Received: 1 January 2021 | Accepted: 23 March 2021

DOI: 10.1002/bab.2159

ORIGINAL ARTICLE

WILEY

Discovery of potent Covid-19 main protease inhibitors using integrated drug-repurposing strategy

Muthu Kumar T | **Rohini K** | **Nivya James** | **Shanthi V** | **Ramanathan K**

Department of Biotechnology, School of Bio-Sciences and Technology, Vellore Institute of Technology, Vellore, India

Correspondence
K. Ramanathan, Department of Biotechnology, School of Bio-Sciences and Technology, Vellore Institute of Technology, Vellore 632 014, India.
Email: 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 significant role of the main protease of Covid-19 for virus replication, we performed a drug-repurposing study using the 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 the DrugBank repository. Further, a hierarchical docking protocol was implemented with the support of the Glide algorithm. The resultant compounds were then examined for their binding free energy against the main protease of Covid-19 by means of the Prime-MM/GBSA algorithm. Most importantly, the machine learning-based AutoQSAR algorithm was used to predict the antiviral activities of resultant compounds. The hit molecules were also examined for their drug-likeness and toxicity parameters through the QikProp algorithm. Finally, the hit compounds activity against the main protease was validated using molecular dynamics simulation studies. Overall, the present analysis yielded two potential inhibitors (DB02986 and DB08573) that are predicted to bind with the main protease of Covid-19 better than currently used drug molecules such as N3 (cocrystallized native ligand), lopinavir, and ritonavir.

KEYWORDS
autoQSAR, DrugBank, glide, molecular dynamics, Prime-MM/GBSA, SARS-CoV-2 main protease

Abbreviations: SARS-CoV-2, Severe Acute Respiratory Syndrome CoronaVirus 2; MM/GBSA, Molecular Mechanics Generalized Born and Surface Area; AutoQSAR, Automatic Quantitative Structure-Activity Relationship; MD, Molecular Dynamics; HIV, Human Immunodeficiency Virus; OPLS, Optimized Potentials for Liquid Simulations; CPH, Common Pharmacophore Hypothesis; HTVS, High-Throughput Virtual Screening; SP, Standard-Precision; XP, Extra-Precision; RMSE, Root-Mean Square Error; SD, Standard Deviation; CNS, Central Nervous System; HOA, Human Oral Absorption; SPC, Simple Point Charge; RMSD, Root-Mean Square Deviations; PCA, Principal Component Analysis; FEL, Gibbs Energy Landscape; IC_{50}, Half Maximal Inhibitory Concentration

1 | INTRODUCTION

The outbreak of coronavirus disease (COVID-19) has raised major health concerns 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 the family Coronoviridae and genus *Betacoronavirus*.[1] 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 (https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/). The virus is 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 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.[2,3] It was also found to exhibit a putatively similar cell entry mechanism and human cell receptor utilization to that of SARS-CoV.[4,5] Considering this apparent similarity, scientists have recently carried out preliminary research to identify potential vaccine targets for COVID-19 based on SARS-CoV immunological studies.[6] However, no specific therapeutics is accessible to treat the infection indicating that only clinical symptoms along with secondary infections could be treated with a repurposed antiviral drug. Hence there is a dire need to develop potent therapeutics and vaccines against SARS-CoV-2.

One of the attractive targets for anti-CoV drug design is coronavirus main protease. It plays a vital role in viral gene expression and replication through proteolytic processing of polyproteins.[7] The crystal structure of SARS-CoV-2 main protease was recently elucidated to enable the designing of specific protease inhibitors.[8] Even though the main proteases of SARS-CoV-2 and SARS-CoV are closely related with a sequence identity of 96.1%[9,10] the drugs developed for SARS-CoV could not be suggested for the treatment as they remained in the preclinical stage.[11] Several studies have been carried out to investigate the inhibitory activity of repurposed drugs for SARS-CoV-2 treatment.[12,13] However, treating this infection with drugs, formerly designed for different targets, might result in adverse side effects and unwanted pharmacological effects.[14] Therefore, in the present study our team has attempted to screen protease inhibitors by explicitly targeting the main protease of SARS-CoV-2. Importantly, we employed a drug-repurposing approach as it helps to identify the hidden feature of the existing drug molecule. Avarol, which is one such notable example, was originally known as an antibacterial compound, which was then identified as a potential drug molecule for Alzheimer's disease and HIV through drug-repurposing approaches.[15–17] These virtual screening strategies have shown great promise in identifying bioactive molecules from large libraries.[18–20] In addition to these approaches, Prime MM/GBSA (molecular mechanics/generalized born surface area) analysis, AutoQSAR techniques, and molecular dynamics (MD) simulation were performed to contemplate more efficacious drugs. We believed that hits resulted from our integrated approach provide a clue to the control of the emerging SARS-CoV-2 pandemic.

Highlights

Given the significant role of the main protease of Covid-19 for virus replication, we performed a drug-repurposing study for novel antiviral therapies using the main protease structure, 6LU7. After screening the DrugBank repository, a hierarchical docking protocol was implemented and candidates were examined for their binding free energy. Machine learning-based AutoQSAR algorithm was then used to predict the antiviral activities of resultant compounds. Finally, hit molecules were examined for their drug-likeness, toxicity, and activity against the main protease. Overall, analysis yielded two potential inhibitors (DB02986 and DB08573) that are predicted to bind with the main protease of Covid-19 better than currently used drug molecules.

2 | MATERIALS AND METHODS

2.1 | Preparation of dataset

The protein structure used in our study was obtained from the Protein Data Bank (PDB ID: 6LU7) and was prepared using Schrödinger's protein preparation wizard.[21] 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 (cocrystallized native ligand; used as the reference compound) was defined for grid generation. Further, a set of nine molecules consisting of the substructure of the cocrystallized ligand of SARS-CoV-2 protease, a cocrystallized 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.[22] These molecules were then cleaned using default specifications of the LigPrep module and utilized for hypothesis generation.[21] Additionally, a phase database was generated from a total of 9591 molecules retrieved from DrugBank and was utilized for virtual screening application.[23]

2.2 | Generation of structure and ligand-based pharmacophore models

The potential ligand–based pharmacophore model was generated with the help of nine main protease inhibitors retrieved from the literature.[22] The nine molecules were initially divided into actives (five molecules) and inactives (four molecules). The protease inhibitors that correspond to SARS-CoV-2 were considered as actives. On the contrary, inhibitors of SARS-CoV-1 were considered as inactives. Note that the present analysis utilizes the high-confident nine molecules for model development as it provides a model with high precision. Subsequently, using the PHASE module of the Schrödinger suite a common pharmacophore hypothesis (CPH) was generated after a stringent scoring and ranking process.[24] Likewise, a structure-based e-pharmacophore model was generated from the XP-docked complex structure information associated with nine other 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 three-dimensional (3D) query for screening the phase database. Finally, hierarchical GLIDE docking consisting of HTVS, SP, and XP was performed against pharmacophore-based screened molecules. This process is of immense importance to distinguish actives from inactives in a virtual screening application.

2.3 | Postscreening 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 the main protease.[25] Despite the number of energy properties generated by the Prime algorithm, the present analysis uses the parameter called free energy of binding to gain insight into the activity of the compounds. Nonetheless, the ligand strain energy, Coulomb energy, and Van der Waals energy were also assessed in filtering of the final hit compounds.

2.4 | Machine learning principles using AutoQSAR

AutoQSAR is a machine-learning algorithm provided by the Schrödinger suite that builds and applies QSAR models through automation.[26] In order to build a predictive model, AutoQSAR takes the one-, two-, and three-dimensional structural data of a molecule along with a property (e.g., 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.[27] 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 in the Supporting Information.

2.5 | Drug likeness and toxicity descriptors

Finally, the set of small molecules resulted from all the screening analyses will be tested for their drug-likeness and toxicity properties. Of note, the pharmaceutically relevant key descriptors such as Stars, central nervous system (CNS), and human oral absorption (HOA) were analyzed. The QikProp algorithm was employed for this purpose.[28] The prediction results from this tool mainly rely on the descriptor's value that corresponds to 95% of the known drugs available in the market. For instance, the Star descriptor provides the number of descriptor outliers, CNS provides the predicted activity of the molecule in the central nervous system, and HOA is the predictor of qualitative human oral absorption. Understanding of these descriptors is undoubtedly of 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.

2.6 | Molecular dynamics

MD simulations were conducted to forecast the protein's dynamic motions with the bounded ligand. Gromacs package version 5.1.2 implemented with GROMOS96 43a1 force field has been employed to generate receptor topologies.[29] On the other hand, the GlycoBioChem PRODRG server was used to generate ligand topologies.[30] The solvation box was designed as the shape of the rhombic dodecahedron type and solvated using the SPC216 (simple point charge) water model. Four chlorine counter ions were added to neutralize the charge of the system during simulation. The steepest descent algorithm was employed to perform the energy minimization of the main protease structure using nsteps and energy step size parameters of 50,000 and 0.02, respectively. Additionally, the system was calibrated with the constant temperature (300 K) and pressure (1 bar) via Berendsen thermostat coupling and default system

pressure coupling, respectively. Each of the equilibration steps was carried out for 100 ps. The dynamic simulation of the complex system was performed for 10 ns after all the preprocessing phases.[31] Root-mean square deviations (RMSD) were estimated using the MD trajectory to learn the variations in protein conformation during the various simulations as well as the total number of intermolecular hydrogen bonds measured for each protein–ligand complex to gain insights into the binding effectiveness.

Importantly principal component analysis (PCA) of the system was performed to evaluate the atomic motions present in the system using the covariance matrix through gromacs utilities such as g_covar, g_anaeig, and g_sham. Eigenvalues and Eigenvectors are generated to determine the directional movements of atoms during different time periods. Finally, Gibbs energy landscape (FEL) was calculated in order to determine the energy minima conformations of ligand during the different motions of the protein.[32]

3 | RESULTS

3.1 | Pharmacophore model generation

The pharmacophore model development was achieved by means of existing main protease inhibitors reported in the literature.[22] Initially, the structure cleaning (protonation and chirality assessment) was accomplished using the LigPrep module. With the set of cleaned-up ligands, a conformational search will be initiated to generate a set of conformers for each ligand by the 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 hypotheses were identified using a tree-based partitioning technique. The hypothesis with the best survival score was utilized to screen the phase database. Here, the hypothesis named AARRH, consisting of two hydrogen bond acceptors (A), two aromatic rings (R), and one hydrophobic group (H), was chosen for further analysis. The generated hypothesis is represented in Figure S1 in the Supporting Information.

3.2 | 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 possessing 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.[33] Consequently, a four-featured model was generated which consists of two hydrogen bond acceptors (A) and two aromatic rings (R) (AARR) (Figure S2 in the Supporting Information). Additionally, during the model generation process, it was made sure that the features should exhibit an energy score higher than –0.8 kcal/mol which was fixed as a threshold. This is essential in prioritizing the crucial sites accountable for effective ligand binding.

3.3 | Virtual screening using the 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, namely, HTVS, SP, and XP. The use of such hierarchical filters and associated parameters was highlighted in our earlier articles.[19,34] 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 multistage docking were filtered out based on the XP score threshold of native ligand (−5.444 kcal/mol). This filter resulted in a total of 65 and 102 compounds from the 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 in eliminating the false positive prediction in the virtual screening application. This integration results in a set of 155 molecules. The resultant compounds with their docking score are reported in Table S2 in the Supporting Information. In the second stage of screening, binding free energy was analyzed for the integrated screened set of compounds. It is worth mentioning that the entire 155 compounds possess better binding free energy values than native ligand (−36.816 kcal/mol). This depicts the strong correlation between the observed docking score and the binding free energy parameter examined from the 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.

TABLE 1 Statistical parameters correspond to ten best models 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

Abbreviations: Kpls, kernel-based partial least squares regression; pls, partial least squares regression; Q^2, predictive squared correlation coefficient; R^2, correlation coefficient; RMSE, root-mean-square error; SD, standard deviation.

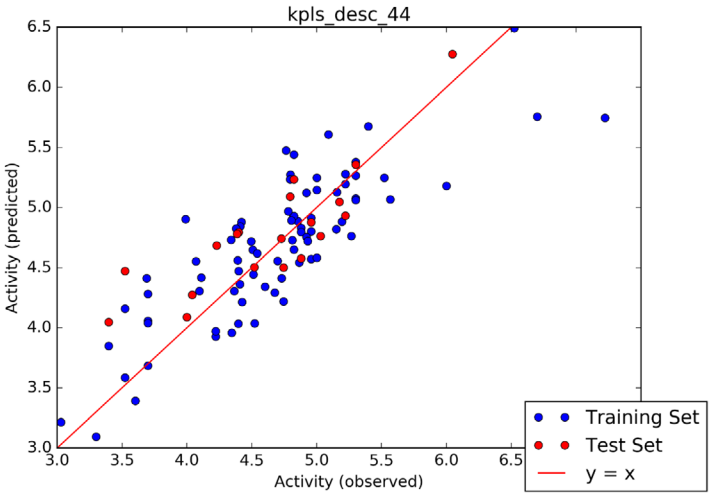

FIG 1 Scatter plot analysis of best model predicted from AutoQSAR

3.4 | AutoQSAR analysis and interaction profiling of the screened hits

The screening process further continued with the aid of a machine learning–based predictive model (pIC_{50} Calculation) generated by the 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 results are shown in Table 1. In particular, we have used the highest score as the criteria to select the best model for the analysis. As stated earlier, the best model showed the lowest SD and RMSE values than the other generated model and thus depicted its reliability in the generated result. The scatter plot depicting predicted pIC_{50} versus experimental pIC_{50} for the best generated model is shown in Figure 1. It is evident from the figure that only a few of the molecules were scattered away from the predicted pattern (trend line in Figure 1). These compounds are likely to be labeled and are unable to fit in a generated QSAR model. Perhaps, we say that these compounds possess different mode/mechanism of interaction pattern with the target protein.[35] However, the generated model displayed a minimum number of outliers and thus we believe that our model (kpls_desc_44) has the ability to predict the in vitro activity (pIC_{50}) of the compounds with high precision.

Here, the entire dataset of 155 compounds was tested for IC_{50} prediction using the best model. The prediction results for the complete set are presented in Table S2 together with the docking score and binding energy. It is interesting to observe that only four 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 four molecules are considered as lead molecules against the main protease of Covid-19 in our study. The binding pattern of these lead molecules together with reference compounds is illustrated in Figure 2.

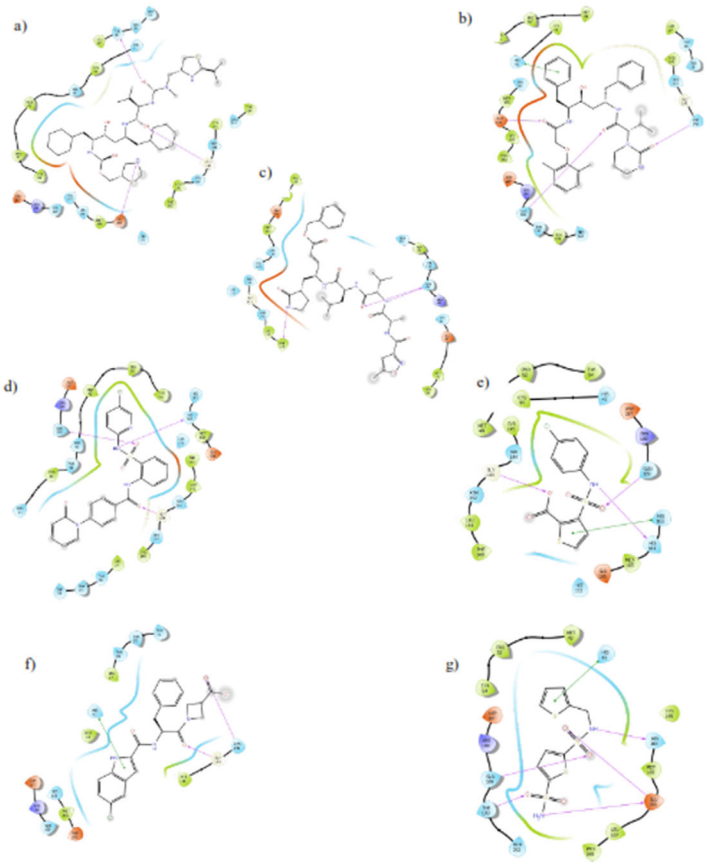

FIG 2 Ligand interaction diagram of references and hit molecules. References; (A) ritonavir; (B) lopinavir; (c) N3–inhibitor (native ligand). Hit molecules: (D) DB07800; (E) DB08573; (F) DB03744; (G) DB02986

3.5 | Drug likeness and toxicity descriptors

Finally, these lead molecules were tested for their drug-likeness and toxicity analysis using the QikProp algorithm. The result is shown in Table 2. It is interesting to note that all the compounds were having satisfactory star values of less than 5. Thus highlighting 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 HOA characteristics than all the reference ligand considered in our study. For instance, ritonavir, native ligand, and lopinavir showed the highest star value of 10, 9, and 3, respectively, in our Qikprop analysis. These data from our study correlate well with reported literature evidences. For instance, evidence states that patients treated with these drugs were more likely to experience side effects like nausea, vomiting, and diarrhea. Although favorable results are achieved in the initial trials, a resistance pattern is reported in recent times in addition to its highly toxic characteristics. Moreover, these drugs certainly cause a lot of adverse effects and chronic medical problems with long-term usage.[36] On the contrary, the hit compounds resulted from our analysis demonstrated an excellent safety profile as observed from the star descriptor than the existing antivirals.

3.6 | Simulation and synergism studies

MD simulation of main protease reference docked structure together with hit complexes such as DB02986, DB03744, DB07800, and DB08573 were performed by gromacs package for 10 ns each to gain insights into the complex structure dynamics perturbations. For instance, ritonavir was used as a reference compound as it has shown better docking score, binding free energy, and pIC_{50} values than the other known compounds considered in our analysis. Initially, the trajectory was analyzed using the RMSD corresponding to proteins Cα atoms to explore the stability of the complex structures. The results are shown in Figure 3. The RMSD values of all the investigated complex structures lie in the range of 1–4 Å. For instance, the ritonavir–protein Cα atoms showed RMSD between ~1.5 and ~2.2 Å over a 10 ns trajectory frame. It is interesting to note that hit compounds such as DB02986 and DB08573 showed a similar trend, stable and constant RMSD value between ~1.5 and 2.2 Å over the simulation period. On the contrary, the complex structures such as DB07800 and DB03744 showed higher deviation in RMSD values and attains ~3.4 Å at the end of the simulation period.

The stability of the complex structures was further explored by means of hydrogen bond analysis over the 10 ns simulation. The results are shown in Figure 4. Analysis

TABLE 2 The collective key parameters corresponds to the reference and hit molecules from our analysis

Compounds	Docking score (kcal/mol)	ΔG_{Bind} (kcal/mol)	ΔG_{Bind} Coulomb	ΔG_{Bind} covalent	ΔG_{Bind} H bond	ΔG_{Bind} Lipo	ΔG_{Bind} packing	ΔG_{Bind} solv. GB	ΔG_{Bind} vdW	Lig. strain energy	Stars	CNS[a]	HoA[b]	Predicted 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 absorption.

of the results is evident that ritonavir is able to maintain less number of hydrogen bonds than the hit compounds screened from our analysis. For instance, hit molecules are able to maintain three to four hydrogen bonds on average for the whole 10 ns simulation with key binding site residues of the main protease structure. In particular, the frequency of hydrogen bond was significantly higher in the case of DB02986 and DB08573 than other studied hits. It is interesting to note that hydrogen bond analysis correlates well with the RMSD behavior of the respective compounds and suggested that these molecules had a strong binding affinity in comparison to the reference molecules.

The compactness and total mobility of the complex structures were further ascertained by employing PCA analysis. The covariance matrix corresponds to all the compounds are plotted in Figure 5. The greater incidence of correlated (red) and anticorrelated (blue) motions between atoms is depicted by the higher color intensity in the covariance matrix. The trace covariance matrix of a protein complex with ritonavir, DB02986, DB03744, DB07800, and DB08573 was found to be 4.34893, 5.06568, 7.0659, 7.30152, and 4.36423, respectively. Also the eigenvalues of the protein complex with ritonavir, DB02986, DB03744, DB07800, and DB08573 were found to be 0.111, 0.136, 0.134, 0.55, and 0.119 nm^2. Interestingly, a trace of the covariance matrix and eigenvalues of DB02986 and DB08573 are relatively equivalent to that of ritonavir which highlights high compactness in comparison to DB03744 and DB07800.

Finally, validation of the compounds using FEL was performed, and results are presented as a contour map in Figure 6. The Gibbs free energy of all the compounds was generated using Cα atoms of the system. It is evident from the figure that ritonavir and DB07800 have one conformation of the global energy minima. Other compounds such as DB02986 and DB03744 exhibit three and two global energy minima conformations, respectively. Most importantly, seven global energy minima conformations have resulted in the case of DB08573 complex structure during the simulation. The highest number of global energy minima conformation associated with DB08573 and DB02986 signifies its binding efficacy against the main protease than reference compounds and other molecules studied in our analysis

Overall, there is a clear-cut correlation was resulted between the results of MD simulation and other analyses implemented in our study. Thus, we hypothesize from our simulation results that DB02986 and DB08573 could be potential repurposed candidate for the treatment and management of the Covid-19 pandemic situation.

The synergic activity of the hit compounds together with ritonavir was examined against the target protein. The choices of the program for the synergism are inspired by the literature.[37] The resultant sequential docked pose of

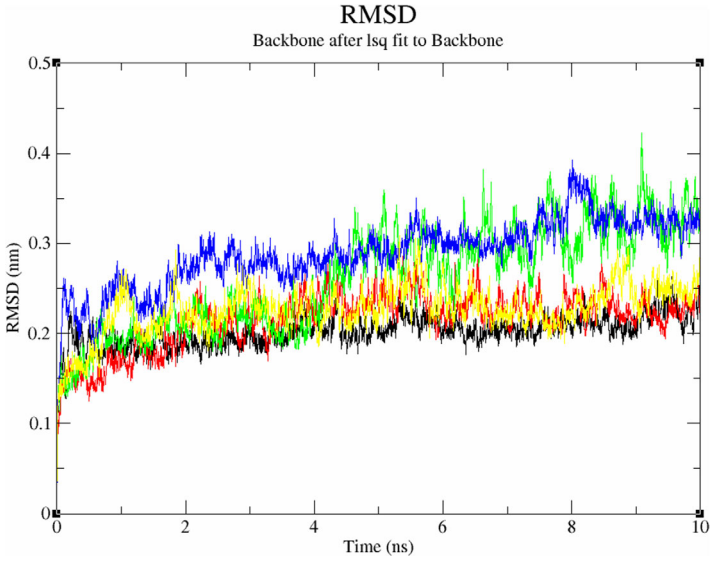

FIG 3 RMSD of the complex structures over the simulation time. Ritonavir (black); DB02986 (red); DB03744 (green); DB07800 (blue); DB08573 (yellow)

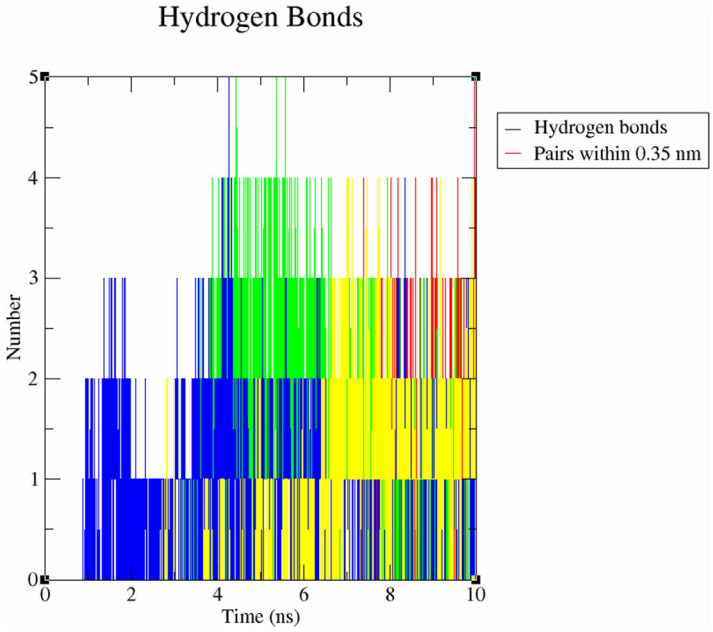

FIG 4 Intermolecular hydrogen bond analysis of complex structures over the simulation time. Ritonavir (black); DB02986 (red); DB03744 (green); DB07800 (blue); DB08573 (yellow)

the complex structures is visualized in Figure 7. Initially, both the hits and reference molecule docked individually against the main protease to understand the mode of interaction and free energy of binding. The process yielded binding energy values of −2.34, −3.93, and −4.64 kcal/mol, respectively, for ritonavir, DB08573, and DB02986. The binding pose was then examined through PyMOL to gain insight into the mechanism of interaction. Moreover, the binding pose in turn utilized for placing the grid position in the sequential docking. It is interesting to note that sequential docking demonstrated a favorable affinity for the resultant compounds. Of note, the binding affinity values of the hit compounds increase to −5.85 and −4.76 kcal/mol when it is docked along with the reference compound. This shows that the drug combination is indeed effective in the treatment than considering each drug separately.

4 | DISCUSSION

A total of four molecules were found to exhibit better docking scores, binding free energy, and pIC_{50} values than existing inhibitors considered in our analysis. Studies highlight that ligand strain energy is one of the significant parameters to be analyzed in the case of drug screening as it depicts an energy cost associated with ligand binding.[38] For instance, the molecules with less strain energy are likely to exhibit better binding free energy. It is important to note that all four compounds resulting from our analysis have yielded less strain energy than existing antivirals studied in our analysis, thus highlighting the tighter binding of hit molecules against the target protein. Moreover, Table 2 highlights that Coulombic and van der Waals interactions provided the most substantial

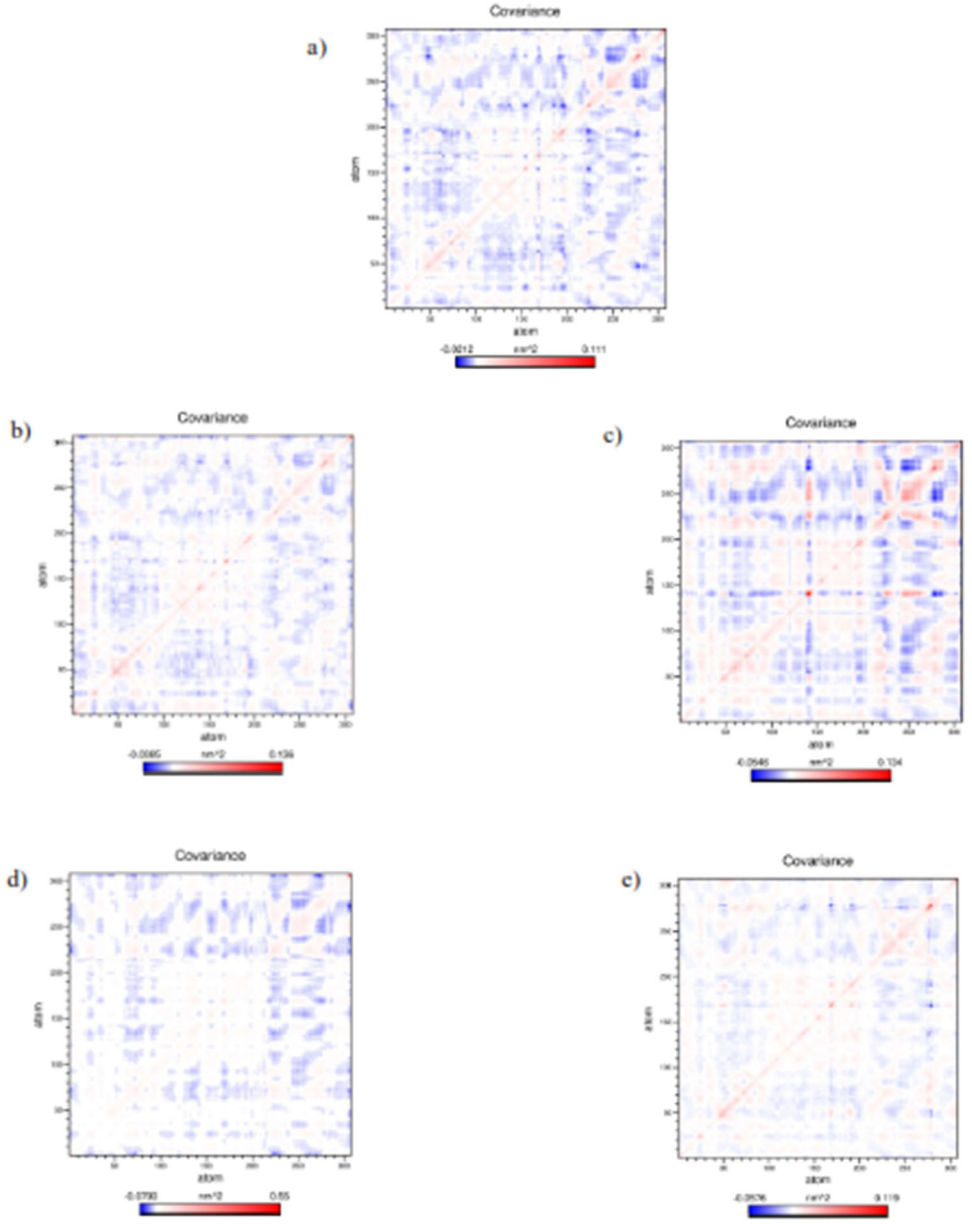

FIG 5 Covariance matrix generated using PCA analysis of complex structures. (A) Ritonavir; (B) DB02986; (C) DB03744; (D) DB07800; (E) DB08573

force for the binding of the inhibitor resulting from our study.

It is evident from Figure 2 that the binding of all four molecules mimics the binding pattern of existing inhibitors. It is interesting to note that GLN189 likely to play a pivotal role in the binding of lead compounds. The detailed listing of binding forces is given in Table S3 in the Supporting Information. Literature evidence highlights that the existence of Π−Π stacking might increase the stability and loading capacity of drugs.[39] Notably, most of the compounds resulting from our study are able to maintain Π−Π stacking in the binding pocket as like Lopinavir. Thus, it can exhibit tighter and stable binding with the main protease of Covid-19.

Furthermore, these four compounds were subjected to MD simulation and the trajectories were explored based on RMSD, hydrogen bond, PCA, and FEL data over 10 ns simulation. Although DB07800 and DB03744 showed promising docking scores and higher binding free energy, they failed to maintain stable conformations with the target protein during the simulation period. On the other hand, lower RMSD values exhibited by hit complexes DB02986 and DB08573 highlight the stable binding with the target protein during the simulation in comparison to DB07800 and DB03744. Note that this result correlates well with observed less ligand strain energy values during MM/GBSA calculation.

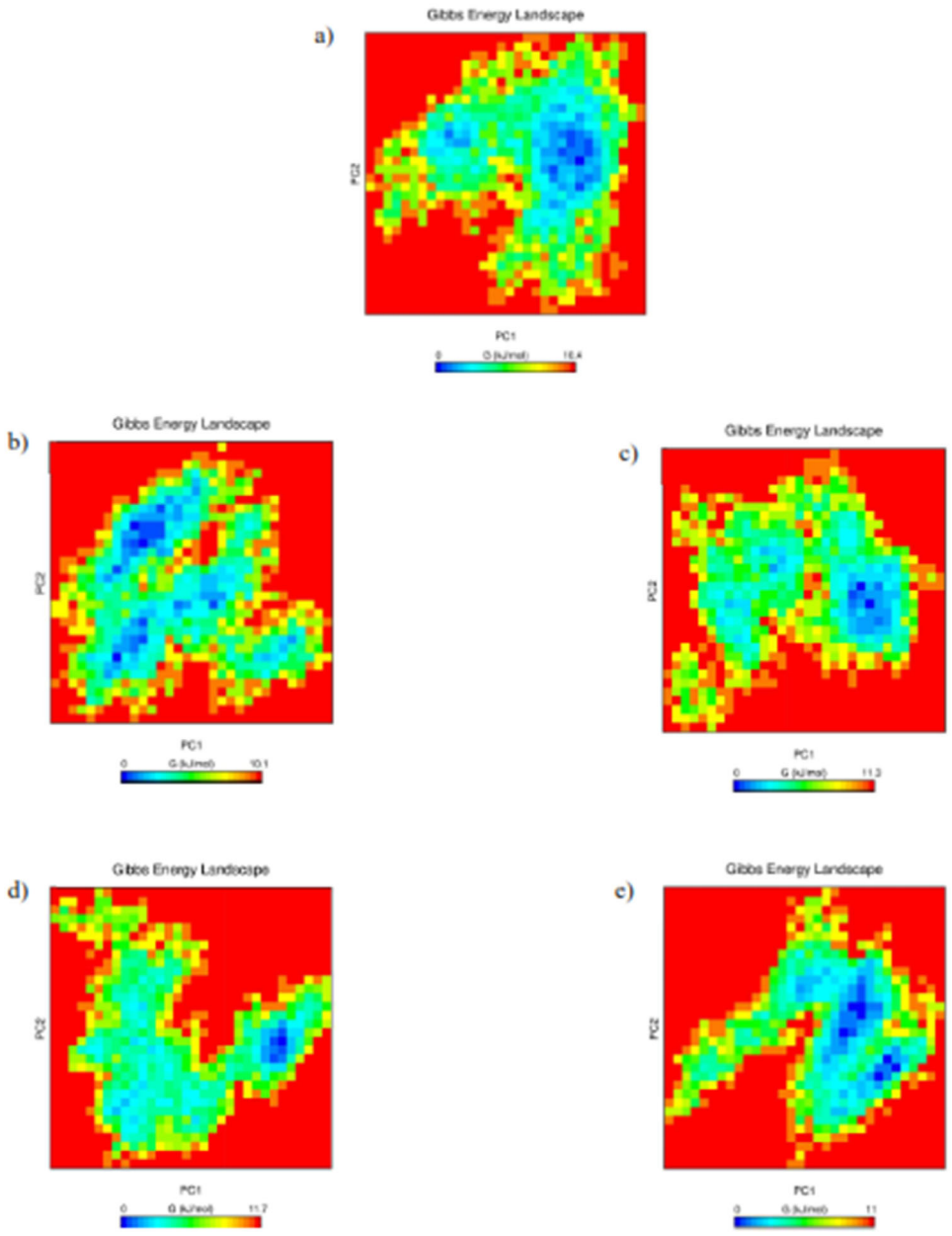

FIG 6 Gibbs energy landscape of complexes over the simulation time. (A) Ritonavir; (B) DB02986; (C) DB03744; (D) DB07800; (E) DB08573

Interestingly, the hydrogen bond data also correlate well with higher lipophilicity values obtained during MM/GBSA calculation and alongside RMSD behavior corresponds to DB02986 and DB08573. It is also evident from the literature that lipophilicity of the compound is crucial for maintaining the hydrogen bond with the partner molecule.[40]

Moreover, the eigenvalues for ritonavir, DB02986, and DB08573 complexes were found to be low as compared to DB03744 and DB07800 complexes which clearly indicate that less fluctuations and high compactness in the DB02986 and DB08573 bound conformations. Importantly, DB02986 and DB08573 are able to mimic the binding conformation of ritonavir as observed by the equivalent trace of the covariance matrix and eigenvalues in the PCA analysis. Altogether, less atomic movements observed in the DB02986 and DB08573 complexes indicate better stability of these compounds with the main protease. On the contrary, the resulting higher values in the case of DB03744 and DB07800 depict protein's larger expansion during its binding. The expansion of protein may thus be responsible for the lower stability and compactness. Finally, the highest number of global energy minima conformation associated with DB08573 and DB02986 clearly depicts the stable binding of these compounds in the binding pocket of the main protease than the other molecules such as ritonavir, DB07800, and DB03744.

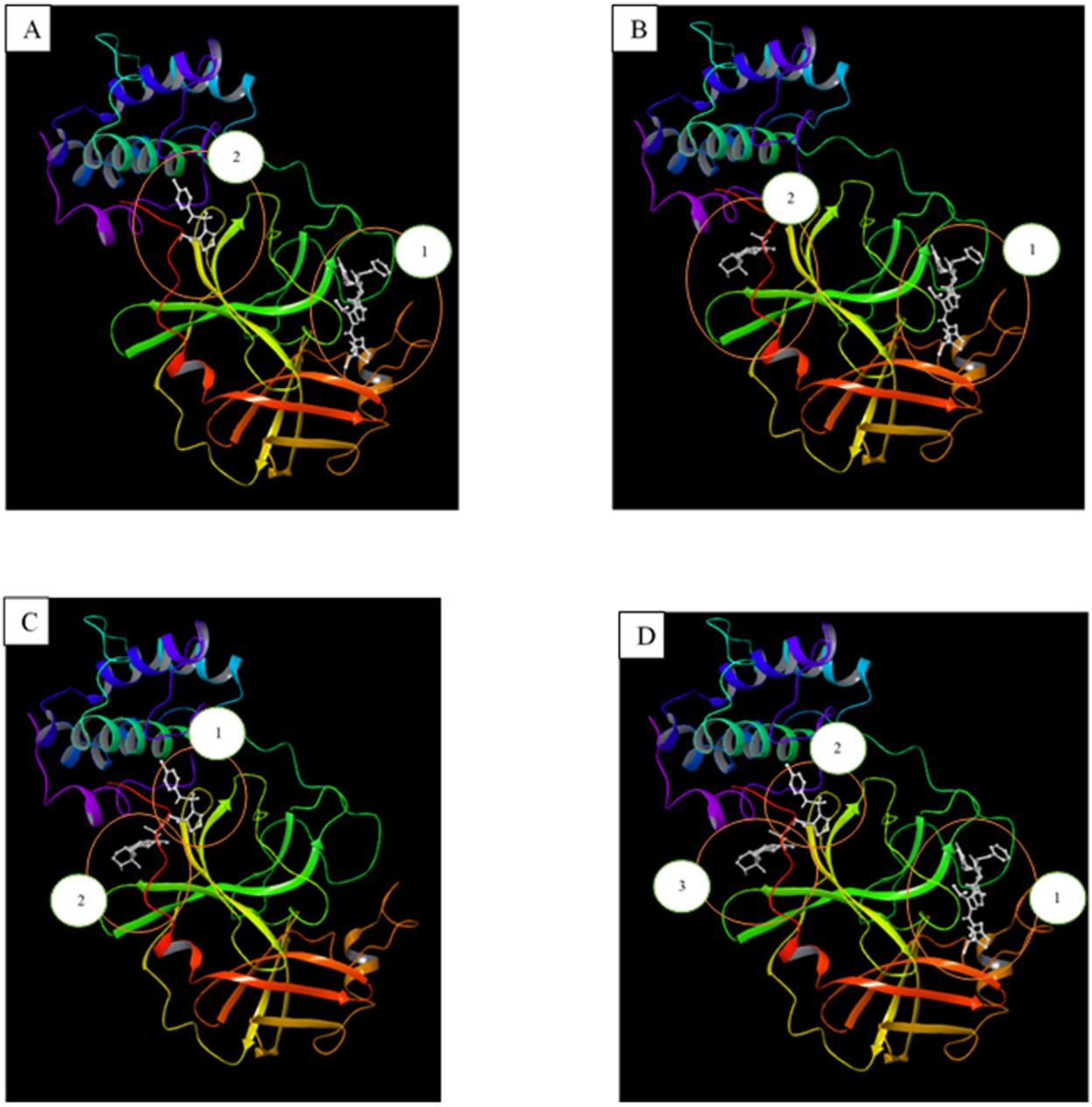

FIG 7 Synergistic binding effect of (A) combination of (1) ritonavir and (2) DB08573; (B) combination of (1) ritonavir and (2) DB02986; (C) combination of (1) DB08573 and (2) DB02986; (D) combination of (1) ritonavir, (2) DB08573, and (3) DB02986. The circle indicates the drug-binding sites and the respective docking scores mentioned below. The sequential order of the docking is represented by a numbers

It is interesting to note that the hit compounds obtained in our study composed of crucial scaffolds, namely sulfonamide and thiophene scaffolds. The 2D structure of hit compounds is shown in Figure 8. For instance, the sulfonamide scaffold is the backbone of the hit compound, DB08573. Notably, sulfonamides moieties can act as potential medicinal molecules in drug discovery and drug development with a broad spectrum of biological applications.[41,42] Of interest here, it has been shown on many occasions that sulfonamide structural units present within various molecules have exhibited interesting antiviral activities. The presence of sulfonamide scaffold within the various molecules is a common factor among the active compounds for combating different infectious viruses.[43–45]

Studies of DB02986 showed that this compound is found to act on carbonic anhydrase. It is thought that carbonic anhydrases have an essential role in the initiation of 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 restricts the binding of the virus in the host cell and even viral replication. Moreover, inhibitors of carbonic anhydrase are also reported to have activity against HIV infection.[46,47] It also highlights that hit compounds displayed significant structural similarity to that of recently

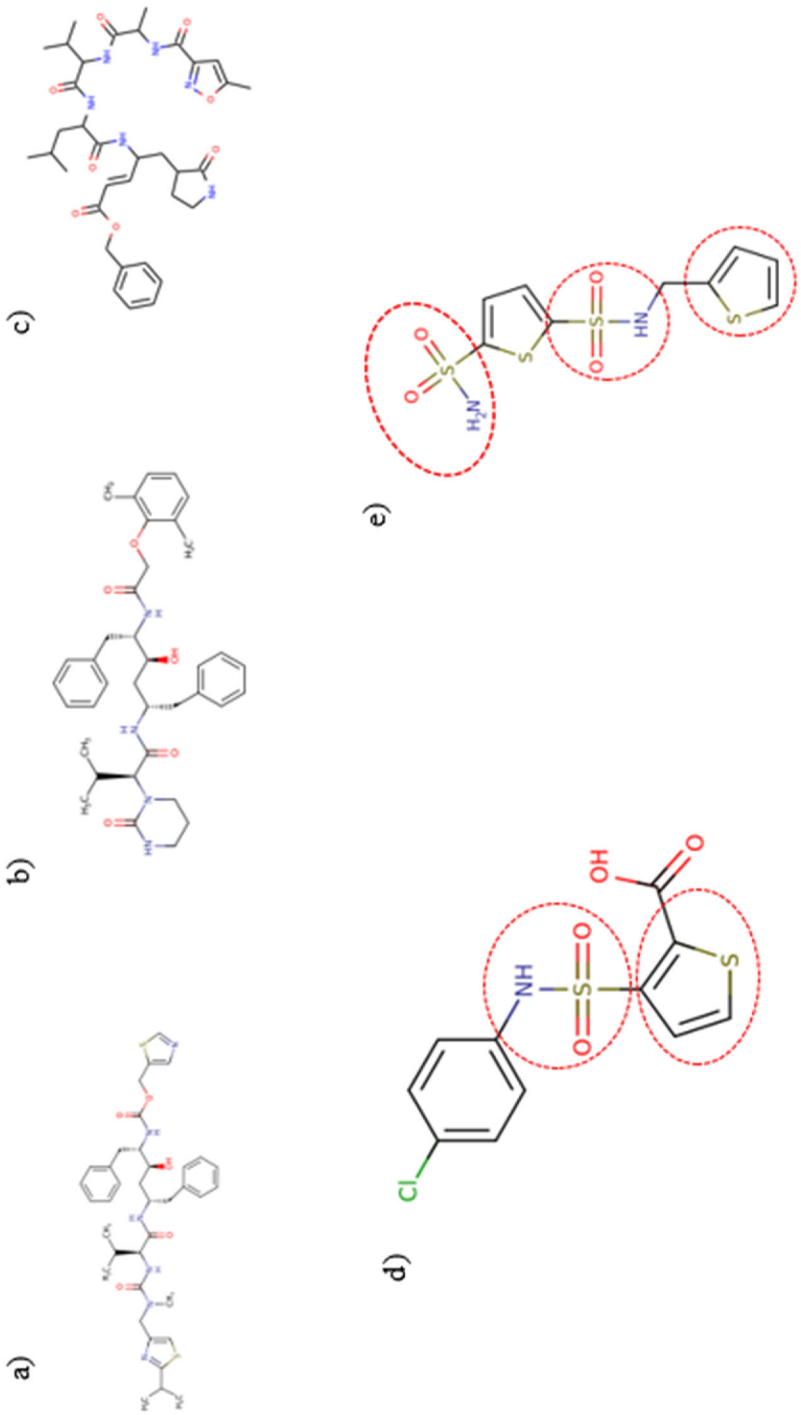

FIG 8 2D structure of reference and hit molecules. References (A) ritonavir; (B) lopinavir; (C) N3 – inhibitor (native ligand). Hit molecules: (D) DB08573; (E) DB02986. Important functional groups are highlighted as the red circle

discovered main protease inhibitors such as lurasidone sulfoxide, benzo[B]thiophene-2-carboxamidine.[48,49] For instance, the Tanimoto coefficient value of more than 0.4 was observed in all the studied cases. Overall, these findings provide evidence for the newly identified compounds that can be potentially developed as drugs for the management of Covid-19 infection.

In conclusion, we made a concerted effort to develop coronavirus therapeutic agents using an integrated machine learning–based drug-repurposing strategy. It is important to emphasize that compounds such as DB08573 (3-[(4-chloroanilino)sulfonyl]thiophene-2-carboxylic acid) and DB02986 (*N*-(2-thienylmethyl)-2,5-thiophenedisulfonamide) exhibit favorable results during MD simulation than cocrystallized native compounds and other existing inhibitors studied in our analysis. The ubiquitous experimental data support that the scaffolds identified in the hits exhibit antiviral 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 studies 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.

ORCID

Shanthi V https://orcid.org/0000-0003-2297-2751
Ramanathan K https://orcid.org/0000-0002-1613-0729

REFERENCES

1. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. Severe acute respiratory syndrome-related coronavirus: the species and its viruses – a statement of the Coronavirus Study Group. bioRxiv. 2020. https://doi.org/10.1101/2020.02.07.937862.
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet North Am Ed. 2020; 395: 565–74.
3. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579: 270–3.
4. Yesudhas D, Srivastava A, Gromiha MM. COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics. Infection. 2021;49(2):199–213.
5. Shanmugam A, Muralidharan N, Velmurugan D, Gromiha MM. Therapeutic Targets and Computational Approaches on Drug Development for COVID-19. Curr Top Med Chem. 2020; 20: 2210–20.
6. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses. 2020 12: 254.
7. Xue X, Yu H, Yang H, Xue F, Wu Z, Shen W, et al. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J Virol. 2008; 82: 2515–27.
8. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. Nature. 2020; 582: 289–93.
9. Xu Z, Peng C, Shi Y, Zhu Z, Mu K, Wang X, et al. 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. https://doi.org/10.1101/2020.01.27.921627.
10. Li Y, Zhang J, Wang N, Li H, Shi Y, Guo G, et al. Therapeutic drugs targeting 2019-nCoV main protease by high-throughput screening. BioRxiv. 2020. https://doi.org/10.1101/2020.01.28.922922.
11. Pillaiyar T, Manickam M, Namasivayam V, Hayashi Y, Jung S-H. An overview of severe acute respiratory syndrome–coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. J Med Chem. 2016 59: 6595–628.
12. Chen YW, Yiu C-PB, Wong K-Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL pro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. F1000Research. 2020;9: 129.
13. Liu X, Wang X-J. 2020 Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. J Genet Genom. 47: 119.
14. Vedani A, Dobler M, Hu Z, Smieško M. 2015 OpenVirtualToxLab—a platform for generating and exchanging in silico toxicity data. Toxicol Lett. 232: 519–32.
15. Tommonaro G, García-Font N, Vitale RM, Pejin B, Iodice C, Canadas S, et al. Avarol derivatives as competitive AChE inhibitors, non hepatotoxic and neuroprotective agents for Alzheimer's disease. Eur J Med Chem. 2016; 122: 326–38.
16. Pejin B, Iodice C, Kojic V, Jakimov D, Lazovic M, Tommonaro G. In vitro evaluation of cytotoxic and mutagenic activity of avarol. Nat Prod Res. 2016; 30: 1293–6.
17. Pejin B, Iodice C, Tommonaro G, De Rosa S. Synthesis and biological activities of thio-avarol derivatives. J Nat Prod. 2008; 71: 1850–3.
18. Wang N, Ren J-X, Xie Y. Identification of novel DHFR inhibitors for treatment of tuberculosis by combining virtual screening with in vitro activity assay. J Biomol Struct Dyn. 2019 37: 1054–61.
19. Rohini K, Ramanathan K, Shanthi V. 2019 Multi-dimensional screening strategy for drug repurposing with statistical framework—a new road to influenza drug discovery. Cell Biochem Biophys. 77: 319–33.
20. Moonsamy S, Bhakat S, Ramesh M, Soliman ME. . Identification of binding mode and prospective structural features of novel Nef protein inhibitors as potential anti-HIV drugs. Cell Biochem Biophys. 2017; 75: 49–64.
21. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013; 27: 221–34.
22. Fischer A, Sellner M, Neranjan S, Smieško M, Lill MA. Potential Inhibitors for novel coronavirus protease identified by

virtual screening of 606 million compounds. Int J Mol Sci. 2020; 21: 3626.
23. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018; 46: D1074-82.
24. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des. 2006; 20: 647–71.
25. Huang N, Kalyanaraman C, Bernacki K, Jacobson MP. Molecular mechanics methods for predicting protein–ligand binding. 2006PCCP 8: 5166–77.
26. Dixon SL, Duan J, Smith E, Von Bargen CD, Sherman W, Repasky MP. AutoQSAR: an automated machine learning tool for best-practice quantitative structure–activity relationship modeling. Future Med Chem. 2016; 8: 1825–39.
27. de Oliveira MT, Katekawa E. On the virtues of automated quantitative structure–activity relationship: the new kid on the block. Future Med Chem. 2018; 10: 335–42.
28. Duffy EM, Jorgensen WL. Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. J Am Chem Soc. 2000; 122: 2878–88.
29. Ali S, Khan FI, Mohammad T, Lan D, Hassan M, Wang Y. Identification and evaluation of inhibitors of lipase from Malassezia restricta using virtual high-throughput screening and molecular dynamics studies. Int J Mol Sci. 2019; 20: 884.
30. Schüttelkopf AW, Van Aalten DM. 2004 PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. Acta Crystallogr Sect D. 60: 1355–63.
31. Mohammad T, Khan FI, Lobb KA, Islam A, Ahmad F, Hassan MI. Identification and evaluation of bioactive natural products as potential inhibitors of human microtubule affinity-regulating kinase 4 (MARK4). J Biomol Struct Dyn. 2019; 37: 1813–29.
32. Cloete R, Akurugu WA, Werely CJ, van Helden PD, Christoffels A. Structural and functional effects of nucleotide variation on the human TB drug metabolizing enzyme arylamine N-acetyltransferase 1. J Mol Graph Model. 2017; 75: 330–9.
33. Therese PJ, Manvar D, Kondepudi S, Battu MB, Sriram D, Basu A, et al. Multiple e-pharmacophore modeling, 3D-QSAR, and high-throughput virtual screening of hepatitis C virus NS5B polymerase inhibitors. J Chem Inf Model. 2014 54: 539–52.
34. James N, Ramanathan K. Discovery of potent ALK inhibitors using pharmacophore—informatics strategy. Cell Biochem. Biophys 2018; 76: 111–24.
35. Verma RP, Hansch C. An approach toward the problem of outliers in QSAR. Bioorg Med Chem. 2005; 13: 4597–621.
36. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. N Engl J Med. 2020; 382: 1787–99.
37. Muralidharan N, Sakthivel R, Velmurugan D, Gromiha MM. Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 protease against COVID-19. J Biomol Struct Dyn. 2020. https://doi.org/10.1080/07391102.2020.1752802
38. Mobley DL, Dill KA. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get. Structure. 2009; 17: 489–98.
39. Shi Y, van Steenbergen MJ, Teunissen EA, Novo L, Gradmann S, Baldus M, et al. Π–Π stacking increases the stability and loading capacity of thermosensitive polymeric micelles for chemotherapeutic drugs. Biomacromolecules. 2013; 14: 1826–37.
40. Chen D, Zhao M, Tan W, Li Y, Li X, Li Y, et al. Effects of intramolecular hydrogen bonds on lipophilicity. Eur J Pharm Sci. 2019; 130: 100–6.
41. Azzam RA, Osman RR, Elgemeie GH. efficient synthesis and docking studies of novel benzothiazole-based pyrimidinesulfonamide scaffolds as new antiviral agents and Hsp90α inhibitors. ACS Omega. 2020; 5: 1640–55.
42. Khan F, Mushtaq S, Naz S, Farooq U, Zaidi A, Bukhari S, et al. Sulfonamides as potential bioactive scaffolds. Curr Org Chem. 2018; 22.
43. Yannopoulos CG, Xu P, Ni F, Chan L, Pereira OZ, Reddy TJ, et al. HCV NS5B polymerase-bound conformation of a soluble sulfonamide inhibitor by 2D transferred NOESY. Bioorg Med Chem Lett. 2004; 14: 5333–7.
44. Ding Y, Smith KL, Varaprasad CV, Chang E, Alexander J, Yao N. Synthesis of thiazolone-based sulfonamides as inhibitors of HCV NS5B polymerase. Bioorg Med Chem Lett. 2007; 17: 841–5.
45. Zhang N, Zhang X, Zhu J, Turpoff A, Chen G, Morrill C, et al. Structure–activity relationship (SAR) optimization of 6-(indol-2-yl)pyridine-3-sulfonamides: identification of potent, selective, and orally bioavailable small molecules targeting hepatitis C (HCV) NS4B. J Med Chem. 2014; 57: 2121–35.
46. Chen IH, Tsai AY, Huang YP, Wu IF, Cheng SF, Hsu YH, et al. Nuclear-encoded plastidal carbonic anhydrase is involved in replication of bamboo mosaic virus RNA in *Nicotiana benthamiana*. Front Microbiol. 2017; 8: 2046.
47. Temperini C, Innocenti A, Guerri A, Scozzafava A, Rusconi S, Supuran CT. 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. Bioorg Med Chem Lett. 2007 17: 2210–5.
48. Thurakkal L, Singh S, Roy R, Kar P, Sadhukhan S, Porel M. An in-silico study on selected organosulfur compounds as potential drugs for SARS-CoV-2 infection via binding multiple drug targets. Chem Phys Lett.(2021) 763: 138193.
49. Chowdhury KH, Chowdhury M, Mahmud S, Tareq AM, Hanif NB, Banu N, et al. Drug repurposing approach against novel coronavirus disease (COVID-19) through virtual screening targeting SARS-CoV-2 main protease. Biology. 2021; 10: 2.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: T MK, K R, James N, V S, K R. Discovery of potent Covid-19 main protease inhibitors using integrated drug-repurposing strategy. Biotechnology and Applied Biochemistry. 2021;68:712–725. https://doi.org/10.1002/bab.2159