

## **Pharmacoinformatics and molecular dynamic simulation studies reveal potential inhibitors of SARS-CoV-2 main protease 3CL<sup>pro</sup>**

Mubarak A. Alamri<sup>a\*</sup>, Muhammad Tahir ul Qamar<sup>b</sup>, Safar M. Alqahtani<sup>a</sup>

*<sup>a</sup>Department of Pharmaceutical Chemistry, College of Pharmacy, Prince Sattam Bin Abdulaziz University, P.O. Box 11323, Alkarj, Saudi Arabia.*

*<sup>b</sup>College of Life Science and Technology, Guangxi University, Nanning 530004, P. R. China*

\*Corresponding author: [m.alamri@psau.edu.sa](mailto:m.alamri@psau.edu.sa)

## Abstract

The SARS-CoV-2 was confirmed to cause the regional outbreak of coronavirus disease 2019 (COVID-19) in Wuhan, China. The 3C-like protease (3CL<sup>pro</sup>), an essential enzyme for viral replication, is a valid target to compact SARS-CoV and MERS-CoV. In this research, an integrated library consisting of 1000 compounds from Asinex Focused Covalent (AFCL) library and 16 FDA-approved protease inhibitors were screened against SARS-CoV-2 3CL<sup>pro</sup>. Top compounds with significant docking scores and making stable interactions with catalytic dyad residues were obtained. The screening results in identification of compound 621 from AFCL library as well as Paritaprevir and Simeprevir from FDA-approved protease inhibitors as potential inhibitors of SARS-CoV-2 3CL<sup>pro</sup>. The mechanism and dynamic stability of binding between the identified compounds and SARS-CoV-2 3CL<sup>pro</sup> were characterized using 50 nanoseconds (ns) molecular dynamic (MD) simulation approach. The identified compounds are potential inhibitors worthy of further development as SARS-CoV-2 3CL<sup>pro</sup> inhibitors/drugs. Importantly, the identified FDA-approved therapeutics could be ready for clinical trials to treat infected patients and help to curb the COVID-19.

**Keywords:** SARS-CoV-2; COVID-19; 3CL protease; Molecular docking; Molecular dynamics and simulations

## Introduction

In last 2 decades, several pathogens spilled over and causes outbreaks. Among them, emergence and reemergence of coronavirus related epidemics widely spread fatal respiratory illness <sup>1</sup>. Coronaviruses are enveloped RNA viruses that are distributed broadly among humans, other mammals, and birds and that cause respiratory, enteric, hepatic, and neurologic diseases <sup>2</sup>. Six coronavirus species are known to cause human disease. Four viruses 229E, OC43, NL63, and HKU1 are prevalent and typically cause common cold symptoms in immunocompetent individuals. The two other strains severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) are zoonotic in origin and have been linked to sometimes fatal illness <sup>3,4</sup>. The SARS-CoV emerged from Guangdong, China in 2003 and affected over 8000 individuals with 774 associated deaths. The MERS-CoV emerged from Saudi Arabia was first reported in 2012 with global mortality rate of 35% (WHO: <https://www.who.int/>).

On December 12, 2019, Wuhan Municipal Health Commission (WMHC), China reported 27 cases of viral pneumonia with 7 of them being critically ill. All of them had history of exposure linked to the Huanan Seafood Wholesale Market where also sold poultry, bats, and snakes <sup>5</sup>. The viral pneumonia outbreak was not caused by SARS-CoV, MERS-CoV, influenza virus, or adenovirus as determined by laboratory tests <sup>6</sup>. A new strain of coronavirus come to limelight <sup>7,8</sup>. On December 30th, The World Health Organization (WHO) temporarily named the viral pneumonia causing pathogen 2019 novel coronavirus (2019-nCoV). On January 30th, 2020, the WHO announced a Public Health Emergency of International Concern (PHEIC) for the 2019-nCoV outbreak. On the Feb 12th, 2020, the WHO permanently named the 2019-nCoV pathogen SARS-CoV-2 and the causing disease into coronavirus disease 2019 (COVID-2019). By February

19<sup>th</sup>, 2020, the death toll reached 2009, with 74,284 laboratory-confirmed cases and 5248 suspected cases. The SARS-CoV-2 also widely spread to over 20 different countries.

Recent studies showed that the SARS-CoV-2 belongs to the beta-corona-virus family and it is closely related to SARS-CoV coronavirus<sup>9</sup>. Similar to other beta-corona-viruses, SARS-CoV-2 produces an 800-kDa polypeptide upon transcription of its genome<sup>5</sup>. This polypeptide is proteolytically cleaved to generate various proteins<sup>5,9</sup>. The proteolytic processing is mediated by papain-like protease (PL<sup>pro</sup>) and 3-chymotrypsin-like protease (3CL<sup>pro</sup>). 3CL<sup>pro</sup> cleaves the polyprotein at 11 distinct sites to generate many of the non-structural proteins which are important in viral replication. Thus, this protease plays a critical role in replication of virus<sup>10,11</sup>. Structure-based activity studies and various high-throughput studies have identified distinct inhibitors of SARS-CoV and MERS-CoV 3CL<sup>pro</sup>. Thus, it is essential to identify novel inhibitors of SARS-CoV-2 3CL<sup>pro</sup>.

Traditional methods for identification of inhibitors are expensive and time consuming. Therefore, the use of *in silico* techniques for identification of inhibitors has gained importance in recent years<sup>12,13</sup>. The available small molecule database could be utilised for either ligand-based or structure based molecular modelling and effective identification of inhibitors<sup>14</sup>. Moreover, 3CL<sup>pro</sup> is highly conserved across coronaviruses, therefore, there is a potential target for identification of compounds that could have broad spectrum anti-viral activity<sup>14</sup>. In this contribution, a combined virtual screening approaches, molecular docking and molecular dynamic simulation were utilized to explore potential inhibitors of SARS-CoV-2 3CL<sup>pro</sup> enzyme as anti-SARS-CoV-2 drugs.

## Materials and Methods

### *Sequence and Structural Alignment Analysis*

A multiple sequence and structure alignment analysis was carried out to find out the evolutionary conserved functional residues among SARS-CoV-2, SARS-CoV and MERS-CoV which could be further targeted as probable targets for the discovery of drug hits. Sequence and 3D structures of SARS-CoV-2 (PDB ID: 6LU7), SRAS-CoV (PDB ID: 2A5I) and MERS-CoV (PDB ID: 5WKK) 3CL<sup>pro</sup> were retrieved from protein data bank (PDB). The 3CL<sup>pro</sup> sequences were aligned using Mega v6.0<sup>15</sup>. To ensure broad spectrum relevance of these protein targets, conserved functional residues recognition within active pockets were analyzed through structural alignment as well. Structural alignment/superposition analysis was done using PyMOL tool<sup>16</sup>.

### *Chemical libraries preparation*

Two chemical libraries were obtained; the commercially available Asinex Focused Covalent (AFCL) library, which consist of 1000 molecules, was retrieved from (<http://www.asinex.com/>) and FDA-approved protease inhibitors including 16 anti-HIV and anti-Hepatitis C antiviral agents were downloaded individually from Pubchem (<https://pubchem.ncbi.nlm.nih.gov/>) in SDF format. Discovery studio visualizer<sup>17</sup> was used to combined both libraries in one SDF file.

### *Structure-based virtual screening*

The chemical libraries were screened against the 3CL<sup>pro</sup> active site within SARS-CoV-2 structure (PDB ID: 6LU7) using Autodock vina in PyRx program<sup>18</sup>. The chemical compounds were initially imported into OpenBable tools in PyRx for energy minimization<sup>19</sup>. The latter, was used also to convert the compounds' SDF files into PDBQT files. The grid box which represent the docking search area, was set to cover the active site of 3CL<sup>pro</sup>. Compounds were ranked based on their

docking scores in Kcal/mole. Autodock tools 1.5.6 program<sup>20</sup> was used to convert the PyRx output files into PDB files. The molecular interactions and binding modes of top poses were determined using Discovery studio Visualizer<sup>17</sup> and Pymol programs<sup>16</sup>.

### ***Molecular docking***

The docking was performed for candidate compounds against the SARS-CoV-2 3CL<sup>pro</sup> structure using Autodock vina 1.1.2 program<sup>21</sup>. Discovery studio Visualizer was used initially to prepare the protein PDB file. Autodock tool 1.5.6 was used to add the polar hydrogens to the protein structure and to convert the PDB files into PDBQT. The same program was used also to obtain the three-dimensional grid box for docking simulation in which the box with size of 24x22x26 was centered using the following dimension; -1.549 x 2.454 x 7.117 to cover the active site along with the essential residues within the binding pocket. Discovery studio Visualizer and Pymol 1.3. programs were used for data analysis.

The predicted inhibitory constant (pK<sub>i</sub>) was calculated using the following equation<sup>22-24</sup> :

$$pK_i = 10^{[\text{Binding energy score} / 1.366]}$$

### ***Molecular dynamic (MD) simulation***

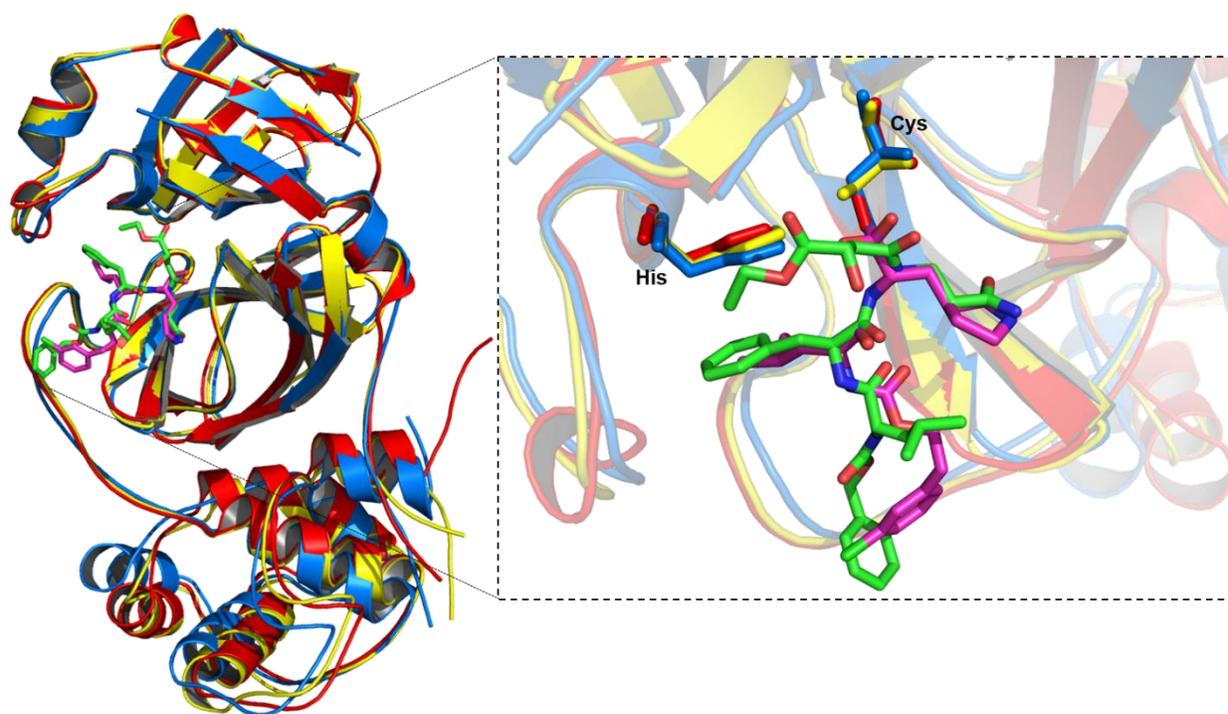
The structure of SARS-CoV-2 3CL<sup>pro</sup> and candidate molecules were prepared for MD simulation using Chimera 1.14<sup>25</sup>. The MD simulation of 3CL<sup>pro</sup>-inhibitor complexes were carried out at 50 ns using Gromacs 2018.1 package<sup>26</sup> using the OPLS-AA/L force field. The parameters of candidate inhibitors were generated by Swissparam online server (<http://www.swissparam.ch/>)<sup>27</sup>. The simulation started by solvating the 3CL<sup>pro</sup>-inhibitor complexes in triclinic box using TIP3P water model. The counter ions were added to the neutralized the system. Periodic boundary conditions were used. The system was energy minimized using a steepest decent algorithm with a

maximum step size of 0.01nm and tolerance of 1000kJ/mol/nm. The system was then equilibrated using NVT and NPT ensemble for 100 ps. Finally, 50 ns production MD was performed for the system. The trajectories were set to be generated every 2 fs and save every 2ps. The 3CL<sup>pro</sup>-inhibitor complexes' results were analyzed for root mean square deviation (RMSD), root mean square fluctuations (RMSF), radius of gyration (Rg) and bond potential energy (BPE).

## Results and discussion

### *Analysis of 3CL<sup>pro</sup> for conserveness among coronaviruses*

The sequence alignment showed that the 3CL<sup>pro</sup> enzyme of new SARS-CoV-2 is 96.08% and 51.82% identical to SRAS (PDB: 2A5I) and MERS CoV (PDB: 5WKK) respectively. The sequence alignment also revealed that catalytic dyad residues His41, Cys145 of 3CL<sup>pro</sup> are conserved among SARS-CoV-2, SRAS-nCoV and MERS-CoV. Furthermore, the structural alignment/superposition of all 3 coronaviruses 3CL<sup>pro</sup> revealed conserved catalytic dyad residues His41, Cys145, inlaid at exactly same position in the binding pocket with an average RMSD nearly 0.12 Å as shown in **Figure 1**. Sequence and structural alignment clearly demonstrated that conserved functional residues within the active pockets of 3CL<sup>pro</sup> among SARS-CoV-2, SRAS-nCoV and MERS-CoV.



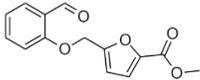
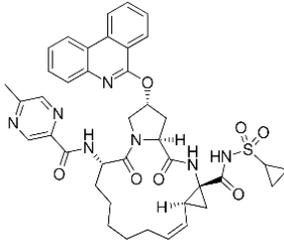
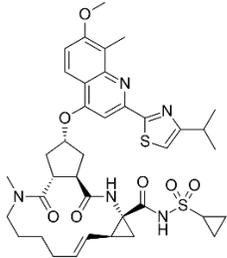
**Figure 1:** (A) Ribbon representation shown for the overlaying of SARS-CoV-2 3CL<sup>pro</sup> (red) (PDB ID: 6LU7), SARS-CoV 3CL<sup>pro</sup> (yellow) bound to an azapeptide covalent inhibitor (green) (PDB: 2A5I) and MERS-CoV 3CL<sup>pro</sup> (blue) bound to GC813 (pink). (B) Ribbon representation shown the catalytic dyad His41, Cys145 within SARS-CoV-2 3CL<sup>pro</sup> and SARS-CoV 3CL<sup>pro</sup> in yellow and red sticks, respectively. Catalytic dyad His42, Cys148 within MERS-CoV 3CL<sup>pro</sup> shown as blue sticks.

### *Structure-based virtual screening*

The generated structure was used to perform the structure-based virtual screening against an integrated library of 1016 compounds including 1000 covalent protease inhibitors and 16 FDA-approved protease inhibitors. The latter approach was applied to make use of both *de novo* drug design as well as the drug repurposing strategies. The screening of both databases reveals three potential inhibitors of SARS-CoV-2 3CL<sup>pro</sup>, namely, 621, Paritaprevir and Simeprevir with high binding energy scores of -13.3, -8.8 and -8.78 (**Table 1**). The predicted inhibitory constants (pKi)

based on the autodock vina docking scores showed that compound 621, Paritaprevir and Simeprevir could inhibit the SARS-CoV-2 3CL<sup>pro</sup> with pKi values of 0.00013, 0.36 and 0.37  $\mu$ M. Interestingly, Paritaprevir and Simeprevir are acylsulfonamide FDA-approved drugs that act as anti-hepatitis C by targeting the NS3/4A protease<sup>28</sup>. Up to our knowledge, there were no clinical data available regarding using of these two compounds to treat SARS-CoV-2, MERS-CoV or SARA-CoV.

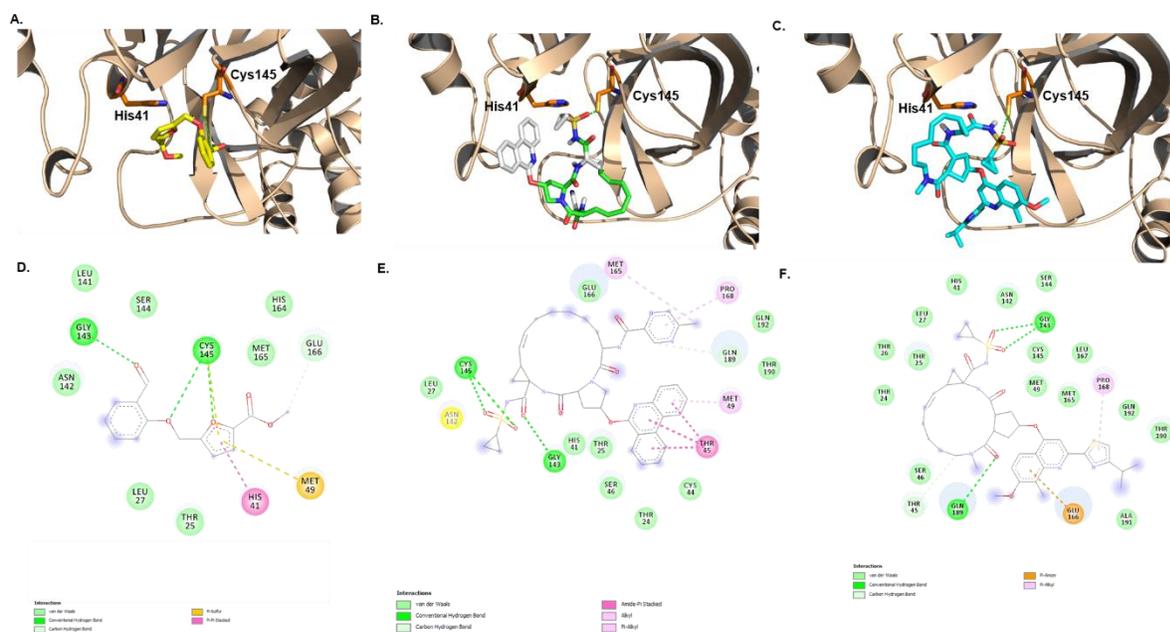
**Table 1:** Chemical structures, binding energy scores, predicted inhibitory constant (pKi) and molecular interactions of identified candidate compounds.

Name	Chemical Structure	Binding Energy Score (Kcal/mole)	pKi ( $\mu$ M)	H-bond	Hydrophobic
621		-13.3	0.00013	Gly143 and Cys145	His41
Paritaprevir		-8.8	0.36	Gly143 and Cys145	Thr45, Met165 and Pro168
Simeprevir		-8.78	0.37	Gly143 and Gln0189	Pro168

### *Molecular interaction and binding mode*

In order to understand the mechanism of interaction of these compounds with SARS-CoV-2 3CL<sup>pro</sup>, an unbiased docking of these compounds into the active site of SARS-CoV-2 3CL<sup>pro</sup>

enzyme was performed. Interestingly, the three compounds adapt the same binding mode within the binding pocket between catalytic dyad His41, Cys145, **Figure 2**. Furthermore, the distance between the reactive group within each compound was 4.1, 4.1 and 4.9Å for, compound 621, Paritaprevir and Simeprevir, respectively. While compound 621 and Paritaprevir were found to form hydrogen bond with Gly143 and Cys145, Simeprevir was found to form hydrogen bonds with Gly143 and Gln 189.



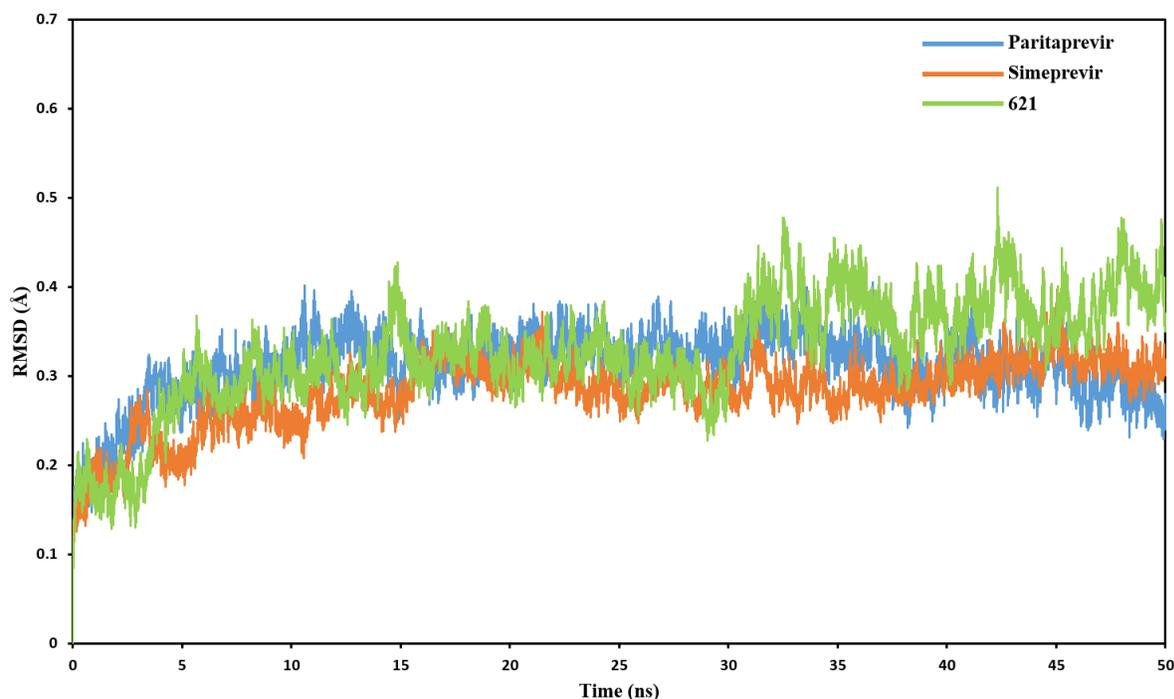
**Figure 2:** Binding mode and molecular interaction of identified candidate compounds with SARS-CoV-2 3CL protease (PDB ID: 6LU7). (A), (B) and (C) showed the binding mode of compound 621, Paritaprevir and Simeprevir, respectively within the active site of SARS-CoV-2 3CL protease. The catalytic dyad His41, Cys145 within the active site is shown in orange sticks. (D), (E) and (F) showed molecular interaction of compound 621, Paritaprevir and Simeprevir, respectively with SARS-CoV-2 3CL protease.

### *Molecular dynamic simulation*

To confirm the docking results and get more insight into the stability of ligand-protein complex, MD simulations were carried out for each SARS-CoV-2 3CL<sup>pro</sup>-inhibitor complex in the solvated states at 50 ns. The results of MD simulations have been examined on the basis of root mean square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration values as a function of time.

#### *Root mean square deviation (RMSD)*

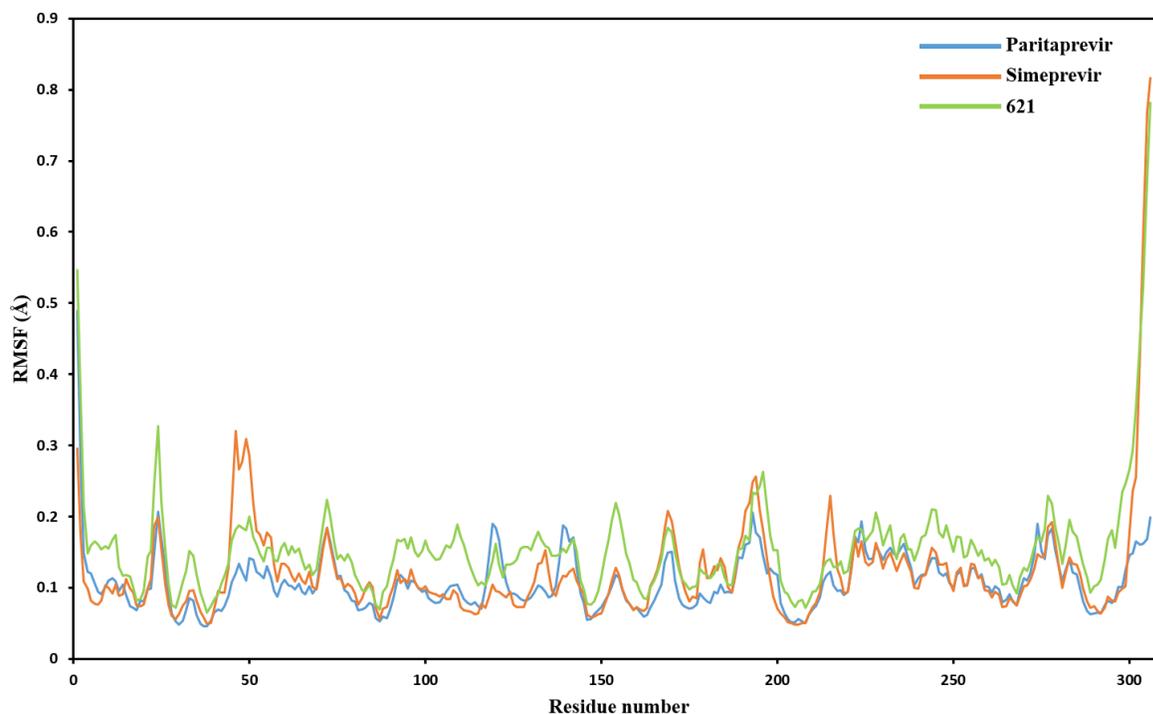
The RMSD measures the direct changes in the protein from the initial coordinates. The RMSD values of the protein backbone in complex with the three potential inhibitors were computed with respect to the initial structure as a frame reference (0 to 50 ns). The RMSD values steadily increased from 0 to 5 ns, and reached equilibration after that throughout the simulation period, especially for Paritaprevir and Simeprevir. The RMSD value for Compound 621 showed oscillations between 29 to 32 ns indicating that the compound was adapting another conformation within the binding pocket (**Figure 3**). The average RMSD values for the last 40 minutes for 621, Paritaprevir and Simeprevir were  $0.32 \pm 0.03$ ,  $0.30 \pm 0.02$  and  $0.35 \pm 0.04$  Å, respectively.



**Figure 3:** Plot of root mean square deviation (RMSD) of C–C $\alpha$ –N backbone vs. simulation time for solvated SARS-CoV-2 3CL protease in complex with the three candidate compounds during 50 ns molecular dynamics simulations.

#### *Root mean square fluctuation (RMSF)*

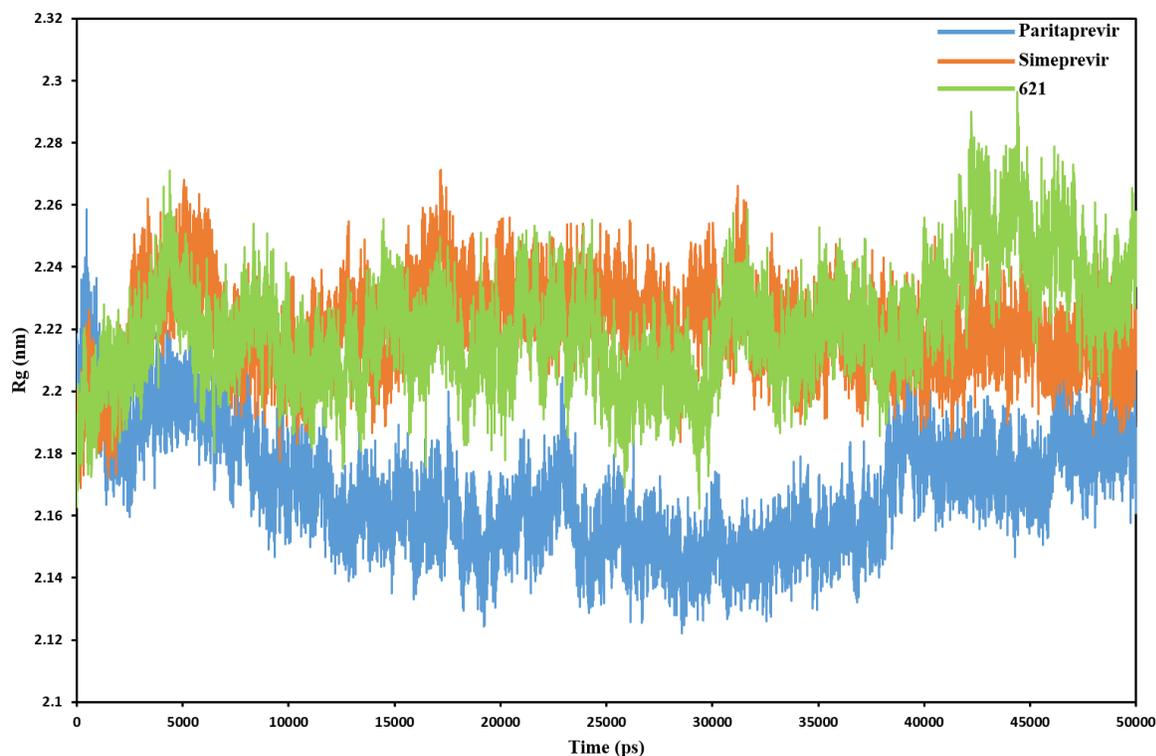
To explore the local protein flexibility, the time average of RMSF values of the 306 amino acids of SARS-CoV-2 3CL<sup>PRO</sup> protein in presence of the three inhibitors over simulation period were calculated. The RMSF values for the three complexes suggested that the catalytic dyad residues (His41 and Cys148) showed less fluctuation in all complexes (**Figure 4**). The average RMSF values were  $0.11 \pm 0.04$ ,  $0.12 \pm 0.012$  and  $0.15 \pm 0.07$  Å for compound 621, Paritaprevir and Simeprevir, respectively.



**Figure 4:** The root mean square fluctuation (RMSF) values of SARS-CoV-2 3CL protease in complex with the three candidate compounds were plotted against residue numbers.

#### *Radius of gyration (Rg)*

The radius of gyration (Rg) of the protein is associated with its size and compactness. The Rg values of three complexes were found to be 2.16 nm at the initial state. The Rg values of the complex of protein with Paritaprevir and Simeprevir were stabilized after initial increase at 5 ns supporting that the systems have reached equilibrium state. In the other hand, the Rg value for compound 621 was decreased from 10 ns to 40 ns and then it slightly increases up to 50 ns. The latter, indicates that the binding of 621 to the protein stabilized its secondary structure (**Figure 5**). The MD simulation results confirmed the stability of all identified compounds at the active site of SARS-CoV-2 3CL<sup>pro</sup>.



**Figure 5:** Plot of radius of gyration (Rg) during 50ns MD simulation of SARS-CoV-2 3CL protease in complex with the three candidate compounds.

## Conclusion

In the current study, pharmacoinformatics and molecular dynamic approaches were utilized to identify inhibitors of SARS-CoV-2 3CL<sup>pro</sup> as treatments for the new outbreak coronavirus disease 2019. Three compounds, compound 621, Paritaprevir and Simeprevir, were identified as potential inhibitors of SARS-CoV-2 3CL<sup>pro</sup> enzyme with predicted inhibitory constant in low micromolar concentration range. The binding affinity, mechanism and stability of binding of these compounds to SARS-CoV-2 3CL<sup>pro</sup> were confirmed by molecular docking and molecular dynamic simulation. Compound 621 could be used as a seed for de novo drug design of potential inhibitors to target the 3CL<sup>pro</sup> enzyme of SARS-CoV-2 as well as MERS-CoV and SARS-CoV. The clinical

agents, Paritaprevir and Simeprevir may also play a role in expediting the drug discovery process and be tested in clinical trials as a treatment for coronavirus disease 2019.

### Acknowledgment

Mubarak A. Alamri and Safar M. Alqahtani would like to thank Prince Sattam Bin Abdulaziz University for providing the necessary tools to conduct this research.

### Conflict of interest

Authors have no conflict of interest to declare.

### References

1. Ji W, Wang W, Zhao X, Zai J, Li XJJoMV. Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. *J Med Virol* 2020.
2. Weiss SR, Leibowitz JL. Coronavirus pathogenesis. In. *Coronavirus pathogenesis. Advances in virus research*: Elsevier; 2011.
3. Su S, Wong G, Shi W, Liu J, Lai AC, Zhou J, Liu W, Bi Y, Gao GFJTim. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* 2016;24(6): 490-502.
4. Cui J, Li F, Shi Z-LJNrM. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 2019;17(3): 181-192.
5. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu RJNEJoM. A novel coronavirus from patients with pneumonia in china, 2019. *N Engl J Med* 2020.
6. Lu H, Stratton CW, Tang YWJJoMV. Outbreak of pneumonia of unknown etiology in wuhan china: The mystery and the miracle. *J Med Virol*.
7. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020.
8. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y et al. A new coronavirus associated with human respiratory disease in china. *Nature* 2020.
9. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, Zhong W, Hao PJSCLS. Evolution of the novel coronavirus from the ongoing wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci* 2020.
10. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen K-Y. Coronaviruses—drug discovery and therapeutic options. *Nature reviews Drug discovery* 2016;15(5): 327.
11. Shirato K, Kawase M, Matsuyama S. Middle east respiratory syndrome coronavirus infection mediated by the transmembrane serine protease tmprss2. *Journal of virology* 2013;87(23): 12552-12561.

12. ul Qamar MT, Maryam A, Muneer I, Xing F, Ashfaq UA, Khan FA, Anwar F, Geesi MH, Khalid RR, Rauf SAJSr. Computational screening of medicinal plant phytochemicals to discover potent pan-serotype inhibitors against dengue virus. *Sci Rep* 2019;9(1): 1-16.
13. Ul Qamar MT, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani SJJotm. Epitope-based peptide vaccine design and target site depiction against middle east respiratory syndrome coronavirus: An immune-informatics study. *J Transl Med* 2019;17(1): 362.
14. McInnes C. Virtual screening strategies in drug discovery. *Current opinion in chemical biology* 2007;11(5): 494-502.
15. Kumar S, Nei M, Dudley J, Tamura KJBib. Mega: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 2008;9(4): 299-306.
16. DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography* 2002;40(1): 82-92.
17. Systèmes D. Biovia, discovery studio modeling environment. Dassault Systèmes Biovia: San Diego, CA, USA 2016.
18. Dallakyan S, Olson AJ. Small-molecule library screening by docking with pyrx. In. *Small-molecule library screening by docking with pyrx*. Chemical biology: Springer; 2015.
19. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open babel: An open chemical toolbox. *Journal of cheminformatics* 2011;3(1): 33.
20. Huey R, Morris GM. Using autodock 4 with autodocktools: A tutorial. The Scripps Research Institute, USA 2008: 54-56.
21. Sanner MF. Python: A programming language for software integration and development. *J Mol Graph Model* 1999;17(1): 57-61.
22. Edwards MP, Price DA. Role of physicochemical properties and ligand lipophilicity efficiency in addressing drug safety risks. In. *Role of physicochemical properties and ligand lipophilicity efficiency in addressing drug safety risks*. Annual reports in medicinal chemistry: Elsevier; 2010.
23. Hopkins AL, Groom CR, Alex A. Ligand efficiency: A useful metric for lead selection. *Drug discovery today* 2004;9(10): 430-431.
24. Murray CW, Erlanson DA, Hopkins AL, Keserü GrM, Leeson PD, Rees DC, Reynolds CH, Richmond NJ. Validity of ligand efficiency metrics. *Med Chem* 2014.
25. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure analysis with ucsf chimera. *BMC bioinformatics* 2006;7(1): 339.
26. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* 2008;4(3): 435-447.
27. Zoete V, Cuendet MA, Grosdidier A, Michielin O. Swissparam: A fast force field generation tool for small organic molecules. *Journal of computational chemistry* 2011;32(11): 2359-2368.
28. McConachie SM, Wilhelm SM, Kale-Pradhan PB. New direct-acting antivirals in hepatitis c therapy: A review of sofosbuvir, ledipasvir, daclatasvir, simeprevir, paritaprevir, ombitasvir and dasabuvir. *Expert review of clinical pharmacology* 2016;9(2): 287-302.