

Enhancing Virtual Screening Efficiency in the Discovery of Human TMPRSS2 Inhibitor for Combating COVID-19

Said A. H. Vuai

The University of Dodoma

Marcelina M. Ogedjo

The University of Dodoma

Isaac Onoka

The University of Dodoma

Mtabazi G. Sahini

The University of Dodoma

Daniel M Shadrack (✉ dmsadrack@gmail.com)

St. John's University of Tanzania

Research Article

Keywords: coronavirus disease 19 (COVID-19) pandemic , health and economic ,TMPRSS2, SARS-CoV-2
S

Posted Date: June 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-613294/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Enhancing virtual screening efficiency in the discovery of human TMPRSS2 inhibitor for combating COVID-19

Said A. H. Vuai¹, Marcelina M. Ogedjo¹, Isaac Onoka¹, Mtabazi G. Sahini¹, Daniel M Shadrack^{2†*}

¹Department of Chemistry, College of Natural and Mathematical Sciences, University of Dodoma, P.O. Box 338 Dodoma, Tanzania.

² Department of Chemistry, Faculty of Natural and Applied Sciences, St. John's University of Tanzania, P.O. Box 47 Dodoma, Tanzania

Abstract

As the coronavirus disease 19 (COVID-19) pandemic continues to pose a health and economic crisis worldwide, the quest for drugs and/or vaccines against the virus continues. The human transmembrane protease serines 2 (TMPRSS2) has attracted attention as a target for drug discovery, as inhibition of its catalytic reaction would result in the inactivation of the proteolytic cleavage of the SARS-CoV-2 S protein. As a result, the inactivation prevents viral cell entry to the host's cell. In this work, we screened and identified two potent molecules that interact and inhibit the catalytic reaction by using computational approaches. Two docking screening experiments were performed utilizing the crystal structure and holo ensemble structure obtained from molecular dynamics in bound form. There is enhancement and sensitivity of docking results to the holo ensemble as compared to the crystal structure. Compound **1** binds more stable than nafamostat by interacting with catalytic triad residue His296 and Ser441, thereby disrupting the already established hydrogen bond interaction. The stability of the ligand-TMPRSS2 complexes was studied by molecular dynamics simulation, and the binding energy was re-scored by using MM-PBSA binding free energy. The obtained compounds may serve as initial point towards the discovery of potent TMPRSS2 inhibitors upon further *in vivo* validation.

1 Introduction

The outbreak of coronavirus disease-19 (COVID-19) caused by a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has resulted in global health and economic crisis. SARS-CoV-2 is a single-stranded RNA virus belonging to the same family with SARS-CoV and the Middle East respiratory syndrome (MERS) [Zhu et al., 2021]. These viruses (SARS-CoV-2, SARS-CoV and MERS-CoV) have been identified as highly pathogenic human coronaviruses, with an ability to infect human hosts [Zhu et al., 2021]. The spread rate of SARS-CoV-2 has been reported to exceed those of MERS-CoV and SARS-CoV [Zhu et al., 2021]. As of 8 June, 2021 approximately 3,764,513 deaths with over 174,822,669 cases were confirmed (<https://www.worldometers.info/coronavirus/>). Until now, there is no effective treatment, although some drugs have been reported and vaccines developed [Shadrack et al., 2021, Zhu

*Correspondence: dmshadrack@gmail.com or mshadrack@sjut.ac.tz (DMS)

et al., 2021]. The long term effects and the associated side effects are yet to be established, thus, scientists across the globe are working to develop effective drugs and vaccines to combat the disease. Among the approaches, drug repurposing has emerged as a promising way to identify drugs, whereby some drugs such as hesperidin and chloroquine have been repurposed for COVID-19 treatment and are in clinical trials [Haggag et al., 2020, Hu et al., 2021]. Many other researchers have devoted efforts in searching for molecules that could block the SARS-CoV-2 entry to human cells.

The SARS-CoV-2 entry into the human cell is mediated by a transmembrane spike (S) glycoprotein [Hu et al., 2021]. Generally, the release of the viral RNA genome into host cells occurs after the S protein binds to the host receptor angiotensin-converting enzyme 2 (ACE2), and upon endocytosis the viral membrane fuses into the host via endosomal membrane [Hu et al., 2021]. The binding and fusion of the viral S protein is mediated and activated by the human proteases [Hu et al., 2021]. The cleavage sites S1/S2 and S' found at the boundary between S1 and S2 subunits are proteolytically cleaved by the proteases (transmembrane protease serine 2 (TMPRSS2), furin and cathepsin L and B) [Zhu et al., 2021]. Although both proteases have shown a potential in priming and activating the S protein, the TMPRSS2 plays an important role in activating the membrane functioning of the SARS-CoV-2 S protein [Hu et al., 2021]. Therefore, small molecules can be targeted to inhibit the proteolytic activity of TMPRSS2, hence blocking the viral cell membrane entrance to the host cell. Targeting the human TMPRSS2 has an advantage of overcoming the drug resistance, unlike targeting the viral proteins. Thus, in the effort to develop new anti-SARS-CoV-2 drugs, the TMPRSS2 is considered as a promising target.

Several research groups have considered TMPRSS2 as the molecular target for identifying small molecules that can bind and inhibit the proteolytic process [Hu et al., 2021, Chen et al., 2021, Huggins, 2020, Peiffer et al., 2021]. For example, nafamostat and camostat, are in clinical trials as inhibitors of TMPRSS2 [Zhu et al., 2021]. Efforts to screen libraries containing small molecules for inhibiting TMPRSS2 are underway in many groups. Hu and co-workers, using *in silico* approaches screened a library and identified potent small molecules [Hu et al., 2021]. In addition, Huggins 2020 reported approved small molecules targetting inhibition of TMPRSS2 *in silico* [Huggins, 2020]. These reports suggest that targeting TMPRSS2 has a great potential to fight against this pandemic.

Atomistic simulation methods have shown a great promise in drug discovery. Methods such as molecular docking are now widely used in screening large libraries of small molecules. However, docking algorithms are limited due to global protein flexibility. Incorporation of full flexible receptor in virtual screening has been suggested and showed improvement in docking score. Several groups have considered the relaxed complex scheme (RCS) docking based approach. In this approach, the apo or holo protein is subjected to molecular dynamics simulation and ensemble structures are extracted from the stable equilibrated structure and used for docking. In an effort to discover a more potent inhibitor of TMPRSS2, our group carried out a combination of *in silico* methods to screen potential candidates from NNPDB database using different screening methods based on the apo and holo ensembles. In this work we found that, holo ensemble structure improved docking score as compared to the crystal structure, the stability of the compound was established by using molecular dynamics and binding free energy calculations based on MM-PBSA.

2 Results

2.1 Screening of natural products

TMPRSS2 is a serine protease responsible for activating and mediating the S-protein of SARS-CoV-2 proteolytic cleavage for viral cell entry through ACE2. SARS-CoV-2 proteolytic cleavage processes result in the formation of S1/S2 and S' fragments which attach to the TMPRSS2 [Zhu et al., 2021, Hu et al., 2021]. The catalytic triad site residues; His296, Asp345 and Ser441 and substrate binding sites residues Asp435, Ser460 and Gly462 forms the human TMPRSS2 active site and are important for the SARS-COV-2 proteolytic cleavage and attachment. Inactivation of TMPRSS2 enzymatic activities requires small molecules that bind strongly to these key residues. To identify novel small molecule inhibitors of TMPRSS2, we screened a database of African medicinal compounds against the catalytic and substrate binding site using the crystal structure (PDB ID: 7MEQ). Screened compounds showed the distribution binding free energy range of -9.6 to -3.2 kcal/mol (see Fig. 1a). Nafamostat, an experimentally known inhibitor of TMPRSS2 was used as a standard and had a binding free energy of -8.1 kcal/mol. The observed binding free energy distribution using the crystal structure suggested potential the compounds that could strongly inhibit the catalytic proteolytic cleavage of TMPRSS2 to SARS-COV-2.

Although molecular docking is an indispensable tool for screening large libraries and predicting the binding modes, the docking algorithms have limitations on accommodating full (global) receptor flexibility, hence may result in generation of false positive lead molecules. To overcome the challenge of global protein flexibility in virtual screening, accommodating receptor flexibility is necessary in drug discovery. A relaxed complex scheme (RCS) employing apo or holo receptor structure has shown a great promise. Our previous experience [Shadrack et al., 2020] shows that holo ensemble improves the binding free energy as compared to the apo receptor ensembles. A similar observation was also reported by [Xu and Lill, 2011]. To perform virtual screening with a flexible receptor, we performed a molecular dynamics simulation of the TMPRSS2 bound with compound **1**, which ranked higher with binding free energy of -9.6 kcal/mol compared to -8.1 kcal/mol of nafamostat. In this work, the free energy was computed from the equilibrated structure using RMSD as a reaction coordinate. The ensemble structures corresponding to minimum free energy were extracted and subjected to docking calculation (Fig.S1).

It should be noted that, during free energy calculations, the first 10 ns were regarded as equilibration period and hence excluded from the analysis. Interestingly, during the holo ensemble screening, we noted two observations; first, the binding free energy generally improved in the holo ensemble as compared to the crystal structure. Fig. 1a shows a shift in normal distribution curve for the holo structure with an improved docking score to -10.4 kcal/mol. Another feature observed was that, some compounds which ranked higher in the crystal structure ranked least during the ensemble screening. This suggests that, accommodating receptor flexibility is necessary in virtual screening which can ultimately help in eliminating false positive binders. For example, compound **1** ranked higher in the crystal structure screening with binding free energy of -9.6 kcal/mol. However, during the RSC, the binding free energy decreased to -7.6 kcal/mol. Similarly, compound **2** with binding free energy of -8.6 kcal/mol in the ensemble structure ranked higher in the holo ensemble screening with binding free energy of -10.6 (Fig 1b). These observations suggest a sensitivity of docking results to the crystal and holo ensemble structure. Generally, the holo ensemble structure improves the docking score as compared to the crystal structure.

To aid our understanding on the inhibition mechanism of nafamostat and compound **1** and **2** which ranked higher during the docking calculations, in Figure 1c-e we show the binding modes of the three molecules inside the TMPRSS2. It is interesting to observe that the guanidinium groups of nafamostat, the sugar moiety of compound **1** and anthraquinone moiety of compound

2 (Fig. 1f) all bound to the substrate site by interacting with Gly462 and Trp461 for compound **2** as shown in Figure 1c. Although the three molecules possess different chemical scaffolds, they showed similar binding modes to TMPRSS2, by interacting with residues at the catalytic triad and substrate site by forming hydrogen bonds with His296, Gly462 and Trp461. Compound **3** also interacted with Val280, in which a similar interaction observed to nafamostat was also reported by [Zhu et al., 2021]. A similar binding mode for nafamostat and compound **3** was observed where they interact with Val280, Thr393 and Gln438 and form hydrogen bonds; such interactions are not observed in compound **1** (Fig. S1). Although the RCS suggested an improvement in docking score to some molecules, the thermodynamic stability and dynamics of the complexes were investigated by means of unbiased classical molecular dynamics (MD) and end-point free energy based on MM-PBSA to ascertain the observed binding free energy.

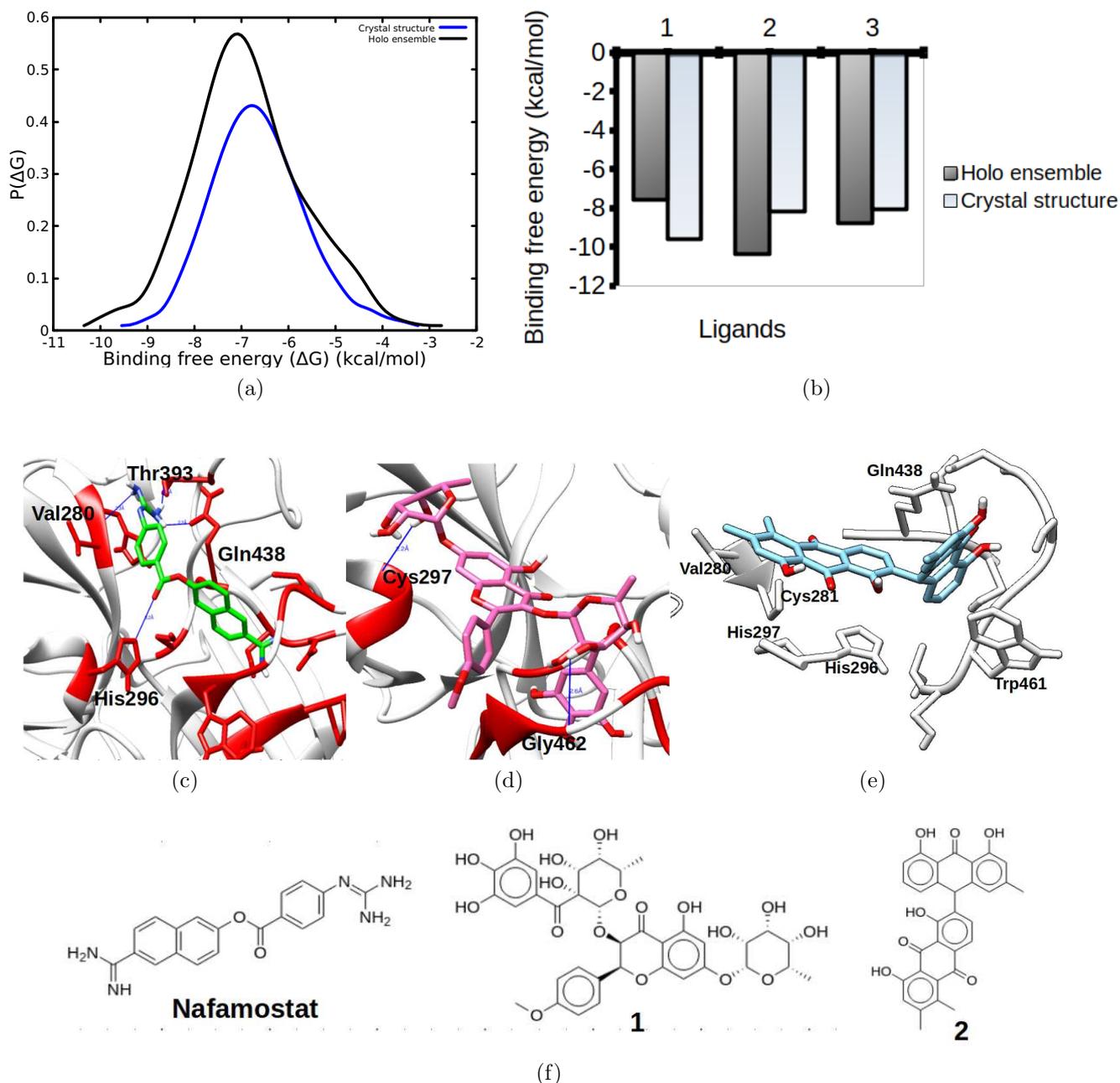


Figure 1: (a) Binding free energy distribution for crystal structure and holo ensemble structure, (b) sensitivity of binding free energy to crystal and holo ensemble. Binding modes of three molecules into the TMPRSS2 active site (c) nafamostat, (d) compound **1**, (e) compound **2**, (f) chemical structures of the selected ligands which ranked higher is crystal structure and holo ensemble.

2.2 Molecular dynamics simulation

Molecular dynamics simulation is an important tool in structure-based drug design, it provides dynamical and related stability of a complex. It should be noted that lower binding free energy obtained from docking calculations does not necessarily imply stability of the complex. In addition, end-point free energy methods are very useful in re-scoring poses obtained from docking calculations. Nafamostat, a known TMPRSS2 inhibitor and compound **1** which ranked the least in holo and higher in crystal structure docking calculations were selected for MD calculations to establish the time-lapse stability.

The general stability of the complexes was attained by measuring the C α root mean square deviation (RMSD) for TMPRSS2 bound with compound **1** and nafamostat. The time dependent RMSD value shows that TMPRSS2-compound **1** complex quickly equilibrated within 10 ns and remained stable with few noticeable fluctuations (Fig. S2). However, TMPRSS2-nafamostat complex equilibrated during the 40 ns and remained stable with few oscillations. The complex formed by TMPRSS2-compound **1** showed an average C α RMSD value of ≈ 0.23 nm and TMPRSS2-nafamostat showed slightly higher average C α RMSD value of ≈ 3.0 nm, the observed C α RMSD suggests that the complex formed by compound **1** is most stable as compared to nafamostat complex (Fig. 2a). The longer equilibration in TMPRSS2-nafamostat complex is due to the conformation orientation and pose changes of nafamostat which occurred during the first 20 ns as confirmed by the ligand pose RMSD (Fig.2b). Such orientations are not observed in TMPRSS2-compound **1** complex. As shown in Fig. S3a,b, the observed stability in compound **1** was enhanced by several hydrogen bonds formed between the residues mediated with interfacial water molecules at the vicinity of the catalytic triad. The number of hydrogen bonds formed between compound **1** are higher than that of nafamostat.

It is important to note that although complex RMSD is useful in assessing the general stability, it does not provide an in-depth understanding on ligand dynamics (pose changes) and stability. To understand whether compound **1** and nafamostat changed from their initial docking pose (association/dissociation), we computed the pose ligand RMSD which predicts drug efficacy *in vitro* or *in vivo*. The time dependent pose RMSD suggests that compound **1** remained in its initial docking pose with few noticeable unbinding (dissociation/association) processes observed from 65 ns (Fig2b). It is important to mention that even when compound **1** dissociated and diffused into solvents with the RMSD distance ≥ 8 nm, it eventually returned to the catalytic triad site (Fig. 2b). Contrary to compound **1**, nafamostat displayed a different binding behavior, (Fig. 2b), in which, during the first 5 ns, it remained in its docking pose then moved to occupy a different pose (Fig.2d). The changes in nafamostat binding site suggests an existence of transient binding pockets in TMPRSS2 which would result in an allosteric inhibition, however, this calls for further exploration. To gain a better understanding on the dynamical stability of compound **1** and nafamostat, the free energy as the function of ligand RMSD and ligand pose RMSD was calculated (Fig. 2c,d). It is observed that, compound **1** although exhibited some dissociation, remained stable as shown in Figure 2c. However, the binding of nafamostat exhibited two RMSD values with the most stable conformation at pose 2 (Fig. 2d). The observed change in binding pose is further confirmed by the distance between His296-nafamostat which formed hydrogen bond in pose 1 (Fig. 2e). Our results are somehow different from the reported finding by [Zhu et al., 2021], who reported the RMSD value for nafamostat to be 0.13 nm. However, in our work we observed two conformational states with RMSD at 0.1 nm and 0.17 nm which still suggests stability. The RMSD value for compound **1** was found to be 0.22 nm, suggesting that all the drugs bind with stability to the TMPRSS2 region.

To understand the backbone stabilities and fluctuations of the two complexes, root mean square fluctuations (RMSF) was computed (Fig. 2f). Two regions with residue 213-224 and 385-395 exhibited high fluctuations in the nafamostat complex. Such fluctuations are not pronounced in the complex formed with compound **1**. This suggests that the binding of compound **1** to TMPRSS2 resulted in structural stabilities unlike nafamostat. The region with residue 460-466 also showed higher fluctuation when bound with nafamostat. Such fluctuations at this region are not observed in the apo and TMPRSS2-compound **1** complex. The observed fluctuations correspond to loops. Generally, the backbone of TMPRSS-compound **1** exhibited lower fluctuations throughout the simulation as compared to the complex formed by TMPRSS2-nafamostat. The compactness of the complexes was assessed by measuring the radius of gyration (Rg) (Fig S4). It is observed that, all the protein showed preservation of

the compactness with an average Rg value below 2.2 nm. Some fluctuations near 65 ns are observed for TMPRSS2-compound **1** which are explained by the dissociation of compound **1**.

The observed dissociation and association of compound **1** could be attributed to; first, the thermal motion of the solvents, when compound **1** dissociated, interacted with solvent molecules until it was attracted by the residues at the TMPRSS2 surface. Secondly, the surface and the inner region of the catalytic region is made up of several negatively charged residues Glu/Asp, the positively charged groups of compound **1** were electrostatically attracted to the vicinity of TMPRSS2 binding substrate and eventually to the catalytic region where stabilizes by forming several hydrogen bond. The observed changes in pose for nafamostat are explained by different conformational and positional changes in the catalytic site before adapting to the final pose. The hydrogen bonds formed at the region stabilized the ligands, hence causing critical disturbances in the already established hydrogen bonds between the catalytic residues His296 and Ser441; a similar observation for nafamostat was reported by [Zhu et al., 2021].

The two residues His296 and Ser441 play an important role in the catalytic reaction for proteolytic cleavage of the S-Protein (Fig. 3a). Any disturbances of these residues would result in the inactivation of the TMPRSS2 catalytic process, hence stopping the virus cell entry. As shown in Fig. 3a, the crystal structure shows that the catalytic triad residues are in the order of hydrogen bond distances. The binding of compound **1** has the role in disturbing the already established His296-Ser441 hydrogen bonding by sitting on the top of Ser441 and forming a π -stacking interaction, thereby inducing a steric hindrance resulting to an increased distance (Fig. 3b,c). However, nafamostat binds differently; it interacts with the loop and β -sheet forming the His296 thereby inducing disturbances in the recognizing Ser441 (Fig. 3d,e). Fig. 3b,d, shows that the binding compound **1** and nafamostat resulted in increased distances between His296 and Ser441 but with different magnitudes. As observed in the snapshot, the two molecules did not induce appreciable changes in Asp345-His296 distances (Fig. S5), suggesting that these compounds work by disturbing the His296-Ser441 interaction. The structural re-orientation of nafamostat resulted in interaction with His279 and Val280 which stabilizes by forming hydrogen bonds; such interactions were previously noted by [Zhu et al., 2021].

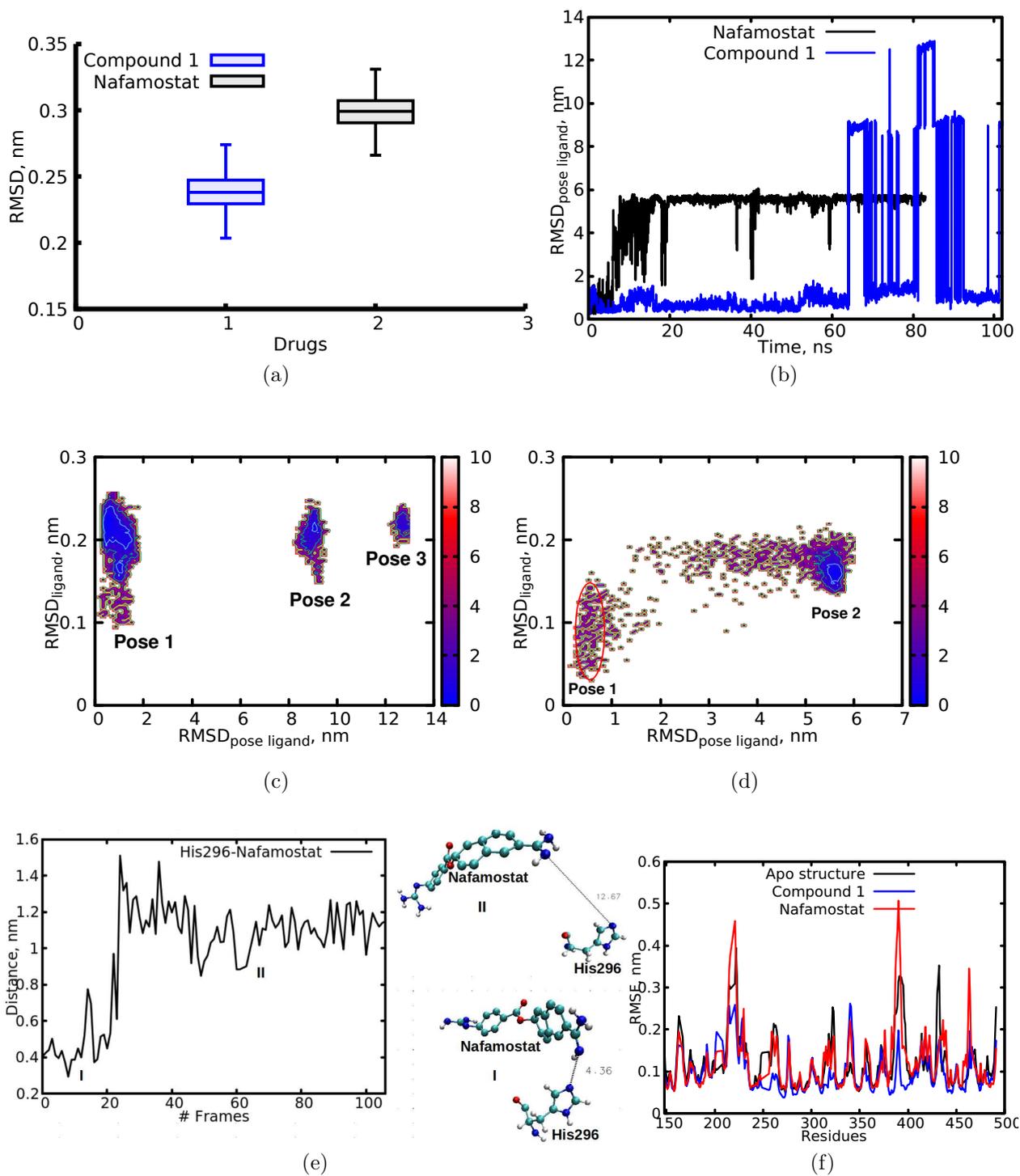
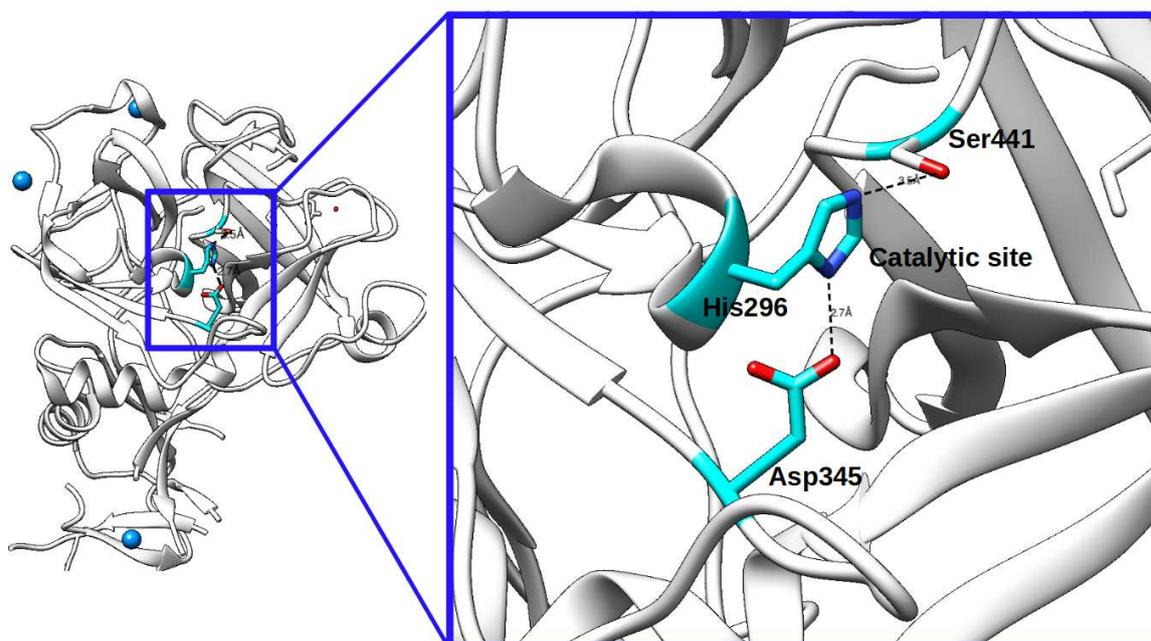
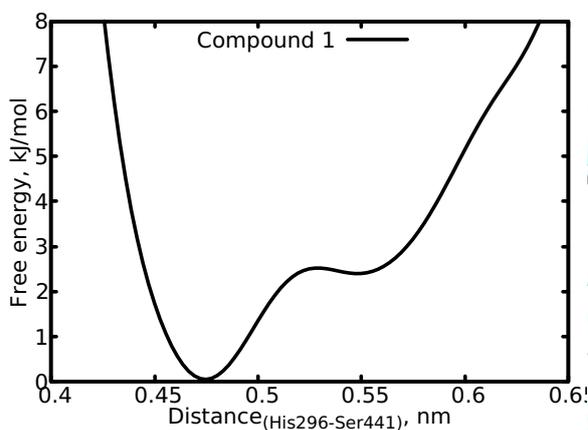


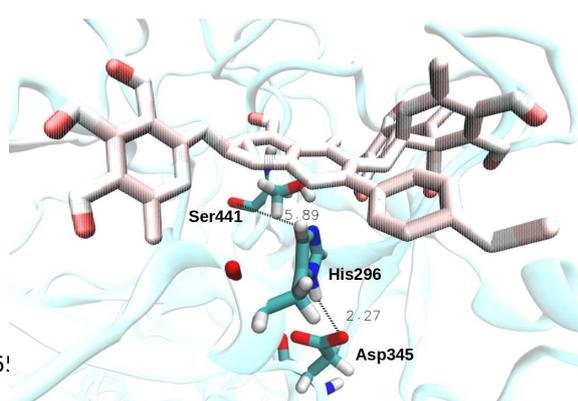
Figure 2: Measured stability of the two complexes: (a) RMSD values for compound 1 and nafamostat, (b) pose ligand RMSD for the two ligands. 2D free energy as the function of ligand RMSD and pose ligand RMSD for (c) compound 1 and (d) nafamostat. (e) RMSF for the two complexes.



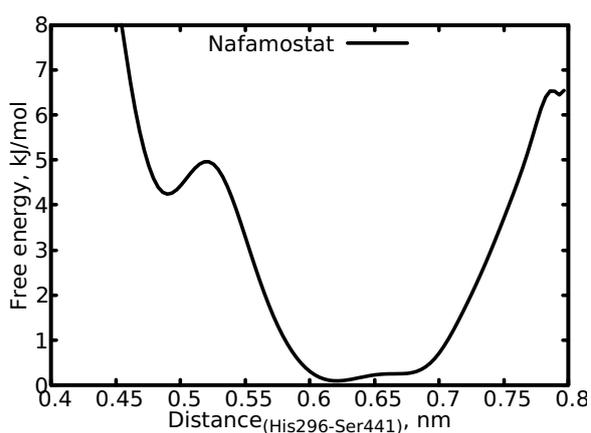
(a)



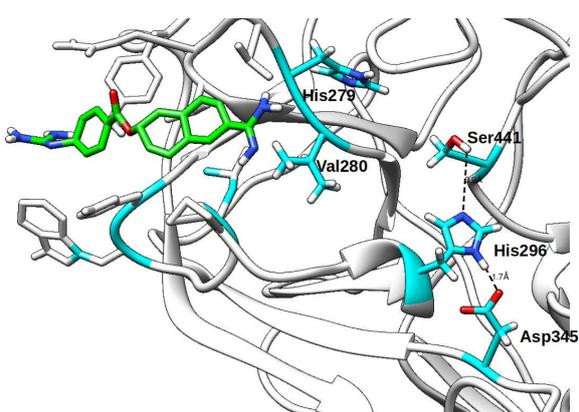
(b)



(c)



(d)



(e)

Figure 3: (a) Residues responsible for the catalytic process in TMPRSS2: (b,d) the distances between His296-Ser441 upon binding of compound 1 and nafamostat, (c,e) snapshot showing the interaction of the compound and change in distance.

2.3 MM-PBSA free energy calculation

The main driving forces responsible for ligand-TMPRSS2 stable interaction were further analysed by MM-PBSA free energy calculations. Table 1 shows that compound **1** binds more strongly than nafamostat; this is in agreement with the observed complex RMSD values (Fig 2a). As shown in Table 1, van der Waals forces are the main driving forces followed by non-polar energy terms. Electrostatic forces contributed least to the interaction while the polar forces resulted in an opposing effect on the ligand-TMPRSS2 interaction. The observed van der Waals forces contribution could be attributed to the group closeness between compound **1**/nafamostat and TMPRSS2 in the binding site.

A closer observation shows that, compound **1** had higher electrostatic contribution as compared to nafamostat. This could be attributed to the presence of many hydroxyl groups in compound **1** which interacts with the catalytic site of TMPRSS2. The free energy decomposition and per residue contribution for the two ligands are shown in Figure 4a,c. The catalytic residue His296, and substrate residues Val280 and Trp461 highly contributed to the interaction with compound **1**. It is observed that, the catalytic residue Ser441 contributed the least in the interaction as compared to His296, and, there is no any interaction contribution observed from Asp345 (Fig. 4b). The distances measured between His296-Asp345 (Fig. S5) also support the per residue decomposition and contribution to the binding free energy analysis. It seems that, compound **1** would work by strongly interacting with His296 and Ser441 thereby blocking their recognition. As mentioned earlier, nafamostat changed its binding pose, and still induced an appreciable distance between His296-Ser441. Per residue contribution and decomposition (Table 2,3) suggests that, nafamostat would bind strongly with the residues Val280, Val275, Leu302 and His307, thereby indirectly inducing disturbances in the His296-Ser441 interaction as shown in Fig 3d and 4d. Furthermore, as indicated in the two tables, the molecular mechanics energy term highly contributed to the stabilization of the complexes.

Table 1: Binding free energy (ΔG (kJ/mol)) calculated by MM-PBSA method

System	ΔE_{vdW}	ΔE_{ele}	ΔE_{polar}	$\Delta E_{\text{non-polar}}$	$\Delta E_{\text{binding}}$
Compound 1	-310.7 ± 23	-24.6 ± 10	158.5 ± 18	-76.2 ± 5	-253 ± 21
Nafamostat	-160.2 ± 16	-3.8 ± 3	48.9 ± 13	-31.2 ± 4	-146.4 ± 14

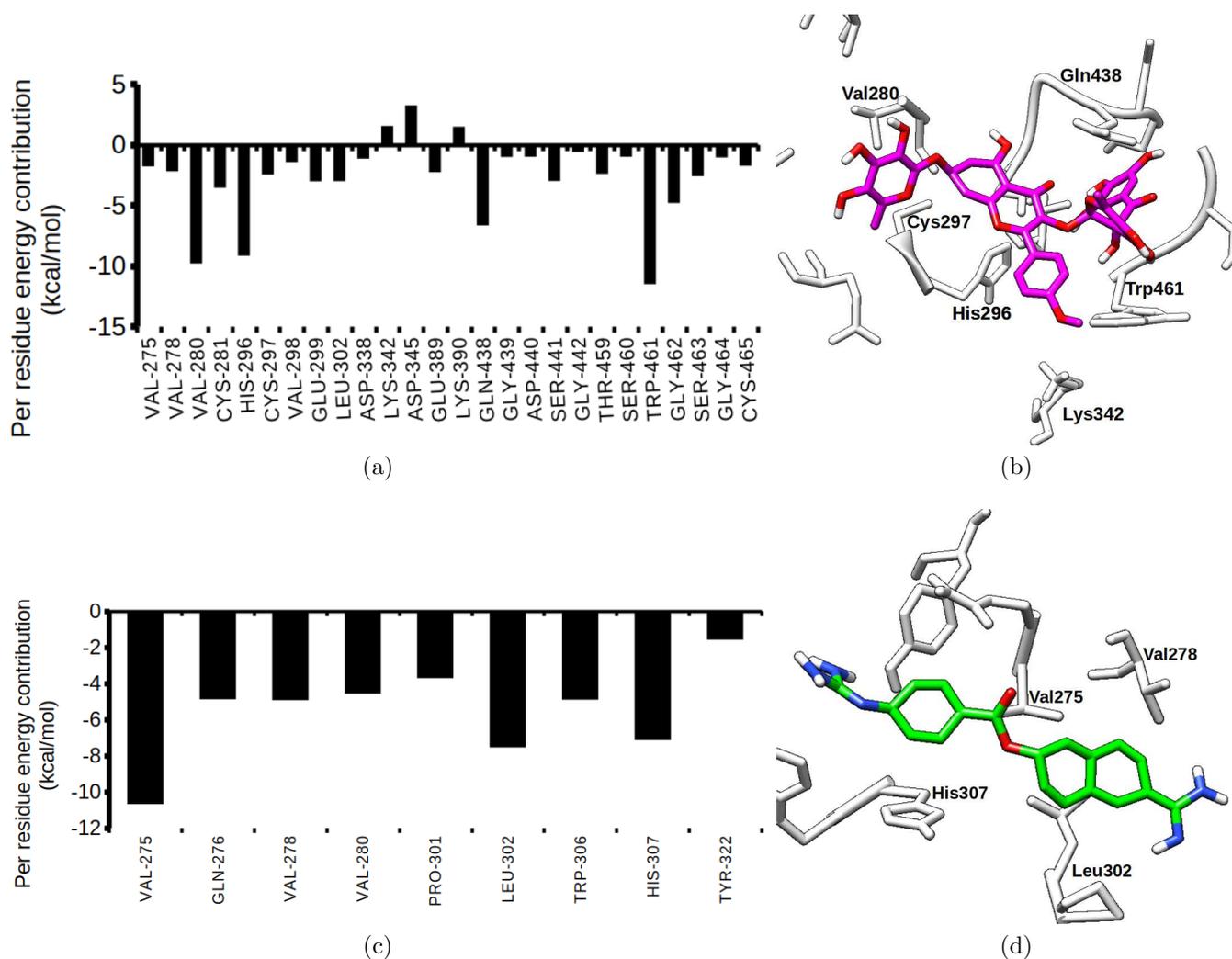


Figure 4: The per residue energy contribution to the binding of: (a) compound 1 and (c) nafamostat, (b,d) highly contributed residues to the binding free energy.

Table 2: The per residue decomposition contribution to binding free energy (ΔG (kJ/mol)) for compound 1

Residues	ΔE_{MM}	ΔE_{polar}	$\Delta E_{non-polar}$	$\Delta E_{binding}$
Val280	-9.9 ± 0.16	2.2 ± 0.08	-2.0 ± 0.02	-9.8 ± 0.14
Cys281	-4.4 ± 0.06	1.2 ± 0.02	-0.3 ± 0.01	-3.5 ± 0.06
His296	-21.2 ± 0.30	15.3 ± 0.2	-3.2 ± 0.04	-9.1 ± 0.30
Glu438	-20.7 ± 0.28	17.8 ± 0.24	-3.8 ± 0.04	-6.6 ± 0.25
Ser441	-6.1 ± 0.14	4.0 ± 0.15	-0.9 ± 0.02	-3.0 ± 0.12
Trp461	-16.7 ± 0.17	8.3 ± 0.09	-3.1 ± 0.05	-11.5 ± 0.22
Gly462	-10.5 ± 0.16	6.8 ± 0.09	-1.1 ± 0.02	-4.8 ± 0.17

Table 3: The per residue decomposition contribution to binding free energy (ΔG (kJ/mol)) for nafamostat

Residues	ΔE_{MM}	ΔE_{polar}	$\Delta E_{non-polar}$	$\Delta E_{binding}$
Val275	-11.2 ± 0.19	1.4 ± 0.03	-0.9 ± 0.02	-10.7 ± 0.18
Leu302	-6.3 ± 0.15	0.14 ± 0.02	-1.3 ± 0.03	-7.5 ± 0.17
His307	-13.1 ± 0.29	7.61 ± 0.21	-1.6 ± 0.04	-7.1 ± 0.23

Water is known to play an important role in protein-ligand interaction and stabilization. The role of water on the binding of compound **1** and nafamostat was also investigated to provide details on the observed binding stability and pose changes. During the first 5 ns of the simulation, it is observed that the head of nafamostat is directly bound at the interface of His296-Ser441. The two residues are observed to form a hydrogen bond which is surrounded by several water molecules; in which two vicinal water molecules are observed to mediate the interaction between nafamostat and protein by forming hydrogen bonds. The hydrogen bond between nafamostat with residues Lys342 and Gln438 are stabilized and mediated by water molecules. However, it is observed that water molecules formed stable interaction with the residues, thereby pushing nafamostat to another position, which is observed during the 40 ns (Fig 5a,b). The head of nafamostat has slightly moved from the catalytic site and interacts strongly with Val275 and with His307 as confirmed with MM-PBSA free energy decomposition. However, such movement is not observed for compound **1**. This is explained by the presence of several hydroxyl groups responsible for the formation of hydrogen bonds at the catalytic site mediated by water molecules (Fig.S3a,b).

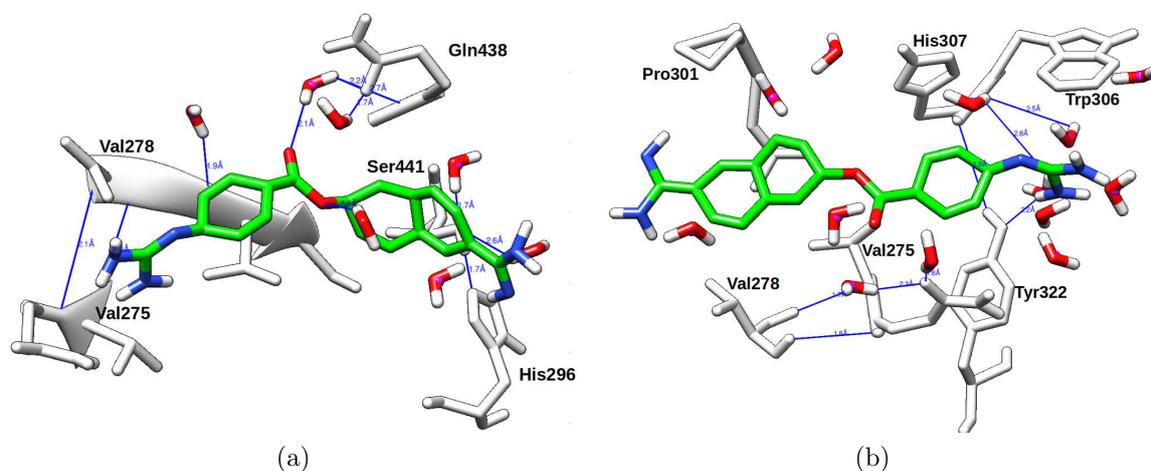


Figure 5: The role of water in mediating and stabilizing the interaction between TMPRSS2 and (a) compound **1**, and (b) nafamostat.

3 Discussion

TMPRSS2 is an important target towards developing an effective drug for COVID-19 treatment. Inhibitions of its proteolytic catalytic process will result in inactivation of the S protein from binding to the ACE2. Targeting inhibition by using small molecules have shown great promise in treatment of various viral diseases. The use of natural products has shown a great potential in the pharmaceutical industry and a number of natural products approved for other diseases has increasingly reported. This work has screened small molecules using computational approaches viz; molecular docking, molecular dynamics and MM-PBSA binding free energy to identify small molecules from African medicinal plant databases that could be used to combat the disease. In the work, we demonstrate that the binding free energy of small molecules can be improved by holo ensemble structure generated from molecular simulation which overcomes the limitations of protein flexibility. The RMSD and RMSF suggested that compound **1** is more stable than nafamostat, the experimentally known TMPRSS2 inhibitor. Vicinal water molecules in the catalytic triad site also played an important role in mediating the interaction between protein and the molecules in a different manner. Having several hydroxyl groups, compound **1** strongly interacted with catalytic site residues mediated with water molecules. For nafamostat, water

molecules are observed to have a stronger affinity with the catalytic residue than nafamostat. Binding free energy based on MM-PBSA confirmed the strong stability of compound **1** with binding free energy of -253 kJ/mol over nafamostat with -146 kJ/mol in the TMPRSS2 pocket. van der Waal forces were found to play big role in the stabilization of the complex. We have also observed an existence of a transient-like pocket in TMPRSS2 as shown by the binding of nafamostat. The methodology applied here suggests that compound **1** could strongly bind and inhibit the catalytic process of TMPRSS2 upon further in vitro and in vivo validation. Given the immediate need to combat COVID-19, the selected compound is proposed for immediate evaluation.

4 Material and Methods

4.1 Relaxed complex scheme based docking screening

Molecular docking using crystal structure was performed by using Autodock Vina [Trott and Olson, 2010]. The structure were obtained from Northern and South African natural compounds (NANPDB) (<https://african-compounds.org/nanpdb/>) in their SMILE file format, energy minimized and finally converted to .pdbqt file format using OpenBabel tool [O’Boyle et al., 2011]. The protein structure was obtained from the RCSB Protein Data Bank (PDB) with the following PDB ID: 7MEQ. Missing hydrogen at pH 7.4 and gasteiger charges were added to the protein structure and then converted to .pdbqt format. Docking screening experiments against TMPRSS2 were performed using our customized shell script. A restricted docking calculation targeting the catalytic and substrate region was performed. Achieving full protein flexibility is a major problem in docking calculations. Some docking tools are able to allow side chain residue flexibility. In this study, relaxed complex scheme based docking screening was performed as described in our previous works [Shadrack et al., 2020]. Briefly, the complex with compound **1** which ranked higher in crystal structure docking was selected and subjected to molecular dynamics simulation of 100 ns. Root mean square deviation (RMSD) was used to assess the stability of the complex, the system RMSD attained equilibrium state during the first 10 ns. To obtain the configurations for docking calculation, we computed the free energy (F) as the function of the RMSD as shown in equation 1.

$$F = -k_B T \ln(P) \quad (1)$$

Where, k_B is the Boltzmann constant, T is the temperature and P is the probability distribution of the RMSD. Structures with RMSD value corresponding to minimum free energy were sampled for docking calculation. As for the crystal structure, ensemble structures were prepared in a similar way and docking screening calculations were performed using AutoDock Vina [Trott and Olson, 2010]. It is important to highlight that our docking calculations were validated using a set of experimental data reported in our recent developed protocol [Shadrack et al., 2020] as well as comparison to root mean-square deviation (RMSD) values of the redocked molecule which gave an acceptable value of 2.5 Å.

4.2 Molecular dynamics (MD) simulation

Classical MD simulation of the complexes was performed as described in our previous works [Shadrack et al., 2020, Shadrack et al., 2021]. Here we describe the methodology briefly, nafamostat and compound **1** were selected for classical MD simulation, the structures obtained from docking calculations were used as the starting coordinates for classical MD simulation. Molecular dynamics for holo proteins was performed using Gromacs 2018 employing Amber03 force field. The systems were energy minimized using the steepest descent algorithm, followed by

equilibration at NVT (constant Number of particles, Volume and Temperature) ensemble for 1000 ps and NPT (constant Pressure) ensemble for 500 ps, with position restraints. The equilibrated system was then subjected to MD production run for 100 ns. During the equilibration stage, both temperature and pressure were maintained using Berendsen method [Nosé, 1986]. For the production stage, Parrinello-Rahman and v-rescale were used for pressure and temperature coupling at 300 K, respectively [Parrinello and Rahman, 1981, Bussi et al., 2007]. Periodic boundary conditions (PBC) were applied in all directions. Particle Mesh Ewald (PME) was used to treat long-range electrostatic interactions with a cutoff distance on 11 Å for both electrostatic and van der Waals interactions, while covalent bonds were constrained using the LINCS algorithm [Cheatham et al., 1995, Hess et al., 1997], a time step of 2 fs was used for all calculations. The MD configuration was used for analysis, gromacs built-in tools and our in-house scripts were used for analysis.

4.3 Molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) binding free energy calculations

Recently, it has been suggested that binding free energy and per-residue energy decomposition are very useful in establishing the underlying inhibition mechanism and biological activities target compound [Gupta et al., 2018]. Herein, free energy calculations and the per-residue decomposition was performed to gain further understanding on the inhibition mechanisms of the studied molecules. MM-PBSA was calculated using *g_mmpbsa* tool [Kumari et al., 2014]. The binding free energy (ΔG_{bind}) for protein-ligand interaction is expressed as

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \quad (2)$$

where, G_{complex} represents the total free energy of the protein-ligand complex while G_{protein} and G_{ligand} are total free energies of the individual protein and ligand in solvent, respectively. The free energy for each individual entity were computed as

$$G_x = \langle E_{\text{MM}} \rangle + \langle G_{\text{sol}} \rangle - T\Delta S \quad (3)$$

where, x stands for the protein or ligand or protein-ligand complex, T and S denote the temperature and entropy, respectively, and $T\Delta S$ refers to the entropic contribution to the free energy in a vacuum. The average molecular mechanics (MM) energy in vacuum is denoted as $\langle E_{\text{MM}} \rangle$, this term includes bonded and non-bonded interactions, $\langle E_{\text{MM}} \rangle$ is calculated based on MM force-field parameters as follows

$$\langle E_{\text{MM}} \rangle = \langle E_{\text{bonded}} \rangle + \langle E_{\text{nonbonded}} \rangle = \langle E_{\text{bonded}} \rangle + (E_{\text{vdW}} + E_{\text{elec}}) \quad (4)$$

where E_{bonded} is bonded interactions consisting of bond, angle, dihedral and improper interactions. The nonbonded interactions ($E_{\text{nonbonded}}$) include both electrostatic (E_{elec}) and van der Waals (E_{vdW}) interactions and are modeled using a Coulomb and Lennard-Jones (LJ) potential function, respectively. The free energy of solvation $\langle G_{\text{sol}} \rangle$ includes G_{polar} and $G_{\text{non-polar}}$ and can be calculated as

$$\langle G_{\text{sol}} \rangle = G_{\text{polar}} + G_{\text{non-polar}} = G_{\text{polar}} + (\gamma \times \text{SASA} + b) \quad (5)$$

where γ is a coefficient related to surface tension, SASA stands for solvent accessible surface area and b is the fitting parameter, $T\Delta S$ is the entropic contributions to free energy. In *g_mmpbsa* an entropic contribution is not considered [Kumari et al., 2014]. The binding free energy was calculated using a single trajectory, where a total of 200 snapshots were evenly extracted at a predetermined time from the first 40 ns and last 40 ns of the trajectory. The solvent and solute dielectric constants were 80 and 2, respectively, and γ was 0.0227, the PB equation was solved by using the linear PBsolver.

Acknowledgments

DMS acknowledge the CHPC lengau for providing access to the computational facilities used in this study. The financial support from the University of Dodoma through JAS program is highly acknowledged.

Authours contribution

SAHV and DMS conceived the study. DMS performed all experiments and data analysis. SAHV,MMO,IO,MGS and DMS wrote the manuscript, revised and approved.

References

- [Bussi et al., 2007] Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101.
- [Cheatham et al., 1995] Cheatham, T. I., Miller, J., Fox, T., Darden, T., and Kollman, P. (1995). Molecular dynamics simulations on solvated biomolecular systems: the particle mesh ewald method leads to stable trajectories of dna, rna, and proteins. *Journal of the American Chemical Society*, 117(14):4193–4194.
- [Chen et al., 2021] Chen, Y., Huang, W.-C., Yang, C.-S., Cheng, F.-J., Chiu, Y.-F., Chen, H.-F., Huynh, T. K., Huang, C.-F., Chen, C.-H., Wang, H.-C., et al. (2021). Screening strategy of tmprss2 inhibitors by fret-based enzymatic activity for tmprss2-based cancer and covid-19 treatment. *American Journal of Cancer Research*, 11(3):827.
- [Gupta et al., 2018] Gupta, A., Chaudhary, N., and Aparoy, P. (2018). Mm-pbsa and per-residue decomposition energy studies on 7-phenyl-imidazoquinolin-4 (5h)-one derivatives: Identification of crucial site points at microsomal prostaglandin e synthase-1 (mpges-1) active site. *International journal of biological macromolecules*, 119:352–359.
- [Haggag et al., 2020] Haggag, Y. A., El-Ashmawy, N. E., and Okasha, K. M. (2020). Is hesperidin essential for prophylaxis and treatment of covid-19 infection? *Medical Hypotheses*, 144:109957.
- [Hess et al., 1997] Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G. (1997). Lincs: a linear constraint solver for molecular simulations. *Journal of computational chemistry*, 18(12):1463–1472.
- [Hu et al., 2021] Hu, X., Shrimp, J. H., Guo, H., Xu, M., Chen, C. Z., Zhu, W., Zakharov, A. V., Jain, S., Shinn, P., Simeonov, A., et al. (2021). Discovery of tmprss2 inhibitors from virtual screening as a potential treatment of covid-19. *ACS Pharmacology & Translational Science*.
- [Huggins, 2020] Huggins, D. J. (2020). Structural analysis of experimental drugs binding to the sars-cov-2 target tmprss2. *Journal of Molecular Graphics and Modelling*, 100:107710.
- [Kumari et al., 2014] Kumari, R., Kumar, R., Consortium, O. S. D. D., and Lynn, A. (2014). g a gromacs tool for high-throughput mm-pbsa calculations. *Journal of chemical information and modeling*, 54(7):1951–1962.
- [Nosé, 1986] Nosé, S. (1986). An extension of the canonical ensemble molecular dynamics method. *Molecular Physics*, 57(1):187–191.

- [O’Boyle et al., 2011] O’Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):33.
- [Parrinello and Rahman, 1981] Parrinello, M. and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190.
- [Peiffer et al., 2021] Peiffer, A. L., Garlick, J. M., Wu, Y., Soellner, M. B., Brooks III, C. L., and Mapp, A. K. (2021). Tmprss2 inhibitor discovery facilitated through an in silico and biochemical screening platform. *bioRxiv*.
- [Shadrack et al., 2021] Shadrack, D. M., Deogratias, G., Kiruri, L. W., Swai, H. S., Vianney, J.-M., and Nyandoro, S. S. (2021). Ensemble-based screening of natural products and fda-approved drugs identified potent inhibitors of sars-cov-2 that work with two distinct mechanisms. *Journal of Molecular Graphics and Modelling*, 105:107871.
- [Shadrack et al., 2020] Shadrack, D. M., Swai, H. S., and Hassanali, A. (2020). A computational study on the role of water and conformational fluctuations in hsp90 in response to inhibitors. *Journal of Molecular Graphics and Modelling*, 96:107510.
- [Trott and Olson, 2010] Trott, O. and 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.
- [Xu and Lill, 2011] Xu, M. and Lill, M. A. (2011). Significant enhancement of docking sensitivity using implicit ligand sampling. *Journal of chemical information and modeling*, 51(3):693–706.
- [Zhu et al., 2021] Zhu, H., Du, W., Song, M., Liu, Q., Herrmann, A., and Huang, Q. (2021). Spontaneous binding of potential covid-19 drugs (camostat and nafamostat) to human serine protease tmprss2. *Computational and structural biotechnology journal*, 19:467–476.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SI.pdf](#)