


 Cite this: *RSC Adv.*, 2021, **11**, 31460

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

 Nadim Ferdous,^a Mahjerin Nasrin Reza,^a Md. Shariful Islam, Md. Tabassum Hossain Emon,^a A. K. M. Mohiuddin^a and Mohammad Uzzal Hossain ^{*c}

The emerging variants of SARS coronavirus-2 (SARS-CoV-2) have been continuously spreading all over the world and have 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, resulting from 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 (M^{pro}) of SARS-CoV-2 are important targets for designing and developing 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-likeness properties for each compound. Luteolin, myricetin and quercetin demonstrated higher affinity for both the proteins and interacted efficiently. To obtain compounds with better properties, we designed three analogues from these compounds and showed their greater druggable properties compared to the parent compounds. Furthermore, we found that the analogues bind to the residues of both proteins, including the recently identified 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 M^{pro} are the most important residues that interacted with the designed inhibitors. Our analysis may provide some breakthroughs to develop new therapeutics to treat the proliferation of SARS-CoV-2 *in vitro* and *in vivo*.

 Received 26th May 2021
 Accepted 30th July 2021

 DOI: 10.1039/d1ra04107j
rsc.li/rsc-advances

Introduction

The first case of 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, and 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 devastating health and economic consequences. This virus is now spread across 219 countries with a total of more than 137 million cases, at the date of writing. The situation of the South-East Asian countries is also frightening as most of the low-income countries of this region lack basic health care facilities, so they 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 countries experiencing cases of the B.1.1.7 (ref. 6) and B.1.351 (ref. 7) variants in South-East Asia.

SARS-CoV-2 is a single-stranded positive-sense RNA virus and its genome size is ~30 kb, 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 (the S1 subunit) is responsible for binding to the host cell receptor and the other (the S2 subunit) is responsible for fusion of the viral and cellular membranes.⁹ It is reported that SARS-CoV-2 utilizes human angiotensin-converting enzyme 2 (hACE2) as the receptor.¹⁰ Reports from fall 2020 revealed the presence of a D614G variant of SGp and it quickly became

^aDepartment of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh

^bDepartment of Biology, University of Kentucky, 101 T.H. Morgan Building, Lexington, KY 40506-022, USA

^cBioinformatics Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka-1349, Bangladesh. E-mail: uzzalbge10044@gmail.com

† Electronic supplementary information (ESI) available: Table S1: molecular docking analysis of seventeen phytochemicals against WT SGp (6LZG) and M^{pro} (6LU7). Table S2: physicochemical and ligand binding properties of luteolin, myricetin, quercetin and the three designed inhibitors. Table S3: ligand binding properties of the three designed inhibitors with SGp variants. Table S4: per residue contribution energy of SGp (WT and mutants) with UN-1, UN-2 and UN-3. Table S5: per residue contribution energy of M^{pro} with UN-1, UN-2 and UN-3. See DOI: 10.1039/d1ra04107j



dominant.¹¹ Recently some emerging variants with rapid transmission capability have been discovered in the 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) variants responsible for causing improved affinity of the viral SGp with cellular receptors.^{6,7} The B.1.351 and P.1 (Brazil) variants (20H/501Y.V2) contain three SGp mutations, K417N, E484K, and N501Y.¹² Another new lineage of SARS-CoV-2, B.1.617 with a combination of L452R and E484Q spike mutations has been reported in India.¹³ Recent studies have shown that the K417N and E484K mutations can enhance RBD-ACE2 binding, making them more transmissible.¹⁴ These two mutations also help the virus escape therapeutically relevant monoclonal antibodies (mAbs).¹⁴ Meanwhile, the UK variant has no obvious effect on mAbs but can also increase the RBD's binding affinity with ACE2.¹⁴ A recent study from California revealed that another mutation from the B.1.617 variant, L452R, can increase infectivity.¹³ All these mutations of SGp have been raising concerns about an 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 has 3 domains. Domain I and domain II contribute one residue to a catalytic dyad with Cys145 and His41, which is then connected by a long loop to domain III.¹⁵ 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.¹⁶ Therefore, SGp and M^{Pro} are attractive drug targets for designing novel inhibitors in order to prevent 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 able to fit both geometrically and chemically to a specific cavity on a target protein.¹⁷ 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.¹⁷ But these methods are high cost and time consuming. In comparison, a modern approach including structure-based drug design with the help of computational and *in silico* 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 have been trying to find the small molecule inhibitors that can restrict the proliferation of several proteins of SARS-CoV-2.¹⁸⁻²² 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 being metabolically unstable; as a result, these compounds do not enter into experimental research over time in a wet laboratory. Also, most of the potent compounds were not able to bind with mutation-susceptible residues, so upon mutation occurring, the compounds might not inhibit the target proteins. Thus, designing non-toxic compounds from existing ones capable of inhibiting multiple variants of pathogens remains one of the

most important challenges in computer aided drug design (CADD).

Addressing these challenges, we selected seventeen known phytochemicals to identify the potential candidates for designing novel inhibitors against SGp (wildtype and mutants) and M^{Pro} of SARS-CoV-2 in this study. Luteolin, myricetin and quercetin showed better results in terms of binding with both SGp and M^{Pro} and in drug-likeness properties than the remaining compounds. Considering binding affinities, interactions with target proteins and the ADMET profile, we designed three novel inhibitors from these three compounds that outnumber all 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 the MM/PBSA approach. We also addressed the key residues of both proteins that are crucial for binding of compounds. As no known therapeutics are available for SARS-CoV-2 to 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 Fig. 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 that emerged in 2002.²³ The Canonical SMILES of these seventeen phytochemicals were collected from the PubChem database and converted into protein data bank files (PDB) using the Online SMILES Translator and Structure File Generator (<https://cactus.nci.nih.gov/translate/>).

Preparation of SGp and M^{Pro}

The structure of the 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 the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB). These structures were collected and the ligand molecules, hetero atoms and water molecules were removed from the structure using Discovery Studio 4.0 client (<http://www.accelrys.com/products/discovery-studio/>). Only chain B of 6LZG and chain A of 6LU7 were kept for further docking analysis. Later, energy minimizations were performed using SwissPdb Viewer²⁴ and polar

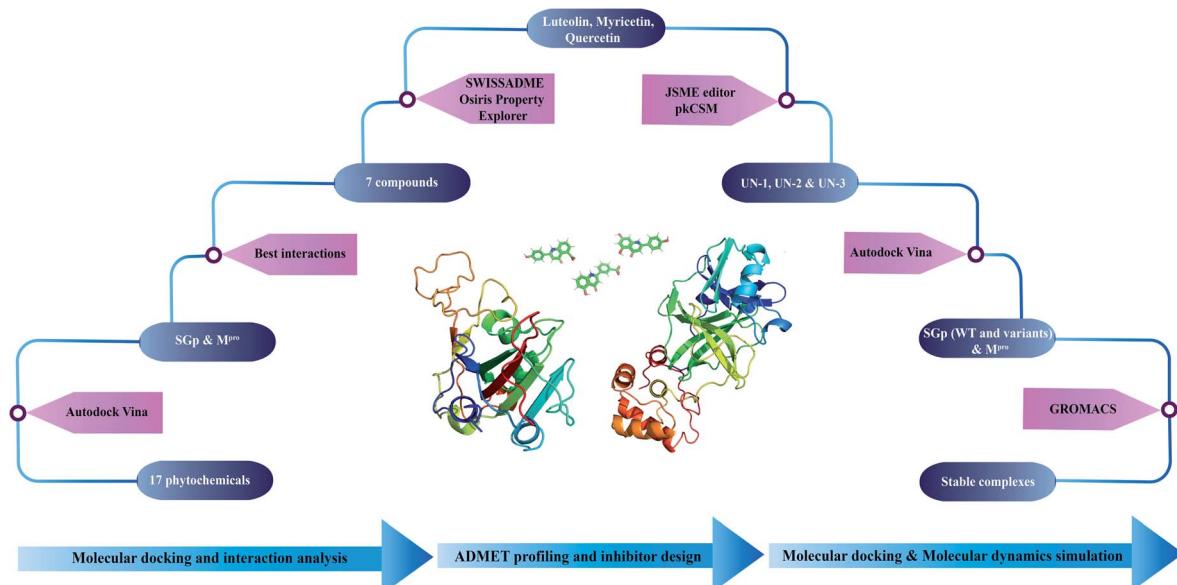


Fig. 1 Schematic workflow of designing novel inhibitors against SGp and M^{pro} of SARS-CoV-2.

hydrogen atoms were added to all these proteins by employing AutoDock Tools.²⁵

Exploration of active site

The active sites of SGp and M^{pro} were explored with the Computed Atlas of Surface Topography of Protein (CASTp) (<http://www.sts.bioengr.uic.edu/castp/>) server which provides an online resource for locating, depicting and measuring concave surface regions on 3D structures of proteins.²⁶

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

Molecular docking simulations were executed with AutoDock tools to determine receptor–ligand interactions. The active site of SGp was enclosed with the parameters of a grid box with $X = 34$, $Y = 34$, $Z = 88$ (center grid box: $X = -40.085$, $Y = 49.129$, $Z = 13.331$; spacing = 0.347 Å) dimensions. In the case of the main protease (M^{pro}), a grid box set to cover all the residues of the active site (center $X = -11.796$, $Y = 21.447$, $Z = 72.229$, dimensions $X = 60$, $Y = 64$, $Z = 60$ and spacing = 0.347 Å). (AutoDock Vina)²⁷ was used to accomplish all the docking simulations with the predetermined parameters. Further, the receptor–ligand interaction was visualized using PyMOL 1.1 (ref. 28) and Discovery studio 4.0 client.

Validation of molecular docking approach

As no respective inhibitors of SGp are available in the PDB database up to now, for validation of docking purposes, we first removed the N3 inhibitor from the M^{pro} (6LU7) and re-docked it in the active site of 6LU7 using AutoDock Vina to act as a positive control. The re-docked pose of N3 was then superimposed onto the reference co-crystallized N3 using PyMOL and the root

mean square deviation (RMSD) was calculated. Afterwards, we analyzed the interactions of a positive control with M^{pro}.

Drug-likeness prediction

Molecular properties based on Lipinski's rule of five and the bioavailability score of each phytochemical docked with both SGp and M^{pro} were calculated on SWISSADME.²⁹ Drug scores were determined with the Osiris property explorer (<https://www.organic-chemistry.org/prog/peo/>).

Novel inhibitor design

In order to improve the anti-COVID-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 the development of an accurate alignment of the side chains and stereochemistry of the designed molecules, all of them were implemented with the Yet Another Scientific Artificial Reality Application (YASARA) force field.³⁰ YASARA conducted molecular corrections of the designed analogs which is vital for structural stability at the molecular level.^{31–33}

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 with the pkCSM server³⁴ and compared with the parent compounds. This is a freely accessible web server (<http://biosig.unimelb.edu.au/pkcs/>) for analyzing *in silico* ADMET properties of compounds. The designed inhibitors were docked with 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 prepared by removing all the ligand molecules. As no known crystal structure of the B.1.617 variant of SGp was available in the database, we carried out the L452R and E484Q mutations in the WT structure (6LZG) using the mutagenesis tool of PyMOL. Further, site-specific docking was performed using Vina covering the four mutation sites. Each designed inhibitor was docked with the four SGp mutants and the 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 (ref. 35) was used to accomplish this task. The ‘pdb2gmx’ script was used to prepare the protein topologies while ligand topologies were generated from the PRODRG server.³⁶ The energy-minimized conformations of the complexes were obtained with 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 was undertaken with a maximum of 50 000 steps and <10.0 kJ mol⁻¹ force. Afterwards, two steps were conducted to equilibrate the systems: an NVT (constant number of particles, volume, and temperature) ensemble and an NPT (constant number of particles, pressure, and temperature) ensemble. Both the NVT and the NPT series were conducted at 300 K temperature and 1 atm for a duration

of 100 picoseconds (ps). V-rescale was selected as the thermostat and 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 the 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 (R_g), number of hydrogen bonds and solvent accessible surface area (SASA) were analyzed to evaluate the stability of the complexes. The graphs were plotted using GRACE software. Further, the MM/PBSA binding free energies were calculated using the ‘g_mmpbsa’³⁷ package of GROMACS followed by a 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, $\Delta G_{\text{binding}}$ = 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 regions of the SGp and M^{pro} of SARS-CoV-2 were identified with the CASTp server. The preeminent active site of SGp was found in areas with 74.091 and a volume of 32.521 amino acids whereas the best active site of 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}

Vina predicted nine possible binding positions as output for each compound. Out of nine possible ligand binding positions,

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

Phytochemicals	Binding energy with SGp (kcal mol ⁻¹)	Interacted residues of SGp	Binding energy with M ^{pro} (kcal mol ⁻¹)	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

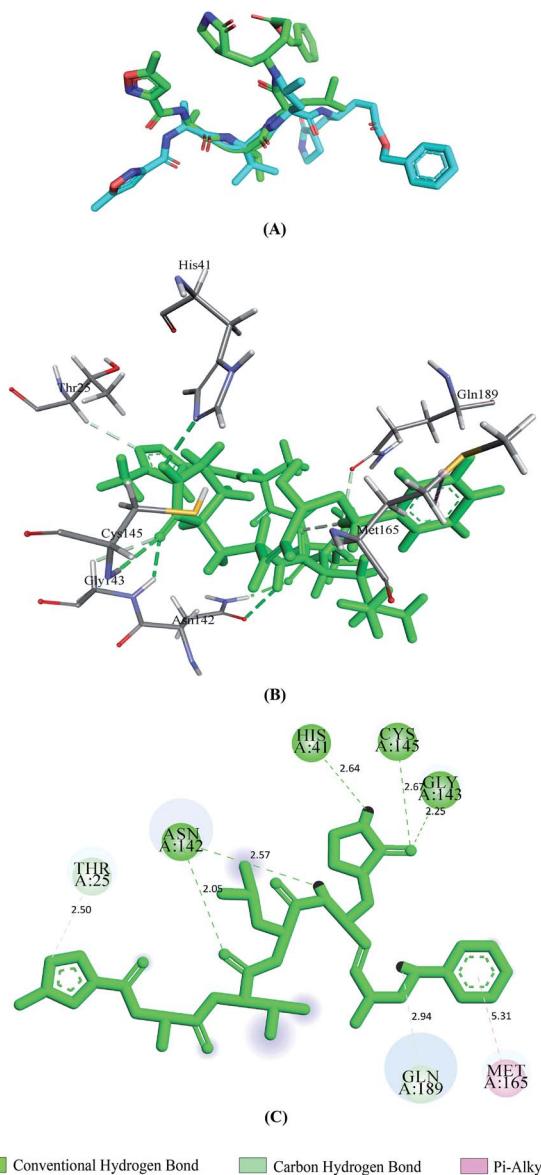


Fig. 2 Validation of docking approach by re-docking N3 with 6LU7. Superimposition of re-docked pose of N3 (light green) and the co-crystallized N3 (cyan) from 6LU7 (A), 3D and 2D images depicting interactions (B and C) between N3 and M^{pro}.

the best one was chosen for each compound based on the lowest docking energy. The docking energy score of all seventeen phytochemicals with both SGp and M^{pro} are shown in Table S1.†

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, N or O > 10, NH or OH > 5	0.17	0.39
Luteolin	Yes; 0 violation	0.55	0.84
Myricetin	Yes; 1 violation: NH or OH > 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, N or O > 10, NH or OH > 5	0.17	0.31

The amino acid interactions of both proteins with these phytochemicals were also identified. Twelve phytochemicals out of seventeen showed favorable interaction with the active site pocket of both 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, excluding the rest. The selected phytochemicals presented lower docking energy ranging from -6.1 to -8.9 kcal mol⁻¹. The N3 showed a docking energy of -6.9 kcal mol⁻¹ 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 shown in Fig. 2, we found that the control N3 had similar interaction with M^{pro} to that we noticed in the interactions of the 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 the control were active site residues and almost all of our selected compounds interacted with these similar residues.

Predicted drug-likeness properties

The SwissADME web server calculated properties based on Lipinski's rule of five and the bioavailability score, while the drug score was determined from the Osiris Property Explorer, as 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 three selected 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, as shown in Fig. 3. The ADMET properties of the designed inhibitors were analyzed for their drug likeliness, drug score, toxicity, bioavailability *etc.*, as shown in Table 3. Among the ADMET properties, the percentage of human intestinal absorption rate has increased for all three designed inhibitors, that is about 94% for both UN-1 and UN-2, about 81% for UN-3. All the



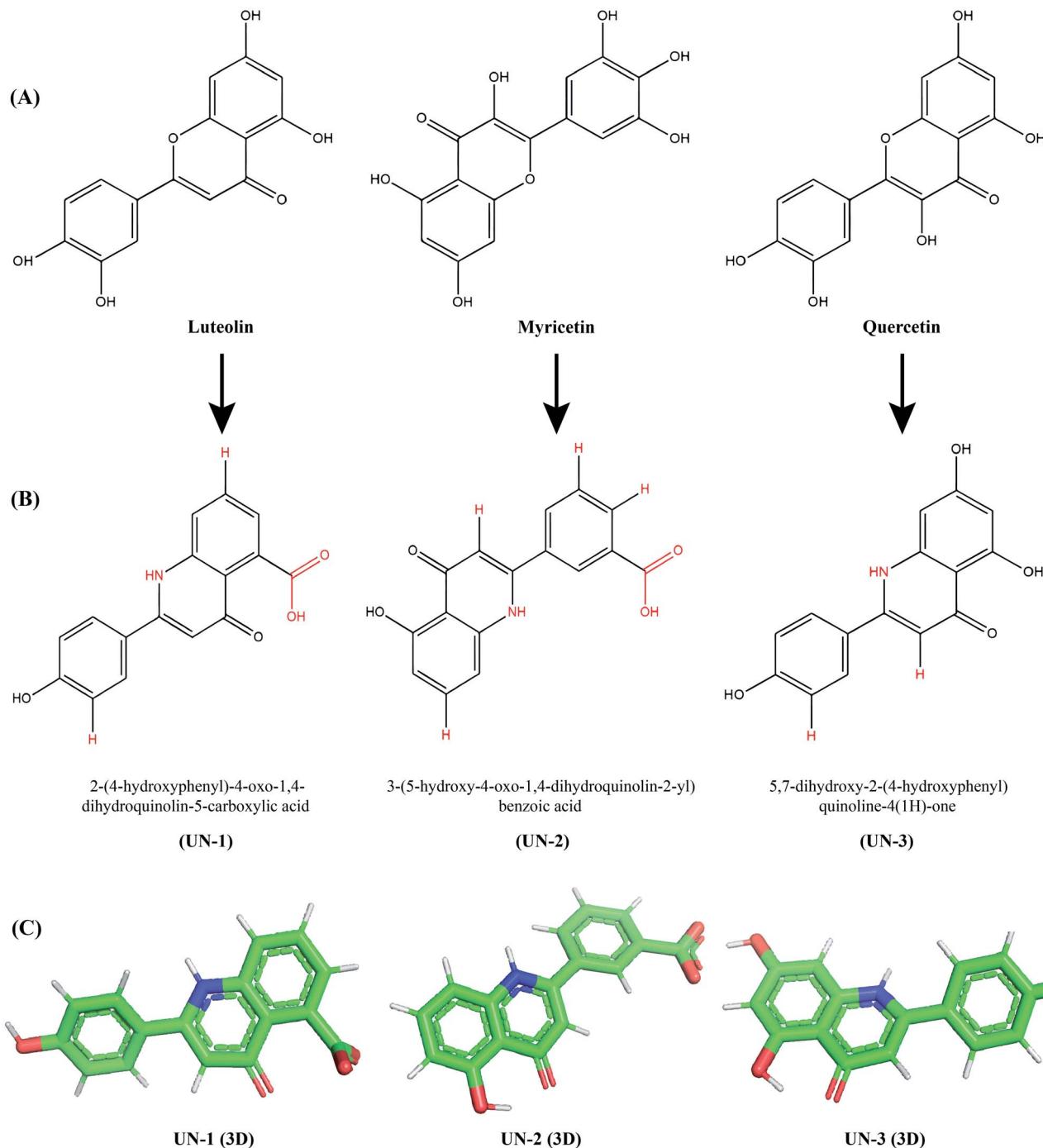


Fig. 3 Parent compounds and newly designed analogues: (A) three parent compounds, (B) three designed analogues drawn in ChemDraw Ultra 12.0 (changed groups are shown in red), (C) energy-minimized three-dimensional structures of the designed analogues visualized in PyMOL.

compounds show Caco-2 permeability between 0.24 and 0.998 $\log P_{app}$ in 10^{-6} cm s^{-1} . No compound seems to cross the blood–brain barrier, as revealed by their negative scores ranging from -0.894 to -1.136 . Also, no compound appears to inhibit hERG I, II receptors. Myricetin and quercetin were predicted to be mutagenic and the latter was also predicted to show tumorigenicity whereas the designed analogues did not show any of these effects. A PAINS alert confirmed the presence of catechol moieties in the parent compounds whereas no

designed inhibitors had any such moieties in their structure. The synthetic accessibility values have significantly decreased in all the designed inhibitors. The bioavailability score has increased for UN-1 and UN-2. In contrast, 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.

Table 3 Comparison of ADMET properties, medicinal chemistry profile and drug likeliness 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
<i>P</i> -Glycoprotein substrate	Yes	Yes	Yes	Yes	Yes	Yes
<i>P</i> -Glycoprotein I inhibitor	No	No	No	No	No	No
<i>P</i> -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 likeliness						
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 ⁻¹)	End energy (kJ mol ⁻¹)
UN-1	-294.0	-331.8
UN-2	-677.9	-735.4
UN-3	-711.7	-827.7

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, the YASARA energy minimization server was employed to minimize the energy of each compound. An energy comparison between all three designed ligands from the START energy to END energy confirms the energy minimization of the structural features (Table 4). The results of binding from Vina show 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 exhibited docking scores of -7.0 kcal mol⁻¹, -5.9 kcal mol⁻¹ and -6.5 kcal mol⁻¹ towards SGp. It was observed that Arg457, Ser459 and Asp467 were common participant residues in all SGp-inhibitor interactions (Fig. 4). The highest number of hydrogen bonding interactions was observed in the quercetin-SGp complex with Ser459, Asp467 and Glu471 residues. Ser459 also yielded hydrogen bonding with the other two complexes. The three designed inhibitors showed -7.5 kcal mol⁻¹, -6.9 kcal mol⁻¹ and -7.9 kcal mol⁻¹ binding energy towards M^{pro}. The common interaction residues include His163, Cys145 and Met165 (Fig. 5). Four hydrogen bonding interactions were observed in the cases of UN-2 and UN-3 whereas three were formed in the case of UN-1 with the active site residues of M^{pro}. From a docking analysis of the designed inhibitors against the four SGp mutants, it was found that UN-1 showed higher binding affinities of -6.3 kcal mol⁻¹ and -6.1 kcal mol⁻¹ against the E484K and L452R mutants than against the other two designed inhibitors. UN-2 and UN-3 had higher binding affinities of -6.8 kcal mol⁻¹ and -5.9 kcal mol⁻¹ against the N501Y and K417N mutants. The

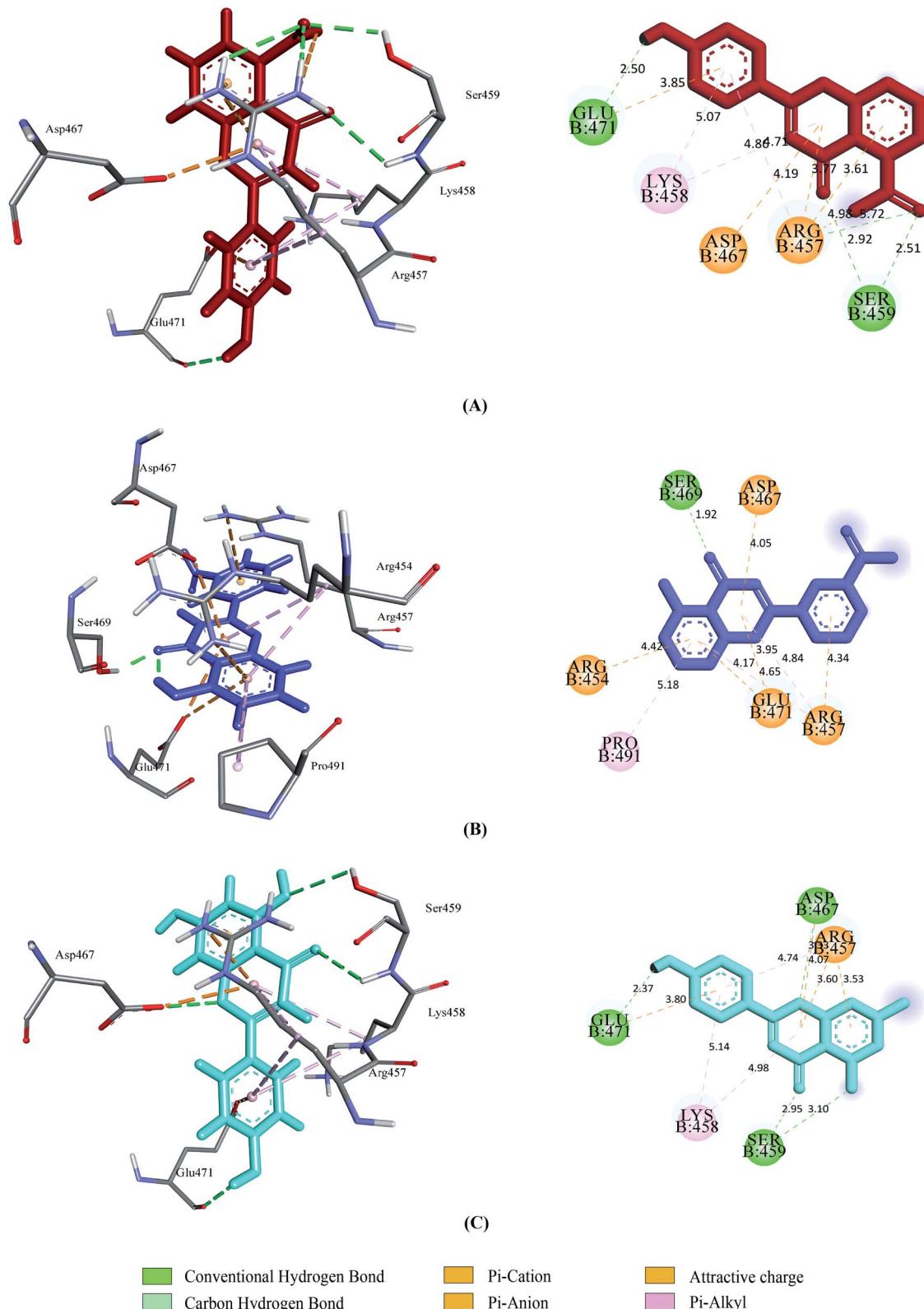


Fig. 4 2D and 3D representation of molecular docking analysis between the SARS-CoV-2 SGp with (A) UN-1, (B) UN-2, (C) UN-3.

N417, K484 and R452 mutant residues formed hydrogen bonding with the inhibitors where the highest number of hydrogen bonds was formed in the binding of UN-1 with the L452R variant (Fig. 6). The physicochemical and ligand binding

properties of the designed compounds with SGp (WT) and M^{PRO} are shown in Table S2.† The docking results of the designed compounds with SGp variants are shown in Table S3.†

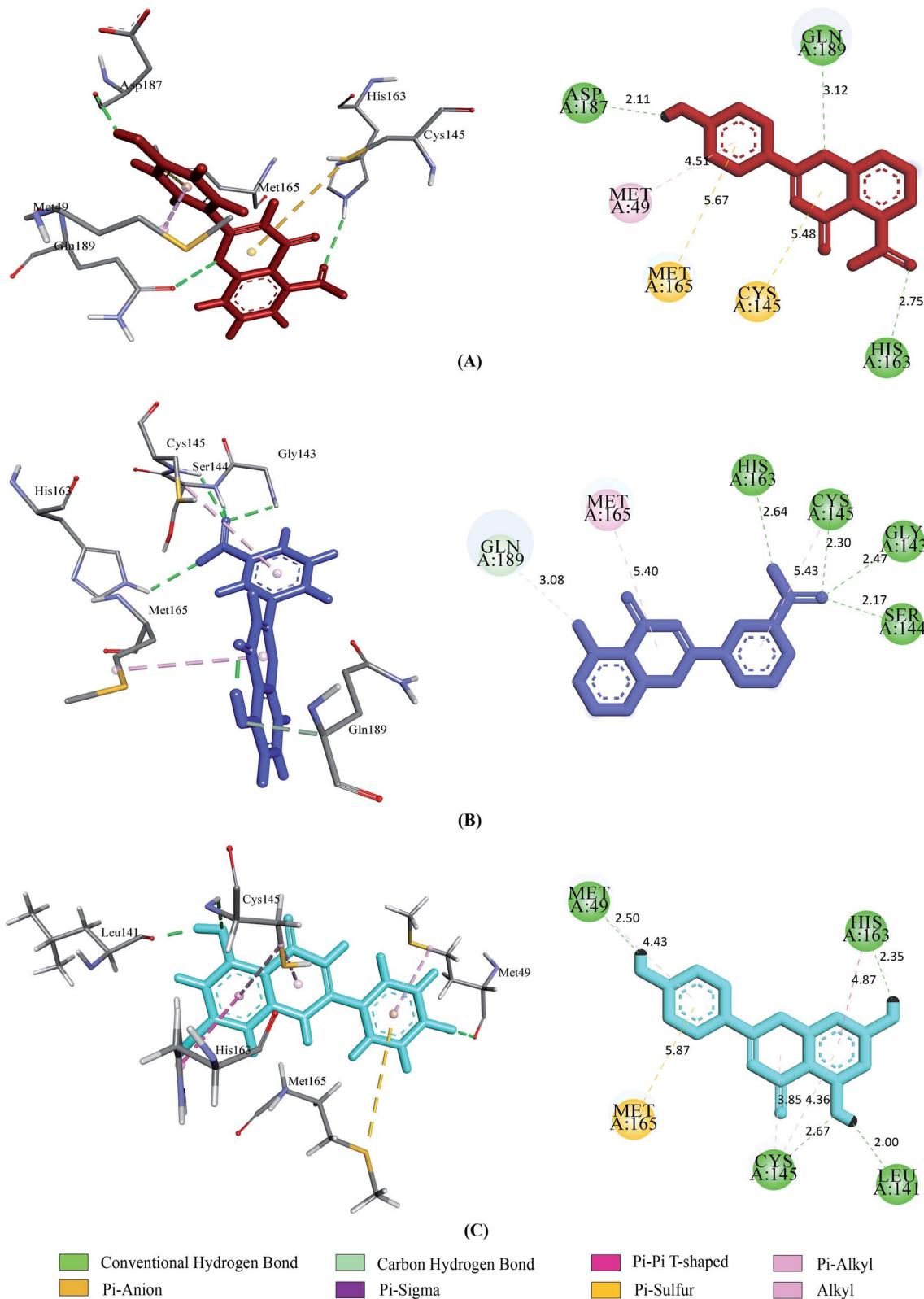


Fig. 5 2D and 3D representation of molecular docking analysis between the SARS-CoV-2 M^{pro} with (A) UN-1, (B) UN-2, (C) UN-3.

Molecular dynamics simulation

The molecular dynamics simulation of the ten protein-inhibitor complexes was performed for a period of 120 ns 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.³⁸ So, in order to examine the dynamic behavior and stability of



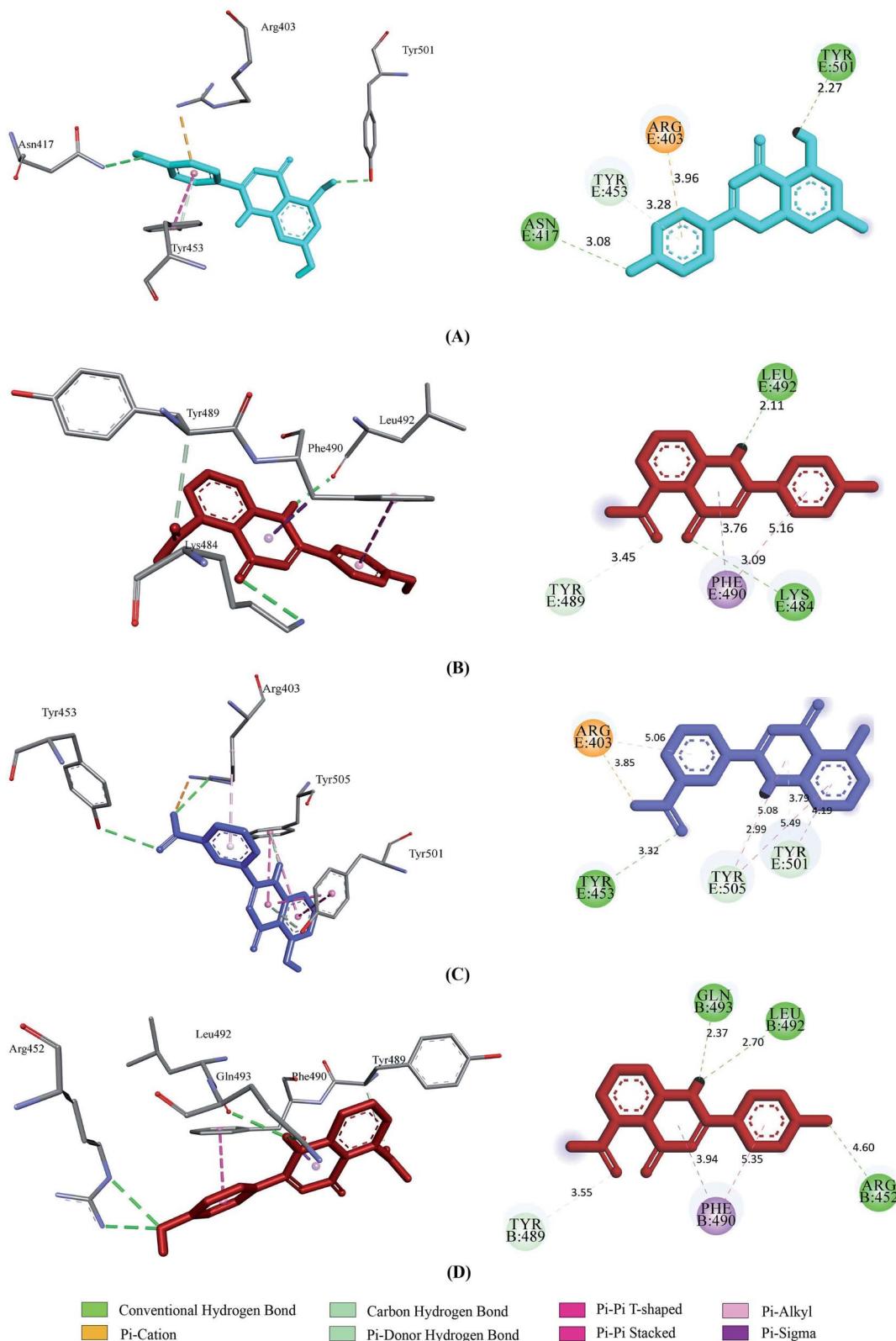


Fig. 6 2D and 3D representation of molecular docking analysis between the SARS-CoV-2 SGp mutants and designed inhibitors, (A) SGp(K417N)-UN-3, (B) SGp(E484K)-UN-1, (C) SGp(N501Y)-UN-2, (D) SGp(L452R)-UN-1.

Table 5 Average values of RMSD, RMSF, R_g , SASA and number of hydrogen bonds of the ten protein–inhibitor complexes

Complex	RMSD (nm)	RMSF (nm)	R_g (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

the complexes throughout the 120 ns simulation period, all ten complexes were analyzed by calculating the RMSD, RMSF, R_g , SASA and hydrogen bonds, as 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 $\text{C}\alpha$ backbone atoms of the ten complexes were

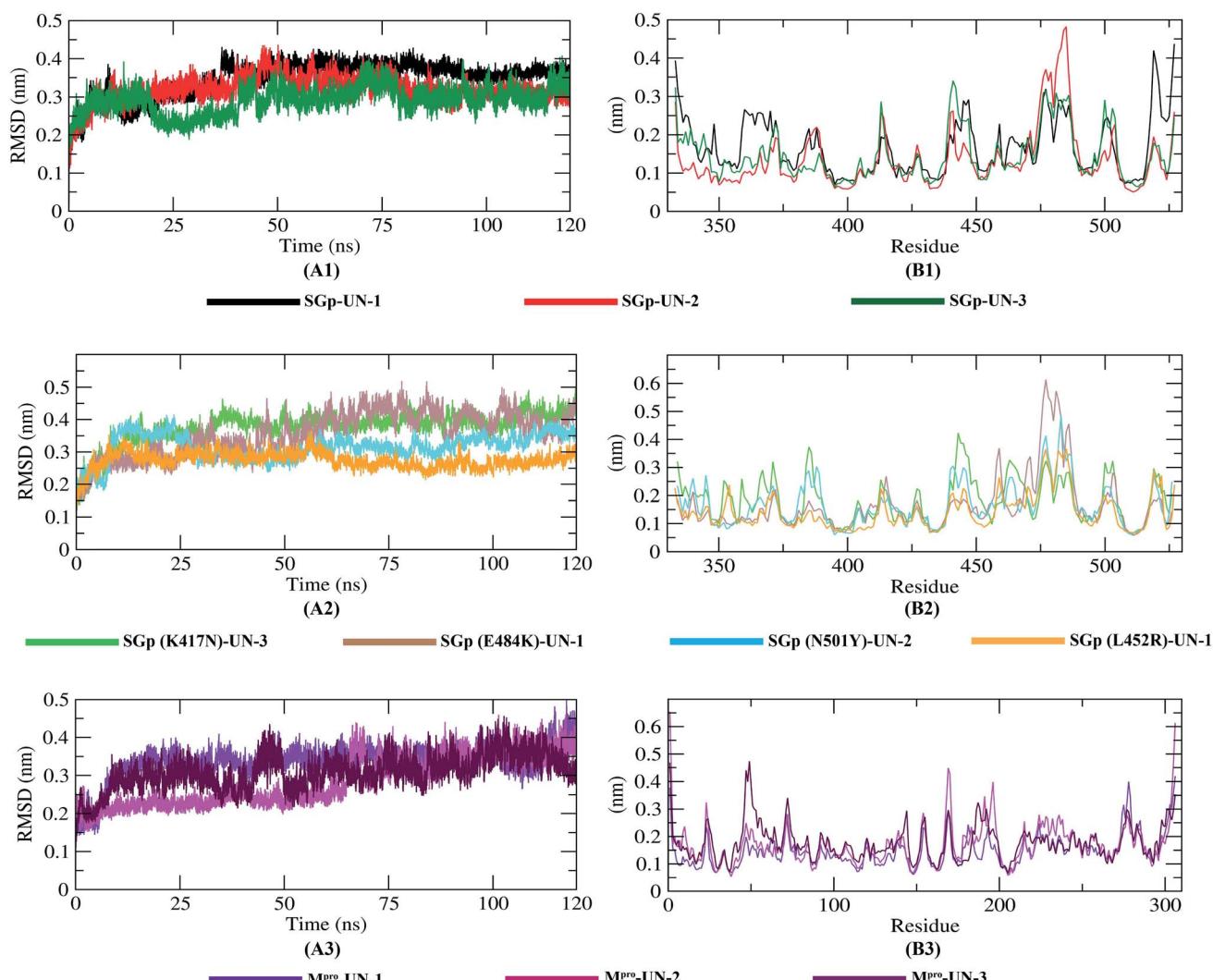


Fig. 7 The RMSD and RMSF of $\text{C}\alpha$ atoms of protein–inhibitor complexes. RMSD and RMSF graph of SGp(WT)–inhibitor complexes (A1–B1), SGp(mutant)–inhibitor complexes (A2–B2) and M^{pro}–inhibitor (A3–B3) complexes from the molecular simulation of 120 ns at 300 K.



calculated with RMSD to detect their stability. It can be observed from Fig. 7 that the SGp(L452R)-UN-1 complex exhibits a lower RMSD than the other nine complexes. The RMSD of SGp(WT)-inhibitor complexes (Fig. 7A1) increases from 50 to 75 ns; however, these values significantly decrease after 80 ns. While assessing the RMSD of SGp(mutant)-inhibitor complexes, a steady increase in RMSD is observed in the SGp(E484K)-UN-1 complex after 50 ns. Nonetheless, deviations for inhibitors complexed to other mutants were less than 0.5 nm (Fig. 7A2). Unlike the SGp-inhibitor complexes, M^{pro} -UN-3 showed a consistent fluctuation over the 35–90 ns period, indicating that UN-3 might change the protein conformation. The remaining two inhibitors bound to M^{pro} exhibited a similar trajectory in the 75–100 ns period and tended to stabilize after 100 ns (Fig. 7A3).

To better understand the regions of proteins that are fluctuating during the simulation, the flexibility of each residue was calculated in terms of RMSF to get a better insight into the

extent to which the binding of the designed inhibitors affects the flexibility of the proteins. It can be understood from Fig. 7B3 that binding with UN-3 makes 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 flexible upon binding with UN-1. Besides, the RMSF values of both the WT and mutant SGp complexes are mostly similar in all areas. Overall, residues such as Ser371, Pro384, Thr385, Lys386, Lys444, Val445, Gly446, Pro463, Phe464, Glu465, Ser477, Asn501 and Val503 are found to be flexible for all the inhibitor-bound SGp complexes.

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

The compactness of the protein–inhibitor complexes is represented by the radius of gyration (R_g). The greater compactness

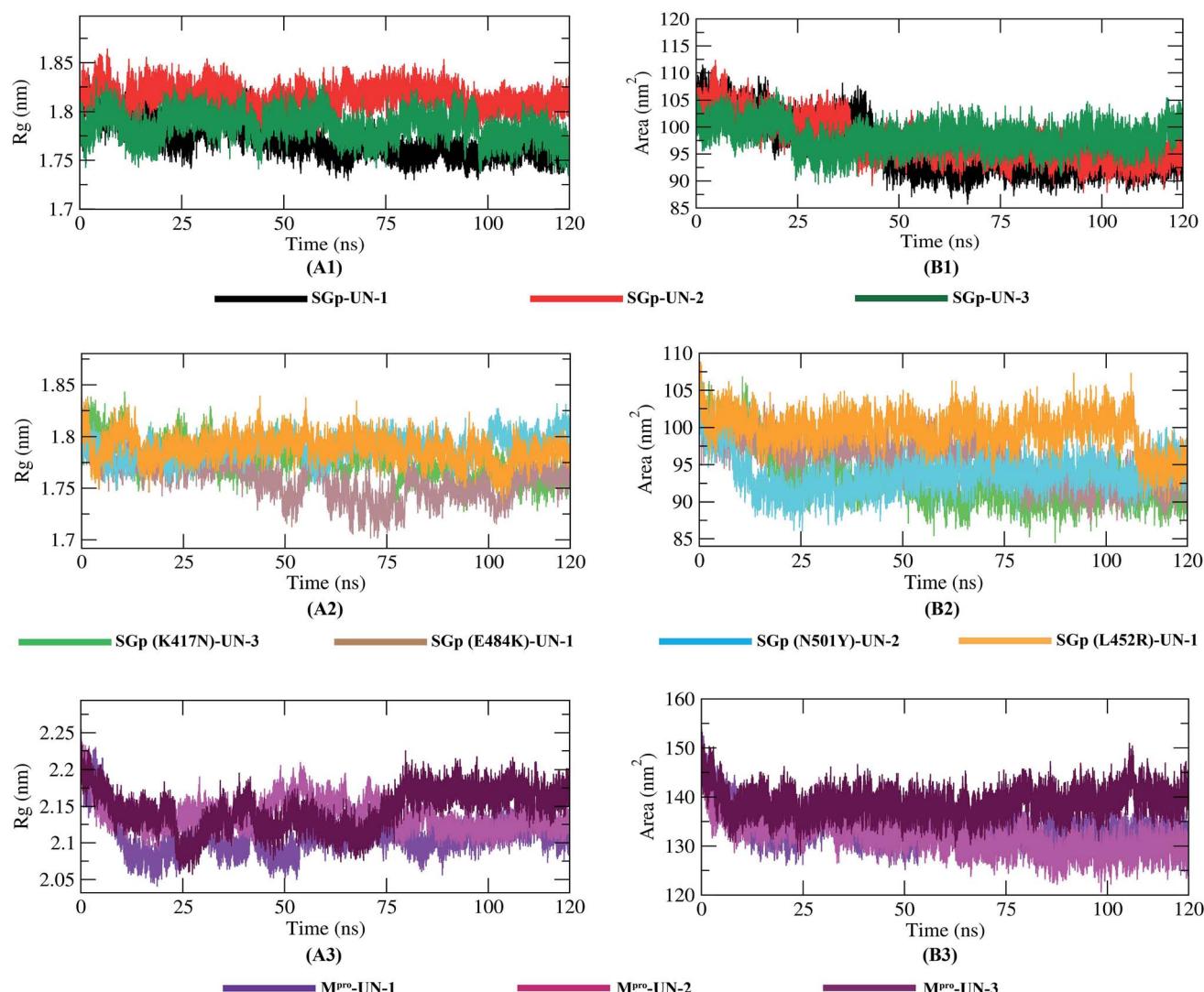


Fig. 8 Radius of gyration (R_g) plot reflecting the compactness of protein–inhibitor complexes (A1–A3) and SASA plot showing the variation in the solvent accessibility of the complexes (B1–B3) during the 120 ns MD simulations.



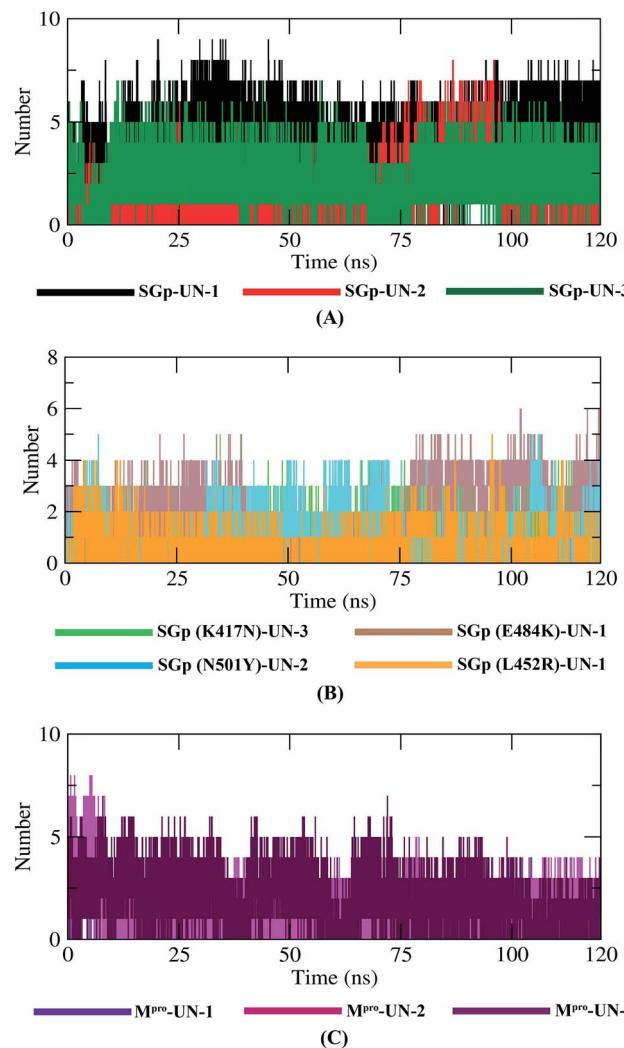


Fig. 9 Plot representing the dynamics observed in the hydrogen bonding patterns for the SGp(WT)–inhibitor (A), SGp(mutant)–inhibitor (B) and M^{pro}–inhibitor (C) complexes.

of a system is indicated by the lower degree of fluctuation throughout the simulation period. The R_g of WT SGp-UN-1 (Fig. 8A1) was found to be the lowest, indicating greater rigidity in contrast to the other complexes. Besides, a higher R_g 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 a greater change in R_g . Inhibitor-bound K417N, N501Y and L452R mutant SGp complexes showed similar R_g throughout the simulation, while inconsistency was observed in the case of the SGp(E484K)-UN-1 complex. The three M^{pro}–inhibitor complexes showed increased R_g from 10 to 75 ns, and the final parts of the trend were consistent. Although the R_g value of M^{pro}-UN-3 was much higher after 75 ns, indicating the looseness of its packing compared to all the other complexes.

Interaction between protein–inhibitor complexes and solvents can be measured with the solvent accessible surface area (SASA) over the simulation time. So, the SASAs of all ten protein–inhibitor complexes were calculated to predict the extent of the conformational changes occurring during the interaction. Plots of SASA value *vs.* time for all the protein–inhibitor complexes are shown in Fig. 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 values than the other complexes. The other inhibitor complexes did not increase in volume during the simulation process but remained in a stable state.

Hydrogen bond analysis

Hydrogen bonding between a protein–inhibitor complex is essential to stabilize the protein structures. Fig. 9 displays the total number of hydrogen bonds present in all 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 the 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 the MM/PBSA method. Apart from the M^{pro}-UN-2 complex, all the

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

Complex	van der Waals 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



complexes showed negative total binding energies, indicating strong binding of the designed compounds with the target proteins (Table 6). In particular, SGp(L452R)-UN-1 complex depicted the lowest binding free energy and a higher binding affinity ($-214.155 \pm 26.323 \text{ kJ mol}^{-1}$), suggesting a more stable conformation of UN-1. The remaining complexes also have

favorable binding energies and, thus, these designed inhibitors could be used as potential lead compounds. A comparison of the binding free energies of all ten complexes was made by plotting the binding energy *versus* time graphs shown in Fig. 10.

Further, we determined the contribution of each residue of SGp and M^{pro} in terms of binding free energy to the interaction

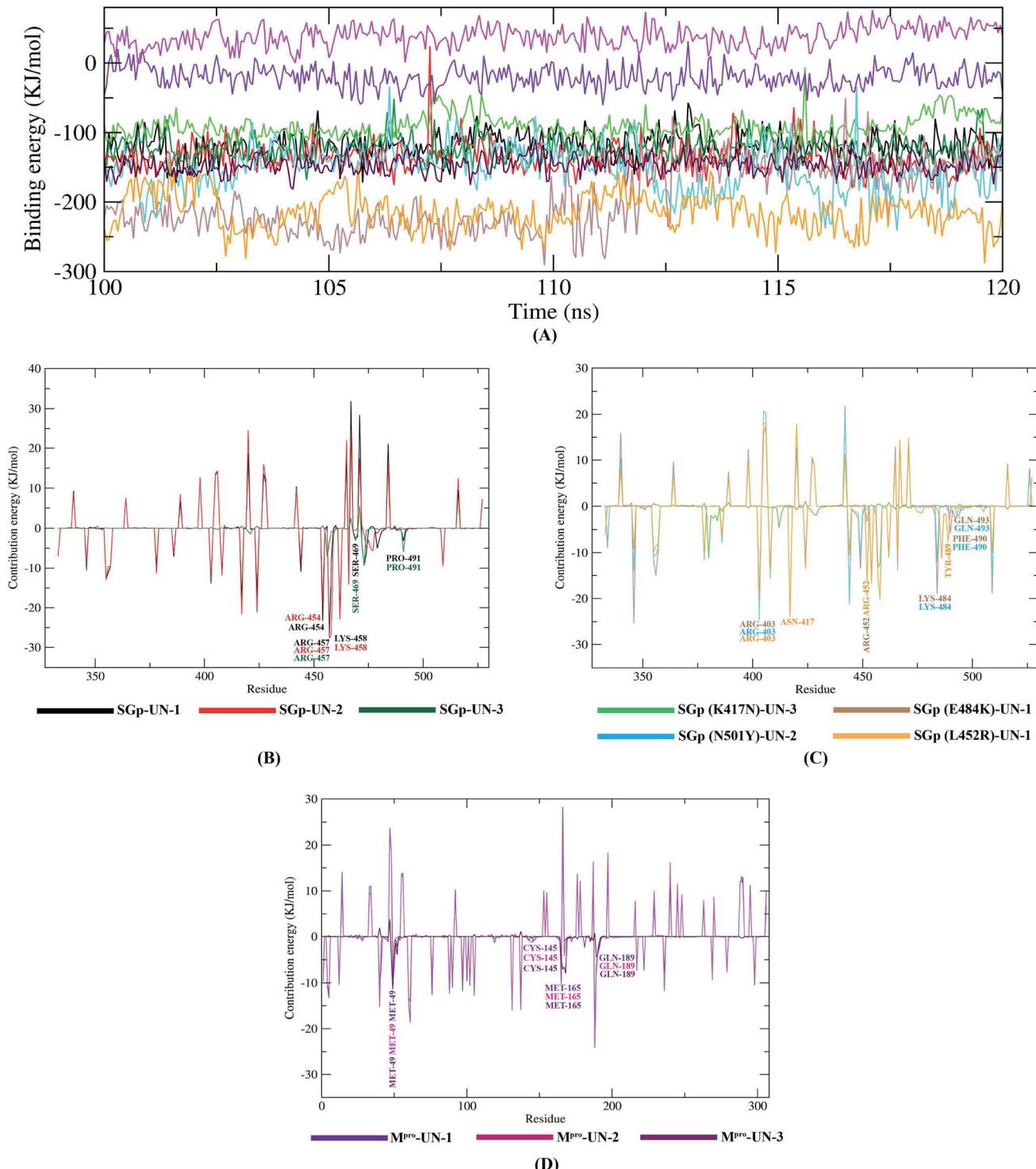


Fig. 10 Graphical representation of the binding free energy of protein–inhibitor complexes (A) and per residue contribution plot for SGp–inhibitors (B), SGp(mutant)–inhibitors (C), M^{pro}–inhibitors (D) complexes.

with the designed inhibitors which was performed by decomposing the total binding free energy of the system into per residue contribution energy (Fig. 10).

Out of the residues showing interaction in molecular docking analysis, the Arg457 residue of SGp exhibited the highest contribution energies of $-26.9792\text{ kJ mol}^{-1}$, $-27.6262\text{ kJ mol}^{-1}$ and $-3.6108\text{ kJ mol}^{-1}$, respectively, with UN-1, UN-2 and UN-3. Several residues of M^{pro} (Met49, CYS145, Met165 and Gln189) also showed favorable contribution energies in interacting with the active site pocket. The Met49 residue showed -9.27 kJ mol^{-1} contribution energy in complexes with UN-1, UN-2 and UN-3: the contribution energies were -5.05 kJ mol^{-1} and -11.4 kJ mol^{-1} , respectively. Many residues of the four mutants showed favorable contribution energies towards the inhibitors. On comparing the complexes, we found that the mutant K484 and R452 residues involved in the interaction showed $-16.8163\text{ kJ mol}^{-1}$ and $-19.0742\text{ kJ mol}^{-1}$ contribution energies 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 energies of other key residues of SGp (both WT and mutants) and M^{pro} are shown in Tables S4 and S5.†

Discussion

Computer aided drug design (CADD) approaches have been utilized very widely in order to increase the efficiency of the course of drug discovery and development. This method requires three-dimensional (3D) structures of the target proteins, where the binding mode, affinity and confirmation of a ligand binding can be detected. Several CADD approaches are being 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.³⁹ Also, the increasing development of robust computational tools allows scientists to assess the pharmacological properties of compounds before entering into experimental research. So, intensive CADD methods were used in this study to design novel inhibitors against COVID-19.

In this study, we emphasized both the binding affinity and ADMET properties of compounds to design potent inhibitors of SGp and M^{pro}. As some newly emerged variants have recently been discovered that are mutated in various regions, we tried to design 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 selected according to their high binding affinity towards the two proteins. It was found that isoflavan 3'-gallate exhibits the highest binding affinity of $-7.2\text{ kcal mol}^{-1}$ while binding with SGp (Table S1†). Conversely, theaflavin 3,3'-digallate exhibits the highest affinity of $-8.9\text{ kcal mol}^{-1}$ and interacted with the highest number of

amino acid residues in the active site of M^{pro}. As we addressed earlier, despite the high binding affinities of these compounds, we further filtered them based on Lipinski's rule of five and their bioavailability score to retain the biologically active compounds for further analysis (Table 2). Luteolin, myricetin and quercetin passed this stage and were 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 $10.6\text{ }\mu\text{M}$) and quercetin was shown to prevent HIV-luc/SARS pseudotyped virus entry in host cells (EC₅₀ of $83.4\text{ }\mu\text{M}$). Myricetin showed an inhibitory effect against SARS-CoV NSP13 ATPase with IC₅₀ values of 2.71 .⁴⁰ A study showed 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.⁴¹ Our computational analysis using up-to-date online-based tools also revealed several toxic properties, such as mutagenicity and tumorigenicity, as well as low intestinal absorption, an inhibitory effect to several isoforms of cytochrome P450 enzyme, and their possession of catechol groups (Table 3). Bearing in mind an improvement in 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 a significant increase in the intestinal absorption rate for all three analogues and they do not cross the blood-brain barrier. UN-1 does not inhibit the CYP2C9 isoform of cytochrome P450 enzyme unlike its parent compound, which is an important detoxification enzyme of the human body. No designed inhibitor shows mutagenic or tumorigenic properties like myricetin or quercetin. There is a very low chance of acquiring long QT syndrome from these designed inhibitors, but none appears to inhibit hERG I, II receptors (Table 3). A significant improvement was also observed in the medicinal chemistry profile of these designed inhibitors. Pan-assay interference compounds (PAINS) are chemical compounds that tend to give false positive results in high-throughput screens and react nonspecifically with several biological targets. Luteolin, myricetin and quercetin had a catechol group in their structures but it was not present in the designed inhibitors, as revealed by the PAINS filter. All the newly designed inhibitors have a low score of synthetic accessibility which denotes that these compounds can be synthesized much more easily in laboratories than the parent compounds. The bioavailability of UN-1 and UN-2 increased slightly and the drug score 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 towards SARS-CoV-2 SGp and M^{pro} active sites. Most of the interacted residues of SGp with the three inhibitors include Arg454, Arg457, Lys458 and Ser459 (Table S2†). Previous research showed that SARS-CoV-2 may interact with the host cell surface glycosaminoglycans (GAG) through



SGp to invade the host cell and it contains three putative GAG-binding motifs where the 453–459 sequence consists of site 1.⁴² 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 bonds and carbon–hydrogen bonds with the inhibitors, indicating strong destabilization of the residues, thus preventing their binding with ACE2. In our docking analysis with M^{pro}, most of the compounds interacted with Met49, His163, Cys145 and Met165 (Table S2†). All the inhibitors bind with Cys145 which is a residue of a 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 SGp, SGp variants and M^{pro}. The RMSD plot suggests that all three inhibitors are stable and the RMSD values did not show any sudden surge throughout the simulation period (Fig. 7). RMSF analysis confirmed that the complexes showed less fluctuation in an acceptable range. The radius of gyration (R_g) of all the protein–inhibitor complexes tend to be in same manner, suggesting that each complex had relatively similar behavior of compactness (Fig. 8). The SASA values depicted that, except for 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 ten complexes throughout the simulation, explaining their conformational stability one more time (Fig. 9). Further, the binding free energies were calculated using the MM/PBSA method for all the complexes and the results show that five complexes showed favorable binding energy except for the M^{pro}-UN-2 complex. The SGp(L452R)-UN-1 complex showed the lowest binding energy among the others (Fig. 10). From the 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 make significant contributions to the binding of the designed inhibitors. Notable contributions of the mutant residues of the SGp variants were also observed, confirming that the inhibitors are capable of stabilizing the variants and interfering in stable binding with ACE2 (Fig. 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. Although the three designed analogues showed tremendous results in inhibiting the two proteins, they have not performed in animal model experiments. 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 the SGp and M^{pro} of SARS-CoV-2 and its variants. The 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, these designed inhibitors might be probable candidates as plausible therapeutics against SARS-CoV-2.

Funding

There are no sources of funding for this study.

Data availability

We have used Online SMILES Translator and Structure File Generator (<https://cactus.nci.nih.gov/translate/>); (Discovery Studio 4.0 client (<http://www.accelrys.com/products/discovery-studio/>); Topography of Protein (CASTp) (<http://www.sts.bioengr.uic.edu/castp/>) server which provides an online resource for locating, depicting and measuring concave surface regions on 3D structures of proteins.²⁶ 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 server³⁴ (<http://biosig.unimelb.edu.au/pkcs/>).

Author contributions

Conceptualization and methodology was performed by MNR, NF and MUH; formal analysis and data curation was performed by MNR, NF and MTHE; writing – original draft prepared by MNR, NF and MUH; writing – review and editing performed by MNR, NF, MUH, AKMM and MSI; supervised by MUH. All authors have read and agreed to submit the final version of the manuscript.

Conflicts of interest

No potential conflict of interest relevant to this article was reported.

Acknowledgements

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. The authors are also highly grateful to Dr Michael Rooney, Trinity College, Dublin, Ireland for the correction of the language, grammar and critical revision for the substantial development of this manuscript.

References

- 1 X. Xu, P. Chen, J. Wang, J. Feng, H. Zhou, X. Li, W. Zhong and P. Hao, *Sci. China: Life Sci.*, 2020, **63**, 457–460.
- 2 D. S. Hui, E. I. Azhar, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T. D. Mchugh, Z. A. Memish, C. Drosten, A. Zumla and E. Petersen, *Int. J. Infect. Dis.*, 2020, **91**, 264–266.



3 H. A. Rothan and S. N. Byrareddy, *J. Autoimmun.*, 2020, **109**, 102433.

4 I. I. Bogoch, A. Watts, A. Thomas-Bachli, C. Huber, M. U. G. Kraemer and K. Khan, *J. Travel Med.*, 2020, **27**(2), DOI: 10.1093/jtm/taaa008.

5 T. Singhal, *Indian J. Pediatr.*, 2020, **87**, 281–286.

6 S. E. Galloway, P. Paul, D. R. MacCannell, M. A. Johansson, J. T. Brooks, A. MacNeil, R. B. Slayton, S. Tong, B. J. Silk, G. L. Armstrong, M. Biggerstaff and V. G. Dugan, *Morb. Mortal. Wkly. Rep.*, 2021, **70**, 95–99.

7 K. Leung, M. H. Shum, G. M. Leung, T. T. Lam and J. T. Wu, *Eurosurveillance*, 2021, **26**, 2002106.

8 Y. Chen, Q. Liu and D. Guo, *J. Med. Virol.*, 2020, **92**, 418–423.

9 A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire and D. Veesler, *Cell*, 2020, **181**, 281–292.

10 X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang and Z. Qian, *Nat. Commun.*, 2020, **11**, 1620.

11 Y. J. Hou, S. Chiba, P. Halfmann, C. Ehre, M. Kuroda, K. H. Dinnon, S. R. Leist, A. Schäfer, N. Nakajima, K. Takahashi, R. E. Lee, T. M. Mascenik, R. Graham, C. E. Edwards, L. V. Tse, K. Okuda, A. J. Markmann, L. Bartelt, A. De Silva, D. M. Margolis, R. C. Boucher, S. H. Randell, T. Suzuki, L. E. Gralinski, Y. Kawaoka and R. S. Baric, *Science*, 2021, **370**, 1464–1468.

12 M. H. S. Paiva, D. R. D. Guedes, C. Docena, M. F. Bezerra, F. Z. Dezordi, L. C. Machado, L. Krokovsky, E. Helvecio, A. F. da Silva, L. R. S. Vasconcelos, A. M. Rezende, S. J. R. da Silva, K. G. d. S. Sales, B. S. L. F. de Sá, D. L. da Cruz, C. E. Cavalcanti, A. d. M. Neto, C. T. A. da Silva, R. P. G. Mendes, M. A. L. da Silva, T. Gräf, P. C. Resende, G. Bello, M. d. S. Barros, W. R. C. do Nascimento, R. M. L. Arcoverde, L. C. A. Bezerra, S. P. Brandão-Filho, C. F. J. Ayres and G. L. Wallau, *Viruses*, 2020, **12**, 1414.

13 V. Tchesnokova, H. Kulakesara, L. Larson, V. Bowers, E. Rechkina, D. Kisiela, Y. Sledneva, D. Choudhury, I. Maslova, K. Deng, K. Kutumbaka, H. Geng, C. Fowler, D. Greene, J. Ralston, M. Samadpour and E. Sokurenko, *bioRxiv Prepr. Serv. Biol.*, 2021, DOI: 10.1101/2021.02.22.432189.

14 C. K. Wibmer, F. Ayres, T. Hermanus, M. Madzivhandila, P. Kgagudi, B. Oosthuysen, B. E. Lambson, T. de Oliveira, M. Vermeulen, K. van der Berg, T. Rossouw, M. Boswell, V. Ueckermann, S. Meiring, A. von Gottberg, C. Cohen, L. Morris, J. N. Bhiman and P. L. Moore, *Nat. Med.*, 2021, **27**, 622–625.

15 M. U. Mirza and M. Froeyen, *J. Pharm. Anal.*, 2020, **10**, 320–328.

16 R. Hilgenfeld, *FEBS J.*, 2014, **281**, 4085–4096.

17 A. Baldi, *Syst. Rev. Pharm.*, 2010, **1**, 99–105.

18 G. Jaiswal and V. Kumar, *PLoS One*, 2020, **15**, e0240004.

19 D. S. N. B. K. Prasanth, M. Murahari, V. Chandramohan, S. P. Panda, L. R. Atmakuri and C. Guntupalli, *J. Biomol. Struct. Dyn.*, 2020, 1–15.

20 H. Puttaswamy, H. G. Gowtham, M. D. Ojha, A. Yadav, G. Choudhir, V. Raguraman, B. Kongkham, K. Selvaraju, S. Shareef, P. Gehlot, F. Ahamed and L. Chauhan, *Sci. Rep.*, 2020, **10**, 20584.

21 Y. Kumar, H. Singh and C. N. Patel, *Syst. Rev. Pharm.*, 2020, **13**, 1210–1223.

22 T. Zi Wei, H. Wang, X. Qing Wu, Y. Lu, S. Hui Guan, F. Quan Dong, C. Le Dong, G. Li Zhu, Y. Zhou Bao, J. Zhang, G. Yu Wang and H. Ying Li, *Chin. J. Integr. Med.*, 2020, **26**, 663–669.

23 M. T. Islam, C. Sarkar, D. M. El-Kersh, S. Jamaddar, S. J. Uddin, J. A. Shilpi and M. S. Mubarak, *Pharmacol. Res.*, 2020, **34**, 2471–2492.

24 N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714–2723.

25 G. M. Morris, H. Ruth, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.

26 W. Tian, C. Chen, X. Lei, J. Zhao and J. Liang, *Nucleic Acids Res.*, 2018, **46**, W363–W367.

27 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2009, **31**, 455–461.

28 M. A. Lill and M. L. Danielson, *J. Comput.-Aided Mol. Des.*, 2011, **25**, 13–19.

29 A. Daina, O. Michielin and V. Zoete, *Sci. Rep.*, 2017, **7**, 42717.

30 E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker and K. Karplus, *Proteins: Struct., Funct., Bioinf.*, 2009, **77**, 114–122.

31 M. J. Sippl, *Proteins: Struct., Funct., Bioinf.*, 1993, **17**, 355–362.

32 G. Vriend and C. Sander, *J. Appl. Crystallogr.*, 1993, **26**, 47–60.

33 J. Kuszewski, A. M. Gronenborn and G. M. Clore, *J. Magn. Reson.*, 1997, **125**, 171–177.

34 D. E. V. Pires, T. L. Blundell and D. B. Ascher, *J. Med. Chem.*, 2015, **58**, 4066–4072.

35 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1–2**, 19–25.

36 A. W. Schüttelkopf and D. M. F. Van Aalten, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2004, **60**, 1355–1363.

37 R. Kumari, R. Kumar and A. Lynn, *J. Chem. Inf. Model.*, 2014, **54**, 1951–1962.

38 V. K. Bhardwaj, R. Singh, J. Sharma, V. Rajendran, R. Purohit and S. Kumar, *J. Biomol. Struct. Dyn.*, 2020, 1–10.

39 L. G. Ferreira, R. N. Dos Santos, G. Oliva and A. D. Andricopulo, *Molecules*, 2015, **20**, 13384–13421.

40 M. Russo, S. Moccia, C. Spagnuolo, I. Tedesco and G. L. Russo, *Chem.-Biol. Interact.*, 2020, **328**, 109211.

41 M. Matsuo, N. Sasaki, K. Saga and T. Kaneko, *Biol. Pharm. Bull.*, 2005, **28**, 253–259.

42 S. Y. Kim, W. Jin, A. Sood, D. W. Montgomery, O. C. Grant, M. M. Fuster, L. Fu, J. S. Dordick, R. J. Woods, F. Zhang and R. J. Linhardt, *Antiviral Res.*, 2020, **181**, 104873.