Taylor & Francis
Taylor & Francis Group

Check for updates

A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2

Rajib Islam[a], Md. Rimon Parves[a], Archi Sundar Paul[a], Nizam Uddin[a,b], Md. Sajjadur Rahman[a,b], Abdulla Al Mamun[c], Md. Nayeem Hossain[a], Md. Ackas Ali[a] and Mohammad A. Halim[a,d]

[a]Division of Computer Aided Drug Design, The Red-Green Research Centre, BICCB, Dhaka, Bangladesh; [b]Department of Chemistry and Biochemistry, South Dakota State University, Brookings, SD, USA; [c]Key Laboratory of Soft Chemistry and Functional Materials of MOE, School of Chemical Engineering, Nanjing University of Science and Technology, Nanjing, China; [d]Department of Physical Sciences, University of Arkansas-Fort Smith, Fort Smith, AR, USA

Communicated by Ramaswamy H. Sarma

ABSTRACT

The main protease of SARS-CoV-2 is one of the important targets to design and develop antiviral drugs. In this study, we have selected 40 antiviral phytochemicals to find out the best candidates which can act as potent inhibitors against the main protease. Molecular docking is performed using AutoDock Vina and GOLD suite to determine the binding affinities and interactions between the phytochemicals and the main protease. The selected candidates strongly interact with the key Cys145 and His41 residues. To validate the docking interactions, 100 ns molecular dynamics (MD) simulations on the five top-ranked inhibitors including hypericin, cyanidin 3-glucoside, baicalin, glabridin, and α-ketoamide-11r are performed. Principal component analysis (PCA) on the MD simulation discloses that baicalin, cyanidin 3-glucoside, and α-ketoamide-11r have structural similarity with the apo-form of the main protease. These findings are also strongly supported by root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (Rg), and solvent accessible surface area (SASA) investigations. PCA is also used to find out the quantitative structure-activity relationship (QSAR) for pattern recognition of the best ligands. Multiple linear regression (MLR) of QSAR reveals the R^2 value of 0.842 for the training set and 0.753 for the test set. Our proposed MLR model can predict the favorable binding energy compared with the binding energy detected from molecular docking. ADMET analysis demonstrates that these candidates appear to be safer inhibitors. Our comprehensive computational and statistical analysis show that these selected phytochemicals can be used as potential inhibitors against the SARS-CoV-2.

Molecular interaction

Binding Pocket
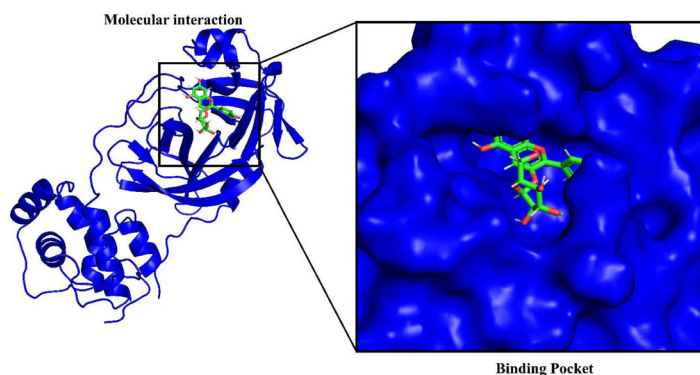
1. Introduction

In December, the first epidemic of 2019 novel coronavirus (SARS-CoV-2) took place in Wuhan city, China (Hasan et al., 2020; Lu et al., 2020; Wu et al., 2020). Since the outbreak, it has been rapidly infecting people around the world and turned into a pandemic. The worldwide number of corona virus cases reached to 2,621,449 with a death toll of 182,991 as of April 22, 2020 (Worldometers.info, 2020). Although it

was started from China, the number of active cases in United States, Spain, Italy, France, Germany, and UK have surpassed the cases identified in China (82,788). It is projected that the total COVID-19 deaths will reach to around 81,114 in US only and approximately a total of 60,000 deaths estimated in Spain, Italy, and France by August 4, 2020 (Team, I. C.-19 Health Service Utilization Forecasting & Murray, 2020).

The Main protease (Mpro) also known as 3-C like protease (3CLpro) received great attention because of its important role in post-translational processing of replicase polyproteins (Boopathi et al., 2020; Elmezayen et al., 2020; Khan, Jha, et al., 2020; Wang et al., 2016). The enzymatic activity of this protein leads to processing of viral polyproteins. The ∼306 amino acid long main protease has high structural and sequence similarity to that of SARS-CoV 3CLpro (Liu & Wang, 2020). The SARS-CoV-2 Main protease (Mpro) monomer consists of N-terminal domain-I, N-terminal domain-II, and C-terminal domain-III (Mirza & Froeyen, 2020). The active site of enzyme contains a catalytic dyad having Cys145 and His41 (Bacha et al., 2004; Khan, Zia, et al., 2020). HIV drugs including lopinavir and ritonavir have been explored recently against MERS-CoV (Sheahan et al., 2020). In addition, peptidomimetic α-ketoamides were designed and synthesized to test their performance against the main proteases of betacoronaviruses, alphacoronaviruses, and the 3CLpro of enteroviruses. In a recent follow-up study, Zhang et al. resolved the crystal structure of Mpro complex of SARS-CoV-2 with modified α-ketoamide inhibitors (Zhang et al., 2020). After the outbreak, several types of drugs alone or with combination have been using in many countries (Colson et al., 2020; Gautret et al., 2020; Wang et al., 2020). However, still there is no definite antiviral drug to fight against the deadly virus.

Herb species and fruits have been serving patients as sources of herbal medicine for a long time through the human history. They contain a wide variety of phytochemicals, such as flavonoids, alkaloids, glucosides, and polyphenolic compounds. These phytochemicals offer a wide range of therapeutic properties and novel scaffolds to design new drugs (Aanouz et al., 2020; Gupta et al., 2020). Among them, some phytochemicals showed high antiviral activity against a number of viral infections. For example, baicalin is an experimentally proved antiviral agent against several viruses, e.g. SARS-CoV-1 (Chen et al., 2004), SARS-CoV-2 (Liu et al., 2020), H1N1-pdm09 (Nayak et al., 2014), and Chikungunya (Oo et al., 2018).

Therefore, an efficient approach is to test the efficacy of antiviral phytochemicals against the 2019 novel SARS Coronavirus (Jassim & Naji, 2003; Li et al., 2020; Naithani et al., 2008; Tong, 2009). Herein, we have selected 40 known phytochemicals isolated from natural herbs and fruits. These phytochemicals have already showed antiviral activity against different viruses. The aim of this study is to explore and identify the binding affinities and interactions of these antiviral phytochemicals against the main protease of SARS-CoV-2 using computational and statistical tools. Molecular docking, molecular dynamics, principal component analysis (PCA), and quantitative structure-activity relationships (QSAR) are performed to assess the performance of these phytochemicals. In addition, absorption, distribution, metabolism, and excretion properties of the best candidates are evaluated.

2. Methods and computational details

2.1. Phytochemicals optimization, protein preparation, and molecular docking

Total forty phytochemicals were selected considering their proved antiviral activities (Table S1). Optimization of the phytochemicals and calculation of vibrational frequency were performed using Gaussian 09 software package (Frisch et al., 2009). The structure of the phytochemicals was optimized at semi-empirical PM6 method (Stewart, 2007). The crystal structure of the main protease was taken from the RSCB Protein Data Bank (PDB ID: 6LU7). Then the crystal structure of the protease was optimized and checked by Swiss-PDB viewer software packages (version 4.1.0) based on their least energy. Some significant factors, such as improper bond order, side chain geometry, and missing hydrogen, were observed in the crystal structure of the protease. PyMol (version 1.1) software package was used to erase all the hetero atoms, water molecules, and inhibitor present in the structure (DeLano, 2002). Finally, the non-covalent interaction of phytochemicals-protease was calculated using Autodock Vina software package for the docking analysis (Trott & Olson, 2009). Using this method, binding affinities of ligand-protease were determined and reported in kcal/mol unit. The grid box in Autodock Vina was generated targeting the active site of the main protease, where the center was at X: −11.76, Y: 15.17, Z: 69.19 and the dimensions of the grid box were, X: 25.22, Y: 29.33 and Z: 29.22 (unit of the dimensions, Å). To get more insights of these results, GOLD (Jones et al., 1997) docking program was also employed.

2.2. Molecular dynamics

The molecular dynamics (MD) simulation was performed on the best five selected phytochemicals obtained from molecular docking study, which helped to get more insight into the protein and docked complexes in biological condition. In this study, MD simulation was conducted by YASARA Dynamics software (Krieger et al., 2012). The AMBER14 force field was employed for this study to describe the macromolecular system (Dickson et al., 2014). Water and Na^+/Cl^- ions were also added to the system. Periodic boundary condition was incorporated to perform the simulation, where the cell size was 20 Å larger than the protease size in all cases. By employing steepest gradient approach (5000 cycles), the initial energy minimization for each system was performed. MD simulations were performed using the PME method to designate long-range electrostatic interactions at a cut off distance of 8 Å, and defining physiological conditions at 310 K, pH 7.4, 0.9% NaCl (Krieger et al., 2006). The simulation temperature was controlled using the Berendsen thermostat, where the pressure kept constant throughout the simulation. A multiple time step algorithm was employed, where the simulation
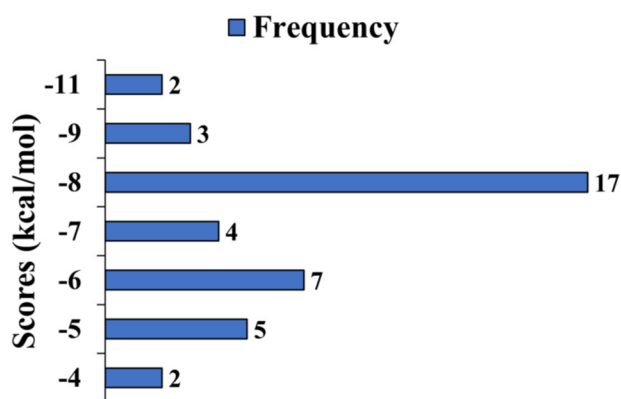
Figure 1. Frequency distribution of 40 phytochemicals over the range of docking scores.

time step was selected as 1.25 fs (Krieger & Vriend, 2015). Finally, MD simulation was performed for 100 ns long and snapshots were saved at every 100 ps into MD trajectory for further analysis. Root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), and total number of H-bond count were calculated from the MD simulations similar to other studies (Elfiky & Azzam, 2020; Enayatkhani et al., 2020; Muralidharan et al., 2020; Pant et al., 2020).

2.3. Pharmacokinetic parameters

Toxicity and some pharmacokinetic parameters were predicted using admetSAR online (Cheng et al., 2012). In early stage of drug discovery, pharmacokinetic study, and data analysis assist scientists to identify safe and effective drug candidates.

2.4. Principal component analysis (PCA) on MD simulation data

The principal component analysis is a widely used unsupervised data reduction method. Here, in this project, the goal of applying PCA method is to highlights the similarity and dissimilarity in the collected structural and energy profile of MD trajectory data (Martens & Naes, 1992; Wold et al., 1987). Using this technique, any structural quality change during MD can be characterized by comparing different drug-protein complexes. The following equation highlights the important components of a PCA model: $X = T_k P_k^T + E$, where, X matrix is expressed as a product of two new matrices, i.e. T_k and P_k, T_k serves as the matrix of scores that represents how samples relate to each other, P_k represents the matrix of loadings which contain information about how variables relate to each other, k is the number of factors included in the model, and E is the matrix of residuals.

These residuals contain the unmodeled variances. Complexes of the main protease with the selected five phytochemicals may have differences with the main protease, i.e. apo-protein, during MD simulations in terms of structural and energy profile. These differences can be detected by the PCA algorithm (Islam et al., 2019). All calculations were performed

Table 1. Docking results of all phytochemicals with main protease of SARS-CoV-2 (AutoDock Vina scores are in kcal/mol and GOLD scores are dimensionless).

Ligand name	AutoDock Vina	GOLD
Hypericin	−10.7	80.15
Pseudohypericin	−10.7	85.31
Cyanidin 3-Glucoside	−8.4	81.71
Baicalin	−8.1	59.19
Glabridin	−8.1	63.68
Glycyrrhizin	−7.9	60.37
α-Ketoamide-11r	−7.8	93.07
Isobavachalcone	−7.8	78.59
Apigenin	−7.7	61.85
Betulinic Acid	−7.6	50.96
Oleuropein	−7.6	78.78
Quercetin	−7.6	66.11
Luteolin	−7.5	60.33
Oleanolic Acid	−7.5	49.4
Psoralidin	−7.5	62.31
Sageone	−7.5	62.02
Ursolic Acid	−7.5	46.43
Cystoketal	−7.4	65.35
Emodin	−7.3	56.4
Dithymoquinone	−7.2	42.44
Rosmarinic Acid	−7.2	70.63
Liquiritigenin	−7.1	58.53
Curcumin	−6.9	70.18
Cinanserin	−6.7	66.07
Safficinolide	−6.6	52.89
Lapachol	−6.3	55
Hibiscus Acid	−5.9	36.75
Gingerol	−5.4	62.7
Hydroxytyrosol	−5.3	44.08
Zingerone	−5.3	48.64
Carvacrol	−5.2	43.9
Cinnamic	−5.2	44.24
Methyl Cinnamate	−5.1	42.05
Thymohydroquinone	−5	47.77
Thymoquinone	−5	42.44
Thymol	−4.9	45.1
Cinnamaldehyde	−4.6	39.1
Ajoene	−4.2	48.47
Allicin	−3.3	37.59
Diallyl Trisulfide	−3.3	41.64

on R platform using in-house developed codes (R Core Team, 2019), and plots were generated using the package ggplot2 (Wickham, 2009). Data were preprocessed using autoscale function before applying PCA algorithm (Martens & Naes, 1992). The final 50 ns of MD trajectory data were used for analyzing the PCA.

2.5. Quantitative structure-activity relationships (QSAR) of phytochemicals

Forty potential ligands were randomly divided into a training set with 25 ligands and test set containing 15 ligands. The test set was utilized as the validation samples. TPSA (topological polar surface area, Å^2), molecular weight, XLogPs-AA, HBD, ROTB count, no. of H, C, O, Cl, N, and F atoms, single bonds (SB) count, double bonds (DB) count, and no. of benzene rings of the drug candidates were the considered as variables. These variables with calculated binding energies were used to correlate with structure-activity relationship using multiple linear regression (MLR) (Fakayode et al., 2009; Liu et al., 2017; Mark & Workman, 2007). Multiple liner regression was performed using XLSTAT (Adinsoft, 2010).

Table 2. Nonbonding interactions of selected five phytochemicals with main protease of SARS-CoV-2 (pose predicted by AutoDock Vina) where, CH = Conventional Hydrogen bond, H = hydrogen bond, C = carbon hydrogen bond, A = alkyl.

Interacting residue	Distance	Bond category	Bond Type
α-Ketoamide-11r			
ASN142	2.49	H	CH
GLY143	2.58	H	CH
GLY143	2.48	H	CH
SER144	2.09	H	CH
SER144	2.13	H	CH
CYS145	2.68	H	CH
PHE140	2.73	H	CH
HIS164	2.89	H	CH
GLY143	2.54	H	CH
HIS41	2.87	Hydrophobic	Pi-Sigma
MET49	4.92	Hydrophobic	Alkyl
Baicalin			
PRO168	2.96	H	CH
GLU166	2.19	H	CH
GLU166	2.53	H	CH
SER144	3.06	H	CH
GLU166	3.01	H	C
GLU166	2.08	H	C
CYS145	5.27	Other	Pi-Sulfur
MET49	5.18	Hydrophobic	Pi-Alkyl
Cyanidin 3-Glucoside			
GLN189	3.02	H	CH
LEU141	2.52	H	CH
THR26	2.84	H	CH
ASP187	2.75	H	CH
GLU166	2.46	H	CH
GLY143	2.84	H	C
MET49	4.93	Hydrophobic	Pi-Alkyl
MET49	4.31	Hydrophobic	Pi-Alkyl
CYS145	5.17	Hydrophobic	Pi-Alkyl
Glabridin			
GLU166	4.10	Electrostatic	Pi-Anion
MET49	3.76	Hydrophobic	Alkyl
MET49	4.86	Hydrophobic	Alkyl
MET165	4.79	Hydrophobic	Pi-Alkyl
HIS41	4.64	Hydrophobic	Pi-Alkyl
HIS41	3.91	Hydrophobic	Pi-Alkyl
Hypericin			
GLU166	2.41	H	CH
LEU141	2.83	H	CH
ASN142	2.95	H	C
GLU166	2.99	H	Pi-Donor H
GLU166	2.69	Hydrophobic	Pi-Sigma
GLN189	2.50	Hydrophobic	Pi-Sigma
MET165	4.32	Hydrophobic	Alkyl
MET165	4.35	Hydrophobic	Pi-Alkyl
CYS145	5.05	Hydrophobic	Pi-Alkyl

3. Results

3.1. Molecular docking

By using Autodock Vina, molecular docking is performed to find out the best candidates among the 40 phytochemicals based on their binding scores. Binding affinities of the phytochemicals are distributed within the range of −3.0 to −4.0, −4.1 to -5.0, −5.1 to −6.0, −6.1 to −7.0, −7.1 to −8.0, −8.1 to −9.0, and −10.0 to −11.0 kcal/mol (Figure 1). Selected compounds are screened primarily using AutoDock vina scoring function to find out the best candidates, then the GOLD suit was employed to understand their fitness. The ChemPLP fitness score is used in GOLD docking, which is the default scoring function of GOLD software program. In GOLD scoring system, higher fitness score indicates the better docking interaction between ligand and protein. The binding affinity and

fitness score of all phytochemicals are shown in Table 1. Based on the best binding affinities, hypericin, cyanidin 3-glucoside, baicalin, and glabridin are selected for further analysis. In this study, α-ketoamide-11r is considered as a control ligand because it is recently reported as a good inhibitor against main protease (Zhang et al., 2020), which shows binding affinity of -7.8 kcal/mol. Hypericin and pseudohypericin show the highest binding affinity of -10.7 kcal/mol. As both of them are structurally similar, only hyperici is selected for further study.

3.2. Molecular interaction of the selected phytochemicals with the main protease

Analysis of the non-covalent interactions of the best five phytochemicals with the main protease reveals that the selected compounds interact with either both (Cys145 and His41) or at least one catalytic residue detected by Autodock Vina, as shown in Table 2 and Figure 2 (interaction detected by GOLD is summarized in Table S2). The α-ketoamide-11r is stabilized by nine hydrogen bonds and two hydrophobic bonds while interacting with the receptor protein. It also forms hydrogen bonds with catalytic residue Cys145 and hydrophobic interaction with His41. Baicalin interacts through six hydrogen bonds, one hydrophobic interaction, and one pi-sulfur interaction with the catalytic residue Cys145. Cyanidin 3-glucoside forms six hydrogen bonding interactions and three hydrophobic interactions in which one hydrophobic interaction is observed with the catalytic residue Cys145. Glabridin gets stabilized by one electrostatic, five hydrophobic interactions, and no hydrogen bonding interaction is observed. Hypericin forms four hydrogen bonding interactions and five hydrophobic interactions in which one pi-alkyl interaction is observed with catalytic residue Cys145.

3.3. Molecular dynamics simulation

The RMSD of alpha carbon atoms of all systems are analyzed to detect their stability. It is observed from Figure 3a that α-ketoamide-11r complex exhibits the lowest RMSD than other complexes. Even RMSD of the apo-protein is slightly higher than the α-ketoamide-11r, which indicates the greater stability of α-ketoamide-11r. The RMSD of baicalin-protein complex reaches to ∼1.63 Å from 10 to 50 ns, however, this value significantly increases after 50 ns and reaches to 2.24 Å. While assessing the RMSD of cyanidin 3-glucoside complex, the steady increase of RMSD is observed after 21 ns (average RMSD >3.0 Å). Nonetheless, this value is decreased eventually, which indicates that cyanidin 3-glucoside may change the protein conformation. Unlike apo-protein and α-ketoamide-11r complex, RMSD of the glabridin complex is mostly stable. But the complex is found exhibiting its increased RMSD from 45.6 to 78 ns (average RMSD >2.74 Å), and subsequently the complex gets stable. Particularly, the hypericin complex shows consistent fluctuation from 11 to 100 ns. This complex also shows the higher deviation of fluctuations throughout the trajectory.
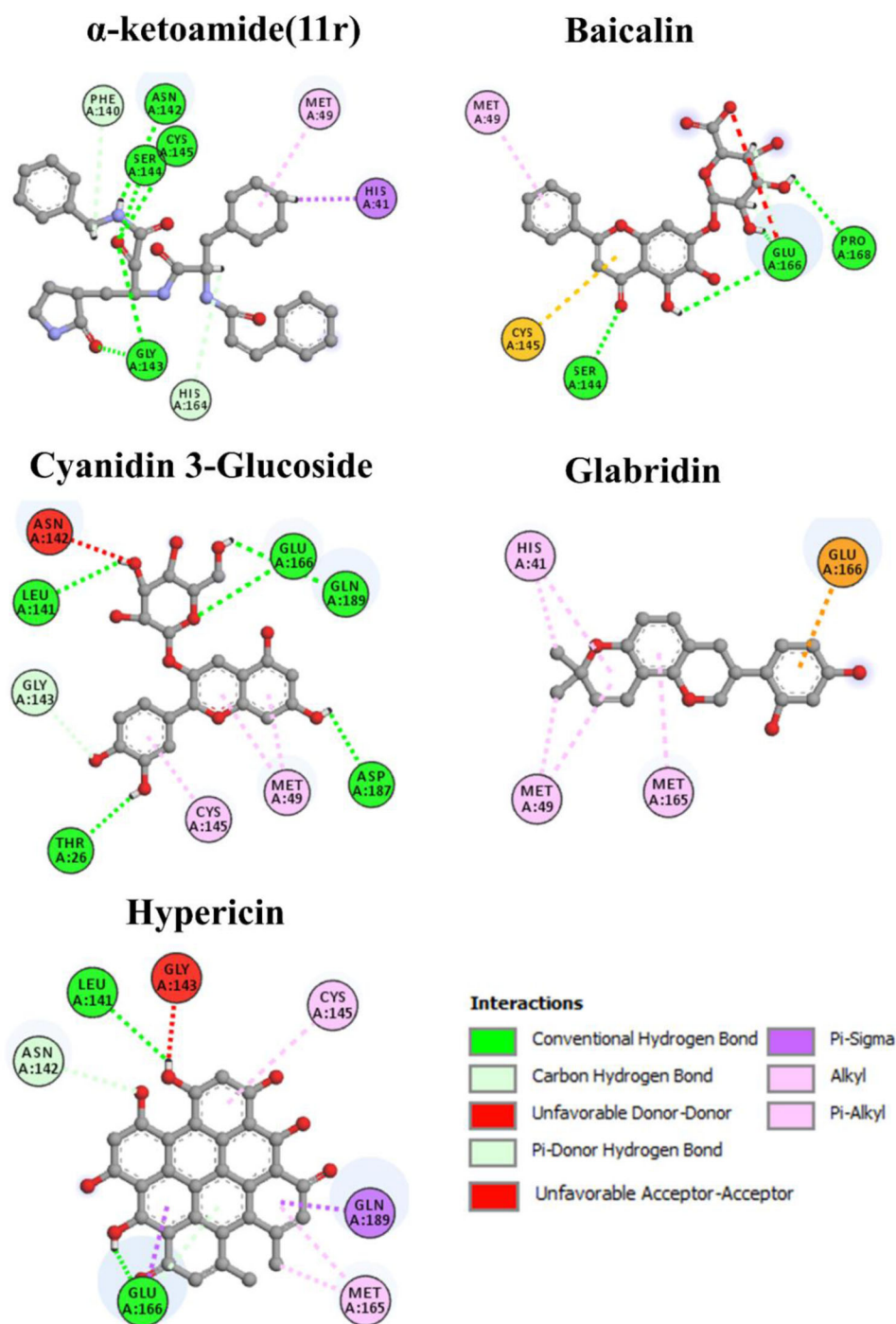
Figure 2. Nonbonding interactions of five selected phytochemicals with the main protease of SARS-CoV-2 (pose predicted by AutoDock Vina).

Since RMSF (root means square fluctuation) helps to understand the region of protein that is being fluctuated during the simulation, the flexibility of each residue is calculated to get better insight on what extent the binding of ligand affects the protein flexibility. It can be understood from Figure 3b that the binding of hypericin makes the protein most flexible in all areas in contrast to apo-protein and the other complexes. Hypericin is found to induce local flexibility at Met49 (S2 pocket), Asn51, Pro52, Tyr154, Asp248, Arg279, and Phe294. The apo-protein structure is found to have the lowest RMSF, which indicates that even in unliganded state, the protein is not very much flexible. Besides, the RMSF values of other complexes including α-ketoamide-11r, baicalin, cyanidin 3-glucoside, and glabridin are mostly similar in all areas. Overall, the residues such as Ile136, Lys137, Gly138, Ser139, Phe140 (S1 pocket), Leu141, Asn142, Asp153, Typ154, Arg222, Ser301, Val303, Thr304, Phe305, and Gln306 are found flexible for both of apo-protein and ligand-bound complexes.

Higher SASA value indicates the expansion of protein volume and a low fluctuation is expected over the simulation time. Binding of any small molecule might change SASA and sometimes could greatly affect the protein structure. The SASA values of alpha ketoamide-11r are found lowest in most of the frames compared to apo-protein. Thus, it can be
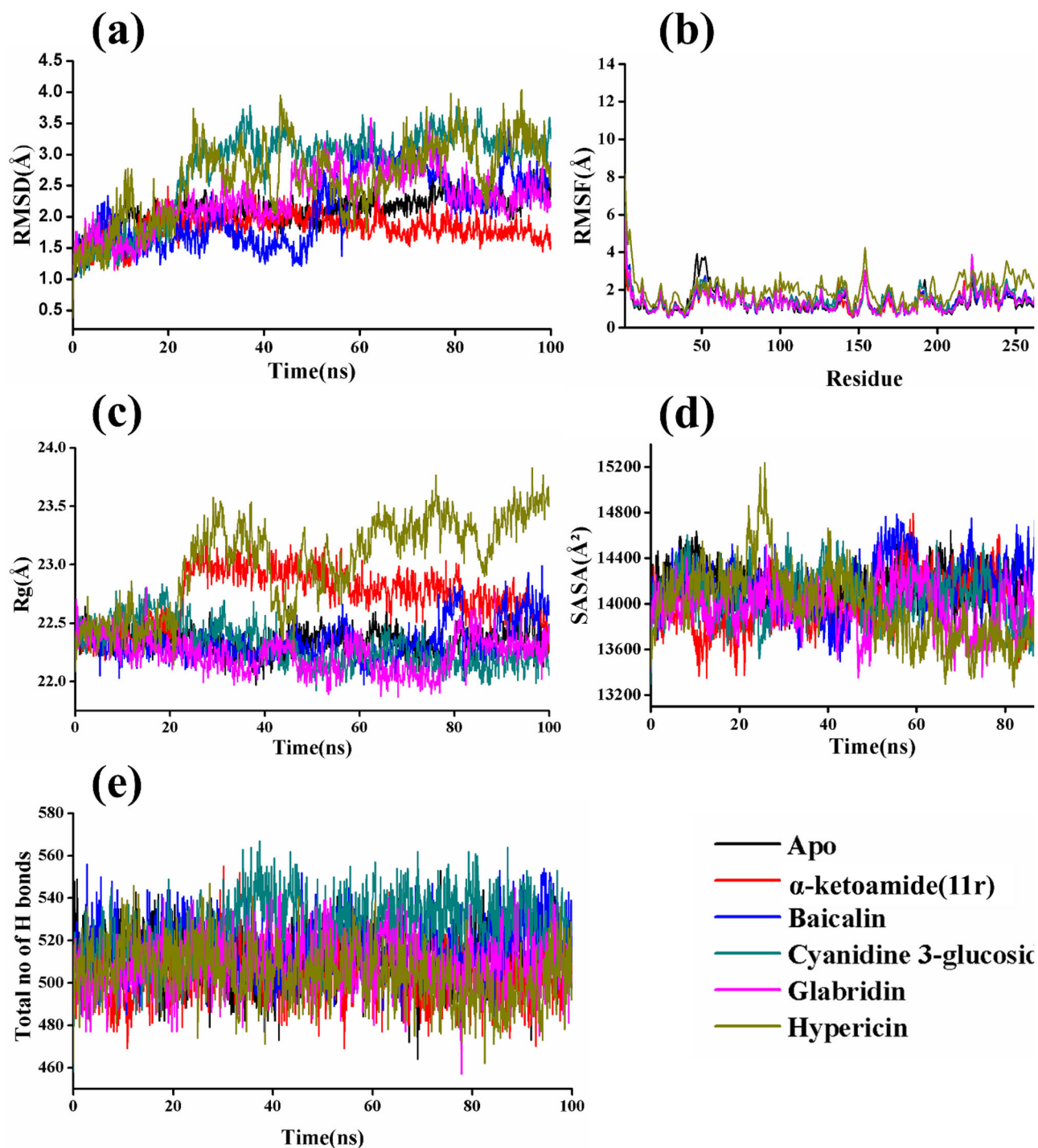
Figure 3. Analysis of RMSD, RMSF, Rg, SASA, and total number of hydrogen bond of apo-protein and selected five phytochemical complexes with protein at 100 ns MD simulations. (a) Root-mean-square deviation (RMSD) of the Cα atoms, (b) RMSF values of the alpha carbon over the entire simulation, where the ordinate is RMSF (Å) and the abscissa is residue, (c) Radius of gyration (Rg) over the entire simulation, where the ordinate is Rg (Å) and the abscissa is time (ns), (d) Solvent accessible surface area (SASA), where the ordinate is SASA (Å²) and the abscissa is time (ns), and (E) Total number of H-bond count throughout the simulation.

suggested that the binding of α-ketoamide-11r potentially could reduce protein expansion.

In addition, the baicalin complex shows normal fluctuations as compared to the apo-protein and α-ketoamide-11r complex. But it shows highest SASA from 80.8 to 100 ns revealing the conformational states with higher protein expansion. Interestingly, cyanidin 3-glucoside complex shows quite similar fluctuations from 60 to 100 ns as exhibited by α-ketoamide-11r complex. It signifies its SASA mediated

behavior as it is also observed by α-ketoamide-11r complex. Although the SASA of glabridin complex is observed in the median till 46 ns compared to other complexes as described earlier, the fluctuations are quite irregular from 66 to 78 ns. The average SASA value of all systems are 14160.7, 13985.1, 14209.3, 14039.9, 13958.4, and 13955.1 for apo-protein, α-ketoamide-11r, baicalin, cyanidin 3-glucoside, glabridin, and hypericin, respectively. The first and second lowest average SASA is found for hypericin and glabridin complexes.

Table 3. Pharmacokinetic parameters of the best phytochemicals.

| Drugs	Carcinogenicity	Rat Acute Toxicity (LD50, (mol/kg)	P-glycoprotein Inhibitor	Blood-brain barrier	Human intestinal absorption	Renal organic cation transporter	P-glycoprotein Substrate
Cyanidin 3-Glucoside	Non carcinogenic	2.6483	Non inhibitor	Positive	Negative	Non inhibitor	Substrate
Hypericin	Non carcinogenic	2.6870	Non inhibitor	Negative	Positive	Non inhibitor	Substrate
Baicalin	Non carcinogenic	2.7357	Non inhibitor	Negative	Positive	Non inhibitor	Substrate
α-Ketoamide-11r	Non carcinogenic	2.3318	Non inhibitor	Negative	Positive	Non inhibitor	Substrate
Glabridin	Non carcinogenic	2.9435	Non inhibitor	Positive	Positive	Non inhibitor	Substrate
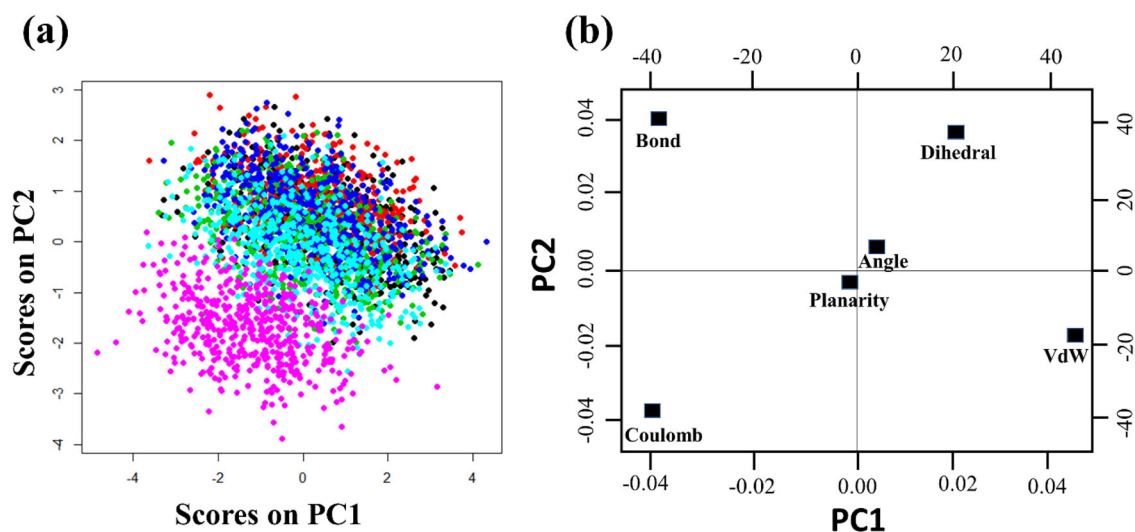
Figure 4. (a) The score plot presented six data clusters in different color, where each dot represented one time point. The clustering is attributable as: apo-protein (black), α-ketoamide-11r complex (red), baicalin (green), cyanidin 3-glucoside complex (blue), glabridin complex (cyan), hypericin complex (magenta), (b) Loading plot from principal components analysis of the energy and structural data.

The radius of gyration (Rg) represents the compactness of a structure. The lower degree of fluctuation with its consistency throughout the simulation indicates the greater compactness and rigidity of a system. The Rg of apo-protein is found almost stable in terms of consistency of fluctuations throughout the simulation. Besides, the Rg of α-ketoamide-11r is increased from 22 to 95.6 ns (average Rg ∼ 22.82). The greater change of Rg might be the result of protein folding, or unique conformational changes. The baicalin complex shows increased Rg from 78 to 100 ns (average Rg ∼ 22.50), and its last parts of the trend are similar to α-ketoamide-11r. In case of cyanidin 3-glucoside complex, the lowest Rg value is observed from 80 to 100 ns, which indicates greater rigidness in contrast to the other complexes. Besides, the glabridin complex seems to have lowest Rg corresponding to its highest rigidity from 0 to 40 ns and 63 ns to 77 ns. The Rg value of hypericin is much higher after 20 ns (average Rg ∼ 23.16) indicating its slackness of packing compared to all other complexes. On the basis of average Rg, the glabridin complex has the lowest Rg, which are calculated as 22.24, while for apo-protein, α-ketoamide-11r, baicalin, cyanidin 3-glucoside, and hypericin, the Rg values are determined as 22.35, 22.74, 22.34, 22.32, and 23.03, respectively. Therefore, the order of compactness and rigidness should be glabridin > cyanidin 3-glucoside > baicalin > apo-protein > α- ketoamide-11r > hypericin.

The number of intermolecular hydrogen bonds in the ligand-protein complex are determined, since it is known to contribute conformational stability. The average number of hydrogen bonds are 513, 504, 518, 526, 508, and 505 for apo-

protein, α-ketoamide-11r, baicalin, cyanidin 3-glucoside, glabridin, and hypericin complexes, respectively. The highest number of hydrogen bonds is observed for cyanidin 3-glucoside complex, whilst the lowest number of hydrogen bonds is observed in α-ketoamide-11r complex over the 100 ns simulation period. The baicalin complex possesses a greater number of hydrogen bonds compared to apo-protein, which shows almost similar trend to that of glabridin complex.

3.4. Pharmacokinetic properties

Total seven pharmacokinetic parameters including carcinogenicity, rat acute toxicity (LD50, mol/kg), p-glycoprotein inhibitor, blood-brain barrier, glycoprotein substrate, organic cation transporter, and human intestinal absorption are tested for the selected best phytochemicals. The results are summarized in Table 3. The results show that the phytochemicals are safer to use. Outcomes are explained in more detail in the discussion section.

3.5. Principal component analysis (PCA) of molecular dynamics

Principal component analysis (PCA) is used to analyze the structural and energy data obtained from MD simulation on protein-phytochemical complexes and apo-protein. Bond distances, bond angles, dihedral angles, planarity, Van der Waals and electrostatic energies were included as variables. The PCA score plot (Figure 4a) reveals different clusters
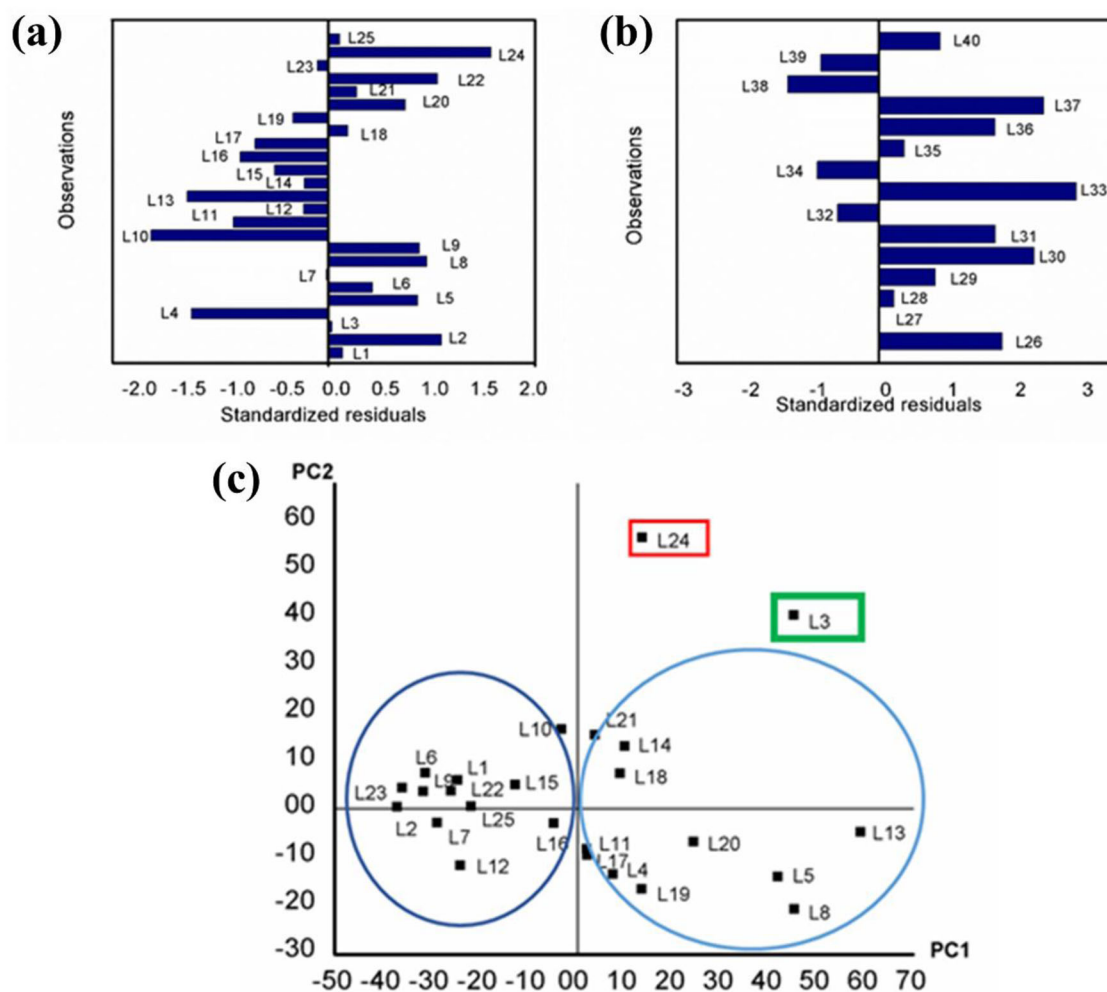
Figure 5. (a) Graphical representation observations vs. standardized residues by MLR (training set), (b) Graphical representation observations vs. standardized residues by MLR (test set), (c) Score plot of PCA analysis for QSAR of ligands.

formation. Among these clusters, α-ketoamide-11r-protein complex (red), cyanidin 3-glucoside-protein complex (blue), and the apo-protein (black) are overlapped. Loading plot of the PCA (Figure 4b) reveals that the dihedral angle energies show positive correlation with these three groups. Baicalin-protein complex (green) demonstrates similar patterns to α-ketoamide-11r and cyanidin 3-glucoside complexes in the score plot, although the energy distribution of baicalin is broader compared to the other two complexes. The fluctuating nature of baicalin-protein complex in the MD simulation could be the reason for its wider distribution in the score plot. However, it can be considered as a potential candidate. The hypericin-protein complex exhibits significant difference compared to the other drug-protein complexes by forming a distinct cluster (magenta). This is reasonable as hypericin shows the highest deviation during complex formation compared to the other candidates. The glabridin-protein complex is also formed a distinct cluster (cyan) in the score plot.

3.6. Quantitative structure-activity relationships (QSAR)

QSAR is a vastly used tool in bioinformatics, drug discovery for the pharmaceutical industry, clinical research, agrochemical, and petrochemical sectors for modeling and predictive pattern analysis (Alam & Khan, 2017; Fakayode et al., 2014; Funar-Timofei et al., 2017). Herein multiple linear regression (MLR) has been used for further analysis. For instance, TPSA (Å^2), molecular weight, XLogP3, H-bond donor count, and H-bond acceptor count of the ligands are the most significant variables on QSAR contributors to the MLR regression (Table S3). PCA is used for pattern recognition of potential ligands, which is represented in this study by PCA score plot (Figure 5c). The first principal component (PC1) shows 57% of the variability in ligand QSAR and 21% of the variation of ligand binding energy. Second principal component (PC2) exhibits 22% in QSAR variability of ligands and the variation of the ligands binding energy is 11%. An interesting grouping of the ligands is observed by the score plot. The ligands with a similar functional group are gathered together side by side on the score plot. For example, ligands (L6, L7, L12, L15, L16, L22, and L25) containing -OH, -COOH, and C=O functional group attached to benzene ring are clustered on the first and second quadrants of the score plot. Only L3 (α-ketoamide-11r) with containing -CONH- and C=O functional groups is placed on the upper-right corner of the score plot.

In contrast, ligands containing isoflavane backbone in (L4, L8, L17, L18, and L19) are clustered on the third and fourth quadrants of the score plot. The observed grouping of

ligands on the score plot is highly significant. In addition, this class of ligands has cyclic ether, -OH, -COOH groups, which may have a significant influence on drug reactivity, chemical behavior, therapeutic effect, and potency. L11 and L13 contain common anthraquinone derivative with fused ring and -OH functional group. This isoflavane and anthraquinone backbone pattern could be a key for further investigations to explore new drugs for COVID-19.

The overarching goal of any MLR model is to predict binding energy of the future drug candidates. The MLR model is developed and therefore is used to predict the binding energy of the test set for validation of drug candidates. The results demonstrate almost the same binding energies obtained from the binding energies from molecular docking. Binding energy (kcal/mol) $= -3.51 + 4.27\text{E-}02*\text{TPSA}$ (Å^2)-6.42E-03*MW(g/mol)-0.25*XLogP3-AA-0.48*HBD-0.64*HBA. Here the R^2 is 0.842 for the training set and 0.753 for the test set. Figure 5a,b show the prediction results versus original binding results of the training set and test set. In Figure 5a, L3 and L7 show lowest residuals. Possible reason could be the presence of the -CONH- and C=O groups in L3, and the carboxylic group and -C=C- present in L7. On the other hand, in the test set L27 (Figure 5b) reveals the lowest standardized residuals may be the presence of hetero N, S, and O atoms. L33 exhibits a high standardized residual may be due to the large asymmetric molecule and their conformational change. The lowest standardized residual indicates that the predicted binding energies have close agreement to the binding energies obtained by molecular docking. The predicted QSAR binding energies and the docking binding energies obtained from the ligands with similar SAR typically have similar chemical reactivity, pharmacological activity, behavior, and efficacy. But the results obtained by the QSAR should be used with caution as the binding energies are not experimental. The model can be further used for rapid screening of COVID-19 candidates for drug discovery in pharmaceuticals and pharmacology research.

4. Discussion

Computer-aided drug design (CADD) has become one of the essential approaches in modern drug discovery as it can significantly minimize the cost and labor involved in the drug discovery process. It accelerates the drug development by allowing the scientists to narrow down the biological and synthetic testing efforts. Moreover, molecular docking, molecular dynamics, QSAR, and ADMET tools have become some of the key parts in the CADD because of their reliable predictions (Talele et al., 2010).

Molecular docking predicts the prevailing binding modes between a ligand and a protein by proposing the hypothesis of how the ligands inhibit the protein and thus it ranks candidate ligands (Morris & Lim-Wilby, 2008). In this study, 40 antiviral phytochemical agents are docked against the main protease of SARS-CoV-2. Among them, five candidates are then selected according to their high binding affinity. It is observed that hypericin, cyanidin 3-glucoside, baicalin,

glabridin, and α-ketoamide-11r could be used as potential inhibitors for the main protease. Hypericin shows the highest binding affinity of −10.7 kcal/mol and forms a pi-alkyl interaction with the catalytic binding residue Cys145. Cyanidin 3-glucoside, baicalin, glabridin, and α-ketoamide-11r also exhibit high binding affinity of −8.4, 8.1, −8.1, and −7.8 kcal/mol, respectively. All the top scored phytochemicals demonstrate strong noncovalent interactions with the binding site residues. More specifically, the selected five phytochemicals form non-covalent interactions with both the two (Cys145 and His41) or at least one of the catalytic residues, and thereby can act as inhibitors of the main protease. QSAR study reveals that the topological polar surface area (TPSA, Å^2), molecular weight, XLogP3, H-bond donor count, and H-bond acceptor count of the ligands are the most significant variables. MLR regression model validated their role in the binding affinity and non-covalent interactions of the ligands with the main protease. It is also predicted from the principal component analysis (PCA) of QSAR that L3 (α-ketoamide-11r) shows the lowest residuals because of the presence of the -CONH- and C=O groups in its chemical structure. Furthermore, QSAR analysis of all the phytochemicals show almost similar binding affinity predicted by the binding affinity of the molecular docking in Autodock Vina.

Molecular dynamics simulation of α-ketoamide-11r shows the lowest RMSD value than the apo-protein and the other complexes, which indicates its greater stability. SASA of α-ketoamide-11r also confirms that it unfolds and stabilizes the main protease. Cyanidin 3-glucoside has lower Rg and SASA values, which indicates that it can make the protease more compact and rigid. Moreover, the lowest RMSF value of the apo-protein indicates its compactness even without forming a complex with a ligand. Except hypericin, all the selected phytochemicals demonstrate similar patterns in RMSF. Hypericin is the most flexible in all areas in contrast to apo-protein and other complexes. Baicalin also shows lower RMSD value and followed the similar trend. Principal component analysis reveals the structural similarity between α-ketoamide-11r-protein and cyanidin 3-glucoside-protein complexes, which is strongly supported by the RMSD, RMSF, Rg, and SASA analysis. PCA analysis unveils that baicalin could be a good candidate to stabilize the main protease as supported by the RMSD analysis.

Recent drug discovery depends on drugs which show good pharmacokinetic properties. To be a promising drug candidate, pharmacokinetic parameters must be optimized so that these can pass standard clinical trial criteria. The best five phytochemicals are found to be noncarcinogenic as evidence obtained from carcinogenicity and rat acute toxicity. Cyanidin 3-glucoside and glabridin might cross the blood-brain barrier. All selected phytochemicals might interact with p-glycoprotein which is a member of ABC transporter family and they will not inhibit organic cation transporter. These parameters provide information about the secretion of drugs through urine. Only cyanidin 3-glucoside could show negative result in terms of human intestinal absorption. These pieces of information may provide necessary data for designing promising and effective inhibitors of the main protease

of SARS-CoV-2 in future (Cheng et al., 2012; Motohashi & Inui, 2013; Mukhametov & Raevsky, 2017).

5. Conclusion

In summary, molecular docking, molecular dynamics, QSAR modeling, PCA, and ADMET tools are successfully employed to determine the best phytochemicals against the main protease of SARS-CoV-2. Among the studied 40 phytochemicals, hypericin, cyanidin 3-glucoside, baicalin, glabridin, and α-ketoamide-11r show the highest binding affinity and strong interactions with both or at least one of the catalytic residues (Cys145 and His41) of the main protease. These compounds show many non-covalent interactions, such as hydrogen bonding, hydrophobic, and electrostatic interactions. MD results show that in the physiological environment, baicalin, cyanidin 3-glucoside, and α-ketoamide-11r are the most stable ligands and they are making a greater number of interactions through hydrogen bonds with the main protease. Pharmacokinetic and ADMET analysis indicate their efficacy as drug molecules. PCA of QSAR results show that the ligands, e.g. L6, L7, L12, L15, L16, L22, and L25 containing -OH, -COOH, and C=O functional groups attached to their benzene ring are grouped together in the first and second quadrants of the score plot. Only α-ketoamide-11r (L3) with -CONH- and C=O functional groups take position on the upper right corner of the score plot. The ligands, e.g. L4, L8, L17, L18, L19, L11, and L13 containing isoflavane backbone, anthraquinone with cyclic ether, fused ring, -OH, and -COOH are placed in the third and fourth quadrants of the score plot. Therefore, PCA analysis successfully divides the studied phytochemicals into two groups. The MLR model calculates the value of R^2, which is 0.842 for the training set and 0.753 for the test set. It means that our proposed model can predict the binding energies compared to the values obtained from molecular docking. It can be concluded that most of the selected phytochemicals show promise and can be used to design effective antiviral drugs against the SARS-CoV-2.

Acknowledgements

Disclosure statement

The authors declare no conflict of interest.

ORCID

Mohammad A. Halim 🔟 http://orcid.org/0000-0002-1698-7044

References

Aanouz, I., Belhassan, A., El Khatabi, K., Lakhlifi, T., El Idrissi, M., & Bouachrine, M. (2020). Moroccan medicinal plants as inhibitors of COVID-19: Computational investigations. *Journal of Biomolecular Structure & Dynamics*, 1–12. https://doi.org/10.1080/07391102.2020.1758790

Adinsoft, S. (2010). *XLSTAT-software, version 10*. Addinsoft.

Alam, S., & Khan, F. (2017). 3D-QSAR studies on maslinic acid analogs for anticancer activity against breast cancer cell line MCF-7. *Scientific Reports*, 7(1). https://doi.org/10.1038/s41598-017-06131-0

Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A., & Freire, E. (2004). Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry*, 43(17), 4906–4912. https://doi.org/10.1021/bi0361766

Boopathi, S., Poma, A. B., & Kolandaivel, P. (2020). Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *Journal of Biomolecular Structure & Dynamics*, 1–14. https://doi.org/10.1080/07391102.2020.1758788

Chen, F., Chan, K. H., Jiang, Y., Kao, R. Y. T., Lu, H. T., Fan, K. W., Cheng, V. C. C., Tsui, W. H. W., Hung, I. F. N., & Lee, T. S. W. (2004). In vitro susceptibility of 10 clinical isolates of SARS coronavirus to selected antiviral compounds. *Journal of Clinical Virology*, 31(1), 69–75. https://doi.org/10.1016/j.jcv.2004.03.003

Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., Lee, P. W., & Tang, Y. (2012). admetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. *Journal of Chemical Information and Modeling*, 52(11), 3099–3105. https://doi.org/10.1021/ci300367a

Colson, P., Rolain, J.-M., Lagier, J.-C., Brouqui, P., & Raoult, D. (2020). Chloroquine and hydroxychloroquine as available weapons to fight COVID-19. *International Journal of Antimicrobial Agents*, 55(4), 105932. https://doi.org/10.1016/j.ijantimicag.2020.105932

DeLano, W. L. (2002). The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific LLC.

Dickson, C. J., Madej, B. D., Skjevik, Å. A., Betz, R. M., Teigen, K., Gould, I. R., & Walker, R. C. (2014). Lipid14: The Amber lipid force field. *Journal of Chemical Theory and Computation*, 10(2), 865–879. https://doi.org/10.1021/ct4010307

Elfiky, A. A., & Azzam, E. B. (2020). Novel guanosine derivatives against MERS CoV polymerase: An in silico perspective. *Journal of Biomolecular Structure & Dynamics*, 1–12. https://doi.org/10.1080/07391102.2020.1758789

Elmezayen, A. D., Al-Obaidi, A., Şahin, A. T., & Yelekçi, K. (2020). Drug repurposing for coronavirus (COVID-19): In silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *Journal of Biomolecular Structure & Dynamics*, 1–12. https://doi.org/10.1080/07391102.2020.1758791

Enayatkhani, M., Hasaniazad, M., Faezi, S., Guklani, H., Davoodian, P., Ahmadi, N., Einakian, M. A., Karmostaji, A., & Ahmadi, K. (2020). Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: An in silico study. *Journal of Biomolecular Structure & Dynamics*, 1–19. https://doi.org/10.1080/07391102.2020.1756411

Fakayode, S. O., Busch, K. W., & Busch, M. A. (2009). Chemometric approach to chiral recognition using molecular spectroscopy. *Chemometrics, Encyclopedia of Analytical Chemistry* (p. 1-32). New Jersey, USA: John Wiley & Sons. https://doi.org/10.1002/9780470027318.a9082

Fakayode, S. O., Mitchell, B. S., & Pollard, D. A. (2014). Determination of boiling point of petrochemicals by gas chromatography-mass spectrometry and multivariate regression analysis of structural activity relationship. *Talanta*, 126, 151–156. https://doi.org/10.1016/j.talanta.2014.03.037

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., & Petersson, G. A. (2009). *Gaussian 09, A.02*. Gaussian, Inc. https://doi.org/111

Funar-Timofei, S., Borota, A., & Crisan, L. (2017). Combined molecular docking and QSAR study of fused heterocyclic herbicide inhibitors of D1 protein in photosystem II of plants. *Molecular Diversity*, 21(2), 437–454. https://doi.org/10.1007/s11030-017-9735-x

Gautret, P., Lagier, J.-C., Parola, P., Hoang, V. T., Meddeb, L., Mailhe, M., Doudier, B., Courjon, J., Giordanengo, V., Vieira, V. E., Dupont, H. T., Honoré, S., Colson, P., Chabrière, E., La Scola, B., Rolain, J.-M., Brouqui, P., & Raoult, D. (2020). Hydroxychloroquine and azithromycin as a

treatment of COVID-19: Results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents*, 105949. https://doi.org/10.1016/j.ijantimicag.2020.105949

Gupta, M. K., Vemula, S., Donde, R., Gouda, G., Behera, L., & Vadde, R. (2020). In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel. *Journal of Biomolecular Structure and Dynamics*, 1–17. https://doi.org/10.1080/07391102.2020.1751300

Hasan, A., Paray, B. A., Hussain, A., Qadir, F. A., Attar, F., Aziz, F. M., Sharifi, M., Derakhshankhah, H., Rasti, B., Mehrabi, M., Shahpasand K., Saboury A. A., & Falahati, M. (2020). A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. *Journal of Biomolecular Structure and Dynamics*, 1–13. https://doi.org/10.1080/07391102.2020.1754293

Islam, M. J., Khan, A. M., Parves, M. R., Hossain, M. N., & Halim, M. A. (2019). Prediction of deleterious non-synonymous SNPs of human STK11 gene by combining algorithms, molecular docking, and molecular dynamics simulation. *Scientific Reports*, 9(1), 16426. https://doi.org/10.1038/s41598-019-52308-0

Jassim, S. A. A., & Naji, M. A. (2003). Novel antiviral agents: A medicinal plant perspective. *Journal of Applied Microbiology*, 95(3), 412–427. https://doi.org/10.1046/j.1365-2672.2003.02026.x

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267(3), 727–748. https://doi.org/10.1006/jmbi.1996.0897

Khan, R. J., Jha, R. K., Amera, G., Jain, M., Singh, E., Pathak, A., Singh, R. P., Muthukumaran, J., & Singh, A. K. (2020). Targeting SARS-CoV-2: A systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2′-O-ribosemethyltransferase. *Journal of Biomolecular Structure & Dynamics*, 1–14. https://doi.org/10.1080/07391102.2020.1753577

Khan, S. A., Zia, K., Ashraf, S., Uddin, R., & Ul-Haq, Z. (2020). Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. *Journal of Biomolecular Structure and Dynamics*, 1–13. https://doi.org/10.1080/07391102.2020.1751298

Krieger, E., Dunbrack, R., Hooft, R., & Krieger, B. (2012). Assignment of protonation states in proteins and ligands: Combining pKa prediction with hydrogen bonding network optimization. *Methods in Molecular Biology*, 819, 405–421. https://doi.org/10.1007/978-1-61779-465-0

Krieger, E., Nielsen, J. E., Spronk, C. A. E. M., & Vriend, G. (2006). Fast empirical pKa prediction by Ewald summation. *Journal of Molecular Graphics and Modelling*, 25(4), 481–486. https://doi.org/10.1016/j.jmgm.2006.02.009

Krieger, E., & Vriend, G. (2015). New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry*, 36(13), 996–1007. https://doi.org/10.1002/jcc.23899

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 368(6490), 489-493. https://doi.org/10.1126/science.abb3221

Liu, C.-H., Chen, X.-Y., Qin, P.-W., Qi, Z.-Q., Ji, M.-S., Liu, X.-Y., Babu, P. V., Li, X.-H., & Cui, Z.-N. (2017). Synthesis, fungicidal activity, and structure activity relationship of β-acylaminocycloalkylsulfonamides against Botrytis cinerea. *Scientific Reports*, 7(1), 42096. https://doi.org/10.1038/srep42096

Liu, H., Ye, F., Sun, Q., Liang, H., Li, C., Lu, R., & Tan, W. (2020). Scutellaria baicalensis extract and baicalein inhibit replication of SARS-CoV-2 and its 3C-like protease in vitro. *BioRxiv Preprint*, 1–18. https://doi.org/10.1101/2020.04.10.035824.

Liu, X., & Wang, X. J. (2020). Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *Journal of Genetics and Genomics*, 47(2), 119–121. https://doi.org/10.1016/j.jgg.2020.02.001

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet*, 395(10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

Mark, H., & Workman, J. (Eds.). (2007). *Chemometrics in spectroscopy* (1st ed.). Academic Press. https://doi.org/10.1016/B978-0-12-374024-3.X5000-4

Martens, H., & Naes, T. (1992). *Multivariate calibration*. John Wiley & Sons, Inc.

Mirza, M. U., & Froeyen, M. (2020). Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease. Nsp12 RNA-dependent RNA polymerase and Nsp13 helicase. Preprints. https://doi.org/10.20944/preprints202003.0085.v1

Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Methods in Molecular Biology (Clifton, N.J.)*, 443, 365–382. https://doi.org/10.1007/978-1-59745-177-2_19

Motohashi, H., & Inui, K. I. (2013). Organic cation transporter OCTs (SLC22) and MATEs (SLC47) in the human kidney. *The AAPS Journal*, 15(2), 581–588. https://doi.org/10.1208/s12248-013-9465-7

Mukhametov, A., & Raevsky, O. A. (2017). On the mechanism of substrate/non-substrate recognition by P-glycoprotein. *Journal of Molecular Graphics and Modelling*, 71, 227–232. https://doi.org/10.1016/j.jmgm.2016.12.008

Muralidharan, N., Sakthivel, R., Velmurugan, D., & Gromiha, M. M. (2020). Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 Protease against COVID-19. *Journal of Biomolecular Structure & Dynamics*, 1–7. https://doi.org/10.1080/07391102.2020.1752802

Naithani, R., Huma, L., Holland, L., Shukla, D., McCormick, D., Mehta, R., & Moriarty, R. (2008). Antiviral activity of phytochemicals: A comprehensive review. *Mini-Reviews in Medicinal Chemistry*, 8(11), 1106–1133. https://doi.org/10.2174/138955708785909943

Nayak, M. K., Agrawal, A. S., Bose, S., Naskar, S., Bhowmick, R., Chakrabarti, S., Sarkar, S., & Chawla-Sarkar, M. (2014). Antiviral activity of baicalin against influenza virus H1N1-pdm09 is due to modulation of NS1-mediated cellular innate immune responses. *Journal of Antimicrobial Chemotherapy*, 69(5), 1298–1310. https://doi.org/10.1093/jac/dkt534

Oo, A., Rausalu, K., Merits, A., Higgs, S., Vanlandingham, D., Bakar, S. A., & Zandi, K. (2018). Deciphering the potential of baicalin as an antiviral agent for Chikungunya virus infection. *Antiviral Research*, 150, 101–111. https://doi.org/10.1016/j.antiviral.2017.12.012

Pant, S., Singh, M., Ravichandiran, V., Murty, U. S. N., & Srivastava, H. K. (2020). Peptide-like and small-molecule inhibitors against Covid-19. *Journal of Biomolecular Structure & Dynamics*, 1–15. https://doi.org/10.1080/07391102.2020.1757510

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Sheahan, T. P., Sims, A. C., Leist, S. R., Schäfer, A., Won, J., Brown, A. J., Montgomery, S. A., Hogg, A., Babusis, D., Clarke, M. O., Spahn, J. E., Bauer, L., Sellers, S., Porter, D., Feng, J. Y., Cihlar, T., Jordan, R., Denison, M. R., & Baric, R. S. (2020). Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. *Nature Communications*, 11(1), 222. https://doi.org/10.1038/s41467-019-13940-6

Stewart, J. J. P. (2007). Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling*, 13(12), 1173–1213. https://doi.org/10.1007/s00894-007-0233-4

Talele, T., Khedkar, S., & Rigby, A. (2010). Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Current Topics in Medicinal Chemistry*, 10(1), 127–141. https://doi.org/10.2174/156802610790232251

Team, I. C.-19 Health Service Utilization Forecasting, & Murray, C. J. (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *MedRxiv*. https://doi.org/10.1101/2020.03.27.20043752

Tong, T. R. (2009). Therapies for coronaviruses. Part 2: Inhibitors of intracellular life cycle. *Expert Opinion on Therapeutic Patents*, 19(4), 415–431. https://doi.org/10.1517/13543770802600698

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461. https://doi.org/10.1002/jcc.21334

Wang, F., Chen, C., Tan, W., Yang, K., & Yang, H. (2016). Structure of main protease from human coronavirus NL63: Insights for wide spectrum anti-coronavirus drug design. *Scientific Reports*, 6(1), 1–12. https://doi.org/10.1038/srep22677

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W., & Xiao, G. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Research*, 30(3), 269–271. https://doi.org/10.1038/s41422-020-0282-0

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis* (1st ed.). Springer-Verlag. https://doi.org/10.1007/978-0-387-98141-3

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1-3), 37–39. https://doi.org/10.1016/0169-7439(87)80084-9

Worldometers.info. (2020). COVID-19 coronavirus pandemic. https://www.worldometers.info/coronavirus/ (Last accessed April 22, 2020).

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y.,

Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*, *368*, 409–412. https://doi.org/10.1126/science.abb3405