



## Identification of potent inhibitors against transmembrane serine protease 2 for developing therapeutics against SARS-CoV-2

Abdulla Al Mamun<sup>a,b</sup>, Farjana Akter<sup>a</sup>, Maksud Khan<sup>a</sup>, Sayeda Samina Ahmed<sup>a</sup>, Md. Giash Uddin<sup>a,c</sup>, Nabila Tabassum Tasfia<sup>a</sup>, Faiyaz Md. Efaz<sup>a</sup>, Md Ackas Ali<sup>a</sup>, Mossammad Umme Chand Sultana<sup>a</sup> and Mohammad A. Halim<sup>d</sup>

<sup>a</sup>Division of Infectious Diseases and Division of Computer Aided Drug Design, The Red-Green Research Centre, Dhaka, Bangladesh; <sup>b</sup>Key Laboratory of Soft Chemistry and Functional Materials of MOE, School of Chemical Engineering, Nanjing University of Science and Technology, Nanjing, P. R. China; <sup>c</sup>Department of Pharmacy, University of Chittagong, Chittagong, Bangladesh; <sup>d</sup>Department of Chemistry and Biochemistry, Kennesaw State University, Kennesaw, GA, USA

Communicated by Ramaswamy H. Sarma

### ABSTRACT

In viral binding and entry, the Spike(S) protein of SARS-CoV-2 uses transmembrane serine protease 2 (TMPRSS2) for priming to cleavage themselves. In this study, we have screened 'drug-like' 7476 ligands and found that over thirty ligands can effectively inhibit the TMPRSS-2 better than the control ligand. Finally, the three best drug agents L1, L2, and L6 were selected according to their average binding affinities and fitting score. These ligands interact with Asp435, Cys437, Ser436, Trp461, and Cys465 amino acid residues. The three best candidates and a reported drug Nafamostat mesylate (NAM) were selected to run 250 ns molecular dynamics (MD) simulations. Various properties of ligand-protein interactions obtained from MD simulation such as bonds, angle, dihedral, planarity, coulomb, and van der Waals (VdW) were used for principal component analysis (PCA) calculation. PCA discloses the evidence of the structural similarities to the corresponding complexes of L1, L2, and L6 with the complex of TMPRSS2(TM) and Nafamostat mesylate (TM-NAM). Moreover, Quantitative structure-activity relationship (QSAR) pattern recognition was generated using PCA for the investigation of structural similarities among the selected ligands. Multiple Linear Regression (MLR) model was built to predict the binding energy compared to the binding energy obtained from molecular docking. The MLR regression model reveals an accuracy of 80% for the prediction of the binding energy of ligands. ADMET analysis demonstrates that these drug agents are appeared to be safer inhibitors. These three ligands can be used as potential inhibitors against the TMPRSS2.

### ARTICLE HISTORY

Received 24 January 2021  
Accepted 8 September 2021

### KEYWORDS

SARS-CoV-2; TMPRSS2; virtual screening; molecular dynamics; QSAR

## 1. Introduction

A major viral respiratory disease appeared in Wuhan, Hubei Province, China, in December 2019 (Jaimes et al., 2020; Wu et al., 2020; Yuan et al., 2020). The initial cluster of infections known as Coronavirus infected not only animals but also humans. Later, it occurred transmission at human-human and spread quickly from China to over the world (Hoffmann et al., 2020; Gorbalenya et al., 2020; L. Zhang et al., 2020). This new corona type virus is known as Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) since eighty-two percent of similar RNA genome analogous found to the SARS coronavirus (SARS-CoV), and both viruses belong to the Betacoronavirus b gene clade (L. Zhang et al., 2020).

The human host cell transmembrane protease serine 2 (TMPRSS2, Uniprot-O153934) which belongs to the transmembrane serine protease (II) family is found in epithelial cells of the human respiratory and gastrointestinal tracts (Bertram et al., 2012; Shrimp et al., 2020). It plays an important role in viral entry and spread in the host (Iwata-

Yoshikawa et al., 2019). The predominant coronavirus antigen is a surface spike glycoprotein (S) which facilitates viral entry by activating the host receptor and negotiating virus-host membrane fusion (Yuan et al., 2020). A two steps mechanism is followed here to activate SARS CoV-2. In the first step, facilitating host-cell entry, viral entry hemagglutinin protein is connected with angiotensin-converting enzyme 2 (ACE2), encoded by the ACE2 gene, that is expressed on respiratory epithelial cells. In the second step, cleavage of hemagglutinin is occurred to activate the internalization of the virus. This second step is proceeded through proteases particularly TMPRSS2, on the host cell (Walls et al., 2020).

It is also assumed that the S protein of SARS CoV-2 is activated by TMPRSS2 and furin. SARS-CoV-2 S protein is cleaved by furin at the S1/S2 site. The S2' site of S protein is cleaved by TMPRSS2 (Bestle et al., 2020). Here, spike protein interacts with ACE2 as the passage receptor and utilizes the cell serine protease TMPRSS2 for S protein preparation (Hoffmann et al., 2020; Kumar et al., 2020). Thus, TMPRSS2 is also responsible for SARS-CoV-2 replication.

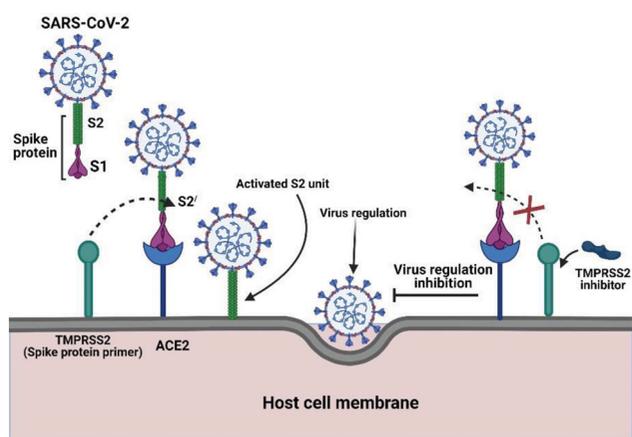


Figure 1. An overview between Spike protein with TMPRSS2 and small molecule.

It has already been shown that TMPRSS2 and HAT split and activate 229E- for cathepsin L- host cell entry. TMPRSS2 triggers diverse human respiratory viruses, therefore, it may stimulate viral propagation in humans (S. Bertram et al., 2013). Nafamostat mesylate, a serine protease inhibitor, inhibits TMPRSS2-dependent host cell entry of MERS-CoV (Yamamoto et al., 2016). Recently, SARS-CoV-2 infection of human lung cells is highly blocked by **nafamostat mesylate** than camostat mesylate while these compounds had inactive against vesicular stomatitis virus infection (Hoffmann et al., 2020). In light of the proven safety of nafamostat mesylate, we consider this as a standard drug for COVID-19 treatment.

This study is conducted to screen a large chemical database with the final goal of identifying the best ligands for design an effective drug for COVID-19. We explore potent ligand candidates, which are found to be more effective than nafamostat mesylate to inhibit the activation of TMPRSS2, and this is primarily responsible for the entrance of SARS CoV-2 (Figure 1).

## 2. Methodology

### 2.1. Modeling of protein structure and structural validation

The 3D structure of the TMPRSS2 protein was unavailable in the PDB database (<http://www.rcsb.org/>) (Berman et al., 2002). Therefore, the amino acid sequences of the protein were collected from the UniProt database (UniProtKB ID **O15393**) (Apweiler, 2009), and the 3D structure of TMPRSS2 was generated by using the SWISS-MODEL server (Waterhouse et al., 2018). Further, the structure validation of the predicted model was investigated by the Phi/Psi Ramachandran plot using PROCHECK Figure 3 (R. A. Laskowski et al., 1993). The quality of the protein model was calculated by employing the Protein Structure Analysis (ProSA) Figure S1 (Wiederstein & Sippl, 2007). Then, the generated structure of targeted protein TMPRSS2 was modelled using the template of Serine protease hepsin (PDB ID: **5CE1**) (Meng et al., 2020) as shown in Figure S2. Moreover, Secondary structure analysis carried out by discovery studio and PDBsum shown in Figure 2A and 2B (Laskowski et al., 2005). Finally, TM-align algorithm was also employed and the

best model was selected based on TM-score (Y. Zhang & Skolnick, 2005).

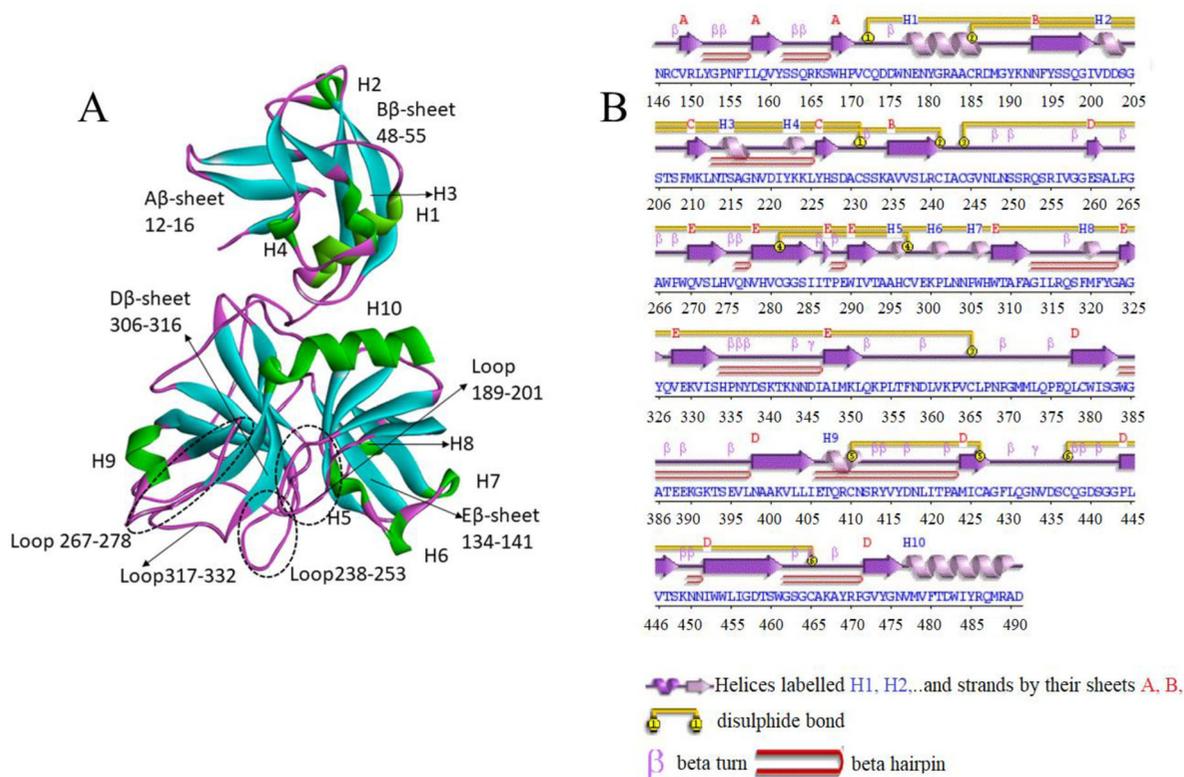
The predicted model structure does not provide information on the location of hydrogens, commonly referred to as the protonation state. The accurate prediction of the correct protonation state, especially within the binding interface, is crucial to accurately predict the correct binding mode and, binding affinity (Onufriev & Alexov, 2013; Petukh et al., 2013). The protonation state of ionizable residues (Asp, Glu, Arg, Lys, Tyr, His, Cys) was set under pH 6.5 based on the pKa values calculated by the H++ server (Anandakrishnan et al., 2012; Gordon et al., 2005).

### 2.2. Virtual screening

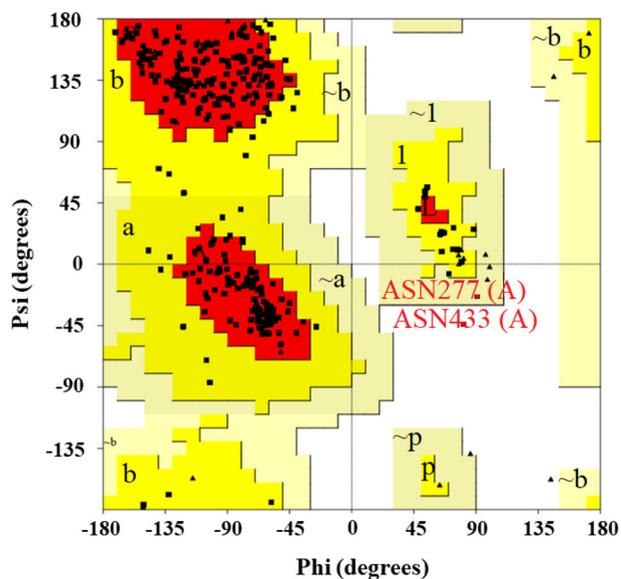
Total 'drug-like' 7476 ligands were predicted against the TMPRSS2 gene (UniProtKB ID O15393) in the ZINC15 database (<http://www.zinc15.docking.org/>) (Apweiler, 2009; Glowacka et al., 2011; Sterling & Irwin, 2015). Zinc15 performed the prediction with a combined method of Similarity Ensemble Approach (SEA) and maximum Tanimoto similarity. It also uses compound annotations from ChEMBL Version 21 (Irwin et al., 2018). In addition, these ligands have the formula mass of less than 700 g/mol, calculated partition coefficient (clogP) ranges from  $-1$  to  $-5$ , and hydrogen bond donors and acceptors are less than 5 and 10, respectively. To optimize the ligands, UFF (universal force field) (Artemova et al., 2016; Zorn et al., 2008) was used with the steepest descent optimization algorithm along with the 2000 number of minimization steps. Using MGL software package in AutoDock Tool (ADT), the protein and ligands PDB files are converted into pdbqt formats. After preparing of protein and ligands, the grid box of AutoDock Vina (using PyRx-0.8 version) was created counting the active binding sites of the targeted TMPRSS2 Model (TM) in which centers were 15.98 Å,  $-7.27$  Å, 24.45 Å, and the dimensions were 18.33 Å, 19.42 Å, 17.66 Å for X, Y, Z directions respectively. The exhaustiveness number is 8, and nine binding modes for every tested ligand.

The ligand's binding affinities were measured in kcal/mol as a negative score unit in which the higher negative score was equal to the more binding affinity. The preliminary virtual screening of 7476 ligands was conducted by AutoDock Vina (Trott & Olson, 2010) docking protocol. Molecular docking simulation of every drug candidate was performed against the targeted active binding sites (HIS296, ASP435, SER436, CYS437, GLN438, GLY439, TRP461, GLY462, GLY464, and CYS465) of TM protein, and the binding sites were collected from various literatures (Hempel et al., 2021; Hussain et al., 2020; Kumar et al., 2020; Singh et al., 2020). According to the measured binding affinity, the top thirty-one potential drug candidates were selected for further study.

To get more insight about the docking analysis of top thirty-one ligands obtained AutoDock Vina, GOLD 5.7 (Genetic Optimization for Ligand Docking) docking (Jones et al., 1997) was also performed to support the AutoDock Vina results. The Hermes visualizer in the GOLD suite was applied to prepare the ligand and protein. For further analysis of



**Figure 2.** (A) 3D structure of TMRSS2 Model (TM) protein indicating helix, beta-sheet, and loop as green, cyan, and purple color respectively. (B) Secondary structure visualization created by PDBSum.



**Figure 3.** Ramachandran plot of TMRSS2 protein. The plot calculations on the 3D models of TMRSS2 protein were calculated with the PROCHECK server. Most favored regions are represented by colored red and additional allowed, generously allowed, and disallowed regions are designated as yellow, light yellow, and white fields, respectively.

docked protein-ligand complexes, Discovery Studio Software (version 19.1.0.18287) was used. AutoDock Vina result showed that L1 (ZINC000116505796, (R)-N-(4-carbamoylbenzyl)-1-(1-hydroxy-2-naphthoyl) pyrrolidine-2-carboxamide), L2 (ZINC000436573789, (R)-N-(3-carbamoylbenzyl)-1-(1-hydroxy-2-naphthoyl)pyrrolidine-2-carboxamide) and L6 (ZINC000026291155, N-(4,6-diphenylpyrimidin-2-yl)cyclohexanecarboxamide) obtained more binding affinity as well as non-covenant interaction compared to NAM

(Nafamostat). Moreover, the pKa values of the potent three ligands were calculated by the chemaxon server (<https://www.chemaxon.com>) shown in Table 1. Therefore, these three and NAM were selected for molecular dynamics simulation.

### 2.3. Molecular dynamics (MD) simulations and binding free energy

Molecular dynamics simulation was performed using the YASARA Dynamics software (Krieger et al., 2004) for Apo form (without ligand) and holo forms (ligand-protein) of TMRSS2 model to understand their conformational dynamics. The AMBER14 force field was used for all calculations (Dickson et al., 2014). The total system had 52, 583 atoms where water molecules were included, which was neutralized by adding NaCl salt at 0.9% concentration (Krieger et al., 2006) at 298 K temperature. At physiological conditions, the Berendsen thermostat was used to control the simulation temperature and it was carried to the short-range Van der Waals and Coulomb interactions, and long-range electrostatic particle-mesh Ewald (PME) method (Darden et al., 1999) was utilized at a cut-off radius of 8 Å.

For the simulation, the periodic boundary condition was included, where the box size was (88.42Åx76.53Åx76.89 Å). Using the time stage 1.25 fs (Krieger & Vriend, 2015), simulation snapshots were saved at every 100 ps. The primary energy minimization process of each simulation system was continued to carry out using the simulated annealing method, the steepest gradient approach was applied in 5000 cycles. Molecular dynamics simulations were performed over 250 ns at constant pressure and Berendsen thermostat.

Energy, bond distance, bond angle, dihedral angle, hydrogen bond, solvent accessible surface area (SASA), radius of gyration (Rg), root mean square fluctuation (RMSF), and root mean square deviation (RMSD) values for backbone and heavy atoms were collected throughout the analysis for further calculations.

PCA is one of the most recognizable and best-known techniques of multivariate analysis. In our study, the main objective of applying this method is to express the correlation among the variables in the collected energy profile of MD trajectory data (R. Islam et al., 2020). PCA represents the multivariate response that is arranged in an X matrix into a product of two new matrices as indicated in the following equation:

$$X = T_k P_k^T + E$$

Here,  $T_k$  is the matrix of scores which represents how samples relate to each other,  $P_k$  is the matrix of loadings which contains 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, which contains the information not retained by the model. Different energy profiles may be observed between complexes of four ligands with protein and main protein, i.e. apo-protein during MD simulation. Through the PCA algorithm, the difference in energy profile can be identified (J. Islam et al., 2019). R platform was used to perform all calculations using in-house developed codes (Team, 2014). The ggplot2 package (Wickham, 2008) was used to generate the plot. Before applying the PCA algorithm (Martens & Ni, 1992), autoscale function was used to preprocess data. The PCA was analyzed by taking the final 150 to 250 ns of MD trajectory data.

Then binding free energy was calculated using PRODIGY-LIG webserver (Vangone et al., 2019). In this server, they predicted the binding free energy based on the number of atomic contacts (ACs) within the distance threshold of 10.5Å and classified the ACs according to the atom involved in the interaction (C = Carbon, O = Oxygen, N = Nitrogen, X = all other atoms). The calculation was conducted for the last 100 ns from MD simulation results of selected three and NAM complexes. The free energy calculation results were obtained using this equation 1:

$$\Delta G_{noelec} = 0.0354707 AC_{NN} - 0.1277895 AC_{XX} - 0.0072166 AC_{CN} - 5.1923181 \quad (1)$$

where,  $AC_{NN}$ ,  $AC_{CN}$ , and  $AC_{XX}$  are the ACs between Nitrogen-Nitrogen, Carbon-Nitrogen, and between all other atoms and polar hydrogens, respectively.  $\Delta G_{noelec}$  is the electrostatic energy calculated through the HADDOCK (van Zundert & Bonvin, 2014) refinement protocol.

#### 2.4. Quantitative structure-activity relationships (QSAR) of ligands

Out of 7476 ligands, the top potential ligands were selected based on the binding energy for the QSAR study. Initially, 21 ligands were arbitrarily selected as the 'training set' and 10 ligands were utilized as the 'test set' or 'validation samples'.

**Table 1.** Calculated pKa values of potent ligands.

Ligands	Breakdown	pKa	Nature
L1	Naphthol	14.59	Basic
	Tertiary amine	15.59	
	Carbonyl (amide)	13.95	
L2	Carbonyl	-0.70	Basic
	Tertiary amine	15.58	
	Naphthol	14.58	
	Carbonyl (amide)	13.94	
L6	Amide (NH <sub>2</sub> )	-1.36	Basic
	Carbonyl	-0.66	
	=NH (carbamimidoyl)	12.78	
Nafamostat	=NH (carbamimidoyl)	12.81	Basic
	Amino	11.65	

Multiple linear regression (MLR) analysis was applied to find the fundamental correlation between the structure-activity relationships with the calculated binding energy values obtained from molecular docking (Fakayode et al., 2009; 2014; Liu et al., 2017). In addition to binding energy, some other important variables like TPSA (topological polar surface area, Å<sup>2</sup>), molecular weight, XLogPs-AA, ROTB count, number of H, C, O, Cl, N, and F atoms, single bonds (SB) count, double bonds (DB) count, and the number of benzene ring were taken into consideration. MLR data processing was carried out using the XLSTAT software package (Adinsoft, 2010).

#### 2.5. Pharmacokinetic properties

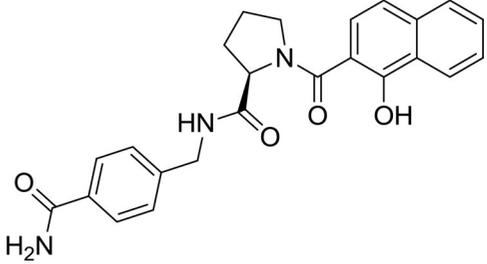
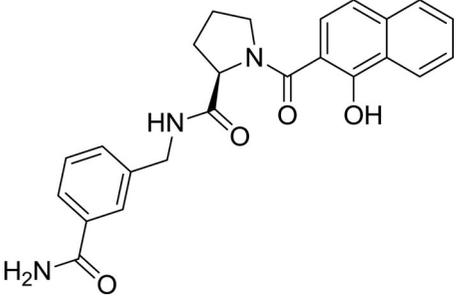
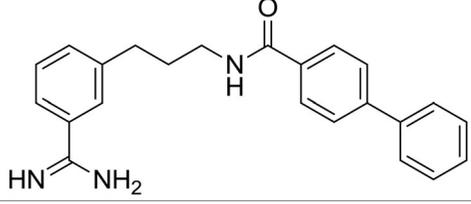
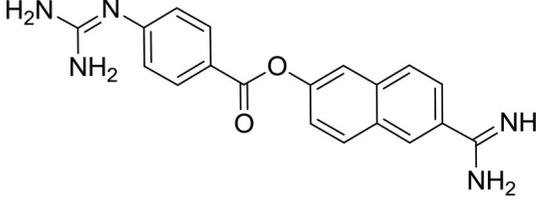
The best possible pharmacokinetics and lethality profile alongside viability are the significant determinants owing to effective medication improvement (Dirar et al., 2016). ADMET (Absorption, Distribution, Metabolism, Excretion & Toxicity) study validates the pharmacokinetics parameters of a drug molecule (Nisha et al., 2016). ADMET structure-activity relationship database, known as AdmetSAR is an online service that gives the data in regards to the lethality, cancer-causing nature of the medications and discovers whether the potential drug candidates follow the Lipinski Rule. In our study, admetSAR 2.0 (<http://lmmd.ecust.edu.cn/admetSAR2/>) was used as a platform to predict ADMET properties. The server was being used by uploading structure data file (sdf) and simplified molecular-input line-entry system (SMILES) strings for the modified drug. Here, the simplified molecular-input line-entry system (SMILES) of the settled ligands was submitted in the admetSAR program to check for lethality. ADMET-related properties were calculated by counting the molecular formula (MF), the blood-brain barrier (BBB), Human either-a-go-go inhibition (hEGI), Human intestinal absorption (HIA), P-glycoprotein inhibitor (PGI), Alogp, Caco-2, Carcinogenicity (binary), mol MW. Due to having significant impacts on the performance and pharmacological activity of the drug of these properties, they were considered in our study.

### 3. Results and discussion

#### 3.1. Validation of predicted TMPRSS2 structure

Model validation is an essential part of homology modeling. The secondary structure analysis shows that the similarity between TMPRSS2 and hepsin is 33.82% with a Q mean

**Table 2.** Docking results of three ligands along with fitting score with TM (AutoDock Vina scores are in kcal/mol).

Entry	Structures	Auto Dock Vina with protonated protein	GOLD
L1		-9.5	74.05
L2		-9.8	68.03
L6		-9.4	88.15
Standard		-8.5	69.51

value of  $-1.43$ . The structural analysis shows that there are 5 sheets, 10 helices, 6 disulphides, and 21 strands are present in the model structure of TMPRSS2 as shown in [Figure 2A](#) and [2B](#).

The stereochemical qualities of the predicted model of TMPRSS2 protein obtained from the SWISS-MODEL server are validated by generating a Ramachandran plot using the PROCHECK (R. A. Laskowski et al., 1993; Waterhouse et al., 2018). The shading on the Ramachandran plot showing in [Figure 3](#) for TMPRSS2 protein is used to indicate different regions such as the red areas reveals the most favorable region for phi-psi combination. In case of better stereochemical quality of any model, maximum residues of protein (about 90%) in the core regions should be found. The plot statistics for TMPRSS2 protein showed 253 residues (84.6%) in the most favored regions, 44 residues (14.7%) are in the additional allowed regions, and also 1 residue (0.3%) in generously allowed regions whereas only 1 residue (0.3%) is in disallowed regions showing in [Figure 3](#) and [Table S1](#). [Table S2](#) represents the main and side chain parameters which also indicates the accuracy of the structure prediction. Other

main chain parameters including omega angle standard deviation, bad contacts/100 residues, zeta angle standard deviation, hydrogen bond energy standard deviation, and overall G factor are determined by PROCHECK. These results also reveal that the SWISS-MODEL generated structure has better structural qualities ([Table S2](#)). Side-chain parameters are summarized as chi1-chi2 side-chain torsion angle combinations for all residues in [Table S3](#) indicate SWISS-MODEL generated structure exhibited a higher score in terms of overall structural accuracy.

A better quality structure typically contains a much lower residue fraction in disallowed regions (Gunasekaran et al., 1996). Therefore, the stereochemical quality of the predicted TMPRSS2 protein model is considered appropriate. The ProSA server is used to recognize the potential errors in the predicted model. Z-score obtained from this server represents the quality of the whole model and the total energy deviation of the structure can be measured from an energy distribution that is derived from random conformations. The Z-score obtained from the server is plotted compared to experimentally found Z-scores of all similar size proteins through NMR and X-

**Table 3.** The average value of RMSD (Å), SASA (Å<sup>2</sup>), Radius of gyration (Å), Number of hydrogen bonds, and Binding free energy (kcal/mol) from 250 ns MD results.

Complex	RMSD (Å)	SASA (Å <sup>2</sup> )	The radius of gyration (Å)	Number of H-bonds (with ligand)	Binding free energy (kcal/mol)
Apo	2.79 ± 0.29	17414.24 ± 274.14	22.06 ± 0.20	–	–
TM-L1	3.70 ± 0.57	16757.20 ± 274.69	22.29 ± 0.19	5 ± 1	–7.67 ± 0.02
TM-L2	3.81 ± 0.43	16890.20 ± 347.80	22.11 ± 0.25	3 ± 1	–8.01 ± 0.24
TM-L6	8.26 ± 1.10	18169.17 ± 330.63	23.72 ± 0.29	4 ± 1	–8.69 ± 0.02
TM-NAM	6.49 ± 1.55	16853.36 ± 401.73	23.09 ± 0.41	3 ± 1	–7.84 ± 0.12

**Table 4.** Predicted binding energy by MLR regression and actual binding energy from molecular docking.

Sample	Predicted binding energy (kcal/mol)	Actual binding energy (kcal/mol)	%RE
L2	–9.25	–9.70	–1.75
L5	–8.96	–9.30	–1.31
L11	–8.95	–9.10	–0.54
L13	–9.07	–9.10	–0.09
L17	–8.71	–8.90	–0.73
L25	–9.16	–8.80	1.42
L26	–9.15	–8.80	1.38
L29	–8.98	–8.70	1.11
L7	–9.13	–9.10	0.13
L14	–8.82	–8.90	–0.28
<b>RMS%RE</b>			<b>1.61</b>

ray (Wiederstein & Sippl, 2007). The z-score obtained for TMPRSS2 protein is –8.67 denoting no significant deviation from the scores calculated for the proteins of closure structure of PDB database showing in Figure S1. Finally, the model structure from SWISS-MODEL is also aligned with the PDB ID:5CE1 by using TM-align server, where amino acid residues are aligned relatively better with SWISS-MODEL (TM-score: 0.96; RMSD 0.91 Å) model. In summary, the results from PROCHECK, ProSA, and TM-align analysis clearly reported that the modeled structure of TMPRSS2 generated by using SWISS-MODEL program was superior in quality. Therefore, further study of the modeled structure of TMPRSS2 is carried out with SWISS-MODEL generated structure.

In the modeled TMPRSS2 structure, Asp, Glu, Arg, Lys, Tyr, His, and Cys residues were protonated under pH 6.5 based on the pKa values. Further, during the protonation state, His296 residue switches to the Hip296 after protonation of the TMPRSS2 model which has a pKa value of 8.91. Histidine (His) residue exhibits three different conformations which can be protonated. The imidazole ring of the His side-chain can be protonated in a neutral confirmation at the ε-nitrogen (HIE) or the δ-nitrogen (HID) or in a charged (+1) conformation (HIP) where both the ε- and δ-nitrogens are protonated (Kim et al., 2013).

### 3.2. Virtual screening and molecular interactions

Structure-based virtual screening is performed to screen about 7476 ligands. A cut-up value –8.7 (the binding affinity of NAM, control ligand) is applied. Thirty-one ligands are shown better binding affinity than the control ligand and they are selected for further analyses. Again, selected 31 compounds were screened three times by AutoDock vina and reported their average binding affinity. At that time, the GOLD suit was used to support the Audock Vina docking results. Higher fitness scores in GOLD docking, as well as

higher binding affinities in AutoDock, refer to the better docking interaction between ligand and protein. Three ligands L1, L2, L6, and NAM are docked with protonated protein TM to compare the binding affinity (Table 2). The binding affinity and fitness score of the top thirty-one ligands are shown in and Table S4. Besides, a standard TMPRSS2 inhibitor such as NAM is subjected to dock to compare the findings with the zinc database ligands. In this study, the binding affinity of 31 ligands is compared to NAM. Docking results show that three ligands have more binding affinities than NAM. The highest binding affinities are observed from the corresponding complexes of L1, L2, and L6 and while NAM shows the binding affinity of –8.5 kcal/mol. The results of the selected ligands are expected as the possible inhibitor of TMPRSS2, clearly representing the selected ligands will display stronger inhibitory effects than standard drugs.

A significant number of non-covalent interactions such as hydrogen bond, hydrophobic and electrostatic interactions are detected in three ligands with TMPRSS2 as shown in Figure 4 and Table S5. Among the selected three ligands, the L1 shows binding affinity –9.5 kcal/mol. In the TM-L1 complex, five hydrogen bonds, eight hydrophobic interactions stabilize the ligand. Three hydrogen bonds and nine hydrophobic interactions stabilize the TM-L2 complex where the TM-L6 complex shows four hydrogen bonds, eight hydrophobic and one electrostatic interaction. Two hydrophobic interactions and one electrostatic interaction are observed for the TM-NAM complex and it shows the lowest binding affinity –8.5 kcal/mol. In the TM-NAM complex, three hydrogen bonds (C–O...H–N) are detected with various amino acids including ASP435 (2.114), TRP461 (2.777), and GLY464 (2.314).

Interestingly, it is observed that the ligands only show hydrogen bonds with the binding site residue ASP435, and GLY464 (Figure 5A). All ligands show important hydrophobic interactions with the active residues. More specifically, the best 3 ligands and NAM form important non-covalent bonds with common residues (TRP461 and GLY462 through pi-Alkyl interaction), and while LYS342, LEU 419 shows Pi- amide interaction (Figure 5B).

### 3.3. Molecular dynamics simulations and free energy calculations

250 ns molecular dynamics simulation was performed for three protonated TM-ligand and TM-NAM complexes. Nafamostat complex (TM-NAM) is used as a control. To recognize the stability of the TM-ligand complexes, the RMSDs of the Cα atoms for the protein were measured and plotted. The protein's activity during the simulation time is depicted in Figure 6 and Table 3.

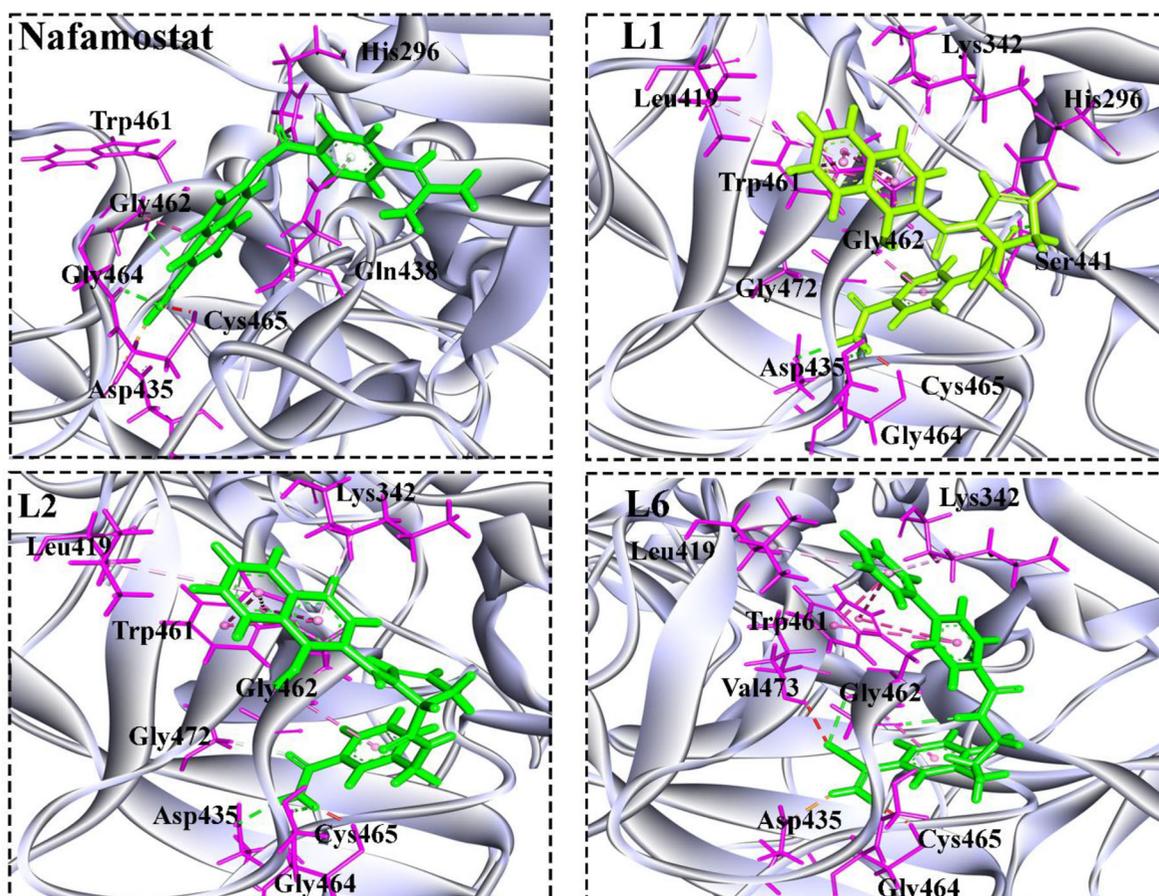


Figure 4. Non-bonding interactions of top three ligands with TM (Pose predicted by AutoDock Vina).

Table 5. Pharmacokinetic properties of selected three ligands.

Drug candidate	MF	AlogP	hEGI	Caco <sup>-2</sup> (cm/s)	BBB	CYP2D6 inhibition	HIA	PGI	Carcinogenicity (binary)	Acute oral toxicity
L1	C <sub>24</sub> H <sub>23</sub> N <sub>3</sub> O <sub>4</sub>	2.57	-0.481	-0.839	+0.886	-0.847	+0.962	+0.768	-0.903	III
L2	C <sub>24</sub> H <sub>23</sub> N <sub>3</sub> O <sub>4</sub>	2.57	-0.441	-0.833	+0.886	-0.847	+0.962	+0.780	-0.903	III
L6	C <sub>23</sub> H <sub>23</sub> N <sub>3</sub> O	4.00	+0.929	-0.729	+0.990	-0.848	+0.885	+0.844	-0.914	III
NAM	C <sub>19</sub> H <sub>17</sub> N <sub>5</sub> O <sub>2</sub>	2.25	-0.438	-0.697	+0.967	-0.880	+0.974	-0.528	-0.745	III

MF-molecular formula; BBB- blood brain barrier; HIA- human intestinal absorption; PGI- p-glycoprotein inhibitor; hEGI- Human either-a-go-go inhibition.

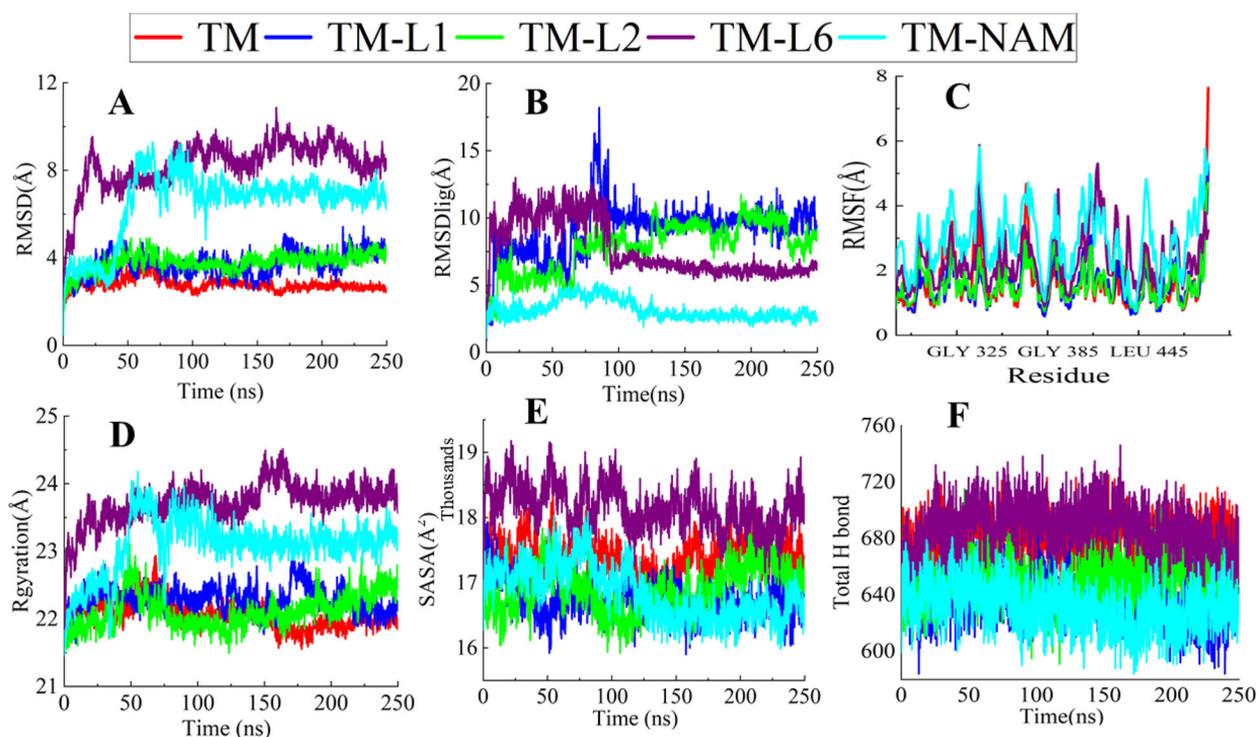
A								B				
Name	HIS296	ASP435	SER441	GLY464	GLY472	TRP461	GLY462	Name	LYS342	LEU419	TRP461	GLY462
TM-L1	Green	Orange	Blue	Yellow	Red			TM-L1	Green	Orange	Orange	Orange
TM-L2		Orange		Yellow	Red			TM-L2	Green	Orange	Orange	Orange
TM-L6		Orange		Yellow			Orange	TM-L6	Green	Orange	Orange	Orange
TM-NAM		Orange		Yellow				TM-NAM			Orange	Orange

Figure 5. Common (A) H- bond (B) Hydrophobic interaction for selected compounds. Here, white color means no residues.

In Figure 6A, RMSDs of apo-form are reported and found lowest of any other complexes (average RMSD 2.79 Å). Lower RMSDs show greater stability. A higher fluctuation of RMSD value is observed in the **TM-L6** complex in comparison to other complexes. The average RMSD value of TM-L6 is 8.26 Å. The average RMSD of TM-L2 is 3.81 Å, and little fluctuation is noticed in RMSD value, however, the ligand significantly changes its position from the initial binding pose over the simulation period (Figure S4A). It is also showed that the

average RMSD value of TM-L1 is 3.70 Å, and it fluctuates very little throughout the MD simulation period. The average RMSD value of TM-NAM is detected at 6.49 Å. Figure 6B shows the RMSDs of the ligand. The average RMSD of the NAM ligand (average RMSD 3.16 Å) shows the lowest value compared to the average RMSD of other ligands including L1 (9.09 Å), L2 (7.86 Å), and L6 (7.74 Å).

The RMSF of TM, TM-L1, TM-L2 and TM-L6 shows little fluctuation throughout the simulation period whereas TM-



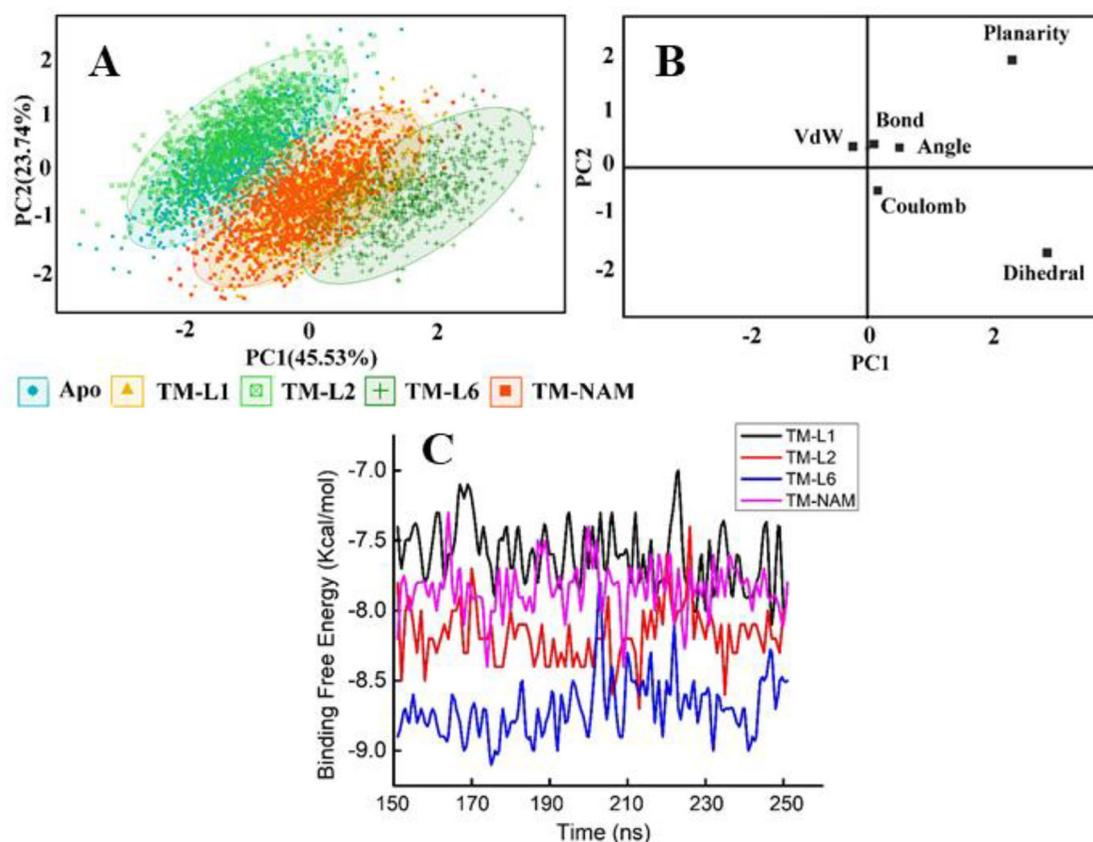
**Figure 6.** (A) Root-mean-square deviation (RMSD) of the  $C\alpha$  atoms, (B) Ligand RMSD of each docked complex, (C) RMSF values of the alpha carbon over the entire simulation, where the ordinate is RMSF ( $\text{\AA}$ ), and the abscissa is residue, (D) Radius of gyration (Rg) over the entire simulation, where the ordinate is Rg ( $\text{\AA}$ ), and the abscissa is time (ns) and (E) Solvent accessible surface area (SASA), where the ordinate is SASA ( $\text{\AA}^2$ ) and the abscissa is time (ns) throughout the simulation, (F) Total number of Hydrogen bonds formed, (G) Molecular Surface Area (MolSA) values of apoprotein in the complexes with selected ligands (TM-L1, TM-L2, and TM-L6) and TM-NAM along the 250 ns MD simulations.

NAM shows the higher fluctuation (Figure 6C). RMSF helps to assess the dynamic activity of amino acid residues in the protein. The flexibility of amino acid residues is measured to provide a deeper insight into the degree to which protein flexibility is influenced by ligand binding. The RMSF of residues (146–255) fluctuated substantially during the interaction between protein and ligands. The average RMSF value of apo-protein, TM-L1, TM-L2, TM-L6, and TM-NAM are 1.82  $\text{\AA}$ , 1.78  $\text{\AA}$ , 1.84  $\text{\AA}$ , 2.89  $\text{\AA}$ , 3.74  $\text{\AA}$  respectively. For the majority of residues, a small RMSF (within 1.84  $\text{\AA}$ ) in the TM-L2 complex is found, a high fluctuation of 2.74  $\text{\AA}$  is identified in the loop region adjacent to the residue 150, and a maximum fluctuation of 5.11  $\text{\AA}$  is noticed in the helix region of residues 220–230. Overall, RMSF of the residues, HIS296, ASP435, CYS437, GLN438, GLY439, TRP461, GLY462, GLY464, and CYS465 are lower for all ligand-bound complexes and apo-protein. It can also be said that the RMSF value of TM-L2 is lower than TM-NAM. The radius of gyration ( $R_g$ ) is used to measure protein structural compactness. The lower degree of fluctuation in the simulation with its consistency implies the increased compactness and rigidity of a system. In Figure 6D; the  $R_g$  of apo-protein (average  $R_g \sim 22.06$ ) is the most stable concerning the consistency of fluctuations in the simulation. The average  $R_g$  value of TM-L1, TM-L2, TM-L6, and TM-NAM is 22.29, 22.11, 23.72, 23.09, respectively. Often the binding of a small molecule could have a significant impact on the structure of the protein and could alter SASA. For all complexes and apo-protein almost equal value of SASA is reported (Figure 6E). The average SASA value of apo-protein, TM-L1, TM-L2, TM-L6, and TM-NAM are 17414  $\text{\AA}^2$ , 16757  $\text{\AA}^2$ ,

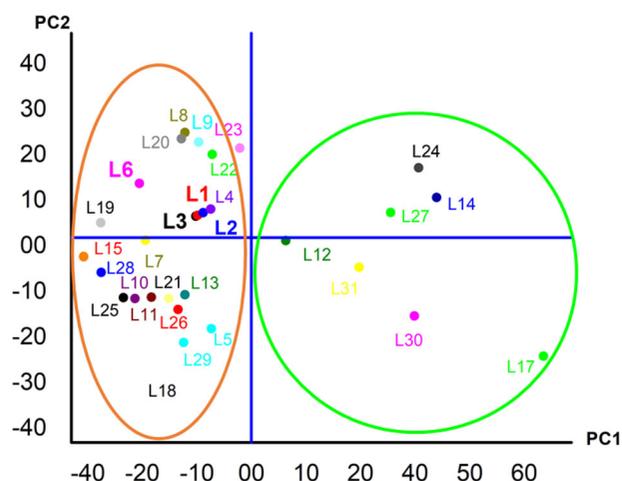
16890  $\text{\AA}^2$ , 18169  $\text{\AA}^2$ , 16853  $\text{\AA}^2$ , respectively. Therefore, it can be proposed that the binding of TM-L2 may decrease the expansion of protein. Intermolecular hydrogen bonds play a role in conformational stability. For all complexes, the numbers of hydrogen bonds are determined. In Figure 6F, the total number of hydrogen bonds is calculated for all complexes. The average total hydrogen bond numbers for apo-protein, TM-L1, TM-L2, TM-L6, and TM-NAM complexes are 680, 637, 650, 687, and 631 respectively. For apo-protein, the highest number of hydrogen bonds is reported. The total hydrogen bond numbers found for TM-L2 are nearly analogous to TM-NAM.

Over the MD simulation period, the hydrogen bond (HB) and hydrophobicity of the corresponding active residues are demonstrated in Figure S3 reveals that during simulation time ASP435 and GLY464 residues always interact with the ligand through HB. On the other side, SER436 and GLY462 amino acids also show the same interaction in MD simulation but exceptions are noticed at 20 ns for SER436 and 10 ns and 70 ns for GLY462 (Figure S4). The TRP461, CYS437, and GLN438 residues show hydrophobic behavior almost continuously during MD simulation time. The rest of the residues are also shown hydrophobicity but comparatively less and follow the discontinuity.

To evaluate the structural quality of protein-ligand complexes, principal component analysis (PCA) is utilized to explore the energy and structural information of four protein-ligand complexes along with apo obtained from MD simulation. The energy and structural information vary with some parameters including bond energies, bond angle



**Figure 7.** (A) The score plot presented five data clusters in a different color, where each dot represented a one-time point. (B) Loading plot from principal components analysis of the energy and structural data, and (C) Binding free energy.



**Figure 8.** Score plot of PCA regression of quantitative structure-activity relationship (QSAR) of ligands.

energies, dihedral angle energies, planarity energies, Van der Waals energies, and Columbic interaction energies are here considered as a variable. Different cluster formation on the PCA plot as shown in [Figure 7A](#) reveals the similarity and dissimilarity between apo and protein-ligand complexes. The PC1 and PC2 score plots are considered in this study. From [Figure 7A](#), it has been observed that the first principal component analysis (PC1) describes 45.53% of the variability and the second principal component (PC2) covers 23.74% of the variability. The scores plot shows that clusters of apo, L2 ligands are overlapped to each other to a great extent. In

addition, by maintaining a small difference from apo and L2 clusters, clusters of L1 and NAM are also overlapped. Interestingly, it is shown from the loading plot ([Figure 7B](#)) that bond energies, bond angle, VdW are found in the same distinct area where the clusters are formed on the scores plot ([Figure 7B](#)). This reveals that these variables may have influenced the formation of different clusters in the scores plot. However, the cluster of L6 shows a noticeable deviation from apo's cluster. During MD simulation both TM-L6 complex exhibit the fluctuating nature that could be responsible for wider distribution in the score plot. Both the TM-L1 and TM-L2 complexes are in the lowest deviation compared to other complexes, especially the TM-L2 complex approximately overlaps with the cluster of apoprotein. Hence, it can be considered as a potential candidate.

[Figure 7C](#) illustrates the binding free energy of all the ligands which are obtained from the prodigy-ligand web server. Here, the TM-L6 complex shows greater stability (around  $-8.9$  kcal/mol) than any other complex.

### 3.4. Quantitative structure-activity relationships (QSAR)

Generally, Quantitative structure-activity relationships studies have been broadly used popular theoretical method for modelling and pattern recognition analysis in drug design for the pharmaceutical industry, toxicology prediction of drug, clinical research, bioinformatics, petrochemical, food, and agrochemicals industry (Alam & Khan, 2017; Berhanu et al., 2012; Luco & Ferretti, 1997; Misra et al., 2017). To get

more insight into the correlation of molecular structure with biological and pharmaceutical activities, multiple linear regression (MLR) has been employed for further analysis. For developing multiple linear regression (MLR), all QSARs of ligands are significant. QSAR study on thirty-one ligands shows that molecular weight, TPSA ( $\text{\AA}^2$ ), XLogP3, H-bond donor count, and H-bond are vital QSAR contributors to the MLR listed in Table S6 and these may play an important role in binding affinity and non-covalent interaction against the TM.

Principal component analysis (PCA) is used for revealing the pattern recognition of potential ligands [Figure 8](#). The first principal component (F1) describes 57.32% of the variability in ligand QSAR. The second principal component (F2) reveals 15.97% in QSAR variability of ligands.

By close observation of the score plot, an interesting grouping is noticed. The ligands which have identical functional groups and rings are assembled on the score plot.

Ligands L1, L2 and L6 show more binding affinity compared to NAM from molecular docking results. Thus, three ligands can be used to design a potential drug to target TM. The ligands L1, L2, L4, L5, L6, L10, L13, L16, L21, and L26 comprise some important groups and rings listed as amide and formamidine groups as well naphthalene, biphenyl, and pyrrolidine rings. These groups and rings are themselves bonded together through a single bond and connected to the benzene ring, hence the ligands are positioned together on the first and second quadrants of the QSAR pattern. It can be concluded that their position in the same quadrants is due to having structural similarities. Therefore, the binding energies are also observed in the range of  $-8.8$  and  $-9.8$  kcal/mol. It is explored that L1, L2, and L6 give the best performance against the targeted protein that supports the generated QSAR pattern. On the other side, L12, L14, L17, L24, L27, L30, and L31 are containing formamidine, diphenylmethyl and propylguanidine groups along with naphthalene, biphenyl, pyrrolidine, and benzothiazole rings. These chemical fragments are attached to the secondary amide group in the corresponding ligand structures and the ligands are clustered together on the third and fourth quadrants for their structural similarities. QSAR pattern reveals that L17 is quite different from the others located on the first quadrant which is compatible with the lower binding affinity obtained from molecular docking. The reason can be due to the presence of more branching in the ligand structure resulting in a larger size, therefore, the interaction between ligand and protein becomes less as ligand size does not fit the binding pocket of the targeted protein. In contrast, L14, L27, and L24 give comparatively more binding energy and lie in the same quadrant of the QSAR pattern. For instance, the presence of dioxosulfide and sulfur groups might enhance the possibility of more interactions. However, L6 is analyzed as the best candidate from molecular dynamics. The reason might be having of formamidine group which contains two nitrogen atoms and one double bond; hence, the density of electrons can be higher and can influence the physical and chemical properties.

The main target of any MLR model is to predict the binding energy of future drug candidates. The training set is used to build an MLR model and therefore it is used to predict the binding energy of the test set for validation of drug candidates. The validation result shows similar binding energy obtained from molecular docking.

$$\begin{aligned} \text{BindingAffinity} = & -9.35574 + 0.04011 \times n_N \\ & - 0.24726 \times n_{\text{double\_bond}} \\ & - 0.15161 \times n_{\text{single\_bond}} \\ & - 0.00575 \times \text{TPSA}(\text{\AA}^2) \\ & + 0.01638 \times \text{MW}(\text{g/mol}) \end{aligned} \quad (2)$$

The binding energy obtained from molecular docking and the predicted binding energy by MLR regression for each ligand and of validation ligands are exhibited in [Table 4](#). The capability of MLR regression to predict accurately the binding energy of the validated ligands is estimated by root-mean-square-relative percent errors (*RMS%RE*). The MLR regression revealed a high accuracy of 80% for the prediction of the binding energy of ligands compared with the binding energy obtained from molecular docking.

### 3.5. Pharmacokinetic properties

Pharmacokinetic properties (absorption, distribution, metabolism, excretion, and toxicity) of the top-ranked three ligands are listed in [Table 5](#) and the rest are in [Table S7](#) by using the admetSAR server (<http://lmmd.ecust.edu.cn/admetSar2/>). Recently, drug discovery emphasizes pharmacokinetically good drugs. Structural optimization is an essential computational feature in the way to design a new drug. It is very important to optimize the pharmacokinetic parameters to find out a promising drug agent that can pass standard clinical trial criteria.

Following the calculation of admetSAR, the selected three ligands are found to be non-carcinogenic as well as show good human intestinal absorption (HIA) which is a major step for transporting drugs to their targets. Again, the pharmacokinetics and pharmacodynamics profile of a drug is greatly affected by a drug transporter called Phosphorylated glycoprotein (P-gp) and also the absorption and bioavailability of a drug are boosted by the P-gp inhibitors. The ligands might have the potential to interact with p-glycoprotein which is characterized as ABC transporter. On the other hand, blood-brain barrier (BBB) is a serious obstacle to drug distribution to the central nervous system (CNS). Approximately the square route of molecular weight is inversely related to the BBB. The drugs having higher molecular weight cross the BBB sufficiently to affect the central nervous system (CNS) (Banks, 2009). From the calculation, it has been found that the selected three ligands L1, L2, and L6 have relatively good BBB penetration probability and thus may not affect CNS function. From the observed data, it can be predicted that the selected ligands may be safe for humans.

## 4. Conclusions

In this study, molecular docking, molecular dynamics, QSAR analysis, PCA, and ADMET tools are employed to detect potent ligands for TMPRSS2, analyzing 7476 ligands from the zinc database. A strong binding affinity is obtained for three ligands (L1, L2 and L6). Significant number of non-covalent interactions such as hydrogen bonding, hydrophobic and electrostatic interactions are detected. MD simulation is conducted to verify the stability of ligands, and found to be more stable, making more hydrogen bond interactions with TMPRSS-2, than the control NAM ligand. Here, L1, L2, L4, L5, L6, L10, L13, L16, L21, and L26 ligands contain amide as well as formamidine groups, naphthalene, biphenyl, and pyrrolidine rings. These groups and rings are themselves bonded together through a single bond and connected to the benzene ring, hence the ligands are positioned together on the first and second quadrants. On the other side, L12, L14, L17, L24, L27, L30, and L31 ligands have formamidine, diphenylmethyl and propylguanidine groups along with naphthalene, biphenyl, pyrrolidine, and benzothiazole rings. These chemical fragments are attached to the secondary amide group in the corresponding ligand structures and therefore, the ligands are clustered together on the third and fourth quadrants. The capability of MLR regression to predict accurately the binding energy of the validated ligands is measured by root-mean-square-relative percent errors (RMS%RE). The MLR regression revealed a high accuracy of 80% for the prediction of the binding energy of ligands compared with the binding energy obtained from molecular docking. It can be concluded that most of the selected ligands have shown promise and can be used to design effective antiviral drugs against the SARS-CoV-2.

## Acknowledgements

We are grateful to our donors who supported to build a computational platform in Bangladesh <http://grc-bd.org/donate/>.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The authors also like to acknowledge The World Academy of Science (TWAS) for providing funding (17-479 RG/CHE/) for this project.

## References

- 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), 1-13. <https://doi.org/10.1038/s41598-017-06131-0>
- Anandakrishnan, R., Aguilar, B., & Onufriev, A. V. (2012). H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, 40(W1), W537-W541. <https://doi.org/10.1093/NAR/GKS375>
- Apweiler, R. (2009). The universal protein resource (UniProt) in 2010. *Nucleic Acids Research*, 38(Suppl. 1), 142-148. <https://doi.org/10.1093/nar/gkp846>
- Artemova, S., Jaillet, L., & Redon, S. (2016). Automatic molecular structure perception for the universal force field. *Journal of Computational Chemistry*, 37(13), 1191-1205. <https://doi.org/10.1002/jcc.24309>
- Banks, W. A. (2009). Characteristics of compounds that cross the blood-brain barrier. *BMC Neurology*, 9(Suppl. 1), 5-9. <https://doi.org/10.1186/1471-2377-9-S1-S3>
- Berhanu, W. M., Pillai, G. G., Oliferenko, A. A., & Katritzky, A. R. (2012). Quantitative structure-activity/property relationships: The ubiquitous links between cause and effect. *ChemPlusChem*, 77(7), 507-517. <https://doi.org/10.1002/cplu.201200038>
- Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., & Zardecki, C. (2002). The protein data bank. *Acta Crystallographica. Section D, Biological Crystallography*, 58(Pt 6 No 1), 899-907. <https://doi.org/10.1107/s0907444902003451>
- Bertram, S., Dijkman, R., Habjan, M., Heurich, A., Gierer, S., Glowacka, I., Welsch, K., Winkler, M., Schneider, H., Hofmann-Winkler, H., Thiel, V., & Pohlmann, S. (2013). TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. *Journal of Virology*, 87(11), 6150-6160. <https://doi.org/10.1128/JVI.03372-12>
- Bertram, S., Heurich, A., Lavender, H., Gierer, S., Danisch, S., Perin, P., Lucas, J. M., Nelson, P. S., Pöhlmann, S., & Soilleux, E. J. (2012). Influenza and SARS-coronavirus activating proteases TMPRSS2 and HAT are expressed at multiple sites in human respiratory and gastrointestinal tracts. *PLoS One*, 7(4), e35876. <https://doi.org/10.1371/journal.pone.0035876>
- Bestle, D., Heindl, M. R., Limburg, H., Pilgram, O., Moulton, H., Stein, D. A., Harges, K., Eickmann, M., Dolnik, O., & Rohde, C. (2020). TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Science Alliance*, 3(9), e202000786. <https://doi.org/10.26508/lsa.202000786>
- Darden, T., Perera, L., Li, L., & Lee, P. (1999). New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure*, 7(3), R55-R60. [https://doi.org/10.1016/S0969-2126\(99\)80033-1](https://doi.org/10.1016/S0969-2126(99)80033-1)
- Dickson, C. J., Madej, B. D., Skjerve, 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>
- Dirar, A. I., Waddad, A. Y., Mohamed, M. A., Mohamed, M. S., Osman, W. J., Mohammed, M. S., Elbadawi, M. A. A., & Hamdoun, S. (2016). In silico pharmacokinetics and molecular docking of three leads isolated from *Tarconanthus camphoratus* L. *International Journal of Pharmacy and Pharmaceutical Sciences*, 8(5), 71-77.
- Fakayode, S. O., Busch, K. W., & Busch, M. A. (2009). Chemometric approach to chiral recognition using molecular spectroscopy. In *Encyclopedia of analytical chemistry*. John Wiley & Sons, Ltd. <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>
- Glowacka, I., Bertram, S., Muller, M. A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T. S., He, Y., Gnirss, K., Niemeyer, D., Schneider, H., Drosten, C., & Pohlmann, S. (2011). Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *Journal of Virology*, 85(9), 4122-4134. <https://doi.org/10.1128/JVI.02232-10>
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., & Ziebuhr, J. (2020). The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV

- and naming it SARS-CoV-2. *Nature Microbiology*, 5(4), 536–544. <https://doi.org/10.1038/s41564-020-0695-z>
- Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., & Onufriev, A. (2005). H<sup>+</sup>: A server for estimating pK<sub>a</sub>s and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, 33(SUPPL. 2), W368–W371. <https://doi.org/10.1093/nar/gki464>
- Gunasekaran, K., Ramakrishnan, C., & Balam, P. (1996). Disallowed Ramachandran conformations of amino acid residues in protein structures. *Journal of Molecular Biology*, 264(1), 191–198. <https://doi.org/10.1006/jmbi.1996.0633>
- Hempel, T., Raich, L., Olsson, S., Azouz, N. P., Klingler, A. M., Hoffmann, M., Pöhlmann, S., Rothenberg, M. E., & Noé, F. (2021). Molecular mechanism of inhibiting the SARS-CoV-2 cell entry facilitator TMPRSS2 with camostat and nafamostat. *Chemical Science*, 12(3), 983–992. <https://doi.org/10.1039/D0SC05064D>
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2), 271–280.e8. <https://doi.org/10.1016/j.cell.2020.02.052>
- Hoffmann, M., Schroeder, S., Kleine-Weber, H., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). Nafamostat mesylate blocks activation of SARS-CoV-2: New treatment option for COVID-19. *Antimicrobial agents and chemotherapy*, 64(6), e00754-20. <https://doi.org/10.1128/AAC.00754-20>
- Hussain, M., Jabeen, N., Amanullah, A., Baig, A. A., Aziz, B., Shabbir, S., Raza, F., & Uddin, N. (2020). Molecular docking between human TMPRSS2 and SARS-CoV-2 spike protein: Conformation and intermolecular interactions. *AIMS Microbiology*, 6(3), 350–360. <https://doi.org/10.3934/microbiol.2020021>
- Irwin, J. J., Gaskins, G., Sterling, T., Mysinger, M. M., & Keiser, M. J. (2018). Predicted biological activity of purchasable chemical space. *Journal of Chemical Information and Modeling*, 58(1), 148–164. <https://doi.org/10.1021/acs.jcim.7b00316>
- Islam, J., Khan, A. M., Parves, R., & Hossain, N. (2019). Prediction of deleterious non-synonymous SNPs of human STK11 gene by combining algorithms, molecular docking, and molecular dynamics simulation. *Scientific Reports*, 1, 16. <https://doi.org/10.1038/s41598-019-52308-0>
- Islam, R., Parves, M. R., Paul, A. S., Uddin, N., Rahman, M. S., Mamun, A., Al, Hossain, M. N., Ali, M. A., & Halim, M. A. (2020). A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, 39(9), 3213–3224. <https://doi.org/10.1080/07391102.2020.1761883>
- Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., & Nagata, N. (2019). TMPRSS2 contributes to virus spread and immunopathology in the airways of murine models after coronavirus infection. *Journal of Virology*, 93(6), 1–15. <https://doi.org/10.1128/JVI.01815-18>
- Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K., & Whittaker, G. R. (2020). Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop. *Journal of Molecular Biology*, 432(10), 3309–3325. <https://doi.org/10.1016/j.jmb.2020.04.009>
- 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>
- Kim, M. O., Nichols, S. E., Wang, Y., & McCammon, J. A. (2013). Effects of histidine protonation and rotameric states on virtual screening of *M. tuberculosis* RmlC. *Journal of Computer-Aided Molecular Design*, 27(3), 235–246. <https://doi.org/10.1007/s10822-013-9643-9>
- Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A., & Vriend, G. (2004). Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins*, 57(4), 678–683. <https://doi.org/10.1002/prot.20251>
- 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>
- Krieger, E., Nielsen, J. E., Spronk, C. A. E. M., & Vriend, G. (2006). Fast empirical pK<sub>a</sub> prediction by Ewald summation. *Journal of Molecular Graphics & Modelling*, 25(4), 481–486. <https://doi.org/10.1016/j.jmkgm.2006.02.009>
- Kumar, V., Dhanjal, J. K., Bhargava, P., Kaul, A., Wang, J., Zhang, H., Kaul, S. C., Wadhwa, R., & Sundar, D. (2020). Withanone and Withaferin-A are predicted to interact with transmembrane protease serine 2 (TMPRSS2) and block entry of SARS-CoV-2 into cells. *Journal of Biomolecular Structure and Dynamics*, 1–13. <https://doi.org/10.1080/07391102.2020.1775704>
- Laskowski, R. A., Chistyakov, V. V., & Thornton, J. M. (2005). PDBsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Research*, 33(Database issue), D266–D268. <https://doi.org/10.1093/nar/gki001>
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291. <https://doi.org/10.1107/S0021889892009944>
- 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>
- Luco, J. M., & Ferretti, F. H. (1997). QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives. *Journal of Chemical Information and Modeling*, 37(2), 392–401. <https://doi.org/10.1021/ci960487o>
- Martens, H., & Ni, T. (1992). *Multivariate calibration*. John Wiley & Sons.
- Meng, T., Cao, H., Zhang, H., Kang, Z., Xu, D., Gong, H., Wang, J., Li, Z., Cui, X., Xu, H., Wei, H., Pan, X., Zhu, R., Xiao, J., Zhou, W., Cheng, L., & Liu, J. (2020). The transmembrane serine protease inhibitors are potential antiviral drugs for 2019-nCoV targeting the insertion sequence-induced viral infectivity enhancement. *BioRxiv*. <https://doi.org/10.1101/2020.02.08.926006>
- Misra, S., Saini, M., Ojha, H., Sharma, D., & Sharma, K. (2017). Pharmacophore modelling, atom-based 3D-QSAR generation and virtual screening of molecules projected for mPGES-1 inhibitory activity. *SAR and QSAR in Environmental Research*, 28(1), 17–39. <https://doi.org/10.1080/1062936X.2016.1273971>
- Nisha, C. M., Kumar, A., Vimal, A., Bai, B. M., Pal, D., & Kumar, A. (2016). Docking and ADMET prediction of few GSK-3 inhibitors divulges 6-bromindirubin-3-oxime as a potential inhibitor. *Journal of Molecular Graphics & Modelling*, 65, 100–107. <https://doi.org/10.1016/j.jmkgm.2016.03.001>
- Onufriev, A. V., & Alexov, E. (2013). Protonation and pK changes in protein-ligand binding. *Quarterly Reviews of Biophysics*, 46(2), 181–209. <https://doi.org/10.1017/S0033583513000024>
- Petukh, M., Stefl, S., & Alexov, E. (2013). The role of protonation states in ligand-receptor recognition and binding. *Current Pharmaceutical Design*, 19(23), 4182–4190. <https://doi.org/10.2174/1381612811319230004>
- Shrimp, J. H., Kales, S. C., Sanderson, P. E., Simeonov, A., Shen, M., & Hall, M. D. (2020). An enzymatic TMPRSS2 assay for assessment of clinical candidates and discovery of inhibitors as potential treatment of COVID-19. *ACS Pharmacology & Translational Science*, 3(5), 997–1007. <https://doi.org/10.1021/acsptsci.0c00106>
- Singh, N., Decroly, E., Khatib, A.-M., & Villoutreix, B. O. (2020). Structure-based drug repositioning over the human TMPRSS2 protease domain: Search for chemical probes able to repress SARS-CoV-2 Spike protein cleavages. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences*, 153, 105495. <https://doi.org/10.1016/j.ejps.2020.105495>
- Sterling, T., & Irwin, J. J. (2015). ZINC 15-ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11), 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>
- Team, R. C. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Austria. Retrieved December 12 from <http://www.R-project.org/>.
- Trott, O., & Olson, A. J. (2010). 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> 19499576
- van Zundert, G. C. P., & Bonvin, A. M. J. J. (2014). Modeling protein–protein complexes using the HADDOCK webserver modeling protein complexes with HADDOCK. In *Protein structure prediction* (pp. 163–179). Springer. [https://doi.org/10.1007/978-1-4939-0366-5\\_12](https://doi.org/10.1007/978-1-4939-0366-5_12)
- Vangone, A., Schaarschmidt, J., Koukos, P., Geng, C., Citro, N., Trellet, M. E., Xue, L. C., & Bonvin, A. M. J. J. (2019). Large-scale prediction of binding affinity in protein-small ligand complexes: the PRODIGY-LIG web server. *Bioinformatics (Oxford, England)*, 35(9), 1585–1587. <https://doi.org/10.1093/bioinformatics/bty816>
- Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Velesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 181(2), 281–292. <https://doi.org/10.1016/j.cell.2020.02.058>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/nar/gky427>
- Wickham, H. (2008). Applied spatial data analysis with r. In *Applied Spatial Data Analysis with R*. Springer. <https://doi.org/10.1007/978-0-387-78171-6>
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(Suppl. 2), 407–410. <https://doi.org/10.1093/nar/gkm290>
- 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>
- Yamamoto, M., Matsuyama, S., Li, X., Takeda, M., Kawaguchi, Y., Inoue, J., & Matsuda, Z. (2016). Identification of nafamostat as a potent inhibitor of Middle East respiratory syndrome coronavirus S protein-mediated membrane fusion using the split-protein-based cell-cell fusion assay. *Antimicrobial Agents and Chemotherapy*, 60(11), 6532–6539. <https://doi.org/10.1128/AAC.01043-16>
- Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C. D., So, R. T. Y., Lv, H., Mok, C. K. P., & Wilson, I. A. (2020). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science (New York, N.Y.)*, 368(6491), 630–633. <https://doi.org/10.1101/2020.03.13.991570>
- 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  $\alpha$ -ketoamide inhibitors. *Science*, 368(6489), 409–412. <https://doi.org/10.1126/science.abb3405>
- Zhang, Y., & Skolnick, J. (2005). TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33(7), 2302–2309. <https://doi.org/10.1093/nar/gki524>
- Zorn, D., Lin, V. S. Y., Pruski, M., & Gordon, M. S. (2008). An interface between the universal force field and the effective fragment potential method. *The Journal of Physical Chemistry. B*, 112(40), 12753–12760. <https://doi.org/10.1021/jp8049729>