Supporting Information: Molecular mechanism of inhibiting the SARS-CoV-2 cell entry facilitator TMPRSS2 with Camostat and Nafamostat

Tim Hempel[1,2], Lluís Raich[1,†], Simon Olsson[1,3,†], Nurit P. Azouz[4,5], Andrea M. Klingler[4], Markus Hoffmann[6,7], Stefan Pöhlmann[6,7], Marc E. Rothenberg[4], and Frank Noé[1,2,8,*]

[1] *Freie Universität Berlin, Department of Mathematics and Computer Science, Berlin, Germany*
[2] *Freie Universität Berlin, Department of Physics, Berlin, Germany*
[3] *Chalmers University of Technology, Department of Computer Science and Engineering, Göteborg, Sweden*
[4] *Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA*
[5] *Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA*
[6] *Infection Biology Unit, German Primate Center – Leibniz Institute for Primate Research, Göttingen, Germany*
[7] *Faculty of Biology and Psychology, University Göttingen, Göttingen, Germany*
[8] *Rice University, Department of Chemistry, Houston, TX, USA*

† Equal contribution
* Corresponding author: frank.noe@fu-berlin.de

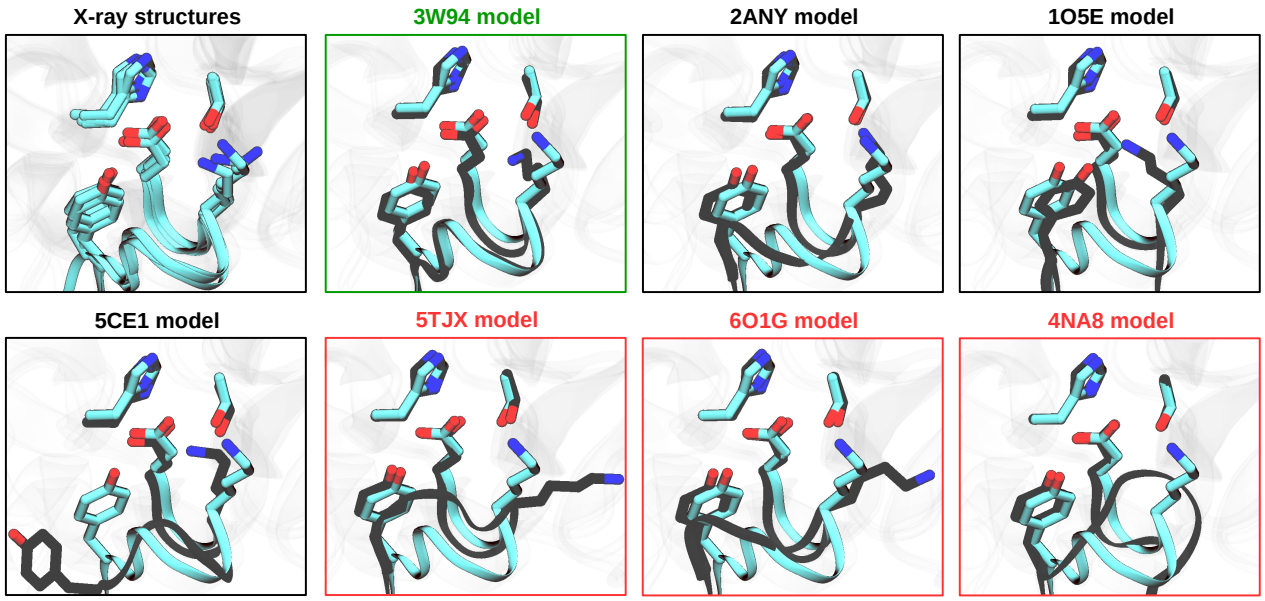

Figure S1: Selection of homology model from Ref. 38 by specific interactions around the catalytic triad. Comparison of TMPRSS2 models (black) and four serine protease structures that contain a lysine residue next to the catalytic aspartate (cyan, PDBs 1EKB, 1FUJ, 3W94 and 4DGJ). Note that the four crystal structures show a very conserved and rigid environment around the catalytic aspartate, with just few fluctuations of the lysine head (K99 in 1EKB). The structural models of TMPRSS2, instead, show a wide variability of conformations of both the backbone and sidechains, with 3W94 being the most conservative model.

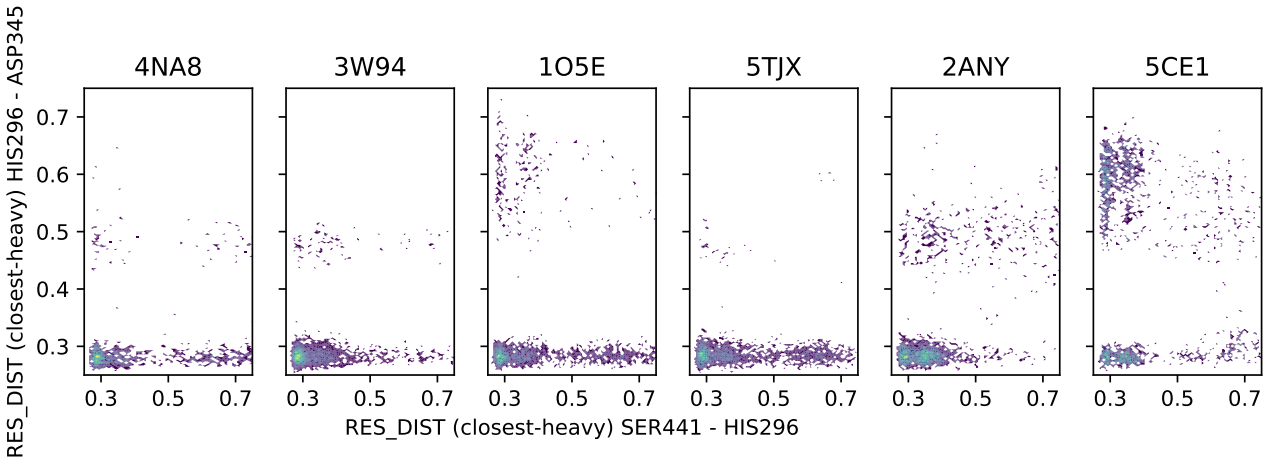

Figure S2: Relevant distances of the catalytic triad for different homology models from Rensi et al. [doi:10.26434/chemrxiv.12009582] as computed from roughly 30μs of MD data for the drug free protein.

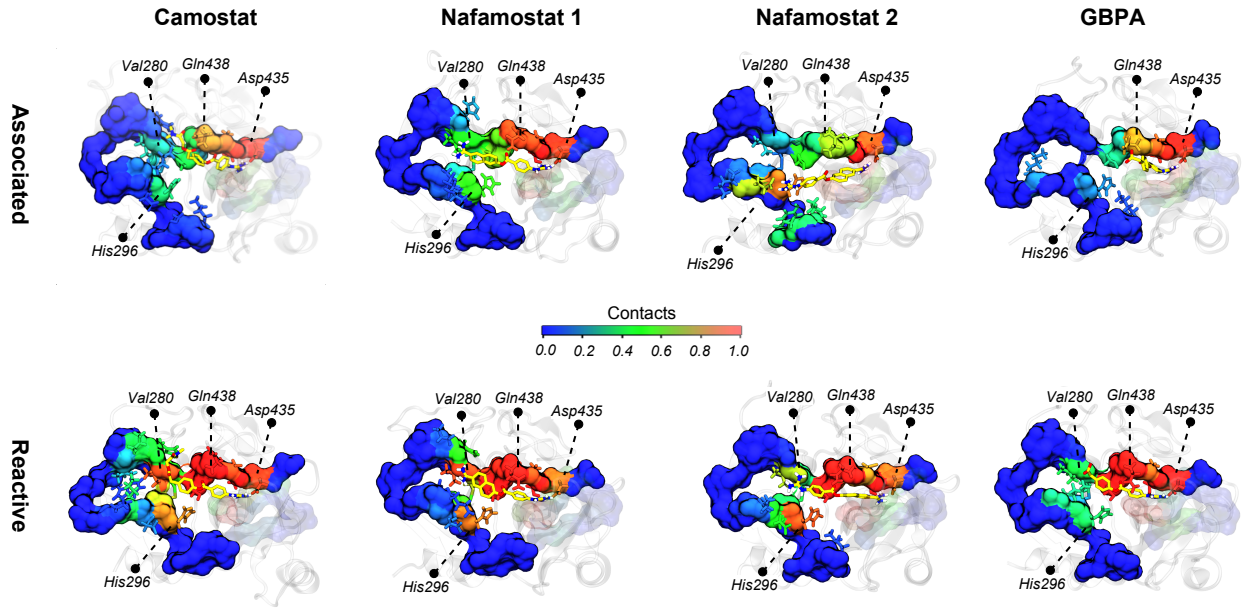

Figure S3: Contact frequency map of TMPRSS2 with camostat, nafamostat, and the camostat product GBPA. Contacts are defined from any atom of the drug to any atom of TMPRSS2 with a distance below 0.35 nm. Residues above with binding frequencies above 0.05 are shown as colored sticks.

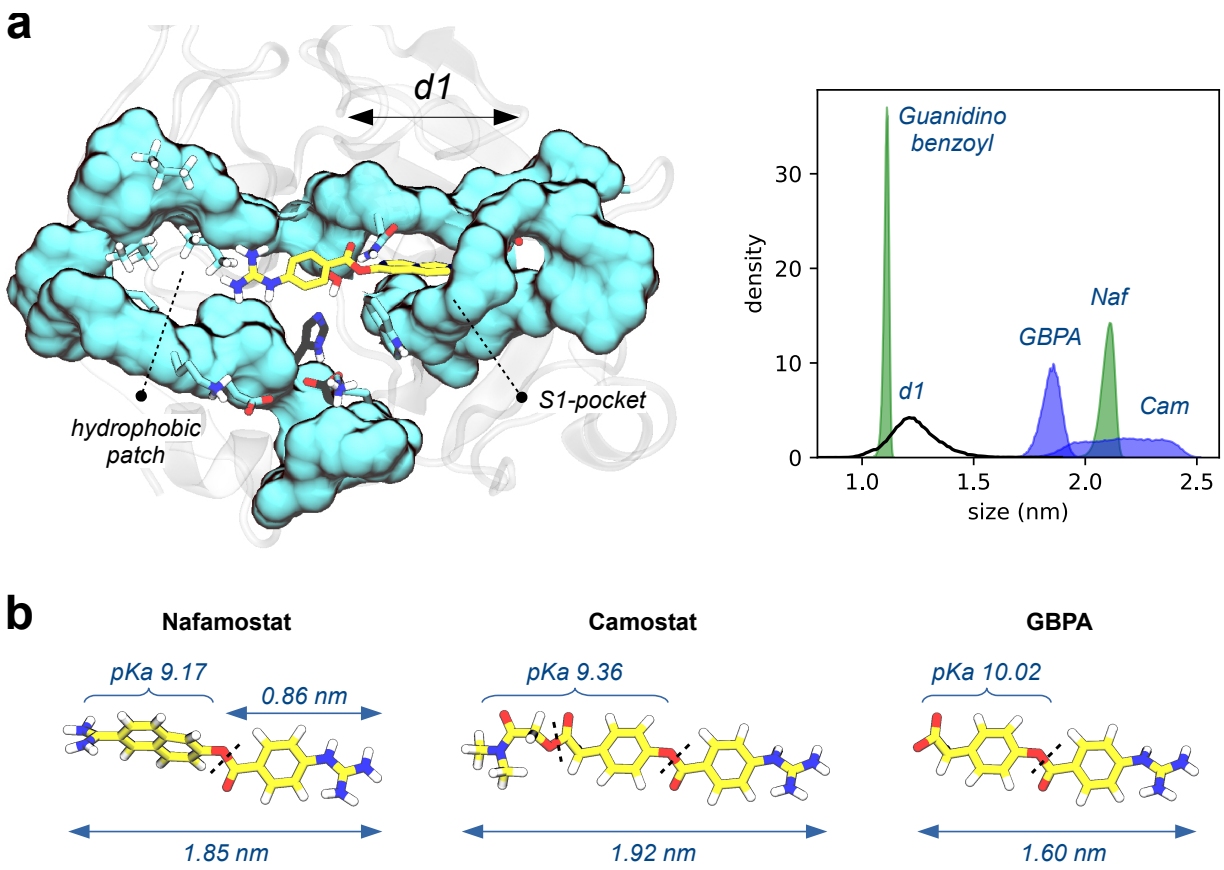

Figure S4: Distribution of TMPRSS2 cavity lengths and end-to-end drug sizes. (a) histogram of drug end-to-end distances and S1 pocket length (d1, minimal distance Ser441(Og)-Asp435(Od1/Od2)). Nafamostat end-to-end distance is show in green, together with the size of the guanidinobenzoyl moiety. Camostat and its metabolized form (GBPA) are shown in blue. Note the wide distribution of camostat, indicating a high degree of flexibility going from compact to extended conformations. Distributions of the drugs are shifted by 0.25 nm to account for the distance of hydrogen bonds. (b) structures of nafamostat, camostat, and GBPA. Note that the three drugs share the same guanidinobenzoyl moiety (right part of each molecule). Hydrolyzable bonds are indicated by dashed lines. Average end-to-end distances by two headed lines. The pKa value of the leaving group (phenol) is shown above each moiety, as predicted by Schrödinger Epik (Schrödinger Release 2020-2).

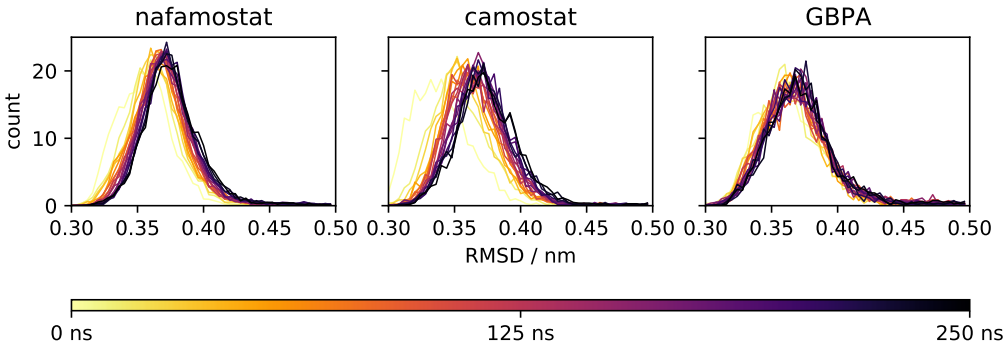

Figure S5: Convergence of root mean squared deviation (RMSD) with trajectory time. As all three datasets include 100s of trajectories, we show the time evolution of RMSD histograms over all available trajectories of a given drug (time as displayed by color code). Please note that the distributions converge over time.

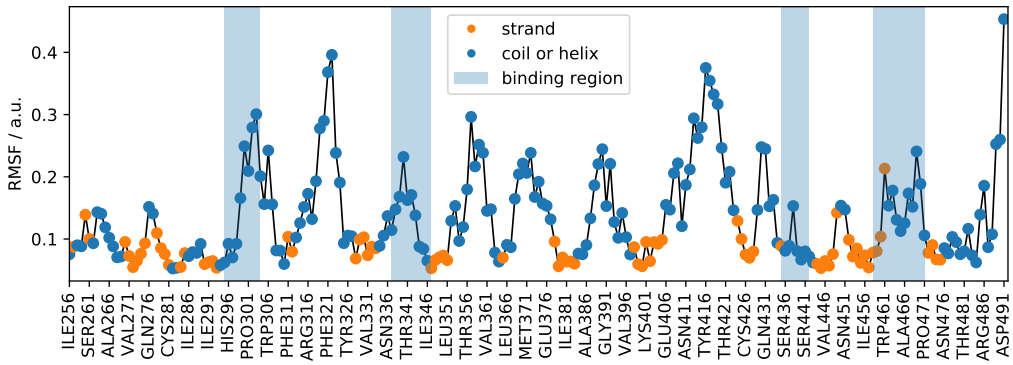

Figure S6: Per-residue root mean squared fluctuations (RMSF) along the protein sequence. The protein core, mainly consisting of β-strands, has a low RMSF, i.e. is rigid compared to coil or helical protein segments such as a large part of the binding region.