

# Spontaneous binding of potential COVID-19 drugs to human serine protease TMPRSS2

Haixia Zhu<sup>1, †</sup>, Wenhao Du<sup>1, †</sup>, Menghua Song<sup>1</sup>, Qing Liu<sup>2</sup>, Andreas Herrmann<sup>3</sup> and Qiang Huang<sup>1, 4, \*</sup>

<sup>1</sup>*State Key Laboratory of Genetic Engineering, Shanghai Engineering Research Center of Industrial Microorganisms, MOE Engineering Research Center of Gene Technology, School of Life Sciences, Fudan University, Shanghai 200438, China*

<sup>2</sup>*State Key Laboratory of Quality Research in Chinese Medicines, School of Pharmacy, Macau University of Science and Technology, Macau, China*

<sup>3</sup>*Institute for Biology and IRI Lifesciences, Humboldt-Universität zu Berlin, 10115 Berlin, Germany*

<sup>4</sup>*Multiscale Research Institute of Complex Systems, Fudan University, Shanghai 201203, China*

<sup>†</sup>These first two authors contributed equally to this work.

\*Corresponding author. Email: huangqiang@fudan.edu.cn

## Abstract

Effective treatment or vaccine is not yet available for combating SARS coronavirus 2 (SARS-CoV-2) that caused the COVID-19 pandemic. Recent studies showed that two drugs, Camostat and Nafamostat, might be repurposed to treat COVID-19 by inhibiting human TMPRSS2 required for proteolytic activation of viral spike (S) glycoprotein. However, their molecular mechanisms of pharmacological action remain unclear. Here, we perform molecular dynamics simulations to investigate their native binding sites on TMPRSS2. We revealed that both drugs could spontaneously and stably bind to the TMPRSS2 catalytic center, and thereby inhibit its proteolytic processing of the S protein. Also, we found that Nafamostat is more specific than Camostat for binding to the catalytic center, consistent with reported observation that Nafamostat blocks the SARS-CoV-2 infection at a lower concentration. Thus, this study provides mechanistic insights into the Camostat and Nafamostat inhibition of the SARS-CoV-2 infection, and offers useful information for COVID-19 drug development.

## Introduction

Coronaviruses (CoVs) are single-stranded RNA viruses that can spread in animals and humans, causing a variety of diseases, such as respiratory, intestinal, kidney, and nervous system diseases (1). To date, three highly pathogenic human coronaviruses (hCoVs) have been identified, including severe acute respiratory syndrome coronavirus (SARS-CoV) (2), Middle East respiratory syndrome coronavirus (MERS-CoV) (3) and the 2019 novel coronavirus (SARS-CoV-2) that emerged in Wuhan in December, 2019 (4). All three viruses have the ability to infect human host cells and thus to transmit from humans to humans. However, according to the data released by WHO, the spread rate of SARS-CoV-2 in humans has significantly exceeded those of SARS-CoV and MERS-CoV (5), leading to a global

pandemic of the coronavirus disease 2019 (COVID-19). Up to October 4, 2020, approximately 34.826 million COVID-19 cases were confirmed worldwide, including 1,032,235 deaths (6). Genomic sequencing and phylogenetic analysis showed that SARS-CoV-2 shares 79.6% sequence identity to SARS-CoV, and belongs to  $\beta$ -coronavirus. Because the overall genome sequence reaches the highest similarity (96.2%) to the bat coronavirus bat-SL-CoV-RaTG13, it was thought that this novel virus might have been transmitted from bats to humans via unknown intermediate host animals (7). As SARS-CoV-2 is a novel human coronavirus, currently there is no effective treatment or vaccine to combat this virus. Therefore, scientists and researchers all over the world are having a race against time to develop effective drugs and vaccines for treating COVID-19 (8), especially drugs that could block the SARS-CoV-2 entry into the human cells, the first step of the viral infection in humans.

The hCoV entry into the human cells is mediated by its trimeric transmembrane spike (S) glycoprotein (9). In general, to release the viral RNA genome into a host cell, the S protein binds to the host receptor angiotensin-converting enzyme 2 (ACE2) on the host cell surface, and upon endocytic uptake the viral membrane fuses with the endosomal membrane (10). In addition to ACE2, human proteases are essential to prime and to activate the S-ectodomain for binding and fusion (11). Two cleavage sites, S1/S2 and S2', located at the boundary between the S1 and S2 subunits, have to be proteolytic processed by different proteases, such as furin, the transmembrane protease serine 2 (TMPRSS2), and cathepsin L and B (12). Cleavage of the S1/S2 site by furin primes the spike protein for efficient binding of S1 to ACE2 through its receptor-binding domain (RBD) attaching the virus to the cell surface (13). TMPRSS2 on the cell surface processes S2', priming the S protein to mediate fusion with the endosomal membrane by its fusion peptide (14). Although both endosomal cathepsins process S2' as well, TMPRSS2 has been shown to be more essential for the S protein priming and thus infection by SARS-CoV-2 (15). In consequence, the TMPRSS2 processing is one of the key steps to activate the membrane function of the SARS-CoV-2 S protein (16). Hence, using drugs to inhibit the proteolytic activity of TMPRSS2 is likely to block the membrane fusion of SARS-CoV-2 (Fig. 1A). Interestingly, TMPRSS2 is a human protease, and as the drug target, will not cause the problem of developing drug resistance like the viral protein targets (17). Indeed, TMPRSS2 is one of the most promising targets for the anti-SARS-CoV-2 drugs (18).

Previous studies have shown that several TMPRSS2 inhibitors could effectively block the hCoV infection (19). Kawase et al. found that Camostat, a drug for treating chronic pancreatitis, can block the SARS-CoV and HCoV-NL63 infections by inhibiting the TMPRSS2 activity (20). Later, Yamaoto et al. screened a library of 1,017 FDA-approved drugs using Dual Split Protein (DSP) reporter fusion assay, and then found that Nafamostat, another related agent for pancreatitis and disseminated intravascular coagulation, can also block the MERS-CoV membrane fusion (21). After the emergence of SARS-CoV-2, scientists soon investigated the ability of these two drugs to block the SARS-CoV-2 infection. Hoffmann et al. firstly showed that Camostat has inhibitory effects on the SARS-CoV-2 in TMPRSS2-expressing human cells (15). More recently, a group at the University of Tokyo confirmed that Nafamostat can block the SARS-CoV-2 fusion at a concentration less than one-tenth that required for Camostat; they reported that 1~10 nM of Nafamostat can significantly inhibit the cell infection of SARS-CoV-2 (22). These results strongly supported

that both Camostat and Nafamostat are potential drugs for treating COVID-19 (23-25). Indeed, several clinical trials for evaluating their therapeutic effects against SARS-CoV-2 are now underway (26). Moreover, these two old drugs have been commercialized for many years; so, once they pass the clinical trials, they may be immediately applied to the COVID-19 treatment.

Although they are undergoing the COVID-19 clinical trials, the molecular mechanisms in which Camostat and Nafamostat inhibit the TMPRSS2 activity remain unclear. To provide guidance for their repurposing for treating COVID-19, it is urgent to answer the following questions: Can Camostat or Nafamostat bind to the catalytic center of TMPRSS2 and then inhibit its proteolytic activity? What are the key molecular interactions in binding process? What are the stable conformations of the drugs bound to TMPRSS2? Besides the on-target binding site (i.e., the catalytic center), are there other off-target binding hotspots on TMPRSS2? To address these questions, we performed atomic-level, unbiased molecular dynamics (MD) simulations to investigate the dynamic binding processes of Camostat and Nafamostat to TMPRSS2. In these simulations, the drug molecules were initially placed at random positions distant from the TMPRSS2 surface; then, without giving any prior knowledge of the drug binding sites, drugs diffused around TMPRSS2 to spontaneously “recognize” their native binding sites on TMPRSS2. Our simulations showed that both Camostat and Nafamostat could spontaneously and stably bind to the TMPRSS2 catalytic center, indicating that the catalytic center is their native binding site. We found that the main driving forces for the binding are the electrostatic attractions between the drug guanidinium group and the surface Asp/Glu residues around the catalytic center. As the drugs enter the catalytic center, van der Waals forces and hydrogen bonds between the drugs and TMPRSS2 stabilize them in the center in an induced-fit way. For this reason, the bound drugs occupy the space required for the substrate binding, and thereby inhibit the proteolytic activity.

## Results

### Building the 3D model of the TMPRSS2 extracellular domain

Human TMPRSS2 protein consists of 492 amino acids and is divided into three domains: the intracellular (aa 1-84), the transmembrane (aa 85-105) and the extracellular (aa 106-492) domain (Fig. 2A) (27). The catalytic site for proteolysis is localized in the extracellular domain (ECD). So far, no three-dimensional (3D) structure of TMPRSS2 has been resolved. Thus, we predicted the atomic model of the TMPRSS2 extracellular domain (TMPRSS2-ECD) by computational methods.

To build the model, we used the Swiss-Model server (<https://swissmodel.expasy.org/>). Amino-acid sequence alignment indicated that the structure of a type II transmembrane trypsin-like serine protease hepsin (PDB ID: 1Z8G) shares the best sequence coverage (89%) and the highest GMQE (Global Model Quality Estimation, 0.64) to TMPRSS2-ECD (fig. S1). In fact, 1Z8G has been widely used as the modeling template of TMPRSS2 (28). Thus, this crystal structure was employed as the template for our model construction. Because this structure lacks the N-terminal segment corresponding to aa 106-145 of TMPRSS2, a 3D structure for the aa 146-492 of TMPRSS2-ECD was built. Notably, the 3D model implies that the distance from the lacking N-terminal segment (aa 106-145) to the active site is greater than 37 Å. Next, to validate the predicted structure, we performed MD simulation with

explicit solvent to further optimize the built TMPRSS2-ECD structure (see Materials and Methods). As shown in fig. S2, the average root mean square deviation (RMSD) of the simulation system in the simulation time  $> 60$  ns is about 5 Å, and the ECD structure maintains a stable state in aqueous solution up to 200 ns. To verify this, we also examined whether the catalytic center of the TMPRSS2-ECD model possesses a conformation identical to that of a typical serine protease. Comparison of the MD snapshot structure at 200 ns with the crystal structure of a typical trypsin (PDB ID: 2PTC) shows that the catalytic triad conformations of the two proteins are very similar, with root mean square deviation (RMSD) of 1.67 Å (fig. S3). As seen, the TMPRSS2 catalytic triad consists of H296, D345 and S441, corresponding to H57, D102 and S195 in trypsin, respectively (29). So, the TMPRSS2-ECD structure obtained from the homology modeling and the MD refinement has a conserved catalytic center of serine proteases. Eventually, we chose the MD snapshot structure at 200 ns as the starting model of TMPRSS2-ECD for the following simulations (Fig. 2A).

Since the experimental structure the TMPRSS2-spike complex is not yet available, to understand the binding mode of the substrate to the TMPRSS2 active site, we compared TMPRSS2-ECD with 2PTC in complex with a peptide inhibitor. Structural alignment indicated that the peptide scissile bond of the protein substrate is located in the catalytic center of TMPRSS2 (Fig. 2B), implying that the catalytic triad of TMPRSS2 likely uses the same proteolytic mechanism of serine proteases to cleave its protein substrate. Based on the orientation of the scissile bond in Fig. 2B, we designated the TMPRSS2 areas occupied by the N-terminal and C-terminal substrate segments of the scissile bond as the N-terminal and C-terminal binding regions, respectively (Fig. 2C).

### **Two drugs spontaneously enter the TMPRSS2 catalytic center**

It is reasonable to hypothesize that any drug molecule that effectively inhibit the proteolysis function of TMPRSS2 has to be bound to the catalytic amino-acids and/or the substrate-binding regions. To test this hypothesis, atomic-level, unbiased MD simulation of the dynamic association process of a given drug with the protein receptor is an effective tool (30, 31). To simulate those association processes for Camostat and Nafamostat in an aqueous environment, we used the mentioned TMPRSS2-ECD structure at 200 ns to establish the simulation systems. As described in Materials and Methods, the protein structure was placed at the center of the simulation box, while the drug molecule (Camostat or Nafamostat) was randomly placed around the protein, with initial position away at least 30 Å from the catalytic triad and at least 15 Å from the protein surface, ensuring that the dynamic binding of the drug to TMPRSS2 is not predetermined by the initial arrangement of components (fig. S4). In the simulations, no bias forces were added to the drug molecules; in other words, the drug movement from the initial position to the catalytic center is completely driven by the interplay of the drug with solvent molecules and TMPRSS2. For each drug, we performed at least 15 independent simulations starting from a random position of drug and lasting at least for 150 ns (Table 1).

To determine whether a drug is bound to the TMPRSS2 catalytic center, we defined the minimum pairwise atomic distance between the drug and the catalytic residues S441 and H296 as “drug distance to the catalytic center” ( $D_{cc}$ ). For  $D_{cc} < 3$  Å lasting at least for 20 ns, the drug molecule is considered to be bound to the catalytic center. Within the simulation

timescale of 150 ns both Camostat and Nafamostat were found to spontaneously bind to the catalytic center with a successful rate of ~40% (Table 1 and fig. S5). The successful binding trajectories showed that Camostat and Nafamostat follow two association pathways: the drug binds to the catalytic center from either the N-terminal binding region (Fig. 3, A and B) or the C-terminal binding region (Fig. 3, C and D). Among them, Camostat prefers to bind to the catalytic center from the N-terminal binding region, Nafamostat prefers to the C-terminal binding region (Table 1) As seen in Fig. 3, both drugs from distantly random positions could associate with TMPRSS2, and finally enter the catalytic center after short periods of conformational adjustment (see also Movies S1 and S2 in Supplementary Materials). In the processes, even if the drugs diffused in the solvent to positions with  $D_{cc} > 60 \text{ \AA}$ , eventually they entered the catalytic center.

To elucidate the driving forces of the binding, we analyzed the successful trajectories that capture the association processes of the drugs with the catalytic center (fig. S5). The simulations showed that whenever the drugs move toward the catalytic center either from the N-terminal binding region or from the C-terminal binding region, their association processes can be divided into two phases, as illustrated by the two typical trajectories in Fig. 3. Initially, driven by the thermal motions of the solvent molecules, the Camostat and Nafamostat drugs diffused randomly around the TMPRSS2 surface until they became attracted to the substrate-binding region (see trajectories in Fig. 3, A and C). Here, as there are many negatively charged Asp/Glu residues in/around this region (fig. S6), the electrostatic interactions between the Asp/Glu oxygens and the positively charged guanidinium group of the drugs attracted the drugs to the TMPRSS2 surface in the vicinity of the substrate-binding region (e.g., panel 2 in Fig. 3, B and D). Next, by forming hydrogen bonds with amino acids in the catalytic center (see below), the drugs continued to adjust their conformations and positions, and eventually achieved the final stable poses in the catalytic center. For example, as illustrated by the trajectory in Fig. 3A, Camostat ever adjusted its head-tail orientation, so that its guanidinium group eventually contacted the catalytic center (Fig. 3B, panel 3). As shown by the trajectory in Fig. 3C, Nafamostat directly approached the catalytic center without any significant head-tail adjustment (Fig. 3D, panel 3). Finally, the hydrogen bonds formed by the drug guanidinium group with the amino acids in the catalytic center stabilized both drugs in the center (Figs. 3, B and D, panel 4). As a result, the hydrogen bonding between S441 and H296 (Fig. 2B) is disrupted by the drug guanidinium group being positioned in the middle of H296 and S441. On-going simulations revealed that the  $D_{cc}$  values of both drugs are almost less than 3 Å, indicating that they are in the stable binding states at least in the time windows of the simulations (Figs. 3, A and C). Note that, besides the trajectories in Fig. 3, drug binding processes of other independent simulation trajectories in fig. S5 are similar. Of course, because of the dynamic nature of the drug binding, all the trajectories are somehow different, but the results are identical: Camostat and Nafamostat become attracted to and spontaneously enter the substrate-binding cavity of TMPRSS2 from the N-terminal or C-terminal binding region, and finally bind to the TMPRSS2 catalytic center.

### **Both drugs are stably bound to the TMPRSS2 catalytic center**

To elucidate the key interacting groups between the drugs and TMPRSS2, we used AutoDock 4.1 semi-empirical free energy function (32) to calculate the free energies of drug binding to

the catalytic center using the snapshots of the drug-TMPRSS2 complexes in the trajectories in Fig. 3. As validated in our previous study (31), this calculation method of binding free-energy is fast and reliable for analyzing large numbers of inhibitor-receptor snapshot complexes in the MD simulations. The calculations for the trajectories in Fig. 3 showed that the drug-TMPRSS2 complex conformations with the lowest binding free energy appeared at ~128 ns (Camostat) and ~66 ns (Nafamostat) (fig. S7), corresponding to  $\Delta G_{bind}$  of  $-9.66 \text{ kcal}\cdot\text{mol}^{-1}$  and  $-9.16 \text{ kcal}\cdot\text{mol}^{-1}$ , respectively. So, although experimental values are not yet available, according to  $\Delta G_{bind} = RT \ln K_d$ , the equilibrium dissociation constants  $K_d$  of the two drugs are predicted in the order of nanomoles.

As shown in Fig. 4, Camostat and Nafamostat embed either in the N-terminal or C-terminal regions of the binding cavity near the catalytic center. The guanidinium groups of the drugs are located between S441 and H296. All three N atoms of the Camostat guanidinium group participate in hydrogen bonding: N1 and N2 form four hydrogen bonds with D440 and S441, and N3 forms a bidentate hydrogen bond with H279 and V280 (Fig. 4A). These hydrogen bonds fix the position of the guanidinium group. In addition, the oxygen atom between two aromatic rings also forms a hydrogen bond with H279, stabilizing the molecular center in the binding cavity. For Nafamostat, the two N atoms of the guanidinium group participate in bonding: N1 and D440 form a bidentate hydrogen bond, and N3 simultaneously interacts with D440, S441 and S460 (Fig. 4B). These hydrogen bonds highly restrict the structural flexibility of the guanidinium group. In addition, compared to the catalytically active conformation in Fig. 2B, the distance between H296 and S441 exceeds 5 Å. Such a distance enlargement is attributed to the steric effects of the bound guanidinium group on the H296 side chain.

To understand the intermolecular forces that maintain the drug-TMPRSS2 complex in the stable binding state, we further analyzed the energy terms contributing to the binding free energies. As shown in Fig. 4C, the main intermolecular forces involved in binding are van der Waals forces, hydrogen bonds, and electrostatic forces. Among them, the van der Waals forces contribute the most. This may be attributed to the matching of the drug shapes to that of the binding cavity, which results in a close contact between the drug atoms and the amino acids in the catalytic center. As mentioned, the drugs also form multiple hydrogen bonds with the amino acids in the catalytic center (Figs. 4, A and B). This restricts their orientation in the catalytic center, and therefore enhances the binding stability. Finally, the electrostatic forces between the positively charged guanidinium group of the drugs and the oxygen atoms in the binding cavity further strengthen the drug-TMPRSS2 binding stability (fig. S6).

In conclusion, the drugs interact with the catalytic amino acids via their guanidinium groups, and form stably bound conformations in the catalytic center, thereby occupy the space required for the substrate binding. By competing with the substrate for the same binding site, Camostat and Nafamostat may inhibit the catalytic activity of TMPRSS2. The guanidinium group appears to be the key pharmacophore of the drugs. Indeed, the conserved arginine (Arg) at the cleavage site of the TMPRSS2 substrate also fully supports this, because the Arg side-chain possesses a guanidinium group. This molecular similarity implies that both drugs binds to the TMPRSS2 in the same way as the substrate does.

### **Nafamostat is more specific for the on-target binding**

To find out all potential high-affinity binding hotspots of the drugs on TMPRSS2, we analyzed the transient complex conformations of the drugs bound to TMPRSS2 in all the independent MD simulations in Table 1. To this end, we firstly sorted out the high-affinity complex conformations with the binding free energy  $\Delta G_{bind} < 6.82 \text{ kcal}\cdot\text{mol}^{-1}$  (corresponding to  $K_d < 10 \text{ }\mu\text{M}$ ). Next, we calculated their pairwise RMSDs, and then classified any two conformations with a pairwise RMSD  $< 4 \text{ \AA}$  into a conformational cluster. We used the conformation with the lowest binding free energy in the cluster as the representative conformation, and regarded its binding site as the TMPRSS2 binding hotspot of the cluster. In Fig. 5, all high-affinity binding hotspots of both drugs are shown. As seen, both Camostat and Nafamostat have certain numbers of high-affinity binding hotspots. Since the high-affinity hotspots outside the substrate-binding region may affect the association efficiency of the drugs with the catalytic center (i.e., the target binding site), we defined the hotspots contacting the substrate-binding region as on-target hotspots, while the others as off-target hotspots. The distribution of the on-target hotspots in Fig. 5 demonstrates that both drugs occupy the substrate-binding region in multiple orientations, suggesting that both drugs are able to compete against the substrate for the binding and thereby inhibit the TMPRSS2 proteolytic activity.

However, both the total numbers of the on-target and off-target hotspots of Camostat are higher than those of Nafamostat (Fig. 5, A and B, panel bottom). Very likely, this can be attributed to the structural difference between two drug molecules. Although the structures of Camostat and Nafamostat are similar, Nafamostat is slightly shorter in size and possess more aromatic rings, resulting in a more rigid structure. Thus, less conformational clusters are possible when binding to TMPRSS2. In contrast, Camostat is slightly longer and more flexible, which may give rise to form more binding poses on TMPRSS2 and thus more binding clusters. As a result, Camostat has more off-target hotspots than Nafamostat. Obviously, this could decrease the binding efficiency of Camostat to the substrate-binding region. Thus, if the numbers of the drug molecules are equal, the proportion of Camostat bound to the substrate-binding region is lower than that of Nafamostat, because of the stronger competition of the off-target hotspots for the binding. In other words, to target the same number of the TMPRSS2 molecules, Nafamostat requires lower concentration than Camostat. This is fully consistent with experimental observations on SARS-CoV-2 membrane fusion showing inhibition by Nafamostat in the range of 1~10 nM while Camostat requires 10~100 nM to achieve a similar extent of inhibition (22).

Moreover, our analysis of the binding free-energy revealed that the lowest binding free energy of Camostat was found for an off-target hotspot with  $-9.93 \text{ kcal}\cdot\text{mol}^{-1}$ , which is lower than all those of the identified on-target hotspots (Fig. 5A). This implies that Camostat could bind to this off-target hotspot with a greater affinity than to the on-target hotspots. This may further reduce the probability of Camostat to bind to the substrate-binding region. In contrast, the binding energy of Nafamostat to the substrate-binding region is lower than those for any off-target site ( $-9.16 \text{ kcal}\cdot\text{mol}^{-1}$  in Fig. 5B), indicating that the catalytic center provides the highest-affinity for Nafamostat. In conclusion, Nafamostat is more specific than Camostat for the on-target binding to the catalytic center. Again, this implies that the antiviral effect of Nafamostat against SARS-CoV-2 is likely better than that of Camostat.

## Discussion

During the current challenging period of accelerated development of COVID-19 therapeutic vaccines and drugs around the world, here we have performed the spontaneous binding simulations to investigate the molecular mechanisms of pharmacological action of two potential COVID-19 drugs - Camostat and Nafamostat. Both drugs are undergoing clinical trials against SARS-CoV-2. Our simulations successfully captured their dynamic association processes with the TMPRSS2 catalytic center (Fig. 3). This demonstrates that both drugs diffusing from randomly initial positions in the solvent, approach the TMPRSS2 surface by attractions of the Asp/Glu residues at the substrate binding site, and finally bind to the catalytic center by the van der Waals forces and hydrogen bonds (Fig. 4). In addition, by mapping the high-affinity hotspots of both drugs on TMPRSS2 (Fig. 5), we found that Nafamostat is more specific for the binding to the TMPRSS2 catalytic center. This provides a molecular explanation for the observation that Nafamostat possesses better inhibition effects than Camostat.

Our study implies that Camostat and Nafamostat are effective inhibitors to block the TMPRSS2-mediated cleavage of the S protein required for the SARS-CoV-2 infection in humans. Moreover, we revealed that the guanidinium group of the drugs is critical for driving the drug associations with the TMPRSS2 catalytic center. When developing new drugs targeting TMPRSS2, one should pay specific attention to this group. No doubt, there are still open questions to be addressed for deeply understanding the molecular mechanisms of drug action. For example, experimental determination of the TMPRSS2 structure in complex with Camostat or Nafamostat will be very helpful for the confirmation of our simulations. Meanwhile, in all the simulations of 150 ns (Table 1), we did not observed drug binding modes similar to that of Camostat in the protease prostatic (33). Because so far no experimental evidence for a covalent binding of the two drugs to TMPRSS2 is available, further studies are needed to clarify this issue.

In summary, this study not only provides mechanistic insights into the Camostat and Nafamostat inhibition of the SARS-CoV-2 infection, but also offers useful information for the repurposing of Camostat and Nafamostat for treating COVID-19.

## Materials and Methods

### Homology modeling of the TMPRSS2-ECD structure

Because no experimental structure is available for TMPRSS2, an initial, all-atom model of TMPRSS2-ECD was first generated by homology modeling via the Swiss-Model server (<https://swissmodel.expasy.org/>). The amino-acid sequence of TMPRSS2 was obtained from UniProt (UniProt ID: O15393; GenBank No: U75329). Then, by uploading the sequence of TMPRSS2-ECD, the 3D model was built using the fully automated modeling mode of the Swiss-Model server. The server searched the existing structure that shares the highest sequence identity to that of TMPRSS2-ECD as the template for building the model. The crystal structure of the extracellular region of the transmembrane serine protease hepsin with a resolution of 1.55 Å (PDB ID: 1Z8G) was found to share the best sequence coverage and the highest GMQE (Global Model Quality Estimation) to TMPRSS2-ECD. So, with this crystal



structure as the template, a 3D atomistic model was eventually constructed for TMPRSS2-ECD (aa 146~492).

### **MD refinement of the TMPRSS2-ECD model**

The homology all-atom model of TMPRSS2-ECD was then optimized by the MD simulation in solution state. The MD simulations were conducted using GROMACS (Ver. 5.1.4). The CHARMM27 force field (34) and the TIP3P water model (35) were employed to model the simulation system. In the simulation system, the all-atom structure of TMPRSS2-ECD was placed in the center of a rectangular water box with a minimal distance of 15 Å from the protein surface to its boundary. Certain numbers of Na<sup>+</sup> and Cl<sup>-</sup> ions were added to the system for setting an ionic concentration of 150 mM and neutralizing the system.

To optimize the system, the energy minimization of the system was first carried out using the steepest descent algorithm for a maximum of 50,000 steps or until the maximum force < 1000 kJ·mol<sup>-1</sup>·nm<sup>-1</sup>. Following the energy minimization, the system was heated by NVT equilibration at 320 K for 100 ps. Then, NPT equilibration for 1 ns was followed by setting the system pressure to 1 bar. In the simulations, the integration time step was 2 fs. The V-rescale method (36) was used to maintain the average temperature of 320 K; and the Parrinello-Rahman barostat (37) was used for maintaining the average pressure of 1 bar. The periodic boundary conditions (PBC) were applied in all three dimensions. The bond lengths and angles were constrained using the LINCS algorithm (38). The Particle Mesh Ewald (PME) method (39) was used for the long-range electrostatics; and a cut-off distance of 10 Å was employed for both short-range electrostatic interactions and van der Waals interactions. Finally, MD refinement simulation was performed in the NVT ensemble for 200 ns. In the simulation, the coordinates of the system atoms were recorded per 10 ps for the analysis.

### **System set-up for spontaneous binding simulations**

As in the above MD refinement, CHARMM27 force field was used to model the protein, and the TIP3P water model was employed for the solvent. The 2D chemical structures of Camostat and Nafamostat were downloaded from PubChem (<https://pubchem.ncbi.nlm.nih.gov>). The ionization state of the drugs at pH~7.0 was predicted using ChemAxon (<https://chemaxon.com/>) and the final 3D structures of two drugs with protons were obtained using MolView (<http://molview.org/>). The drugs were modeled using the General Amber Force Field (GAFF) (40). Their topology and parameter files were generated using tLEaP module of AmberTools17 and converted to GROMACS-compatible files with ACPYPE (41).

To build the initial system for simulating a drug association with TMPRSS2, the refined structure of TMPRSS2-ECD was placed in the center of a rectangular water box, with a distance at least 15 Å from protein surface to the box boundary. Then, a drug molecule (Camostat or Nafamostat) was placed at a random position in the solvent, at least 30 Å away from the TMPRSS2 catalytic center. The system was added certain numbers of Na<sup>+</sup> and Cl<sup>-</sup> ions for setting the ionic concentration to 150 mM and neutralization. Next, the systems were minimized and equilibrated using the same control methods of temperature and pressure as those in the above MD refinement. Finally, for each drug 15 independent simulations were

performed in the NVT ensemble, and each simulation lasted at least for 150 ns. In the simulations, the coordinates of the system atoms were recorded per 100 ps for the analysis.

### Time-dependent drug distance to the catalytic center ( $D_{cc}$ )

We determined whether a drug (Camostat or Nafamostat) binds to the TMPRSS2 catalytic center in the simulations by defining a drug distance to the catalytic center ( $D_{cc}$ ). In the simulations, the TMPRSS2 region within 3 Å of S441 and H296 is regarded as the catalytic center (Fig. 5, region in cyan). Then, in a given frame of the MD trajectories, the minimum distance from any atom of the drug to any atom of the S441 and H296 is defined as  $D_{cc}$ . In a given simulation, if  $D_{cc} < 3$  Å, the drug molecule is considered to enter the catalytic center in that MD frame. If this state lasts for more than 20 ns, the simulation will be considered as a successful trajectory to capture the spontaneous binding process of the drug to the catalytic center.

### Binding free energy calculation

We used AutoDockTools (42) to calculate the free energy of a drug (Camostat or Nafamostat) that bind to TMPRSS2-ECD in the simulation snapshots. AutoDockTools uses the AutoDock 4.1 semi-empirical free energy force field to estimate the free energy of a small-molecular drug binding to a protein receptor based on their drug-protein complex. With the given complex, the binding free energy ( $\Delta G_{bind}$ ) equals to the change in the free energy of the drug from the unbound state to the bound state:

$$\Delta G_{bind} = (\Delta G_{intermol} + \Delta G_{intramol} + \Delta G_{tor}) - \Delta G_{unbound} \quad (1)$$

where the first two terms are the energies of the drug-protein complex in the bound state, consisting of the intermolecular and intramolecular free energies. The third term is the conformational entropy change of the drug in the binding and is directly calculated from the sum of the torsional degrees of freedom. The fourth term is the reference energy of the drug in its unbound state and is defined as 0.0 kcal·mol<sup>-1</sup>. Of them, the intermolecular energy involves in van der Waals, hydrogen bonding, desolvation and electrostatic contributions, and is calculated by:

$$\Delta G_{intermol} = \Delta G_{vdW} + \Delta G_{H-bond} + \Delta G_{desolv} + \Delta G_{elec} \quad (2)$$

For the drug-TMPRSS2 complex in a given MD frame, Python programs in AutoDockTools were directly used to calculate the AutoDock 4.1 parameters of the drug and TMPRSS2 and the binding free energy according to standard procedures.

### References

1. S. Su, G. Wong, W. F. Shi, J. Liu, A. C. K. Lai, J. Y. Zhou, W. J. Liu, Y. H. Bi, G. F. Gao, Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* **24**, 490-502 (2016).
2. N. S. Zhong, B. J. Zheng, Y. M. Li, L. L. M. Poon, Z. H. Xie, K. H. Chan, P. H. Li, S. Y. Tan, Q. Chang, J. P. Xie, X. Q. Liu, J. Xu, D. X. Li, K. Y. Yuen, J. S. M. Peiris, Y. Guan, Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353-1358 (2003).
3. A. M. Zaki, S. van Boheemen, T. M. Bestebroer, A. D. Osterhaus, R. A. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814-1820 (2012).
4. N. Zhu, D. Y. Zhang, W. L. Wang, X. W. Li, B. Yang, J. D. Song, X. Zhao, B. Y. Huang, W. F. Shi, R. J. Lu, P. H. Niu, F. X. Zhan, X. J. Ma, D. Y. Wang, W. B. Xu, G. Z. Wu, G. G. F. Gao, W. J. Tan, A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727-733 (2020).

5. *Coronavirus disease (COVID-2019) situation reports*(WHO Situation Report Series No. 45, World Health Organization, Geneva, Switzerland, 2020).
6. Johns Hopkins University. World Map. <https://coronavirus.jhu.edu/map.html> (2020).
7. P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. D. Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao, Z. L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
8. BioRender. COVID-19 Vaccine & Therapeutics Tracker. <https://biorender.com/covid-vaccine-tracker> (2020).
9. A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Velesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281-292 (2020).
10. W. H. Li, M. J. Moore, N. Vasilieva, J. H. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, M. Farzan, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454 (2003).
11. G. Simmons, P. Zmora, S. Gierer, A. Heurich, S. Pöhlmann, Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. *Antiviral Res.* **100**, 605-614 (2013).
12. J. A. Jaimes, J. K. Millet, G. R. Whittaker, Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 site. *Iscience* **23**, (2020).
13. W. Tai, L. He, X. Zhang, J. Pu, D. Voronin, S. Jiang, Y. Zhou, L. Du, Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: Implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol.* **17**, 613-620 (2020).
14. X. Ou, W. Zheng, Y. Shan, Z. Mu, S. R. Dominguez, K. V. Holmes, Z. Qian, Identification of the fusion peptide-containing region in betacoronavirus spike glycoproteins. *J. Virol.* **90**, 5586-5600 (2016).
15. M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N. H. Wu, A. Nitsche, M. A. Muller, C. Drosten, S. Pöhlmann, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271-280 (2020).
16. S. Matsuyama, N. Nao, K. Shirato, M. Kawase, S. Saito, I. Takayama, N. Nagata, T. Sekizuka, H. Katoh, F. Kato, M. Sakata, M. Tahara, S. Kutsuna, N. Ohmagari, M. Kuroda, T. Suzuki, T. Kageyama, M. Takeda, Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7001-7003 (2020).
17. D. R. Kuritzkes, Drug resistance in HIV-1. *Curr. Opin. Virol.* **1**, 582-589 (2011).
18. H. Zhou, Y. Fang, T. Xu, W. J. Ni, A. Z. Shen, X. M. Meng, Potential therapeutic targets and promising drugs for combating SARS-CoV-2. *Br. J. Pharmacol.* **177**, 3147-3161 (2020).
19. L. W. Shen, H. J. Mao, Y. L. Wu, Y. Tanaka, W. Zhang, TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. *Biochimie* **142**, 1-10 (2017).
20. M. Kawase, K. Shirato, L. van der Hoek, F. Taguchi, S. Matsuyama, Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *J. Virol.* **86**, 6537-6545 (2012).
21. M. Yamamoto, S. Matsuyama, X. Li, M. Takeda, Y. Kawaguchi, J. Inoue, Z. Matsuda, 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. *Antimicrob. Agents Chemother.* **60**, 6532-6539 (2016).
22. M. Yamamoto, M. Kiso, Y. Sakai-Tagawa, K. Iwatsuki-Horimoto, M. Imai, M. Takeda, N. Kinoshita, N. Ohmagari, J. Gohda, K. Semba, Z. Matsuda, Y. Kawaguchi, Y. Kawaoka, J. I. Inoue, The anticoagulant Nafamostat potently inhibits SARS-CoV-2 S protein-mediated fusion in a cell fusion assay system and viral infection in vitro in a cell-type-dependent manner. *Viruses* **12**, 629 (2020).
23. D. L. McKee, A. Sternberg, U. Stange, S. Laufer, C. Naujokat, Candidate drugs against SARS-CoV-2 and COVID-19. *Pharmacol. Res.* **157**, 104859 (2020).
24. R. K. Guy, R. S. DiPaola, F. Romanelli, R. E. Dutch, Rapid repurposing of drugs for COVID-19. *Science* **368**, 829-830 (2020).
25. M. Yamaya, H. Nishimura, X. Deng, A. Kikuchi, R. Nagatomi, Protease inhibitors: Candidate drugs to inhibit severe acute respiratory syndrome coronavirus 2 replication. *Tohoku J. Exp. Med.* **251**, 27-30 (2020).
26. D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074-D1082 (2018).
27. B. Lin, C. Ferguson, J. T. White, S. Wang, R. Vessella, L. D. True, L. Hood, P. S. Nelson, Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. *Cancer Res.* **59**, 4180-4184 (1999).
28. R. P. Vivek-Ananth, A. Rana, N. Rajan, H. S. Biswal, A. Samal, In silico identification of potential natural product inhibitors of human proteases key to SARS-CoV-2 infection. *Molecules* **25**, (2020).

29. D. E. Afar, I. Vivanco, R. S. Hubert, J. Kuo, E. Chen, D. C. Saffran, A. B. Raitano, A. Jakobovits, Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. *Cancer Res.* **61**, 1686-1692 (2001).
30. R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis, Y. Shan, H. Xu, D. E. Shaw, Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13118-13123 (2011).
31. Q. Liu, A. Herrmann, Q. Huang, Surface binding energy landscapes affect phosphodiesterase isoform-specific inhibitor selectivity. *Comput. Struct. Biotechnol. J.* **17**, 101-109 (2019).
32. R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell, A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **28**, 1145-1152 (2007).
33. G. Spraggon, M. Hornsby, A. Shipway, D. C. Tully, B. Bursulaya, H. Danahay, J. L. Harris, S. A. Lesley, Active site conformational changes of prostatin provide a new mechanism of protease regulation by divalent cations. *Protein Sci.* **18**, 1081-1094 (2009).
34. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586-3616 (1998).
35. D. J. Price, C. L. Brooks, A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* **121**, 10096-10103 (2004).
36. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690 (1984).
37. M. Parrinello, A. Rahman, Polymorphic transitions in single-crystals - a new molecular-dynamics method. *J. Appl. Phys.* **52**, 7182-7190 (1981).
38. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463-1472 (1997).
39. T. Darden, D. York, L. Pedersen, Particle Mesh Ewald - an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092 (1993).
40. J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157-1174 (2004).
41. A. W. Sousa da Silva, W. F. Vranken, ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **5**, 367 (2012).
42. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785-2791 (2009).

## Acknowledgments

We thank Dr. Yi Wang for helpful discussion on the project. Computational resources were provided by the Shanghai Supercomputer Center.

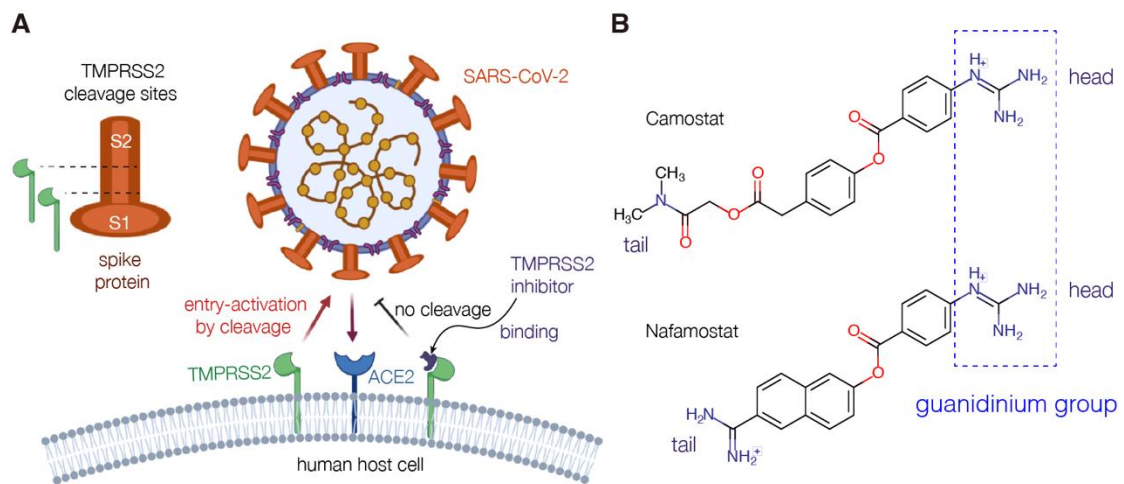
**Funding:** This work was partially supported by the grants from the National Major Scientific and Technological Special Project for ‘Significant New Drugs Development’ (2018ZX09J18112), the National Natural Science Foundation of China (91430112, 31671386 and 31971377), the Shanghai Municipal Science and Technology Major Project (2018SHZDZX01), and Corona Virus Pre Exploration Project by Berlin University Alliance GC2 Global Health (to A.H.).

**Author contributions:** Q.H. and A.H. initialized the project; Q.H. supervised the study; H.Z., W.D. and M.S. prepared the simulation systems; H.Z. and W.D. performed the simulations; W.D., H.Z., Q.L., A.H. and Q.H. carried out the data analysis; W.D. and Q. H. generated the figures and movies; H.Z., W.D., A.H. and Q.H. wrote the manuscript.

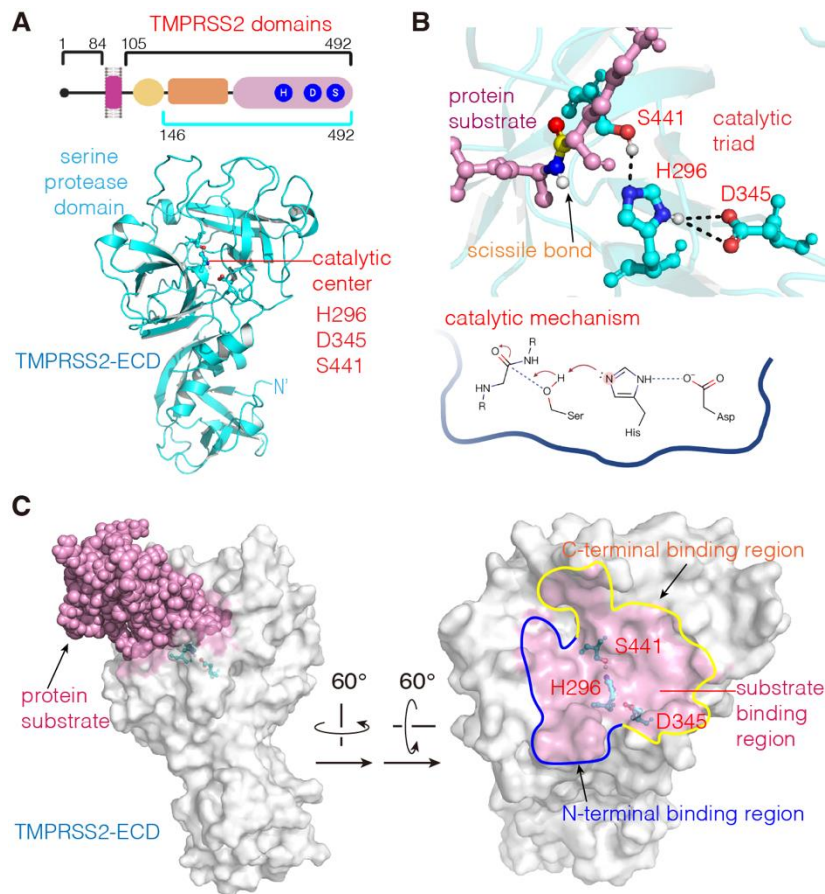
**Competing interests:** The authors declare that they have no competing interests.

**Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

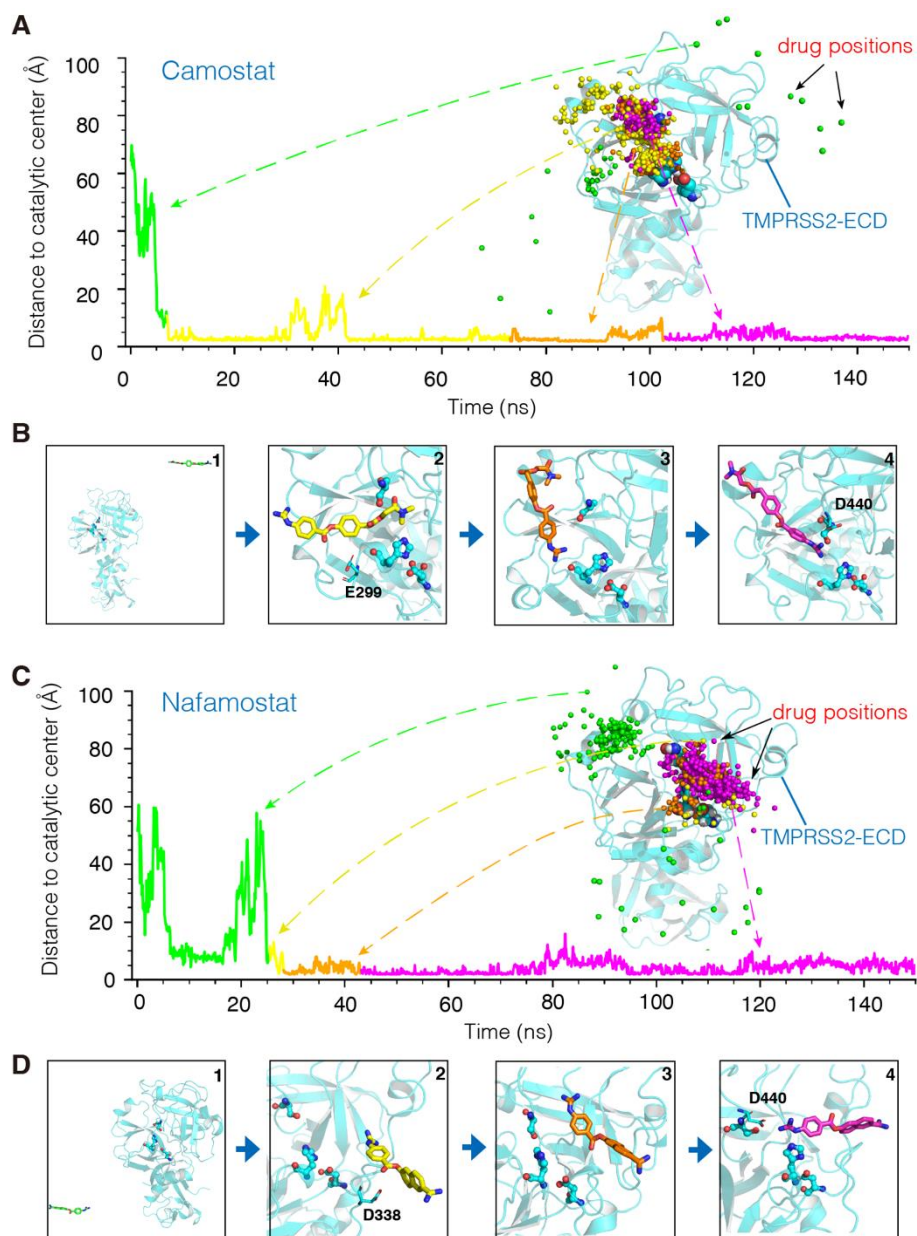
## Figures



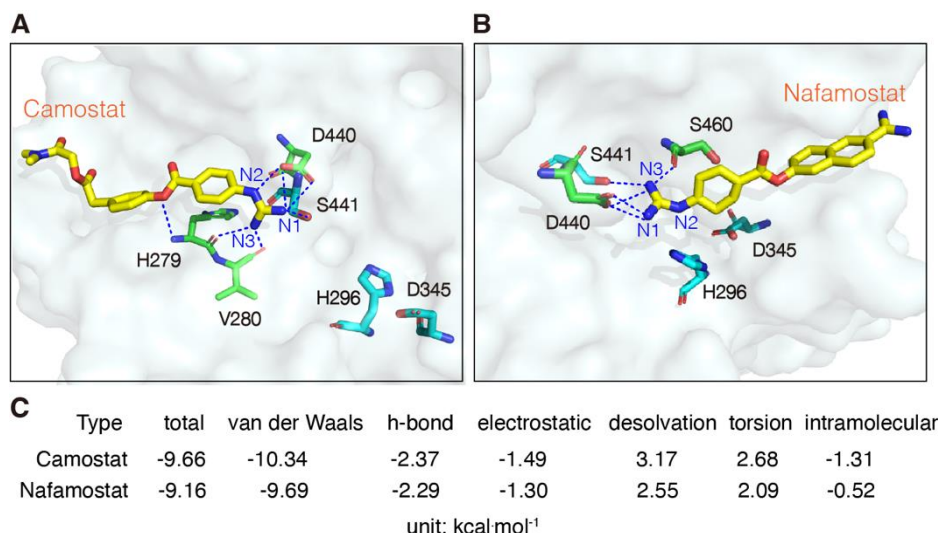
**Fig. 1. TMPRSS2 activation of SARS-CoV-2 entry into human host cell and two potential COVID-19 drugs targeting TMPRSS2.** (A) The schematic diagram of TMPRSS2 activation of SARS-CoV-2 into the host cell. The spike protein of SARS-CoV2 is cleaved by human proteases at the S1/S2 boundary and/or within the S2 subunit with conserved Arginine residues. TMPRSS2 inhibitors could block the entry activating process. (B) The chemical structures of two potential COVID-19 drugs that inhibit the cleavage of the spike protein by TMPRSS2: Camostat and Nafamostat. The guanidinium group of the drug (Camostat or Nafamostat) is defined as its head, and the other terminus as its tail.



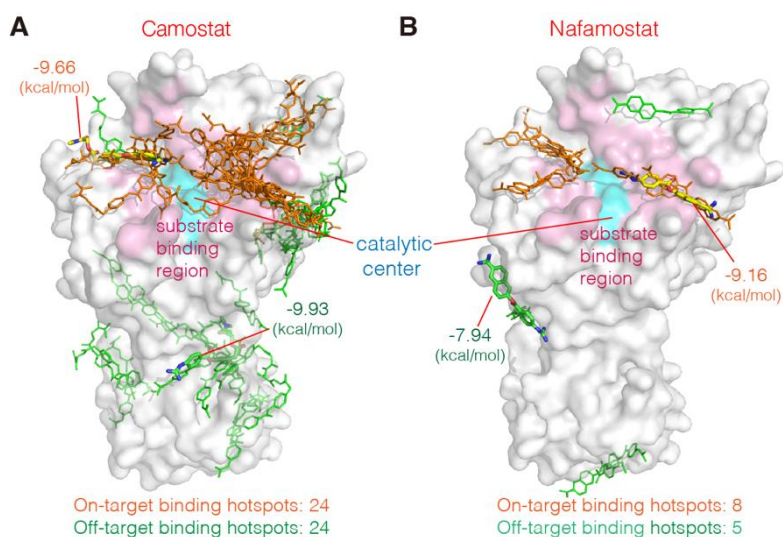
**Fig. 2. The 3D structure of TMPRSS2-ECD and its catalytic center.** (A) The 3D atomic model of TMPRSS2-ECD constructed via the Swiss-Model server. (B) The catalytic center with catalytic triad consisting of H296, A345 and S441, and corresponding catalytic mechanism for cleaving the protein substrate. (C) The substrate-binding region. TMPRSS2-ECD is represented by the surface model (white) and the substrate is shown as pink spheres. The substrate-binding region in the right panel (pink) is defined as TMPRSS2 atoms being in a distance to the substrate  $< 3 \text{ \AA}$ . The N- and C-terminal binding regions are enclosed by the curves in blue and yellow, respectively.



**Fig. 3. Spontaneous binding processes to the catalytic center of TMPRSS2.** (A) Time-dependent Camostat distance to the catalytic center ( $D_{cc}$ ). TMPRSS2-ECD is represented by the cartoon in cyan, and corresponding drug positions represented by the drug atoms closest to the catalytic center (spheres in colors). See also Movie S1 in Supplementary Materials. (B) Representative conformations of bound Camostat. (C) Time-dependent Nafamostat distance to the catalytic center ( $D_{cc}$ ). TMPRSS2-ECD is represented by the cartoon in cyan, and corresponding drug positions represented by the drug atoms closest to the catalytic center (spheres in colors). See also Movie S2 in Supplementary Materials. (D) Representative conformations of bound Nafamostat.



**Fig. 4.** The binding poses of drugs at the catalytic center with the lowest binding free energy. (A and B) The binding poses of Camostat and Nafamostat, respectively. Hydrogen bonds are shown as dotted lines. (C) Intramolecular and intermolecular terms of the lowest binding free energies corresponding to the binding poses in A (Camostat) and B (Nafamostat).



**Fig. 5.** The high-affinity binding hotspots of two drugs on TMPRSS2 revealed in the MD simulations. (A) The binding hotspots of Camostat. (B) The binding hotspots of Nafamostat. Each hotspot is represented by line model of the lowest-energy pose of corresponding conformational cluster of the drug (Camostat or Nafamostat). We defined the hotspots contacting the substrate-binding region as the on-target binding hotspots (orange), while the others as the off-target binding hotspots (green). Corresponding hotspots with the lowest binding energies are indicated by stick models with the energy values.

**Table**

**Table 1. Summary of spontaneous binding simulations.**

Drug	Number of independent simulations	Trajectory number of drugs bound to catalytic center	Number of binding from the N-terminal binding region	Number of binding from the C-terminal binding region	The lowest binding energy at catalytic center (kcal·mol <sup>-1</sup> )
Camostat	15	6	4	2	-9.66
Nafamostat	15	6	2	4	-9.16



# Supplementary Materials

## Spontaneous binding of potential COVID-19 drugs to human serine protease TMPRSS2

Haixia Zhu<sup>1, †</sup>, Wenhao Du<sup>1, †</sup>, Menghua Song<sup>1</sup>, Qing Liu<sup>2</sup>, Andreas Herrmann<sup>3</sup> and Qiang Huang<sup>1, 4, \*</sup>

<sup>1</sup>State Key Laboratory of Genetic Engineering, Shanghai Engineering Research Center of Industrial Microorganisms, MOE Engineering Research Center of Gene Technology, School of Life Sciences, Fudan University, Shanghai 200438, China

<sup>2</sup>State Key Laboratory of Quality Research in Chinese Medicines, School of Pharmacy, Macau University of Science and Technology, Macau, China

<sup>3</sup>Institute for Biology and IRI Lifesciences, Humboldt-Universität zu Berlin, 10115 Berlin, Germany

<sup>4</sup>Multiscale Research Institute of Complex Systems, Fudan University, Shanghai 201203, China

<sup>†</sup>These first two authors contributed equally to this work.

\*Corresponding author. Email: huangqiang@fudan.edu.cn

### Table of Content

Fig. S1. Amino-acid sequence alignment of TMPRSS2-ECD and serine protease hepsin.

Fig. S2. Time-dependent RMSDs of the TMPRSS2-ECD model in the MD refinement.

Fig. S3. Structural alignment of the catalytic amino acids of TMPRSS2 and trypsin.

Fig. S4. A system for the spontaneous binding simulations.

Fig. S5. Time-dependent drug distances to the catalytic center of TMPRSS2 ( $D_{cc}$ ).

Fig. S6. 3D distributions of the Aps/Glu residues and oxygen atoms in/around the substrate-binding cavity of TMPRSS2.

Fig. S7. Binding free energies of the drug-TMPRSS2 complexes formed in the spontaneous binding simulations in Fig. 3.

Movie S1. A typical spontaneous binding trajectory of Camostat to TMPRSS2.

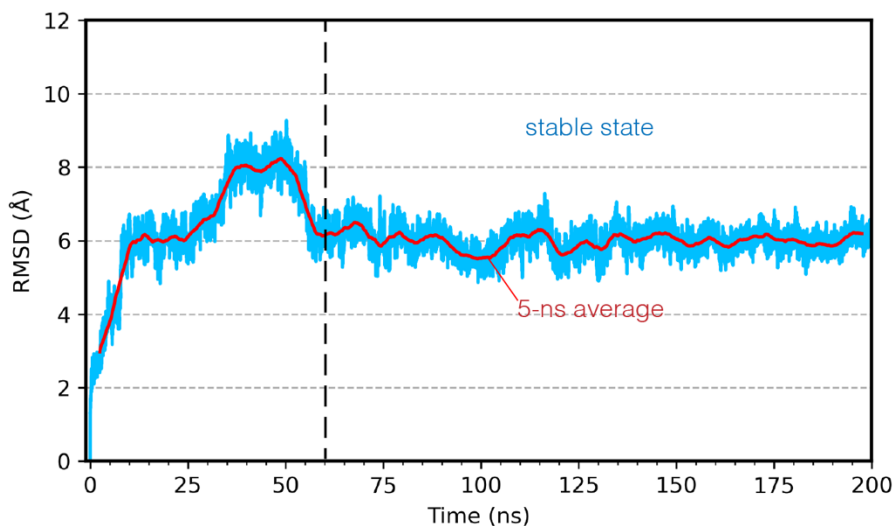
Movie S2. A typical spontaneous binding trajectory of Nafamostat to TMPRSS2.

```

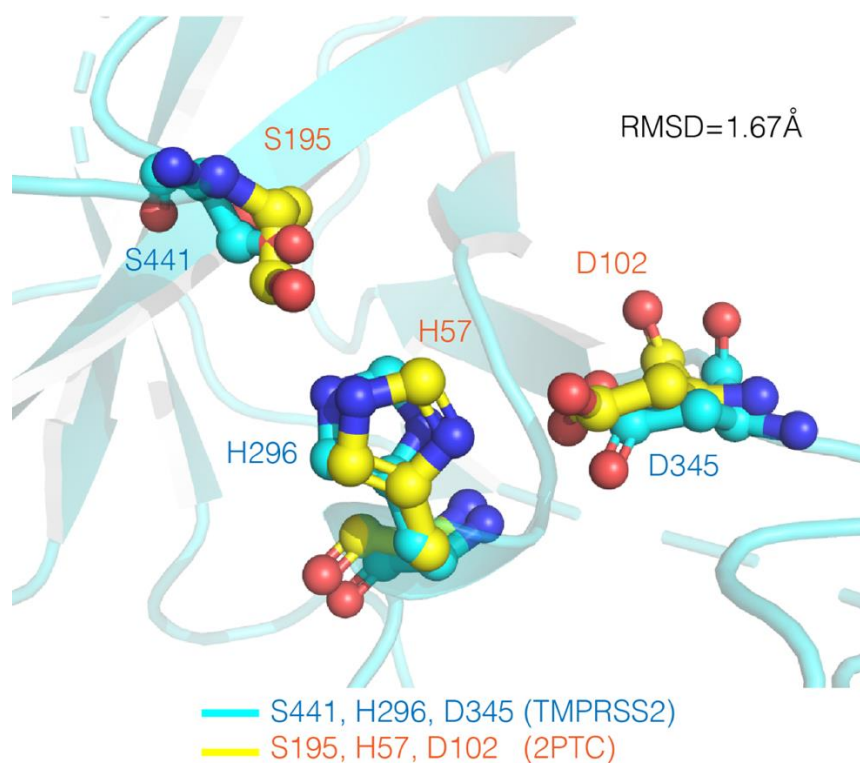
TMPRSS2  -----NRCVRLYGPNFILQVYSSQRKSWHPVCQDDWNEHYGRAACRDMGYKNNFYSSQGI 200
1Z8G      SDQEPLYPVQVSSADARLMVFDKTEGTWRLLCSSRSNARVAGLSCEEMGFLRALTHSELD 60
           *:: . : * *:: . : *:: *:: . : * . . : *:: *:: . : *::
TMPRSS2  VDDSG---STSFMKLNLSAGNVDI---YKKLYHSDACSSKAVVSLRCIACGVNLSNRQS 254
1Z8G      VRTAGAAGTSGFFCVDE--GRLPHTQRLLEVISVCDPCRGRFLAaicQDCGRR--KLPVD 116
           * : * : : * : : : * : : : : * . : : * * * . . .
TMPRSS2  RIVGGESALPGAWPWQVSLHVQNVHVCSSIIITPEWIVTAAHCVEKPLNPNWHWTAFAGI 314
1Z8G      RIVGGRDTSIGRWPWQVSLRYDGAHLCCGSLSGDWVLTAAHCFFERNRVLRSRWRVFAGA 176
           *****: * *****: : : : * : : : : : : * : : * : * *
TMPRSS2  LRQSFMFYGAGYQVEKVIS-----HPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPN 368
1Z8G      VAQASPH-GLQLGVQAVVYHGGYLPRDPNSEENSNDIALVHLSPLPLTEYIQPVCLPA 235
           : * : . * : * : : : : * : : * : : * : : * : : * : : * : *
TMPRSS2  PGMLLQPEQLCWISGWGATEEKGKTSEVLNAAKVLLETQRCNSRYVYDNLITPAMICAG 428
1Z8G      AGQALVDGKICTVTGWGNTQYYGQAGVLEARVPIISNDVCNGADFYGNQIKPKMFCAG 295
           * * : : * : : * : * : : * : * : * : * : : * : * : * : * : *
TMPRSS2  FLQGNVDSQCQDSSGGPLVTS----KNNIWWLIGDTSWGSACAAYRPGVYGNVMVFTDWI 484
1Z8G      YPEGIDACQDSSGGPFVCEDSISRTPRWRLCGIVSWGTCALAQKPGVYTKVSDFREWI 355
           : : * : * : * : * : * : * . : . * * * . * : * : * : * : * : *
TMPRSS2  YRQMRADG----- 492
1Z8G      FQAIKTHSEASGMVTQL 372
           : : : : : .

```

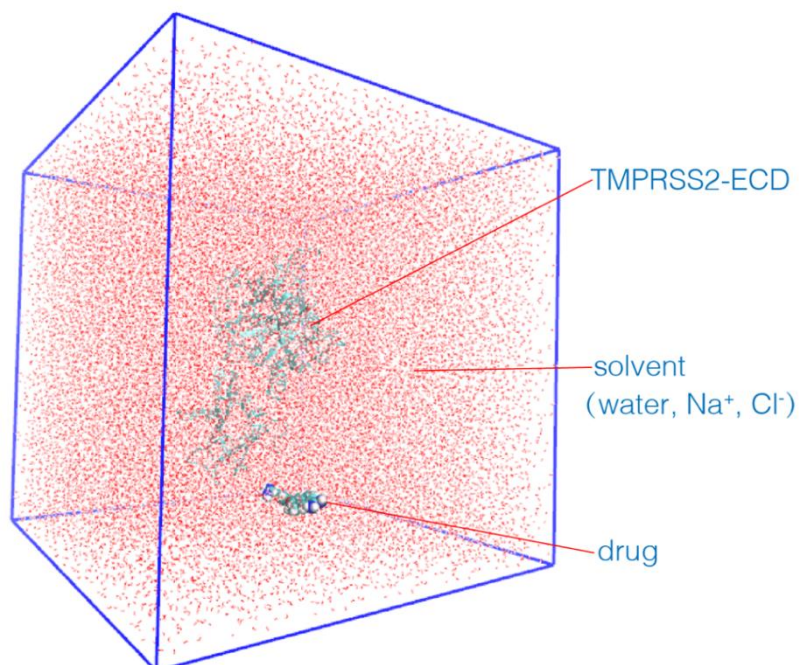
**Fig. S1. Amino-acid sequence alignment of TMPRSS2-ECD and serine protease hepsin.** The sequence of the serine protease is taken from PDB ID: 1Z8G. Conserved residues are indicated by a (\*), strongly similar residues by a (:), and weakly similar residues by a (.). The color codes of the residues are: basic, DE, red; acidic, KR, pink; polar, CGHNQSTY, green and hydrophobic, AFILMPVW, red.



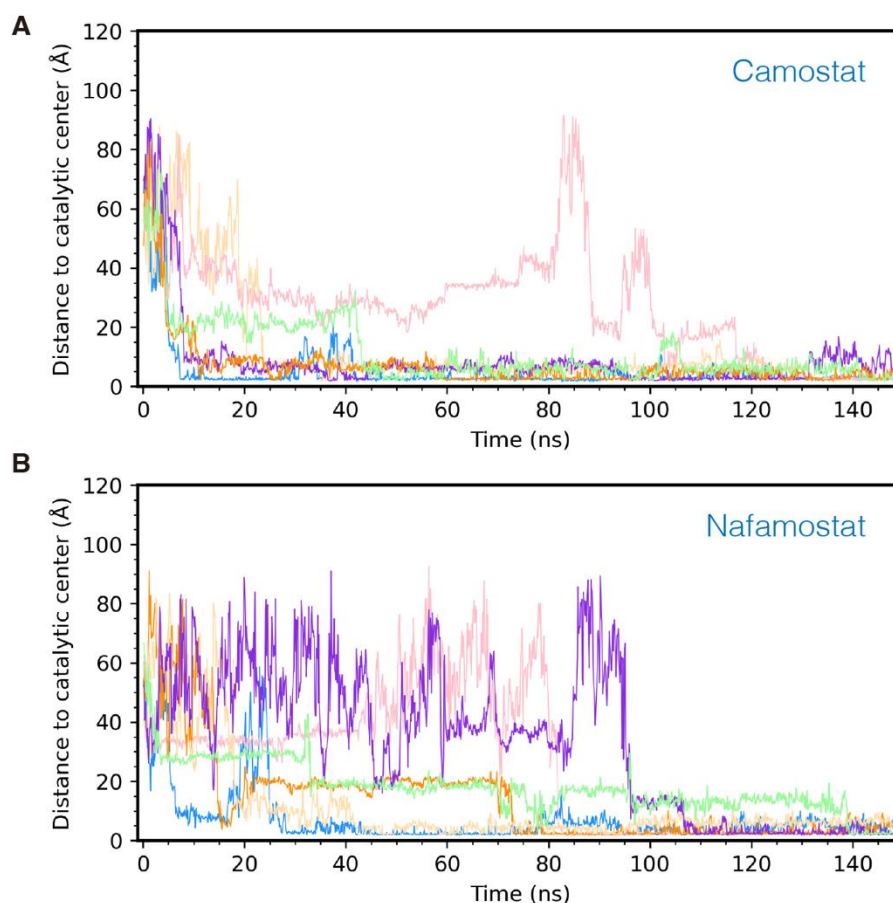
**Fig. S2. Time-dependent RMSDs of the TMPRSS2-ECD model in the MD refinement.** The initial structure of TMPRSS2-ECD is the reference structure for the RMSD calculations.



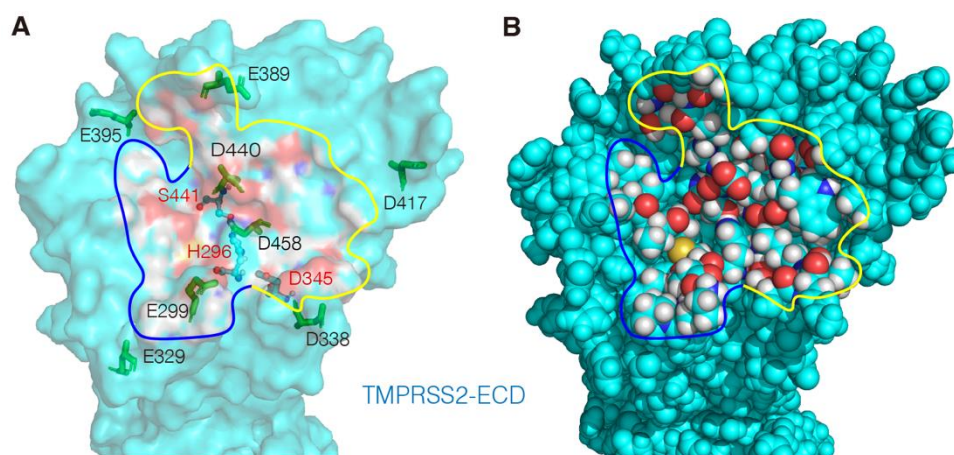
**Fig. S3. Structural alignment of the catalytic amino acids of TMPRSS2 and trypsin.** The PDB ID of the trypsin structure is 2PTC (yellow). The RMSD of the catalytic triads of two structures is 1.67 Å (where TMPRSS2 in cyan, and trypsin in yellow).



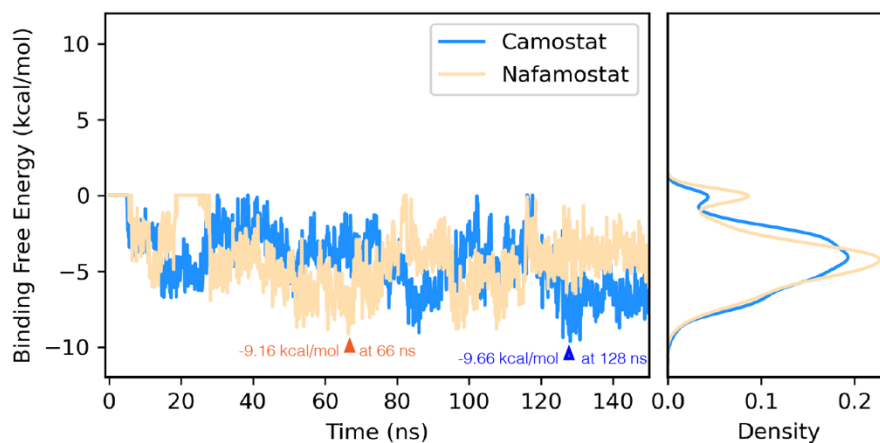
**Fig. S4. A system for the spontaneous binding simulations.** TMPRSS2-ECD is represented by the cartoon in cyan, the drug molecule (Camostat or Nafamostat) is represented by the spheres in cyan, and the solvent (water, Na<sup>+</sup> and Cl<sup>-</sup>) are represented by the sticks in red.



**Fig. S5.** Time-dependent drug distances to the catalytic center of TMPRSS2 ( $D_{cc}$ ). (A) Camostat  $D_{cc}$  in its 6 successful binding trajectories of 15 independent simulations. (B) Nafamostat  $D_{cc}$  in its 6 successful binding trajectories of 15 independent simulations.



**Fig. S6.** 3D distributions of the Aps/Glu residues and oxygen atoms in/around the substrate-binding cavity of TMPRSS2. (A) The acidic amino acids (Asp/Glu) in/around the N-terminal and C-terminal binding regions are shown as sticks in green. (B) The oxygen atoms in the binding regions are shown as spheres in red.



**Fig. S7. Binding free energies of the drug-TMPRSS2 complexes formed in the spontaneous binding simulations in Fig. 3.** The drug-TMPRSS2 complex conformations with the lowest binding free energy appear at ~128 ns (Camostat) and ~66 ns (Nafamostat), respectively.

## Movies

### Movie S1. A typical spontaneous binding trajectory of Camostat to TMPRSS2.

Please see the MPEG4 file: *camo\_binding.mp4*. For the sake of clarity, solvent molecules and ions are removed.

### Movie S2. A typical spontaneous binding trajectory of Nafamostat to TMPRSS2.

Please see the MPEG4 file: *nafa\_binding.mp4*. For the sake of clarity, solvent molecules and ions are removed.