Molecular mechanism of inhibiting the SARS-CoV-2 cell entry facilitator TMPRSS2 with camostat and nafamostat†

Tim Hempel, Lluís Raich, Simon Olsson, Nurit P. Azouz, Andrea M. Klingler, Markus Hoffmann, Stefan Pohlmann, Marc E. Rothenberg and Frank Noé*

The entry of the coronavirus SARS-CoV-2 into human lung cells can be inhibited by the approved drugs camostat and nafamostat. Here we elucidate the molecular mechanism of these drugs by combining experiments and simulations. In vitro assays confirm that both drugs inhibit the human protein TMPRSS2, a SARS-CoV-2 spike protein activator. As no experimental structure is available, we provide a model of the TMPRSS2 equilibrium structure and its fluctuations by relaxing an initial homology structure with extensive 330 microseconds of all-atom molecular dynamics (MD) and Markov modeling. Through Markov modeling, we describe the binding process of both drugs and a metabolic product of camostat (GBPA) to TMPRSS2, reaching a Michaelis complex (MC) state, which precedes the formation of a long-lived covalent inhibitory state. We find that nafamostat has a higher MC population than camostat and GBPA, suggesting that nafamostat is more readily available to form the stable covalent enzyme–substrate intermediate, effectively explaining its high potency. This model is backed by our in vitro experiments and consistent with previous virus cell entry assays. Our TMPRSS2–drug structures are made public to guide the design of more potent and specific inhibitors.

1 Introduction

In December 2019 several cases of unusual and severe pneumonia were reported in the city of Wuhan, China. These cases were traced back to a new coronavirus, SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2); the disease is called COVID-19.1 As of October 11, 2020 there are over 37 million confirmed COVID-19 cases and more than 1 million deaths,2 with both numbers likely to be severe underestimates. Given estimates of the infection mortality rate of 0.4 to 1.4% (ref. 3–5) the virus has the potential to kill tens of millions of people unless efficient vaccines or drugs are available.

As other coronaviruses,6–9 SARS-CoV-2 exploits host proteins to initiate cell-entry, in particular TMPRSS2 and ACE2, two membrane-bound proteins expressed in the upper and lower respiratory tract.10–13 TMPRSS2 contains an extracellular trypsin-like serine-protease domain that can proteolytically activate the spike (S) protein on the surface of SARS-CoV-2 viral particles14 (Fig. 1). While in certain cell lines, the S-protein can also be activated by the endo/lysosomal pH-dependent cysteine protease cathepsin L,14,15 virus entry into human airway cells16 seems to depend on TMPRSS2 but not cathepsin L. Consistently, epidemiological data of prostate cancer patients...
undergoing androgen-deprivation therapies, which lowers TMPRSS2 levels, indicate a lower risk of contracting the SARS-CoV-2 infection.\textsuperscript{17} We further note that low concentration levels of TMPRSS2 are observed in children and infants, possibly explaining lower risks of severe COVID-19 infections in younger age groups.\textsuperscript{18}

TMPRSS2 is also exploited by other coronaviruses and influenza A viruses for activation of surface glycoproteins, viral spread, and pathogenesis.\textsuperscript{19–25} TMPRSS2 knock-out mice have no phenotype in the absence of infection,\textsuperscript{26} indicating that inhibiting TMPRSS2 function might not be associated with substantial unwanted side effects. As a result, TMPRSS2 is a promising therapeutic target in the context of influenza A and coronavirus infection, including SARS-CoV-2. Since TMPRSS2 is host encoded and thus genetically stable, treatment should be associated with a low risk of drug resistance.

Here, we study the structural basis and molecular mechanisms of TMPRSS2 inhibition by nafamostat, camostat, and its metabolic product 4-[4-guanidinobenzoyloxy]phenylacetic acid (GBP). These guanidinobenzoyl-containing drugs are approved for human use in Japan and have been demonstrated to inhibit SARS-CoV-2 cell-entry.\textsuperscript{14,27–29} A recent survey of FDA approved drugs further found nafamostat to be an effective inhibitor of SARS-CoV-2 infection in human lung-cell cultures.\textsuperscript{30} We report experimental measurements demonstrating that nafamostat and camostat inhibit TMPRSS2 activity by using our recently established cell-based assay,\textsuperscript{31} consistent with \textit{in vitro} enzymatic TMPRSS2 activity assays.\textsuperscript{12}

Despite the hopes associated with TMPRSS2 inhibition, we are, as yet, lacking an experimental structure. We here go beyond the previous dependence on homology models by an extensive 330 microseconds of high-throughput all-atom molecular dynamics (MD) simulations and Markov modeling. This approach provides an ensemble of equilibrium structures of the protein–drug complex and also drug binding kinetics. We show that nafamostat, camostat, and GBP are covalent inhibitors with an identical covalent complex, but their different inhibitory activity can be explained by different populations of their Michaelis complex preceding the covalent complex. These findings, combined with the simulation structures that we make publicly available, provide an important basis for developing more potent and specific TMPRSS2 inhibitors.

2 Results

\textbf{Camostat and nafamostat inhibit the catalytic activity of TMPRSS2}

First we confirm that camostat and nafamostat are TMPRSS2 inhibitors. To this end, we employ our recently reported activity assay\textsuperscript{33} of the full-length TMPRSS2 protein on the surface of live cells with both inhibitors [Fig. 2A]. Briefly, we transfected the human cell-line HEK-293T with a TMPRSS2 expression vector. We then measured the protease activity of the transfected cells using the fluorogenic peptide substrate BOC-QAR-AMC, following incubation of the cell with increasing inhibitor concentrations. Peptide-digestion induced a minimal increase in fluorescent signal in control cells without exogenous TMPRSS2 expression (un-normalized mean enzyme activity = 2.4), while TMPRSS2 over-expression resulted in a much faster peptide digestion (un-normalized mean enzyme activity = 12.8). Therefore, our assay is mostly specific for TMPRSS2.\textsuperscript{34} Significantly lower enzyme activity at higher drug concentrations can thus be attributed to TMPRSS2 inhibition.

For both camostat and nafamostat, we see a clear dose-dependent inhibition and estimate their respective IC50
values to $142 \pm 31 \text{nM}$ and $55 \pm 7 \text{nM}$ (Fig. 2C). Our results are consistent with the finding that both drugs inhibit cell entry of SARS-CoV-2 and other coronaviruses, and that nafamostat is the most potent inhibitor.\textsuperscript{27,28,32}

Note that in humans, camostat is rapidly processed to 4-(4-guanidinobenzoyloxy)phenylacetic acid (GBPA) (Fig. 2A).\textsuperscript{33} It has been recently shown that GBPA also inhibits TMPRSS2 and cell entry of SARS-CoV-2 viruses, although slightly less efficiently than camostat.\textsuperscript{29} Hence, we subsequently study the molecular interactions between TMPRSS2 and all three compounds: camostat, GBPA, and nafamostat.

Equilibrium structures of TMPRSS2 in complex with camostat and nafamostat

We now set off to investigate the molecular mechanism of TMPRSS2 inhibition by nafamostat, camostat, and its metabolite GBPA. No TMPRSS2 crystal structure is available to date, however it has been shown that all-atom MD simulations can reliably model the equilibrium structures of proteins when (i) a reasonable model is available as starting structure, and (ii) simulations sample extensively, such that deficiencies of the starting structure can be overcome.\textsuperscript{35–39}

Here, we initialize our simulations with recent homology models of the TMPRSS2 protease domain and with camostat/nafamostat docked to them.\textsuperscript{40} Trypsins adopt a common fold and share an active-site charge relay system whose structural requirements for catalytic activity are well understood;\textsuperscript{41} we select our MD model consistent with these structural requirements. In particular, we focus on systems with Asp435 (substrate recognition) deprotonated and His296 (catalytic function) in a neutral form ($N_\text{d}$ protonated), as well as on the interactions of a charged lysine nearby the catalytic Asp345 (Fig. S1 and S2).\textsuperscript{†}

In order to avoid artifacts of the initial structural model and to simulate the equilibrium ensemble of the TMPRSS2–drug complexes, we collected a total of 100 $\mu$s of simulation data for TMPRSS2–camostat, 50$\mu$s for TMPRSS2–GBPA, and 180 $\mu$s for TMPRSS2–nafamostat. Every drug dataset has converged RMSD distributions (Fig. S5) and samples various drug poses and multiple association/dissociation events. Using Markov modeling\textsuperscript{42–46} we derive the structures of the long-lived (meta-stable) states and characterize protein–drug binding kinetics and thermodynamics.

We find TMPRSS2 has flexible loops around the binding site but maintains stable structural features shared by other trypsin-like proteases (Fig. 3A and S6). After formation of a non-covalent substrate–enzyme complex (binding step, Fig. 2B), trypsins cleave peptide-like bonds in two catalytic steps, assisted by a conserved catalytic triad (Asp345, His296, and Ser441 in TMPRSS2). The first step involves the formation of a covalent acyl–enzyme intermediate between the substrate and Ser441.\textsuperscript{43} During this step, His296 serves as a general base to deprotonate...
the nucleophilic Ser441, and subsequently as a general acid to protonate the leaving group of the substrate. The second step involves the hydrolysis of the acyl–enzyme intermediate, releasing the cleaved substrate and restoring the active form of the enzyme (Fig. 2B).

Along these two steps, the so-called “oxyanion hole”, formed by the backbone NHs of Gly439 and Ser441, helps to activate and stabilize the carbonyl of the scissile bond. Another important structural feature is the S1 pocket, which contains a well conserved aspartate (Asp435) that is essential for substrate binding and recognition. At the opposite site of the S1 pocket, a loop containing a hydrophobic patch delimits the binding region of substrates within enzymatic active site. All these structural elements, known to play crucial roles in the function of serine proteases, are generally stable and preserved in our equilibrium structures (cf. Fig. S2 and S6†).

Structural basis of TMPRSS2 inhibition by camostat and nafamostat

Drugs with a guanidinobenzoyl moiety can inhibit trypsins by mimicking their natural substrates (Fig. 3B). Indeed, the ester group, resembling a peptide bond, can react with the catalytic serine with rates that are orders of magnitude faster, forming the acyl–enzyme intermediate. In contrast to peptide catalysis, the drug’s guanidinobenzoyl group stays covalently linked to the catalytic serine with a small off-rate, rendering it an effective chemical inhibitor. Note that in its inhibited state, the TMPRSS2 active site is modified such that protease activity is disabled, preventing SARS-CoV-2 S-protein cleavage.

The present MD simulations sample different conformations of the complex formed by the enzyme with each of the drugs that precede the covalent substrate–enzyme complex. We can, therefore, elucidate their binding and how specific interactions stabilize different modes. However, please note that our simulations do not simulate the covalent complex’s formation. All binding modes mimic interactions made between trypsins and their natural substrates, in which lysine heads interact with a conserved aspartate in the S1 pocket (Asp435, Fig. 3B). In camostat, its metabolic product GPBA, and nafamostat, the role of the lysine heads is taken by the guanidinium heads which bind in the S1 pocket and also interact with Asp435 (Fig. 3C–E). However, the guanidinium–Asp435 salt bridge is formed and broken transiently especially for camostat and GBPA (Fig. 3G–I), indicating that these drugs are not optimized for the TMPRSS2 pocket (Fig. S4†).

Nafamostat also binds in a “reverse” orientation where the amidinium head binds into the S1 pocket and interacts with Asp435 (Fig. 3F). In this orientation, the guanidinium head mainly interacts with Glu299, with the drug reactive center slightly displaced from the oxyanion hole, while the “forward” orientation (Fig. 3E) keeps the amidinium head mainly nearby Val280, with the ester center well positioned for the reaction (Fig. S3†). This observation is in agreement with several crystal structures of acyl–enzyme intermediates between different trypsins and guanidinobenzoyl molecules bound to the S1 pocket (e.g. PDBs 2AH4, 3DFL, 1GBT†). There are also “inverse substrates” known to react with rates comparable to the ones of normal esters, suggesting that the inverted nafamostat orientation may also be reactive.41

A fraction of the bound-state structures resembles a reactive Michaelis complex (MC) which fulfills the necessary criteria for catalysis of the inhibitory acyl–enzyme complex: small distances of (i) the drug ester carbon to catalytic serine oxygen, and (ii) the catalytic serine hydrogen to catalytic histidine nitrogen (Methods). We observe that besides Asp435 binding to the S1 pocket, drugs in the MC state are particularly stabilized by the oxyanion hole. Our model predicts that nafamostat has the highest MC state population followed by camostat and GBPA (Fig. 4), an order that coincides with the one of experimental drug binding affinities.29 We note that the relative free energies of binding to the MC states are significantly different between nafamostat (2.1 ± 0.1 kcal mol⁻¹) and the other drugs (2.8 ± 0.1 kcal mol⁻¹ and 3.1 ± 0.2 kcal mol⁻¹ for camostat and GBPA, respectively), with the bootstrap sample distributions of camostat and its metabolite displaying a partial overlap.

Whereas the contact patterns of camostat and nafamostat associated states are similar, the leaving group in the inverted nafamostat conformation shows contacts predominantly with residues E299 and Tyr337 (Fig. S3†). GBPA, due to its shorter length, has less contacts to residues outside of the S1 pocket. In the reactive MC state, interestingly, all tested drugs display similar contact patterns overall, and their leaving groups bind in between Val280 and His296, with their ester group in contact with Ser441 (Fig. S3†).

Kinetic mechanism of TMPRSS2 inhibition by camostat, GBPA, and nafamostat

Finally, we investigate the molecular basis for the greater inhibition by nafamostat and formulate starting points for designing new and more efficient covalent TMPRSS2 inhibitors following these leads.

To illustrate the reversible binding of camostat, its product GBPA, and nafamostat to TMPRSS2, we used our Markov models to simulate long time-scale trajectories of 50 µs (Fig. 3G–I). We see a clear correlation between tight inhibitor–Asp435 interactions and contact formation between catalytic serine and the inhibitor ester group, potentially forming a reactive complex. In other words, the binding of reactive drugs in the S1 pocket favors the interactions necessary for a catalytically competent MC.

We estimate the dissociation constants for the non-covalent complex, i.e. the ratio of dissociated state and non-covalent complex populations, to be between 6 and 9 mM for the three drugs. Even though our IC50-measurements include other processes and thus are not straightforward to compare, IC50-values in the 10–100s nanomolar range (i.e. 4–5 orders of magnitude smaller, Fig. 2C) are a strong indicator that the major source of inhibition cannot be the non-covalent complex, but is rather the longer-lived covalent acyl–enzyme complex. However, as all three drugs yield identical acyl–enzyme complexes, the differences in TMPRSS2 inhibition can only
arise from either (1) the formation or population of their MCs, or (2) differences in the catalytic rate $k_{\text{cat}}$ of acylation.

Interestingly, we observe that the MSM-predicted populations of the MCs in nafamostat, camostat, and GBPA have approximate ratios of 6 : 2 : 1, respectively, as well as a significantly higher on-rate for nafamostat (Fig. 4). A simple three-state kinetic model of dissociated state, MC and covalent complex shows that the overall association constant ($K_a$, ratio of inhibited versus apo protein states) directly scales with the association constant of the MC ($K_{\text{M}}$, ratio of MC versus dissociated states) by a constant factor (Methods):

$$K_a = K_{\text{M}} \frac{k_{\text{cat}}}{k_{\text{dis}}}$$

Simply speaking, this indicates that nafamostat is a better inhibitor because it is more often found in the reactive MC state, and is therefore more likely to be attacked by the catalytic serine oxygen and enter the long-lived acyl–enzyme inhibitor complex.

Moreover, we note that the $k_{\text{cat}}$ of acylation of these drugs may depend on their leaving group $pK_a$'s. Leaving groups with a low $pK_a$ will require less assistance from acid catalysis and will be easily displaced by the nucleophilic serine, favoring the formation of the acyl–enzyme intermediate. We expect the leaving group of nafamostat to have a lower $pK_a$ than the one of camostat, following the values of similar molecules such as naphthol (9.57 (ref. 52)) and 4-methylphenol (10.26 (ref. 53)), respectively. Indeed, these comparative insights are backed by computational $pK_a$ predictions for nafamostat (9.17), camostat (9.36), and GBPA (10.02) (Fig. S4†). We note that these predictions are made in aqueous solution, which could differ slightly from the estimates in the enzyme due to the different environment. Nonetheless, we expect the $pK_a$ values to be in the same relative order given that the three compounds have similar contacts with the enzyme in the reactive state (Fig. S3†).

This suggests that the $k_{\text{cat}}$ of acylation will be slightly faster for nafamostat in particular compared to the camostat metabolic product GBPA, further contributing to nafamostat's superior inhibition of TMPRSS2.

3 Discussion

Camostat and nafamostat are promising drug candidates for a COVID-19 treatment strategy. Here we have combined cell-based assays, extensive molecular simulations, and Markov modeling to unravel the molecular action principle of these drugs and provide data that may help to improve them further.
Our binding assays provide evidence that both inhibitors directly act on TMPRSS2 and that nafamostat is more potent compared to camostat, and this qualitative difference is in agreement with complementary in vitro assays on purified protein construct or cell-entry assays.\textsuperscript{27,28} We note that the absolute IC50 values differ between these three assay types, reflecting differences in experimental conditions and which function is being inhibited and measured.

While no crystallographic structure of TMPRSS2 is available, we provide extensive 330 microseconds of all-atom MD simulations starting from a homology model that generate stable equilibrium structure ensembles of the protein–drug complexes. These simulations sample multiple association/dissociation events and various drug poses in the protein active site. Our analyses show that the non-covalent complexes of nafamostat, camostat, and its metabolic product GBPA are relatively short-lived, suggesting that the main inhibitory effect is due to the formation of the long-lived covalent acyl–enzyme complex between the drug’s guanidinobenzoyl moiety and the catalytic serine of TMPRSS2.

Although the MC state is not the main cause of inhibition, its population directly translates into the potency of the inhibitor, as higher MC population corresponds to a higher catalytic rate and therefore yields a larger population of inhibited enzyme. Consistently with the higher potency of nafamostat, it is found to have a threefold more stable MC compared to camostat, and sixfold compared to GBPA. A second contribution may be the $pK_a$ of drug leaving groups, affecting the rate of enzyme acylation.

Our detailed models of the thermodynamic and kinetics of inhibitor binding highlight the bound state’s heterogeneity, with both drugs adopting multiple distinct poses. We note the importance of residue Asp435 in the conserved S1-pocket, which stabilizes the MC state and helps to orient the reactive molecules in a conformation that is suited for catalysis. Nafamostat has two groups that can potentially bind into the S1 pocket, whereas camostat has only one. However, we find that the population of S1 associated states are similar between nafamostat, camostat, and GBPA, suggesting that non-covalent inhibition is likely a minor contribution to the overall inhibition of TMPRSS2.

We conclude that the design of future TMPRSS2 inhibitors with increased potency and specificity should incorporate the following points:

First, stabilizing the non-covalent complex with the TMPRSS2 active site is beneficial for both, covalent and non-covalent inhibitors. As S1 pocket binding is a major contribution to the stability of the non-covalent complex, effective drugs may contain hydrogen bond donors and positively charged moieties that could interact principally with Asp435, but also with different backbone carbonyls of the loops that compose the cavity (e.g. from Trp461 to Gly464).

Second, for covalent inhibitors, we must consider that the catalytic serine is at a distance of around 1.3 nm from Asp435. Thus, the reactive center of an effective drug and its S1-interacting moieties should be within that distance. We note that, even though all three molecules fit well in the overall active site, the guanidinobenzoyl moiety is slightly shorter than the ideal size of the TMPRSS2 cavity (Fig. S4†). We further suggest that a drug should be size-compatible to the hydrophobic patch on the S1 distal site (Fig. 3A and S4†). We speculate that drugs with a large end to end distance and high rigidity may not fit well in the described TMPRSS2 scaffold, and in particular, might be significantly less reactive.

Third, optimizing the $pK_a$ of the drug’s leaving group might be beneficial for improving covalent TMPRSS2 inhibitors. The first step of the reaction would be faster, and the acetyl–enzyme intermediate would accumulate. We note that the deacetylation off-rate must be very low, ideally on the order of magnitude of guanidinobenzoyl moiety containing drugs.

Finally, we make our simulated equilibrium structures of TMPRSS2 in complex with the simulated drugs available, hoping they will be useful to guide future drug discovery efforts.

### 4 Materials and methods

#### TMPRSS2 activity assays

The TMPRSS2 activity assay was described previously.\textsuperscript{24} Briefly, we transfected HEK-293T with a PLX304 plasmid containing the open reading frame (ORF) sequence of TMPRSS2 which encodes for the full length protein (492 amino acids). Control experiments are conducted with PLX304 plasmids.

Eighteen hours later, we replaced the media to either PBS alone or PBS in the presence of varying concentrations of candidate inhibitors camostat and nafamostat. Fifteen minutes later, we added the fluorogenic substrate BOC-QAR-AMC to the wells to induce a measurable signal of enzyme activity. We measured the fluorescent signal immediately after adding the substrate, in 15 minutes intervals for a total time of 150 minutes.\textsuperscript{31} A baseline proteolytic activity of control cells was measured; we hypothesize that this is because of proteolytic cleavage of the substrate by endogenous transmembrane proteases. However, the TMPRSS2 overexpression cells have significantly increased proteolytic activity compared with control cells.\textsuperscript{31}

To validate the exogenous expression of TMPRSS2, we performed western-blot analysis of cell lysates from TMPRSS2 overexpressing cells and control cells. A 60 kDa band was observed in TMPRSS2 overexpressing cells but not in control cells, which is the expected molecular weight of TMPRSS2 protein after post transcriptional modifications, indicating that the target protein has been successfully expressed.

#### IC50 estimation

We used a generalized log-logistic dose–response model

$$f(x, (b, c, d, e)) = c + \frac{d - c}{1 + e^{(\ln(b) - \ln(x))}}$$

with the concentration $x$, $c$ and $d$ representing the lower and upper limits, $b$ steepness of the curve, and $e$ to estimate IC50 values.\textsuperscript{34}

Upper and lower limits were set to the means computed from control experiments with no drug (upper limit) and PLX
plasmid (no TMPRSS2; background noise). We used scipy's curve fitting algorithms to extract the IC50 with error estimates.

Molecular dynamics simulations

MD simulations were run with OpenMM 7.4.0 (ref. 56) using the CHARMM 36 force field (2019 version). Camostat and nafamostat structures were taken from PubChem® with PubChem CIDs 4413 (nafamostat) and 2536 (camostat), respectively, and modeled with the CHARMM general force field (CGenFF v. 4.3). We generated our MD setups with CharmmGULP. We initiate a simulation box of side length 7.5 nm with a NaCl ion concentration of 0.1 mol l⁻¹ at neutral charge and the TIP3P water model. The setups contain 12038 (camostat), 12030 (GBPNA), and 12039 (nafamostat) water molecules, respectively.

We run simulations in the NPT ensemble and keep the temperature at 310 K (physiological temperature) and the pressure at 1 bar. We use a Langevin integrator with 5 fs integration step and heavy hydrogen approximation (4 amu). PME electrostatics, rigid water molecules, and a 1 nm cutoff for non-bonded interactions are used. Simulation times vary between 100 and 500 ns and accumulate to 100 μs (camostat), 50 μs (GBPNA), 180 μs (nafamostat), respectively. Structures were visualized using VMD.

Due to the lack of a crystal structure for TMRPSS2, MD simulations were seeded from a homology model. It is taken from ref. 40, model 3W94 is chosen based on preulsive MD analyses that showed that 3W94 has the most stable catalytic triad configuration (Fig. S1 and S2). The construct includes amino acids 256 to 491 of the full sequence, corresponding to the catalytic chain except for a C-terminal Glycine missing due to homology modeling against a shorter sequence. MD simulations are seeded as follows: equilibrated docking poses (highest scorers of ref. 40) of the ligand were generated in a preursive run using another homology model. We note that the used camostat docking pose resembles the one described by ref. 63. This data set was equilibrated with local energy minimization, 100 ps simulations with 2 fs time steps in NVT and NPT ensemble subsequently. Frames are selected based on a preliminary metastability analysis, protein conformation is constraint to 3W94 homology model using a constraint force minimizing minRMSD. Production run MD simulations are started from these poses, i.e., from the same protein configuration and with 77 (nafamostat) and 60 (camostat) ligand docking poses, respectively. To ensure convergence of sampling statistics, we run multiple adaptive runs of simulations, seeding new simulations with coordinates associated with sparsely sampled states.

We later added the camostat metabolite GBPNA by following the same setup procedure. Due to its similarity to camostat, we seeded production simulations from representative structures of the camostat stage 1 Markov model (described below) using 200 representative structures.

Markov modeling

We model the binding and unbinding rates in a two step procedure using Markov state type models. First, we describe drug unbound and associated states using a hidden Markov model (HMM). Second, we define a reactive state by using distance cutoffs.

In detail, in the first stage we define distance features between drug guanidinium group and TMRPSS2 Asp435 (minimal distance), drug amidinium group and TMRPSS2 Asp435 (minimal distance, nafamostat only). We further use a binary “reactive” distance feature defined by drug ester carbon to catalytic Ser441-OG, and catalytic serine (HG) to catalytic histidine (NE2) and a threshold of 0.35 nm. If both last mentioned distances are below the threshold, both nucleophilic attack of the serine-OG to the drug ester group and proton transfer from serine to histidine are possible, thus defining the reactive state.

We discretize this space into 243 (camostat), 240 (GBPNA), and 490 (nafamostat) states using regular spatial clustering and use an HMM at lag time 5 ns with 5 (camostat, GBPNA) or 8 (nafamostat) hidden states. Nafamostat yields two metastable S1 associated states encoding for both binding directions, camostat/GBPNA a single one, that are defined by being at salt bridge distance to Asp435. We note no significant correlation between the hidden states and the reactive state, i.e. reactivity is not metastable. Also note that in contrast to later modeling stages, reactivity according to this HMM does not necessitate S1 pocket binding. The described HMMs are used to generate the (non-equilibrium) time series presented in Fig. 3G-I. Besides distance to D435, we also show a reactivity coordinate which we define as the mean of (a) drug ester carbon to catalytic serine oxygen and (b) catalytic serine hydrogen to catalytic histidine nitrogen. Reactivity, i.e. when both reactive distances are within range, is indicated with red markers (MC state).

In the second stage, we split the HMM bound states into reactive and non-reactive by combining HMM Viterbi paths and the reactive state trajectories to one single discrete trajectory consisting of 3 states. We define the S1 associated states by filtering the Viterbi paths of the HMM according to S1 association. We use the reactivity trajectories to further bisect the S1 associated state into reactive and non-reactive states, yielding a three state discretization of the drug binding mode. Note that the S1-reactive state is a subset of the reactive state in the stage 1 HMM model.

We estimate a reversible maximum likelihood Markov state model (MSM) from the stage 2 trajectories as described in ref. 45. We report the stationary probability vector as well as transition rates. The latter are approximated using the matrix logarithm approximation of scipy to compute the transition rate matrix $\Gamma$ from the transition probability matrix $T$ using the definition $T = \exp(\Gamma r)$ with the lag time $r$. We found that all reported quantities are converged with respect to the lag time above $r = 500$ ns which was thus chosen as the model lag time. Errors are estimated by bootstrapping validation using a random sample (with replacement) of the stage 2 trajectory data. All MSM/HMM analyses were conducted using the PyEMMA 2 software (version 2.5.7).

Dissociation constants $K_d = \frac{P_{\text{unbound}}}{P_{\text{bound}}}$ from the non-covalent state were estimated from this model and amount to 5.95 mM (4.60, 7.30) for camostat, 8.45 mM (5.81, 11.65) for...
GBPA, and 6.07 mM (5.55, 6.93) for nafamostat (68% confidence intervals).

**Kinetic model**

Simplifying the binding kinetics into a three-state model describing the binding to/dissociation from the Michaelis complex (ligand concentration $c$ and rates $k_{on}$, $k_{off}$), catalytic rate of entering the covalent complex (apo state ($k_{cat}$) and dissociation to the apo state ($k_{dis}$), the kinetics are described by the rate matrix:

$$K = \begin{bmatrix} -ck_{on} & c_{on} & 0 \\ k_{off} & -k_{off} - k_{cat} & k_{cat} \\ k_{dis} & 0 & -k_{dis} \end{bmatrix}$$ (2)

with the (unnormalized) equilibrium distribution

$$\pi = \begin{bmatrix} k_{dis}(k_{off} + k_{cat}) \\ k_{dis} + k_{cat} \\ k_{dis}/k_{cat} \\ 1 \end{bmatrix}$$ (3)

The overall dissociation constant is then:

$$K_d = \frac{\pi_1}{\pi_2 + \pi_3} = \frac{k_{dis}(k_{off} + k_{cat})}{k_{cat}(k_{dis} + k_{cat})}$$ (4)

The non-covalent dissociation constant of the Michaelis complex:

$$K_{d}^{M} = \frac{\pi_1}{\pi_2} = \frac{k_{off}}{c_{on} + k_{cat}}$$ (5)

The dissociation constant scales as:

$$K_{d} = K_{d}^{M} \frac{k_{dis}}{k_{cat} + k_{dis}}$$ (6)

and thus the association constant scales with the stability of the Michaelis complex by a constant factor given by the rates of chemical catalysis and dissociation:

$$K_{a} = K_{d}^{M} \frac{k_{cat} + k_{dis}}{k_{dis}}$$ (7)

**Software and data availability**

Structural ensembles of camostat, GBPA, and nafamostat binding poses are published online at https://github.com/noegroup/tmprss2_structures.

**Author contributions**


**Conflicts of interest**

M. E. R. is a consultant for Pulm One, Spoon Guru, ClostratioBio, Serpin Pharm, Allakos, Celgene, Astra Zeneca, Arena Pharmaceuticals, GlaxoSmithKline, Guidepoint, and Suvretta Capital Management, and has an equity interest in the first five listed, and royalties from reslizumab (Teva Pharmaceuticals), PEESSv2 (Mapi Research Trust) and Utah Date. M. E. R. is an inventor of patents owned by Cincinnati Children’s Hospital.

**Acknowledgements**

We acknowledge financial support from Deutsche Forschungsgemeinschaft DFG (SFB/TRR 186, Project A12), the European Commission (ERC CoG 772230 “ScaleCell”), the Berlin Mathematics Center MATH+ (AA1-6) and the Federal Ministry of Education and Research BMBF (BIFOLD and RAPID Consortium, 01KI1723D). Stefan Pölzmann was supported by DFG (PO 716/11-1) and BMBF (01KI1723D). We are grateful for in-depth discussions with John D. Chodera (MSKCC New York), Matthew D. Hall (NIH), Katarina Elez, Robin Winter, Tuan Le, Moritz Hoffmann (FU Berlin), and the members of the JEDI COVID-19 grand challenge.

**References**

10 C. G. K. Ziegler, et al., SARS-CoV-2 Receptor ACE2 is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and is Detected in Specific Cell Subsets across Tissues, Cell, 2020, 181(5), 1016–1035.

11 S. Lukassen, et al., SARS-CoV-2 Receptor ACE2 and TMPRSS2 are Primarily Expressed in Bronchial Transient Secretory Cells, EMBO J., 2020, 39(10), e105114.


18 B. A. Schuler et al., Age-Determined Expression of Priming Protease TMPRSS2 and Localization of SARS-CoV-2 Infection in the Lung Epithelium, 2020, bioRxiv, DOI: 10.1101/2020.05.22.111187.


22 C. Tarnow, et al., TMPRSS2 is a Host Factor that is Essential for Pneumotropism and Pathogenicity of H7N9 Influenza A Virus in Mice, J. Virol., 2014, 88(9), 4744–4751.


29 M. Hoffmann et al., Camsotstat Mesylate Inhibits SARS-CoV-2 Activation by TMPRSS2-Related Proteases and Its Metabolite GBPA Exerts Antiviral Activity, 2020, bioRxiv, DOI: 10.1101/2020.08.05.237651.


31 N. P. Azouz, A. M. Klingler, and M. E. Rothenberg, Alpha 1 Antitrypsin is an Inhibitor of the SARSCoV2–Priming Protease TMPRSS2, 2020, bioRxiv, DOI: 10.1101/2020.06.23.167544.


57. R. B. Best, et al., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone \( \phi, \psi \) and Side-Chain \( \chi_1 \) and \( \chi_2 \) Dihedral Angles, *J. Chem. Theory Comput.*, 2012, **8**(9), 3257–3273.


