infection. 2020 Mar 1;80(3):350-71.
7. Ferrey AJ, Choi G, Hanna RM, Chang Y, Tantisattamo E, Ivaturi K, Park E, Nguyen L, Wang B, Tonthat S, Rhee CM. A Case of Novel Coronavirus Disease 19 in a Chronic Hemodialysis Patient Presenting with Gastroenteritis and Developing Severe Pulmonary Disease. American Journal of Nephrology. 2020 Mar 28:1-6.
8. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell research. 2020 Mar;30(3):269-71.
9. Colson P, Rolain JM, Raoult D. Chloroquine for the 2019 novel coronavirus. Int J Antimicrob Agents. 2020 Feb 15.
10. Gao J, Tian Z, Yang X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Bioscience trends. 2020.
11. Varshney KK, Varshney M, Nath B. Molecular Modeling of Isolated Phytochemicals from Ocimum Sanctum Towards Exploring Potential Inhibitors of SARS Coronavirus Main Protease and Papain-Like Protease to Treat COVID-19. Available at SSRN 3554371. 2020 Mar 14.
12. Chandel V, Raj S, Rathi B, Kumar D. In Silico Identification of Potent COVID-19 Main Protease Inhibitors from FDA Approved Antiviral Compounds and Active Phytochemicals through Molecular Docking: A Drug Repurposing Approach.
13. Giri R, Bhardwaj T, Shegane M, Gehi BR, Kumar P, Gadhave K, Oldfield CJ, Uversky VN. When Darkness Becomes a Ray of Light in the Dark Times: Understanding the COVID-19 via the Comparative Analysis of the Dark Proteomes of SARS-CoV-2, Human SARS and Bat SARS-Like Coronaviruses. bioRxiv. 2020 Jan 1.

14. Pendyala B, Patras A. In silico Screening of Food Bioactive Compounds to Predict Potential Inhibitors of COVID-19 Main protease (Mpro) and RNA-dependent RNA polymerase (RdRp).
15. Shanker A, Bhanu D, Alluri A. Analysis of Whole Genome Sequences and Homology Modelling of a 3C Like Peptidase and a Non-Structural Protein of the Novel Coronavirus COVID-19 Shows Protein Ligand Interaction with an Aza-Peptide and a Noncovalent Lead Inhibitor with Possible Antiviral Properties.
16. Kandeel M, Al-Nazawi M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. Life Sciences. 2020 Apr 3:117627.
17. Knehans T, Schüller A, Doan DN, Nacro K, Hill J, Güntert P, Madhusudhan MS, Weil T, Vasudevan SG. Structure-guided fragment-based in silico drug design of dengue protease inhibitors. Journal of computer-aided molecular design. 2011 Mar 1;25(3):263-74.
18. Kapetanovic IM. Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. Chemico-biological interactions. 2008 Jan 30;171(2):165-76.
19. Zoete V, Grosdidier A, Michielin O. Docking, virtual high throughput screening and in silico fragment based drug design. Journal of cellular and molecular medicine. 2009 Feb;13(2):238-48.
20. Cheng F, Li W, Liu G, Tang Y. In silico ADMET prediction: recent advances, current challenges and future trends. Current topics in medicinal chemistry. 2013 Jun 1;13(11):1273-89.
21. Ganesh R, Kannan I. Molecular Docking Study of Certain Plant Alkaloid Derivatives as Inhibitors of Various Drug Targets of Alzheimer's Disease. Biomedical and Pharmacology Journal. 2017 Sep 25;10(3):1489-94.
22. Rithvik Ganesh, In silico identification of Tumor Necrosis Factor Alpha inhibitors as treatments for Alzheimer's Disease, Int J Com Bio & Bioinfo, (2020); 1(1), 01 - 03.
23. Raghu G, Karunanithi A, Kannan I. Molecular docking study on curcumin and its derivatives as inhibitors of BACE1 in the treatment of Alzheimer's disease. National Journal of Physiology, Pharmacy and Pharmacology. 2018;8(2):244-50.
24. Bode W, Huber R. Ligand binding: proteinase-protein inhibitor interactions: Current opinion in structural biology 1991, 1: 45–52. Current Opinion in Structural Biology. 1991 Feb 1;1(1):45-52.
25. Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: current status and future challenges. Proteins: Structure, Function, and Bioinformatics. 2006 Oct 1;65(1):15-26.
26. Tian K, Shao M, Wang Y, Guan J, Zhou S. Boosting compound-protein interaction prediction by deep learning. Methods. 2016 Nov 1;110:64-72.
27. Wan F, Zeng J. Deep learning with feature embedding for compound-protein interaction prediction. bioRxiv. 2016 Jan 1:086033.

28. Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR. Protein–ligand scoring with convolutional neural networks. Journal of chemical information and modeling. 2017 Apr 24;57(4):942-57.
29. Liu X, Zhang B, Jin Z, Yang H, Rao Z. The crystal structure of COVID-19 main protease in complex with an inhibitor N3.
30. Schrodinger LL. The PyMOL molecular graphics system. Version. 2010 Aug;1(5):0.
31. Bolton EE, Wang Y, Thiessen PA, Bryant SH. PubChem: integrated platform of small molecules and biological activities. InAnnual reports in computational chemistry 2008 Jan 1 (Vol. 4, pp. 217-241). Elsevier.
32. Chung NC, Miasojedow B, Startek M, Gambin A. Jaccard/Tanimoto similarity test and estimation methods. arXiv preprint arXiv:1903.11372. 2019 Mar 27.
33. Fligner MA, Verducci JS, Blower PE. A modification of the Jaccard–Tanimoto similarity index for diverse selection of chemical compounds using binary strings. Technometrics. 2002 May 1;44(2):110-9.
34. Seeliger D, de Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. Journal of computer-aided molecular design. 2010 May 1;24(5):417-22.
35. Kiss R, Sandor M, Szalai FA. http://Mcule. com: a public web service for drug discovery. Journal of cheminformatics. 2012 Dec 1;4(S1):P17.
36. Tsubaki M, Tomii K, Sese J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics. 2019 Jan 15;35(2):309-18.
37. Liu H, Sun J, Guan J, Zheng J, Zhou S. Improving compound–protein interaction prediction by building up highly credible negative samples. Bioinformatics. 2015 Jun 15;31(12):i221-9.
38. Kwon S, Yoon S. Deepcci: End-to-end deep learning for chemical-chemical interaction prediction. InProceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics 2017 Aug 20 (pp. 203-212).
39. Sargunam P, Sridharan S. In-Silico Drug Designing of Novel Morpholino Based Physcion Drug Candidate and Investigation of Inhibition Effects on Covid-19 RNA Dependent-RNA Polymerase Non Structural Protein 12 (Nsp 12) with ADMET Study.
40. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews. 1997 Jan 15;23(1-3):3-25.
41. Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. Journal of combinatorial chemistry. 1999 Jan 12;1(1):55-68.

42. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. Journal of medicinal chemistry. 2002 Jun 6;45(12):2615-23.
43. Egan WJ, Merz KM, Baldwin JJ. Prediction of drug absorption using multivariate statistics. Journal of medicinal chemistry. 2000 Oct 19;43(21):3867-77.
44. Muegge I. Selection criteria for drug like compounds. Medicinal research reviews. 2003 May;23(3):302-21.
45. VSDNA NS, Muniappan M, Kannan I, Viswanathan S. Phytochemical analysis and docking study of compounds present in a polyherbal preparation used in the treatment of dermatophytosis. Current Medical Mycology. 2017 Dec;3(4):6.
46. Delaney JS. ESOL: estimating aqueous solubility directly from molecular structure. Journal of chemical information and computer sciences. 2004 May 24;44(3):1000-5.
47. Ali J, Camilleri P, Brown MB, Hutt AJ, Kirton SB. Revisiting the general solubility equation: in silico prediction of aqueous solubility incorporating the effect of topographical polar surface area. Journal of chemical information and modeling. 2012 Feb 27;52(2):420-8.
48. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties