



# Identification of some dietary flavonoids as potential inhibitors of TMPRSS2 through protein–ligand interaction studies and binding free energy calculations

Jibin K. Varughese<sup>1</sup> · Kavitha J<sup>1</sup> · Sindhu K. S<sup>2,1</sup> · Dhiya Francis<sup>1</sup> · Joseph Libin K. L<sup>1</sup> · Abi T. G<sup>1,2</sup>

Received: 29 December 2021 / Accepted: 27 April 2022 / Published online: 25 May 2022  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

The continuing threat of COVID-19 and deaths need an urgent cost-effective pharmacological approach. Here, we examine the inhibitory activity of a group of dietary bioactive flavonoids against the human protease TMPRSS2, which plays a major role in SARS CoV-2 viral entry. After the molecular docking studies of a large number of flavonoids, four compounds with high binding scores were selected and studied in detail. The binding affinities of these four ligands, Amentoflavone, Narirutin, Eriocitrin, and Naringin, at the active site of the TMPRSS2 target, were investigated using MD simulations followed by MM-PBSA binding energy calculations. From the studies, a number of significant hydrophobic and hydrogen bonding interactions between the ligands and binding site amino residues of TMPRSS2 are identified which showcase their excellent inhibitory activity against TMPRSS2. Among these ligands, Amentoflavone and Narirutin showed MM-PBSA binding energy values of  $-155.57$  and  $-139.71$  kJ/mol, respectively. Our previous studies of the inhibitory activity of these compounds against the main protease of SARS-CoV2 and the present study on TMPRSS2 strongly highlighted that Amentoflavone and Naringin can exhibit promising multi-target activity against SARS-CoV-2. Moreover, due to their wide availability, no side effects, and low cost, these compounds could be recommended as dietary supplements for COVID patients or for the development of SARS-CoV-2 treatments.

**Keywords** TMPRSS2 · COVID-19 · Flavonoids · Molecular docking · Molecular dynamics

## Introduction

The coronavirus disease 2019 (COVID-19), caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), has emerged as a global pandemic and is continuing as a threat to human beings [1]. SARS-CoV-2 is a beta coronavirus similar to SARS-CoV and MERS viruses. Compared to the other two pathogenic human coronaviruses, SARS-CoV-2 has extraordinary transmissibility which underscores the urgent need for pharmacological

approaches to combat the virus [2]. Cost-effective treatments should be developed quickly to support the available vaccines for avoiding chaos all over the world.

The SARS-CoV-2 virus is enveloped and it contains four major structural proteins—the spike protein (S), nucleocapsid protein (N), membrane protein (M), and the enveloped protein (E). Once the virus enters the host cell, its RNA is released to the cell cytoplasm and there it undergoes replication and gets converted into an effector protein by viral proteases. Our previous studies were focused on the main protease (MPro) of SARS-CoV-2 which plays a key role in viral replication and transcription. We have used computational tools to study the flavonoids like Amentoflavone, Naringin, Eriocitrin, and Narirutin, and they were found to possess strong inhibitory activity against the MPro receptor [3]. After this, we have been thoroughly investigating whether these flavonoids can be used to inhibit those protein targets which assist the viral cell entry. Here we present a selection of relevant compounds that inhibit them as it would prevent the viral load [4].

✉ Sindhu K. S  
sindhuxavier05@gmail.com

✉ Abi T. G  
abitg@shcollege.ac.in

<sup>1</sup> Department of Chemistry, Sacred Heart College (Autonomous) Thevara, Kochi, Kerala 682013, India

<sup>2</sup> Department of Chemistry, Morning Star Home Science College, Angamaly, Kerala 683573, India

Recently, it is reported that the cell entry of SARS-CoV-2 occurs as a result of binding of its spike glycoprotein S to the human host cell ACE2 receptor and it is a human protease TMPRSS2 which primes and activates S protein for its binding and fusion [2]. The spike protein has 2 subunits—S1 and S2. The S1 subunit with a receptor-binding domain binds to ACE2 receptor, whereas S2, the membrane-anchored subunit, helps in the fusion of virus and host membrane. From the studies, it is clear that TMPRSS2, the endothelial cell surface protein, helps in SARS-CoV-2 viral entry and infection. The S1 subunit binds to the host cell receptor ACE2, which is followed by the priming of spike protein by the host TMPRSS2. TMPRSS2 breaks the viral S-protein at the right upstream of fusion peptide and results in membrane fusion. This points out the importance of inhibiting TMPRSS2 to avoid the host cell entry of SARS-CoV-2 virus [1]. Another reason which makes TMPRSS2 a promising target is the fact that it is a human protease, thereby this eliminates the problem of causing drug resistance like viral protein targets [2].

Many attempts were made to develop a potential drug that can target TMPRSS2. The repurposing efficacy of some drugs were studied, and from the studies, Camostat [2], a drug for treating chronic pancreatitis, was verified as a potential inhibitor of TMPRSS2. Later, after the screening of several FDA-approved drugs, Nafamostat, the drug for pancreatitis and disseminated intravascular coagulation, was found to block the SARS-CoV-2 fusion at a concentration less than that required for Camostat [2]. Apart from these proven antiviral drugs, a natural compound Lutonarin [5] and the polyphenols like Mangiferin, Glucogallin, and Phlorizin [6] were also identified as safe TMPRSS2 inhibitors.

In this study, we focused on investigating the efficacy of bioactive compounds like flavonoids in inhibiting the serine protease “TMPRSS2.” Flavonoids are powerful antioxidants and they are found to be effective against diseases such as cancer, obesity, hypertension, and other diseases [7]. The literature studies also underline the effectiveness of flavonoids as potential antiviral compounds. Even though the crystal structure of TMPRSS2 was available (PDB ID 7MEQ) [8], the structure is incomplete with missing coordinates of co-crystallized ligand (Nafamostat) and missing amino residues from residue numbers 164–166, 203–207, 217–220, and 250–255. The use of this incomplete crystal structure in docking and simulation studies may lead to wrong results. So we choose this crystal structure as a template and generated its complete structure using homology modeling. The virtual screening of a library of bioactive flavonoids was conducted based on molecular docking towards the binding site of homology modeled TMPRSS2 structure. Among the screened flavonoids, Amentoflavone, Narirutin, Eriocitrin,

and Naringin showed strong inhibitory activity against the receptor.

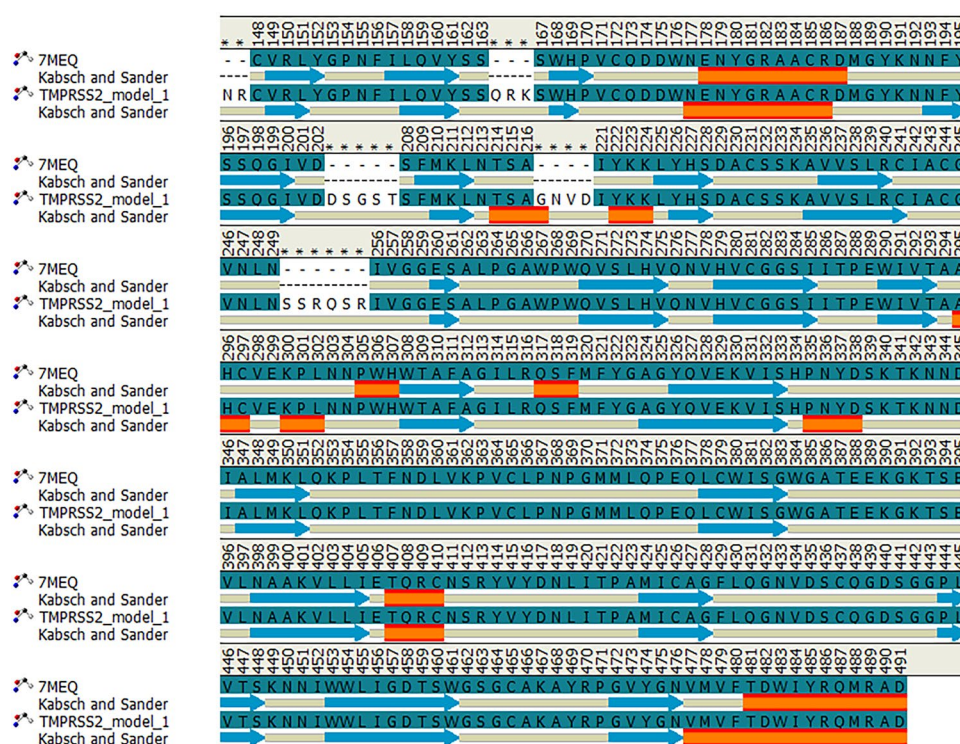
The bioflavonoid Amentoflavone can be isolated from the Ginkgo biloba plant, whereas the flavonoids Naringin is mostly found in citrus fruits. These two dietary flavonoids, which possess strong antiviral activity, are available as nutritional supplements from different commercial sources [3]. Eriocitrin is also a citrus flavonoid found in lemon peel which has antioxidant and enzyme inhibitory activity [9]. Narirutin is a flavanone originally isolated from *C. paradisi* that has anti-inflammatory and antidepressant-like activities [10]. These natural flavonoids have several benefits over other therapeutic agents (a) available as dietary supplements, (b) rarely have side effects (c), cost-effective, (d) can be easily absorbed in the intestine, and (e) can be used effectively for home care patients with mild symptoms. These four selected compounds were then subjected to molecular dynamics simulations followed by MM-PBSA calculations. Their nonbonding interactions with the protein residues were also investigated. The MM-PBSA binding free energy calculations and nonbonding interaction studies along with docking results confirmed that Amentoflavone and Narirutin have maximum antiviral activity against TMPRSS2 protein.

## Materials and methods

### Preparation of target protein using homology modeling

The target TMPRSS2 structure [11] was prepared using the homology modeling method in which the sequence of TMPRSS2 was obtained from the UniProt database (Id O15393). As per the UniProt data, the amino acid sequence which belongs to the catalytic region of TMPRSS2 (256 to 491) was selected as our target sequence. The crystal structure of TMPRSS2 (PDB ID 7MEQ) [8] with some missing residues (Fig. 1) and 98.5% sequence identity is used as the standard template. Using the homology modeling tools of the SWISS-MODEL web server [12, 13] and Modeller 9.2 [14] interface in UCSF Chimera, we generated the 3D structure of TMPRSS2 target—modeled structures 1 and 2, respectively. We checked the reliability and consistency of the modeled structure obtained from the SWISS-MODEL web server using the validation parameters such as QMEAN [15] (normal score = 0.89 and z score = -0.55) and GMQE value = 0.93. The discrete optimized protein energy (DOPE) [16] score can be employed for selecting the best structure from a group of models built by the Modeller. The modeled structure with the negative value of DOPE score (-1.12) is an indication of stable protein structure with minimum energy value. The RMSD of the two aligned model structures is only 0.5 Å. We further validated the stereochemical

**Fig. 1** Amino acid sequence and Kabsch and Sander secondary structure representation of 7MEQ template and modelled structure of Tmprss2



quality of both modeled structures from their Ramachandran plots (RC-plot) using the Procheck web server [17]. The percentage of residues belonging to the most favored regions of the Ramachandran plot was found to be 86.9% and 86.8%, respectively, for modeled structures of SWISS-MODEL and Modeller 9.2. The modeled structure obtained from SWISS-MODEL was utilized for further studies, and their structural and energetic validations were performed using the molecular dynamic simulation of 50 ns duration.

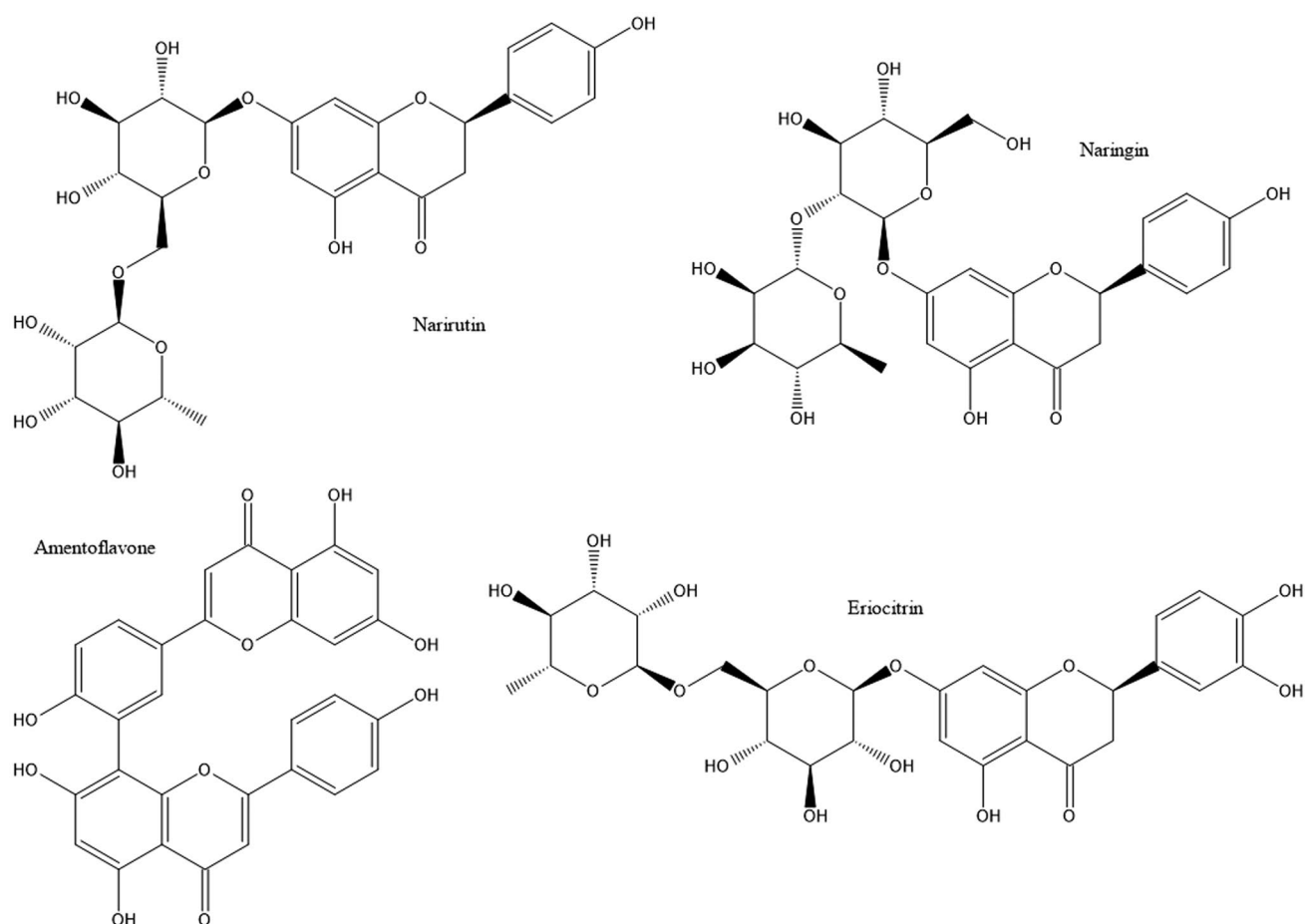
### Preparation of ligands and molecular docking calculations

The structure of the Tmprss2 target obtained from homology modeling is used in molecular docking studies. After the screening of a group of flavonoids using docking studies, Amentoflavone, Narirutin, Eriocitrin, and Naringin with high binding free energy (docking score) with the target were chosen for extensive study and analysis. The representative 2D structures of these flavonoids are illustrated in Fig. 2. The structure data files (SDF) of these flavonoids were downloaded from the Pubchem database and a DFT-based structure optimization was carried out using the B3LYP/SVP [18] method and Turbomole package [19]. The molecular docking calculations were performed using Autodock Vina software [20]. The Cartesian coordinates of the center of the catalytic triad (HIS 296, ASP 345, and SER 441) in modeled structure 1 are fixed as the center of the grid box in Vina docking calculations. There were 6400

grid points per map where 40 grid points each were fixed in x, y, and z dimensions, respectively. We fixed the grid point spacing as 1 Å and exhaustiveness as 8 in our current calculation. The best binding poses with higher docking scores were identified and then subjected to further MD studies. For the validation of docking studies, we truncate the residues of modeled structures 1 and 2 from 256–491 as the crystal structure (7MEQ) [8] consists of continuous residues from 256–491 and performed the docking of all ligands to these structures. Furthermore, we calculated the RMSD of the best docking pose of each ligand with the crystal structure and two model structures of Tmprss2.

### Molecular dynamics simulations

The molecular dynamics simulations were executed using the conformations of the protein–ligand complex generated from molecular docking calculations. GROMACS 2018.1 molecular dynamics package [21] was employed to conduct the simulations. We have used OPLS all-atom force field parameters [22] for the protein–ligand system in the simulation procedure. The LigParGen server [23] provides an interface for generating OPLS force field parameters for organic ligands with partial atomic charges in the most commonly used molecular dynamics format. The structure (PDB format) of the best docking pose of all ligands gets protonated at physiological pH (7.4) and is submitted to LigParGen server to obtain the ligand coordinate and topology files for protein–ligand MD simulation. We used protonated



**Fig. 2** Representation of selected flavonoids

protein–ligand system at constant pH (7.4) for the MD simulation. The neutralization of the protein–ligand complex structure was done by adding sodium ions to it followed by its solvation using SPC/E water [24]. To optimize the solvated system, it was subjected to an energy minimization process using the steepest-descent minimization algorithm [25]. To avoid the collapse of the simulating system during the production run, we carried out a 1 ns simulation of NVT and NPT equilibration MD run. We applied a position restraining force on the heavy atoms of two coupled groups (protein–ligand and solvent-ions) during the process. This enables the proper equilibration of solvent molecules around the protein–ligand system, and a steady state of energy, temperature, and pressure is achieved. The system arrived to correct temperature and pressure with a proper orientation of protein–ligand system to simulate after NVT and NPT equilibration, respectively.

The equilibrated system was then subjected to a 100-ns final production run with 2 fs time step. The temperature and pressure of the system were fixed as 300 K and

1 bar, respectively. They were maintained throughout using Berendsen thermostat and Berendsen barostat [26]. The short-range forces of the simulating system were computed using the Verlet neighbor list and long-range electrostatic forces using the particle-mesh Ewald summation method [27]. We also performed a replicate MD simulation of 100 ns duration with temperature condition as 310 K (physiological temperature). The protein–ligand complex structure (PDB) obtained from the final frame of the simulation at 300 K is utilized as the initial structure for the MD replicate simulation at 310 K. After every 10 ps, the coordinates from the trajectory were written so that a sufficient number of frames were obtained for analysis. The stabilization of protein–ligand structure was monitored through various trajectory analysis tools such as RMSD, RMSF, and radius of gyration. The hydrogen bonding and non-bonding analysis information throughout the MD trajectory are extracted using the GROMACS tools and Pycontact [28] tool. We found the consistency of results obtained in post MD trajectory analysis of both

simulations at 300 K and 310 K, respectively. Herein, we report the results of the replicate simulation at 310 K.

### MM-PBSA binding free energy calculations

MM-PBSA method, developed by Kollman and coworkers, can be used to evaluate the binding affinity of a ligand to a target protein in terms of binding free energy [29]. It is proved to be useful in drug designing, studying the stabilities of macromolecular conformations, and identifying the hotspots [30]. It is also helpful in investigating the dominant binding interactions between ligand and protein [31]. This method proves to be effective in the drug discovery process, particularly for finding inhibitors against target proteins [32]. We used MM-PBSA implemented `g_mmpbsa` [33] code handling GROMACS MD trajectories to perform binding energy calculations. The energy values corresponding to protein (P), ligand (L), and protein–ligand complex (C) conformations were obtained from the MD trajectory using a single trajectory approach. The binding free energy of the protein–ligand complex in the solution phase was calculated using Eq. (1).

$$\Delta G_{bind} = G_c - (G_p + G_L) \quad (1)$$

The free energy of any individual component is given by:

$$G = E_{MM} + G_{solv} - TS$$

$$E_{MM} = E_{bonded} + E_{elec} + E_{vdW} \quad (2)$$

The ( $E_{MM}$ ) term in Eq. (2) represents the energy of molecular mechanics (MM) potential with both bonding and non-bonding (Van der Waals + electrostatic) terms. The value of  $E_{bonded}$  energy can be taken as zero under the assumption that the bound and unbound forms of protein and ligand conformations in the single trajectory method are similar. The TS term is the conformational entropy term associated with complex and isolated protein is calculated in the vacuum environment. Instead of considering absolute binding free energy, we focused on the contribution of individual residues of protein and ligands to the individual components of  $E_{MM}$  and  $G_{solv}$  terms given in Eqs. (2), (3), and (4). As the change in entropy term does not affect the relative binding energy of ligands, it was neglected [34].

$$G_{solv} = G_{PB} + G_{SA} \quad (3)$$

$$G_{SA} = G_{SASA} + b \quad (4)$$

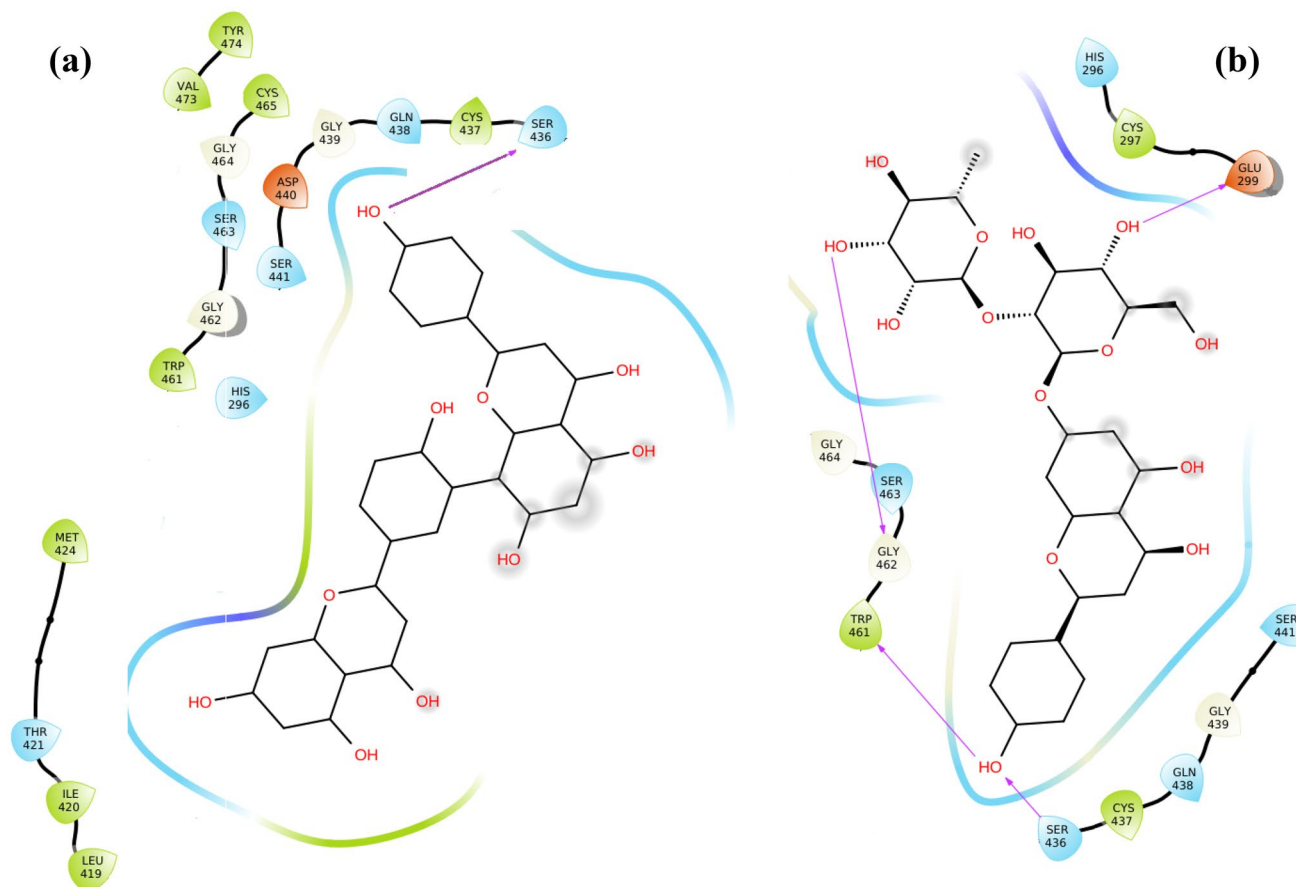
The free energy of solvation ( $G_{solv}$ ) in Eq. (3) is evaluated as the sum of the non-linearized version of the Poisson-Boltzmann Equation (GPB) that gives the energy of polar interactions. The nonpolar energy ( $G_{SA}$ ) term is calculated

using the solvent-accessible surface area (SASA). The term in Eq. (4) is a coefficient related to the surface tension of the solvent, whereas  $b$  stands for the fitting parameter. The dielectric constants used for solvent, solute, and vacuum in the `g_mmpbsa` calculations have the values 80, 2, and 1, respectively, and the solvent probe radius is 1.4 Å. We conducted the MM-PBSA calculations using the frames obtained from the 80–100 ns interval of equilibrated MD trajectory. Furthermore, we conducted 5 more MD replicate simulations of 20 ns duration using the protein–ligand conformations obtained from the 100 ns MD trajectory and validated the MM-PBSA results. The studies related to the energy decomposition per residues that contributed to the estimated MM-PBSA binding energy of ligand in the protein–ligand complex were also conducted.

## Results and discussion

### Molecular docking studies

The 2D representation of the best docking poses of 4 selected flavonoids and their interactions at the active site of TMPRSS2 modeled structure 1 obtained from Auto-dock Vina calculations is depicted in Figs. 3 and 4. The protein residues which show hydrogen bond/Van der Waals interactions with the ligands in these docking poses were extracted using the freeware Maestro 12.5 (Schrödinger Release 2019–3: Maestro, Schrödinger, LLC, New York, NY). The higher negative docking scores of Amentoflavone (−9.2 kcal/mol), Narirutin (−9.1 kcal/mol), Eriocitrin (−9 kcal/mol), and Naringin (−8.1 kcal/mol) are indications of their potential binding towards the TMPRSS2 target. The catalytic binding site of TMPRSS2 [8, 11] comprises amino acid residues from residue number 256–492 and the residues SER441, HIS296, and ASP345 form the catalytic triad. The ligand Amentoflavone with the highest binding affinity showed prominent Van der Waals interactions with the catalytic residues HIS296 and SER441. Furthermore, these residues exhibit either Van der Waal or hydrogen bonding interactions with all other flavonoids discussed in the current work. However, the catalytic residue ASP345 shows interactions only with Narirutin and Eriocitrin. The other amino acid residues MET424, LEU419, GLN438, TRP461, CYS465, TYR474, VAL473, and CYS437 also show Van der Waals interactions with Amentoflavone whereas the residue SER436 forms a hydrogen bond. The residues responsible for Van der Waal interactions with Narirutin are HIS296, ASP345, CYS437, GLN438, TRP461, and CYS465 and that form hydrogen bonding interactions are SER436 and SER441. Eriocitrin was found to form hydrogen bonds with the residues GLY462 and SER436, and hydrophobic contacts with the residues like ASP345, GLU299, GLU389,



**Fig. 3** 2D representation of best docking poses and protein–ligand interactions of **a** Amentoflavone and **b** Naringin. Pink lines indicate hydrogen bonds between ligand and residues. All other residues exhibit hydrophobic interactions with the ligand

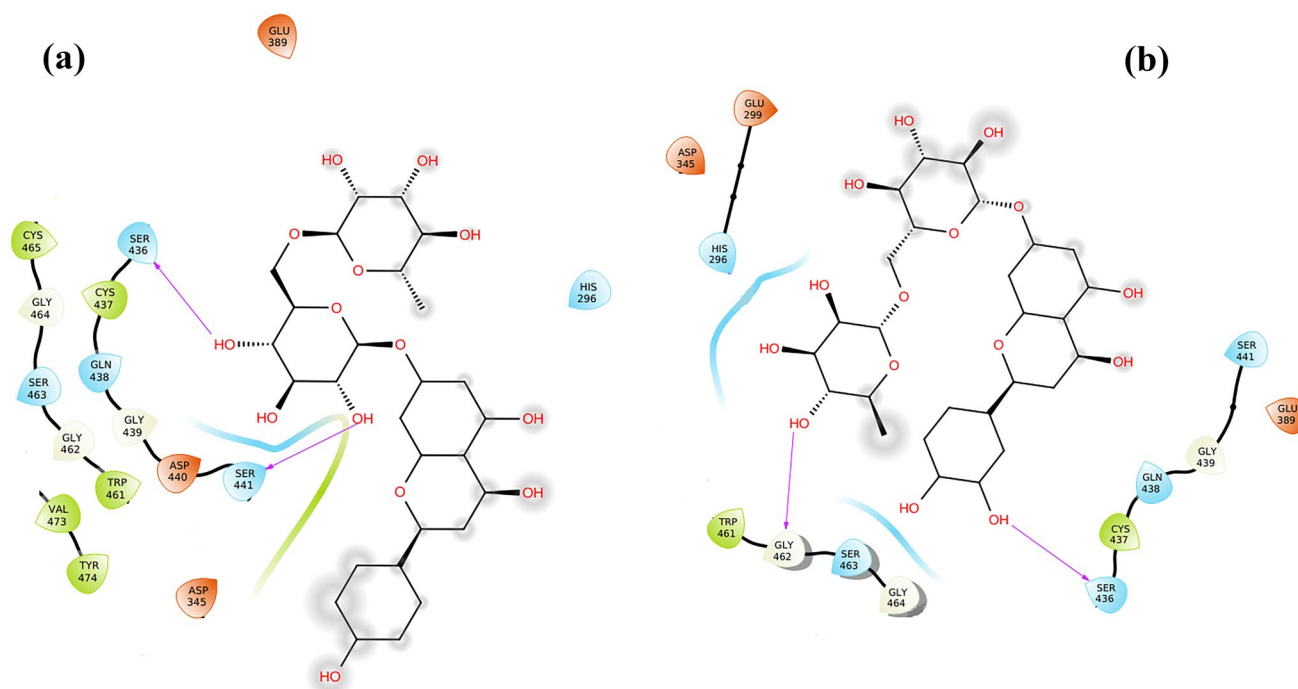
TRP461, HIS296, GLN438, SER441, and GLY464. Naringin was the ligand with the least docking score with Van der Waals interactions owned by the residues CYS297, HIS296, SER463, GLN438, SER441, and GLY464, and hydrogen bond interactions with the residues GLU299, GLY462, TRP461, and SER436. More hydrogen bonding interactions were observed in the docking pose of Naringin compared to Eriocitrin, Narirutin, and Amentoflavone.

For a benchmarking, we also carried out the docking calculations of known drugs Nafamostat and Camostat [2] which are active against TMPRSS2. The docking scores of Nafamostat and Camostat were found to be  $-8.4$  kcal/mol and  $-6.6$  kcal/mol, respectively, and were closer to the reported values [34]. We also analyzed the amino acid residues involved in the binding interactions between the TMPRSS2 target and ligands reported in the previous studies [35, 36]. We also found significant nonbonding interactions between the reported residues HIS296 and SER441 and the selected flavonoids in our study. Other than these, GLN438, ASP435, GLU299, and TRP461 residues of TMPRSS2 were reported to have significant interactions

with the ligands Gabexate, Camostat, and Nafamostat, respectively [35, 36]. The residues GLN438 and TRP461 also showed Van der Waals interactions with the selected flavonoids and GLU299 showed Van der Waals interaction with Eriocitrin and hydrogen bond interactions with Naringin. The analysis shows that our results are in good agreement with the reported literature. The validation of docking results (scores, RMSD and protein–ligand interactions) of all ligands in modeled structures and crystal structure of TMPRSS2 are provided in the supplementary information. We found that the RMSD of the best docking pose of all ligands in modeled structures 1 and 2 with respect to crystal structure is less than  $2 \text{ \AA}$ . Identification of the crucial residues from the docking calculations belonging to the catalytic site of TMPRSS2 gave insights for further analysis.

### Stability profile analysis of MD trajectory

In this section, we reported the results of trajectory analysis of 100 ns MD simulation carried out at 310 K. The steady energies of the simulating protein–ligand system with

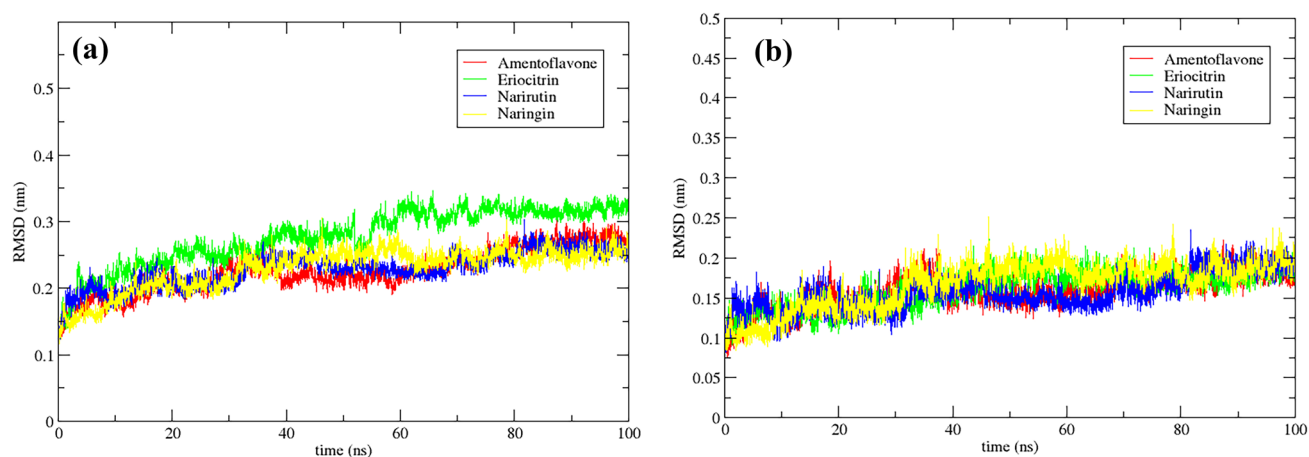


**Fig. 4** 2D representation of best docking poses and protein–ligand interactions of **a** Narirutin and **b** Eriocitrin. Pink lines indicate hydrogen bonds between ligand and residues. All other residues exhibit hydrophobic interactions with the ligand

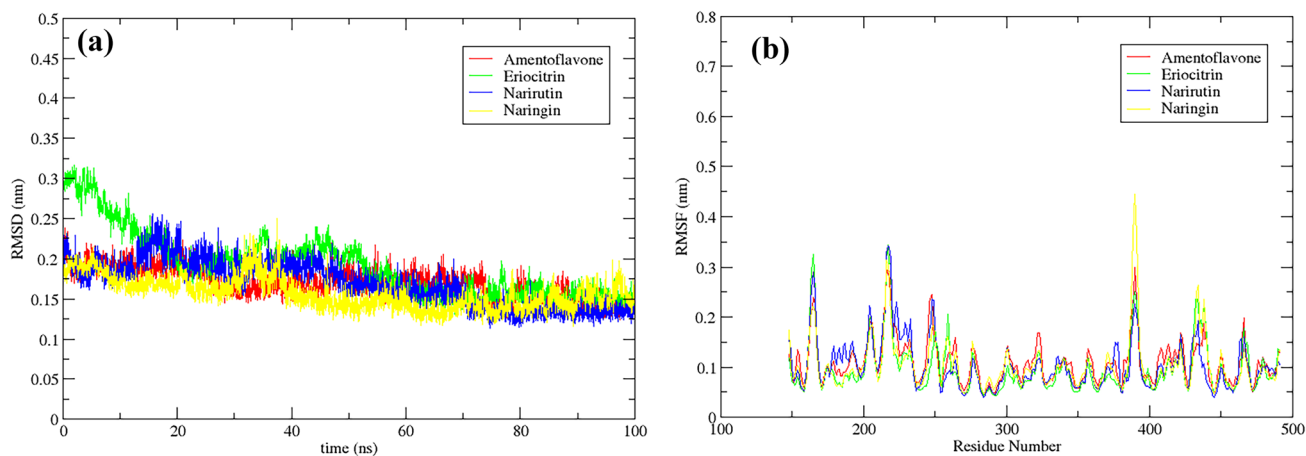
respect to time were observed and proper equilibration was ensured. Further detailed RMSD calculations of simulating complex system and protein backbone with respect to the equilibrated initial structure were conducted and plotted in Fig. 5a, b, respectively.

The RMSDs of Amentoflavone, Narirutin, and Eriocitrin complex lie within the range of 0.15–0.3 nm, which indicates the higher stability of these complexes in a solvated system. However, for the Eriocitrin complex system initially, there is a hike in the deviation and then comes down

to a steady plot in course of time. When analyzing the RMSDs of the protein backbone (Fig. 5b), minor fluctuations were observed for all protein–ligand systems within 0.2 nm. We also calculated the RMSD of the complex w.r.t the average structure from the MD trajectory and were plotted in Fig. 6a. The Narirutin complex showed negligible deviation from its average structure with an RMSD of 0.7 nm. The RMSDs of all protein–ligand systems showed minimum deviation within the range of 0.15–0.25 nm from their average structure except in the case of the Eriocitrin



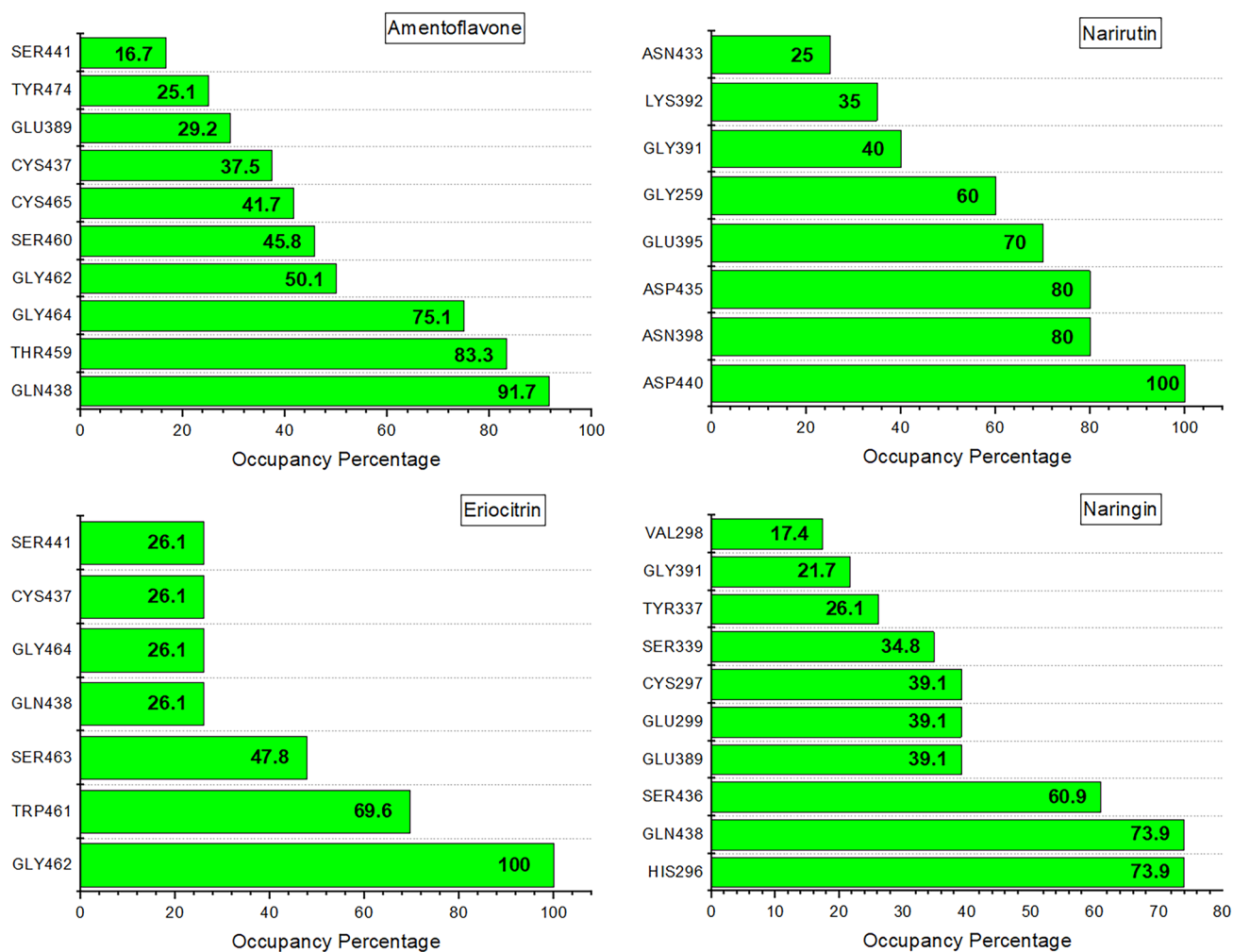
**Fig. 5** RMSD plots of **a** protein ligand complexes and **b** protein backbone in these complexes



**Fig. 6** **a** RMSD plots of complex with respect to average structure and **b** RMSF of residues of Tmprss2

complex (0.3 nm) at the beginning of the simulation. All steady RMSD plots with minimum deviations confirmed their stable structure in the 100 ns MD trajectory. Next,

we analyzed the contribution of individual protein residues to the structural fluctuations and their intensity in terms of RMSF (Fig. 6b). The protein residues in all complexes



**Fig. 7** Protein–ligand hydrogen bonding interaction profile of protein ligand complexes from MD trajectory

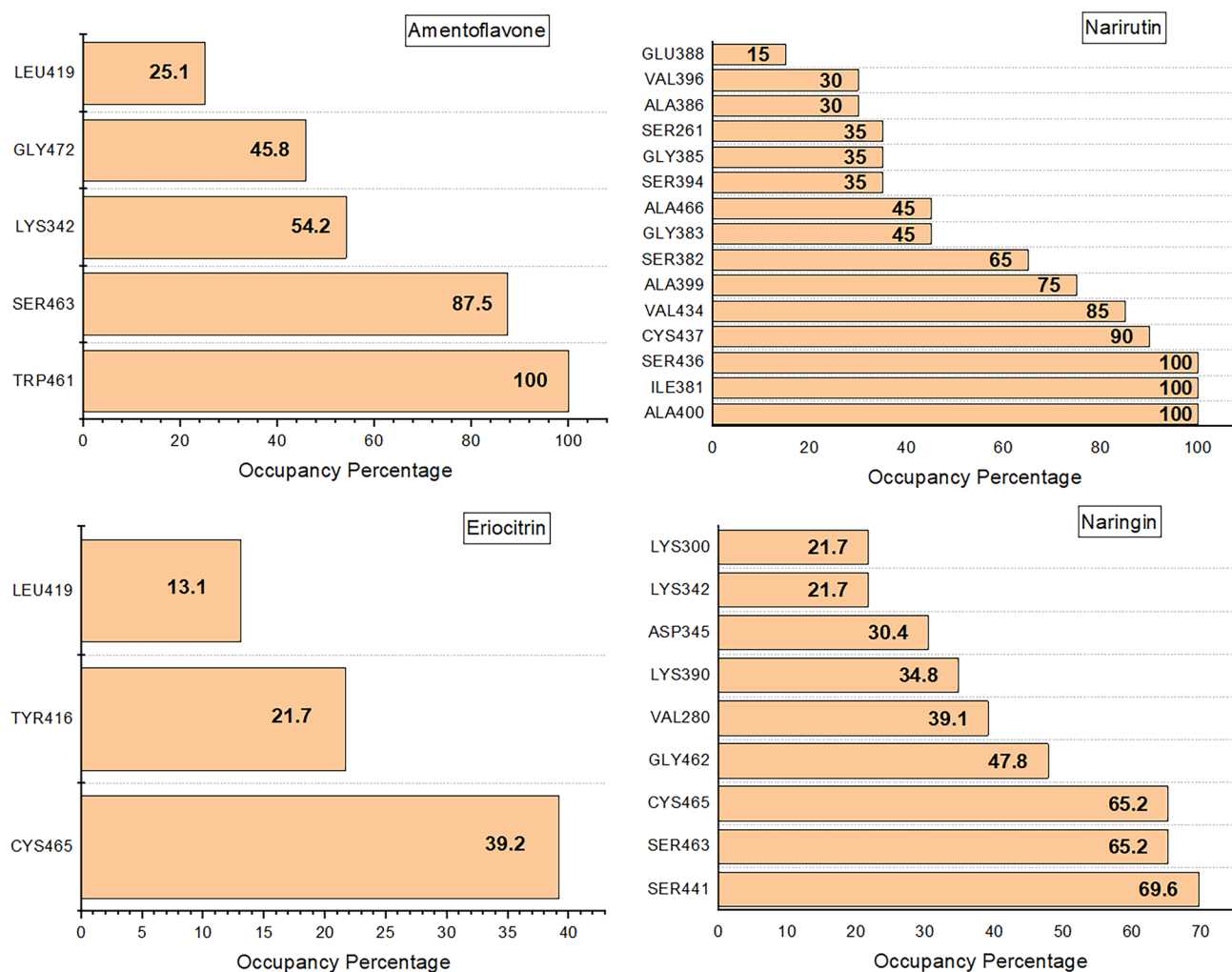


showed minimum fluctuations within the range of 0.45 nm. An average of 4–5 residues (LYS390, GLU389, GLY217, ALA216, and ARG165) showed comparatively higher fluctuations in the range 0.3–0.45 nm throughout the simulation. The lesser number of fluctuating residues is again an indication of stable protein–ligand systems in MD simulation.

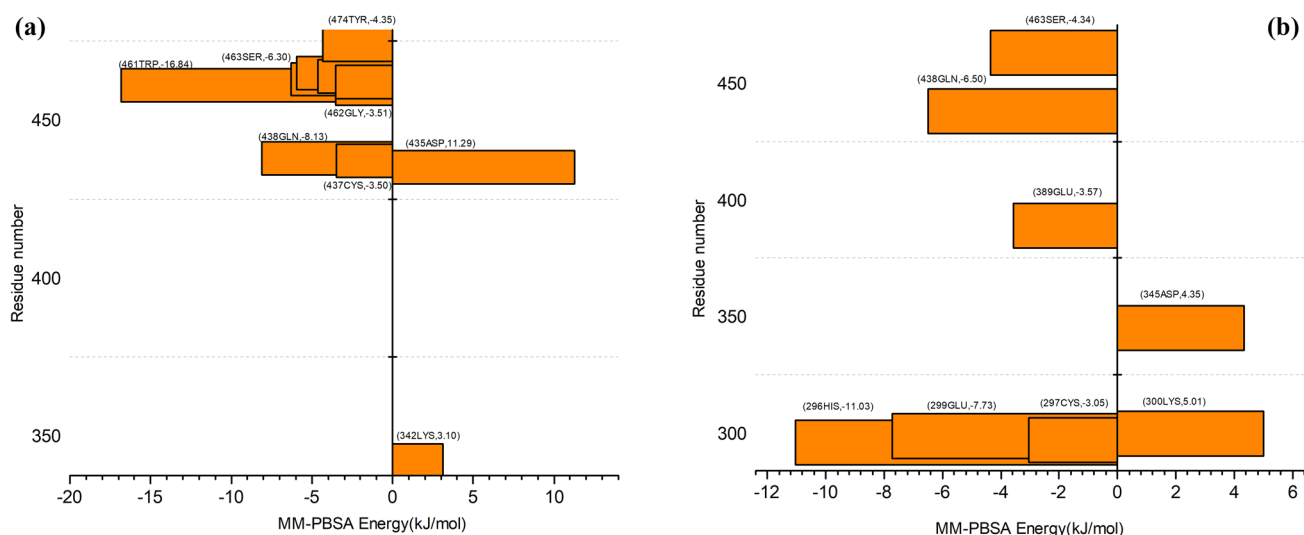
### Contact map analysis

The hydrophobic and hydrogen bonding interactions between ligands and TMPRSS2 target in the MD trajectory were analyzed using GROMACS and Pycontact tools [28]. The bond distance range for all hydrogen bond interactions was fixed as 1.5 to 3.5 Å and the minimum D-H-A bond angle in a hydrogen bond was fixed as 120°. The maximum cutoff bond distance for all hydrophobic interactions was fixed as 5.0 Å. The classes of interactions we considered in hydrophobic interactions include pi-stacking, pi-cation,

pi-alkyl, pi-amide, and other vdW interactions. The occupancy percentage of all interactions in the entire trajectory at 100 ps intervals is depicted in Figs. 7 and 8, respectively. The hydrophobic-vdW interactions play a vital role in effective protein–ligand binding. TRP461 was found as a crucial residue with a 100% occupancy, which showed a significant hydrophobic interaction with Amentoflavone through pi-amide stacking interaction. The other residues that form hydrophobic interactions with the Amentoflavone ligand are in the order of SER463 > LYS342 > GLY472 > LEU419. These residues interact with Amentoflavone through vdW contacts. The residues GLN438 and THR459 show high hydrogen bond occupancy of 91.7% and 83.3%, respectively, with Amentoflavone. The occupancy of other amino acid residues follows the order GLY464 > GLY462 > SER460 > CYS465 > CYS437 > GLU389 > TYR474. The occupancy percentage of hydrogen-bonded residues of Narirutin complex is in the order ASP440 > ASN398 = ASP435 > GLU395 > GLY259 > GLY391 > LYS392 > AS



**Fig. 8** Protein–ligand hydrophobic interaction profile of protein ligand complexes from MD trajectory



**Fig. 9** MM-PBSA binding energy decomposition of TMPRSS2 residues in **a** Amentoflavone complex and **b** Naringin complex

N433, whereas its hydrophobic residues follow the order ALA400 = ILE381 = SER436 > CYS437 > VAL434 > ALA399 > SER382. Among these residues, ALA400 and ILE381 interact with Narirutin via alkyl interactions and all other residues interact through vdW contacts. When it comes to Eriocitrin, strong hydrogen bonding interactions are shown by the residues GLY462, TRP461, and SER463 with occupancy percentages of 100%, 69.6%, and 47.8%, respectively, whereas the order of hydrophobic interaction is CYS465 > TYR416 > LEU419. The residue CYS465 interacts with Eriocitrin via alkyl interactions, whereas TYR416 and LEU419 interact through other vdW contacts. In the Naringin complex, the residues HIS296 and GLN438 show high hydrogen bond occupancy of 73.9% which is followed by SER436. The residues GLU389, GLU299, and CYS297 are found to have an occupancy percentage of 39.1%. Based on the profile, the significant residues that contribute to hydrophobic interactions with the Naringin ligand are in the order SER441 > SER463 = CYS465 > GLY462 > VAL280 > LYS390 > ASP345. The type of hydrophobic interactions in these residues are found to be pi-alkyl (LYS 390), pi-sulfur (CYS465), and other vdW (SER441, SER463, GLY462, VAL280, ASP345), respectively.

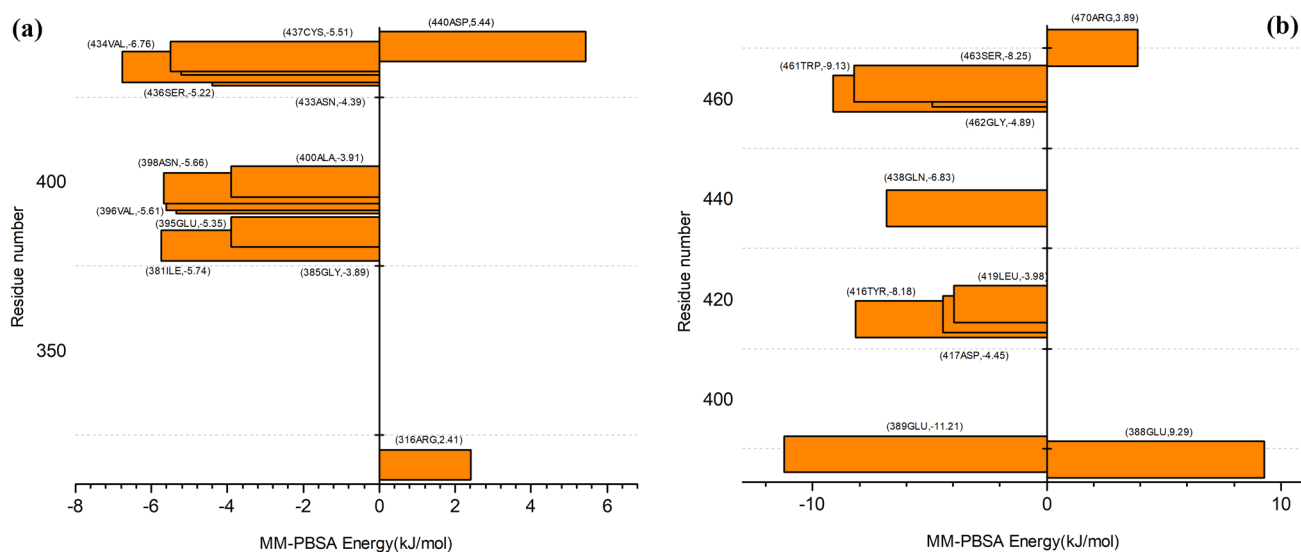
The catalytic triad SER441, ASP345, and HIS296 were found to be involved in significant non-bonding interactions with all the selected ligands.

### MM-PBSA binding energies

We reported here the results of MM-PBSA calculations using 1000 snapshots obtained from the MD-trajectory between 81 and 100 ns at 20 ps intervals. The MM-PBSA

binding energies obtained from the 20 ns MD replicates (provided in the supporting information) are found to be closer to the results reported in Table 1. It is the Van der Waal energy (vdW-E) that contributes most towards the total negative binding energy value of all four ligands. Along with this, the electrostatic energy (ESE) and the solvent accessible surface area energy (SASA-E) also contribute to the MM-PBSA binding energy with negative values. However, the increase in the polar solvation energy (PSE) with positive energy values reduces the total negative value of the binding energy. The increase in PSE is a consequence of the unfavorable interaction of residues with solvent molecules. The summation of the energies contributed by both protein and ligand residues to the components such as vdW-E, ESE, PSE, and SASA-E, respectively, gave the net MM-PBSA binding energy. The energy contribution of individual protein–ligand residues to the total binding energy was calculated using the code “energy2bfac” implemented in g\_mmpbsa.

The amino acid residues which contribute to the MM-PBSA binding energy with negative energy values are termed hotspot residues and those which contribute positive energy values are bad contact residues. From the MM-PBSA protein residue-free energy decomposition analysis, we identified the hot spots and bad contact residues of all four complexes. Among the four ligands, the binding energy of Amentoflavone towards the active site of TMPRSS2 is found to possess the highest negative binding energy value of  $-155.48$  kJ/mol. The energy contribution of Amentoflavone residue to this total binding energy is calculated as  $-85.25$  kJ/mol which is higher compared to other ligand residues.



**Fig. 10** MM-PBSA binding energy decomposition of TMPRSS2 residues in **a** Narirutin complex and **b** Eriocitrin complex

### Hot spot and bad contact residues in Amentoflavone complex (Fig. 11a)

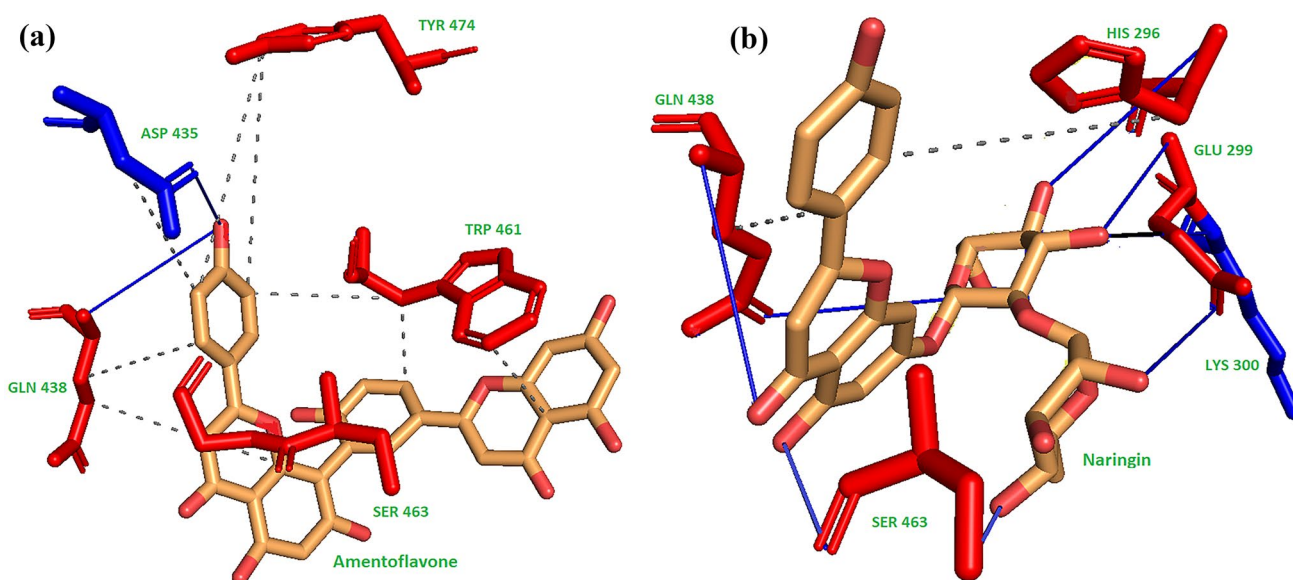
In the case of Amentoflavone, the hot spot residues (Figs. 9a and 11a) with negative binding energy are TRP461 (−16.84), GLN438 (−8.13), SER463 (−6.30), TYR474 (−4.35), and CYS437 (−3.50), respectively. However, ASP435, a bad contact residue that creates steric clashes inside the complex, contributes positive binding energy of 11.29 kJ/mol. The contact map analysis gave evidence that the protein residue GLN438 interacts with Amentoflavone using hydrogen bond interactions whereas TRP461 and SER463 through hydrophobic interactions. The residues that form hydrogen bonds with Amentoflavone contributed to the electrostatic energy part of MM-PBSA energy. TRP461 exhibits pi-amide stacking interaction and the other two residues have vdW contacts. These three hydrophobic contacts gave their contribution to the vdW component of total BE. A higher energy contribution of TRP461 residue (−16.49 kJ/mol) is pointing out its crucial involvement in the catalytic activity of TMPRSS2. The potential binding of Amentoflavone to TMPRSS2 is again confirmed from these results.

### Hot spot and bad contact residues in Narirutin complex (Fig. 12a)

The next potential inhibitor of TMPRSS2 is found to be Narirutin with the protein–ligand binding energy of −139.71 kJ/mol. The ligand residue contribution, in this case, is −76.98 kJ/mol. The hot spot residues (Figs. 10a and 12a) with negative binding energy contribution are identified as VAL434 (−6.76), ILE381 (−5.74), ASN398 (−5.66), VAL396 (−5.61), CYS437 (−5.51), GLU395 (−5.35), SER436 (−5.22), ASN433 (−4.39), ALA400 (−3.91), and GLY385 (−3.89). As per the non-bonding interaction profile, the higher occupancy of hydrophobic contacts of the residues ALA400, ILE381, SER436, CYS437, and VAL434 with Narirutin contributes significantly to the vdW component of MM-PBSA binding energy. Meanwhile, ASN398 and GLU395 residues in the Narirutin complex contribute through hydrogen bonding interactions. The ASP440 residue with unfavorable interactions resulted in a contribution of a positive binding energy value of 14.0 kJ/mol to MM-PBSA energy.

**Table 1** Binding energy values and individual component energy calculated with MM-PBSA method for ligands. All reported energy values have a standard deviation within the range of 10 kJ/mol

Sl No	Ligand	vdW-E	ESE	PSE	SASA-E	BE	SD
1	Amentoflavone	−213.469	−46.366	124.164	−19.902	−155.573	8.514
2	Narirutin	−188.967	−90.898	161.379	−21.223	−139.710	9.392
3	Eriocitrin	−210.970	−83.975	188.026	−23.705	−130.624	8.312
4	Naringin	−165.118	−90.722	160.086	−21.081	−116.834	10.451
5	Camostat	−86.431	−21.941	62.013	−11.823	−58.182	7.964
6	Nafamostat	−64.647	−60.267	90.462	−10.467	−44.920	10.495



**Fig. 11** Hotspot and bad contact residues in **a** Amentoflavone and **b** Naringin complex. Amino acid residue representation with red color and blue color indicate hot spot residues and bad contact resi-

due respectively. Dotted, blue and black lines indicate hydrophobic, hydrogen bond and steric clash respectively between amino acid residues and ligand

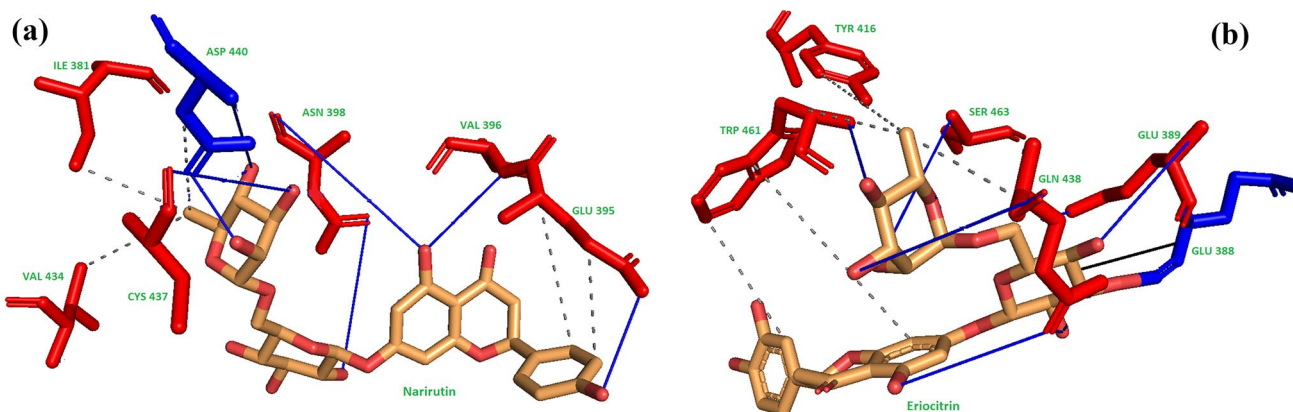
#### Hot spot residues in Eriocitrin complex (Fig. 12b)

Coming to the Eriocitrin complex, the total binding energy and the ligand residue contribution are  $-130.624$  kJ/mol and  $-79.58$  kJ/mol, respectively. The hotspot residues (Figs. 10b and 12b) and their binding energy contribution, in this case, include GLU389 ( $-11.21$ ), TRP461 ( $-9.13$ ), SER463 ( $-8.25$ ), TYR416 ( $-8.18$ ), GLN438 ( $-6.83$ ), GLY462 ( $-4.89$ ), ASP417 ( $-4.45$ ), and LEU419 ( $-3.98$ ), respectively. Two residues (GLU388 and ARG470) create steric clashes and hinder the effective protein–ligand binding. From the contact map analysis, it is found that TRP461,

GLY462, and SER463 interact with Eriocitrin through hydrogen bonding interactions and the residues TYR416 and LEU419 through hydrophobic-vdW interactions.

#### Hot spot residues in Naringin complex (Fig. 11b)

When the TMPRSS2–Naringin complex was analyzed, it is found that the residues (Figs. 9b and 11b) HIS296 ( $-11.03$ ), GLU299 ( $-7.73$ ), GLN438 ( $-6.50$ ), SER463 ( $-4.34$ ), GLU389 ( $-3.57$ ), and CYS297 ( $-3.05$ ) contribute to the negative binding energy, whereas the residues



**Fig. 12** Hotspot and bad contact residues in **a** Narirutin and **b** Eriocitrin complex. Amino acid residue representation with red color and blue color indicate hot spot residues and bad contact residue respec-

tively. Dotted, blue and black lines indicate hydrophobic, hydrogen bond and steric clash respectively between amino acid residues and ligand

ASP345 and LYS300 show unfavorable interactions. The contact map analysis results showed that the interaction of the protein residues GLN438, HIS296, GLU389, GLU299, and CYS297 with Naringin are through hydrogen bonds. The residues SER463 and GLY462 show hydrophobic-vdW contact with the receptor. This complex has the least negative value of  $-116.83$  kJ/mol and the ligand contribution was  $-73.23$  kJ/mol.

The rank of the ligands in the order of their inhibitory activity was found to be as Amentoflavone > Narirutin > Eriocitrin > Naringin. For a better comparison, we calculated MM-PBSA binding free energy of Camostat and Nafamostat using the same MM-PBSA method. The binding energies of Camostat and Nafamostat were found to be only  $-58.182$  kJ/mol and  $-44.92$  kJ/mol, respectively. The plots of MM-PBSA energy decomposition analysis of these compounds are given in the supporting data. We also analyzed the hot spot residues of TMPRSS2 obtained from the MM-PBSA calculations of other protein–ligand systems reported earlier [35, 36]. The MM-PBSA energy decomposition studies of TMPRSS2 residues in Camostat/Gabexate complexes identified GLN438 as the key hot spot residue. Similarly, TRP461 was identified as a hot spot residue when UKI-1 complexed with TMPRSS2 [34] and GLU389 was found as a hot spot residue when Bromhexine hydrochloride complexed with TMPRSS2 [36].

These hot spot residues reported earlier were found to be significant in our calculations also and they contributed with negative MM-PBSA binding energies in our calculations. The common hot spot residues of TMPRSS2 that interact with all the selected ligands are found to be GLN438, SER463, GLU389, GLY462, TRP461, and LEU419. The major roles of these residues in effective protein–ligand binding were previously observed in our non-bonding and molecular docking calculations. Concluding this section, we observed that all four flavonoids are better inhibitors than Nafamostat and Camostat in terms of MM-PBSA binding energies.

## Conclusion

The use of natural compounds as pharmacological agents will be highly beneficial in preventing the widespread of COVID-19. The major finding of the work is the identification of some selected flavonoids as potential inhibitors of human protease TMPRSS2, which is one of the potential targets of SARS CoV-2. By inhibiting TMPRSS2, they help in preventing viral entry and thus reducing the viral load inside the cell. Through molecular docking, MD simulation, contact map analysis, and the MM-PBSA calculations, we have identified Amentoflavone and Narirutin as

the flavonoids with very high inhibitory activity against TMPRSS2. These flavonoids bind well to the active site of the protein via hydrophobic and hydrogen bond interactions. All the flavonoids reported are natural non-toxic compounds with proven drug potential against several health issues. Also, most of the reported flavonoids are dietary compounds and therefore they can be used by patients without the fear of side effects. Since all these flavonoids are widely available at a low cost, they can be used for further studies to assess their pharmacological relevance.

## Associated content

We provide the following data and figures in the supporting file (a) B3LYP/SVP optimized coordinates of ligands used for docking, (b) Vina results of docking between modeled TMPRSS2 structures 1 and 2 (from Modeler) and selected ligands, (c) docking poses of selected flavonoids, (d) docking poses of Camostat and Nafamostat, (e) RMSD of best docking pose of ligands in modeled structures 1 and 2 with respect to 7MEQ, (f) MM-PBSA results obtained from MD simulation at 300 K, (g) MM-PBSA results of MD replicates, (h) Ramachandran plots of modeled TMPRSS2 structures, and (i) PDB file of modeled TMPRSS2 structure.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11224-022-01955-7>.

**Acknowledgements** We acknowledge DST-FIST Delhi, India, for providing the financial assistance of setting computational facilities in Heartian Center for Molecular Modeling (HCMM) of Sacred Heart College Thevara. We are also thankful to Dr. Girinath Pillai (Scientist, Nyro Research India) and Dr. Rajesh M (Assistant Professor, Department of English, Sacred Heart College Thevara) for giving the valuable guidance.

**Author contribution** Dr. Abi, Dr Sindhu, and Jibin K Varughese designed the research project and performed the whole theoretical and computational calculations. Kavitha J, Dhiya Francis, and Joseph Libin K L compiled the manuscript. Dr. Abi supervised the whole preparation of the manuscript.

**Availability of data and material** The preparation of target protein used in the article is performed with the help of the data obtained from RCSB Protein databank and Uniprot database. The ligand structures are obtained from Pubchem database.

**Code availability** We used licensed (academic) version of Turbomole in the DFT optimization calculation of ligands. Except Turbomole, all other molecular modeling tools used to execute computational and theoretical calculations are free-wares.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

1. Barge S, Jade D, Gosavi G, Talukdar NC, Borah J (2021) *Eur J Pharm Sci* 162:105820. <https://doi.org/10.1016/j.ejps.2021.105820>
2. Zhu H, Du W, Song M, Liu Q, Herrmann A, Huang Q (2021) *Comput Struct Biotechnol J* 19:467. <https://doi.org/10.1016/j.csbj.2020.12.035>
3. Varughese JK, Joseph Libin KL, Sindhu KS, Rosily AV, Abi TG (2021) Investigation of the inhibitory activity of some dietary bioactive flavonoids against SARS-CoV-2 using molecular dynamics simulations and MM-PBSA calculations. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2021.1891139>
4. Baby K, Maity S, Mehta CH, Suresh A, Nayak UY, Nayak Y (2021) *Eur J Pharmacol* 896:173922. <https://doi.org/10.1016/j.ejphar.2021.173922>
5. Roomi M, Khan Y (2020) Potential compounds for the inhibition of TMPRSS2. *ChemRxiv*. Cambridge, Cambridge Open Engage
6. Singh R, Gautam A, Chandel S, Ghosh A, Dey D, Roy S, Ravichandiran V, Ghosh D (2020) Protease inhibitory effect of natural polyphenolic compounds on SARS-CoV-2: An in silico study. *Molecules* 25:4604. <https://doi.org/10.3390/molecules25204604>
7. Middleton E Jr, Kandaswami C, Theoharides TC (2000) *Pharmacol Rev* 52:673
8. Fraser BJ, Beldar S, Seitova A, Hutchinson A, Mannar D, Li Y, Kwon D, Tan R, Wilson RP, Leopold K, Subramaniam S, Halabelian L, Arrowsmith CH, Bénard F (2021) *bioRxiv*: 2021.06.23.449282. <https://doi.org/10.1101/2021.06.23.449282>
9. Horowitz RM, Gentili B (1960) *J Am Chem Soc* 82:2803. <https://doi.org/10.1021/ja01496a033>
10. Mizelle JW, Dunlap WJ, Hagen RE, Wender SH, Lime BJ, Albach RF, Griffiths FP (1965) *Anal Biochem* 12:316. [https://doi.org/10.1016/0003-2697\(65\)90097-7](https://doi.org/10.1016/0003-2697(65)90097-7)
11. Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, Gallagher T (2011) *J Virol* 85:873. <https://doi.org/10.1128/jvi.02062-10>
12. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) *Nucleic Acids Res* 46:W296. <https://doi.org/10.1093/nar/gky427>
13. Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L (2017) *Schwede T* 45:D313. <https://doi.org/10.1093/nar/gkw1132>
14. Webb B, Sali A (2016) *Curr Protoc Bioinformatics* 54: 5.6.1. <https://doi.org/10.1002/cpbi.3>
15. Benkert P, Biasini M, Schwede T (2010) *Bioinformatics* 27:343. <https://doi.org/10.1093/bioinformatics/btq662>
16. Shen MY, Sali A (2006) *Protein Sci* 15:2507. <https://doi.org/10.1110/ps.062416606>
17. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) *J Appl Crystallogr* 26:283. <https://doi.org/10.1107/S0021889892009944>
18. Becke AD (1988) *Phys Rev A* 38:3098. <https://doi.org/10.1103/PhysRevA.38.3098>
19. Bauernschmitt R, Häser M, Treutler O, Ahlrichs R (1997) *Chem Phys Lett* 264:573. [https://doi.org/10.1016/S0009-2614\(96\)01343-7](https://doi.org/10.1016/S0009-2614(96)01343-7)
20. Trott O, Olson AJ (2010) *J Comput Chem* 31:455. <https://doi.org/10.1002/jcc.21334>
21. Berendsen HJC, van der Spoel D, van Drunen R (1995) *Comput Phys Commun* 91:43. [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)
22. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) *J Am Chem Soc* 118:11225. <https://doi.org/10.1021/ja9621760>
23. Dodda LS, Cabeza de Vaca I, Tirado-Rives J, Jorgensen WL (2017) *Nucleic Acids Research* 45: W331. <https://doi.org/10.1093/nar/gkx312>
24. Toukan K, Rahman A (1985) *Phys Rev B* 31:2643. <https://doi.org/10.1103/PhysRevB.31.2643>
25. Chen C, Huang Y, Ji X, Xiao Y (2013) *J Chem Phys* 138:164122. <https://doi.org/10.1063/1.4799236>
26. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) *J Chem Phys* 81:3684. <https://doi.org/10.1063/1.448118>
27. Di Pierro M, Elber R, Leimkuhler B (2015) *J Chem Theory Comput* 11:5624. <https://doi.org/10.1021/acs.jctc.5b00648>
28. Scheurer M, Rodenkirch P, Siggel M, Bernardi RC, Schulten K, Tajkhorshid E, Rudack T (2018) *Biophys J* 114:577. <https://doi.org/10.1016/j.bpj.2017.12.003>
29. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) *Acc Chem Res* 33:889. <https://doi.org/10.1021/ar00003j>
30. Zoete V, Irving MB, Michielin O (2010) *J Mol Recognit* 23:142. <https://doi.org/10.1002/jmr.1005>
31. Li S, Xu W, Chu S, Ma N, Liu S, Li X, Wang T, Jiang X, Li F, Li Y, Zhang D, Luo Q, Liu J (2019) *Chemistry – A European Journal* 25: 10350. <https://doi.org/10.1002/chem.201901480>
32. Genheden S, Ryde U (2015) *Expert Opin Drug Discov* 10:449. <https://doi.org/10.1517/17460441.2015.1032936>
33. Kumari R, Kumar R, Lynn A (2014) *J Chem Inf Model* 54:1951. <https://doi.org/10.1021/ci500020m>
34. Hou Q, Zhang L (2020) *Langmuir* 36:1813. <https://doi.org/10.1021/acs.langmuir.9b03832>
35. Hu X, Shrimp JH, Guo H, Xu M, Chen CZ, Zhu W, Zakharov AV, Jain S, Shinn P, Simeonov A, Hall MD, Shen M (2021) *ACS Pharmacology & Translational Science* 4:1124. <https://doi.org/10.1021/acspsci.0c00221>
36. Sonawane KD, Barale SS, Dhanavade MJ, Waghmare SR, Nadaf NH, Kamble SA, Mohammed AA, Makandar AM, Fandilolu PM, Dound AS, Naik NM, More VB (2021) *Informatics in medicine unlocked* 24:100597. <https://doi.org/10.1016/j.imu.2021.100597>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.