Taylor & Francis
Taylor & Francis Group

Check for updates

Combination of QSAR, molecular docking, molecular dynamic simulation and MM-PBSA: analogues of lopinavir and favipiravir as potential drug candidates against COVID-19

Md. Oliullah Rafi[a]* (ID), Gourab Bhattacharje[b]*, Khattab Al-Khafaji[c], Tugba Taskin-Tok[c,d] (ID), Md. Almujaddade Alfasane[e], Amit Kumar Das[b], Md. Anowar Khasru Parvez[f,g] and Md. Shahedur Rahman[a] (ID)

[a]Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Jashore, Bangladesh; [b]Department of Biotechnology, Indian Institute of Technology-Kharagpur, Kharagpur, India; [c]Department of Chemistry, Faculty of Arts and Sciences, Gaziantep University, Gaziantep, Turkey; [d]Department of Bioinformatics and Computational Biology, Institute of Health Sciences, Gaziantep University, Gaziantep, Turkey; [e]Department of Botany, University of Dhaka, Dhaka, Bangladesh; [f]Department of Microbiology, Jahangirnagar University, Dhaka, Bangladesh; [g]Treasurer Office, Pabna University of Science and Technology, Pabna, Bangladesh

Communicated by Ramaswamy H. Sarma

ABSTRACT

Pandemic COVID-19 infections have spread throughout the world. There is no effective treatment against this disease. Viral RNA-dependent RNA polymerase (RdRp) catalyzes the replication of RNA from RNA and the main protease (M^{pro}) has a role in the processing of polyproteins that are translated from the RNA of SARS-CoV-2, and thus these two enzymes are strong candidates for targeting by anti-viral drugs. Small molecules such as lopinavir and favipiravir significantly inhibit the activity of M^{pro} and RdRp in vitro. Studies have shown that structurally modified lopinavir, favipiravir, and other similar compounds can inhibit COVID-19 main protease (M^{pro}) and RNA-dependent RNA polymerase (RdRp). In this study, lopinavir and its structurally similar compounds were chosen to bind the main protease, and favipiravir was chosen to target RNA-dependent RNA polymerase. Molecular docking and the quantitative structure-activity relationships (QSAR) study revealed that the selected candidates have favorable binding affinity but less druggable properties. To improve the druggability, four structural analogues of lopinavir and one structural analogue of favipiravir was designed by structural modification. Molecular interaction analyses have displayed that lopinavir and favipiravir analogues interact with the active site residues of M^{pro} and RdRp, respectively. Absorption, distribution, metabolism, excretion and toxicity (ADMET) properties, medicinal chemistry profile, and physicochemical features were shown that all structurally modified analogues are less toxic and contain high druggable properties than the selected candidates. Subsequently, 50 ns molecular dynamics simulation of the top four docked complexes demonstrated that CID44271905, a lopinavir analogue, forms the most stable complex with the M^{pro}. Further MMPBSA analyses using the MD trajectories also confirmed the higher binding affinity of CID44271905 towards M^{pro}. In summary, this study demonstrates a new way to identify leads for novel anti-viral drugs against COVID-19.

Abbreviations: 3D: 3-dimensional; ADMET: absorption, distribution, metabolism, excretion and toxicity; CADD: computer-aided drug design; : ; LogS: logarithm of solubility; M^{pro}: main protease; MD: molecular dynamics; MMPBSA: molecular mechanics Poisson Boltzmann surface area; PDB: protein data bank; QSAR: quantitative structure-activity relationships; RdRp: RNA-dependent RNA polymerase; RMSD: root mean square deviation; RMSF: root mean square fluctuation; SARS-CoV-2: severe acute respiratory syndrome coronavirus-2; TPSA: topological polar surface area

Highlights

- The main protease (M^{pro}) and the RNA-dependent RNA polymerase (RdRp) are two very important enzymes of the replication machinery of SARS-CoV-2.
- Lopinavir, a protease inhibitor, and Favipiravir, known to inhibit RNA polymerase, are two established antiviral drugs.
- Structural analogues of Lopinavir and Favipiravir were designed and their binding with M^{pro} and RdRp of SARS-CoV-2 was studied computationally.
- Among the three Lopinavir analogues that showed favorable binding, CID44271905 emerged as the best candidate to inhibit M^{pro}.

1. Introduction

COVID-19 has become a rapidly emerging public health issue which is caused by the outbreak of severe acute respiratory syndrome corona virus-2 (SARS-CoV-2) (Porcheddu et al., 2020). As of 6th June 2020, COVID-19 disease has been reported in 215 countries with 6,915,119 confirmed cases and 399,938 total deaths according to worldometers database (https://www.worldometers.info/coronavirus/). Fever, cough, nausea, shortness of breath are common symptoms of this disease and it spread in a community by the transmission from person to person (Chan et al., 2020; Samad et al., 2020). Clinical studies revealed that male and old patients have a higher case of fatality rate (CFR) than the others (Yang et al., 2020). Apart from respiratory dysfunction COVID-19 patient has a higher rate of kidney dysfunction (Li et al., 2020). There are no applicable anti-viral drugs, vaccines available for treating COVID-19.

The impact of COVID-19 is varying from country to country. Every day, the number of newly infected and death is increasing exponentially. So, there is an urgent essential need to discover out effective vaccine, drug, or therapies that can be used for the combating COVID-19. It is a lengthy process to discover a new drug. Thus, existing therapeutic molecules and their analogues with favorable drug properties would be the best solution for COVID-19 crisis.

SARS-CoV-2 is a positive-sense single-stranded RNA virus, it has many structural proteins (accessory proteins and spike glycoprotein) and contains various non-structural proteins that are encoded by the viral genome. 3-Chymotrypsin-like protease, helicase, papain-like protease, and RNA-dependent RNA polymerase (RdRp) are the main non-structural protein of this virus (Zumla et al., 2016). RdRp is an essential enzyme for the replication of SARS-CoV-2. Several antiviral drugs have been developed against zika, hepatitis C, and other coronaviruses by targeting viral-specific RdRp (Elfiky, 2020). A recent study explored that anti-viral drug remdesivir and favipiravir which are applicable against several RNA virus diseases. These drugs effectively inhibit RdRp and RNA polymerase and have essential effects against SARS-CoV-2 replication *in vitro* (Furuta et al., 2013; Harrison, 2020; Wang et al., 2020; Zhang et al., 2020).

The viral protease (M^pro) encoded by retroviral RNA genome which is essential for SARS-CoV-2 replication. This enzyme is the best candidates for targeting antiviral drugs (Tözsér, 2010; Zhao et al., 2014). Restricting the M^pro protein activity will block the viral replication (Li & De Clercq, 2020; Zhang et al., 2020). Protease inhibitor drugs can inhibit protease enzymes and reduce the number of viral particles. Nowadays, some of these drugs can work against COVID-19. Lopinavir and ritonavir are two HIV-1 protease inhibiting drugs, that were identified to be capable of inhibiting SARS-CoV-2 main protease (Khaerunnisa et al., 2020). These two drugs were used against HIV protease and could be used as a homologous target as the previous SARS-CoV main protease has 96.1% of similarity with the SARS-CoV-2 main protease (Khaerunnisa et al., 2020). In the United States, four hepatitis C virus (HCV) protease inhibitor and its similar drugs were identified as the treatment against the SARS-CoV-2 (de Leuw & Stephan, 2017; Ezat et al., 2019).

Recently, computer-aided drug design (CADD) has been successfully used in drug discovery. In comparison to traditional methods, this method significantly reduces the efforts, time, and expenses. Computational drug screening can search compound libraries to identify potential drug (Lionta et al., 2014). In this study, we have designed structural analogues of lopinavir and favipiravir to target the main protease (M^pro) and the RNA dependents RNA polymerase (RdRp) of SARS-CoV-2, respectively. We have also analyzed their physicochemical properties as a possible drug against COVID-19 by various computational methods, and finally obtained a lopinavir analogue, CID44271905, as a very good candidate for further experimental analyses.

2. Methods and materials

Chemoinformatics methods were applied to find potential inhibitor against the relevant enzymes and assessing *in silico* techniques. The selection and designing of the potential drug were executed in many steps: (i) QSAR analysis and structural modification of lopinavir and favipiravir, (ii) molecular docking of compounds with targeted proteins, (iii) ADMET analysis, and (IV) molecular dynamic simulation and MM-PBSA to stability and estimate ligand-binding affinities (Figure 1). This *in silico* analysis may provide the potential drugs which will be able to inhibit SARS-CoV-2.

2.1. Protein selection and preparation

The three-dimensional crystal structure of the M^pro (PDD ID: 6LU7) and RdRp (PDB ID: 6M71) were retrieved from the protein data bank (PDB) (https://www.rcsb.org/) (Berman et al., 2000). The PDB structures were not suitable for molecular docking studies (Kalia et al., 2017), therefore water molecules and ligands were removed from the PDB structure by Discovery Studio (DS) 3.5 (Accelrys, 2013; Kalia et al., 2017). The focused proteins of SARS-CoV-2, M^pro and RdRp, contain two chains (chains A and C) and four chains (chains A, B, C, and D), respectively. The large number of amino acids residues involved in the active sites were found in each of their A chains and these identified chains were saved as the PDB file format for further studies, where other chains were removed by using DS subprotocol.

2.2. Compounds selection, structural modification, and quantitative structure-activity relationships analysis

Based on the antiviral activities, we selected the both lopinavir and favipiravir drugs and their similar compounds (Figure 2) after running Quantitative structure-activity relationship (QSAR). QSAR analysis of these compounds was performed by Osiris properties explorer (https://www.organic-chemistry.org/prog/peo/) (Sander, 2001) and Molinspiration (https://www.molinspiration.com/cgi-bin/properties) (Banerjee et al., 2018). QSAR analysis of lopinavir revealed that the selection of four compounds that are highly similar to lopinavir (Figure 2). We modified the structure of lopinavir and favipiravir to create four lopinavir structural analogues (CID10009410, CID44271905, CID3010243, and CID271958) and one favipiravir analogue (CID89869520). CID10009410 (lopinavir analogue)
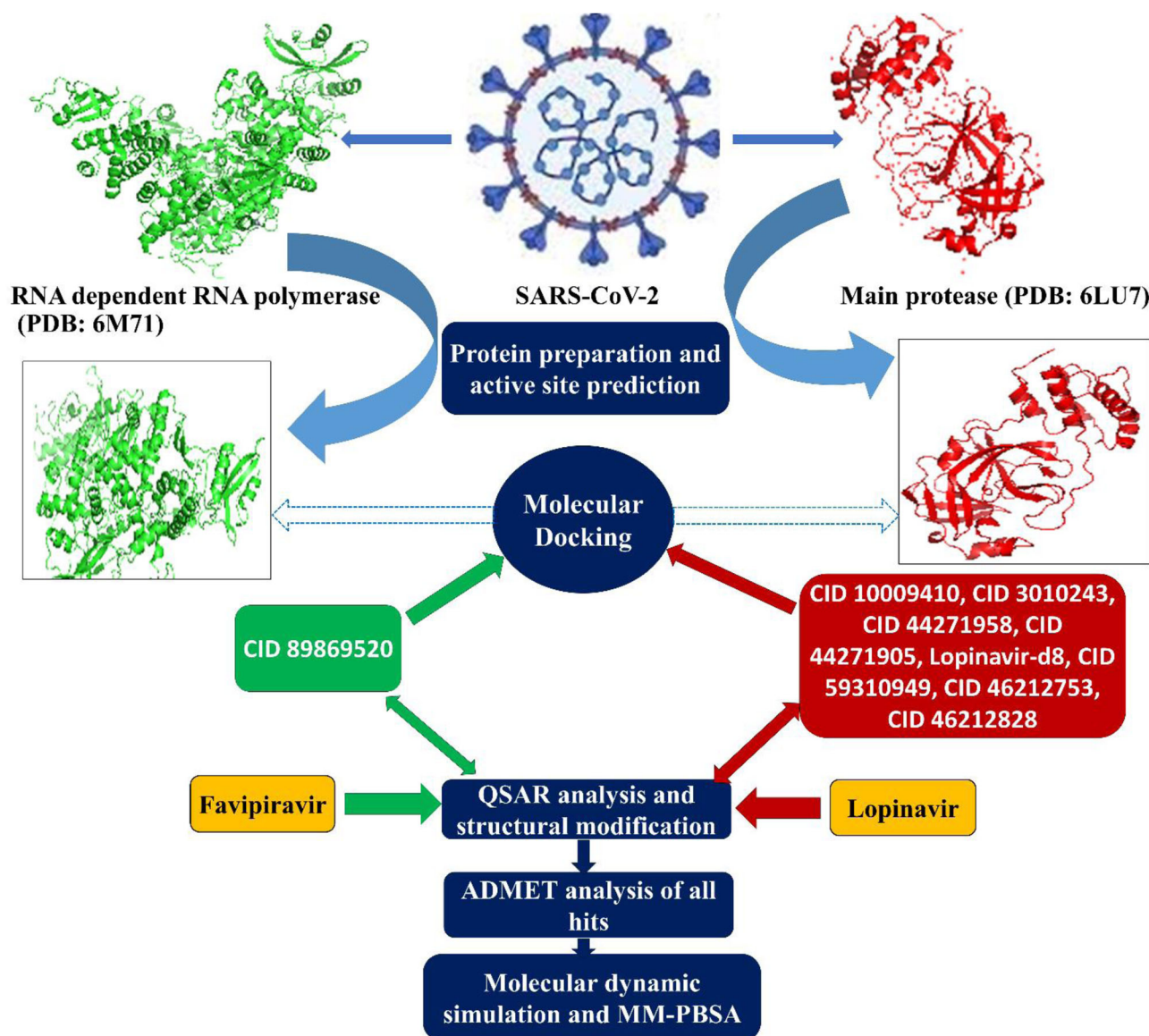
Figure 1. Overall workflow of the study.

and CID89869520 (favipiravir analogue) were generated by adding –F and –CH3 groups at the end of their two dimensional (2D) structures, respectively. CID44271905 (lopinavir analogue) and CID44271958 (lopinavir analogue) were generated by removing trimethyl-benzene fragment and adding 1,3,5- trimethyl-benzene and benzene fragments into the 2D structure of lopinavir and the structure of CID3010243 were generated by removing tetrahydro-pyrimidionepropylene urea fragment and adding 2-imidazolidone fragments into lopinavir (Figure 3). The structural modifications were carried out with help of ZINC database (https://zinc.docking.org/substances) and Osiris properties explorer. The structure of all selected and designed compounds was downloaded from Pubchem database (https://pubchem.ncbi.nlm.nih.gov/) as SDF format (Cheng et al., 2014).

2.3. Ligand preparation and energy minimization

The ligand preparation and energy minimization were performed by PyRx virtual screening tool (Dallakyan & Olson,

2015). We have used Universal Force Field (UFF) to minimize the energy of all ligands, with the help of Open Babel of PyRx software converted all ligands into the PDB format. Afterwards, ligands were first optimized and converted to PDBQT format using the graphical user interface of PyRx. This PDBQT format is suitable for molecular docking.

2.4. Molecular docking

Molecular docking of all the compounds was analyzed using PyRx software by Autodock wizard (Dallakyan & Olson, 2015). The prepared protein structures, provided by the Autodock wizard panel, were used to create the macromolecules. For molecular docking, ligands were considered to be flexible and the proteins were considered to be rigid. The grid parameters were generated using the Auto Grid engine in PyRx, this grid box was used to predict the amino acids in the active site of the protein that interacts with the ligands. The grid box for M^{pro} (6LU7_A) was set with the dimension of $X = 13.20$, $Y = 13.20$ and $Z = 13.20$ angstrom (Å) and center
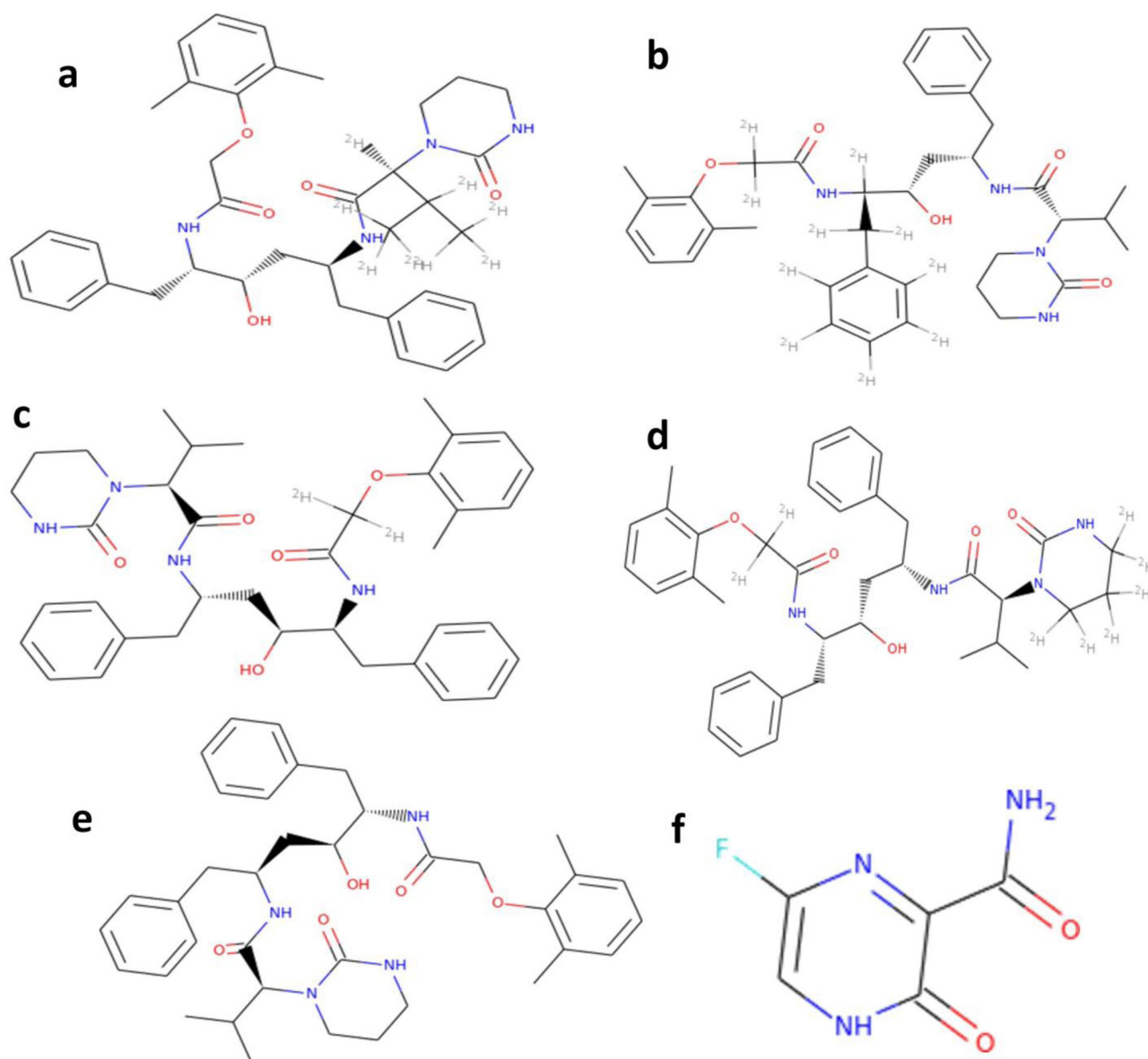
Figure 2. Two-dimensional chemical structures of the selected compounds (a) Lopinavir –d8; (b) CID 4621253; (c) CID 46212828; (d) CID 59310494; (e) Lopinavir; (f) Favipiravir.

of grid box were X= –11.13, Y = 13.51, Z = 70.22 and the grid box for RdRp (6M71_A) was set with the dimension of X = 17.34, Y = 17.34 and 17.34 angstrom and center of grid box were X = 125.02, Y = 131.9 and Z = 133.3. Lopinavir, four structurally modified lopinavir analogues and the compounds those highly similar to lopinavir (lopinavir, lopiravir-d8, CID59310949, CID46212753, CID46212828, CID10009410, CID3010243, CID4271958 and CID44271905) were targeted for 6LU7_A. Favipiravir and its structurally modified analogue CID89869520 were targeted for 6M71_A. Vina software was run in exhaustiveness = 8. As a result of these processes, docking results were visualized by PyMol 1.1 (Lill & Danielson, 2011) and Discovery studio visualizer 2019.

2.5. Pharmacoinformatics analysis

We have used various tools for pharmacoinformatics elucidation of selected and newly designed compounds. Absorption,

distribution, metabolism, excretion and toxicity (ADMET) and the pharmacophoric library screening properties were analyzed by pkCSM server (http://biosig.unimelb.edu.au/pkcsm/) (Pires et al., 2015), admet SAR (http://lmmd.ecust.edu.cn/admetsar2/) (Cheng et al., 2019), Molinspiration (https://www.molinspiration.com/cgi-bin/properties) (Banerjee et al., 2018), Swiss ADME (http://www.swissadme.ch/) (Daina et al., 2017), ProTox-2 (http://tox.charite.de/protox_II/) (Banerjee et al., 2018).

2.6. Molecular dynamic simulation

Molecular dynamics (MD) simulations on top four docked complexes (lopinavir analogues) were conducted using the GROMACS version 2019.4 (Abraham et al., 2015), with Gromos54a7 all-atom force field (Schmid et al., 2011) on Ubuntu operating system (version 18.04). SPC water model was chosen to simulate molecular dynamics (MD) in explicit solvation. To calculate the dimension of the rectangular box and the number of ions required to maintain
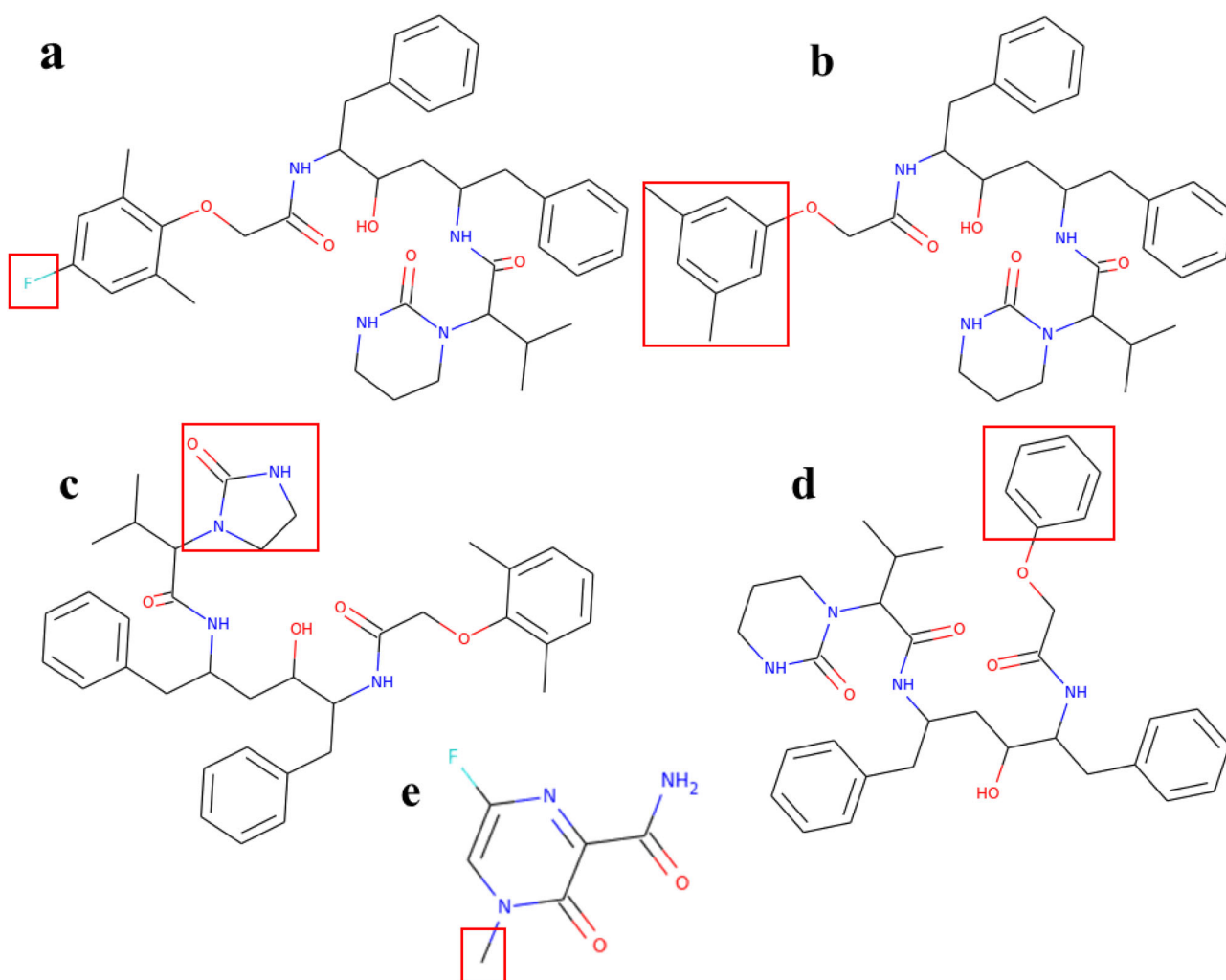
Figure 3. Two-dimensional chemical structures of the newly designed (a) CID10009410; (b) CID44271905; (c) CID3010243; (d) CID44271958 and (e) CID89869520. CID 10009410 (Lopinavir analogues) and CID 89869520 (Favipiravir analogues) were modified by adding –F and –CH3 groups at the end, respectively. CID 44271905 (Lopinavir analogue) and CID 44271958 (Lopinavir analogue) were modified by removing trimethyl-benzene fragment and adding 1,3,5-trimethyl-benzene and benzene fragments, respectively. The structure of CID 3010243 was modified by removing tetrahydro-pyrimidionepropylene urea fragment and adding 2-imidazolidone fragment. All structural modifications have been shown inside the red box.

a physiological NaCl concentration of 0.16 M Packmol package (Martinez et al., 2009) was used. The rectangular box dimensions for periodic boundary conditions, while keeping a minimum distance from any atom to the boundary of the box at 1 nm, were calculated to be 7.17 nm × 8.64 nm x 8.02 nm. The charge parameters of the ligand were obtained from the ProDRG webserver (Schüttelkopf & van Aalten, 2004). To neutralize the system, 41 Na$^+$ atoms and 37 Cl$^-$ atoms were added to the solution. Steepest descent algorithm was used for energy minimizations and maximum force, F_{max} was set to not exceed 1000 kJ/mol.nm. The system was equilibrated with temperature 300 K and pressure 1 bar by two consecutive 100 ps simulations with canonical NVT ensembles and isobaric NPT ensembles respectively. The protein and the docked ligands were restrained independently during equilibration and were thermostat coupled for the entire simulation. MD simulations were run for 50 ns with stable temperature and pressure with a time step of 2 fs and long-range interaction cut-off of 1 nm. Trajectories were analyzed using GROMACS tools. The graphs depicting the results of the simulation were also plotted using Origin2020b (OriginLab Corporation, USA).

2.7. Principal component analysis

Principal component analysis (PCA) is a statistical technique that reduces the complexity of the dynamics data and extracts the collective and correlated motions of the atoms of the biological macromolecules. PCA was carried out on snapshots stored every 1 ps of the 50 ns simulations. Covariance matrices of C α atoms were constructed to capture the essential collective motions of the protease with and without ligands. Positive sign of the entries in the covariance matrix signifies correlated motion whereas negative sign indicates anticorrelated motion between two C α atoms. Covariance matrices were then diagonalized to produce set of eigenvectors with respective eigenvalues. The eigenvalues represent the relevance of their corresponding eigenvectors in the system dynamics, where the eigenvectors with the largest eigenvalue signifies most relevant motions. Principal components (PCs) are obtained by taking the projection of the displacement of the C α atoms at each time point onto the eigenvectors. Gromacs 2019.4 tools, "gmx covar" and "gmx anaeig" were used to generate the covariance matrices, eigenvectors, and two-dimensional plots of PC1 versus PC2.

2.8. MMPBSA analysis

Binding free energy of the protein-ligand complex can be calculated as follows:

$$\Delta E_{MMPBSA} = E_{complex} - (E_{protein} + E_{ligand})$$

where $E_{complex}$ is the total MMPBSA energy of protein-ligand complex, $E_{protein}$ and E_{ligand} are the total solution free energies of the isolated protein and ligand, respectively. Now, each individual total free energy can be expressed as:

$$E_a = <E_{MM}> - TS + <E_{solvation}>$$

where a is either protein or ligand or protein-ligand complex. $<E_{MM}>$ is the ensemble average value of the molecular mechanic's potential energies in vaccum. T is absolute temperature and S represents entropy, and together TS contribute to the entropic contribution to the total free energy in vaccum. Calculation of the accurate values of entropic contribution (TS) is computationally expensive procedure and adds little to the overall free energy values, and thus generally omitted in calculating ΔE_{MMPBSA}. $<E_{solvation}>$ is the average free energy of solvation. The $<E_{MM}>$ in the above equation is sum of both bonded and nonbonded interactions of the molecules. The bonded interactions include bond energies, bend energies, dihedral energies and improper torsion energies. The nonbonded interactions include van der Waals energy (E_{vdW}) and electrostatic energy (E_{elec}).

$$<E_{MM}> = E_{bonded} + E_{nonbonded} = E_{bonded} + (E_{vdW} + E_{elec})$$

Now, E_{vdW} and E_{elec} are calculated using Lennard-Jones potential and Coulomb potential, respectively. As number of bonds formed or broken during a MD simulation is zero, E_{bonded} part contributes nothing to overall ΔE_{MMPBSA}. Calculation of entropic contribution is computationally more demanding and adds very little to the system information. In addition, estimation of entropic contribution values varies significantly to predict any stable interaction using MMPBSA analysis. Average free energy of solvation contains two parts, polar (E_{polar}) and nonpolar ($E_{nonpolar}$):

$$E_{solvation} = E_{polar} + E_{nonpolar}$$

Polar solvation energy (E_{polar}) is estimated using Poisson-Boltzmann (PB) equation. Linear model was used to solve the Poisson-Boltzmann equation (Baker et al., 2001) to calculate the polar part of the solvation energy (E_{polar}). Further, solvent accessible surface area (SASA) method was employed to calculate the nonpolar part of the solvation energy ($E_{nonpolar}$). In this model, it is assumed that $E_{nonpolar}$ is linearly dependent on the SASA and written as follows:

$$E_{nonpolar} = gA + b$$

where γ is a parameter representing surface tension of the solvent, A is the calculated SASA of the molecule and b is a fitting parameter. Surface tension constant (γ) was taken 0.0226778 kJ/mol·Å^2 and the SASA constant was also taken 3.84928 kJ/mol for fitting. Binding energies were extracted using MmPbSaStat.py program and residue-specific contributions towards binding were obtained from the MmPbSaDecomp.py program (Kumari et al., 2014). From the original MD trajectory of 50 ns, a shorter trajectory, consisting of last 20 ns, was used for the MMPBSA calculations where trajectory parameters were stored for every 50 ps.

3. Results

3.1. Quantitative structure-activity relationships (QSAR) analysis and structural modification

The QSAR method explains the properties like topologic, electronic, steric, and hydrophobic of numerous molecules, it has been determined through experimental methods, and only more recently by computational methods (Park et al., 2008; Suh et al., 2002). Osiris properties explorer was shown drug score and drug-likeness of all compounds. The selected four compounds displayed low drug score and drug-likeness. On the other hand, structurally modified compounds were also obeying QSAR properties including physicochemical (TPSA, cLogP, MW, and solubility) and toxicological molecular properties and some of them contain better drug scores than lopinavir and favipiravir (Figure 4). Drug score of CID3010243, CID44271958, CID44271905 and CID89869520 is 0.19, 0.36, 0.26, and 0.95, respectively where lopinavir and favipiravir drug score is 0.17 and 0.93. The newly designed compounds based on lopinavir and favipiravir drugs also contain better drug-likeness properties than selected compounds (Lopinavir-d8, CID 4621253, CID 46212828, CID 59310494, Lopinavir, Favipiravir) (Figure 5).

3.2. Molecular docking analysis

The Vina wizard has displayed nine possible binding positions as an output for each compound. The favorable binding affinity was estimated by finding the results of less than 1.0 Å in positional root mean square deviation (RMSD). The highest binding energy (most negative) was measured as the ligand with maximum binding affinity. The selected five compounds efficiently bind to the main protease of SARS-CoV-2. Docking energies of lopiravir-d8, CID59310949, CID46212753, CID46212828, and lopinavir were −12.5, −12.4, −10.9, −10.5, −7.00 kcal/mol, respectively and the binding affinity of favipiravir with RNA dependent RNA polymerase was −4.9 kcal/mol (Table 1). Moreover, these compounds, the docking results of structurally modified analogues of the remaining five drugs show favorable interaction with both targets, M^{pro} and RdRp. Structural analogues of lopinavir CID10009410, CID3010243, CID44271958, and CID44271905 binding affinities against the main protease protein were −8.1, −7.7, −7.4, and −7.3 kcal/mol, respectively (Table 2).

The amino acid interactions between main protease as target and these compounds as ligands were also identified (Figure 6–9). His41, Met149, Cys145, Gln189, Met164, and Leu27 residues in main protease were found in binding interactions with the four designed analogues (CID10009410, CID44271905, CID3010243, CID44271958). Among these, His41, Met149, and Cys145 residues having been in active site of the main protease were common interactions between all designed compounds and the related target.
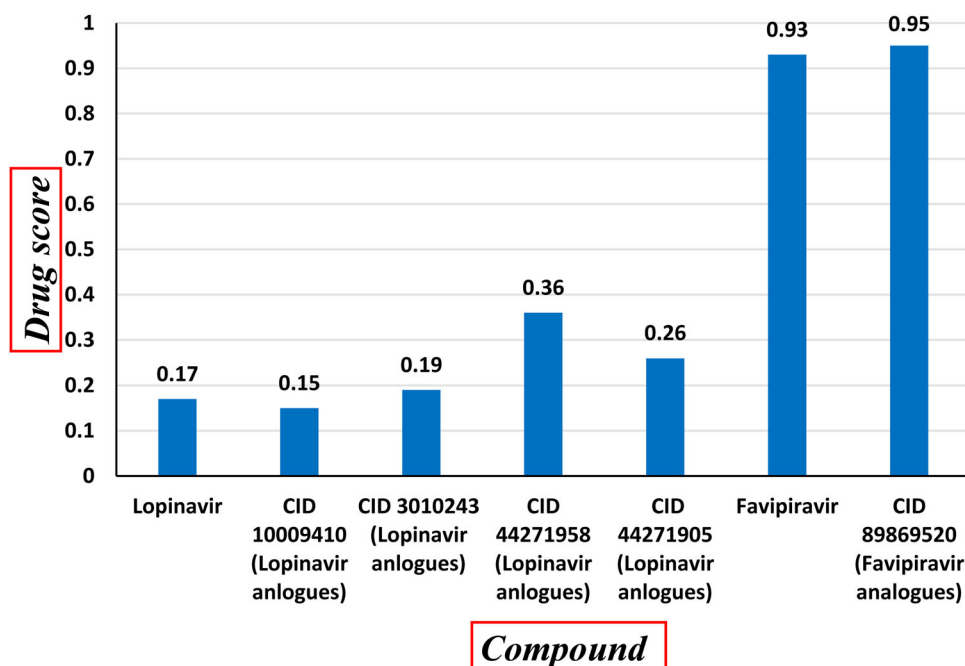
Figure 4. Drug score of five structurally modified analogues, lopinavir and favipiravir.
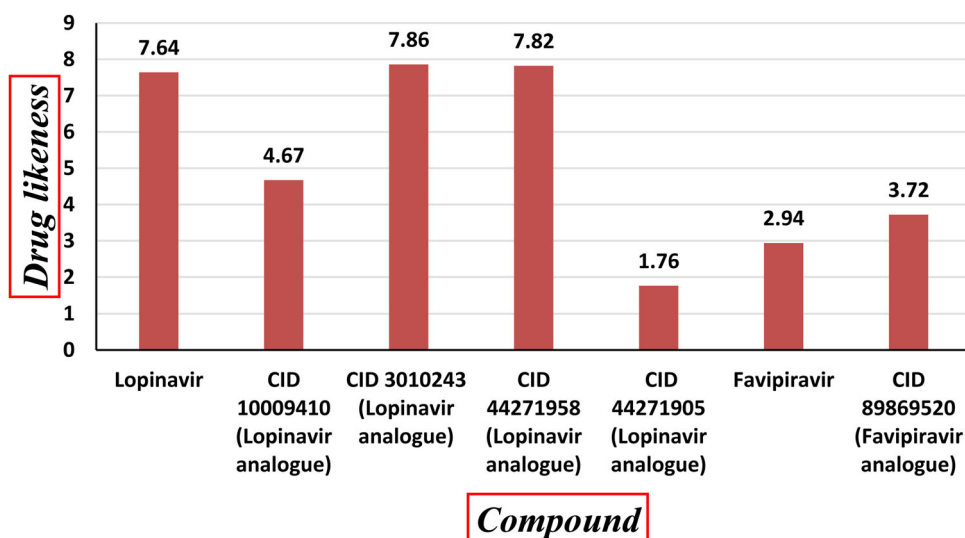

**Figure 5.** Druglikeness of five structurally modified analogues, lopinavir and favipiravir.

Table 1. Molecular docking selected compounds with the target protein.

S. no.	Compound ID	Compound name	Target protein	Binding energy(kcal/mol)
1.	CID46212486	Lopinavir-d8	Main protease of SARS-CoV-2 (PDB:6LU7) A chain	−12.5
2.	CID46212753	(2S)-3-methyl-2-(2-oxo-1,3-diazinan-1-yl)-N-[(2S,4S,5S)-5,6,6-trideuterio-5-[[2,2-dideuterio-2-(2,6-dimethylphenoxy)acetyl]amino]-4-hydroxy-6-(2,3,4,5,6-pentadeuteriophenyl)-1-phenylhexan-2-yl]butanamide	Main protease of SARS-CoV-2 (PDB:6LU7) A chain	−10.9
3.	CID46212828	(2S)-N-[(2S,4S,5S)-5-[[2,2-dideuterio-2-(2,6-dimethylphenoxy)acetyl]amino]-4-hydroxy-1,6-diphenylhexan-2-yl]-3-methyl-2-(2-oxo-1,3-diazinan-1-yl)butanamide	Main protease of SARS-CoV-2 (PDB:6LU7) A chain	−10.5
4.	CID59310949	(2S)-2,3,4,4,4-pentadeuterio-N-[(2S,4S,5S)-5-[[2-(2,6-dimethylphenoxy)acetyl]amino]-4-hydroxy-1,6-diphenylhexan-2-yl]-2-(4,4,5,5,6,6-hexadeuterio-2-oxo-1,3-diazinan-1-yl)-3-(trideuteriomethyl)butanamide	Main protease of SARS-CoV-2 (PDB:6LU7) A chain	−12.4
5.	CID92727	Lopinavir	Main protease of SARS-CoV-2 (PDB:6LU7) A chain	−7.00
6.	CID492405	Favipiravir	RdRp of SARS-CoV-2 (PDB: 6M71) A chain	−4.9

Table 2. Molecular docking of structurally modified analogues and amino acid interaction with protein.

S. no.	Compound ID	Compound name	Target protein	Binding energy (kcal/mol)	Binding residues of Protein	Residues found in the active site	Num of H-bond and involve residues
1	CID10009410 (Lopinavir analogue)	(2R)-N-[(2R,4R,5R)-5-[[2-(4-fluoro-2,6-dimethylphenoxy)acetyl]amino]-4-hydroxy-1,6-diphenylhexan-2-yl]-3-methyl-2-(2-oxo-1,3-diazinan-1-yl)butanamide	Main protease of SARS-CoV-2 (PDB:6LU7) A chain	−8.1	His41, Met149, Gln189, Glu166, Met164, Leu144, Gly143, Cys145,Ser144 and Leu27	Leu27, His41, Met149, Leu14, Gly143, Ser144, Cys145 and Glu166	6. Cys145,Glu166, Gln189, Ser144 and Leu141
2.	CID3010243 (Lopinavir analogue)	(2S)-N-[(2S,4S,5S)-5-[[2-(2,6-dimethylphenoxy)acetyl]amino]-4-hydroxy-1,6-diphenylhexan-2-yl]-3-methyl-2-(2-oxoimidazolidin-1-yl)butanamide	Main protease of SARS-CoV-2 (PDB:6LU7) A chain	−7.7	Cys145,Met165, Met149, His41, Asp187,Gln189, Thr190, Tyr54 and Asn142	His41,Asn142, Cys145 and Met165	5. Asn142,Cys145, Tyr54, Asp187 and Gln189
3.	CID44271958 (Lopinavir analogue)	(2R)-N-[(2R,4R,5R)-4-hydroxy-5-[(2-phenoxyacetyl)amino]-1,6-diphenylhexan-2-yl]-3-methyl-2-(2-oxo-1,3-diazinan-1-yl)butanamide	Main protease of SARS-CoV-2 (PDB:6LU7) A chain	−7.4	Thr190, Pro168, Asn142,Gly143, Cys145, His41, His164, Met165, Met149, Arg188, and Gln189	Cys145, His41, Met165, Met149, Gly143, Asn142	7. Asn142,Gln189, Gly143, Arg188,Met165, His164 and Thr190
4.	CID44271905 (Lopinavir analogue)	(2R)-N-[(2R,4R,5R)-5-[[2-(3,5-dimethylphenoxy)acetyl]amino]-4-hydroxy-1,6-diphenylhexan-2-yl]-3-methyl-2-(2-oxo-1,3-diazinan-1-yl)butanamide Leu141,Cys145,Met165,Glu166, and Met49	Main protease of SARS-CoV-2 (PDB:6LU7) A chain 4. Leu141, Cys145, Glu166 and Thr190	−7.3	Leu27, His41, Leu141,Cys145, Met165,Glu166, Thr90, Asp188, Arg189, Met149	Leu27, His41,	
5.	CID89869520 (Favipiravir analogue)	6-Fluoro-4-methyl-3-oxopyrazine-2-carboxamide	RdRp of SARS-CoV-2 (PDB: 6M71) A chain	−5.1	Lys798, Trp617, Trp800, Asp760, Asp761	Lys798, Trp617, Trp800, Asp760, Asp761	4. Lys798, Trp617, Trp800 and Asp761
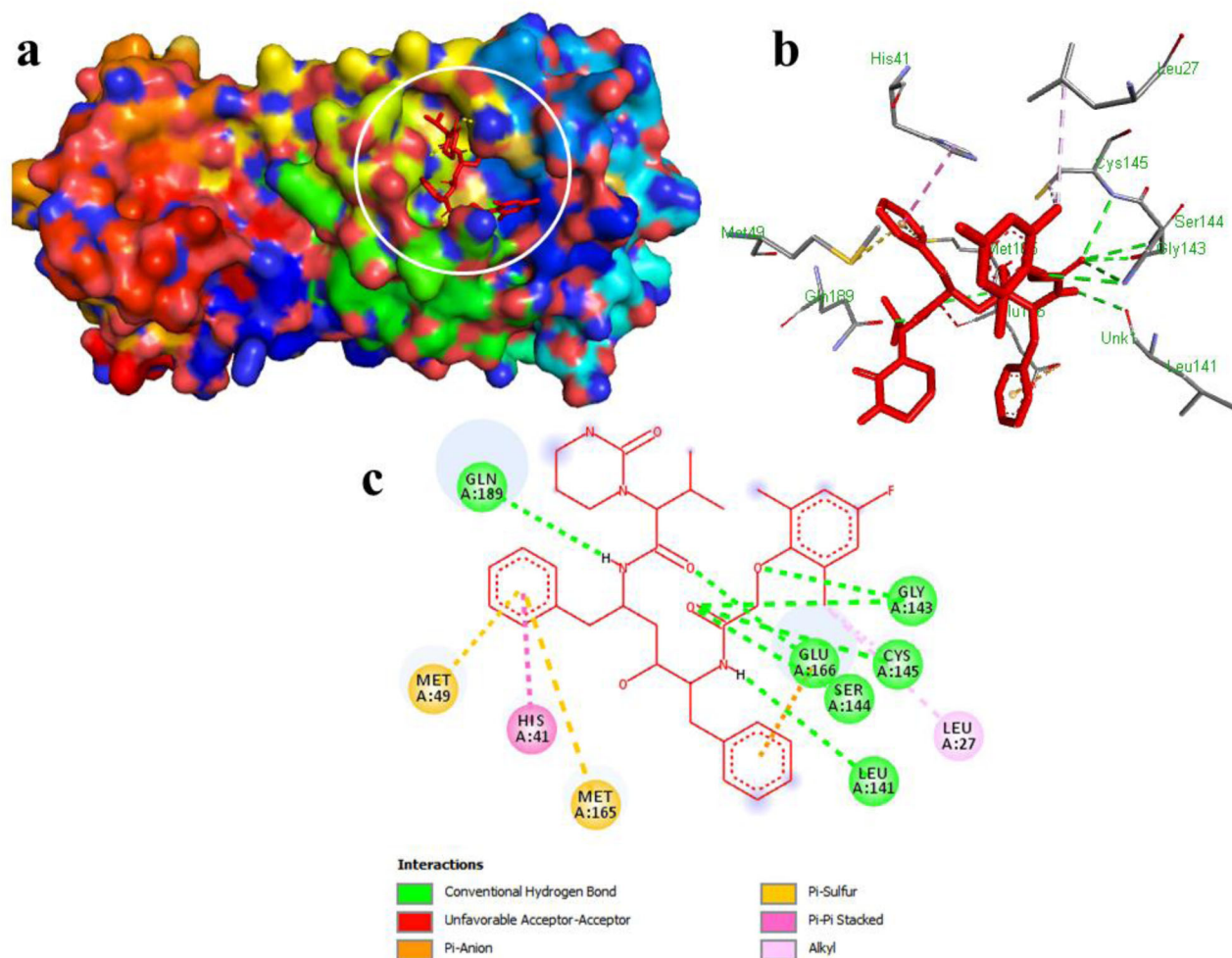
Figure 6. Interaction between CID10009410 (Lopinavir analogue) and main protease. (a) CID10009410 and main protease complex; (b) three-dimensional (3D) binding mode of the complex; (c) two-dimensional binding mode of the complex.

On the other hand, favipiravir analogue CID89869520 binding affinity with RdRp (6M71_A) of SARS-CoV-2 was −5.1 kcal/mol, this value is more negative than favipiravir binding affinity. This analogue binds with Lys798, Trp617, Trp800, Asp760, and Asp761 residues of RdRp, (Figure 10) these amino acids also found in the active site residues of RdRp.

Docking analysis confirmed that all designed lopinavir and favipiravir analogues bind in the same binding site of the main protease and RdRp of SARS-CoV-2, respectively (Figure 11).

3.3. Pharmacoinformatics analysis

The pharmacophore and ADMET properties of the compounds were analyzed for their toxicity, pharmacokinetics, physiochemical properties, water-solubility, lipophilicity, drug score, and drug-likeness using various server and software mention in Section 2 and the results are listed in Table 3, and supplementary material Tables 1 and 2. Among the ADMET properties, selected four compounds which were targeted for the main protease of SARS-CoV-2 show moderate pharmacokinetics properties. Human intestinal absorption is highest for lopinavir and CID46212828 (66.4% and 65.33%), these two drugs also show the highest Caco-2 permeability (0.96 and 0.93).

On the other hand, the compounds that formed by structural modification of lopinavir and favipiravir show good pharmacokinetics and QSAR properties than selected six compounds. Among them, human intestinal absorption is highest for CID44271905 and CID89869520 (favipiravir analogue) (0.764 and 0.98 out of 1) while Caco-2 permeability rate is in between (−0.008 and 1.53) log Papp in 10^{-6} cm/s. Blood–brain distribution of these drugs is less than zero. All candidates show a very high metabolic score. However, out of structurally modified analogues, CID10009410 and CID3010243 showed irritating effects (supplementary material Table 2). Ligand Properties of all structural analogues were found in an acceptable range. Consensus Log Po/w of four candidates is less than 5. Furthermore, their drug score and drug-likeness revealed that these candidates are more likely to be used as drugs.

3.4. Molecular dynamic simulation

Molecular dynamics simulations were performed to check the stability of the docked complexes by using GROMACS software. The main protease without any ligand and with ligands, CID3010243, CID10009410, CID44271905 and CID44271958, were simulated for 50 ns in an explicit solvation system. A
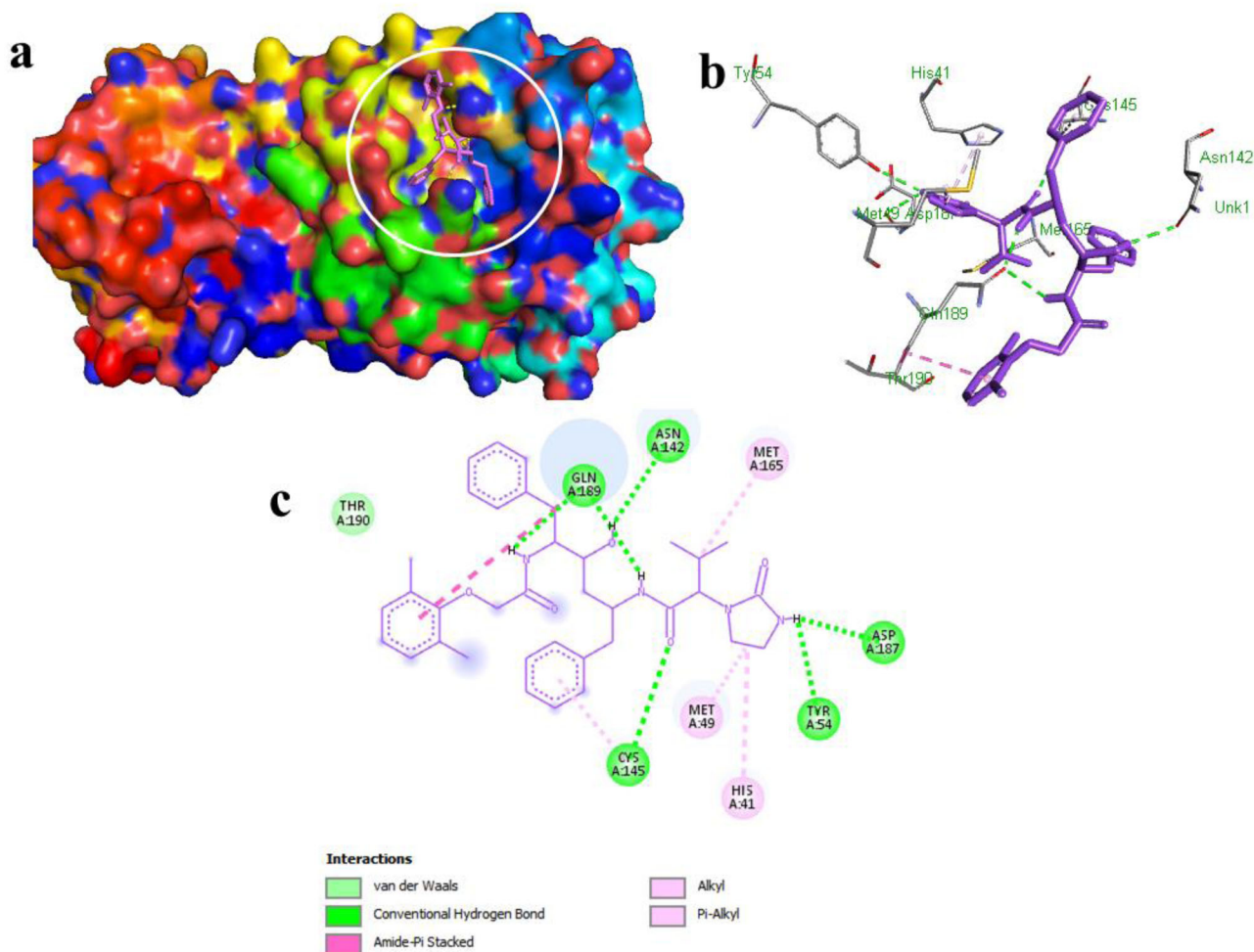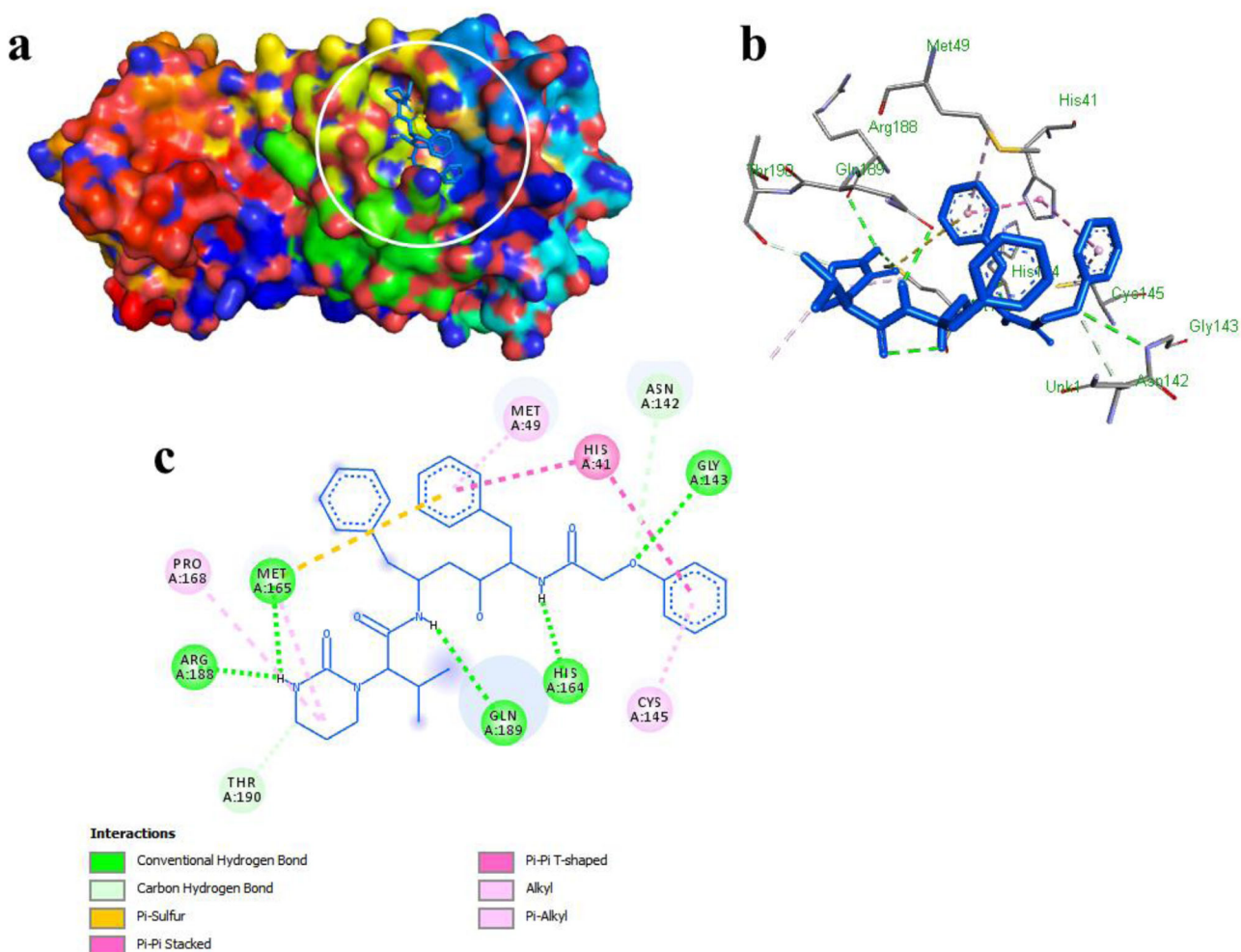
Figure 7. Interaction between CID3010243 (Lopinavir analogue) and main protease (a) CID3010243 and main protease complex. (b) three-dimensional binding mode of the complex. (c) two-dimensional binding mode of the complex.

physiological salt concentration was also maintained. The RMSD values of the protein backbone with or without ligands indicated stabilization after ∼5 ns of the simulation (Figure 12). These RMSD values showed similar changes in the protein backbones from the starting structures (<0.45 nm). It can also be observed that apo-form of the main protease showed the most deviation in the RMSD values, whereas, the docked complexes showed similarly or lower RMSD values, indicating that the docked ligands stabilize the protease structure to some extent. The RMSD values of the non-hydrogen atoms of each of the ligands were calculated to check if the ligands moved significantly during the simulation (Figure 13). These RMSD values demonstrate that once these ligands stabilized, their position did not change much. It can also be observed that CID3010243 and CID44271958 showed the maximum RMSD values indicating their position changed the most, whereas CID10009410 and CID44271905 showed the minimum RMSD values indicating the least change in their position compared to the other docked complexes. The average fluctuation of the position is calculated by root mean square fluctuation (RMSF) values of each residue to check the mobility or flexibility of the residues of a protein during a simulation. The RMSF values of the main protease with and without ligands are depicted in Figure 14. As can be expected, terminal residues of main

protease, apo form or ligand-bound, showed higher RMSF values, confirming their mobile nature. Compared to apo-protease structure, while bound to CID3010243, residues 49 showed higher mobility, whereas residues 52, 140,169 and 189 showed higher mobility. Similarly, the residue 120 showed higher flexibility while bound to CID44271958 and the residues 186–190 showed higher flexibility while bound to CID44271905.

The difference in the higher RMSF values of various residues may indicate difference in the binding nature of the respective ligands and subsequent influences on the protein structure and dynamics. Thus, although similar RMSF values are observed in Figure 14, there are some residues for which changes can be observed upon binding of different ligands. In order to investigate the overall effect of the ligands on the protease structure, radii of gyration (R_g) of the main protease were calculated for each complex (Figure 15). R_g values of the apo-protease and protease complexed with CID3010243, CID10009410, CID44271905, and CID44271958 are, respectively 2.300 nm, 2.238 nm, 2.232 nm, 2.207 nm, and 2.251 nm. Based on this observation, it can be said that the protease structure becomes the most compact upon biding with CID44271905, as indicated by the lowest R_g value. How different ligand interacts with the main protease is an interesting question, and the number of hydrogen bonds formed

Figure 8. Interaction between CID44271958 (Lopinavir analogue) and main protease (a) CID44271958 and main protease complex; (b) three-dimensional binding mode of the complex; (c) two-dimensional binding mode of the complex.

between the protease and the ligands may indicate the difference in binding pattern of various ligands with the main protease. We plotted the number of hydrogen bonds formed by protease with different ligands during the simulation in Figure 16.

It can be observed that CID44271905 and CID44271958 form more hydrogen bonds compared to CID3010243 and CID10009410. The average number of hydrogen bonds formed during the 50 ns simulation, over 50,000-time points, between the main protease and the ligands, CID3010243, CID10009410, CID44271905, and CID44271958, are respectively 1.10, 1.03, 2.77, and 1.63. Although the docked complexes showed similar binding energies, this analysis shows that number of hydrogen bonds formed in aqueous solution is quite different for all four ligands. This can probably be ascribed to the fact that docking is carried out without any solvent, whereas, the effect of solvent can have a profound effect in ligand binding.

3.5. Principal component analysis (PCA)

PCA analyses were performed on the MD trajectories of the main protease with or without the ligands. Although the total motion of the $C\sigma\upsilon\beta > \dot{\alpha}/\sigma\upsilon\beta >$ atoms were dispersed over 918

eigenvectors, top 10 eigenvectors, sorted by their corresponding eigenvalues, were responsible for more than 70% of the overall collective motions (Figure 17). Combined contributions of the eigenvectors with top two eigenvalues were 57%, 52%, 48%, 40%, and 50% for the protease molecule with the ligands CID3010243, CID10009410, CID44271905 and CID44271958, and the apo protease, respectively. It was also noted that more number of eigenvectors were involved in the overall motions of the protease complexed with the compound CID44271958 than the apo-protease or the other protease complexes. This can probably be explained by enhanced local motions for the protease-CID44271958 complex due to its more extended structure, as was shown by the R_g analysis (Figure 15). Projections of displacement vectors of the C α atoms onto eigenvector 1 and 2 are called principal component (PC) 1 and 2, respectively. PC1 and PC2 were plotted together to visualize the essential subspace of collective motions (Figure 18). It can be observed that the range of the PC1 values are larger for apo-protease (Figure 18(a)), compared to other protease complexes, indicating that the apo-protease is relatively more flexible with respect to the motion depicted by eigenvector 1. It can also be observed that the spread of both PC1 and PC2 of the protease complexed with CID44271905 and CID44271958 are comparatively smaller, indicating more restricted motions of the backbone C α atoms,
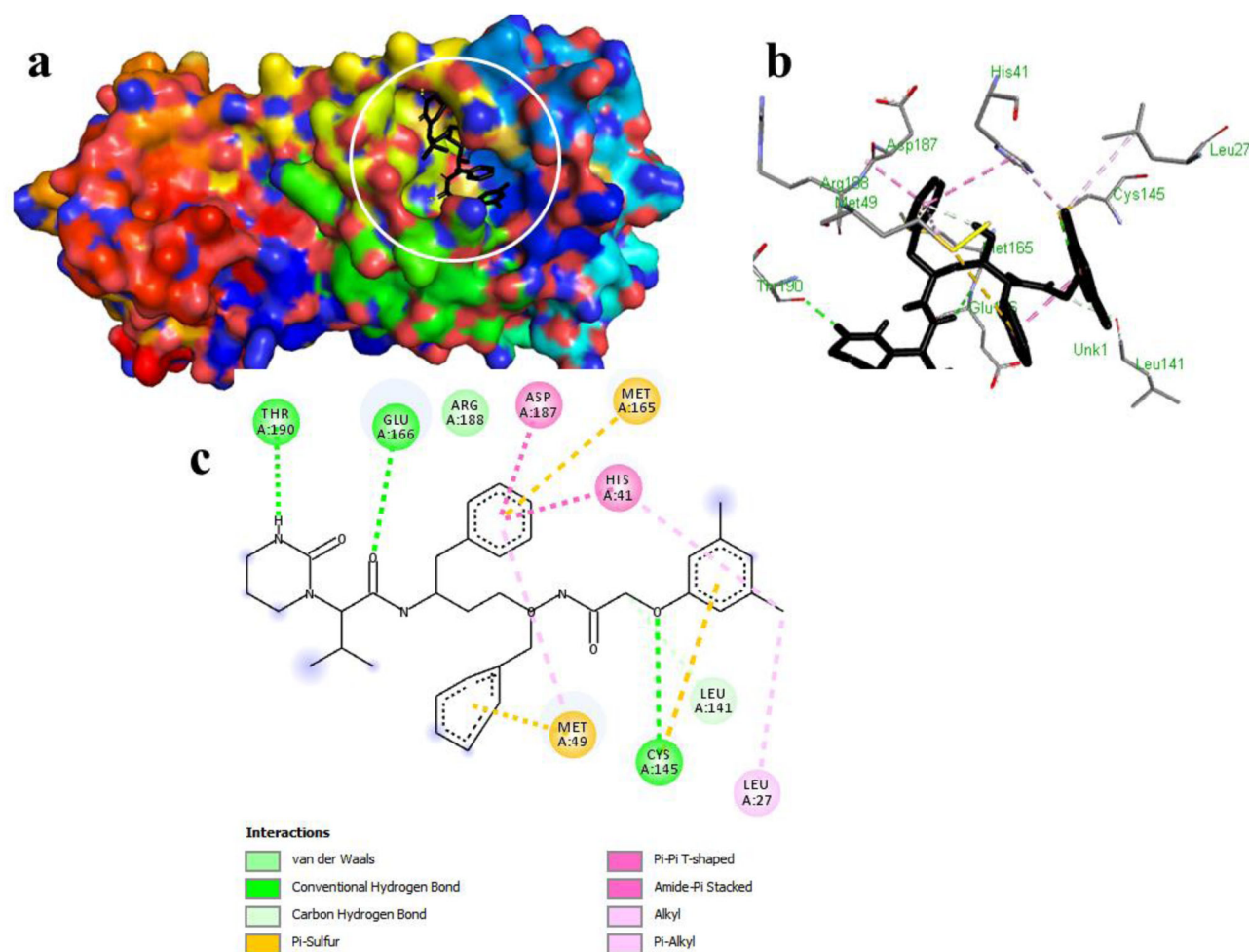
Figure 9. Interaction between CID44271905 (Lopinavir analogue) and main protease (a) CID44271905 and main protease complex; (b) three-dimensional binding mode of the complex; (c) two-dimensional binding mode of the complex.

probably due to more number of intermolecular hydrogen bonds formed between the protein and the ligands. A quantitative analysis of the overall flexibility of the protease molecule without any ligand and with the ligands can be performed using the trace of the diagonalized covariance matrix. The traces were obtained to be 9.08, 10.24, 9.44, 8.55 and 5.48 for the apo-protease and for the protease complexed with CID3010243, CID10009410, CID44271905, and CID44271958, respectively. This indicates decreased flexibility of the protease while complexed with CID44271905 and CID44271958, probably due to more compact structure facilitated by the intermolecular hydrogen bonds. Thus, PCA analyses indicate that the overall motions of the protease complexed with the ligands CID44271905 and CID44271958 are more restricted compared to the apo-protease and other protease complexes studied here.

3.6. Analysis of protein-ligand interaction energy

Beside molecular dynamics simulations, g_mmpbsa tool was also applied to estimate the binding free energies of the protein-ligand systems (Kumari et al., 2014). One of the popular methods to estimate the interaction energies is Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) analysis. This method uses molecular dynamics simulation

trajectories to predict binding free energies (ΔE_{MMPBSA}) of protein-protein, protein-ligand, or protein-DNA systems.

We performed the MMPBSA analysis on the 30–50 ns part of the MD trajectory of each protease-ligand complex at an interval of 50 ps. Total binding energies of the protease complexes with ligands CID3010243, CID10009410, CID44271905, and CID44271958 are plotted in Figure 19. It can be observed that binding free energies of protease compounds with CID3010243, CID10009410, and CID44271958 alter between two different values, owing to differences in the mode of binding.

It appears from the binding free energy analysis that there are two states of ligand binding for each of the ligand and there are interchanges between these two states. However, in the case of the protease-CID44271905, this jump between two states in ΔE_{MMPBSA} is much less frequent, which indicate a more stable binding. This can also be observed in the total values of ΔE_{MMPBSA}, as shown in Table 4. These values are -28.06 ± 17.95 (kcal/mol), -15.44 ± 21.02 (kcal/mol), -41.13 ± 10.91 (kcal/mol), and -0.66 ± 12.85 (kcal/mol) for CID3010243, CID10009410, CID44271905, and CID44271958. It can also be observed that the van der Waal's energy (E_{vdW}), electrostatics energy (E_{elec}), and nonpolar solvation energy ($E_{nonpolar}$) contribute significantly toward stronger binding of CID44271905, as shown by their respective
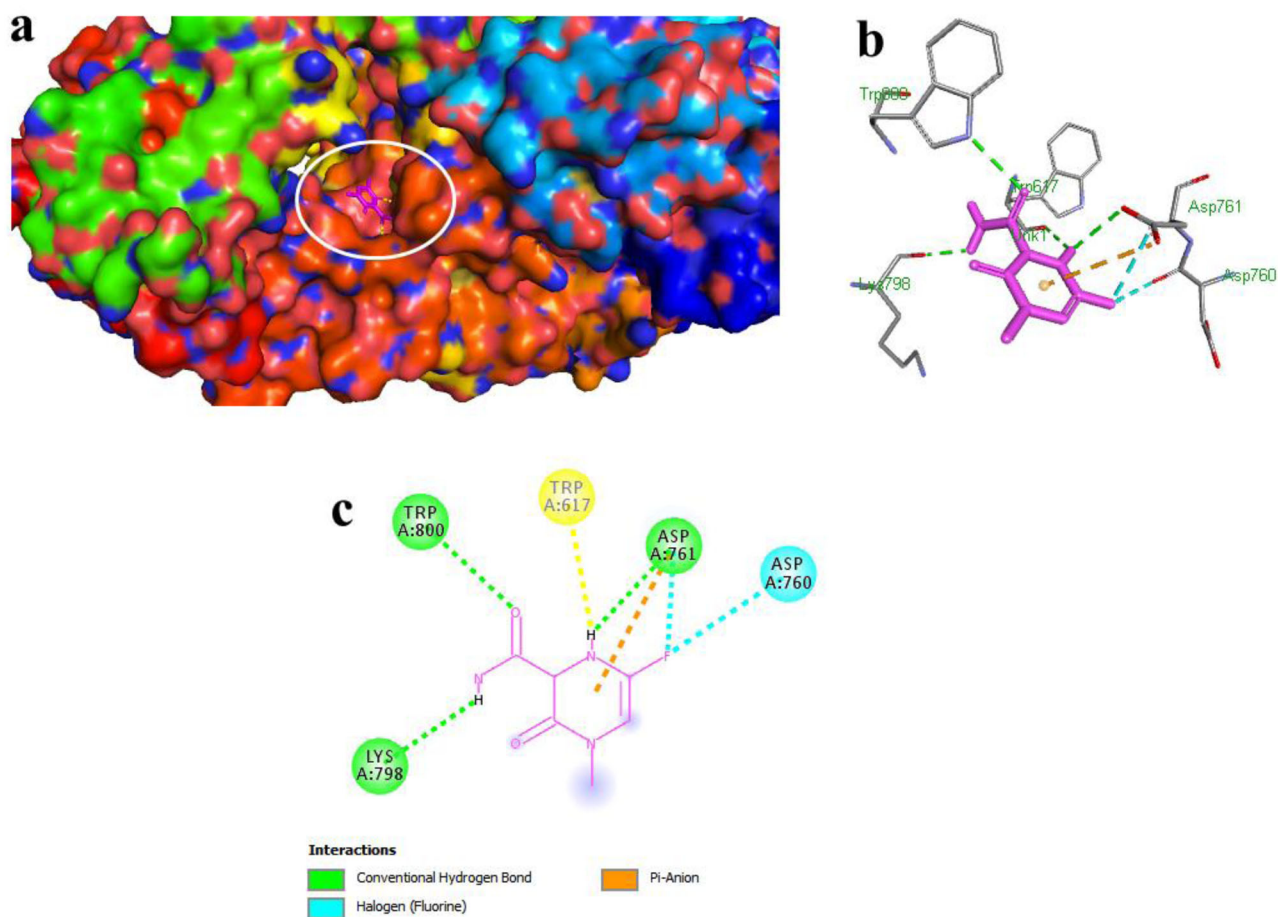
Figure 10. Interaction between CID89869520 (Favipiravir analogue) and RNA dependents RNA polymerase (RdRp) (a) CID89869520 and RdRp complex; (b) three-dimensional binding mode of the complex; (c) two-dimensional binding mode of the complex.

average values of -54.70 ± 13.62 kcal/mol, -12.64 ± 3.89 kcal/mol, and -5.42 ± 1.36 kcal/mol. This also indicates that non-polar interaction energy ($E_{vdW} + E_{nonpolar}$) contributes strongly towards binding, suggesting that hydrophobic interaction between CID44271905 and the protease plays a crucial role. Negative electrostatics energy also supports the observation that intermolecular hydrogen bonds are also favored.

It should also be noted that E_{elec} between the protease and the ligands CID3010243, CID10009410, CID44271905, and CID44271958 are respectively -3.00 ± 3.01 kcal/mol, -0.17 ± 0.52 kcal/mol, -12.64 ± 3.89 kcal/mol, and -0.91 ± 2.73 kcal/mol. This also supports our previous observations from the hydrogen bond analysis (Figure 16), which showed that CID44271905 formed most number of hydrogen bonds among the other ligands. In all four protease-ligand complexes, polar solvation energy opposes the binding, but compensated by van der Waal's energy, electrostatics energy and nonpolar solvation energy. We also wanted to check which residues of main protease contribute significantly towards binding of the CID44271905 (Figure 18). As can be observed, a few residues, such as, residues 47, 49, 50, 165, 167 and 185 contribute strongly toward binding of CID44271905 to main protease. However, we should note that these contributions include a sum total of molecular potential energy, polar solvation energy and nonpolar solvation energy. From the docking analyses (Figure 9), we observed that the residues M49 and

M165 were actively involved in the intermolecular interaction. It should also be noted that residues E166, D187, R188 and T190 were also observed to interact with the CID44271905, where we find the nearby residues to play crucial role by MMPBSA analysis, thus supporting the important role of these residues in binding of the ligand CID44271905. Thus, although MMPBSA analyses indicate that CID3010243, CID10009410, and CID44271905 form stable complexes, CID44271905 form the most stable complex among the lopinavir analogs with the main protease of SARS-CoV-2.

4. Discussion

Computer-aided drug design (CADD) has been an active area of research. Computational compound screening has been extensively used in recent years (Bajorath, 2002; Chien et al., 2002). This study was based on interrelation between drug and SARS-CoV-2 enzymes (M^{pro} and RdRp) and in silico screening, designing, identifying new drugs against COVID-19. Lopinavir is a protease inhibiting antiviral drug and favi-piravir is a viral RNA dependent RNA polymerase inhibiting drug, these two compounds are a potential candidate for the treatment of COVID-19 (Furuta et al., 2017). The new drug candidates were identified by similarity searching the structure of lopinavir (Figure 2). All potential selected drug candidate compounds docked with the active site cleft
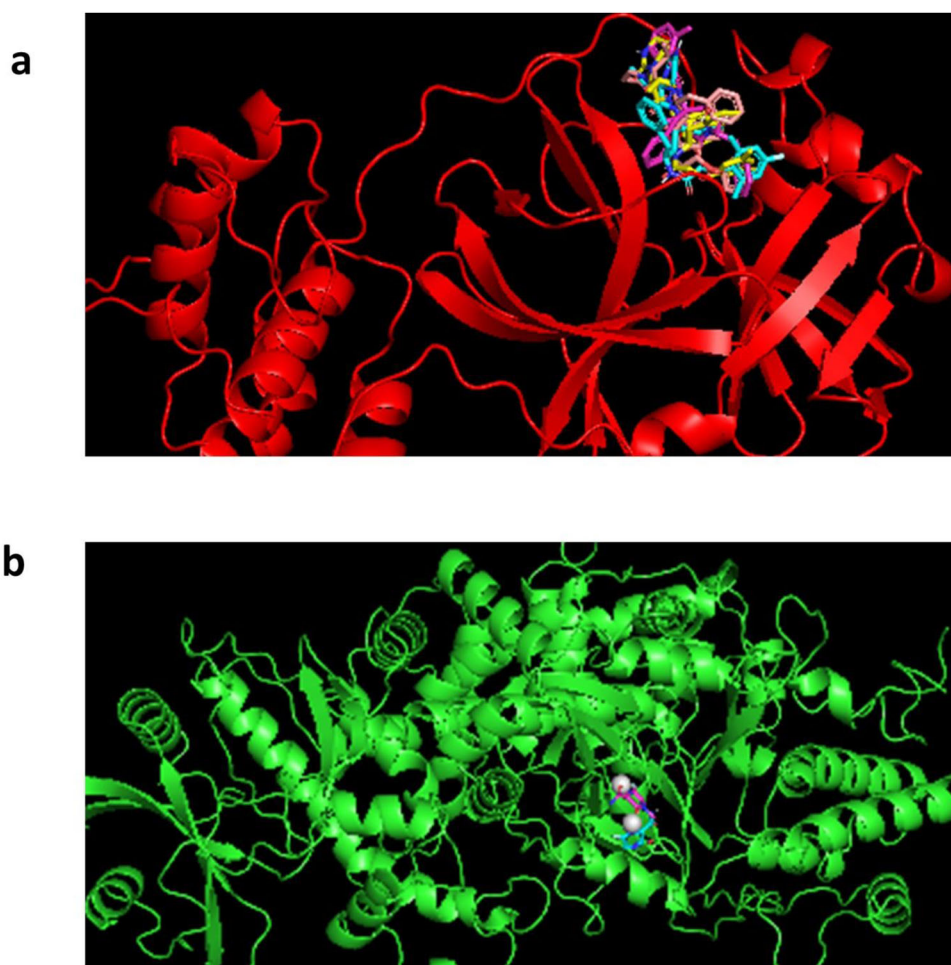
Figure 11. Binding representation of all structurally modified analogues. (a) Four Lopinavir analogues bind with the same position of main protease (M^pro), His41, Met149 and Cys145 were the common binding residues among these drug compounds; (b) one Favipiravir analogue binds the active site residues of RdRp.

amino residues of both targets, Mpro and RdRp. The docking energies displayed that the most desired binding affinity. Among them, lopinavir-d8 and CID59310949 showed the highest binding affinity (−12.5 and −12.4 kcal/mol) with M^pro. Lopinavir-d8 is a labelled selective HIV-1 protease inhibiting drug which is an analogue of ritonavir (Kumar et al., 1999; Sham et al., 1998), this drug may act against COVID-19.

The ADMET properties assess the pharmacokinetics and pharmacology for absorption, distribution, metabolism, excretion, and toxicity that described the proper disposition of drug-like compounds within the body of an organism (Balani et al., 2005). All drugs selected show a moderate to high absorption rate and Caco-2 permeability. Highest absorption rate and Caco-2 permeability found in CID46212828 (65.39%) and 0.93, respectively. None of the drugs shows CYP2D6 inhibitor activity. Among all selected candidate, four drugs show an irritating effect in the field of toxicity (supplementary material Table 1).

For improving drug-like properties of lopinavir and favipiravir, especially drug score, drug-likeness, polarity, and reducing toxicity, the structural modification was done on lopinavir and favipiravir and designed five structural analogues. CID10009410 (lopinavir analogues) and CID89869520 (favipiravir analogues) were modified by adding –F and –CH3

groups at the end, respectively. CID44271905 (lopinavir analogue) and CID44271958 (lopinavir analogue) were modified by removing trimethyl-benzene fragment and adding 1,3,5-trimethyl-benzene and benzene fragments and the structure of CID3010243 were modified by removing tetrahydro-pyrimidionepropylene urea fragment and adding 2-imidazolidone fragment (Figure 3). The pharmacophore analysis revealed that except for two drugs (CID10009410 and CID3010243) there is no toxicity found in designed compounds (supplementary material Table 2). Among these compounds, four candidates show a high drug score and drug-likeness than lopinavir and favipiravir (Figures 4 and 5). The drug score of CID3010243, CID44271958, CID44271905 and CID89869520 is 0.19, 0.36, 0.26, and 0.95, respectively. Some antiviral drugs such as remdesivir, ritonavir and lopinavir drug score are 0.05, 0.13 and 0.17, respectively,

The docking analysis of these compounds determined their favorable binding affinity against to their target proteins. Lopinavir analogues interact with amino acids in the active site of the main protease, and one favipiravir analogue bind with residues in active site of RNA dependent RNA polymerase (Figures 6–10). The interactions between the selected target proteins and ligands were mediated by several hydrogens, hydrophobic, electrostatic bond. The

Table 3. Physicochemical properties, lipophilicity, water-solubility, drug-likeness and medicinal chemistry profile of structurally modified analogues.

| Properties	CID10009410 (Lopinavir analogue)	CID 3010243 (Lopinavir analogue)	CID 44271958 (Lopinavir analogue)	CID 44271905 (Lopinavir analogue)	CID 89869520 (Favipiravir analogue)
Physicochemical properties					
Molecular weight (g/mol)	646.79	614.77	600.75	628.80	171.13
Num. rotatable bonds	17	17	17	17	1
Num. H-bond acceptors	6	5	5	5	4
Num. H-bond donors	4	4	4	4	1
TPSA (A*2)	120	120	120	120	77.98
Lipophilicity					
Consensus Log $P_{o/w}$ [a]	4.73	4.21	3.83	4.55	−0.19
Water solubility					
Log S [a] (water solubility)	ESOL: −6.80 (PS) Ali: −8.31 (PS) SILICOS-IT: −10.31 (IS)	ESOL: −6.33 (PS) Ali: −7.84 (PS) SILICOS-IT: −9.89 (PS)	ESOL: −6.02 (PS) Ali: −7.46 (PS) SILICOS-IT: −9.32 (PS)	ESOL: −6.64 (PS) Ali: −8.21 (PS) SILICOS-IT: −10.05 (IS)	ESOL: −0.97 (VS) Ali: −0.80 (VS) SILICOS-IT: −0.99 (S)
Medicinal chemistry					
PAINS	0 alert	0 alert	0 alert	0 alert	0 alert
Brenk	0 alert	0 alert	0 alert	0 alert	0 alert
Lead-likeness	No; 3 violation, MW > 480, MR > 130, #atom > 70, XLOGP 3 > 3.5	No; 3 violation, MW > 480, MR > 130, XLOGP 3 > 3.5	No; 3 violation, MW > 480, MR > 130, XLOGP 3 > 3.5	No; 3 violation, MW > 350, Rotors > 7, XLOGP 3 > 3.5	No;1 violation, MW < 250
Synthetic accessibility	5.73	5.52	5.35	5.63	2.21
Druglikeness					
Lipinski	No; 1 violation, MW > 500	No; 1 violation, MW > 500	No; 1 violation, MW > 500	No; 1 violation, MW > 500	Yes; 0 violation
Ghose	No; 3 violation, MW > 480, MR > 130, #atom > 70	No; 3 violation, MW > 480, MR > 130, #atom > 70	No; 3 violation, MW > 480, MR > 130, #atom > 70	No; 3 violation, MW > 480, MR > 130, #atom > 70	No; 3 violation, WLOGP<0.4, MR < 40 #atom < 20
Veber	No; 1 violation: Rotors > 10	No; 1 violation: Rotors > 10	No; 1 violation: Rotors > 10	No; 1 violation: Rotors > 10	Yes
Egan	Yes	Yes	Yes	Yes	Yes
Muegge	No; 3 violations: MW > 600, XLOGP3 > 5, Rotors > 15	No; 3 violations: MW > 600, XLOGP3 > 5, Rotors > 15	No; 3 violations: MW > 600, XLOGP3 > 5, Rotors > 15	No; 3 violations: MW > 600, XLOGP3 > 5, Rotors > 15	No; 1 violation: MW < 250
Bioavailability score	0.55	0.55	0.55	0.55	0.55
Drug-likeness	4.67	7.86	7.82	1.76	3.72
Drug score	0.15	0.19	0.36	0.26	0.95

[a]Consensus Log $P_{o/w}$ is the average value of iLOGP, XLOGP3, WLOGP, MLOGP and SILICOS-IT.
PS = poorly soluble, VS = very soluble, S = soluble.
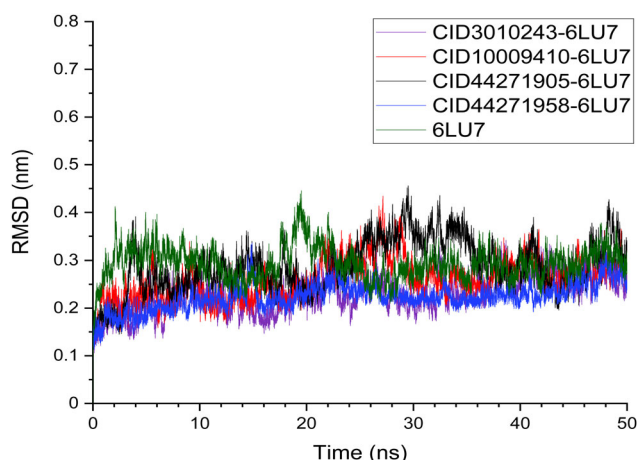
Figure 12. Root mean square deviation (RMSD) of the backbone atoms of the main protease with ligands including CID3010243, CID10009410, CID44271905, and CID 44271958, and without ligand show that each of the five systems was stabilized during the 50 ns MD simulation.
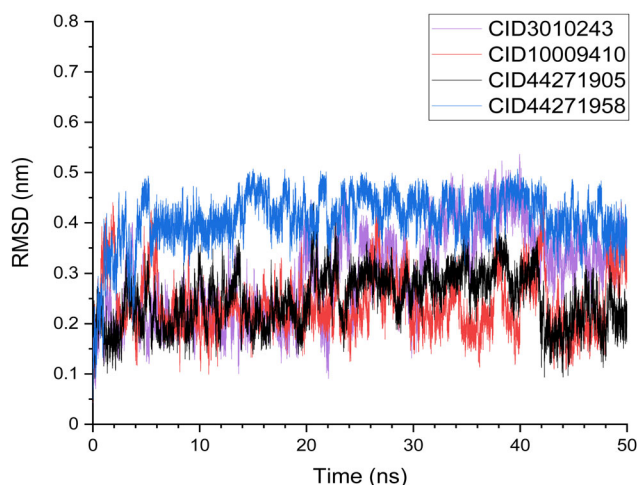
Figure 13. Root mean square deviation (RMSD) of the nonhydrogen atoms of the ligands, CID3010243, CID10009410, CID44271905 and CID44271958, show that the position of the ligands did not fluctuate significantly during the simulation.
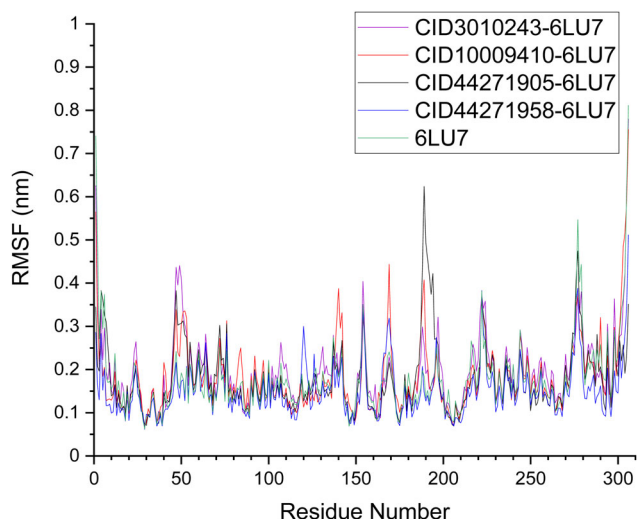

Figure 14. Root mean square fluctuation (RMSF) of the protease residues show that similar residues were involved while interacting with the ligands.
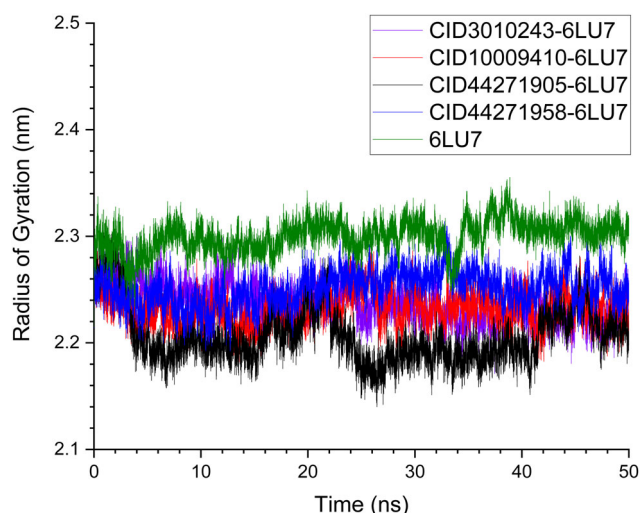

Figure 15. Radius of Gyration (R_g) analysis shows that each ligand induces compactness to the protease structure; however, binding of CID44271905 demonstrated the most compact structure.

interactions of designed drug candidates confirmed the M^{pro} and RdRp modulation. Subsequently, 50 ns molecular dynamic simulations of the top four protein-ligand complexes (Lopinavir analogues) were run in explicit solvent condition. It was observed that all simulations were stable as indicated by the protease backbone (Figure 12) and the distance of the ligand from the starting position did not change significantly after stabilization (Figure 13). It was also observed that the ligands CID10009410 and CID44271905 showed the least displacement from their starting structures (Figure 13), which indicates that the docked positions were also more stable for them in the aqueous environment compared to the other two lopinavir analogues. The RMSF analysis (Figure 14) showed similar characteristics of all the complexes, supporting similar nature of the ligands and their physiochemical properties. In the meantime, radius of gyration (R_g) analysis (Figure 15) revealed that CID44271905 induced the most compact conformation of the main protease than the other ligands, hinting a possible stronger binding compared to other complexes. Hydrogen bond analyses also revealed that this lopinavir analogue, CID44271905, on average of the 50,000 trajectory points each separated by 1 ps, formed the greatest number of hydrogen bonds with the main protease (Figure 16). Principal component analyses (PCA) also indicated that CID44271905 and CID44271958 relatively restrict the overall collective motion of the protease compared to the apo-protease (Figure 18). We also analyzed the relative binding energies (ΔE_{MMPBSA}) by MMPBSA analysis (Figures 19 and 20 and Table 4), which also showed that CID44271905 formed the most stable complex with the protease ($\Delta E_{MMPBSA} = -41.13 \pm 10.91$ kcal/mol), compared to the other complexes of the protease and CID3010243 ($\Delta E_{MMPBSA} = -28.06 \pm 17.95$ kcal/mol), CID10009410 ($\Delta E_{MMPBSA} = -15.44 \pm 21.02$ kcal/mol) and CID44271958 ($\Delta E_{MMPBSA} = -0.66 \pm 12.85$ kcal/mol). It should also be noted that apart from CID44271905, both CID3010243 and CID10009410 also form a stable binding, and can also be taken forward for
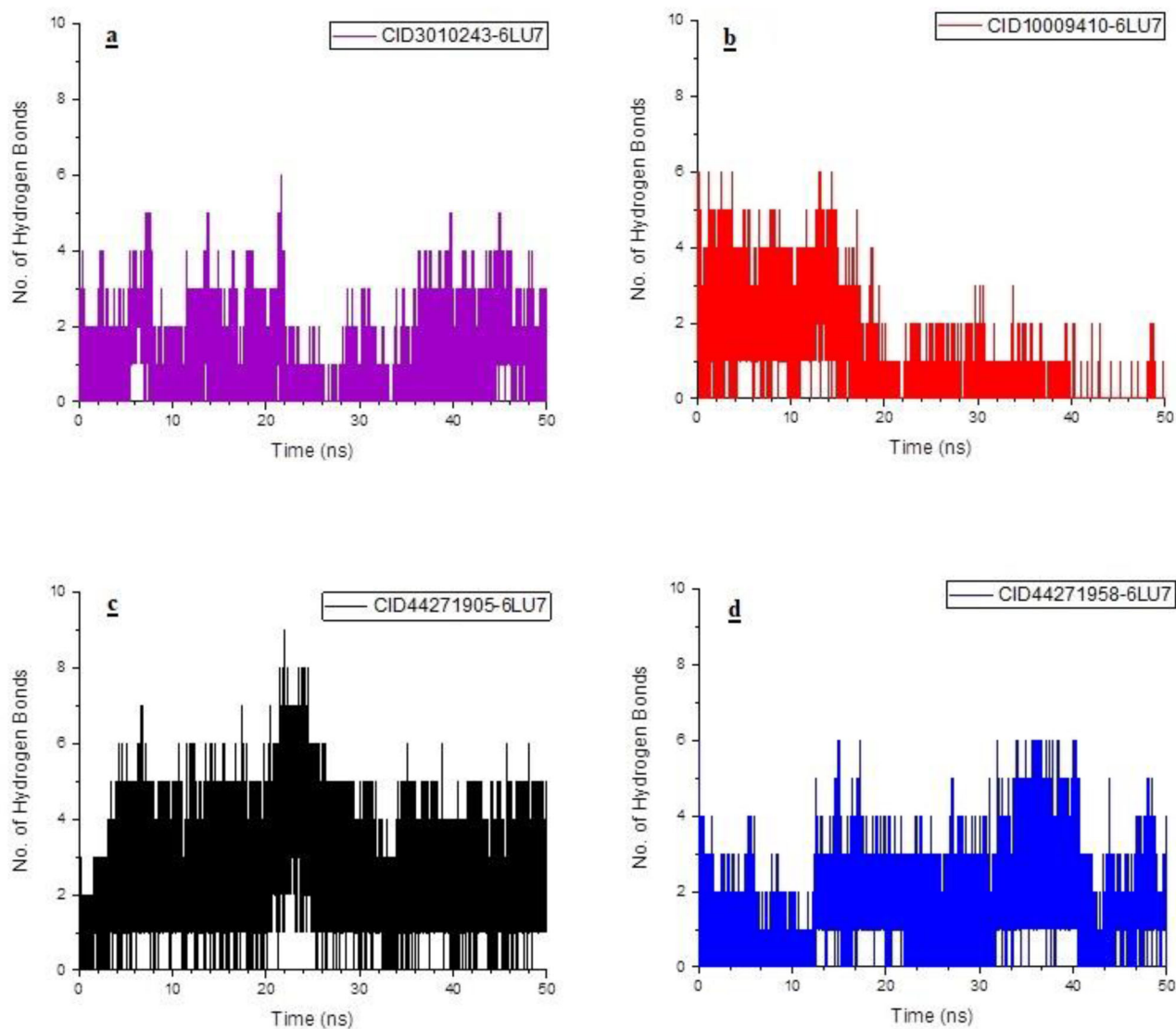
Figure 16. Number of hydrogen bonds formed during the course of simulation by the main protease with the ligands (a) CID3010243; (b) CID10009410; (c) CID44271905 and (d) CID44271958 indicates that CID44271905 forms the most stable structure.
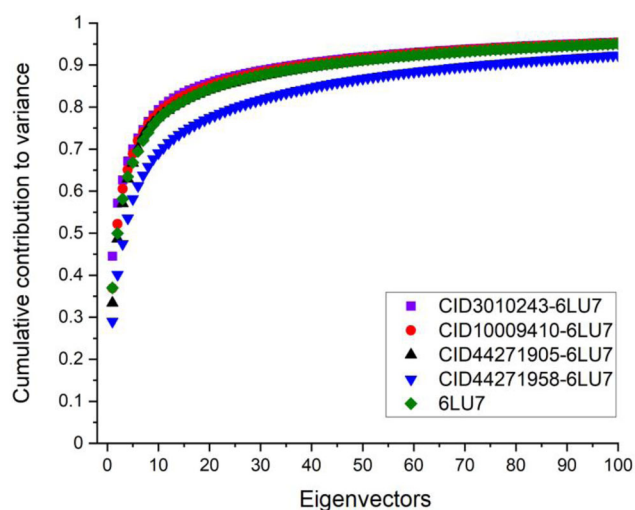

Figure 17. Cumulative contribution of the 100 most relevant eigenvectors to the variance of the overall motion of the protease without ligand and with ligands CID3010243, CID10009410, CID44271905, and CID44271958.

further experimental studies. It was also observed that non-polar interaction energy, i.e. sum of van der Waal's energy and nonpolar solvation energy, contribute strongly towards the binding of all the four ligands and the protease, indicating hydrophobic forces favor the formation of the complexes. Although docking analyses revealed that the four complexes of the viral main protease with CID3010243, CID10009410, CID44271905, and CID44271958, showed similar binding energy values, binding free energy studies by molecular dynamics simulations and MMPBSA analysis revealed that their binding differs in aqueous solution. Although, we have identified three structural analogues of lopinavir (CID3010243, CID10009410 and CID44271905) as potentially strong drug candidates, our computational analyses strongly indicate that CID44271905 showed the most promise to bind to the main protease of SARS-CoV-2, and we recommend if the binding of these compounds and their physiological efficacy can be further tested by *in vitro* and *in vivo* experiments.
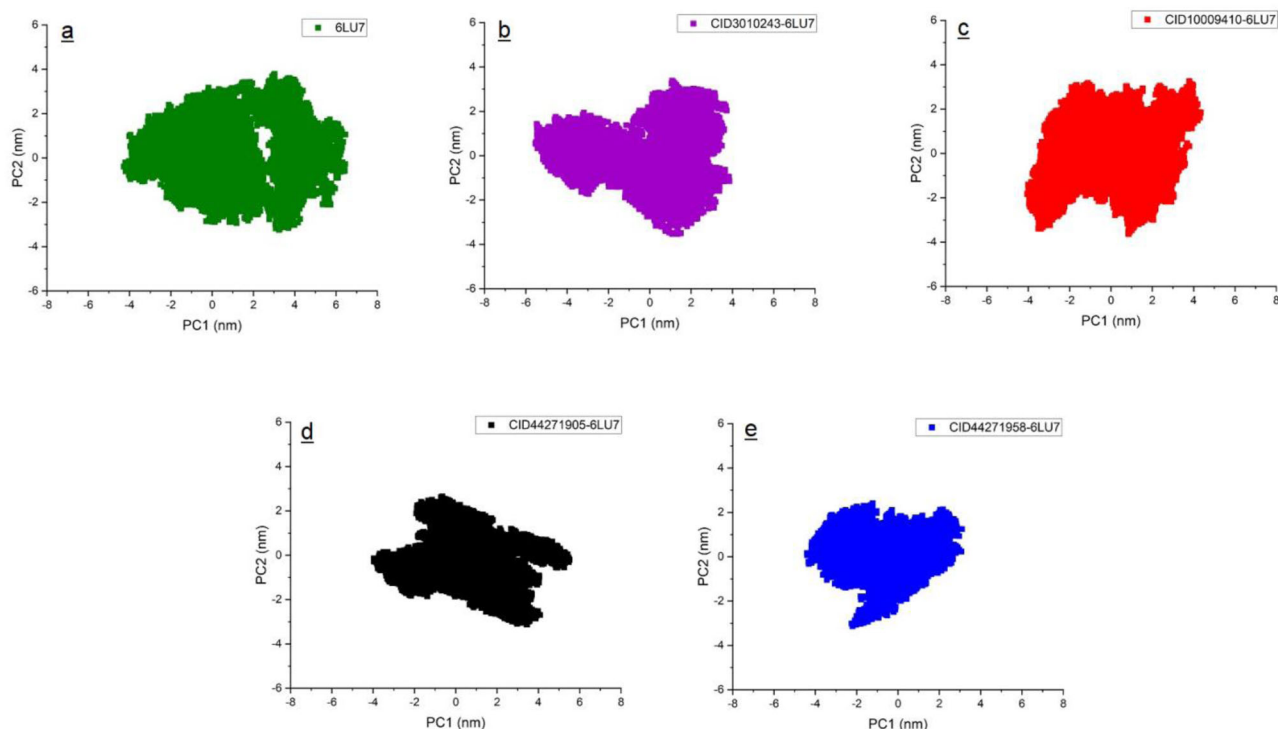
Figure 18. PCA scatter plots show the projections of the displacement of C_a atoms along the first eigenvector and the second eigenvector at each time point for (a) the apo-protease, and protease with (b) CID3010243; (c) CID10009410; (d) CID44271905, and (e) CID44271958 respectively.
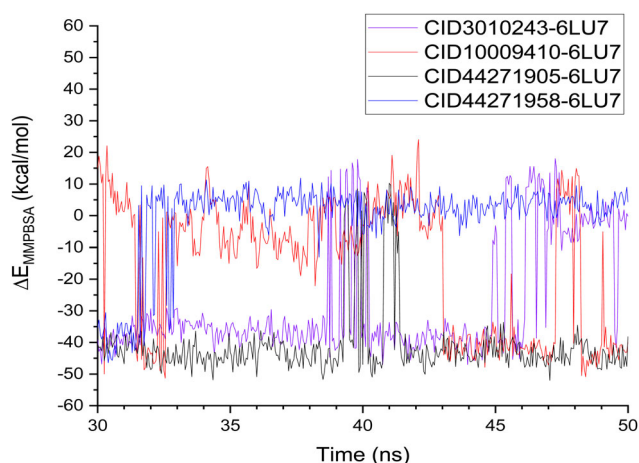
Figure 19. Calculation of MMPBSA energy (ΔE_{MMPBSA}) of the main protease and the ligands, CID3010243, CID10009410, CID44271905 and CID 44271958 show that the lowest binding energy was observed for CID44271905.
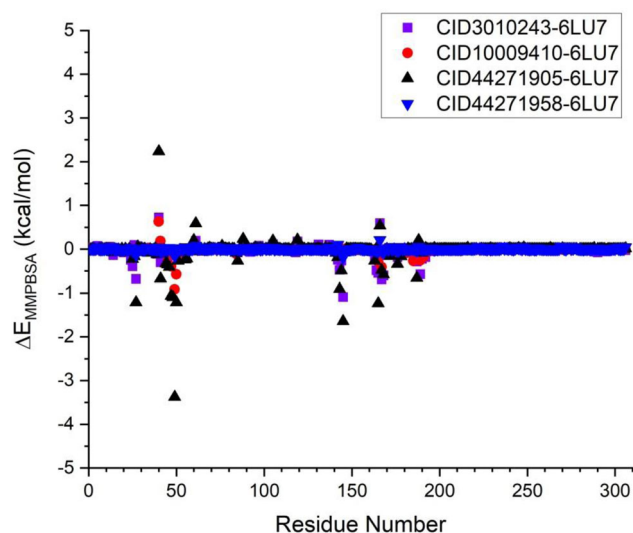
Figure 20. Residue wise decomposition of binding free energies obtained from theMMPBSA analyses show similar residues are involved in binding the ligands .

Table 4. MMPBSA energy (ΔE_{MMPBSA}) of binding for the protease complexed with ligands CID3010243, CID10009410, CID44271905, and CID44271958.

Criteria	CID3010243	CID10009410	CID44271905	CID44271958
Van der Waal energy (kcal/mol)	−40.82 ± 23.07	−19.85 ± 27.20	−54.70 ± 13.62	−5.40 ± 16.01
Electrostatic energy (kcal/mol)	−3.00 ± 3.01	−0.17 ± 0.52	−12.64 ± 3.89	−0.91 ± 2.73
Polar solvation energy (kcal/mol)	19.70 ± 10.87	6.59 ± 12.55	31.63 ± 8.26	6.14 ± 8.47
Nonpolar solvation energy (kcal/mol)	−3.94 ± 2.20	−2.02 ± 2.76	−5.42 ± 1.36	−0.49 ± 1.57
ΔE_{MMPBSA} (kcal/mol)	−28.06 ± 17.95	−15.44 ± 21.02	−41.13 ± 10.91	−0.66 ± 12.85

5. Conclusions

In the present study, we used bioinformatics and computational studies to identify, design, and propose five oral-based novel anti-viral inhibitors for the treatment of COVID-19. The pharmacophore assessment suggests the potential drug-like properties of these compounds. Additionally, we propose more studies includes synthesis and biological activities due to these compounds (CID3310243, CID10009410, and CID44271905) are structurally much more efficient.

Acknowledgment

Thanks to Prof. Dr. Esin Akı Yalcin and the research group for technical support.

Disclosure statement

The authors declare that they have no conflict of interest.

Competing interests

The authors declare that they have no competing interests.

Ethical approval

This is an *in silico* study and does not involve any human, animal, plant or clinical experiments. Thus, does not require ethical approval.

ORCID

Md. Oliullah Rafi ⓘ http://orcid.org/0000-0002-1707-655X
Tugba Taskin-Tok ⓘ http://orcid.org/0000-0002-0064-8400
Md. Shahedur Rahman ⓘ http://orcid.org/0000-0002-1707-655X

Data availability statement

The authors declare that all the data supporting the findings of this study are available within the article.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, *1-2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Accelrys Software Inc., Discovery Studio 3.5, San Diego 2013.

Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery*, *1*(11), 882–894. https://doi.org/10.1038/nrd941

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences United States of America*, *98*(18), 10037–10041. https://doi.org/10.1073/pnas.181342398

Balani, S. K., Miwa, G. T., Gan, L.-S., Wu, J.-T., & Lee, F. W. (2005). Strategy of utilizing in vitro and in vivo ADME tools for lead optimization and drug candidate selection. *Current Topics in Medicinal Chemistry*, *5*(11), 1033–1038. https://doi.org/10.2174/1568026605774297038

Banerjee, P., Eckert, A. O., Schrey, A. K., & Preissner, R. (2018). ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Research*, *46*(W1), W257–W263. https://doi.org/10.1093/nar/gky318

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C.-Y., Poon, R. W.-S., Tsoi, H.-W., Lo, S. K.-F., Chan, K.-H., Poon, V. K.-M., Chan, W.-M., Ip, J. D., Cai, J.-P., Cheng, V. C.-C., Chen, H., Hui, C. K.-M., & Yuen, K.-Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. The *Lancet*, *395*(10223), 514–523. https://doi.org/10.1016/S0140-6736(20)30154-9

Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., Lee, P. W., & Tang, Y. (2019). Correction to "admetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties". *Journal of Chemical Information and Modeling*, *59*(11), 4959. https://doi.org/10.1021/acs.jcim.9b00969

Cheng, T., Pan, Y., Hao, M., Wang, Y., & Bryant, S. H. (2014). PubChem applications in drug discovery: A bibliometric analysis. *Drug Discovery Today*, *19*(11), 1751–1756. https://doi.org/10.1016/j.drudis.2014.08.008

Chien, A., Foster, I., & Goddette, D. (2002). Grid technologies empowering drug discovery. *Drug Discov Today*, *7*(20 Suppl), S176–S180. https://doi.org/10.1016/s1359-6446(02)02369-3

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*, 42717. https://doi.org/10.1038/srep42717

Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. In *Chemical biology* (pp. 243–250). Berlin, Germany: Springer.

de Leuw, P., & Stephan, C. (2017). Protease inhibitors for the treatment of hepatitis C virus infection. *GMS Infectious Diseases*, *5*, 1-14.

Elfiky, A. A. (2020). Ribavirin, remdesivir, sofosbuvir, galidesivir, and tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life Sciences*, *253*, 117592. https://doi.org/10.1016/j.lfs.2020.117592

Ezat, A. A., Elfiky, A. A., Elshemey, W. M., & Saleh, N. A. (2019). Novel inhibitors against wild-type and mutated HCV NS3 serine protease: An in silico study. Virus Disease, *30*(2), 207–213. https://doi.org/10.1007/s13337-019-00516-7

Furuta, Y., Gowen, B. B., Takahashi, K., Shiraki, K., Smee, D. F., & Barnard, D. L. (2013). Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral Research*, *100*(2), 446–454. https://doi.org/10.1016/j.antiviral.2013.09.015

Furuta, Y., Komeno, T., & Nakamura, T. (2017). Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proceedings of the Japan Academy, Series B*, *93*(7), 449–463. https://doi.org/10.2183/pjab.93.027

Harrison, C. (2020). Coronavirus puts drug repurposing on the fast track. *Nature Biotechnology*, *38*(4), 379–381. https://doi.org/10.1038/d41587-020-00003-1

Kalia, M., Singh, P. K., Yadav, V. K., Yadav, B. S., Sharma, D., Narvi, S. S., Mani, A., & Agarwal, V. (2017). Structure based virtual screening for identification of potential quorum sensing inhibitors against LasR master regulator in *Pseudomonas aeruginosa*. *Microbial Pathogenesis*, *107*, 136–143. https://doi.org/10.1016/j.micpath.2017.03.026

Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S., & Soetjipto, S. (2020). Potential inhibitor of COVID-19 main protease (Mpro) from several medicinal plant compounds by molecular docking study. *Preprints*, 20944, 1–14. https://doi.org/10.20944/preprints202003.0226.v1

Kumar, G. N., Jayanti, V., Lee, R. D., Whittern, D. N., Uchic, J., Thomas, S., Johnson, P., Grabowski, B., Sham, H., Betebenner, D., Kempf, D. J., Denissen, J. F. (1999). In vitro metabolism of the HIV-1 protease inhibitor ABT-378: Species comparison and metabolite identification. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, *27*(1), 86–91.

Kumari, R., Kumar, R., & Lynn, A., Open Source Drug Discovery Consortium. (2014). g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*(7), 1951–1962. https://doi.org/10.1021/ci500020m

Li, G., & De Clercq, E. (2020). Therapeutic options for the 2019 novel coronavirus (2019-nCoV). In. *Nature Reviews. Drug Discovery*, *19*(3), 149–150. https://doi.org/10.1038/d41573-020-00016-0

Li, Z., Wu, M., Guo, J., Yao, J., Liao, X., Song, S., Han, M., Li, J., Duan, G., Zhou, S., Wu, X., Zhou, Z., Wang, T., Hu, M., Chen, X., Fu, Y., Lei, C., Dong, H., Xu, C., Hu, Y., Han, M., … Yan, J. (2020). Caution on kidney dysfunctions of 2019-nCoV patients. *MedRxiv*. https://doi.org/10.1101/2020.02.08.20021212

Lill, M. A., & Danielson, M. L. (2011). Computer-aided drug design platform using PyMOL. *Journal of Computer-Aided Molecular Design*, *25*(1), 13–19. https://doi.org/10.1007/s10822-010-9395-8

Lionta, E., Spyrou, G., K., Vassilatis, D., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: Principles, applications and recent advances. *Current Topics in Medicinal Chemistry*, *14*(16), 1923–1938. https://doi.org/10.2174/1568026614666140929124445

Martinez, L., Andrade, R., Birgin, E. G., & Mart\'\Inez, J. M. (2009). PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, *30*(13), 2157–2164. https://doi.org/10.1002/jcc.21224.

Park, D. S., Kim, J. M., Lee, Y. B., & Ahn, C. H. (2008). QSID Tool: A new three-dimensional QSAR environmental tool. *Journal of Computer-Aided Molecular Design*, *22*(12), 873–883. https://doi.org/10.1007/s10822-008-9219-2

Pires, D. E. V., Blundell, T. L., & Ascher, D. B. (2015). pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of Medicinal Chemistry*, *58*(9), 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104

Porcheddu, R., Serra, C., Kelvin, D., Kelvin, N., & Rubino, S. (2020). Similarity in case fatality rates (CFR) of COVID-19/SARS-COV-2 in Italy and China. *The Journal of Infection in Developing Countries*, *14*(02), 125–128. https://doi.org/10.3855/jidc.12600

Sander, T. (2001). OSIRIS property explorer. *Organic Chemistry Portal.* https://www.organic chemistry.org/prog/peo/

Samad, A., Ahammad, F., Nain, Z., Alam, R., Imon, R. R., Hasan, M., & Rahman, M. S. (2020). Designing a multi-epitope vaccine against SARS-CoV-2: An immunoinformatics approach. *Journal of Biomolecular Structure and Dynamics*, 1–17.

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., & van Gunsteren, W. F. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*, *40*(7), 843–856. https://doi.org/10.1007/s00249-011-0700-9

Schüttelkopf, A. W., & van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(Pt 8), 1355–1363. https://doi.org/10.1107/S0907444904011679

Sham, H. L., Kempf, D. J., Molla, A., Marsh, K. C., Kumar, G. N., Chen, C.-M., Kati, W., Stewart, K., Lal, R., Hsu, A., Betebenner, D., Korneyeva, M., Vasavanonda, S., McDonald, E., Saldivar, A., Wideburg, N., Chen, X., Niu, P., Park, C., … Norbeck, D. W. (1998). ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrobial Agents and Chemotherapy*, *42*(12), 3218–3224. https://doi.org/10.1128/AAC.42.12.3218

Suh, M.-E., Park, S.-Y., & Lee, H.-J. (2002). Comparison of QSAR methods (CoMFA, CoMSIA, HQSAR) of anticancer 1-N-substituted imidazoquino-line-4, 9-dione derivatives. *Bulletin-Korean Chemical Society*, *23*(3), 417–422.

Tözsér, J. (2010). Comparative studies on retroviral proteases: Substrate specificity. Viruses, *2*(1), 147–165. https://doi.org/10.3390/v2010147

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W., & Xiao, G. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Research*, *30*(3), 269–271. https://doi.org/10.1038/s41422-020-0282-0

Yang, X., Yu, Y., Xu, J., Shu, H., Xia, J., Liu, H., Wu, Y., Zhang, L., Yu, Z., Fang, M., Yu, T., Wang, Y., Pan, S., Zou, X., Yuan, S., & Shang, Y. (2020). Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retro-spective, observational study. *The Lancet. Respiratory Medicine*, *8*(5), 475–481. https://doi.org/10.1016/S2213-2600(20)30079-5

Zhang, Z.-J., Wu, W.-Y., Hou, J.-J., Zhang, L.-L., Li, F.-F., Gao, L., Wu, X.-D., Shi, J.-Y., Zhang, R., Long, H.-L., Lei, M., Wu, W.-Y., Guo, D.-A., Chen, K.-X., Hofmann, L. A., & Ci, Z.-H. (2020). Active constituents and mecha-nisms of Respiratory Detox Shot, a traditional Chinese medicine pre-scription, for COVID-19 control and prevention: Network-molecular docking-LC-MSE analysis. *Journal of Integrative Medicine*, *18*(3), 229–241. https://doi.org/10.1016/j.joim.2020.03.004

Zhao, J., Li, K., Wohlford-Lenane, C., Agnihothram, S. S., Fett, C., Zhao, J., Gale, M. J. J., Baric, R. S., Enjuanes, L., Gallagher, T., McCray, P. B. J., & Perlman, S. (2014). Rapid generation of a mouse model for Middle East respiratory syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(13), 4970–4975. https://doi.org/10.1073/pnas.1323279111

Zumla, A., Chan, J. F. W., Azhar, E. I., Hui, D. S. C., & Yuen, K.-Y. (2016). Coronaviruses - drug discovery and therapeutic options. *Nature Reviews. Drug Discovery*, *15*(5), 327–347. https://doi.org/10.1038/nrd.2015.37