Discovery of Potent Covid-19 Main Protease Inhibitors using Machine learning based Virtual Screening strategy

T. Muthu Kumar, K. Rohini, Nivya James, V. Shanthi and K. Ramanathan*

Department of Biotechnology

School of Bio Sciences and Technology

Vellore Institute of Technology, Vellore – 632014, Tamil Nadu, India.

* Corresponding author. Tel.: +91 4162202538; Fax: +91 4162243092.

E-mail: kramanathan@vit.ac.in

Abstract

The emergence and rapid spreading of novel SARS-CoV-2 across the globe represent an imminent threat to public health. Novel antiviral therapies are urgently needed to overcome this pandemic. Given the great role of main protease of Covid-19 for virus replication, we performed drug repurposing study using recently deposited main protease structure, 6LU7. For instance, pharmacophore- and e-pharmacophore-based hypotheses such as AARRH and AARR respectively were developed using available small molecule inhibitors and utilized in the screening of DrugBank repository. Further, hierarchical docking protocol was implemented with the support of Glide algorithm. The resultant compounds were then examined for its binding free energy against main protease of Covid-19 by means of Prime-MM/GBSA algorithm. Most importantly, the resultant compounds antiviral activities were further predicted by machine learning-based model generated by AutoQSAR algorithm. Finally, the hit molecules were examined for its drug likeness and its toxicity parameters through QikProp algorithm. Overall, the present analysis yielded four potential inhibitors (DB07800, DB08573, DB03744 and DB02986) that are predicted to bind with main protease of Covid-19 better than currently used drug molecules such as N3 (co-crystallized native ligand), Lopinavir and Ritonavir.

Key words: SARS-CoV-2 main protease; Glide; Prime-MM/GBSA; AutoQSAR; DrugBank

Introduction

The outbreak of coronavirus disease (COVID-19) has raised major health concern to humans worldwide. The novel Severe Acute Respiratory Syndrome Corona Virus-2 (SARS-CoV-2) has been identified as the causative pathogen which belongs to family *Coronoviridae* and genus *Betacoronavirus* [1, 2]. The global pandemic initiated in late December 2019 in Wuhan, capital of China's Hubei Province. Since then, it has swiftly spread across the world claiming thousands of lives. As of 24th March 2020, a total of 372757 confirmed cases have been reported internationally with 16231 cases of deaths [3]. The virus most likely originated from a zoonotic transmission from bats to humans and now has progressed to transmit from humans to humans. Faced with such a forbidding situation, World Health Organization (WHO) is also determined to working together with transport, travel and tourism sectors on emergency preparedness and response. Therefore, there is an imminent necessity to understand this novel virus and develop various measures to control its spread.

Recent studies have proposed that the full-length genome of SARS-CoV-2 is quite similar to SARS-CoV based on phylogenetic analysis [4, 5]. It was also found to exhibit putatively similar cell entry mechanism and human cell receptor utilization to that of SARS-CoV [4, 6, 7]. Considering this apparent similarity scientists have recently carried out preliminary research to identify potential vaccine target for COVID-19 based on SARS-CoV immunological studies [8]. However, no specific therapeutics is accessible to treat the infection indicating that, only clinical symptoms along with secondary infections could be treated with repurposed antiviral drug. Hence there is a dire need to develop potent therapeutics and vaccines against SARS-CoV-2.

One of the attractive target for anti-CoV drug design is coronavirus main protease. It plays a vital role in the viral gene expression and replication through proteolytic processing of polyproteins [9]. The crystal structure of SARS-CoV-2 main protease was recently elucidated

to enable designing of specific protease inhibitors [10]. Even though the main proteases of SARS-CoV-2 and SARS-CoV are closely related with a sequence identity of 96.1% [11], [12], the drugs developed for SARS-CoV could not be suggested for the treatment as they remained in pre-clinical stage [13]. Several studies have been carried out to investigate the inhibitory activity of repurposed drugs for SARS-CoV-2 treatment [14, 15]. However, treating this infection with drugs, formerly designed for different targets, might result in adverse side effects and unwanted pharmacological effect [16]. Therefore, in the present study our team has made an attempt to screen protease inhibitors by specifically targeting the main protease of SARS-CoV-2. Most importantly, we have employed drug repurposing approaches by integrating pharmacophore and e-pharmacophore based screening to retrieve the novel and potent compounds against SARS-CoV-2. These virtual screening strategies have shown great promise in identifying bioactive molecules from large libraries [17 - 19]. In addition to these approaches, Prime MMGBSA (Molecular Mechanics/Generalized Born Surface Area) analysis and AutoQSAR techniques were performed to contemplate more efficacious drugs. We believed that hits resulted from our integrated approach provides clue for the control of the emerging SARS-CoV-2 pandemic.

Materials and Methods

Preparation of Dataset

The protein structure used in our study was obtained from Protein Data Bank (PDB ID: 6LU7) and was prepared using Schrödinger's protein preparation wizard [20]. Hydrogen bond optimizations, water removal, protein structure correction and finally protein energy minimization using OPLS_2005 force field were carried out during the preparation. Subsequently, the position of N3 (co-crystallized native ligand; used as the reference compound) was defined for grid generation. Further, a set of 9 molecules consisting of

substructure of the co-crystallized ligand of SARS-CoV-2 protease, a co-crystallized ligand, three substructures of equivalent SARS-CoV-1 protease inhibitors and four known SARS-CoV-1 inhibitors with low binding affinity were extracted from the literature [21]. These molecules were then cleaned using default specifications of LigPrep module and utilized for hypothesis generation [20]. Additionally, a phase database was generated from a total of 9591 molecules retrieved from DrugBank and was utilized for virtual screening (VS) application [22].

Generation of structure and ligand based pharmacophore models

The potential ligand based pharmacophore model was generated with the help of 9 main protease inhibitors retrieved from the literature [21]. The 9 molecules were initially divided into actives (5 molecules) and inactives (4 molecules). The protease inhibitors corresponds to SARS-CoV-2 were considered as actives. On the contrary, inhibitors of SARS-CoV-1 were considered as inactives. Note that present analysis utilizes the high-confident 9 molecules for model development as it provides model with high precision. Subsequently, using PHASE module of Schrödinger suite a common pharmacophore hypothesis (CPH) was generated after a stringent scoring and ranking process [23]. Likewise a structure based e-pharmacophore model was generated from the XP docked complex structure information associated with 9 existing inhibitors. By selecting only favorable sites that contributed more to the Glide XP energy terms, a CPH was constructed. Both the generated CPHs were used as a 3D query for screening the phase database. Finally, hierarchical GLIDE docking consisting of HTVS, SP and XP were performed against pharmacophore based screened set of molecules. This process is of immense importance to distinguish actives from inactives in a virtual screening application.

Post Screening Analysis

The XP screened out molecules underwent Prime MM/GBSA analysis where their binding energies were estimated in order to examine fine levels of compounds activity against main protease [24]. Despite the number of energy properties generated by Prime algorithm, the present analysis uses the parameter called free energy of binding to gain insight into the compounds activity. Nonetheless, the ligand strain energy, Coulomb energy and Van der Waals energy were also assessed in filtering the final hit compounds.

Machine Learning Principles using AutoQSAR

AutoQSAR is a machine-learning algorithm provided by Schrödinger suite that builds and applies QSAR models through automation [25]. In order to build a predictive model, AutoQSAR takes the 1D, 2D and 3D structural data of a molecule along with a property (eg: IC_{50}) to be modeled, as an input . It will then compute the fingerprints and descriptors using machine-learning statistical methods for creating a predictive QSAR model. The predictive accuracy of the model is evaluated using various parameters such as ranking score, Root Mean Square Error (RMSE), Standard Deviation (SD), Q^2 and R^2 values [26]. It is worth mentioning that the present analysis utilizes a total of 100 3C-like proteinase inhibitors for predictive model development. The details of the molecules along with their pIC_{50} values are presented in Table S1.

Drug likeness and Toxicity Descriptors:

Finally, the set of small molecules resulted from all the screening analysis will be tested for its druglikenss and toxicity properties. Of note, the pharmaceutically relevant key descriptors such as Stars, CNS and HOA were analyzed. The QikProp algorithm was employed for this purpose [27]. The prediction results from this tool is mainly relies on the descriptors value corresponds to the 95% of the known drugs available in the market. For instance, Star descriptor provides the number of descriptor outliers, CNS provides the predicted activity of

the molecule in central nervous system and HOA is the predictor of qualitative human oral absorption. It is certain that understanding these descriptors is utmost importance to overcome the clinical failure of the resultant compounds which in turn reduce the time and resources associated with the overall drug development process.

Results and Discussion

Pharmacophore Model Generation

The pharmacophore model development was achieved by means of existing main protease inhibitors reported in the literature [21]. Initially, the structure cleaning (protonation and chirality assessment) was accomplished using LigPrep module. With the set of cleaned-up ligands, conformational search will be initiated to generate a set of conformers for each ligand by ConfGen option. The generated conformers are grouped into actives and inactives and then pharmacophoric sites were created. The pharmacophores from all conformations of the ligands in the active set are examined. Subsequently, Common pharmacophores hypothesis were identified using a tree-based partitioning technique. The hypothesis with best survival score was utilized to screen the phase database. Here, the hypothesis named AARRH, consisting of two hydrogen bond acceptors, two aromatic rings and one hydrophobic group, was chosen for further analysis. The generated hypothesis is represented in Fig. S1.

E-pharmacophore Model Generation

On the contrary, the energy-based model is generated by considering the Glide XP docked structures of protein complexes as an input. The ligand docked poses of all the fragments and their contributions in ligand binding will be taken into account in this approach of model generation. The hypothesis was built by mapping energetic terms of Glide XP on pharmacophoric features. The structural and energy information present in between the

protein and ligand molecule were used to compute these energetic terms [28]. Consequently, a four featured model was generated which consists of two hydrogen bond acceptors and two aromatic rings (AARR) (Fig. S2). Additionally, during the model generation process, it was made sure that the features should exhibit an energy score greater than – 0.8 kcal/mol which was fixed as a threshold. This is essential in prioritizing the important sites accountable for effective ligand binding.

Virtual Screening using Docking Algorithm and Prime MMGBSA Analysis

In the initial stage of screening, the DrugBank database was screened independently using AARRH and AARR hypotheses to retrieve hit molecules with similar pharmacophore features. A total of 1000 hit molecules were retrieved from each hypothesis. These hits were then taken for Glide docking studies. In this process the hits were ranked using a three step hierarchical process, viz HTVS, SP and XP. The use of such hierarchical filters and associated parameters were highlighted in our earlier articles [18, 29, 30]. Initially, HTVS was carried out and 500 (50%) of the high scoring compounds were selected for the SP docking. Finally a total of 250 (50%) molecules from SP docking were subjected to XP docking. The compounds from multi-stage docking were filtered out based on the XP score threshold of native ligand (-5.444 kcal/mol). This filter resulted a total of 65 and 102 of compounds from pharmacophore- and e-pharmacophore-based hypothesis respectively. Subsequently, the results from both of our models were integrated. It is certain that integrating multiple hypotheses will be useful to eliminate the false positive prediction in the virtual screening application. This integration resulted a set of 155 molecules. The resultant compounds with its docking score are reported in Table 1. In the second stage of screening, binding free energy were analyzed for the integrated screened set of compounds. It is worth mentioning that entire 155 compounds possesses better binding free energy values than native ligand (-36.816 kcal/mol). This depict the strong correlation between the observed docking

score and binding free energy parameter examined from Prime algorithm. In essence, 149 compounds from this set likely to have better binding energy than Ritonavir, the currently used molecules in the treatment of Covid-19 infection.

AutoQSAR analysis and Interaction profiling of the screened hits

The screening process further continued with the aid of machine learning-based predictive model (pIC_{50} Calculation) generated by AutoQSAR module of Schrodinger. For instance, the machine learning model was developed with the help of 3C-like proteinase inhibitors retrieved from BindingDB. The algorithm generated 10 best models and the result is shown in Table 2. In particular, we have used highest score as the criteria to select the best model for the analysis. As stated earlier, the best model showed lowest SD and RMSE values than the other generated model and thus depicts its reliability in the generated result. The scatter plot depicting predicted pIC_{50} versus experimental pIC_{50} for best generated model was shown in Fig. 1.

Here, the entire data set of 155 compounds were tested for IC_{50} prediction using the best model. The prediction results for the complete set is presented in Table 1 together with docking score and binding energy. It is interesting to observe that only 4 compounds such as DB07800, DB08573, DB03744 and DB02986 showed better pIC_{50} values than native ligands and other existing drugs such as Ritonavir and Lopinavir. Thus, these 4 molecules considered as lead molecules against main protease of Covid-19 in our study. The binding pattern of these lead molecules together with reference compounds are illustrated in Fig. 2.

Drug likeness and Toxicity Descriptors:

Finally, these lead molecules were tested for its drug likeness and toxicity analysis using QikProp algorithm. The result is shown in Table 3. It is interesting to note that all the compounds having satisfactory star values of less than 5. Thus highlights that majority of the pharmaceutically relevant descriptors are found to be in the acceptable range for the screened

hit compounds. Additionally, the lead molecules demonstrated significant human oral absorption characteristics than all the reference ligand considered in our study. For instance, Ritonavir, native ligand and Lopinavir showed highest star value of 10, 9 and 3 respectively in our Qikprop analysis. This data from our study is correlates well with reported literature evidences. For instance, evidences states that patients treated with these drug were more likely to experience side effects like nausea, vomiting and diarrhea. Although positive results are achieved in the initial trials, resistance pattern is reported in the recent times in addition to its highly toxic characteristics. Moreover, these drugs certainly cause lot of adverse effects and chronic medical problems with long term usage [31]. On the contrary, the hit compounds resulted from our analysis demonstrated excellent safety profile as observed from the star descriptor than the existing antivirals.

Discussions

A total of 4 hit molecules were proposed from our study against main protease of Covid-19. The hit compounds exhibit better docking score, binding free energy and pIC_{50} values than existing inhibitors. Studies highlights that ligand strain energy is one of the significant parameter to be analysed in the case of drug screening [32]. It is an energy cost associated with ligand binding. Hence, the molecules with less strain energy likely to exhibit better binding free energy. It is important to note that all the hit compounds resulted from our analysis yielded less strain energy than existing anti-virals studied in our analysis. Thus highlights the tighter binding of hit molecules against the target protein. Moreover, Table 3 also highlights that Coulombic and van der Waals interactions provided the most substantial force for the binding of the inhibitor resulted from our study.

The binding of all the 4 hit molecules mimics binding pattern of existing inhibitors. It is interesting to note that GLN189 likely to play pivotal role in the binding of lead compounds.

The detailed listing of binding forces are given in Table S2. Literature evidence highlights that the existence of Π−Π stacking might increases the stability and loading capacity drugs [33]. Notably, most of the hits resulted from our study able to maintain Π−Π stacking in the binding pocket as like Lopinavir. Overall, it is certain that the key residues proposed in our study will facilitate the design of novel compounds in the future.

It is interesting to note that the hit compounds obtained in our study composed of crucial scaffolds, namely; benzamide indole and azetidine scaffolds. The two dimensional structure of hit compounds shown in Fig. 3. For instance, benzamide scaffold is the backbone of the hit compound, DB07800. On the other hand, indole and azetidine are the functional moieties in the case of hit compound, DB03744. The literature evidence highlight that benzamide derivative such as AH0109 exhibits potent anti-HIV-1 activity and has capability of disrupting the replication of HIV-1 strains that are resistant to the routinely used anti-HIV-1 drugs [34].

Importantly, the azetidine containing rimantadine analogues were found to exhibit 10 to 20 fold more potent activity against influenza A H2N2 virus than amantadine [35]. Of note, Indole group exemplifies one of the utmost privileged scaffolds in drug discovery. Currently, an amazing number of indole-containing compounds are in different clinical phases. In particular, Indole scaffold is widely used in antiviral research. One of the famous of marketed indole-containing antiviral drug is Arbidol, which is also in practice the management of Covid-19 infection [36 - 38].

Studies of DB02986 showed that this compound is found to act on carbonic anhydrase. It is thought that carbonic anhydrases have an important role in the initiation of the viral replication. It is likely that inhibition of the carbonic anhydrase increases the concentration of hydrogen ions intracellularly and decreases the pH. This decrease in pH in turn restrict the binding of virus in the host cell and even viral replication. Moreover, inhibitors of carbonic

anhydrase are also reported to have an activity against HIV infection [39, 40]. Taken together, these findings provide evidences for the newly identified compounds that can potentially be developed as drugs for the management of Covid-19 infection.

Conclusion

Here, we made a concerted effort to develop coronavirus therapeutic agents using integrated machine learning-based drug repurposing strategy. It is important to emphasize that compounds such as DB07800 (N-(2-(((5-Chloro-2-Pyridinyl)Amino)Sulfonyl)Phenyl)-4-(2-oxo-1(2H)-Pyridinyl)Benzamide), DB08573 (3-[(4-Chloroanilino)Sulfonyl]Thiophene-2-Carboxylic acid), DB03744 (Cp403700, (S)-1-{2-[(5-Chloro-1h-Indole-2-Carbonyl)-Amino]-3-Phenyl-Propionyl}-Azetidine-3-Carboxylate) and DB02986 (N-(2-Thienylmethyl)-2,5-Thiophenedisulfonamide) found to possesses better glide Score, binding free energy and predicted IC_{50} values alongside satisfactory drug likeness than co-crystallized native compounds and other existing inhibitors. The ubiquitous experimental data supports that the scaffolds identified in the study exhibit anti-viral activities and hence demonstrating the reliability of our results. Hopefully, we have proposed some useful candidates for SARS-CoV-2 main protease inhibitors. However, further experimental study on these compounds will be necessary to confirm the conclusions.

Acknowledgment

The authors thank VIT for providing ‘VIT SEED GRANT’ for carrying out this research work.

Conflict of Interests

The authors declare that they have no conflict of interests.

References

1. World Health Organization (WHO). Novel Coronavirus - China. Geneva: WHO; 2020. https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/
2. Gorbalenya AE., 2020. Severe acute respiratory syndrome-related coronavirus–The species and its viruses, a statement of the Coronavirus Study Group. bioRxiv 2020.02.07.937862.
3. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/
4. Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L. and Chen, H.D., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579(7798), pp.270-273.
5. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N. and Bi, Y., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet, 395(10224), pp.565-574.
6. Letko, M., Marzi, A. and Munster, V., 2020. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nature microbiology, pp.1-8.
7. Hoffmann, M., Kleine-Weber, H., Krüger, N., Mueller, MA., Drosten, C. and Pöhlmann, S., 2020. The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. bioRxiv 2020.01.31.929042
8. Ahmed, S.F., Quadeer, A.A. and McKay, M.R., 2020. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses, 12(3), p.254.

9. Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q. and Zhang, X.C., 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. Journal of virology, 82(5), pp.2515-2527.
10. Liu, X., Zhang, B., Jin, Z., Yang, H. and Rao, Z., 2020. The crystal structure of COVID-19 main protease in complex with an inhibitor N3. http://www.rcsb.org/structure/6LU7.
11. Xu, Z., Peng, C., Shi, Y., Zhu, Z., Mu, K., Wang, X. and Zhu, W., 2020. Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. bioRxiv 2020.01.27.921627
12. Li, Y., Zhang, J., Wang, N., Li, H., Shi, Y., Guo, G., Liu, K., Zeng, H. and Zou, Q., 2020. Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening. bioRxiv 2020.01.28.922922
13. Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. and Jung, S.H., 2016. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. Journal of medicinal chemistry, 59(14), pp.6595-6628.
14. Chen, YW., Yiu, CP. and Wong, KY., 2020. Prediction of the 2019-nCoV 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. https://doi.org/10.26434/chemrxiv.11831103.v2.
15. Liu, X. and Wang, XJ., 2020. Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines. bioRxiv 2020.01.29.924100.

16. Vedani, A., Dobler, M., Hu, Z. and Smieško, M., 2015. OpenVirtualToxLab—a platform for generating and exchanging in silico toxicity data. Toxicology letters, 232(2), pp.519-532.
17. Moonsamy, S., Bhakat, S., Ramesh, M. and Soliman, M.E., 2017. Identification of binding mode and prospective structural features of novel Nef protein inhibitors as potential anti-HIV drugs. Cell biochemistry and biophysics, 75(1), pp.49-64.
18. Rohini, K., Ramanathan, K. and Shanthi, V., 2019. Multi-Dimensional Screening Strategy for Drug Repurposing with Statistical Framework—A New Road to Influenza Drug discovery. Cell Biochemistry and Biophysics, 77(4), pp.319-333.
19. Wang, N., Ren, J.X. and Xie, Y., 2019. Identification of novel DHFR inhibitors for treatment of tuberculosis by combining virtual screening with in vitro activity assay. Journal of Biomolecular Structure and Dynamics, 37(4), pp.1054-1061.
20. Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R. and Sherman, W., 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. Journal of computer-aided molecular design, 27(3), pp.221-234.
21. Fischer, A., Sellner, M., Neranjan, S., Lill, MA. and Smieško, M., 2020. Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 687 Million Compounds. https://doi.org/10.26434/chemrxiv.11923239.v1
22. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z. and Assempour, N., 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic acids research, 46(D1), pp.D1074-D1082.
23. Dixon, S.L., Smondyrev, A.M., Knoll, E.H., Rao, S.N., Shaw, D.E. and Friesner, R.A., 2006. PHASE: a new engine for pharmacophore perception, 3D QSAR model

development, and 3D database screening: 1. Methodology and preliminary results. Journal of computer-aided molecular design, 20(10-11), pp.647-671.

24. Huang, N., Kalyanaraman, C., Bernacki, K. and Jacobson, M.P., 2006. Molecular mechanics methods for predicting protein–ligand binding. Physical Chemistry Chemical Physics, 8(44), pp.5166-5177.
25. Dixon, S.L., Duan, J., Smith, E., Von Bargen, C.D., Sherman, W. and Repasky, M.P., 2016. AutoQSAR: an automated machine learning tool for best-practice quantitative structure–activity relationship modeling. Future medicinal chemistry, 8(15), pp.1825-1839.
26. de Oliveira, M.T. and Katekawa, E., 2017. On the Virtues of Automated QSAR The New Kid on the Block. arXiv preprint arXiv:1711.02639.
27. Duffy, E.M. and Jorgensen, W.L., 2000. Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. Journal of the American Chemical Society, 122(12), pp.2878-2888.
28. Therese, P.J., Manvar, D., Kondepudi, S., Battu, M.B., Sriram, D., Basu, A., Yogeeswari, P. and Kaushik-Basu, N., 2014. Multiple e-pharmacophore modeling, 3D-QSAR, and high-throughput virtual screening of hepatitis C virus NS5B polymerase inhibitors. Journal of chemical information and modeling, 54(2), pp.539-552.
29. Rohini, K. and Shanthi, V., 2018. Hyphenated 3D-QSAR statistical model-drug repurposing analysis for the identification of potent neuraminidase inhibitor. Cell biochemistry and biophysics, 76(3), pp.357-376.
30. James, N. and Ramanathan, K., 2018. Discovery of potent ALK inhibitors using pharmacophore-informatics strategy. Cell biochemistry and biophysics, 76(1-2), pp.111-124.

31. Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., Ruan, L., Song, B., Cai, Y., Wei, M. and Li, X., 2020. A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. New England Journal of Medicine.
32. Mobley, D.L. and Dill, K.A., 2009. Binding of small-molecule ligands to proteins:“what you see” is not always “what you get”. Structure, 17(4), pp.489-498.
33. Shi, Y., van Steenbergen, M.J., Teunissen, E.A., Novo, L., Gradmann, S., Baldus, M., van Nostrum, C.F. and Hennink, W.E., 2013. Π–Π stacking increases the stability and loading capacity of thermosensitive polymeric micelles for chemotherapeutic drugs. Biomacromolecules, 14(6), pp.1826-1837.
34. Chen, L., Ao, Z., Jayappa, K.D., Kobinger, G., Liu, S., Wu, G., Wainberg, M.A. and Yao, X., 2013. Characterization of antiviral activity of benzamide derivative AH0109 against HIV-1 infection. Antimicrobial agents and chemotherapy, 57(8), pp.3547-3554.
35. Zoidis, G., Fytas, C., Papanastasiou, I., Foscolos, G.B., Fytas, G., Padalko, E., De Clercq, E., Naesens, L., Neyts, J. and Kolocouris, N., 2006. Heterocyclic rimantadine analogues with antiviral activity. Bioorganic & medicinal chemistry, 14(10), pp.3341-3348.
36. Leneva, I.A., Russell, R.J., Boriskin, Y.S. and Hay, A.J., 2009. Characteristics of arbidol-resistant mutants of influenza virus: implications for the mechanism of anti-influenza action of arbidol. Antiviral research, 81(2), pp.132-140.
37. Boriskin, Y.S., Leneva, I.A., Pecheur, E.I. and Polyak, S.J., 2008. Arbidol: a broad-spectrum antiviral compound that blocks viral fusion. Current medicinal chemistry, 15(10), pp.997-1005.

38. Zhang, M.Z., Chen, Q. and Yang, G.F., 2015. A review on recent developments of indole-containing antiviral agents. European journal of medicinal chemistry, 89, pp.421-441.
39. Chen, I., Tsai, A.Y., Huang, Y.P., Wu, I., Cheng, S.F., Hsu, Y.H. and Tsai, C.H., 2017. Nuclear-encoded plastidal carbonic anhydrase is involved in replication of Bamboo mosaic virus RNA in Nicotiana benthamiana. Frontiers in Microbiology, 8, p.2046.
40. Temperini, C., Innocenti, A., Guerri, A., Scozzafava, A., Rusconi, S. and Supuran, C.T., 2007. Phosph (on) ate as a zinc-binding group in metalloenzyme inhibitors: X-ray crystal structure of the antiviral drug foscarnet complexed to human carbonic anhydrase I. Bioorganic & medicinal chemistry letters, 17(8), pp.2210-2215.

Table Legends

TABLE 1 Docking score, Binding energy and pIC_{50} values of Screened hits molecules

TABLE 2 Statistical parameters corresponds to ten best model generated by AutoQSAR.

TABLE 3 The collective key parameters corresponds to the reference and hit molecules from our analysis.

Figure Legends

FIGURE 1 Scatter plot analysis of best model predicted from AutoQSAR.

FIGURE 2 Ligand interaction diagram of references and hit molecules. References - a) Ritonavir ; b) Lopinavir ; c) N3 – Inhibitor (Native Ligand). Hit molecules - d) DB07800 ; e) DB08573 ; f) DB03744 ; g) DB02986.

FIGURE 3 2D structure of reference and hit molecules. References - a) Ritonavir ; b) Lopinavir; c) N3 – Inhibitor (Native Ligand) . Hit molecules - d) DB07800 ; e)DB08573 ; f) DB03744 ; g) DB02986. Important functional groups are highlighted as red circle.

TABLE 1 Docking score, Binding energy and pIC_{50} values of Screened hits molecules

S. No	**DrugBank - ID**	**Docking score (kcal/mol)**	**ΔG Bind (kcal/mol)**	**Predicted pIC_{50}**
1	DB02375	-9.107	-49.332	5.47
2	DB08230	-8.464	-43.044	5.298
3	DB12784	-8.439	-46.629	3.27
4	DB09298	-8.403	-53.381	4.664
5	DB04216	-8.174	-43.275	5.282
6	DB07903	-8.119	-64.959	4.787
7	DB01858	-8.112	-59.842	4.125
8	DB13781	-8.057	-58.554	4.518
9	DB02224	-7.941	-42.874	5.148
10	DB07439	-7.815	-45.165	4.574
11	DB08775	-7.723	-50.41	4.582
12	DB07039	-7.67	-44.844	4.817
13	DB02537	-7.583	-47.838	4.511
14	DB11841	-7.529	-46.507	4.778
15	DB07795	-7.457	-44.796	5.169
16	DB08138	-7.301	-54.598	3.837
17	DB08066	-7.269	-40.134	5.147
18	DB07419	-7.183	-54.207	4.029
19	DB01852	-7.174	-43.482	5.012
20	DB11698	-7.087	-40.811	4.283
21	DB08573	-7.052	-40.254	5.882
22	DB07991	-7.027	-58.267	5.116

23	DB08229	-7.009	-38.27	5.3
24	DB07800	-7.005	-54.044	5.548
25	DB08424	-6.927	-43.028	3.965
26	DB14086	-6.919	-40.063	5.001
27	DB12507	-6.855	-42.375	4.954
28	DB12672	-6.822	-55.413	4.695
29	DB07587	-6.806	-46.706	4.243
30	DB12301	-6.768	-51.556	4.847
31	DB07194	-6.768	-54.274	4.361
32	DB08067	-6.73	-49.22	4.985
33	DB11967	-6.718	-57.339	4.239
34	DB05772	-6.688	-51.108	4.65
35	DB12140	-6.687	-54.71	3.584
36	DB07750	-6.68	-46.262	4.179
37	DB04903	-6.638	-59.378	5
38	DB09082	-6.636	-60.372	4.766
39	DB07589	-6.627	-51.155	4.5
40	DB07685	-6.62	-56.332	5.25
41	DB07676	-6.615	-49.46	4.545
42	DB00231	-6.61	-49.926	5.291
43	DB03916	-6.609	-51.124	4.325
44	DB08517	-6.577	-44.173	4.511
45	DB02241	-6.569	-56.316	4.678
46	DB06397	-6.549	-52.239	4.782
47	DB04200	-6.543	-40.777	4.546

48	DB07054	-6.541	-44.817	3.516
49	DB07315	-6.513	-44.896	4.847
50	DB08454	-6.498	-56.19	4.344
51	DB04030	-6.497	-59.602	5.253
52	DB11824	-6.459	-50.469	4.437
53	DB00972	-6.438	-50.359	4.793
54	DB12941	-6.427	-60.855	5.029
55	DB04811	-6.417	-45.287	4.882
56	DB07352	-6.398	-40.273	4.866
57	DB12078	-6.383	-54.886	4.362
58	DB08345	-6.35	-65.857	4.65
59	DB08968	-6.344	-60.232	5.208
60	DB07809	-6.328	-57.424	4.54
61	DB11696	-6.318	-50.386	4.987
62	DB07288	-6.274	-50.755	5.166
63	DB07941	-6.272	-55.102	4.09
64	DB08282	-6.257	-57.374	5.229
65	DB07794	-6.254	-44.029	4.54
66	DB05956	-6.248	-48.81	3.275
67	DB01094	-6.236	-38.898	4.813
68	DB14034	-6.226	-54.481	5.154
69	DB07421	-6.221	-49.145	4.507
70	DB04669	-6.197	-57.198	4.37
71	DB02558	-6.187	-46.331	5.243
72	DB07157	-6.167	-43.229	4.746

73	DB02194	-6.166	-55.182	4.88
74	DB12207	-6.15	-45.258	4.043
75	DB07500	-6.146	-39.329	4.559
76	DB00522	-6.146	-51.823	5.246
77	DB07649	-6.137	-46.083	4.059
78	DB08488	-6.136	-49.881	5.164
79	DB01128	-6.128	-39.381	3.967
80	DB04861	-6.125	-49.294	3.469
81	DB03467	-6.109	-37.932	4.926
82	DB11885	-6.1	-49.185	4.017
83	DB07042	-6.1	-50.681	4.62
84	DB12448	-6.075	-39.864	4.973
85	DB07065	-6.063	-62.437	4.797
86	DB12233	-6.063	-47.742	5.159
87	DB06983	-6.057	-49.81	4.292
88	DB06837	-6.038	-40.556	5.297
89	DB06914	-6.022	-52.964	4.511
90	DB08531	-6.012	-56.319	4.666
91	DB08233	-6.012	-51.402	4.636
92	DB12183	-6.008	-51.037	4.144
93	DB08450	-6.002	-41.838	4.249
94	DB03509	-5.991	-47.54	5.144
95	DB07469	-5.96	-49.56	4.872
96	DB07468	-5.957	-42.67	4.924
97	DB07785	-5.948	-42.986	4.573

98	DB07779	-5.948	-43.432	4.579
99	DB07766	-5.934	-52.316	4.591
100	DB02046	-5.92	-55.748	5.139
101	DB07257	-5.917	-43.817	5.15
102	DB08137	-5.913	-53.883	4.806
103	DB06919	-5.91	-53.654	5.363
104	DB08538	-5.908	-43.935	3.506
105	DB08159	-5.904	-41.77	5.266
106	DB12963	-5.893	-48.754	4.969
107	DB07789	-5.887	-51.108	4.526
108	DB12535	-5.87	-53.539	3.924
109	DB08042	-5.866	-47.802	3.858
110	DB08701	-5.863	-50.988	5.321
111	DB07470	-5.854	-44.823	4.962
112	DB01136	-5.834	-42.977	4.212
113	DB01136	-5.834	-42.977	4.212
114	DB12100	-5.828	-50.762	4.736
115	DB07247	-5.818	-51.825	5.112
116	DB13874	-5.817	-51.635	4.286
117	DB11838	-5.814	-53.126	4.877
118	DB07423	-5.802	-53.099	4.705
119	DB09295	-5.8	-43.682	4.282
120	DB12848	-5.792	-51.389	4.04
121	DB12047	-5.792	-42.523	4.586
122	DB08536	-5.792	-46.002	4.37

123	DB08340	-5.791	-41.883	4.567
124	DB11877	-5.783	-56.326	3.791
125	DB07845	-5.765	-43.48	4.083
126	DB12200	-5.758	-49.984	5.203
127	DB02611	-5.758	-47.782	4.996
128	DB02986	-5.75	-45.017	6.039
129	DB12021	-5.736	-44.565	4.707
130	DB09330	-5.729	-61.252	4.152
131	DB08883	-5.71	-58.954	4.74
132	DB03744	-5.71	-38.836	5.656
133	DB07595	-5.688	-57.132	4.346
134	DB08459	-5.681	-55.417	4.855
135	DB04518	-5.642	-44.649	4.173
136	DB08064	-5.638	-54.695	4.957
137	DB06237	-5.629	-42.562	4.375
138	DB08392	-5.607	-46.21	3.966
139	DB07324	-5.598	-44.017	4.54
140	DB08143	-5.595	-51.824	4.629
141	DB02830	-5.586	-49.851	4.911
142	DB07379	-5.583	-51.098	4.6
143	DB07267	-5.566	-51.25	4.117
144	DB07210	-5.515	-58.707	4.441
145	DB07755	-5.499	-44.172	4.179
146	DB13016	-5.496	-45.466	4.931
147	DB00904	-5.494	-54.03	4.692

148	DB08738	-5.483	-37.956	5.324
149	DB03601	-5.48	-39.733	4.953
150	DB06927	-5.48	-42.356	4.687
151	DB06589	-5.475	-55.857	5.039
152	DB12465	-5.468	-48.798	4.321
153	DB07878	-5.467	-45.682	4.786
154	DB09198	-5.459	-48.496	4.369
155	DB11952	-5.448	-48.708	5.111

TABLE 2 Statistical parameters corresponds to ten best model generated by AutoQSAR

Model code	**Score**	**SD**	**R^2**	**RMSE**	**Q^2**
kpls_desc_44	0.7022	0.4104	0.6949	0.3662	0.5604
pls_44	0.6457	0.4598	0.6170	0.3674	0.6481
kpls_radial_25	0.6406	0.4361	0.6391	0.4143	0.4996
pls_45	0.5958	0.4677	0.5926	0.4383	0.5320
Kpls_desc_25	0.5957	0.4711	0.5897	0.4359	0.4461
kpls_desc_200	0.5928	0.4230	0.6684	0.4266	0.5385
kpls_molprint2	0.5920	0.4589	0.5923	0.4430	0.4885
pls_20	0.5896	0.4578	0.6115	0.4399	0.5092
kpls_dpsc_20	0.5928	0.4230	0.6684	0.4266	0.5385
Kpls_sesc_45	0.5790	0.4869	0.5466	0.4188	0.5726

TABLE 3 The collective key parameters corresponds to the reference and hit molecules from our analysis.

Compounds	**Docking score (kcal/mol)**	**ΔdG Bind (kcal/mol)**	**ΔdG Bind Coulomb**	**ΔdG Bind Covalent**	**ΔdG Bind Hbond**	**ΔdG Bind Lipo**	**ΔdG Bind Packing**	**ΔdG Bind Solv GB**	**ΔdG Bind vdW**	**Lig Strain Energy**	**stars**	**CNS[a]**	**HoA[b]**	**pIC$_{50}$**
Ritonavir	-7.529	-39.34	-20.57	5.506	-2.68	-16.04	0	37.716	-43.271	27.905	10	-2	1	4.625
Lopinavir	-5.907	-39.186	-6.528	1.124	-2.123	-18.651	-3.508	25.659	-35.158	22.952	3	-2	1	5.535
N3 inhibitor	-5.444	-36.816	-21.566	5.249	-1.797	-11.098	-0.52	40.063	-47.148	18.938	9	-2	1	5.49
DB07800	-7.005	-54.044	-17.867	3.381	-2.042	-14.955	-3.982	25.544	-44.124	16.931	1	-2	3	5.548
Db03744	-5.71	-38.836	0.75	14.349	-1.461	-14.117	-2.798	1.666	-37.225	12.068	0	-2	2	5.565
Db08573	-7.052	-40.254	18.054	3.051	-1.891	-11.498	-3.794	-6.955	-37.221	3.14	1	-1	3	5.882
Db02986	-5.75	-45.017	-18.601	6.469	-2.485	-10.941	-2.579	15.522	-32.402	9.043	1	-2	2	6.039

[a] - Central Nervous System ; [b] - Human Oral Absorbtion

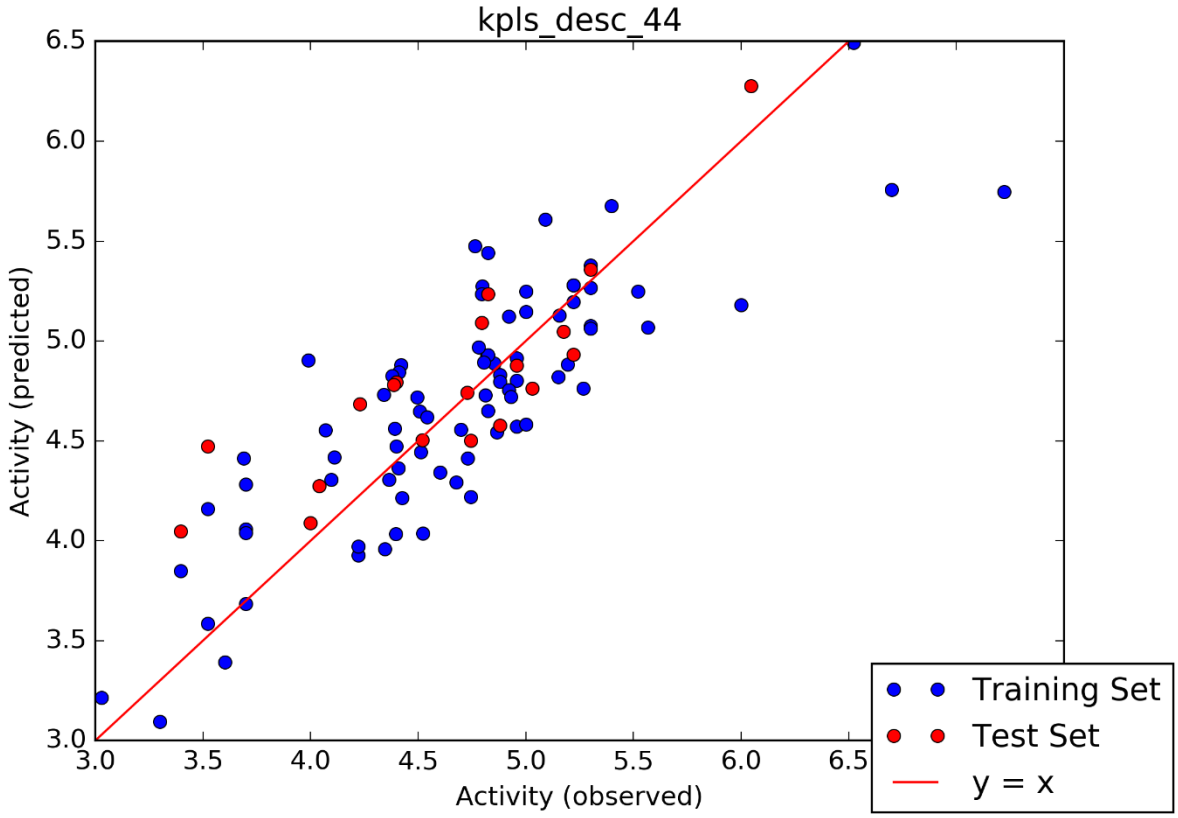


FIGURE 1 Scatter plot analysis of best model predicted from AutoQSAR.

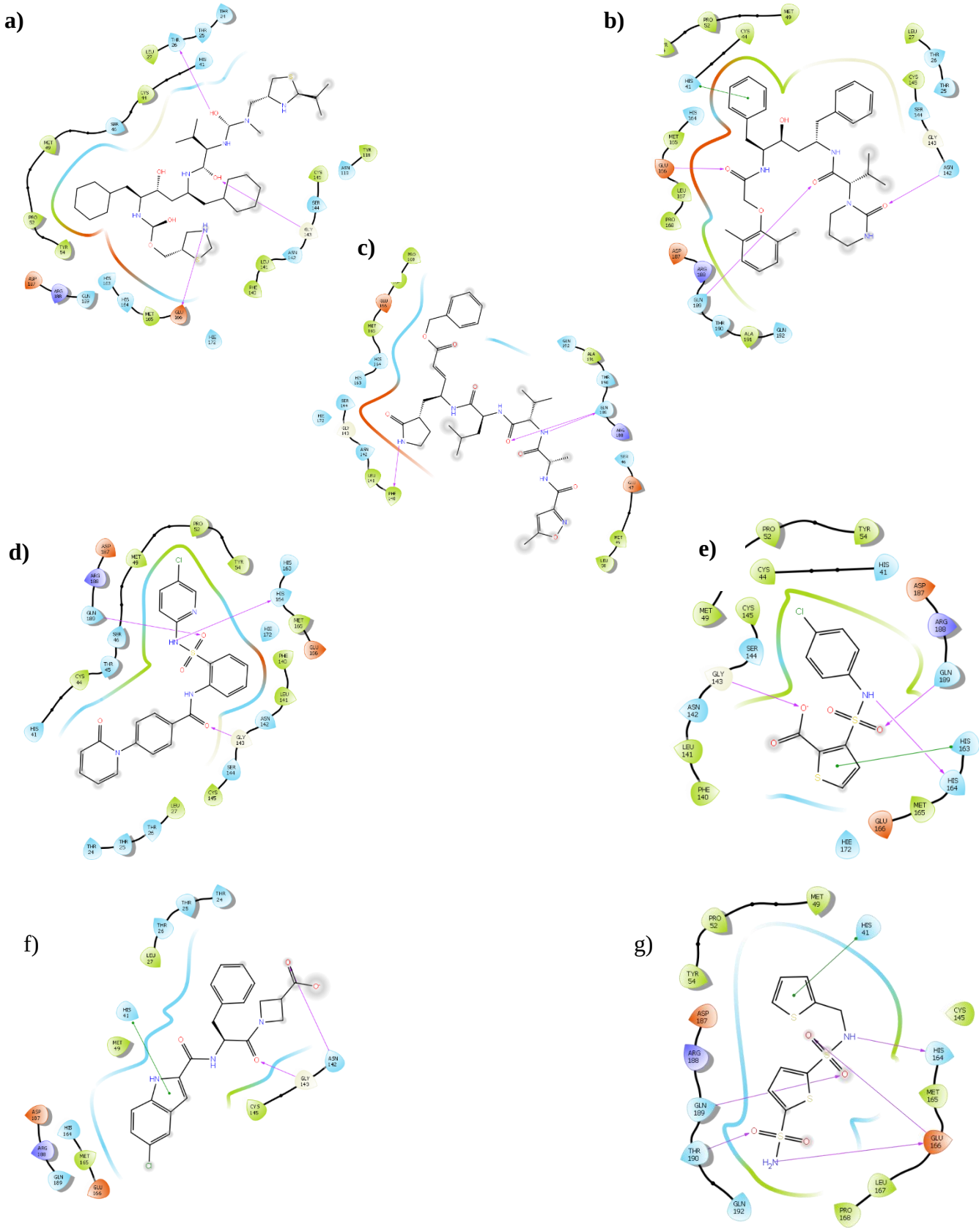
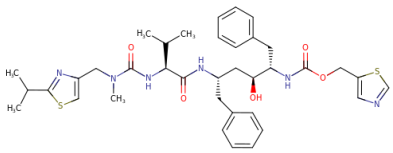
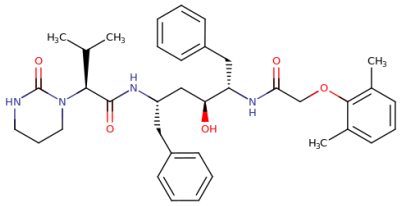


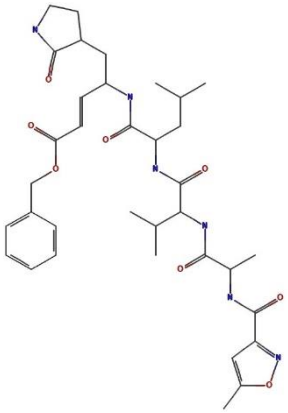
FIGURE 2 Ligand interaction diagram of references and hit molecules. References - a) Ritonavir ; b) Lopinavir; c) N3 - Inhibitor (Native Ligand) . Hit molecules - d) DB07800 ; e)DB08573 ; f) DB03744 ; g) DB02986



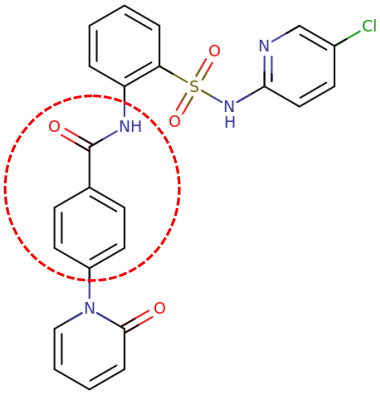
a)



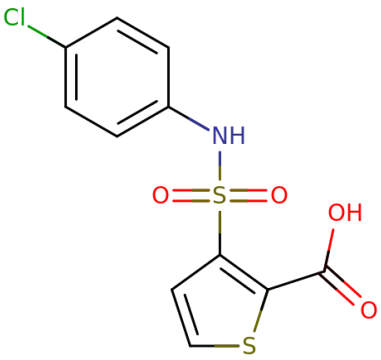
b)



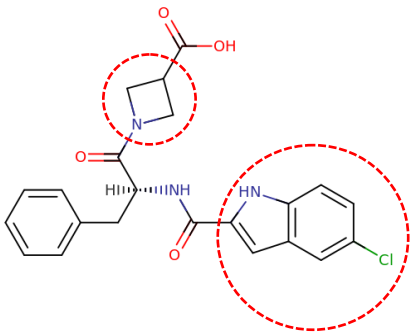
c)



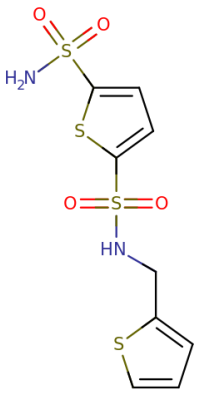
d)



e)



f)



g)

FIGURE 3 2D structure of reference and hit molecules. References - a) Ritonavir ; b) Lopinavir; c) N3 – Inhibitor (Native Ligand) . Hit molecules - d) DB07800 ; e)DB08573 ; f) DB03744 ; g) DB02986. Important functional groups are highlighted as red circle.