

Newly designed analogues from SARS-CoV inhibitors mimicking the druggable properties against SARS-CoV-2 and its novel variants

Nadim Ferdous

Mawlana Bhashani Science and Technology University

Mahjerin Nasrin Reza

Mawlana Bhashani Science and Technology University

Md. Shariful Islam (✉ sharifbge@gmail.com)

University of Kentucky

Md. Tabassum Hossain Emon

Mawlana Bhashani Science and Technology University

Mohamed A. Nassan

Turabah University College, Taif University

Eman Ibrahim Ahmed

Jouf University

Gaber El-Saber Batiha

Damanhour University

A. K.M. Mohiuddin

Mawlana Bhashani Science and Technology University

Mohammad Uzzal Hossain

National Institute of Biotechnology

Research Article

Keywords: SARS-CoV-2, Novel variants, Phytochemicals, Spike glycoprotein, Main protease, Drug discovery

Posted Date: May 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-515155/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The emerging variants of SARS Coronavirus-2 (SARS-CoV-2) has been continuously spreading all over the world and raised global health concerns. The B.1.1.7 (United Kingdom), P.1 (Brazil), B.1.351 (South Africa) and B.1.617 (India) variants resulted due to multiple mutations in the spike glycoprotein (SGp), are resistant to neutralizing antibodies and enable increased transmission. Hence, new drugs might be of great importance against the novel variants of SARS-CoV-2. The SGp and main protease (Mpro) of SARS-CoV-2 are important targets to design and develop antiviral compounds for new drug discovery. In this study, we selected seventeen phytochemicals and later performed molecular docking to determine the binding interactions of the compounds with the two receptors and calculated several drug likeliness properties for each compound. Luteolin, myricetin and quercetin demonstrated higher affinity for both the proteins and interacted efficiently. To get better compounds, we designed three analogues from these compounds and showed the greater druggable properties than the parent compounds. Furthermore, we found that the analogues bind to the residues of both proteins including the recent novel variants of SARS-CoV-2. The binding study was further verified by molecular dynamics (MD) simulation and molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) approaches by assessing the stability of the complexes. MD simulations revealed that Arg457 of SGp and Met49 of Mpro are the most important residues that interacted with the designed inhibitors. Our analysis may give some breakthroughs to develop new therapeutics to treat the proliferation of SARS-CoV-2 in vitro and in vivo.

Introduction

The first case of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was reported on December 30, 2019, in Wuhan city of Hubei province in China [1–4]. The symptoms of COVID-19 caused by SARS-CoV-2 are usually fever, cough, sore throat, breathlessness among others [5]. The World Health Organization (WHO) announced a Public Health Emergency across the globe for this outbreak. The emergence of novel rapidly transmissible variants of SARS-CoV-2 threatens to prolong this pandemic creating a devastating health and economic consequences. This virus is now spread across 219 countries with a total of more than 137 million cases, till the date of this writing. The situation of the South-East Asian countries is also frightening. Most of the low-income countries of this region lack basic health care facilities, thereby are failing to combat the pandemic. As of May 4th, Bangladesh has a death toll of 11,766 with 767,914 confirmed cases, being one of the most severely affected country experiencing the cases of B.1.1.7 [6] and the B.1.351 [7] variants in South-East Asia.

SARS-CoV-2 is a single-stranded positive-sense RNA virus and the genome size is ~30kb, which is the largest among all RNA viruses [8]. SARS-CoV-2 entry into host cells is mediated by the spike (S) glycoprotein which comprises two functional subunits, one (S1 subunit) is responsible for binding to the host cell receptor and the other (S2 subunit) is responsible for fusion of the viral and cellular membranes [9]. It is reported that SARS-CoV-2 uses human angiotensin-converting enzyme 2 (hACE2) as the receptor [10]. Reports from fall 2020 showed a D614G variant of SGp but it quickly became dominant [11]. Recently some emerging variants with rapid transmission capability have discovered in UK, Brazil and

South Africa. The UK variant (B.1.1.7) shares the N501Y mutation with the B.1.351 (South Africa) and the B.1.1.28 (Japan) variant responsible for causing improved affinity of the viral SGp with cellular receptors [6,7]. The B.1.351 variant (20H/501Y.V2) contains three SGp mutations, K417N, E484K, and N501Y [12] while the P.1 (Brazil) variant also contains three substitutions at the same RBD (receptor-binding domain) residues as B.1.351 [13]. Another new lineage of SARS-CoV-2, B.1.617 with combination of L452R and E484Q spike mutations have reported in India [14]. Recent studies showed that the K417N and E484K mutation can enhance the RBD-ACE2 binding, making them more transmissible [15]. These two mutations also help virus escape the therapeutically relevant monoclonal antibodies (mAbs) [15]. While U.K. variant has no obvious effect on mAbs but can also increase the RBD's binding affinity with ACE2[15]. A recent study from California revealed that another mutation from the B.1.617 variant, L452R can increase infectivity [14]. All these mutations of SGp have been raising concerns about increase in number of individuals re-infected by SARS-CoV-2 threatening the efficacy of current vaccines.

The main protease (M^{Pro}) of SARS-CoV-2, also called the 3-C like protease ($3CL^{Pro}$) consists of 306 amino acids and have 3 domains. Domain I and Domain II contribute one residue to a catalytic dyad having Cys145 and His41, then connected by a long loop to Domain III [16]. It is one of the best characterized drug targets among coronaviruses because this enzyme is essential for processing the polyproteins that are translated from the viral RNA [17]. Therefore, SGp and M^{Pro} are attractive drug target to design novel inhibitors preventing viral attachment and replication.

The process of drug discovery requires a multi-disciplinary effort to design effective and commercially feasible drugs against any pathogens. The objective of drug design is to find a chemical compound having able fit both geometrically and chemically to a specific cavity on a target protein[18]. The chemical compounds can be found in nature or synthesized in laboratories. This compound becomes a drug available to patients after passing several animal tests and different phases of human clinical trials [18]. But these methods are high cost and time consuming. In comparison, modern approach including structure-based drug design with the help of computational methods can be carried out within a short period of time which is cost effective and has speeded up the drug discovery process. From the beginning of the COVID-19 pandemic, researchers and computational biologists are trying to find the small molecule inhibitors that can restrict the proliferation of several proteins of SARS-CoV-2 [19–23]. Despite having a large number of natural compounds that can bind and interfere with the significant proteins of SARS-CoV-2, many of them possess toxic properties as well as metabolically unstable, as a result, these compounds do not enter into experimental research over time in wet laboratory. Also, most of the potent compounds were not able to bind with mutation susceptible residues, so upon occurring mutation, the compounds might not inhibit the target proteins. Thus, designing non-toxic compounds from the existing ones capable of inhibiting multiple variants of pathogens remain one of the most important challenges in computer aided drug design (CADD).

Addressing the challenges, we selected seventeen known phytochemicals to identify the potential candidates for novel inhibitor design against SGp (wildtype and mutants) and M^{Pro} of SARS-CoV-2 in this study. Luteolin, myricetin and quercetin showed better results in binding with both the SGp and M^{Pro} and

in drug likeliness properties than the remaining compounds. Considering binding affinities, interactions with target proteins and ADMET profile, we designed three novel inhibitors from these three compounds that outnumbers all the seventeen compounds in interfering with the function of both proteins including the SGp variants. Additionally, we have analyzed and validated the stability of protein-inhibitor complexes using MD simulation and MM/PBSA approach. We also addressed the key residues of both proteins that are crucial in binding of compounds. As no known therapeutics are available for SARS-CoV-2 till date, the results of this study might be valuable references for further experimental research.

Materials And Methods

A step-wise protocol was followed to design the novel inhibitors against SGp and M^{pro} of SARS-CoV-2. The work flow is shown in **Figure 1**.

Selection and preparation of phytochemicals

Seventeen phytochemicals with antiviral activity were selected for this study. The phytochemicals are Aloe-emodin, Amentoflavone, Apigenin, Beta-sitosterol, Betulonic acid, Curcumin, Hesperetin, Hinokinin, Indigo, Isotheaflavin 3'-Gallate, Luteolin, Myricetin, Niclosamide, Quercetin, Savinin, Scutellarein and Theaflavin 3,3'-Digallate. All these compounds were found to have inhibitory efficacy against several proteins of SARS-CoV emerged in 2002 [24]. The Canonical SMILES of these seventeen phytochemicals were collected from the PubChem database. Then they were converted into protein data bank files (PDB) using Online SMILES Translator and Structure File Generator (<https://cactus.nci.nih.gov/translate/>).

Preparation of SGp and M^{pro}

The structure of spike receptor-binding domain (PDB ID: 6LZG) at 2.50 Å resolution and M^{pro} (PDB ID: 6LU7) bound to an inhibitor N3 (N-[(5-methylisoxazol-3-yl)carbonyl]alanyl-L-valyl-N~1~-((1R,2Z)-4-(benzyloxy)-4-oxo-1-[[[(3R)-2-oxopyrrolidin-3-yl]methyl]but-2-enyl]-L-leucinamide) at 2.16 Å resolution of SARS-CoV-2 are available in Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB). These structures were collected and the ligand molecules, hetero atoms and the water molecules were removed from the structure using Discovery Studio 4.0 client (<http://accelrys.com/products/discovery-studio/>). Only the chain B of 6LZG and chain A of 6LU7 were kept for further docking analysis. Later, energy minimizations were performed using SwissPdb Viewer [25] and polar hydrogen atoms were added to all these proteins using AutoDock Tools [26].

Exploration of active site

The active sites of SGp and M^{pro} were explored by the Computed Atlas of Surface Topography of Protein (CASTp) (<http://sts.bioengr.uic.edu/castp/>) server which provides an online resource for locating, depicting and measuring concave surface regions on 3D structures of proteins [27].

Molecular docking of phytochemicals against wildtype (WT) SGp and M^{pro}

Molecular docking simulations were executed by AutoDock tools to determine receptor-ligand interactions. The active site of SGp was enclosed with the parameters of grid box with X=34, Y=34, Z=88 (Center grid box: X = -40.085, Y = 49.129, Z = 13.331; Spacing = 0.347Angstrom) dimensions. In case of the main protease (M^{Pro}), Grid Box was also set to cover all the residues of active site (center X= -11.796, Y = 21.447, Z = 72.229, dimensions X = 60, Y = 64, Z = 60 and spacing = 0.347 Angstrom). AutoDock Vina (ADV) [28] was used to accomplish all the docking simulations with the predetermined parameters. Further, the receptor-ligand interaction was visualized by Pymol 1.1 [29] and Discovery studio 4.0 client.

Validation of molecular docking approach

No respective inhibitors of SGp are available in the PDB database till now. So, for validation of docking purposes, first we removed the N3 inhibitor from the M^{Pro} (6LU7) and re-docked in the active site of 6LU7 using AutoDock Vina to use as positive control. The re-docked pose of N3 was then superimposed on to the reference co-crystallized N3 using PyMOL and the root mean square deviation (RMSD) was calculated. Afterwards, we analyzed the interactions of positive control with the M^{Pro}.

Drug likeliness prediction

Lipinski's rule of five and the bioavailability score of each phytochemical docked with both the SGp and M^{Pro} were calculated on SWISSADME [30]. Drug score was determined from Osiris property explorer (<https://www.organic-chemistry.org/prog/peo/>).

Novel inhibitor design

For improving the antiCOVID-19 activity of the phytochemicals, novel inhibitors were designed by generating their analogs. JSME structure editor, Osiris property explorer and SWISSADME were used to complete the steps. For development of an accurate alignment of the side chains and stereochemistry of designed molecules, all of them were implemented by Yet Another Scientific Artificial Reality Application (YASARA) force field [31]. YASARA conducted molecular corrections of the designed analogs which is vital for structural stability at the molecular level [32–34].

ADMET studies and molecular docking with SGp (WT and mutants) and M^{Pro}

The ADMET (absorption, distribution, metabolism, excretion and toxicity) properties of the designed analogs were determined by pkCSM server [35] and compared with the parent compounds. This is a freely accessible web server (<http://biosig.unimelb.edu.au/pkcsml/>) for in silico ADMET properties analysis of compounds. The designed inhibitors were docked with the SGp (WT and mutants) and M^{Pro}. Considering the rapid transmissibility, we selected the K417N, E484K, N501Y and L452R mutant SGp to test the inhibition efficacy of the designed compounds. The structure of the receptor binding domain of SARS-CoV-2 B.1.351 variant SGp in complex with COVOX-222 and EY6A Fabs (PDB ID: 7NXA) at 2.50 Å resolution is available in the PDB database. The structure was then prepared by removing all the ligand molecules. As no structure of the B.1.617 variant SGp was available in the database, we carried out the

L452R and E484Q mutation in the WT structure (6LZG) using mutagenesis tool of Pymol. Further, site specific docking was performed using AutoDock Vina covering the four mutation sites. Each designed inhibitor was docked with the four SGp mutants and results were analyzed.

Molecular dynamics (MD) simulations and MM/PBSA calculation

Molecular dynamics (MD) simulations were performed to assess the structural stability of protein-inhibitor complexes at an atomistic level. The GROningen MACHine for Chemical Simulations (GROMACS) version 5.1.4 [36] was used to accomplish this task. The 'pdb2gmx' script was used to prepare the protein topologies while ligand topologies were generated from PRODRG server [37]. The energy minimized conformations of complexes were obtained by the GROMOS96 43a1 force field and then solvated with a single point charge (SPC) water model in a rectangular box where every structure was placed in the center at least 1.0 nm from the box edges. The 'gmx genion' script was used to neutralize the net charges in the systems. By employing the steepest descent minimization algorithm, the energy minimization of all the complexes were done with a maximum of 50,000 steps and < 10.0 kJ/mol force. Afterwards, two steps were conducted to equilibrate the systems: a NVT (constant number of particles, volume, and temperature) ensemble and a NPT (constant number of particles, pressure, and temperature) ensemble. Both the NVT and the NPT series were conducted at a 300 K temperature and 1 atm for the duration of 100 picoseconds (ps). V-rescale was selected as the thermostat whereas Parrinello-Rahman was selected as the barostat of the performed simulation. Finally, the production simulation was performed at 300 K for a duration of 120 nanoseconds (ns) in the supercomputing system of National Institute of Biotechnology, Savar, Bangladesh with a time step of 2 fs, and the structural coordinates were saved after every 10 ps. Thereafter, root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), number of hydrogen bonds and solvent accessible surface area (SASA) were analyzed to evaluate the stability of the complexes. The graphs were plotted using the GRACE software. Further, The MM/PBSA binding free energies were calculated by using the 'g_mmpbsa' [38] package of GROMACS followed by final MD run. The binding energies in this method were calculated with the following equation:

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}})$$

Here, the $\Delta G_{\text{binding}}$ = the total binding energy of the complex, G_{protein} = the binding energy of free protein, and G_{ligand} = the binding energy of unbounded ligand.

Results

Predicted Active Site

The active site region of the SGp and M^{Pro} of SARS-CoV-2 were identified by the CASTp server. The preeminent active site of the SGp was found in areas with 74.091 area and a volume of 32.521 amino

acids whereas the best active site of the M^{Pro} was found in areas with 351.125 and a volume of 319.370 amino acids.

Molecular Docking Analysis of phytochemicals with SGp and M^{Pro}

AutoDock Vina predicted nine possible binding positions as output for each phytochemical. Out of nine possible ligand binding positions, the best one was chosen for each phytochemical based on the lowest docking energy. The docking energy score of all the seventeen phytochemicals with both the SGp and M^{Pro} are shown in Supplementary Table S1. The amino acid interactions of both proteins with these phytochemicals were also identified. Twelve phytochemicals out of seventeen interacted with the active site pocket of both the SGp and M^{Pro} and seven were bonded to at least five active amino acid residues of both proteins (Table 1). These seven phytochemicals were considered for further analysis and rest of them were excluded. The selected phytochemicals presented less docking energy ranging from -6.1 to -8.9 Kcal/mol. The N3 showed docking energy of -6.9 Kcal/mol with M^{Pro}. The re-docked N3 was then superimposed onto the native co-crystallized N3 using PyMOL and a low RMSD of 0.226 Å was observed. From the interaction analysis showed in **Figure 2**, we found that the control N3 had similar interaction with M^{Pro} as we noticed in the interactions of co-crystallized N3-M^{Pro} complex. Thr25, His41, Asn142, Gly143, Cys145, Met165 and Gln189 are the interacting amino acids in the active site pocket and about five hydrogen bonds were formed. All the interacted residues of M^{Pro} with control were active site residues and almost all of our selected compounds interacted with these similar residues.

Predicted drug likeliness properties

SwissADME calculated properties based on Lipinski's rule of-five and the bioavailability score, whereas drug score was determined from Osiris Property Explorer, shown in Table 2. Based on higher bioavailability and drug score (bioavailability score >0.50 and drug score >0.30), three phytochemicals, luteolin, myricetin and quercetin were chosen. Further, these three phytochemicals were used as lead compounds to generate three novel analogue inhibitors.

Designed novel inhibitors and ADMET profiling

To improve the binding affinity and ADMET properties of the selected three compounds, we designed three analogues named UN-1 (2-(4-hydroxyphenyl)-4-oxo-1,4-dihydroquinolin-5-carboxylic acid), UN-2 (3-(5-hydroxy-4-oxo-1,4-dihydroquinolin-2-yl)benzoic acid) and UN-3 (5,7-dihydroxy-2-(4-hydroxyphenyl)quinoline-4(1H)-one). Several groups have been replaced with NH, COOH and H groups to generate these analogues shown in **Figure 3**. The ADMET properties of the designed inhibitors were analyzed for their drug likeness, drug score, toxicity, bioavailability etc. shown in Table 3. Among the ADMET properties, the percentage of human intestinal absorption rate has been increased for all the three designed inhibitors, that is about 94% for both UN-1 and UN-2, about 81% for UN-3. All the compounds show Caco-2 permeability in between 0.24 – 0.998 log Papp in 10⁻⁶ cm/s. No compound seems to cross the blood brain barrier as revealed by their negative score ranging from -0.894 to -1.136. Also, no

compound appears to inhibit hERG I, II receptors. Myricetin and quercetin were predicted as mutagenic and the later was also predicted to show tumorigenicity whereas the designed analogues did not show any of these effects. PAINS alert confirmed presence of catechol moieties in the parent compounds whereas no designed inhibitors had any such moieties in their structure. Synthetic accessibility values have been significantly decreased in all the designed inhibitors. Bioavailability has increased for UN-1 and UN-2. Contrarily, drug scores have increased for UN-2 and UN-3. All these results indicate that UN-1, UN-2 and UN-3 are more likely to be used as inhibitors than their parent compounds in terms of pharmacological properties.

Molecular docking analysis of designed inhibitors of SGp (WT and mutants) and M^{PRO}

Prior to the molecular docking simulation of the three designed inhibitors, YASARA energy minimization server was employed to minimize the energy of each compound. The energy comparison between all three designed ligands from the START energy to END energy confirms the energy minimization of structural features. The results of binding from AutoDock vina shows that the three designed inhibitors bind to both SGp (WT) and M^{PRO} with higher affinity than the parent compounds. UN-1, UN-2 and UN-3 showed docking score of -7.0 kcal/mol, -5.9 kcal/mol and -6.5 kcal/mol towards the SGp. It was observed that Arg457, Ser459 and Asp467 were common participant residues in all SGp-inhibitor interactions (Figure 4). Highest number of hydrogen bonding interaction was observed in quercetin-SGp complex with Ser459, Asp467 and Glu471 residues. Ser459 yielded hydrogen bonding with the other two complexes also. The three designed inhibitors showed -7.5 kcal/mol, -6.9 kcal/ml and -7.9 kcal/mol binding energy towards M^{PRO}. The common interaction residues include His163, Cys145 and Met165 (Figure 5). Four hydrogen bonding interactions were observed in case of UN-2 and UN-3 whereas three were formed in case of UN-1 with the active site residues of M^{PRO}. From the docking analysis of designed inhibitors against the four SGp mutants, it was found that UN-1 showed higher binding affinity of -6.3 kcal/mol and -6.1 kcal/mol against the E484K and L452R mutants than the other two designed inhibitors. UN-2 and UN-3 had higher binding affinity of -6.8 kcal/mol and -5.9 kcal/mol against the N501Y and K417N mutants. The N417, K484 and R452 mutant residues formed hydrogen bonding with the inhibitors where highest number of hydrogen bonds were formed in the binding of UN-1 with L452R variant (Figure 6). Physicochemical and ligand binding properties of the designed compounds with SGp (WT) and M^{PRO} are shown in Supplementary Table S2. The docking results of the designed compounds with SGp variants are shown in Supplementary Table S3.

Molecular dynamics simulation

The Molecular dynamics simulation of the ten protein-inhibitor complexes was performed for 120 ns long to analyze the physical movement of atoms at the atomistic level. The dynamic behavior of a protein can give rise to conformational changes which might affect its actual biological functioning [39]. So, in order to examine the dynamic behavior and stability of the complexes throughout the 120 ns simulation period, all the ten complexes were analyzed by calculating the RMSD, RMSF, Rg, SASA and hydrogen bonds shown in Table 5.

Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analysis

The dynamic movements of atoms and conformational variations of C α backbone atoms of the ten complexes were calculated by RMSD to detect their stability. It is observed from Figure 7 that SGp (L452R)-UN-1 complex exhibits the lowest RMSD than other nine complexes. The RMSD of SGp (WT)-inhibitor complexes (Figure 7A1) increases from 50 to 75 ns, however, these values significantly decrease after 80 ns. While assessing the RMSD of SGp (mutant)-inhibitor complexes, the steady increase of RMSD is observed in SGp(E484K)-UN-1 complex after 50 ns. Nonetheless, inhibitors complexed to other mutants deviated lower than 0.5 nm (Figure 7A2). Unlike the SGp-inhibitor complexes, M^{Pro}-UN-3 showed consistent fluctuation from 35-90 ns period indicating that UN-3 might change the protein conformation. The remaining two inhibitors bound M^{Pro} exhibited similar trajectory in the 75-100 ns period and tend to stabilize after 100 ns (Figure 7A3).

To understand the region of proteins that are being fluctuated during the simulation, the flexibility of each residue was calculated in terms of RMSF to get better insight on what extent the binding of designed inhibitors affects the flexibility of proteins. It can be understood from Figure 7B3 that the binding of UN-3 makes the M^{Pro} most flexible in all areas in contrast to other complexes. UN-2 is found to induce local flexibility at Met49, Glu47, Asn72, Asp187, Arg188, Gln192 and Gly302. The SGp (E484K) structure is found to have the lowest RMSF, which indicates that the protein is not very much flexible upon binding of UN-1. Besides, the RMSF values of both the WT and mutant SGp complexes are mostly similar in all areas. Overall, the residues such as Ser371, Pro384, Thr385, Lys386, Lys444, Val445, Gly446, Pro463, Phe464, Glu465, Ser477, Asn501 and Val503 are found flexible for all the inhibitor-bound SGp complexes.

Radius of gyration (Rg) and solvent accessible surface area (SASA)

The compactness of the protein-inhibitor complexes is represented by the radius of gyration (Rg). The lower degree of fluctuation throughout the simulation period indicates the greater compactness of a system. The Rg of WT SGp-UN-1 (Figure 8A1) was found to be the lowest indicating greater rigidity in contrast to the other complexes. Besides, higher Rg of SGp-UN-2 was observed compared to the other two SGp-inhibitor complexes. Phenomena such as protein folding, or unique conformational changes result in greater change in Rg. Inhibitors bound K417N, N501Y and L452R mutant SGp complexes showed similar Rg throughout the simulation while inconsistency was observed in case of SGp (E484K)-UN-1 complex. The three M^{Pro}-inhibitor complexes show increased Rg from 10 to 75 ns, and its last parts of the trend were consistent. Although the Rg value of M^{Pro}-UN-3 was much higher after 75 ns indicating its looseness of packing compared to all other complexes.

Interaction between protein-inhibitor complexes and solvents can be measured by solvent accessible surface area (SASA) over the simulation time. So, SASA of all the ten protein-inhibitor complexes were calculated to predict the extent of the conformational changes occurred during the interaction. The plot of SASA value vs. time for all the protein-inhibitor complexes are shown in Figure 8. Interestingly, SGp(WT)-UN-3, SGp(L452R)-UN-1 and M^{Pro}-UN-3 featured an expansion of the surface area showing relatively

higher SASA value than other complexes. The other inhibitor complexes did not increase in volume during the simulation process although remained in a stable state.

Hydrogen bond analysis

Hydrogen bonding between a protein-inhibitor complex is essential to stabilize the structures. Figure 9 displays the total number of hydrogen bonds present in all the ten complexes calculated after the 120 ns production run. It was observed that the highest number of conformations of SGp (WT) complexes formed up to five hydrogen bonds, mutant SGp complexes formed up to two hydrogen bonds and M^{Pro} complexes formed up to four hydrogen bonds throughout the simulation. UN-1 and UN-2 formed the highest number of hydrogen bonds with SGp (WT) and M^{Pro} respectively during the simulation period. Interestingly, UN-1 and UN-2 also formed highest number of hydrogen bonds with E484K and N501Y mutant SGp complexes.

Post dynamics binding free energy results

The binding free energy (MM/PBSA) of the last 20 ns with an interval of 50 ps was calculated from MD trajectories using MM/PBSA method. Apart from the M^{Pro}-UN-2 complex, all the complexes showed the total binding energies in negative, indicating strong binding (Table 6). In particular, SGp(L452R)-UN-1 complex depicted the lowest binding free energy and higher binding affinity (-214.155 ± 26.323 kJ mol⁻¹), suggesting a more stable conformation of UN-1. The remaining complexes also have favorable binding energy and thus, these designed inhibitors could be used as potential lead compounds. A comparison of the binding free energies of all the ten complexes were made by plotting the binding energy versus time graphs shown in Figure 10.

Further, we examined the contribution of each residue of SGp and M^{Pro} in terms of binding free energy to the interaction with the designed inhibitors that was performed by decomposing the total binding free energy of the system into per residue contribution energy (Figure 10).

Out of the residues that showed interaction in molecular docking analysis, Arg457 residue of SGp exhibited highest contribution energy of -26.9792 kJ/mol, -27.6262 kJ/mol and -3.6108 kJ/mol respectively with UN-1, UN-2 and UN-3. Several residues of M^{Pro} (Met49, CYS145, Met165 and Gln189) also showed favorable contribution energy that interacted with the active site pocket. The Met49 residue showed -9.27 kJ/mol contribution energy in complex with UN-1, UN-2 and UN-3, the contribution energy was -5.05 kJ/mol and -11.4 kJ/mol respectively. Many residues of the four mutants showed favorable contribution energy towards the inhibitors. On comparing the complexes, we found that the mutant K484 and R452 residues involved in the interaction showed -16.8163 kJ/mol and -19.0742 kJ/mol contribution energy in complex with UN-1. This suggests that the binding of UN-1 to these mutants could interfere with the binding of SGp as the residues would not be available to interact with ACE2. The contribution energy of other key residues of SGp (both WT and mutants) and M^{Pro} are shown in Supplementary table S4 and Supplementary table S5.

Discussion

Computer aided drug design (CADD) approaches have been utilizing very widely in order to increase the efficiency of the drug discovery and development course. This method requires three-dimensional (3D) structure of the target proteins, where binding mode, affinity and confirmation of a ligand binding can be detected. Several approaches of CADD are evaluated as promising techniques in modern drug discovery. The key parts in CADD are molecular docking and molecular dynamics simulation that can be used to perform virtual screening of compounds to analyze how the ligands bind and inhibit a target and confirm their stability towards the target proteins [40]. Also, the increasing development of robust computational tools allow scientists to assess pharmacological properties of compounds before entering into experimental research. So, intensive CADD methods were used in this study to design novel inhibitors from against COVID-19.

In this study, we emphasized on both the binding affinity and ADMET properties of compounds to design the potent inhibitors of SGp and M^{Pro}. As recently, some newly emerged variants have been discovered that undergo mutation in various regions, we tried to design the inhibitors capable of binding with both the key active site residues and the mutant residues of two proteins, so that these designed inhibitors can inhibit several strains of SARS-CoV-2. For this, seventeen phytochemicals with antiviral effects were selected and docked against the two proteins. Among them, twelve candidates were then selected according to high binding affinity in the active site pocket of the two proteins. It was found that Isotheaflavin 3'-Gallate exhibit highest binding affinity of -7.2 kcal/mol while binding with the SGp (Supplementary Table S1). Conversely, Theaflavin 3,3'-Digallate exhibit highest affinity of -8.9 kcal/mol and interacted with the highest number of amino acid residues in the active site of the M^{Pro}. As we addressed earlier, despite having high binding affinity of these compounds, we further filtered them based on Lipinski's rule of five and bioavailability score to keep the biologically active compounds for further analysis (Table 2). Luteolin, myricetin and quercetin passed this stage and considered as lead compounds to design novel inhibitors against the two significant proteins of SARS-CoV-2. To our knowledge, this is the first study to design novel inhibitors from these three flavonoids via in silico approaches against both SGp and M^{Pro}.

Among the selected three compounds, luteolin and quercetin, proved their capacity to block the entry of SARS-CoV in a dose-dependent manner (EC₅₀ value of 10.6 μ M) and quercetin showed to prevent HIV-luc/SARS pseudotyped virus entry in host cells (EC₅₀ of 83.4 μ M). Myricetin showed inhibitory effect against SARS-CoV NSP13 ATPase with IC₅₀ values of 2.71 [41]. A study shows that luteolin is toxic to human lung embryonic fibroblasts (TIG-1) and human umbilical vein endothelial (HUVE) cells and quercetin was more toxic to HUVE cells than several flavonoids [42]. Our computational analysis using the up-to-date online-based tools also revealed several toxic properties such as mutagenicity and tumorigenicity as well as low intestinal absorption, inhibitory effect to several isoforms of cytochrome P450 enzyme, presence of catechol group in them (Table 3). Keeping in mind to improve these properties,

we designed three analogues named UN-1, UN-2 and UN-3 and tested their binding affinity towards the two proteins and profiled their ADMET properties.

We found significant increasing in intestinal absorption rate for all the three analogues and they do not cross the blood brain barrier. UN-1 does not inhibit CYP2C9 isoform of cytochrome P450 enzyme unlike its parent compound, which is an important detoxification enzyme of human body. No designed inhibitor shows mutagenic or tumorigenic properties as myricetin and quercetin. There is a very low chance of acquiring long QT syndrome from these designed inhibitors, none appears to inhibit hERG I, II receptors (Table 3). Significant improvement was observed in medicinal chemistry profile of these designed inhibitors also. The Pan-assay interference compounds (PAINS) are chemical compounds that tends to give false positive results in high-throughput screens and react nonspecifically with several biological targets. Luteolin, myricetin and quercetin possessed a catechol group while it was not present in the designed inhibitors revealed from PAINS filter. All the newly designed inhibitors have low score of synthetic accessibility which denotes that these compounds can be synthesized much more easily in laboratories than the parent compounds. Bioavailability of UN-1 and UN-2 has increased slightly and drug score has increased for UN-2 and UN-3 (Table 3).

Upon testing the ADMET properties, we found that all the designed inhibitors exhibited similar and, in some cases, better binding affinities for residues in the SARS-CoV-2 SGp and M^{Pro} active site. Most of the interacted residues of SGp with the three inhibitors include Arg454, Arg457, Lys458 and Ser459 (Supplementary Table S2). Previous research showed that SARS-CoV-2 may interact with host cell surface glycosaminoglycans (GAG) through the SGp to invade host cell and it contains three putative GAG-binding motifs where the 453-459 sequence consists the site 1 [43]. This endorses the theory that the designed compounds may interfere with the interaction of GAGs by binding with the GAG-binding motif. Several residues of SGp variants including the mutants interacted with the designed inhibitors. The mutant residues formed hydrogen bond and carbon hydrogen bond with the inhibitors indicating strong destabilization of the residues and thus, preventing their binding with ACE2. In our docking analysis with M^{Pro}, and most of the compounds interacted with Met49, His163, Cys145 and Met165 (Supplementary Table S2). All the inhibitors bind with Cys145 which is a residue of catalytic dyad of M^{Pro} located in a cleft between the two domains. UN-2 and UN-3, each formed a hydrogen bond with CYS145. All the other interacted residues with the designed inhibitors serve as substrate binding residues of M^{Pro}.

Finally, MD simulation results confirmed the stability of the three designed inhibitors with the SGp, SGp variants and M^{Pro}. The RMSD plot suggests that all the three inhibitors are stable and RMSD values did not show any sudden surge throughout the simulation period (Figure 7). RMSF analysis confirmed that the complexes showed less fluctuations in acceptable range. The radius of gyration (Rg) of all the protein-inhibitor complexes were tend to be in same manner suggesting that each complex had relatively similar behavior of compactness (Figure 8). The SASA values depicted that except the M^{Pro}-UN-3 complex, the remaining complexes did not significantly increase in volume. A good number of hydrogen bonds was observed in all the ten complexes throughout the simulation explaining their conformational stability one

more time (Figure 9). Further, the binding free energies were calculated using MM/PBSA method for all the complexes and results show that five complexes showed favorable binding energy except the M^{Pro}-UN-2 complex. The SGp(L452R)-UN-1 complex showed the lowest binding energy among the others (Figure 10). From, per residue interaction energy profile, it can be concluded that residues Arg454, Arg457, Lys458, Ser469 and Pro491 of SGp and Met49, Cys145, Met165 and Gln189 of M^{Pro} play an essential role in protein-inhibitor stabilization and making significant contributions to the binding of the designed inhibitors. Notable contribution of the mutant residues of SGp variants were also observed confirming that the inhibitors are capable of stabilizing the variants and interfere in stable binding with ACE2 (Figure 10). Thus, these designed inhibitors have the ability to bind with several variants of SARS-CoV-2 and inhibit their activity.

As the present study has been conducted through intensive computational analysis, there might be some limitations. The three designed analogues although showed tremendous results in inhibiting the two proteins, yet they have not been performed in animal model experiment. So, adequate experimental wet lab validations are needed to confirm their therapeutic efficacy.

Conclusion

Our study proposed, designed and tested three novel inhibitors, UN-1, UN-2 and UN-3 against SGp and M^{Pro} of SARS-CoV-2 and its variants. Binding of these inhibitors to the target proteins may eventually create hindrance in entering host cells and in viral replication. Hence, after further in vitro and in vivo experiments, it can be concluded that these designed inhibitors can be used as therapeutics against the SARS-CoV-2.

Declarations

Author Contributions

Conceptualization and Methodology was performed by NF, MR, MN, MSI, AKMM, MUH; Formal analysis and Data curation was performed by TE, NF, MR and MSI; Writing-original draft prepared by NF, MR, EIA, AKMM, MUH and MSI; Writing-review and editing performed by MUH, GB, MN, EIA and MSI; Supervised by MUH and MSI. All authors have read and agreed to submit the final version of the manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Acknowledgments

The authors are grateful to Bioinformatics Division of National Institute of Biotechnology, Bangladesh for providing their supercomputing system to perform molecular dynamics simulation and other necessary equipment required in the study.

Availability of data sets

We have used Online SMILES Translator and Structure File Generator (<https://cactus.nci.nih.gov/translate/>); (Discovery Studio 4.0 client (<http://accelrys.com/products/discovery-studio/>); Topography of Protein (CASTp) (<http://sts.bioengr.uic.edu/castp/>) server which provides an online resource for locating, depicting and measuring concave surface regions on 3D structures of proteins [27]. Osiris property explorer (<https://www.organic-chemistry.org/prog/peo/>); The ADMET (absorption, distribution, metabolism, excretion and toxicity) properties of the designed analogs were determined by pkCSM server [35] (<http://biosig.unimelb.edu.au/pkcsml/>);

Funding

The authors would like to acknowledge and to thank the Taif University Researchers Supporting Project number (TURSP-2020/71), Taif University, Taif, Saudi Arabia.

References

1. Xu X, Chen P, Wang J, et al. 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; 63:457–460
2. Hui DS, I Azhar E, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health – The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int. J. Infect. Dis.* 2020; 91:264–266
3. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J. Autoimmun.* 2020; 109:
4. Bogoch II, Watts A, Thomas-Bachli A, et al. Pneumonia of unknown aetiology in Wuhan, China: Potential for international spread via commercial air travel. *J. Travel Med.* 2020; 27:
5. Singhal T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J. Pediatr.* 2020; 87:281–286
6. Galloway SE, Paul P, MacCannell DR, et al. Emergence of SARS-CoV-2 B.1.1.7 Lineage – United States, December 29, 2020–January 12, 2021. *MMWR. Morb. Mortal. Wkly. Rep.* 2021; 70:95–99
7. Leung K, Shum MH, Leung GM, et al. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance* 2021; 26:2002106
8. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.* 2020; 92:418–423

9. Walls AC, Park YJ, Tortorici MA, et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 2020; 181:281-292.e6
10. Ou X, Liu Y, Lei X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* 2020; 11:
11. Hou YJ, Chiba S, Halfmann P, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* (80-.). 2021; 370:1464–1468
12. Paiva MHS, Guedes DRD, Docena C, et al. Multiple Introductions Followed by Ongoing Community Spread of SARS-CoV-2 at One of the Largest Metropolitan Areas of Northeast Brazil. *Viruses* 2020; 12:1414
13. Wang P, Nair MS, Liu L, et al. Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *Nature* 2021; 1–6
14. Tchesnokova V, Kulakesara H, Larson L, et al. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. *bioRxiv Prepr. Serv. Biol.* 2021; 2021.02.22.432189
15. Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* 2021; 27:622–625
16. Mirza MU, Froeyen M. Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. *J. Pharm. Anal.* 2020; 10:320–328
17. Hilgenfeld R. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.* 2014; 281:4085–4096
18. Baldi A. Computational approaches for drug design and discovery: An overview. *Syst. Rev. Pharm.* 2010; 1:99–105
19. Jaiswal G, Kumar V. In-silico design of a potential inhibitor of SARS-CoV-2 S protein. *PLoS One* 2020; 15:e0240004
20. Prasanth DSNBK, Murahari M, Chandramohan V, et al. In silico identification of potential inhibitors from Cinnamon against main protease and spike glycoprotein of SARS CoV-2. *J. Biomol. Struct. Dyn.* 2020; 1–15
21. Puttaswamy H, Gowtham HG, Ojha MD, et al. In silico studies evidenced the role of structurally diverse plant secondary metabolites in reducing SARS-CoV-2 pathogenesis. *Sci. Rep.* 2020; 10:
22. Kumar Y, Singh H, Patel CN. In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *J. Infect. Public Health* 2020; 13:1210–1223
23. Wei T zi, Wang H, Wu X qing, et al. In Silico Screening of Potential Spike Glycoprotein Inhibitors of SARS-CoV-2 with Drug Repurposing Strategy. *Chin. J. Integr. Med.* 2020; 26:663–669
24. Islam MT, Sarkar C, El-Kersh DM, et al. Natural products and their derivatives against coronavirus: A review of the non-clinical and pre-clinical data. *Phyther. Res.* 2020; 34:2471–2492

25. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 1997; 18:2714–2723
26. Morris GM, Ruth H, Lindstrom W, et al. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 2009; 30:2785–2791
27. Tian W, Chen C, Lei X, et al. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* 2018; 46:W363–W367
28. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 2009; 31:NA-NA
29. Lill MA, Danielson ML. Computer-aided drug design platform using PyMOL. *J. Comput. Aided. Mol. Des.* 2011; 25:13–19
30. Daina A, Michielin O, Zoete V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 2017; 7:
31. Krieger E, Joo K, Lee J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins Struct. Funct. Bioinforma.* 2009; 77:114–122
32. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Bioinforma.* 1993; 17:355–362
33. Vriend G, Sander C. Quality control of protein models: Directional atomic contact analysis. *J. Appl. Crystallogr.* 1993; 26:47–60
34. Kuszewski J, Gronenborn AM, Clore GM. Improvements and Extensions in the Conformational Database Potential for the Refinement of NMR and X-ray Structures of Proteins and Nucleic Acids. *J. Magn. Reson.* 1997; 125:171–177
35. Pires DEV, Blundell TL, Ascher DB. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* 2015; 58:4066–4072
36. Abraham MJ, Murtola T, Schulz R, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015; 1–2:19–25
37. Schüttelkopf AW, Van Aalten DMF. PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2004; 60:1355–1363
38. Kumari R, Kumar R, Lynn A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 2014; 54:1951–1962
39. Bhardwaj VK, Singh R, Sharma J, et al. Identification of bioactive molecules from tea plant as SARS-CoV-2 main protease inhibitors. *J. Biomol. Struct. Dyn.* 2020; 1–10
40. Ferreira LG, Dos Santos RN, Oliva G, et al. Molecular docking and structure-based drug design strategies. *Molecules* 2015; 20:13384–13421
41. Russo M, Moccia S, Spagnuolo C, et al. Roles of flavonoids against coronavirus infection. *Chem. Biol. Interact.* 2020; 328:109211

42. Matsuo M, Sasaki N, Saga K, et al. Cytotoxicity of flavonoids toward cultured normal human cells. *Biol. Pharm. Bull.* 2005; 28:253–259
43. Kim SY, Jin W, Sood A, et al. Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. *Antiviral Res.* 2020; 181:104873

Tables

Table 1: Molecular docking results of the selected seven phytochemicals with SGp and M^{pro}.

Phytochemicals	Binding energy with SGp (kcal/mole)	Interacted residues of SGp	Binding energy with M ^{pro} (kcal/mole)	Interacted residues of M ^{pro}
Aloe-emodin	-6.1	Arg454, Phe456, Arg457, Lys458, Asp467, Glu471	-7.5	His41, Leu141, Asn142, Cys145, Glu166, Arg188,
Isotheaflavin 3'-Gallate	-7.2	Arg454, Arg457, Lys458, Asp467, Ser469	-7.2	Thr26, His41, Ser46, Ser144, Leu141, Asn142, Cys145, His163, Glu166
Luteolin	-7.0	Arg454, Arg457, Lys458, Asp467, Ser469, Glu471	-7.4	His41, Asn142, Cys145, Arg188, Thr190, Gln192
Myricetin	-6.3	Arg454, Arg457, Lys458, Asp467, Glu471	-7.4	Leu141, Gly143, Ser144, Cys145, His163, Met165
Niclosamide	-6.2	Arg454, Arg457, Lys458, Asp467, Ser469, Glu471	-7.0	Thr26, Leu141, Gly143, Ser144, Cys145, Glu166
Quercetin	-6.3	Arg454, Arg457, Lys458, Asp467, Ser469, Glu471	-7.3	Leu141, Gly143, Ser144, Cys145, His163, Met165, Arg188
Theaflavin 3,3'-Digallate	-6.5	Arg454, Phe456, Arg457, Lys458, Asp467, Ser469, Glu471	-8.9	His41, Ser46, Leu141, Gly143, Ser144, Cys145, Met165, Glu166, Gln189, Thr190

Table 2: Bioavailability score and drug score of the selected phytochemicals.

Phytochemicals	Lipinski filter	Bioavailability score	Drug score
Aloe-emodin	Yes; 0 violation	0.55	0.21
Isotheaflavin 3'-Gallate	No; 3 violations: MW>500, NorO>10, NHorOH>5	0.17	0.39
Luteolin	Yes; 0 violation	0.55	0.84
Myricetin	Yes; 1 violation: NHorOH>5	0.55	0.46
Niclosamide	Yes; 0 violation	0.55	0.14
Quercetin	Yes; 0 violation	0.55	0.30
Theaflavin 3,3'-Digallate	No; 3 violations: MW>500, NorO>10, NHorOH>5	0.17	0.31

Table 3: Comparison of ADMET properties, medicinal chemistry profile and drug-likeness of the three designed analogues with their parent compounds.

Properties	Luteolin	UN-1	Myricetin	UN-2	Quercetin	UN-3
Absorption						
Intestinal absorption (human)	81.13	94.757	65.93	94.506	77.207	81.577
Caco2 permeability	0.096	0.687	0.095	0.998	-0.229	0.24
P-glycoprotein substrate	Yes	Yes	Yes	Yes	Yes	Yes
P-glycoprotein I inhibitor	No	No	No	No	No	No
P-glycoprotein II inhibitor	No	No	No	No	No	No
Distribution						
Fraction unbound (human)	0.168	0.186	0.238	0.121	0.206	0.123
BBB permeability	-0.907	-1.032	-1.493	-0.894	-1.098	-1.136
CNS permeability	-2.251	-2.231	-3.709	-2.34	-3.065	-2.327
Metabolism						
Inhibitory substrate to	CYP1A2, CYP2C9	CYP1A2	CYP1A2	CYP1A2	CYP1A2	CYP1A2, CYP2C9
Excretion						
Total Clearance	0.495	0.592	0.422	0.671	0.407	0.545
Renal OCT2 substrate	No	No	No	No	No	No
Toxicity						
AMES toxicity	No	No	No	No	No	No
hERG I inhibition	No	No	No	No	No	No
hERG II inhibition	No	No	No	No	No	No
Mutagenicity	No	No	Yes	No	Yes	No
Tumorigenicity	No	No	No	No	Yes	No
Medicinal Chemistry Profile						
PAINS	1 alert: catechol_A	0 alert	1 alert: catechol_A	0 alert	1 alert: catechol_A	0 alert
Brenk	1 alert: catechol	0 alert	1 alert: catechol	0 alert	1 alert: catechol	0 alert
Lead-likeness	Yes	Yes	Yes	Yes	Yes	Yes
Synthetic accessibility	3.02	2.17	3.27	2.41	3.23	2.25
Drug likeness						
Bioavailability Score	0.55	0.56	0.55	0.56	0.55	0.55
Drug Score	0.84	0.80	0.46	0.51	0.30	0.85

Table 4: Energy minimization score of newly designed inhibitors, UN-1, UN-2 and UN-3.

Inhibitors	START energy (KJ/mol)	END energy (KJ/mol)
UN-1	-294.0	-331.8
UN-2	-677.9	-735.4
UN-3	-711.7	-827.7

Table 5: Average values of RMSD, RMSF, Rg, SASA and number of hydrogen bonds of the ten protein-inhibitor complexes.

Complex	RMSD (nm)	RMSF (nm)	Rg (nm)	SASA (nm ²)	Number of hydrogen bonds
SGp - UN-1	~0.34	~0.41	1.77	~96	~3
SGp - UN-2	~0.32	~0.27	1.81	~97	~1
SGp - UN-3	~0.28	~0.28	1.78	~98	~3
SGp (K417N) - UN-3	~0.23	~0.26	~1.78	~95	~1
SGp (E484K) - UN-1	~0.21	~0.16	~1.78	~98	~2
SGp (N501Y) - UN-2	~0.18	~0.23	~1.79	~98	~2
SGp (L452R) - UN-1	~0.14	~0.23	~1.79	~101	~0.5
M ^{pro} - UN-1	~0.33	~0.39	2.11	~133	~2
M ^{pro} - UN-2	~0.26	~0.63	2.13	~132	~2
M ^{pro} - UN-3	~0.29	~0.40	2.14	~138	~3

Table 6: Binding free energy calculations (MM/PBSA) for ten protein-inhibitor complexes.

Complex	Van der Waal energy (kJ mol ⁻¹)	Electrostatic energy (kJ mol ⁻¹)	Polar solvation energy (kJ mol ⁻¹)	SASA energy (kJ mol ⁻¹)	Binding energy (kJ mol ⁻¹)
SGp-UN-1	-153.355 +/- 14.684	-298.188 +/- 32.518	345.938 +/- 36.742	-13.150 +/- 0.636	-118.756 +/- 18.878
SGp-UN-2	-96.001 +/- 8.894	-104.597 +/- 35.037	71.040 +/- 42.894	-7.920 +/- 0.863	-137.478 +/- 20.273
SGp-UN-3	-184.212 +/- 11.486	-45.100 +/- 9.213	120.643 +/- 16.284	-14.204 +/- 0.853	-122.874 +/- 16.319
SGp(K417N)- UN-3	-111.776 +/- 11.526	-32.936 +/- 8.607	63.459 +/- 14.781	-9.823 +/- 0.737	-91.076 +/- 15.572
SGp(E484K)- UN-1	-93.881 +/- 25.697	-349.662 +/- 86.527	258.331 +/-110.155	-10.181 +/- 2.068	-195.394 +/- 46.276
SGp(N501Y)- UN-2	-75.876 +/- 16.915	-186.460 +/- 65.676	117.115 +/- 74.476	-8.016 +/- 1.498	-153.236 +/- 31.770
SGp(L452R)- UN-1	-97.354 +/- 15.815	-181.414 +/- 60.448	73.267 +/- 65.259	-8.654 +/- 1.224	-214.155 +/- 26.323
M ^{pro} -UN-1	-197.339 +/- 11.919	126.365 +/- 31.436	64.367 +/- 27.287	-14.173 +/- 0.782	-20.780 +/- 15.306
M ^{pro} -UN-2	-139.783 +/- 11.157	139.668 +/- 23.276	50.376 +/- 26.112	-11.851 +/- 0.864	38.410 +/- 14.909
M ^{pro} -UN-3	-180.918 +/- 12.496	-32.152 +/- 14.250	80.529 +/- 18.979	-14.278 +/- 0.898	-146.820 +/- 12.549

Figures

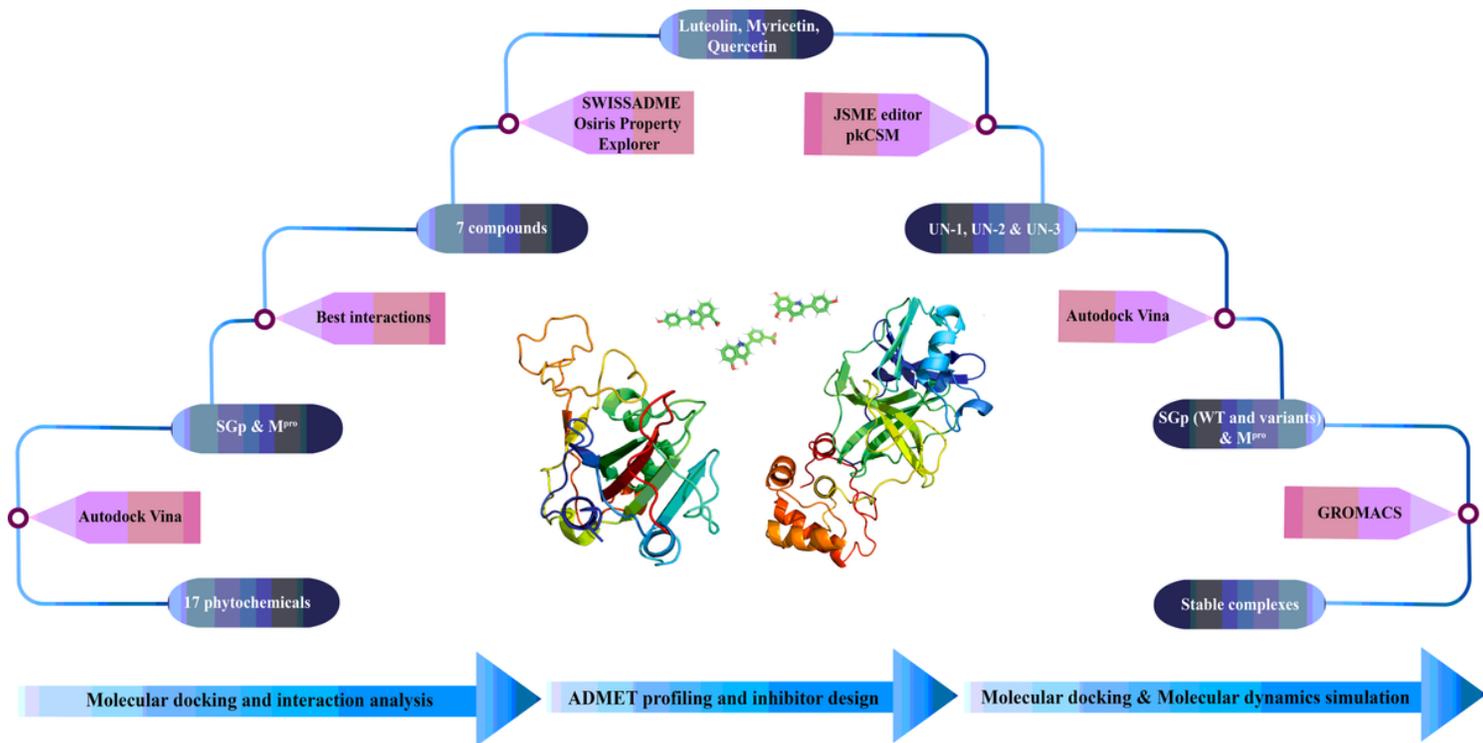


Figure 1

Schematic workflow of designing novel inhibitors against SGp and Mpro of SARS-CoV-2.

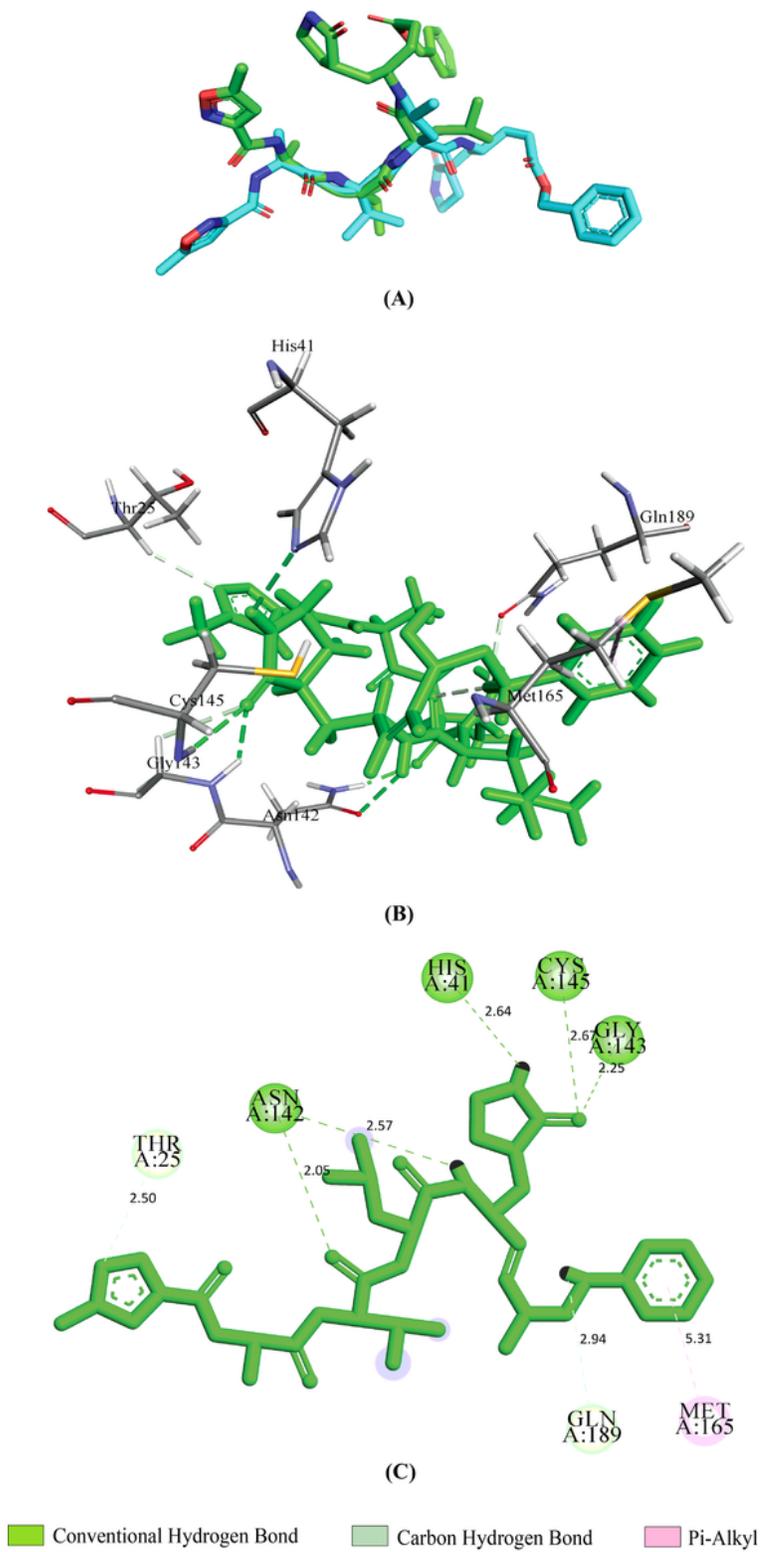


Figure 2

Validation of docking approach by re-docking N3 with 6LU7. Superimposition of re-docked pose of N3 (light green) and the co-crystallized N3 (cyan) from 6LU7 (A), 3D and 2D images depicting interactions (B-C) between N3 and Mpro.

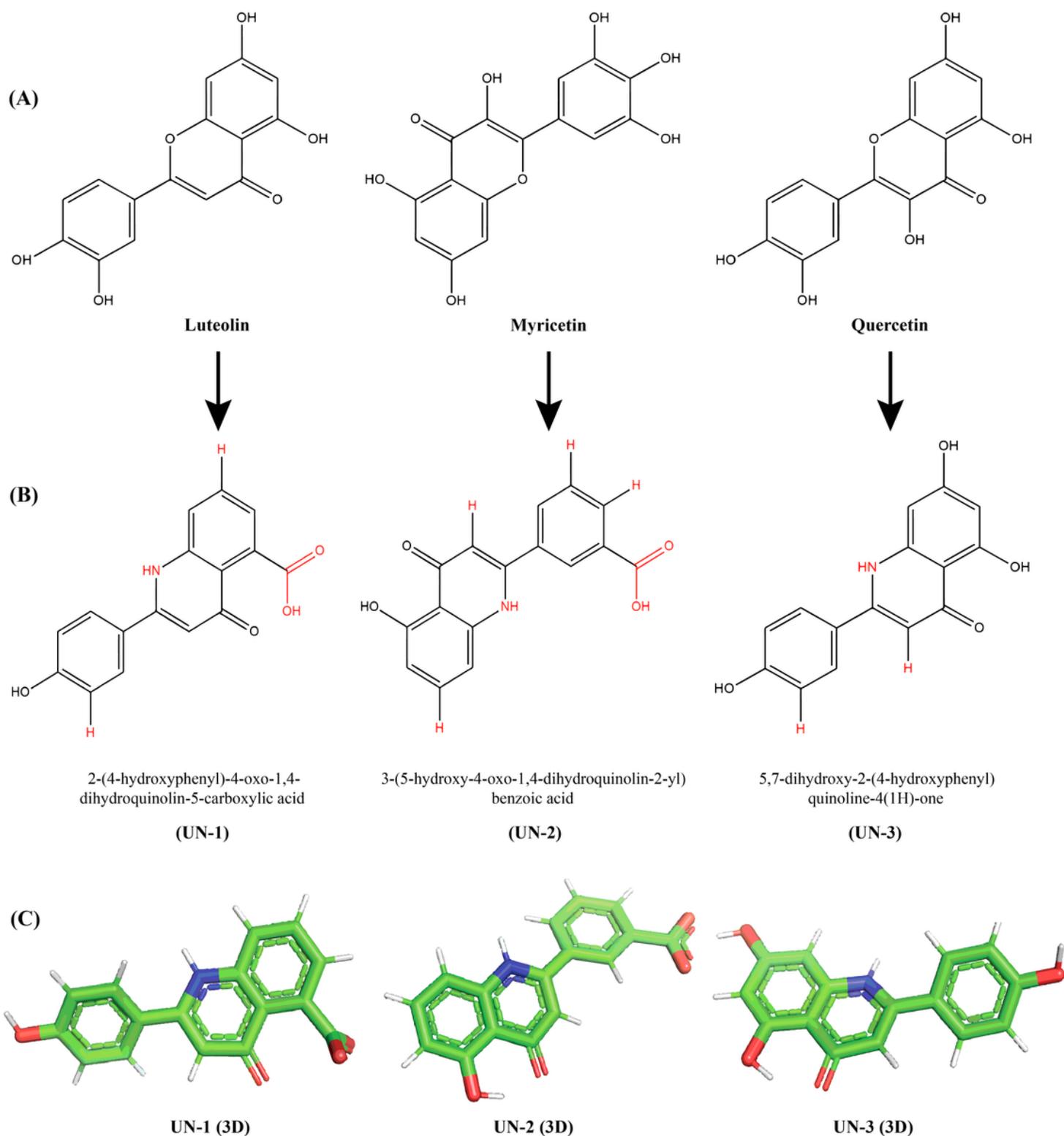
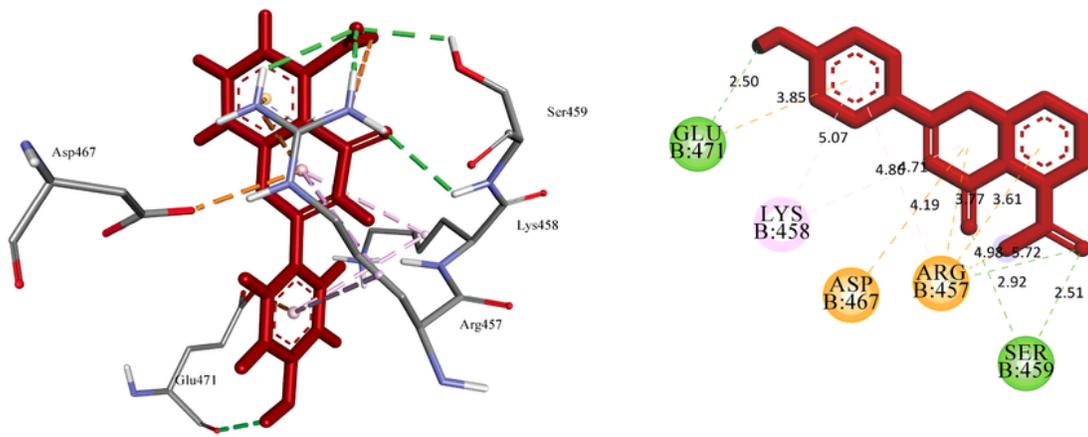
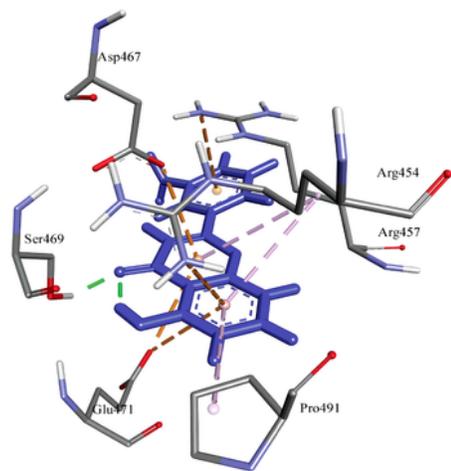
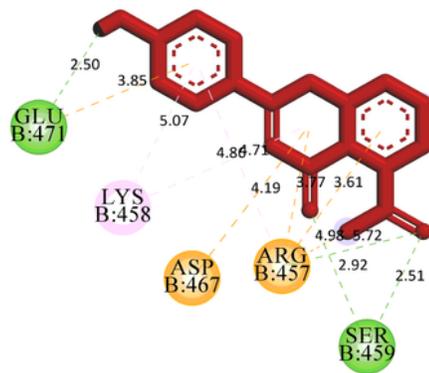


Figure 3

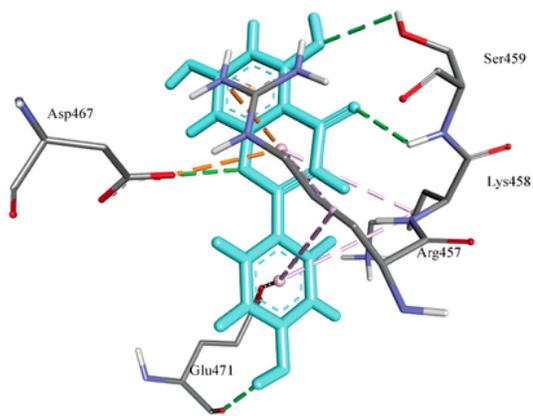
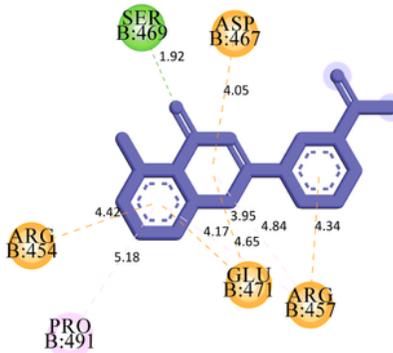
Parent compounds and newly designed analogues. (A) three parent compounds, (B) three designed analogues drawn in Chemdraw Ultra 12.0 (changed groups are shown in red), (C) energy minimized three-dimensional structures of the designed analogues visualized in Pymol.



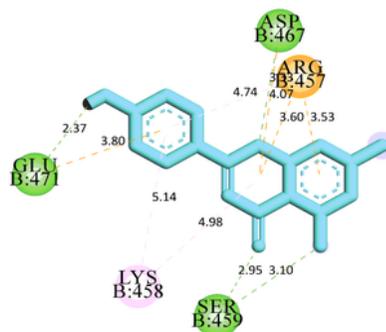
(A)



(B)



(C)



- | | | |
|----------------------------|-----------|-------------------|
| Conventional Hydrogen Bond | Pi-Cation | Attractive charge |
| Carbon Hydrogen Bond | Pi-Anion | Pi-Alkyl |

Figure 4

2D and 3D representation of molecular docking analysis between the SARS-CoV-2 SGp with (A) UN-1, (B) UN-2, (C) UN-3.

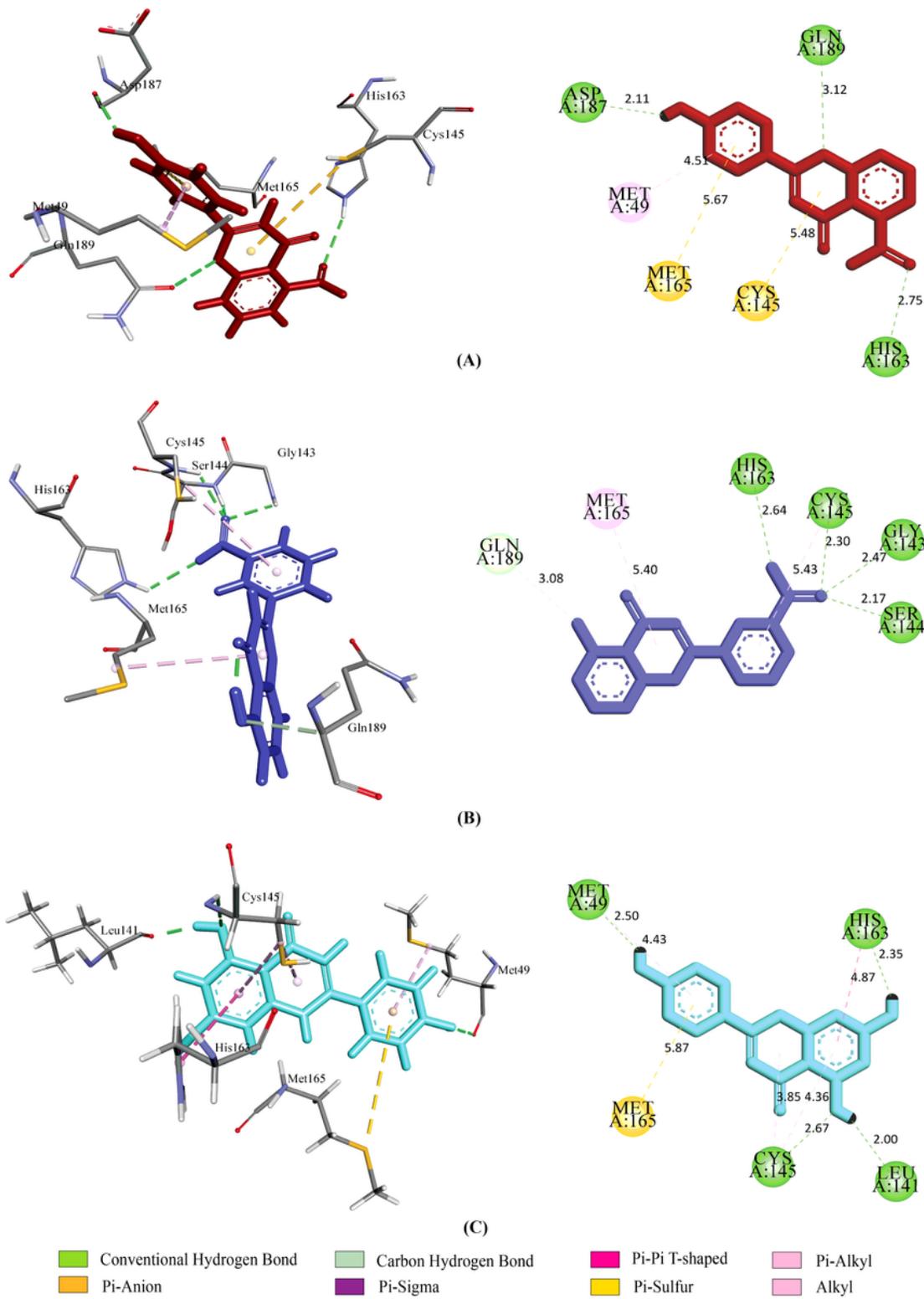


Figure 5

2D and 3D representation of molecular docking analysis between the SARS-CoV-2 Mpro with (A) UN-1, (B) UN-2, (C) UN-3.

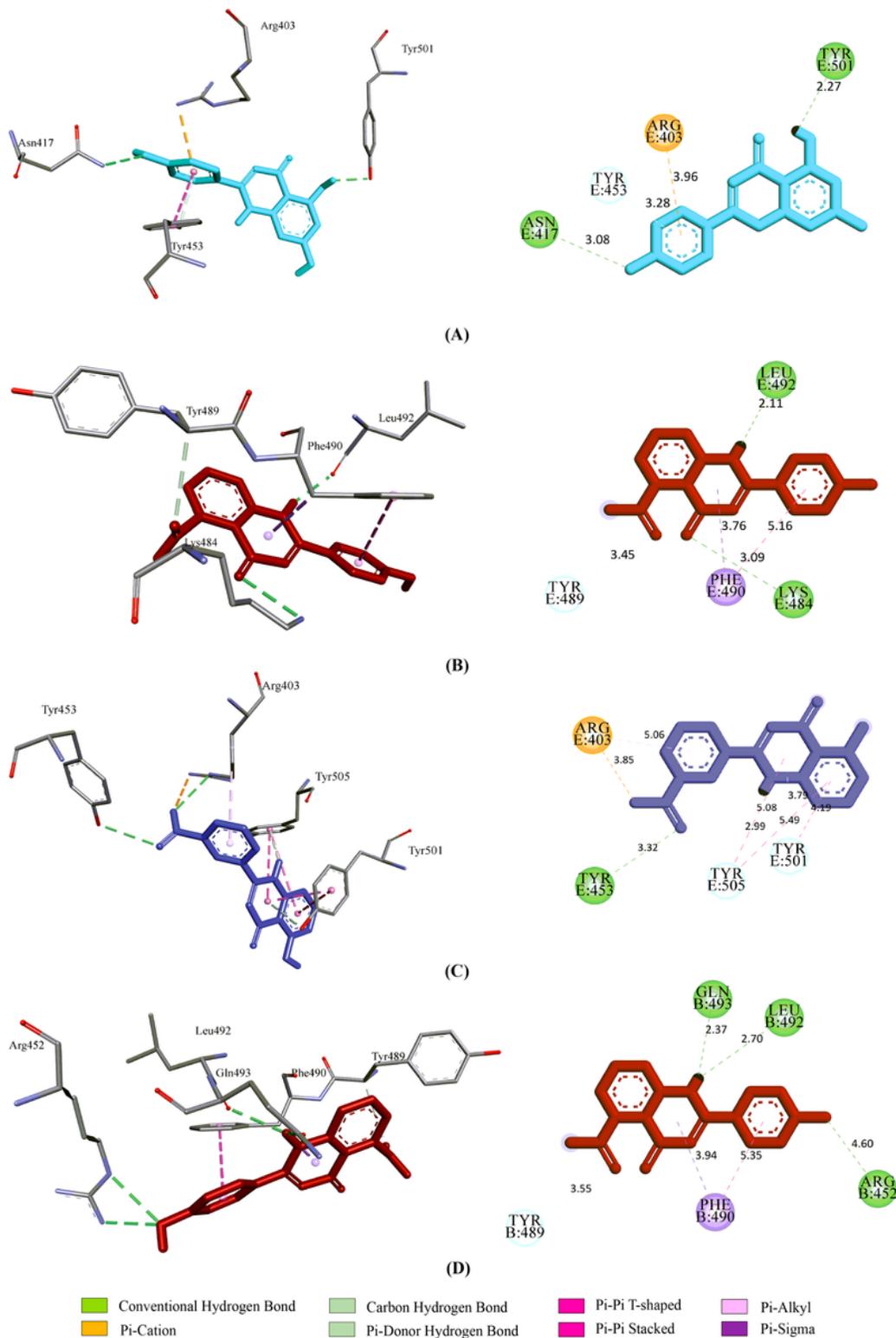


Figure 6

2D and 3D representation of molecular docking analysis between the SARS-CoV-2 SGp mutants and designed inhibitors, (A) SGp (K417N)-UN-3, (B) SGp (E484K)-UN-1, (C) SGp (N501Y)-UN-2, (D) SGp (L452R)-UN-1.

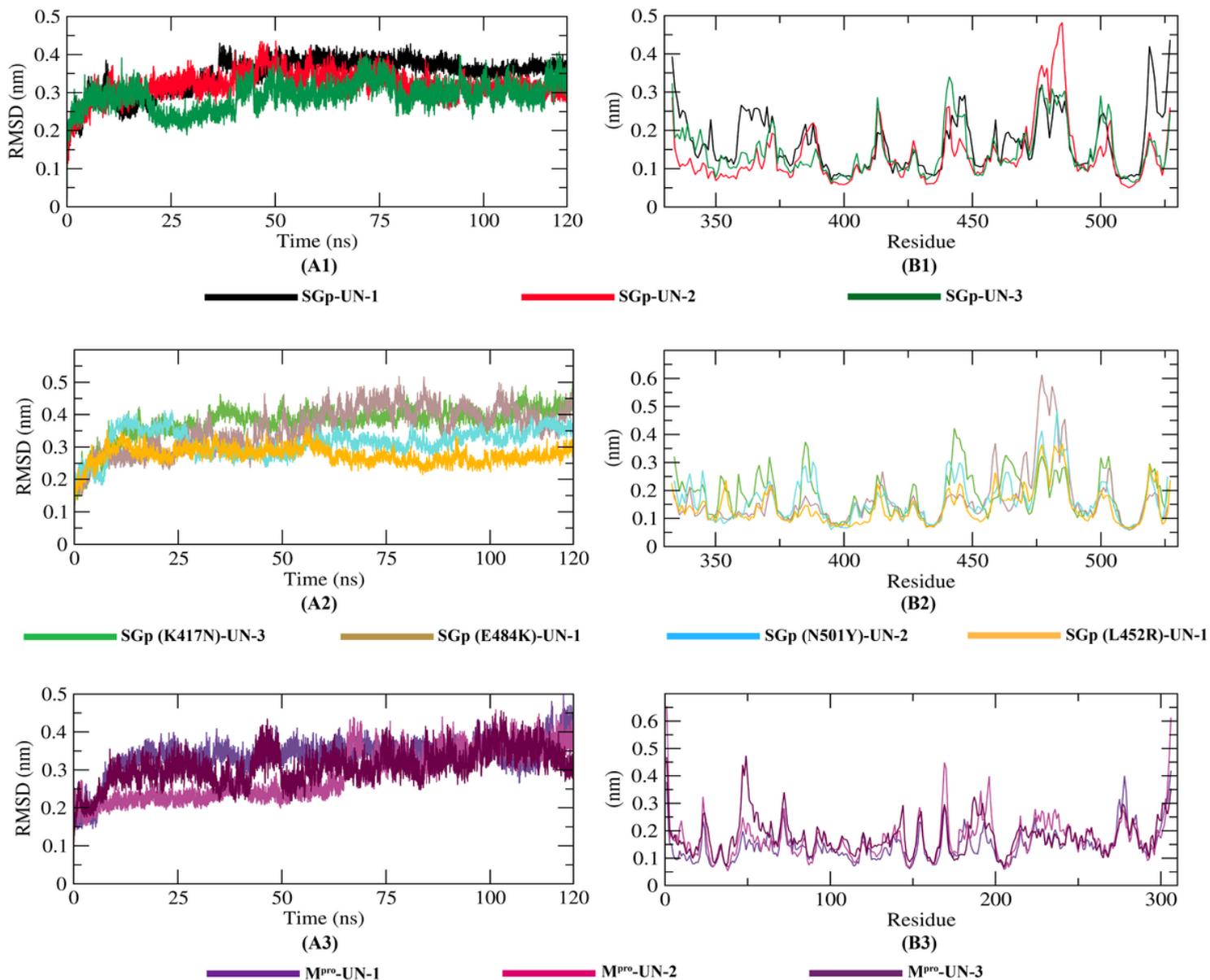


Figure 7

The RMSD and RMSF of Ca atoms of protein-inhibitor complexes. RMSD and RMSF graph of SGp (WT)-inhibitor complexes (A1-B1), SGp (mutant)-inhibitor complexes (A2-B2) and Mpro-inhibitor (A3-B3) complexes from the molecular simulation of 120 ns at 300 K.

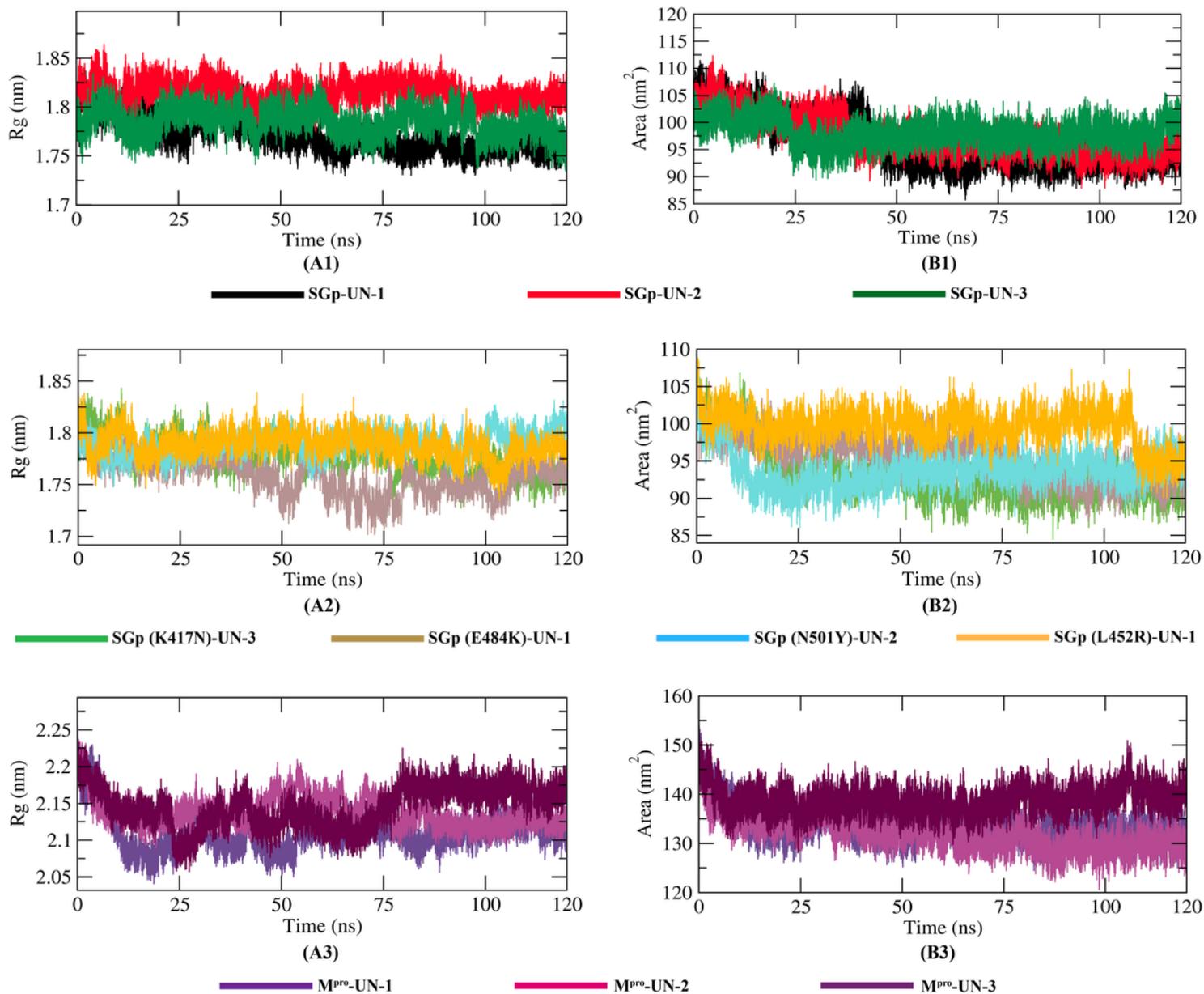


Figure 8

Radius of Gyration (Rg) plot reflecting the compactness of protein-inhibitor complexes (A1-A3) and SASA plot showing the variation in the solvent accessibility of the complexes (B1-B3) during the 120 ns MD simulations.

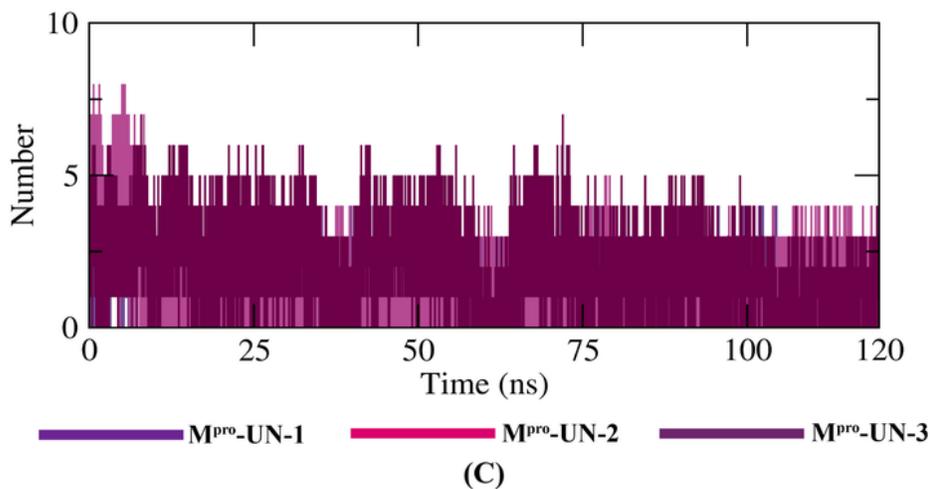
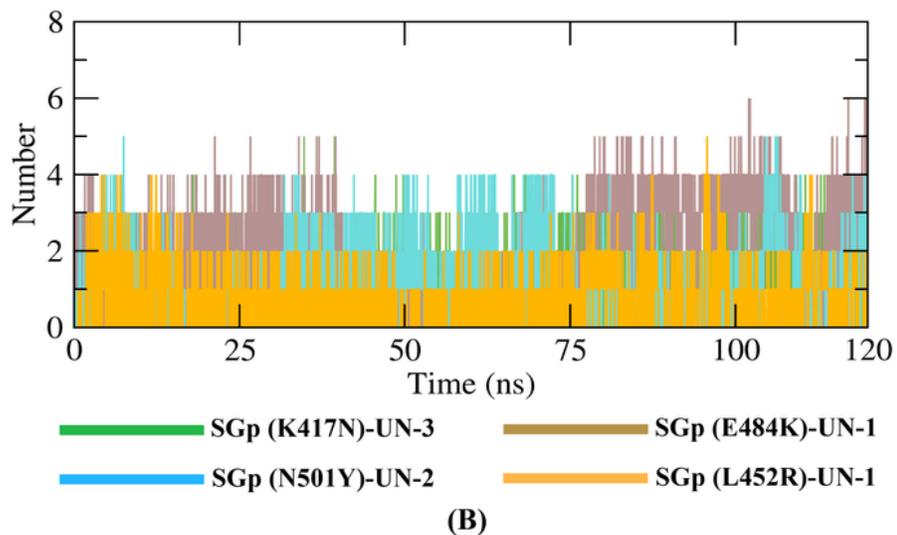
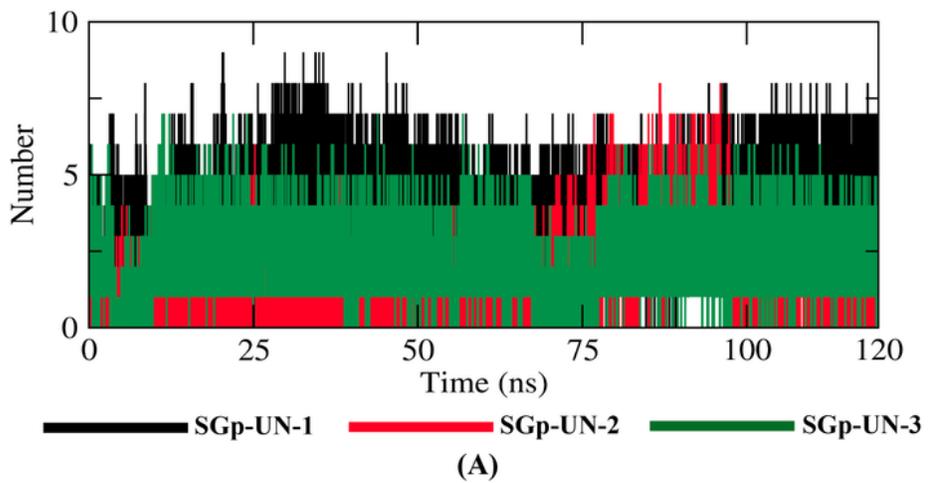


Figure 9

Plot representing the dynamics observed in the hydrogen bonding patterns for the SGp(WT)-inhibitor (A), SGp(mutant)-inhibitor (B) and Mpro-inhibitor (C) complexes.

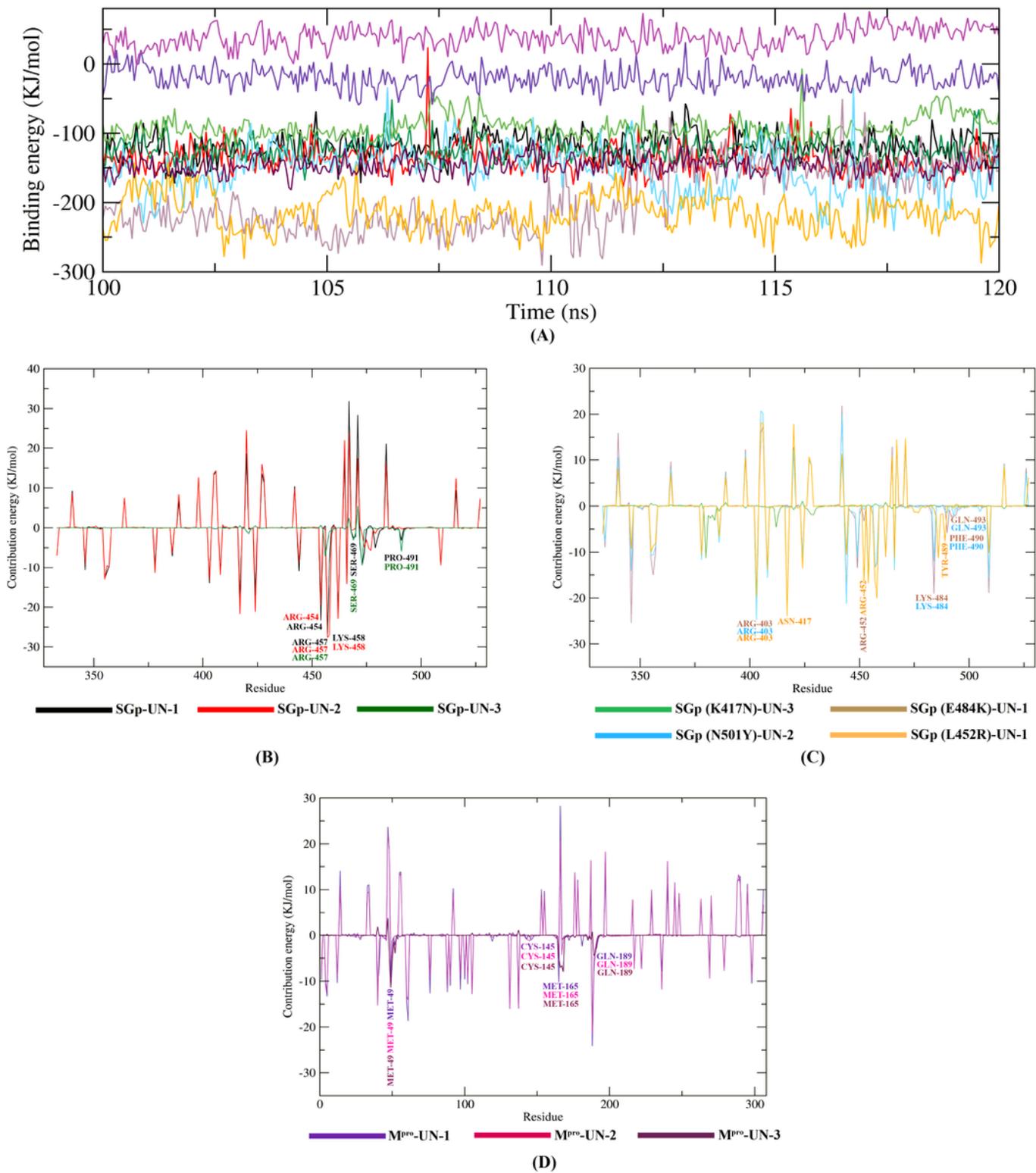


Figure 10

Graphical representation of the binding free energy of protein-inhibitor complexes (A) and per residue contribution plot for SGp-inhibitors (B), SGp(mutant)-inhibitors (C), Mpro-inhibitors (D) complexes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFileSci.Reports.docx](#)