Supercomputer-aided Drug Repositioning at Scale: Virtual Screening for SARS-CoV-2 Protease Inhibitor

Sangjae Seo[a,*], Jung Woo Park[a], Dosik An[a], Junwon Yoon[a], Hyojung Paik[a], and Soonwook Hwang[a]

[a] Korean Institute of Science and Technology Information, Republic of Korea

*Corresponding Author: sj.seo@kisti.re.kr

ABSTRACT

Coronavirus diseases (COVID-19) outbreak has been labelled a pandemic. For the prioritization of treatments to cope with COVID-19, it is important to conduct rapid high-throughput screening of chemical compounds to repurposing the approved drugs, such as repositioning of chloroquine (Malaria drug) for COVID-19. In this study, exploiting supercomputer resource, we conducted high-throughput virtual screening for potential repositioning candidates of the protease inhibitor of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Using the three dimensional structure of main protease (Mpro) of SARS-CoV-2, we evaluated binding affinity between Mpro and drug candidates listed in SWEETLEAD library and ChEMBL database. Docking scores of 19,168 drug molecules at the active site of Mpro were calculated using Autodock Vina package. Among the calculated result, we selected 43 drug candidates and ran molecular dynamics (MD) simulation to further

investigate protein-drug interaction. Among compounds that bind to the active site of SARS-CoV-2, we finally selected the 8 drugs showing the highest binding affinity; asunaprevir, atazanavir, dasabuvir, doravirine, fosamprenavir, ritonavir, voxilaprevir and amprenavir, which are the antiviral drugs of hepatitis C virus or human immunodeficiency virus. We expect that the present study provides comprehensive insights into the development of antiviral medication, especially for the treatment of COVID-19.

INTRODUCTION

Coronaviruses are enveloped RNA viruses, which cause severe respiratory illness. Since the first outbreak of Severe Acute Respiratory Syndrome (SARS) in 2002[1,2], coronavirus has become a formidable threat to public health. The total number of global infection of Middle East Respiratory Syndrome (MERS)[3,4] confirmed by WHO reached 2,484 cases, with 858 deaths. A new type of novel coronavirus, named Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2), first identified in Wuhan, Hubei, China, in December 2019, and it has rapidly spread across the world. Up to now, according to situational report of WHO released April 8, 2020, due to coronavirus disease 2019 (COVID-19), 1,282,931 cases have been confirmed, and 72,774 deaths were reported. Unfortunately, no vaccine or antiviral drugs have been developed that clearly demonstrate effectiveness in the treatment of COVID-19, and drug candidates for treatment are undergoing clinical trials worldwide. In particular, for rapid development of therapies to COVID-19, investigations of efficacy of approved drugs, known as drug repositioning, is being actively conducted.

Viral replication process essentially requires the cleavage of polyproteins, which is catalyzed by a protease[5]. Thus, it has been suggested that inhibiting the active site of protease is an effective treatment of viral infection[6–8]. Owing to recent intensive efforts to unveil the structure

of SARS-CoV-2, the three-dimensional structure of main protease (Mpro) of SARS-CoV-2 (PDB ID: 6LU7) was resolved[9]. The Mpro forms a dimer, and the residues of CYS41 and HIS145 were identified as an active site[10]. Also several studies investigated the marketed drugs as for protease inhibitor (PI) of SARS-CoV-2[10–14]. However, to our knowledge, there was no large-scale virtual screening attempted for discovering leads for PI of SARS-CoV-2 using both docking calculations and massive molecular dynamics (MD) simulations.

In this study, high-throughput virtual screening of drug molecules was conducted on Nurion, KISTI's Cray CS500 flagship supercomputer. We performed docking calculation on known drug molecules and MD simulations on selected protein-ligand complexes to obtain accurate screening results. The purpose of this study was two-fold: 1) to evaluate drug molecules in terms of drug repositioning/repurposing for COVID-19 treatment, and 2) to provide insightful information in atomic details for the development of antiviral drug of SARS-CoV-2.

METHOD

Dataset

The drug molecules were obtained from SWEETLEAD library[15] and the drug library of ChEMBL database[16]. The SWEETLEAD library contains 9,127 chemical structures of which are approved drugs, isolates from traditional medicinal herbs, and regulated chemicals. ChEMBL database is a manually curated open-source database and contains more than 100 million bioactive molecules. We chose the drug compounds only, which are 10,041 entries, including both FDA approved drugs and investigational compounds. The SDF file of SWEETLEAD library was converted into MOL2 and PDB formatted files using Open Babel[17]. Since ChEMBL does not contain three-dimensional coordinates, we searched SMILES codes

from ChEMBL IDs, and converted the SMILES codes into 3D structures using UCSF Chimera[18].

Docking calculation

Molecular docking studies were performed using AutoDock Vina package[19]. The required input files for Autodock Vina were prepared using AutoDock Vina plugin of UCSF Chimera[18] with default options. The three-dimensional coordinates (X, Y, Z) of center of grid box was set to (-20.5987, 28.2394, 46.3799) with the search size of (25.0, 20.0, 25.0). The center of grid box locates near the center of mass (COM) of residues of HIS41 and CYS145, and the size of box was determined to sufficiently cover the active site.

Simulation details

The initial configurations of protein and drug molecules were extracted from the result of Autodock Vina. Using our in-house code, we converted the Autodock Vina results to PDB compatible format and added missing hydrogen atoms to the drug molecules. The systems for MD simulation were prepared using CHARMM-GUI[20]. The topology and parameters for the drug molecules were automatically generated using CGenFF[21,22]. The Mpro was modelled with the CHARMM 36 force field[23]. Each simulation box was solvated with TIP3P water molecules, and additional NaCl ions were placed in the box to neutralize systems. All MD simulations were performed using GROMACS software[24]. The temperature and pressure were maintained at 303K and 1atm, respectively. Nose-Hoover thermostat[25,26] and Parinello-Rahman barostat[27,28] were used to control temperature and pressure, respectively. The simulation box was modelled in octahedral shape with isotropic coupling scheme. Electrostatic interactions

were calculated with the particle mesh Ewald method[29,30]. Lennard-Jones potentials were truncated using a force switching cutoff approach. All bonds involving hydrogen atoms were constrained using the LINCS algorithm[31]. We equilibrated the systems according to the protocol presented by CHARMM-GUI[20]. We conducted energy minimization of the systems using steepest descent algorithm, and another 200 ps simulation was run for the equilibration of the system in *NVT* ensemble. We performed production run of 500 ns for each system in *NPT* ensemble with 2 fs time step.

Analyses

The analyses were conducted using VMD[32] and in-house code written with Python, Scipy and MDAnalysis[33]. In order to evaluate the stability of selected drug candidates at the active site, we measured the distance between COM of drug molecule, and COM of residues of HIS41 and CYS145. Binding criteria was determined by COM-COM distance. If the COM-COM distance is smaller than 1.5 nm, we regarded that the ligand is bound to the active site of Mpro. Then, we calculated bound ratio, which is defined by the ratio of residence (bound) time and unbound time. The binding free energy of protein-ligand complex was calculated using *g_mmpbsa* package[34]. For each protein-ligand complex MD simulation, 100 snapshots were extracted from the frames of which ligands are bound.

RESULTS and DISCUSSION

Computational framework

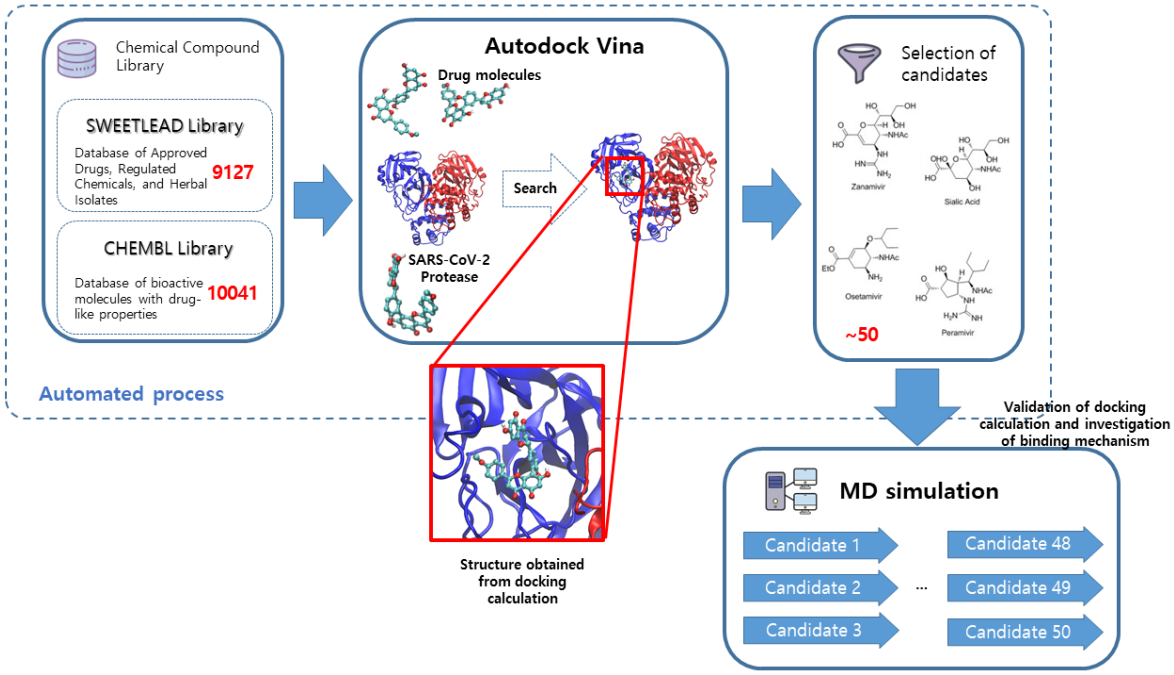

Figure 1. Computational framework for high-throughput virtual screening. Numbers in red represents the number of molecules at each step

For rapid evaluation of drug candidates, we developed a computational framework, which is suitable for parallel calculation on supercomputer resource. Figure 1 presents the developed framework. Our in-house code automatically converts SDF files from each drug compound library, and generates input scripts for job submission. All docking calculations were conducted in a parallel manner, which enables rapid virtual screening. Once docking calculation is finished, the result files from Autodock Vina were analyzed, and the summary report was generated. The drug molecules, which were selected based on the docking scores and criteria we assigned, are then converted into MD input files. We automated the screening process with bash scripts and python codes. This study was conducted on KISTI's Nurion supercomputer

equipped with Intel Xeon Phi 7250 and 96GB memory each node (8,305 nodes in total). The entire computation was finished in a week on 512 nodes (64 MPI ranks/node). Considering that a similar sized system takes 40-100 ns/day on a single node with GPU[35], it will require at least 200 days (i.e., performing 100ns/day, 40 jobs in serial) to finish whole calculations on PC. Exploiting supercomputer resource and our parallel framework enabled us to significantly reduce computation time.

Docking calculation

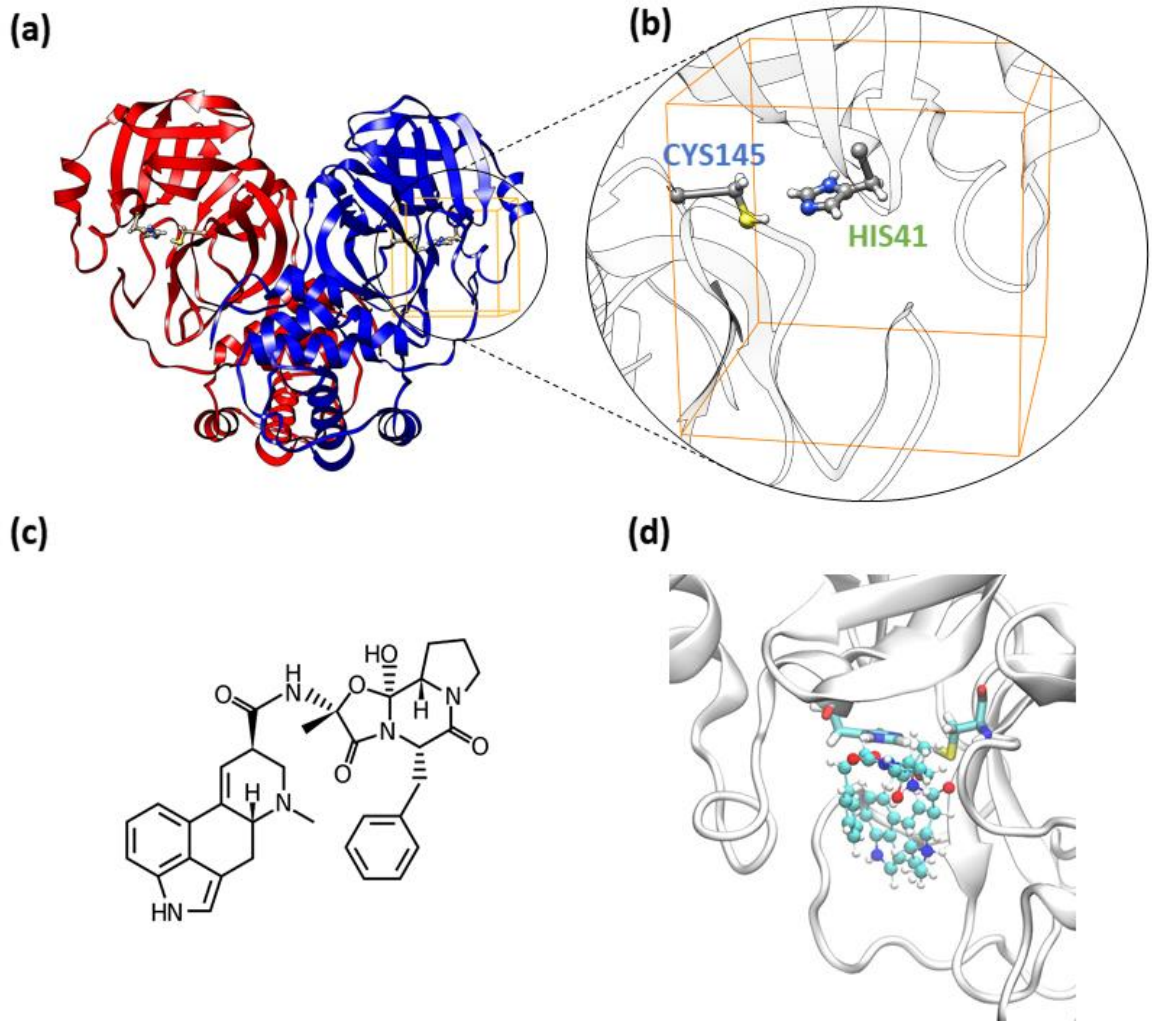

Figure 2. Molecular structure of the main protease (Mpro) of SARS-CoV-2 and a potential

protease inhibitor. (a) Each monomer of Mpro is represented in different colors. (b) The main active site of Mpro and search box (orange line) are highlighted. (c) Chemical structure of ergotamine, which showed highest docking score (d) Binding conformation of ergotamine (Ball and stick) at active site of Mpro.

We conducted docking calculations at the active site of Mpro of SARS-CoV-2. Figure 2 depicts the structure of Mpro and its binding site. We assigned calculation box near the residues of CYS41 and HIS145 (Fig 2(b)). The original lists extracted from SWEETLEAD library and ChEMBL database contained 19,168 molecules. Figure 2(c, d) illustrates the chemical structure and the conformation of ergotamine, which showed highest score. Although there are several hydrogen bond donors found in the ergotamine structure, the calculated ergotamine structure did not participate in hydrogen bonding with Mpro.

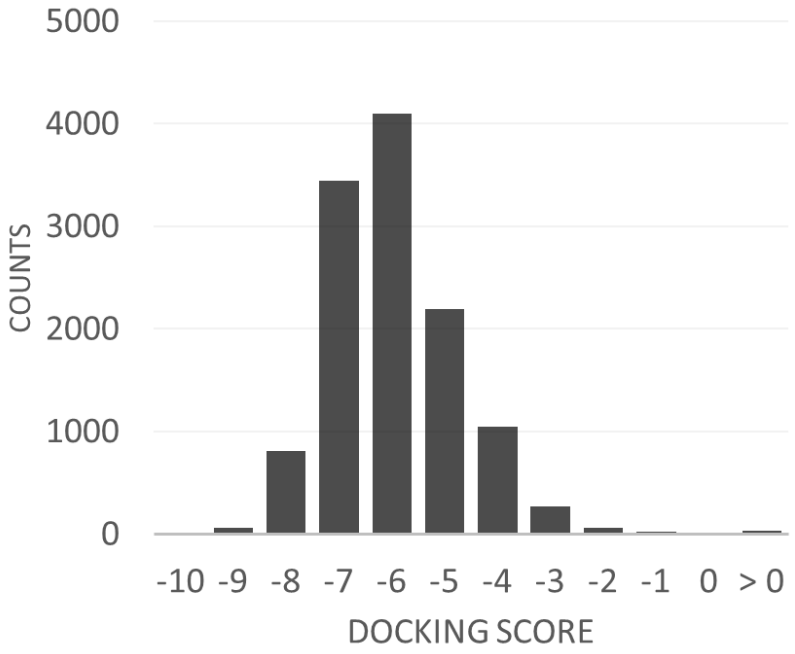

Figure 3. Docking score distributions of drug molecules

Figure 3 depicts the histogram of docking scores (See SI for full lists). Removing repetitive items and cases of calculation failure, 12,057 molecules were remained. As shown in Fig. 3, many drug molecules show high affinity to the active site of Mpro of SARS-CoV-2. 8,428 molecules showed docking scores larger than -6.0. Also, it should be noted that the differences of docking scores between each molecule are very small. For example, 3,448 and 4,102 molecules are included in the bins of -7 and -6, respectively. Generally, docking calculation enables us to compare the relative preference of drug molecules to proteins within reasonable calculation time. However, considering that the differences of docking scores are very small, and the results are sensitive to the options that user chose, it is difficult to solely rely on the result of docking calculation to distinguish which one is prior to the other for the purpose of drug repositioning. In other words, docking calculation can give a brief knowledge for drug repositioning, but further analyses and cross-validation are still required. Thus, we conducted MD simulation on the selected molecules to verify the result of docking calculation, and also to investigate docking mechanisms.

Molecular dynamics simulation

We ran MD simulations of selected molecules to verify the result of docking calculation and investigate binding mechanism of PIs. The fluctuation of root-mean-square deviation (RMSD) of Mpro indicates that Mpro did not undergo significant conformational changes in the course of simulation (See SI).

In order to select molecules for MD simulation, we put highest priority on various inhibitors for antiviral treatment, antibiotics for pneumonia, and a couple of vitamins, which showed high docking scores. In addition, the drugs underwent clinical trials for the treatment of COVID-19, such as remdesivir and hydroxychloroquine, were also tested. Table 1 lists the selected drug candidates and its calculated scores.

Table 1. List of selected molecules for MD simulation and its simulation results

Score (kcal/mol)	Name	MD Simulation status[a]	Description[b]
-8.6	SIMEPREVIR	UNBOUND	HCV[d] ns3/4a protease inhibitor
-8.3	GLECAPREVIR	UNBOUND	HCV ns3/4a protease inhibitor
-8.2	FOLIC ACID	UNBOUND	A synthetic form of vitamin B9
-7.9	DARUNAVIR	UNBOUND	HIV[e] protease inhibitor
-7.9	DASABUVIR	TIGHT	HCV non-nucleoside ns5b inhibitor
-7.8	ZOFENOPRIL	UNBOUND	Angiotensin-converting enzyme (ACE) inhibitor.
-7.8	TELAPREVIR	UNBOUND	HCV ns3/4a protease inhibitor
-7.6	REMDESIVIR	UNBOUND	Experimental, a potential treatment for Ebola virus
-7.5	ELVITEGRAVIR	UNBOUND	HIV-1 integrase strand transfer inhibitor
-7.5	GEMIFLOXACIN	UNBOUND	Inhibitor of both DNA gyrase and topoisomerase IV (bacterial pneumonia)
-7.4	DARUNAVIR ETHANOLATE	UNBOUND	HIV protease inhibitor
-7.4	DEXLANSOPRAZOLE	UNBOUND	Proton pump inhibitor
-7.4	AMPRENAVIR	LOOSE	HIV protease inhibitor
-7.4	RITONAVIR	TIGHT	HIV protease inhibitor
-7.3	PARECOXIB	UNBOUND	Cyclooxygenase-2 (COX-2) inhibitor
-7.3	ATAZANAVIR	TIGHT	HIV protease inhibitor

-7.3	FOSAMPRENAVIR	TIGHT	HIV protease inhibitor
-7.3	ROFLUMILAST	UNBOUND	Phosphodiesterase-4 (PDE-4) inhibitor
-7.2	CEFTOBIPROLE	UNBOUND	Treatment of complicated skin infections and pneumonia.
-7.2	ETRAVIRINE	UNBOUND	HIV-1 non-nucleoside reverse transcriptase inhibitor
-7.2	SULINDAC	UNBOUND	COX-1 and COX-2 inhibitor
-7.2	CALCIPOTRIENE HYDRATE	UNBOUND	A synthetic derivative of calcitriol or Vitamin D.
-7.2	DORAVIRINE	TIGHT	HIV-1 non-nucleoside reverse transcriptase inhibitor
-7	TRIFLURIDINE	UNBOUND	Herpes simplex virus type 1 and 2 thymidine phosphorylase inhibitor
-7	ENALAPRILAT	UNBOUND	ACE inhibitor
-6.8	BOCEPREVIR	UNBOUND	HCV ns3/4a protease inhibitor
-6.8	TRIMETREXATE	UNBOUND	Dihydrofolate reductase inhibitor (pneumonia)
-6.7	GRAZOPREVIR	UNBOUND	HCV ns3/4a protease inhibitor
-6.5	SULFACYTINE	UNBOUND	Antibiotics inhibitors of *p*-aminobenzoic acid
-6.5	ASUNAPREVIR	TIGHT	HCV ns3 protease inhibitor
-6.4	LEFAMULIN	UNBOUND	Inhibitor binds to the peptidyl transferase center of the 50S subunit of the bacterial ribosome(bacterial community-acquired pneumonia)
-6.3	PENTAMIDINE	UNBOUND	Topoisomerase inhibitor (pneumonia)
-6.2	BENZYLPENICILLIN	UNBOUND	Antibiotic used to treat a number of bacterial infections including pneumonia
-6.2	HYDROXYCHLOROQUINE	UNBOUND	Treatment of uncomplicated malaria

-6.1	EDOXUDINE	UNBOUND	Herpes simplex virus type 1 and 2 selective inhibitor
-6.1	GANCICLOVIR	UNBOUND	Potent inhibitor of the herpesvirus family including cytomegalovirus.
-6.1	OSELTAMIVIR	UNBOUND	Influenza viruses a (including pandemic H1N1) and b neuraminidase inhibitor
-6.1	PERAMIVIR	UNBOUND	Influenza virus neuraminidase inhibitor, .
-5.8	TAZOBACTAM	UNBOUND	Beta-lactamase inhibitor (bacterial pneumonia)
-5.8	ZALCITABINE	UNBOUND	Inhibitor of HIV replication
-5.5	LAMIVUDINE	UNBOUND	HIV-1 and hepatitis b virus reverse transcriptase inhibitor
-4	VOXILAPREVIR	TIGHT	HCV ns3/4a protease inhibitor
-3.4	VELPATASVIR	UNBOUND	HCV NS5A inhibitors

[a] We assigned TIGHT state if the bound ratio was larger than 0.95, and UNBOUND state otherwise. LOOSE state indicates that the ligand was detached from the active site in chain A, and bound again to the active site in chain B.

[b] Description is referred from the DrugBank database[36].

[d] Hepatitis C virus (HCV)

[e] Human immunodeficiency virus (HIV)

Firstly, docking calculation and MD simulation did not show consistent result. For example, docking score of simeprevir shows highest score among the selected molecules, but it was unbound from the Mpro. On the other hands, voxilaprevir shows relatively small docking score, but it was tightly bound to the active site of Mpro. We suspect that the inconsistency is mainly due the inaccuracy of docking calculation, and there is also a possibility that incorrect parameters are assigned to drug molecule due to the automated parameterization using CGenFF.

MD simulation results show that asunaprevir, atazanavir, dasabuvir, doravirine, fosamprenavir, ritonavir and voxilaprevir, which are inhibitors of HCV or HIV, are tightly bound at the active site. It indicates that those drugs can be a potential candidate for the treatment of COVID-19. It is noteworthy that hydroxychloroquine, which is actively being studied as a treatment for COVID-19, was stabilized at the location predicted by Autodock Vina. However, since the binding location was at distance from the active site, which was out of our criteria, we labeled it as UNBOUND. Intriguingly, although amprenavir was detached from the active site in chain A at the beginning of simulation, after around 400 nanoseconds, amprenavir molecule was bound to the active site in chain B. Thus, we also considered that amprenavir also can be a potential inhibitor of Mpro. Nonetheless, the cases of amprenavir indicate that we need long simulation time to accurately predict the binding affinity.

To further investigate the binding mechanism of the candidate inhibitors, we analyzed the binding free energy and hydrogen bonds. We measured binding free energy using molecular mechanics Poisson-Boltzmann surface area (MMPBSA) approach. Table 2 lists the energetic contributions to the binding free energy, including van der Waals (E_{vdW}), electrostatic interaction (E_{elec}), polar solvation (Polar), and non-polar solvation energy (SASA) terms. The contribution from electrostatic interaction is relatively small, and the tight binding of drug molecules on Mpro is mainly due to van der Waals interactions and. Also, hydrogen bonding

between ligands and the active sites (HIS41 and CYS145 residues) is not considered to be an important contributor to the stability of protein-ligand interaction (Table 3).

Table 2. Binding free energy analysis of drugs (kcal/mol)

	E_{vdW}	E_{elec}	Polar	SASA	ΔG_{bind}
DASABUVIR	-38.2±4	-7.3±3.2	29.5±5.7	-4.5±0.5	-20.4±3.1
ASUNAPREVIR	-57.9±4.1	-12.4±3.2	50.6±3.2	-6.7±0.4	-26.5±4.8
DORAVIRINE	-39.6±2.7	-7.2±2.2	27.8±3.6	-3.7±0.2	-22.7±3.2

Table 3. Hydrogen bonds analysis of candidate drugs

	Donor	Acceptor	Occupancy
AMPRENAVIR	HSD41	Ligand	0.11%
ASUNAPREVIR	CYS145	Ligand	1.40%
ATAZANAVIR	-	-	-
DASABUVIR	HSD41	Ligand	0.28%
DORAVIRINE	HSD41	Ligand	0.14%
FOSAMPRENAVIR	CYS145	Ligand	0.55%
RITONAVIR	Ligand	HSD41	4.10%
VOXILAPREVIR	HSD41	Ligand	2.23%

CONCLUSION

In this study, exploiting a supercomputer resource, we searched for potential drug candidates of SARS-CoV-2. For efficient and accurate virtual screening of drug molecules, we presented a novel computational framework, which can accelerate drug repositioning study. Approximately 20,000 drug molecules were evaluated to assess its binding affinity to the active site of the Mpro using Autodock Vina software. Docking calculation indicated that many of

approved drugs show high affinity to Mpro. In order to validate the docking results, we selected 43 drug molecules and performed MD simulation. Unlike docking calculation results, MD simulation showed that most of selected drugs have less binding affinity to Mpro. This result indicates that detailed scrutiny is required to correctly interpret the result of docking calculation. The MD simulation results presented that some FDA approved antiviral drugs can be used as a potential treatment for COVID-19. Despite the promising results from this study, further validations of the result, such as *in vitro* and *in vivo* experiment, and clinical test, are still necessary. In addition, longer simulation time and/or enhanced sampling techniques (*i.e.,* replica exchange MD) are needed to be performed to obtain more accurate and detailed information of protein-ligand interaction.

ACKNOWLEDGEMENT

This work was supported by the National Supercomputing Center with supercomputing resources including technical support TS-2020-RE-0012. We would like to express our appreciation to Juyong Lee (Department of Chemistry, Kangwon National University, South Korea) for helpful discussion.

REFERENCE

1. Peiris, J. S. M., Guan, Y. & Yuen, K. Y. Severe acute respiratory syndrome. *Nat. Med.* **10**, S88–S97 (2004).

2. Groneberg, D. A. *et al.* Treatment and vaccines for severe acute respiratory syndrome. *Lancet Infect. Dis.* **5**, 147–155 (2005).

3. Shapiro, M. *et al.* Middle East respiratory syndrome coronavirus: review of the current situation in the world. *Disaster Mil. Med.* **2**, 9 (2016).

4. Shehata, M. M., Gomaa, M. R., Ali, M. A. & Kayali, G. Middle East respiratory syndrome coronavirus: a comprehensive review. *Front. Med.* **10**, 120–136 (2016).

5. Kohl, N. E. *et al.* Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci.* **85**, 4686–4690 (1988).

6. Chou, K.-C. Prediction of Human Immunodeficiency Virus Protease Cleavage Sites in Proteins. *Anal. Biochem.* **233**, 1–14 (1996).

7. Flexner, C. HIV-Protease Inhibitors. *N. Engl. J. Med.* **338**, 1281–1293 (1998).

8. Lamarre, D. *et al.* An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**, 186–189 (2003).

9. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L., Xu, W., Xiao, G., Qin, C., H. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *RCSB PDB* doi:10.2210/pdb6lu7/pdb.

10. Tahir Ul Qamar, M., Alqahtani, S., Alamri, M. & Chen, L.-L. *Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants*. *Journal of Pharmaceutical Analysis* (2020). doi:10.1016/j.jpha.2020.03.009.

11. Chang, Y.-C. *et al. Potential Therapeutic Agents for COVID-19 Based on the Analysis of Protease and RNA Polymerase Docking*. (2020). doi:10.20944/preprints202002.0242.v1.

12. Khaerunnisa, S., Kurniawan, H., Awaluddin, R., Suhartati, S. & Soetjipto, S. *Potential Inhibitor of COVID-19 Main Protease (Mpro) From Several Medicinal Plant Compounds by Molecular Docking Study*. (2020). doi:10.20944/preprints202003.0226.v1.

13. Chandel, V., Raj, S., Rathi, B. & Kumar, D. *In Silico Identification of Potent COVID-19 Main Protease Inhibitors from FDA Approved Antiviral Compounds and Active Phytochemicals through Molecular Docking: A Drug Repurposing Approach*. (2020). doi:10.20944/preprints202003.0349.v1.

14. Fischer, A., Sellner, M., Neranjan, S., Lill, M. & Smiesko, M. *Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 687 Million Compounds*. (2020). doi:10.26434/chemrxiv.11923239.

15. Novick, P. A., Ortiz, O. F., Poelman, J., Abdulhay, A. Y. & Pande, V. S.

SWEETLEAD: an In Silico Database of Approved Drugs, Regulated Chemicals, and Herbal Isolates for Computer-Aided Drug Discovery. *PLoS One* **8**, e79568 (2013).

16. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).

17. O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).

18. Pettersen, E. F. *et al.* UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

19. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–61 (2010).

20. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

21. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* NA-NA (2009) doi:10.1002/jcc.21367.

22. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).

23. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).

24. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).

25. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).

26. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).

27. Parrinello, M. & Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **45**, 1196–1199 (1980).

28. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

29. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

30. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

31. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

32. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol.*

Graph. **14**, 33–38 (1996).

33. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

34. Kumari, R., Kumar, R. & Lynn, A. g_mmpbsa —A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **54**, 1951–1962 (2014).

35. Kutzner, C. *et al.* More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **40**, 2418–2431 (2019).

36. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).