Identification of Severe Acute Respiratory Syndrome Coronavirus-2 inhibitors through in silico structure-based virtual screening and molecular interaction studies

Iqbal Azad | Tahmeena Khan | Akhilesh Kumar Maurya | Mohd. Irfan Azad | Nidhi Mishra | Amer M. Alanazi

1Department of Chemistry, Integral University, Lucknow, India
2Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj, India
3Department of Chemistry, Jamia Millia Islamia, New Delhi, India
4Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia

Correspondence
Tahmeena Khan, Department of Chemistry, Integral University, Kursi Road,Lucknow 226026, UP, India. Email: tahminakhan30@yahoo.com

Abstract
The novel coronavirus Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) or COVID-19 has caused a worldwide pandemic. The fatal virus has affected the health of human beings as well as the socio-economic situation all over the world. To date, no concrete medicinal solution has been proposed to combat the viral infection, calling for an urgent, strategic, and cost-effective drug development approach that may be achievable by applying targeted computational and virtual screening protocols. Immunity is the body’s natural defense against disease-causing pathogens, which can be boosted by consuming plant-based or natural food products. Active constituents derived from natural sources also scavenge the free radicals and have anti-inflammatory activities. Herbs and spices have been used for various medicinal purposes. In this study, 2,96365 natural and synthetic derivatives (ligands) belonging to 102 classes of compounds were obtained from PubChem and assessed on Lipinski’s parameters for their potential bioavailability. Out of all the derivatives, 3254 obeyed Lipinski’s rule and were virtually screened. The 115 top derivatives were docked against SARS-CoV-2, SARS-CoV, MERS-CoV, and HCoV-HKV1 main proteases (Mpros) as receptors using AutoDock Vina, AutoDock, and iGEMDOCK 2.1. The lowest binding energy was exhibited by ligands 2 and 6 against all the four Mpros. The molecular dynamic simulation was also performed with ligand 6 using the GROMACS package. Good bioactivity scores, absorption, distribution, metabolism, excretion, and toxicity profile and drug-like pharmacokinetic parameters were also obtained. Hydroxychloroquine was used as the control drug.

KEYWORDS
docking, Lipinski, protease, SARS-CoV-2, simulation, virtual screening

1 | INTRODUCTION
Betacoronaviruses (βCoVs) are the genera of coronaviruses (CoVs), which have received significant importance due to their potent virulence. Human infecting coronaviruses include HCoV-OC43, HKU and 229E5, Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). Out of these, the SARS and MERS coronaviruses were highly pathogenic and associated with high mortality. In December 2019, the third Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) was identified from China as the newest pathogenic virus of the πCoV lineage. By March 2020, the World Health Organization...
inhibitor remdesivir and research has suggested that nucleoside ana-
although clinical trials are underway with RNA polymerase (RdRP)
marketed drugs against SARS-CoV, SARS-CoV-2, and MERS-CoV,
hijacking mechanism for drug development. To date, there are no
SARS-CoV-2 Mpro inhibitors. SARS-CoV-2 Mpro interacts with the
and computational tools have been explored to develop probable
strategies include molecular dynamics (MD) and virtual screening11
and computational tools have been explored to develop probable
SARS-CoV-2 Mpro inhibitors. SARS-CoV-2 Mpro interacts with the
active sites and releases the RNA into the host cell and infects the
same.12 It replicates itself and mainly affects the respiratory sys-
tem. Hence, there is an imperative need to identify a SARS-CoV-2
Mpro inhibitor. The CoV Mpros are enveloped positive-stranded RNA
viruses with the largest viral RNA genomes ranging between 27 and
31 kb.13 The CoV replicate gene has two polyproteins, termed pp1a
and pp1b, which govern the replication and transcription.14 Most mat-
uration cleavage events are governed by the CoV Mpro or 3CL prote-
ase or 3CLpro, a three-domain (domains I to III) cysteine protease.15
The structures of CoV Mpro revealed the formation of an active homodimer.16,17 the 1-7 N-terminal residues of the Mpro have an
important role in the proteolytic action18 and the C-terminal domain
is essential for dimerization.19 Computational tools help in under-
standing the interaction between small molecules, also known as
ligands and receptors.20 Molecular docking and MD simulation find
application in the assessment of the binding affinity of the probable
lead molecule with the receptor. Molecular docking studies predict
the binding energy when a ligand binds to the receptor and the inter-
action affinity.21 MD simulation is used to simulate a protein to know
the dynamic behavior of the system. The simulating entities are taken
in a box either at a fixed or varying temperature. Replica exchange
molecular dynamics (REMD) can be used to study the inhibition of the
protease with a variation in temperature.22

The present study is aimed at the development of a probable
inhibitor of CoV Mpros, including SARS-CoV-2, with the help of in sil-
ico and virtual screening of the bioactive constituents (natural and
synthetic derivatives) of some commonly occurring herbs and spices
against the target protein. Spices and herbs have found a prominent
role as flavoring and aroma inducing agents. They are also useful as
coloring or preservative agents and used for medicinal purposes.23
They also offer an array of benefits against acute, chronic, and non-
communicable ailments. Their active constituents are rich in sulfur-
containing compounds, alkaloids, tannins, terpenoids, vitamins, flavo-
noids, etc.24 They also contain saponins, enzymes, and minerals,
including selenium, copper, manganese, zinc, chromium, iodine, etc.
Spices also possess exceptional antioxidant activity as they are rich in
phenolic compounds.25 Other than antioxidant potential, the bioactive
constituents are also responsible for anti-inflammatory, anti-tumor-
genic, and anti-carcinogenic activities along with blood sugar and
cholesterol-lowering activities.26 Studies support that spice consump-
tion reduces the risk of mortality associated with cancer, heart, and
respiratory diseases. A recent study in China with 487 375 partici-
pants related to the consumption of spicy foods as part of their diet
concluded the inclusion of spices in food to improve and maintain
health.27 The present study took into account the computational
pharmacokinetics and drug-likeness assessment of the hit molecules,
in search of a potential inhibitor of the SARS-CoV, SARS-CoV-2,
HCoV-HKU1, and MERS-CoV Mpros. The protein-ligand binding
was studied through molecular docking and MD simulation was performed
to measure the stability of the best protein-ligand complex.

2 | MATERIAL AND METHODS

The crystal structure of the target protein is the most accurate model
for the evaluation of protein-ligand interaction. Several important
criteria need to be considered before the selection of the appropriate
crystal structure of the target protein viz. (a) minimum possible resolu-
tion or maximum quality of the target protein; (b) identification of the
crystal structure based on domain completeness. The full structure of
the studied protein must be available, and there should not be any
gap between the residues, as partial domain leads to inaccurate analy-
sis, (c) any mutation in the target protein of interest must be avoided,
(d) the side chains of the crystal structure must be small so that the
ligand does not interact with them, (e) identification of the complete
structure of the target protein is necessary, as partial domain would
lead to an incorrect evaluation, and the confirmation must be done
based on its sequence analysis, (f) presence of crystalline water/co-
factor/ligands also favors the selection of desirable target proteins, as
the attachment of a ligand shows that the target protein is mimicking
the natural system and the very fact supports the presence of binding
sites or active pockets, even though later on they can be easily
removed, (g) active sites can be easily identified. Based on these

(Addresses correspondence to: Shekhar Azad, Salar Jung
Medical College, Osmania University, Hyderabad. E-mail:
azadshekar@gmail.com)

(To whom correspondence should be addressed: AVS
Kandeepan, School of Life Sciences, Indian Institute of
Science Education and Research, Pune 411 008.
E-mail: avskandeepan@gmail.com)
parameters, the SARS-CoV-2 (PDB ID: 6LU7), SARS-CoV (1Q2W), MERS-CoV (4RSP), and HCoV-HKV1 (3D23) M\textsuperscript{pro}\textsubscript{5} were selected from the RCSB protein data bank. The structures of the ligands were obtained from PubChem and assessed on Lipinski’s rule of five for their bioavailability. Virtual screening was done through VEGA ZZ 3.2.0, MGL Tools 1.5.6 and AutoDock Vina 1.1.2. AutoDock 4.2.6 and iGEMDOCK 2.1 were used for docking evaluation. The grid parameters used for the molecular docking studies are represented in Table S2 along with their coordinates. The most interactive and suitable grids were, Grid IV for SARS-CoV, Grid II for HCoV-HKU1, Grid I for MERS-CoV, and Grid II for SARS-CoV-2 M\textsuperscript{pro}\textsubscript{5} (Table S1), respectively. The best conformations were identified based on the liberated energy upon interaction with the target M\textsuperscript{pro}\textsubscript{5}. Top scoring derivatives were analyzed for their drug-like character on several parameters, including bioactivity score and ADMET profile. Molinspiration 2018 (www.molinspiration.com) was used to predict the bioactivity scores. OSIRIS Property Explorer 5.2.1 was used to assess the toxicity, admetSAR 2.0 (http://mirmd.ecust.edu.cn/admetSAR1/predict/) and swissADME (http://www.swissadme.ch/) were used to assess the ADMET properties. BIOVIA Discovery Studio 2017R2 and PyMol 1.8 were used to visualize the structures.

2.1 | Identification of ligands

In developing countries like India and China, a plant-based medicine system has been used for ages for the improvement of health and well-being. WHO has predicted that the majority of the world population relies upon medicine based on herbal sources.\textsuperscript{28} Herbs and spices have been used as food additives\textsuperscript{29} and derived from roots, barks, seeds, fruits, flowers, etc.\textsuperscript{30} They are used to enhance aroma, as flavoring agents and for various medicinal purposes. Their bioactive constituents have the potential to decrease the risk of diabetes, obesity, cancer, and cardiovascular disease.\textsuperscript{31} Their bioactive constituents have innumerable medicinal benefits like purgative, laxative, expectorant, carminative, diuretic, etc. Spices have a range of functions and protect the human body against many illnesses and disorders.\textsuperscript{32} In this study, 2,96,365 plant-based natural and synthetic derivatives were extracted from PubChem (Table S2), for further exploration of their binding potential with the CoV M\textsuperscript{pro}\textsubscript{5}. The docking interactions were also studied against the two physiological ligands (MPD and PRD) and hydroxychloroquine (HCQ) as a reference drug.

2.2 | Identification of receptor

Spike, envelop, membrane, protease, and nucleocapsid proteins are some of the potential target sites explored for drug designing against coronaviruses. Sixteen nonstructural polyproteins (NSPs) containing PP1a and PP1b are found in the genome of coronavirus encoded by replicase.\textsuperscript{33} The release of NSPs is facilitated by the protease. The cysteine protease or M\textsuperscript{pro}\textsubscript{5} or 3-C like protease (3CL\textsuperscript{pro}) leads to the cleavage at the C-terminal of the polyprotein. The Cys-His dyad is responsible for the protease activity.\textsuperscript{34} The protease cleaves at 11 specific sites in the p1 region of PP1a and PP1ab facilitating replication of the virus and release of the NSPs. Human immunodeficiency virus (HIV) protease inhibitors like lopinavir and ritonavir have shown inhibitory action against M\textsuperscript{pro}\textsubscript{5}. In this study, the M\textsuperscript{pro}\textsubscript{5} was selected as the target protein, and through flexible docking, the binding interaction with the selected ligands was unleashed.

2.3 | Ligand preparation

Crystalline structures of the selected ligands were obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov). Overall, 2,76,225 derivatives were found in the database. To obtain the potential lead entity, the ligands were filtered on Lipinski’s rule of bioavailability, which is essential to be known in a promising drug moiety to know how the dose reaches the bloodstream, upon oral administration.\textsuperscript{35} Of all the selected ligands, 3254 obeyed Lipinski’s rule and were further selected for virtual screening and molecular docking. Further optimization was done through Merck Molecular Force Field and MGL tools 1.5.6 was used for the preparation of ligand parameters. 2D and 3D structures of the ligands were viewed in ChemDraw 3D.

2.4 | Receptor preparation

Replicase polyprotein 1a (pp1a) and replicase polyprotein 1b (pp1b) exist in the genome of coronavirus.\textsuperscript{36} The viral 30 kb RNA genome open reading frame (ORF1a/b) consists of polyproteins (pp1a and pp1b), accessory, spike, small envelope, matrix, and nucleocapsid proteins. In this study, pp1a was downloaded from the PDB. The open reading frame is important for defining the target genomic sequence.\textsuperscript{37} Open reading frame 1a (ORF1a) was obtained from the Protein Data Bank. The structural procurement of target proteins was carried out by removing water, hit atom, ligands, and ions. Target proteins were also subjected to structural filtration through Chimera 1.12 and energy minimization was done through the default root mean deviation square (RMSD) and assisted model building with energy refinement (AMBER) force field 14SB. For molecular docking, target structures were modified by adding polar hydrogen atoms and Kollman charges, whereas nonpolar hydrogen atoms were merged. Polar hydrogen atoms are usually involved in hydrogen bonding leading to dipole-dipole and other weak interactions, whereas non-polar hydrogens are involved in van der Waals interactions, which do not have very significant importance. The Kollman charges depict the template value for an amino acid, derived from the corresponding electrostatic potential. Although the binding sites of the ligands were identified by the sequence analysis of the PDB file, yet for accurate confirmation and validation, the interior pocket and cavities of the target protein were also obtained from CASTp 3.0 (Computed Atlas of Surface Topography of proteins) (Figure S1).\textsuperscript{38}
2.5 | Molecular docking interactions through AutoDock Vina, AutoDock, and iGEMDOCK

Molecular docking helps in establishing the most suitable interaction profile of the ligand with the target protein and is used to estimate the energy released during the interaction. AutoDock is an open-source and automated docking program developed by Molecular Graphics Lab, Scripps Research Institute, La Jolla, CA 92037, United States. It is applicable for the calculation of the interactions profile of biological macromolecules such as proteins/enzymes and ligands. AutoDock suite offers the minimum binding energy of interaction between the ligand and the receptor protein. The calculation of the binding energy is based on the scoring function, which in turn is based on the AMBER force field as well as linear regression analysis using the Lamarckian genetic algorithm. The default settings of AutoDock are commonly applied to comprehend the interaction profile of the ligand with the target. For each docking application, AutoDock is run several times to have several docked conformations of the ligand with the receptor protein. iGEMDOCK is another software used for molecular docking, virtual screening, and postscreening investigation. The iGEMDOCK success rate is 78% (RMSD <2.0 Å). Accurate docking was performed by iGEMDOCK, by setting the population size at 800, 80 generations in 10 number of solutions as the genetic algorithm parameters.

2.6 | MD simulation

The ligand 6 (PubChem CID: 20183791) and its complexes with SARS-CoV-2, SARS-CoV-, MERS-CoV, and HCoV- HKU1 Mpros were subjected to MD simulations to analyze the conformational flexibility and binding stability by using GROMACS (2018.2). The force field CHARMM 36 and CGenFF 4.1 were used to generate the topology, atomic, and charge parameters. The box sizes of complex systems were generated in the cubic unit cell. The box sizes of SARS-CoV-2, SARS-CoV, HCoV-HKU1, and MERS-CoV Mpro complex systems were set at 893.32, 831.3, 878.63, and 956.33 nm³ respectively. TIP3P water model was selected for the generation of protein topology files in the CHARMM 36 force field. Further, complexes of the ligand 6 with the target proteases were built and a dodecahedron edge box was generated. An aqueous environment was created around each complex. The system was neutralized using Na⁺ and Cl⁻ ions. The complex was energetically minimized using the steepest descent algorithm. After the minimization, it was equilibrated at constant temperature (300 K) and pressure (1 bar) to generate a conserve ensemble of Canonical ensemble (NVT): moles (N), volume (V) and temperature (T) and Isothermal-Isobaric (NPT) ensemble: moles (N), pressure (P) and temperature (T) for 100 ps time. After the equilibration, the complex system was subjected to run for 100 ns time of simulation with a step of 2 femtoseconds (fs) using leapfrog algorithm. After the
completion of the runs, the generated trajectories were analyzed for the structural deviation between the ligand and the complex system and graphs were generated using MATLAB.

3 | RESULTS AND DISCUSSION

The diagrammatic representation of the study is given as Figure 1.

3.1 | Virtual screening

The main aim of virtual screening is to assess a large collection of molecules against a specific target in a time and cost-effective manner. The virtual screening was performed using VEGA ZZ 3.2.0 and AutoDock Vina with 3254 ligands, out of which 115 best ligands were selected for docking studies.

3.2 | Molecular docking

AutoDock Tools were used for the multiple grid docking evaluations and docking validation was done through iGEMDOCK. The results of the best five ligands and the released binding energy upon interaction are represented in Tables 1 and 2. After molecular docking, the top 13 ligands were selected for drug-likeness analysis (Figure 2). Two physiological ligands (4 seconds)-2-methyl-2,4-pentanediol or MPD and n-[5-(methylisoxazol-3-yl]carbonyl]alaninyl-l-valyl-n-1-~-(1r,2z)-4-(benzoyloxy)-4-oxo-1-[[3r]-2-oxopyrrolidin-3 yl[methyl]but2-aryl]-l leucinamide or PRD) were used as reference and HCQ was used as the control drug. The minimum binding energy was exhibited by ligands 2 and 6 against the CoV Mpro's. The chemical structures of the 13 best ligands along with their chemical identification number are given as Figure 2. The docking interactions of the ligands are shown in Figure 3, along with the interacting amino acids (Figure 4). The heteroatoms of the ligands were involved in the hydrogen bonding, with the target proteins. Ligand 2 has two rings and an NH group in its structure. This functional group is well-known for its tendency to form hydrogen bond and is mainly responsible for the polar interactions between the ligand and the target Mpro. In addition, in the presence of aromatic rings, in its structure, ligand 2 was also involved in non-polar interactions. In the same way, N and O atoms of triazole and isoxazole rings of the ligand 6 were involved in the formation of hydrogen bonding as well as the dipole-induce dipole, dispersion force, London dispersion force, or van der Waals forces with the amino acids present in the active site. Owing to the presence of two-azole rings, ligand 6 may have potential inhibitory action against the viral Mpro. Most interacting amino acid residues are presented in Table S3 and Figure 5, in which hydrogen bond-forming residues are also included along with the bond distance and interacting atoms. The docking results were further validated through iGEMDOCK, and the results of the best two ligands (2, 6) are summarized in Table S3.

### Table 1 The binding energy of hit compound at different grid map of the target protein

| Identified hits | Identified hits | Identified hits | Identified hits | Identified hits | Identified hits |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Derivative      | Grid I          | Grid II         | Grid III        | Grid IV         | Grid V          |
| 1Q2W (SARS-CoV) | –10.0           | –10.1           | –9.8            | –9.8            | –9.8            |
| 6               | –10.4           | –10.5           | –10.6           | –9.9            | –10.6           |
| 11              | –9.4            | –9.5            | –9.5            | –9.9            | –9.9            |
| 3               | –9.2            | –9.4            | –9.4            | –8.7            | –7.1            |
| 5               | –8.7            | –9.8            | –7.9            | –8.4            | –7.4            |
| 2               | –10.2           | –10.1           | –10.1           | –8.9            | –8.9            |
| 12              | –10.0           | –10.1           | –10.1           | –8.9            | –10.1           |
| 6               | –9.1            | –8.3            | –8.3            | –8.4            | –8.4            |
| 5               | –8.2            | –7.9            | –7.9            | –8.6            | –8.6            |
| 10              | –8.0            | –6.9            | –6.9            | –7.7            | –7.7            |
| 11              | –9.3            | –9.3            | –9.3            | –8.9            | –8.9            |
| 12              | –8.9            | –9.2            | –9.2            | –10.0           | –10.0           |

The molecular docking results were in tandem with some of the previous computational and virtual screening studies undertaken in search of a probable Mpro inhibitor. However, most of the earlier studies have either used food and drug administration (FDA) approved drugs or naturally occurring phytochemicals as probable Mpro inhibitor (Table 3). The present study mainly explored natural as well as synthetic bioactive constituents of herbs and spices as prospective Mpro inhibitors, not only against SARS-CoV-2 but also against the other known coronaviruses. Another differentiating point of the study was the usage of three docking software (AutoDock Vina, AutoDock, and iGEMDOCK) for the binding energy calculation and validation.

3.3 | Pharmacokinetic study of the hit molecules

3.3.1 | Bioactivity score

Molinspiration was used for the analysis of bioactivity score against known drug targets like G protein-coupled receptors (GPCR) ligand, ion channel modulator (ICM), nuclear receptor ligand (NRL), kinase...
inhibitor (Ki), protease inhibitor (Pi), and enzyme inhibitor (Ei). The software has a broad range of cheminformatics tools associated with the molecular operation and requires simple handling and uses . smiles/sdf file (http://www.molinspiration.com). Molinspiration does molecular calculation with the help of the existing information in the molecular database and controls fragment-based virtual screening.47 A drug target can be a receptor protein, which is usually essential for the natural system. Some most acceptable drug targets are -

1. GPCR
2. Enzymes (mainly protein kinases, proteases, esterases, and phosphatases)
3. Ligand-gated ion channel
4. Ion channels Modulators

Almost all the top 13 ligands displayed moderate to good bioactivity score, showing their potential drug-like nature. Ligand 6 and 13 had positive bioactivity score against all the selected drug targets (Table 4).

3.4 Pharmacokinetic and drug-likeness assessment

Evaluation of ADMET properties is a prerequisite for the pre-clinical analysis.48 The drug development procedure involves substantial time and money. Computational and in silico prediction for important pharmacokinetic properties is an accurate and time-effective process. Different computational models have been used to predict ADMET (absorption, distribution, mechanism, excretion, and toxicity) parameters of lead compounds during the early stage of drug discovery, which may facilitate the development of new drugs with better ADMET properties. admetSAR is a free online database, that is, regularly updated and provides access to 96 000 unique compounds covering more than 50 categories with ADMET-allied properties49 such as blood-brain barrier (BBB) penetration, human intestinal absorption (HIA). The BBB restricts the passage of substances in the central nervous system and also regulates the movement of substances that are crucial for the healthy maintenance of the CNS and its functioning, thereby acting as a mediator between the CNS and blood. For effective drug development, it is almost essential to understand the physiology of BBB and the ability of the drug to cross over it. HIA is an important ADME parameter because of its relevance associated with the transportation of the drug to its target. The intestinal absorption of a molecule also influences the bioavailability of the drug. SwissADME is another web-based free tool, established by the molecular modeling group of the Swiss Institute of Bioinformatics (SIB). Cheminformatics has established many molecular descriptors based on chemical structures, including oral bioavailability and carcinogenicity.50,51 Ames toxicity assessment, which is used to assess the potential mutagenicity associated with a test chemical by treating the mutated Salmonella typhimurium strains with it. The top 13 ligands had positive HIA and BBB crossing potential. No carcinogenicity was found to be associated with any of the 13 ligands (Table 5).

### Table 2

| Derivative | 1Q2W | 3D23 | 4RSP | 6LU7 |
|-----------|------|------|------|------|
|           | BE   | IC   | BE   | IC   |
| 1         | -6.4 | 20.42| -5.41| 108.14|
| 2         | -7.1 | 6.2  | -6.93| 8.32 |
| 3         | -6.5 | 13.45| -6.70| 12.32|
| 4         | -4.18| 864.86| -4.55| 465.33|
| 5         | -6.45| 18.76| -6.30| 24.03|
| 6         | -7.06| 6.68 | -6.83| 9.84 |
| 7         | -5.81| 54.83| -5.59| 79.75|
| 8         | -5.92| 45.47| -5.04| 201.18|
| 9         | -5.95| 43.46| -5.93| 45.06|
| 10        | -6.61| 14.28| -6.58| 15.09|
| 11        | -6.81| 10.24| -6.88| 8.99 |
| 12        | -7.29| 4.57 | -6.21| 28.1 |
| 13        | -6.08| 34.84| -5.58| 81.69|
| MPD*      | -2.2 | 24.46| -2.48| 15.09|
| PRD*      | -3.54| 2.54 | -4.12| 1.01 |
| HCQ*      | -3.5 | 2.72 | -3.95| 1.27 |

Note: BE, binding energy; IC, inhibition constant (μM); MPD, ([4 seconds]-2-methyl-2,4-pentanediol; PRD, 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-ethyl)-l-leucinamide; HCQ, hydroxychloroqui. IC (mM).
To assess the HIA and BBB crossing ability, the Egan’s egg was used to create the BOILED egg model, which accurately predicts the physicochemical descriptors. The BOILED-Egg’s yolk represents the physicochemical space for a high probability of BBB permeability, whereas the Egg-white represents the physicochemical space for high HIA probability. The blue points on the plot are predicted to be actively effluxed by P-gp (PGP +) and red dots show the nonsubstrate of P-gp (PGP-). P-gp protects the central nervous system from xenobiotics. Points’ located within the BOILED-Egg’s yolk represent passive permeation of the molecules through the BBB as well as passive absorption by the gastrointestinal tract. The ligand 6, lied inside the white ellipse and the BOILED-Egg’s yolk, as shown by the findings (Figure 5).

Several parameters like molecular weight (MW), polar surface area, etc. influence the drug-likeness character. Large molecules have difficulty crossing over the biological membranes and it has been reported that 80% of the known drugs have their MW below 450 Da. The Polar surface area and the number of hydrogen bonds increase with the MW. Hydrogen bonding also has a vital significance in biological systems, as extensive hydrogen bonding causes a delay in the movement of drugs from gut to blood. Drugs having partial rigid character generally possess good to excellent oral bioavailability. LogP
FIGURE 3  A, Docking interactions of ligand 2 with A: 1Q2W, B: 3D23, C: 4RSP, and D: 6LU7; B, Docking interactions of ligand 6 with A: 1Q2W, B: 3D23, C: 4RSP, and D: 6LU7.
FIGURE 4  Interacting amino acids and released binding energy of ligand 2 and 6
represents the solubility of the proposed compounds affecting the capacity of a molecule to pass through the cell membrane. The MW of all the 13 ligands was within the limit of 500 and TPSA was below 140 Å² (Table 6).

The bioavailability radar is a pictorial representation and assessment of drug-likeness of a molecule. It is plotted for six physicochemical properties viz. LIPO (Lipophilicity): \(-0.7 < \text{XLOGP}_3 < +5.0\); SIZE: 150 g/mol < MW < 500 g/mol; POLAR (Polarity): 20Å² < TPSA <130 Å²; INSOLU (Insolubility): 0 < Log S (ESOL) < 6; INSATU (Insaturation): 0.25 < Fraction Csp3 < 1; FLEX (Flexibility): 0 < Num, rotatable bonds <9. The pink area represents the optimal range for each parameter. The colored zone in the radar shows the best suited physiochemical space for oral bioavailability. The bioavailability radar predicted promising results for ligand 6 (Figure 6).

3.5 | MD simulation analysis

MD simulation is performed to know the movement of atoms and molecules over a certain time. The simulation predicts the dynamic behavior of the interaction between the target and receptor. The results may be used to analyze the ligand’s confirmation and energy change upon interaction. The docked M\(_{\text{pro}}\) complexes with ligand 6 were simulated using the GROMACS package. The stability of the protein-ligand conformation can be estimated by the deviations observed during the simulation time. The smaller the deviations, stable would be the protein structure. The RMSD analysis is used for the assessment of the deviation in the backbone structure of the protein during the simulation time. The smaller the deviation between the starting and final conformations, the more stable would be the structure of the protein. RMSD was calculated for the C-alpha (Ca) to check the stability. Simulated complex systems showed no significant change in the Ca atoms of the ligated SARS-CoV, MERS-CoV, and SARS-CoV-2 M\(_{\text{pro}}\)s from the initial structure, whereas the ligated HCoV- HKU1 M\(_{\text{pro}}\) showed a significant change. The complexes of SARS-CoV, MERS-CoV, and SARS-CoV-2 M\(_{\text{pro}}\)s with the ligand 6 showed the maximum deviation around 0.20 nm from the initial structure, whereas the complex of HCoV-HKU1 M\(_{\text{pro}}\) with ligand 6 showed a deviation greater than 0.45 nm, depicting the instability of the HCoV-HKU1 complex in comparison to the other proteins. The overall complex trajectories were stable during the time of simulation (Figure 7A). The RMSD plot of the ligand suggested its stable interaction with the target M\(_{\text{pro}}\)s during the time of simulation except for MERS-CoV M\(_{\text{pro}}\) which showed a deviation of 1.2 nm (Figure 7B). The high deviation in the ligand RMSD plot of MERS-CoV M\(_{\text{pro}}\) might be due to the unstable interaction with the protein, as the trajectory visualization showed that the ligand moved to positions other than the binding pocket.

The radius of gyration (Rg) depicts the atoms mass-weighted root mean square distance of atoms from the center of mass and shows the stability of the complex. The fitness of folding, compactness, and target’s structure, for the simulation time of 100 ns are presented in the Rg plot. During the simulation time, the ligated MERS-CoV and SARS- CoV-2 M\(_{\text{pro}}\)s showed constant Rg values of 2.22 nm and the ligated SARS-CoV M\(_{\text{pro}}\) showed less fluctuation in the Rg value viz. 2.25 nm, as compared to the initial time to 35 ns. The ligated HCoV-HKU1 M\(_{\text{pro}}\) showed a higher Rg value viz., 2.30 nm after the initial time (5 ns) (Figure 8). The higher the Rg value, the greater would be the change in the structural compactness and folding of the protein.
Solvent-accessible surface area (SASA) of ligated Mpro was studied to assess the changes in the hydrophobic and hydrophilic residues. The target-ligand complexes showed almost similar SASA values between 168 and 170 nm², except ligated SARS-CoV Mpro, showing a lesser value (163 nm²) during the simulation time (Figure 9).

To further analyze the complex’s structural stability and the effect of the ligand on the flexibility of the structure of the target, some previous virtual studies to identify Mpro inhibitor were done. The results are presented in Table 3.

| S. No. | Target/PDB ID | Docking tool | Number of ligands | Lead molecule | Binding Energy (kcal/mol) | References |
|--------|---------------|--------------|-------------------|---------------|--------------------------|------------|
| 1      | SARS-CoV-2 2/6LU7 | AutoDock Vina | 100 (Fungal metabolites) | Pyranonigrin A | −7.3 | 40 |
| 2      | SARS-CoV-2 3CLpro/6Y2F; SARS-CoV 3CLpro/3TNT; MERS-CoV 3CL/5WKK | AutoDock Vina | 263 (Phytochemicals) and 75 (FDA approved antiviral drugs) | Glecaprevir (DB13879) | −9.3 (6Y2F); −9.7 (3TNT); | 41 |
|        |               |              |                   | Daclatasvir (DB09102) | −9.2 (6Y2F); −9.3 (3TNT); |           |
|        |               |              |                   | Paritaprevir (DB09297) | −9.3 (6Y2F); −8.7(5WKK) |           |
|        |               |              |                   | Atazanavir (DB01072) | −9.2 (3TNT); −9.1(5WKK) |           |
|        |               |              |                   | Vincapuse (11646359) | −7.5 (6Y2F); −8.3 (3TNT); −9.7(5WKK) |           |
|        |               |              |                   | Alloyoimine (120716) | −8.0 (6Y2F); −9.0 (3TNT); −8.6(5WKK) |           |
|        |               |              |                   | Gummadiol (10308017) | −7.8 (6Y2F); −8.4 (3TNT); −8.4(5WKK) |           |
|        |               |              |                   | 18-Hydroxy-3-epi-alpha-yohimbine (102004710) | −8.1 (6Y2F); −8.0 (3TNT) |           |
| 3      | SARS-CoV-2 Mpro/6Y2F | AutoDock Vina | 77 (FDA approved antiviral drugs) | Lopinavir-Ritonavir | −10.6 | 42 |
|        |               |              |                   | Tipranavir | −8.7 |           |
|        |               |              |                   | Raltegravir | −8.3 |           |
| 4      | SARS-CoV-2 Mpro/5R7Y | AutoDock Vina | 61 (SARS 3CLpro inhibitors) | Molecule ID 19 | −8.6 | 43 |
|        |               |              |                   | Molecule ID 43 | −8.6 |           |
|        |               |              |                   | Molecule ID 44 | −8.5 |           |
|        |               |              |                   | Molecule ID45 | −8.6 |           |
|        |               |              |                   | Molecule ID 49 | −8.5 |           |
|        |               |              |                   | Molecule ID 58 | −8.8 |           |
| 5      | SARS-CoV-2 2/6LU7 | Glide | 3992 (FDA approved drugs) and 175 815 (Asinex BioDesign Library) | Ritonavir | −8.01 | 44 |
|        |               |              |                   | Disubstituted pyrazole-based (1) | −9.02 |           |
|        |               |              |                   | Disubstituted pyrazole-based (2) | −8.725 |           |
|        |               |              |                   | Cyclic amide based (12) | −8.273 |           |
| 6      | SARS-CoV-2 Mpro/7BQY | AutoDock Vina | >100 | Camptothecin gimatecan | −11.07 | 45 |
|        |               |              |                   | Lamellarin D | −11.32 |           |
|        |               |              |                   | Leopolic acid A | −12.22 |           |
|        |               |              |                   | Stachys ehrenbergii | −13.07 |           |
| 7      | SARS-CoV-2 Mpro/6LU7 | AutoDock Vina | 1051 (FDA approved drugs) | R428 | −10.5 | 46 |
|        |               |              |                   | Teniposide | −9.8 |           |
|        |               |              |                   | VS5584 | −9.4 |           |
|        |               |              |                   | Setileuton | −8.5 |           |
protein, the root mean square fluctuation (RMSF), per residue of all the target proteins’ backbone, was studied. To understand the changes related to the ligand, the RMSF plots of the ligated Mpros were generated for ligand 6 and evaluated for atom wise fluctuations in the molecule. The RMSF plot suggested that the target-ligand complex had higher fluctuations and low stability (Figures 10 and 11).

The interaction strength between the ligand and the target was quantified by the analysis of hydrogen bonding and non-bonding energies, throughout the simulation time. Ligand 6 interacted and formed hydrogen bonds with the target Mpros for the whole simulation time (Figure 11).

Interaction energy and stability of the complexes were studied by the analysis of Coulomb and Lennard-Jones interaction energies. The ligand 6 exhibited better Coulomb-Short Range (Coulumb-SR) binding energy (−175 kJ/mol) with SARS-CoV and HCoV- HKU1 Mpros. The Coulumb-SR interactions with SARS-CoV-2 Mpro showed binding energy of 100 kJ/mol from the starting to 10 ns and between 60 and 100 ns (Figure 12A). Between 10 and 60 ns, the ligand did not show any Coulomb-SR binding with MERS-CoV Mpro. Ligand 6 showed better Lennard-Jones interaction energy (Figure 12B) with SARS-CoV and HCoV- HKU1 Mpros, exhibiting better Lennard-Jones interaction energies with SARS-CoV and HCoV- HKU1 Mpros between 10 and 60 ns (Figure 12B).

### Table 4: Bioactivity score

| Derivative | GPCR | ICM | KI | NRL | PI | EI |
|------------|------|-----|----|-----|----|----|
| 1          | 0.33 | 0.76| 0.41| 0.45| 0.38| 0.24|
| 2          | 0.14 | 0.34| 0.25| 0.42| 0.41| 0.38|
| 3          | 0.10 | 0.22| 0.21| 0.01| 0.14| 0.09|
| 4          | 0.09 | 0.22| 0.26| 0.16| 0.10| 0.16|
| 5          | 0.04 | 0.26| 0.37| 0.33| 0.07| 0.09|
| 6          | 0.53 | 0.29| 0.13| 0.06| 0.23| 0.21|
| 7          | 0.46 | 0.50| 0.85| 0.43| 0.47| 0.44|
| 8          | 0.62 | 0.67| 1.07| 0.70| 0.62| 0.53|
| 9          | 0.54 | 0.51| 0.94| 0.51| 0.53| 0.48|
| 10         | 0.56 | 0.57| 0.96| 0.59| 0.52| 0.46|
| 11         | 0.10 | 0.20| 0.27| 0.18| 0.10| 0.01|
| 12         | 0.14 | 0.47| 0.59| 0.47| 0.04| 0.10|
| 13         | 0.14 | 0.07| 0.51| 0.12| 0.01| 0.46|

**Abbreviations:** El, enzyme inhibitor; GPCR, G-protein-coupled receptor; ICM, ion channel modulator; KI, kinase inhibitor; NRL, nuclear receptor ligand; PI, protease inhibitor.

### Table 5: Absorption, distribution, mechanism, excretion, and toxicity (ADMET) analysis

| Derivative | HIA | HOB | Caco-2 | PPB | BBB | AMES | Carcino | Biodeg |
|------------|-----|-----|--------|-----|-----|------|----------|--------|
| 1          | +/0.9850 | −/0.6286 | −/0.8385 | 0.936 | +/0.9401 | −/0.5700 | −/0.9714 | −/0.8750 |
| 2          | +/0.9683 | +/0.5571 | −/0.7407 | 1.090 | +/0.9890 | −/0.5900 | −/0.8857 | −/0.9500 |
| 3          | +/0.9708 | +/0.5143 | −/0.6371 | 1.116 | +/0.9793 | −/0.6800 | −/0.8031 | −/0.7500 |
| 4          | +/0.9514 | +/0.6714 | −/0.5257 | 1.199 | +/0.9855 | −/0.5900 | −/0.9286 | −/0.9500 |
| 5          | +/0.9791 | +/0.6571 | −/0.8786 | 1.115 | +/0.9749 | −/0.6800 | −/0.7745 | −/0.9250 |
| 6          | +/0.9866 | +/0.6286 | −/0.5850 | 0.874 | +/0.9857 | −/1.0000 | −/0.5900 | −/0.8500 |
| 7          | +/0.9785 | +/0.5714 | +/0.5949 | 1.321 | +/0.9779 | −/0.6300 | −/0.8000 | −/0.9000 |
| 8          | +/0.9409 | −/0.5286 | −/0.6278 | 1.272 | +/0.9730 | +/0.6600 | −/0.8714 | −/0.9250 |
| 9          | +/0.9490 | +/0.6571 | −/0.6630 | 1.330 | +/0.9740 | −/0.6900 | −/0.8000 | −/0.9250 |
| 10         | +/0.9618 | +/0.6286 | −/0.5962 | 1.250 | +/0.9730 | +/0.5400 | −/0.8000 | −/0.9750 |
| 11         | +/0.9878 | +/0.7857 | −/0.7812 | 1.347 | +/0.9781 | −/0.6700 | −/0.7429 | −/0.9500 |
| 12         | +/0.9899 | +/0.6714 | −/0.8151 | 0.947 | +/0.9763 | −/0.5200 | −/0.8286 | −/0.8750 |
| 13         | +/0.9660 | −/0.5000 | +/0.5349 | 0.953 | +/0.9698 | +/0.6600 | −/0.8714 | −/0.9000 |

**Abbreviations:** AMES, Ames mutagenesis; BBB, blood–brain barrier; Biodeg, biodegradability; Caco-2, Caco-2 permeability; Carcino, carcinogenesis; HIA, human intestinal absorption; HOB, human oral bioavailability; PPB, plasma protein binding (Unit = 100%).
CONCLUSION

The main protease enzyme is chymotrypsin-like cysteine, which plays an important role in governing the replication and transcription of coronavirus. In this study, virtual interaction of main proteases of SARS-CoV-2, SARS-CoV, MERS-CoV, and HCoV-HKV1 with active constituents of some common herbs and spices, as the plant-based medicinal system is as old as mankind itself. The derivatives were explored and filtered through Lipinski’s parameters. HCQ was used as a reference drug. The interactions showed positive results with some of the derivatives and the best interaction was observed in the case of dihydroquinoline derivative, 3,4-Dihydro-2H-quinolin-1-yl-[5-methyl-1-(3-phenyl-2,1-benzoxazol-5-yl)triazole-4 yl]methanone and obtained from Flinder rose (caper), belonging to Capparidacea.

### Table 6

| Derivative | MW     | nRB | nOH | nOHNH | miLogP | TPSA  | Abs% | nAtoms | MR     | Lipinski violation |
|------------|--------|-----|-----|-------|--------|-------|------|--------|--------|-------------------|
| 1          | 445.48 | 5   | 8   | 3     | 3.70   | 113.43| 69.87| 33     | 113.56 | 0                 |
| 2          | 435.49 | 4   | 5   | 0     | 5.13   | 77.06 | 82.41| 33     | 129.04 | 0                 |
| 3          | 452.50 | 4   | 5   | 0     | 4.65   | 57.69 | 89.10| 31     | 118.17 | 0                 |
| 4          | 352.29 | 3   | 4   | 2     | 2.98   | 58.20 | 88.92| 25     | 85.34  | 0                 |
| 5          | 410.50 | 3   | 7   | 2     | 3.10   | 93.09 | 76.88| 29     | 115.48 | 0                 |
| 6          | 400.52 | 4   | 4   | 2     | 5.36   | 46.29 | 93.03| 30     | 124.85 | 0                 |
| 7          | 431.43 | 3   | 3   | 0     | 5.22   | 37.38 | 96.10| 31     | 111.73 | 1                 |
| 8          | 434.45 | 3   | 8   | 0     | 4.01   | 101.67| 73.92| 32     | 121.66 | 0                 |
| 9          | 458.44 | 4   | 6   | 0     | 5.02   | 83.20 | 80.20| 33     | 120.60 | 0                 |
| 10         | 440.50 | 3   | 6   | 0     | 5.30   | 83.20 | 80.20| 33     | 133.10 | 0                 |
| 11         | 451.59 | 4   | 6   | 2     | 4.80   | 82.27 | 80.62| 32     | 132.49 | 0                 |
| 12         | 466.54 | 5   | 7   | 3     | 4.63   | 99.66 | 74.62| 35     | 145.90 | 0                 |
| 13         | 425.44 | 2   | 7   | 2     | 4.85   | 96.96 | 75.55| 32     | 126.59 | 0                 |

Abbreviations: Abs %, percentage of absorption; miLog P, logarithm of compound partition coefficient between n-octanol and water; MR, molecular refractivity; MW, molecular weight; nAtoms, number of atoms; nON, number hydrogen bond acceptor; nOHNH, number hydrogen bond donor; nRB, number of rotatable bonds; TPSA, topological polar surface area.

**Figure 6** Representation of bioavailability radar for the selected lead derivatives

4 | CONCLUSION
family and used as a seasoning agent and the fruit is used as a pickle and 6-[2-(4-Phenyl-1,2,3,6-tetrahydropyridin-2-yl)ethyl]-6,7,8,9-tetrahydro-5H-[1,3]dioxolo[4,5-b]carbazole, a potential derivative of carbazole, which is obtained from sweet neem belonging to Rutaceae family is also a potent antioxidant agent. The findings pave way for further exploration of the active constituents of natural plant-based resources for an effective treatment process to combat SARS-CoV-2 and other microbial infections with minimal side effects.
The authors are thankful to the Research and Development Committee, Integral University for the support. Dr A. M. Alanazi acknowledges the Research Supporting Project Number (RSP-2020/261), King Saud University, Riyadh, Saudi Arabia. Dr Nidhi Mishra is thankful to the Central Computing facility, IIIT-Allahabad.

CONFLICT OF INTEREST
The authors declare they have no competing interests.

DATA AVAILABILITY STATEMENT
The data that supports the findings of this study are available in the supplementary material of this article.
57. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem. 2002;45:2615-2623.

58. Azad I, Nasibullah M, Khan T, Hassan F, Akhter Y. Exploring the novel heterocyclic derivatives as lead molecules for design and development of potent anticancer agents. J Mol Graph Model. 2018;81:211-228. https://doi.org/10.1016/j.jmgm.2018.02.013.

59. Aier I, Varadwaj P, Raj U. Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. Sci Rep. 2016;6:34984. https://doi.org/10.1038/srep34984.

60. Maurya AK, Mulpuru V, Mishra N. Discovery of novel coumarin analogs against the α-glucosidase protein target of diabetes mellitus: pharmacophore-based QSAR, docking, and molecular dynamics simulation studies. ACS Omega. 2020;5(50):32234-32249. https://doi.org/10.1021/acsomega.0c03871.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: I. Azad, T. Khan, A. K. Maurya, M. Irfan Azad, N. Mishra, and A. M. Alanazi, Identification of Severe Acute Respiratory Syndrome Coronavirus-2 Inhibitors through in silico structure-based virtual screening and molecular interaction studies, J Mol Recognit (2021), 1-17. https://doi.org/10.1002/jmr.2918