Screening and Discovery of Small Molecule Inhibitors Targeting SARS-Cov-2 Main Protease

Ying Wang, Daixi Li, Yuqi Zhu

Abstract— **The novel coronavirus that emerged at the end of 2019 (SARS-CoV-2) has had a great impact on people's health and lives. The role of the main protease of SARS-CoV-2 is to cleave the viral protein. Studies have shown that the main protease is functionally specific and evolutionarily highly conserved and could be a target for the development of antiviral drugs. There is still no specific drug for the treatment of novel coronavirus. In this paper, a new small molecule Chembl2_3 with higher binding free energy than the natural ligand was screened by computer-aided technology, demonstrating the potential of this small molecule as a therapeutic drug for novel coronavirus, and the method also provides a reference for small molecule drug design.**

***Index Terms*—molecular docking; main protease; small molecule drug; SARS-CoV-2.**

I. Introduction

Novel coronavirus pneumonia caused by a novel coronavirus in late 2019(COVID-19), has had a huge impact on people's lives and society. SARS-CoV-2 belongs to the coronavirus family and is the seventh known human coronavirus in the same family after HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, Middle East respiratory syndrome coronavirus and SARS-CoV[1]. Common symptoms include fever, dry cough, shortness of breath, and generalized pain, which in severe cases may lead to organ failure and death. Transmission of the virus is mainly through respiratory droplets, so it is essential to reduce or avoid contact with infected individuals. Precautions include regular mask wearing, social distancing, avoiding people other than close relatives, self-isolation, and indoor ventilation. Because of its strong infectious and transmissible nature, deaths have exceeded six million cases worldwide. The latest variant of the COVID-19 virus is JN.1 (BA.2.86.1.1)[2]. Emerging in 2023, it quickly became the dominant strain in the United States and other regions. JN.1, a descendant of Omicron, is characterized by its high degree of mutation and an exceptionally high transmission capability[3]. While a significant portion of the globe has adapted to a new normal following the pandemic, existing prevention and treatment approaches for it remain insufficiently precise.

Manuscript received March 17, 2025

Ying Wang, School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai ,China

Daixi Li, School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, China

Yuqi Zhu, School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, China

Several drugs are recommended for the treatment of COVID-19 according to the 9th edition of the World Health Organization's guidelines for living published on January 14, 2022. In the extensive array of antiviral medications on the market, a select few have garnered FDA approval for managing COVID-19 symptoms. These include Remdesivir, Tocilizumab, Baricitinib, and the combined formulation of Nirmatrelvir and Ritonavir, known commercially as Paxlovid. Mpro inhibitors include Paxlovid and Ensitrelvir. The two molecules in question serve as emblematic examples of covalent and noncovalent Mpro inhibitors. Despite their availability, both Nirmatrelvir and Ensitrelvir come with inherent constraints. Nirmatrelvir cannot be used independently and must be administered alongside ritonavir, which inhibits the human CYP3A4 enzyme to extend the half-life of Nirmatrelvir in the body. Similarly, Ensitrelvir, a strong inhibitor of CYP3A4, could pose risks for patients with pre-existing medical conditions[4]. Therefore, there is a need to develop additional potential inhibitors against SARS-CoV-2 that offer opportunities for COVID-19 treatment. Empirical studies involving trial and error are frequently costly, which is why medication development takes so long and costs so much money. To meet the urgent need for developing anti-COVID-19 drugs, rapid discovery of potential candidate compounds through computer technology is essential[5].

These polyproteins translated by viral code are processed by two virus-encoded proteases, papain-like protease (PLpro) and main protease (3 Chymotrypsin like protease, Mpro, also known as 3CLpro), and release 16 non-structural proteins (nsp). Each nsp serves a different function in the virus life cycle and is crucial to the virus. Mpro cleaves no less than 11 multiprotein sites and releases the nsp4-nsp16, including Mpro, RdRp, and Helicase (Hel). The rest of the genome encodes four structural proteins (spike (S), membrane (M), envelope (E), and nucleocapsid (N)) and nine auxiliary proteins[6].SARS-CoV-2 main protease is a cysteine protease with a relative molecular mass of 34,000 and has 96% similarity to SARS-CoV Mpro. Mpro is a key CoV enzyme that plays an important role in the cleavage of viral proteins. When the virus enters the cell, the RNA virus takes over the host cell to express structural and nonstructural proteins, and Mpro cleaves large viral polyproteins into mature viral proteins. To date, SARS-CoV-2 has produced many mutants and a comparison of SARS-CoV-2 Mpro with Omicron Mpro sequence comparisons based on the World Health Organization Designated Variant of Concern (VOC) strains on GISAID (http://gissaid.org/) shows that it is highly conserved. Mpro is also considered an attractive target for inhibiting viral replication due to some of its specific

functions in the transmission of viral infections, which make it an ideal target for drug development[7].

In recent years, the application of molecular simulation methods in chemoinformatics (chemical molecules) and structural bioinformatics (proteins) in the drug discovery process has yielded very impressive results, Meanwhile, the development of bioinformatics tools and computer capabilities can accelerate the drug discovery process by shortening timelines and predicting their potential affinity to many drug targets[8, 9]. Imran et al.[10] used computer-based screening simulations of natural flavonoids and found that flavonoids such as baicalin and kaempferol showed stronger bonding affinity to the SARS-CoV-2 primary protease site. Also, bioflavonoids, isoprenylated flavonoids, flavones, and flavanones showed better interaction and binding affinity at the active site of the main protease, leading to the conclusion that flavonoids from some medicinal plants could be effective candidates against SARS-CoV-2 main protease. Manish et al. [11] used molecular dynamics and alchemical methods to confirm that theaflavins could inhibit SARS-CoV-2 main protease. Masand et al.[12]screened from a food database and found that two derivatives of spermidine exhibited high affinity to COVID-19 main protease and could be used as candidates for the treatment of COVID-19.

In this paper, computer-aided drug design and virtual screening methods were used to explore additional candidate compounds for the treatment of COVID-19. The molecular database was subjected to virtual screening, followed by subjecting top ranked main protease and small molecule complexes to molecular dynamics simulations. The simulated complexes were analyzed for their binding modes, and binding free energies. Additionally,the literature crystal structures were compared to screen potential inhibitors targeting the main protease that provided more references for the discovery of anti-novel coronavirus drugs.

II. Materials and Methods

A. SARS-CoV-2 Main Protein and Ligand Preparation

The homodimeric complex of SARS-CoV-2 Mpro (PDB ID: 6LU7)[13] were downloaded from the RCSB PDB database (http://www.rcsb.org/) and used as the receptor protein for molecular docking after removing the crystalline water molecules and small molecule ligands using PyMOL which is a Python-based molecular visualization software[14-16]. Small molecule compounds were downloaded from the ChEMBL database (http://www.ebi.ac.uk/chembl/) for building a molecular database. For verification, we docked the crystal ligand (PDB ID: 6LU7) in the protein again, and then using PyMOL, separated the natural ligand from the receptor, and then docked the natural ligand into the same receptor as a control experimental group.

B. Virtual Screening

AutoDock[17] stands out as a freely accessible and open-source molecular docking software that enjoys widespread adoption across academics. Its core strengths lie in the innovative integration of "fast grid-based energy evaluation" alongside an "efficient search for torsional degrees of freedom", making it a tool of choice for many researchers.

This sophisticated balance between delivering highly accurate predictions and managing computational resources efficiently allows AutoDock to proficiently forecast the interactions between ligands and biomolecular targets, offering a valuable blend of precision and practicality in the domain of molecular docking. An initial screening of the small molecule library was performed using Lipinski's principle[18] (MW<500da, H-bond acceptors<10, H-bond donors<5) to exclude molecules that were not suitable drug candidates, and the number was reduced from the starting about 1,300,000 to 179 small molecules, reducing the scope and reducing the computational effort.

The receptor proteins and small molecule compounds were converted into pdbqt format using the AutoDock tool. Box information and coordinate information were generated using PyMOL software, the box size of the SARS-CoV-2 main protease binding pocket was set to 19.3 Å× 29.9Å× 21.4Å, the grid coordinates for the main protease binding site determined as (x = −9.75, y = 11.45, z = 68.9). Autodock Vina affinity scoring program was called and combined with Python script for molecular docking and screening. Finally, 179 small molecule compounds with a binding affinity greater than -10kJ/mol were obtained, and the results showed that all small molecules were in the binding site and not off-target.The small molecules and receptor proteins screened in the previous step are imported into the Autogrow program. This program has an improved algorithm that allows semi-automated screening of designed protein inhibitors to generate new small molecule structures using affinity as a standard.

C. Molecular Dynamics Simulation

Molecular dynamics simulations are carried out using GROMACS 2020.4 software. Before the MD simulations, to describe clearly the natural ligand N3 and the generated small molecules referred to as N3 and Chembl2_3 in the subsequent discussion analysis, an ff14SB force field was established for the main protease and a GAFF force field for the ligand, us ing antechamber to assign ligand charges. The complexes were dissolved in a TIP3P water model box and appropriate amounts of positive and negative ions were added to make the system neutral. The steepest descent and conjugate gradient algorithm are used to minimize the energy of the system, the shake algorithm is used to constrain the hydrogen atom bond length, and the particle network ewald sum (PME) method is used to deal with the long-distance electrostatic interaction between the main protease and the ligand. For each system, the cut-off value of non-bond interaction is set to 10 Å, the total pressure of the system is set to 1 atm, the temperature is set to 300 K, the time interval is 2 fs, and the total simulation time is 500 ns.

D. Combination of Free Energy

The study uses YANK to calculate the binding free energy of the master protease and ligand. It is a platform for the alchemical free energy calculations. Free energy is a function of state, and YANK calculates free energy following a thermodynamic cycle, which is based on the principle of converting the protein and ligand from the unbound state to

the bound state through a series of non-physical (or alchemical) intermediate states. The initial phase involves severing the interaction between the ligand and its surrounding solution, which results in a change in free energy denoted as $\Delta G^{solv}_{elec+vdw}$.

Consequently, the ligand ceases to interact with its environment. To ensure the ligand's position and orientation remain akin to those in its bound state, a constraint is applied, with its associated free energy being determined in accordance with the methodology outlined by Boresch[19]. Following this, the protein and ligand, albeit constrained, are introduced into a solution not as separate entities but in a configuration resembling their bound state. During this stage, the absence of interaction between the ligand and the protein means the change in free energy (ΔG) is zero. The next step allows for the interaction of the ligand with the solution, facilitating the calculation of $\Delta G^{prot}_{elec+vdw}$. Despite this interaction, the ligand remains constrained, distinguishing it from a true bound state. To address this, the constraints are systematically lifted in a sequence of simulations, enabling the determination of ΔG^{prot}_{restr}. Completing these calculations across the entire alchemical cycle allows for the derivation of the binding free energy as the system transitions from an unbound to a fully bound state. Jama et al.[20]successfully employed the YANK approach to compute the combined free energy, yielding impressive outcomes.

III. Results

A. *Analysis of Docking*

A virtual screen finally yielded 179 small molecule structures with a binding affinity of -10 kJ/mol with a ring, which was entered as ligands into the Autogrow program for several iterations, resulting in a new small molecule with a binding affinity of -11.4 kJ/mol. The 6LU7 structure is formed by the binding of SARS-CoV-2 Mpro to the inhibitor N3, which is known from It is known from the published literature that the main protease consists of three structural domains, with N3 interacting with its residues to achieve stability. The addition of S on residue Cys145 to vinyl C forms a covalent bond, which plays a key role in inhibiting protein activity. His163, Gly143, Glu166, and Phe140 interact with N3 molecules by forming multiple hydrogen bonds. In N3 molecule, Leu is embedded in His41, Met49, Met165, and Tyr54 to form hydrophobic molecular bags, while Ala is surrounded by the side chains of Met165 and Leu167, resulting in hydrophobic interactions.

The 6LU7 crystal structure was used for recovered molecular docking and the binding pattern of recovered docking was observed in PyMOL. As can be seen in Fig. 3, the natural ligand N3 docked successfully, and the ligand Chembl2_3 docked in the same position.

To analyse the mechanism of interaction between bioactive molecules and enzyme targets, biological activity was predicted by analysing the pattern of intermolecular interactions. To predict the activity of SARS-CoV-2, the important sites in the Mpro of SARS-CoV-2 that form the active site pocket are located in His41, Met49, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166.The recurrence of residues Asn142, Leu141, Ser144 , and Cys145 suggests that they may be required for inhibitor binding[21].

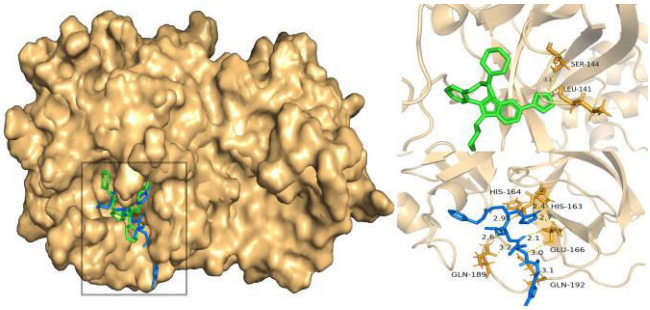

Fig. 1. 3D docking diagram of N3, Chembl2_3 and main protease, blue is N3, green is ligand Chembl2_3.

Fig. 1 shows a visualization of the interaction pattern of the ligand Chembl2_3 with the main protease. The ligand Chembl2_3 forms hydrogen bonding interactions with residues Leu141, Ser144, and hydrogen bonding interactions formed by Gln189, Arg188, Asp187, Met165, His164, Cys145, Thr25, His41, Leu27, Asn142, Met49, Glu166, His163, His172 and Phe140 form hydrophobic interactions. In addition to the traditional hydrogen bonds, there are also carbon-hydrogen bonds.

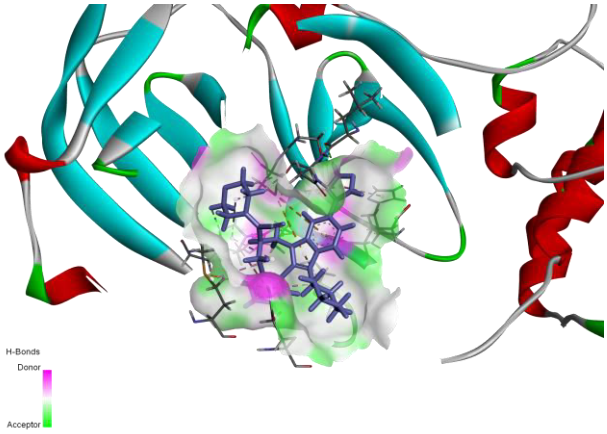

Fig. 2. presents the surface of SARS-CoV-2 Mpro colored by hydrogen bond type in complex with compound Chembl2_3.

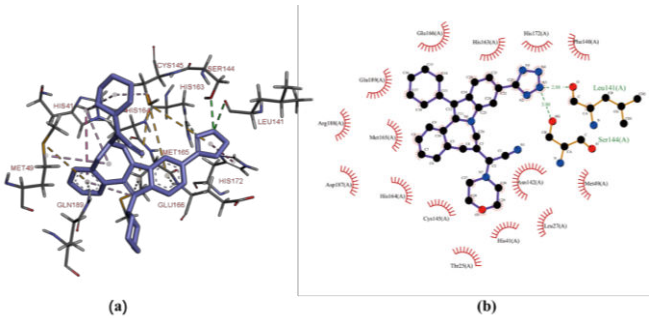

Fig. 3. (a)3D visualization of the interaction between ligand Chembl2_3 and main protease. **(b)** 2D visualization of the interaction between ligand Chembl2_3 and main protease.

B. *Molecular Dynamics Trajectory Analysis*

The RMSD and RMSF parameters can be used to comprehend the volatility and stability of the protein complexes during the simulation. From Fig. 4, it can be observed that the protein-ligand complex is relatively stable throughout the simulation, the small molecule ligand complex is relatively stable in the early stage, and the RMSD image fluctuates between 0.1nm-0.15nm from the beginning of the simulation to around 300ns of the simulation, then it undergoes a large fluctuation around 320ns before leveling off again, and the whole RMSD is below 0.3nm. According to Byura, the conformational stability of protein-ligand complexes is indicated by RMSD values less than 3 Å [22].

The RMSF trajectory is indicative of the flexibility, receptor stability, stiffness, and denseness of the protein during the 500 ns simulation, with higher RMSF values indicating flexible residues and lower RMSF values indicating more stable residues [7]. The RMSF values for the whole main protease ranged from 0 to 0.65 Å. The higher fluctuation of the terminal amino acid residues may be because the N and C termini move faster than the other parts [8]. The solvent-accessible surface area (SASA) calculates the surface area of the protein-ligand complex that can interact with the solvent molecule. A decrease in the SASA value of a protein indicates a decrease in exposure to the solvent and an increase in densification. The SASA values of the protein-ligand complexes were low and remarkably stable throughout the simulation, indicating no structural changes in the complexes and, on the other hand, the high compactness of the complexes. The SASA results indicate that the main protease forms stable complexes with small molecule ligands. The radius of rotation (Rg) is also an indicator of the degree of denseness of the protein-ligand complex, and there is an inverse relationship between the rg value and the protein densities. As can be seen from the graph, the structure of the main protease ligand complex shows a small change in Rg value throughout the simulation, implying no significant conformational changes in the complex system. It can therefore be concluded that the docking conformation of the main protease with the small molecule ligand Chembl2_3 is stable.

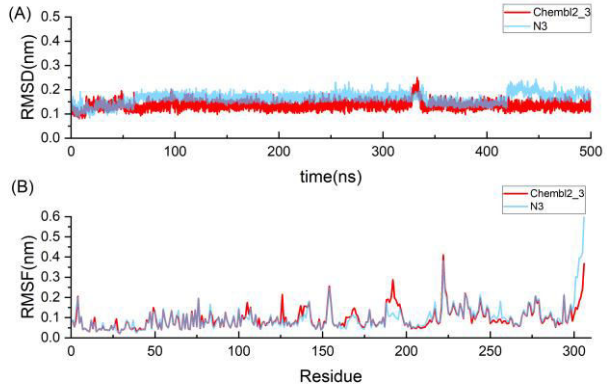

Fig. 4.(A) The RMSD trends of SARS-CoV-2 Mpro in complex with the ligands. **(B)**The RMSF of SARS-CoV-2 Mpro residues in complex with the ligands.

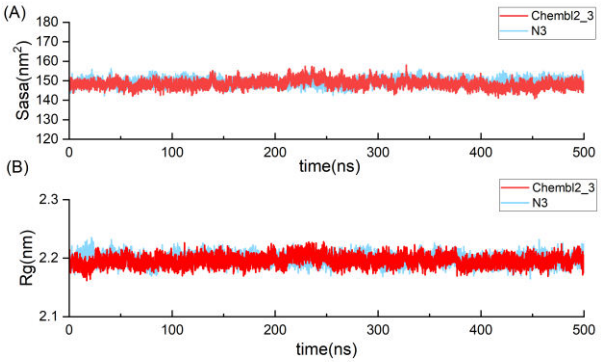

Fig. 5. (A) SASA peaks of SARS-CoV-2 Mpro in complex with the ligands. **(B)** Radius of gyration (Rg) trends of SARS-CoV-2 Mpro in complex with the corresponding ligands.

C. *Free Energy Analysis*

The free energy of binding of the main protease and the original ligand was calculated as a control group, and the results showed that the free energy of the control group was -6.01 kcal/mol, while the free energy of binding of the main protease and the small molecule was -11.56 kcal/mol, which indicated that the binding of the main protease and the small molecule ligand Chembl2_3 was stronger and more stable.

IV. Conclusion

In this study, we employed an integrative approach consisting of virtual screening, molecular docking, and molecular dynamics simulations to identify several small molecules from a comprehensive database with potential inhibitory effects, culminating in the development of a novel compound that exhibits enhanced inhibition of the primary protease. The molecular dynamics simulations underscored the ability of this compound to achieve stable docking with the receptor protease, while the calculated free energy interactions substantiated a robust binding affinity between the ligand and its receptor. We hope the finding can offer optimistic prospects for the ongoing research into effective COVID-19 inhibitory drugs.

References

[1] M.U. Mirza and M. Froeyen. Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. Journal of Pharmaceutical Analysis[J].2020. 10(4): 320-328.

[2] Y. Kaku, K. Okumura, M. Padilla-Blanco, et al. Virological characteristics of the SARS-CoV-2 JN.1 variant. bioRxiv[J]. 2023: 2023.2012.2008.570782.

[3] M.K. Hemo and M.A. Islam. JN.1 as a new variant of COVID-19 - editorial. Ann Med Surg (Lond) [J].2024. 86(4): 1833-1835.

[4] K. Khatua, Y.R. Alugubelli, K.S. Yang, et al. Azapeptides with unique covalent warheads as SARS-CoV-2 main protease inhibitors. Antiviral Research[J].2024. 225: 105874.

[5] S. Li, L. Wang, J. Meng, et al. De Novo design of potential inhibitors against SARS-CoV-2 Mpro. Computers in Biology and Medicine[J]. 2022. 147: 105728.

[6] C. Cao, B. Xu, X. Hu, et al. Research Progress of Antiviral Drug Targeting the Main Protease of SARS-CoV-2. Chinese Journal of Organic Chemistry[J].2022. 42(7).

[7] A.H. Harisna, R. Nurdiansyah, P.H. Syaifie, et al. In silico investigation of potential inhibitors to main protease and spike protein of SARS-CoV-2 in propolis. Biochemistry and Biophysics Reports[J]. 2021. 26: 100969.

[8] D. Dwarka, C. Agoni, J.J. Mellem, et al. Identification of potential SARS-CoV-2 inhibitors from South African medicinal plant extracts using molecular modelling approaches. South African Journal of Botany[J]. 2020. 133: 273-284.
[9] S.Chtita, A. Belhassan, M. Bakhouch, et al.QSAR study of unsymmetrical aromatic disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods. Chemometrics and Intelligent Laboratory Systems[J].2021. 210: 104266.
[10] M. Imran, S. Iqbal, A. Hussain, et al. In silico screening, SAR and kinetic studies of naturally occurring flavonoids against SARS CoV-2 main protease. Arabian Journal of Chemistry[J].2022. 15(1): 103473.
[11] M. Manish, S. Mishra, A. Anand, et al. Computational molecular interaction between SARS-CoV-2 main protease and theaflavin digallate using free energy perturbation and molecular dynamics. Computers in Biology and Medicine[J]. 2022. 150: 106125.
[12] V.H. Masand, M.F. Sk, P. Kar, et al. Identification of Food Compounds as inhibitors of SARS-CoV-2 main protease using molecular docking and molecular dynamics simulations. Chemometrics and Intelligent Laboratory Systems[J]. 2021. 217: 104394.
[13] B. Mercorelli, J. Desantis, M. Celegato, et al. Discovery of novel SARS-CoV-2 inhibitors targeting the main protease Mpro by virtual screenings and hit optimization. Antiviral Research[J]. 2022. 204: 105350.
[14] Schrodinger, Llc, The PyMOL Molecular Graphics System, Version 1.8. 2015.
[15] Schrodinger, Llc, The JyMOL Molecular Graphics Development Component, Version 1.8. 2015.
[16] Schrodinger, Llc, The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8. 2015.
[17] O. Trott and A.J. Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem[J]. 2010. 31(2): 455-461.
[18] C.A. Lipinski. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discovery Today:Technologies[J].2004.1(4): 337-341.
[19] S. Boresch, F. Tettinger, M. Leitgeb, et al. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. The Journal of Physical Chemistry B[J]. 2003. 107(35): 9535-9551.
[20] M. Jama, M. Ahmed, A. Jutla, et al. Discovery of allosteric SHP2 inhibitors through ensemble-based consensus molecular docking, endpoint and absolute binding free energy calculations. Computers in Biology and Medicine[J].2023. 152: 106442.
[21] H. Hadni, A. Fitri, A.T. Benjelloun, et al. Evaluation of flavonoids as potential inhibitors of the SARS-CoV-2 main protease and spike RBD: Molecular docking, ADMET evaluation and molecular dynamics simulations. Journal of the Indian Chemical Society[J]. 2022. 99(10): 100697.
[22] S. Beura and P. Chetti. In-silico strategies for probing chloroquine based inhibitors against SARS-CoV-2. Journal of Biomolecular Structure and Dynamics[J].2021. 39(10): 3747-3759.