Received: 12 April 2023 | Revised: 23 June 2023 | Accepted: 26 June 2023

DOI: 10.1002/prot.26548

RESEARCH ARTICLE

PROTEINS WILEY

Conformational response to ligand binding of TMPRSS2, a protease involved in SARS-CoV-2 infection: Insights through computational modeling

Giorgia Frumenzio[1,2] | Balasubramanian Chandramouli[1] | Neva Besker[3] | Alessandro Grottesi[3] | Carmine Talarico[4] | Francesco Frigerio[5] | Andrew Emerson[1] | Francesco Musiani[2]

[1]Super Computing Applications and Innovation, Department HPC, CINECA, Casalecchio di Reno, Italy

[2]Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

[3]Department HPC, CINECA, Rome, Italy

[4]EXSCALATE – Dompé Farmaceutici SpA, Naples, Italy

[5]Department of Physical Chemistry, Eni SpA, San Donato Milanese, Italy

Correspondence
Balasubramanian Chandramouli, Super Computing Applications and Innovation, Department HPC, CINECA, via Magnanelli 6/3, 40033 Casalecchio di Reno, Italy.
Email: b.chandramouli@cineca.it

Francesco Musiani, Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Viale G. Fanin 40, 40127 Bologna, Italy.
Email: francesco.musiani@unibo.it

Funding information
Ministero dell'Istruzione, dell'Università e della Ricerca (RFO grant 2020); Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine

Abstract

Thanks to the considerable research which has been undertaken in the last few years to improve our understanding of the biology and mechanism of action of SARS-CoV-2, we know how the virus uses its surface spike protein to infect host cells. The transmembrane prosthesis, serine 2 (TMPRSS2) protein, located on the surface of human cells, recognizes the cleavage site in the spike protein, leading to the release of the fusion peptide and entry of the virus into the host cells. Because of its role, TMPRSS2 has been proposed as a drug target to prevent infection by the virus. In this study, we aim to increase our understanding of TMPRSS2 using long scale microsecond atomistic molecular dynamics simulations, focusing on the conformational changes over time. The comparison between simulations conducted on the protein in the native (apo) and inhibited form (holo), has shown that in the holo form the inhibitor stabilizes the catalytic site and induces rearrangements in the extracellular domain of the protein. In turn, it leads to the formation of a new cavity in the vicinity of the ligand binding pocket that is stable in the microsecond time scale. Given the low specificity of known protease inhibitors, these findings suggest a new potential drug target site that can be used to improve TMPRSS2 specific recognition by newly designed inhibitors.

KEYWORDS
membrane proteins, molecular dynamics, Nafamostat, SARS-CoV-2, TMPRSS2 protease

1 | INTRODUCTION

Coronaviruses are enveloped viruses with large (27–31 kb) single-stranded positive-sense RNA genomes.[1,2] Past outbreaks of severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome (MERS) were merely the preamble to the pandemic unleashed by the coronavirus SARS-CoV-2. The latter was initially isolated and identified in December 2019 in Wuhan (China)[3,4] and in the following years it caused a tremendous impact on global health and economic systems. To date (beginning of 2023), the WHO reports that more than 750 million people have been infected and almost 6.8 million have died.[5] SARS-CoV-2 belongs to the

β-coronavirus genus of Coronaviridae family of the order Nidovirales.[6] The virion is defined by four structural proteins: spike, envelope, membrane, and nucleocapsid. The latter covers the genome, which in turn is enclosed by a lipid bilayer formed by the other three proteins. All of these are involved in the mechanism of virus entry into the host cell. The recognition and the binding to the host cell are performed by the spike protein, which binds to the target cell through specific interactions with cellular receptors.[7,8] From a structural point of view, the spike protein is an elongated homotrimer and added to the spherical nucleocapsid gives the virus the typical crown-like appearance. Each subunit of the spike protein is composed of two domains, S1 and S2 (Figure 1A).[7,8]

The S1 domain contains the Receptor-Binding Domain (RBD), which interacts closely with the receptor on the host cell, while the second subunit (S2) is responsible for membrane fusion, an event that allows effective access. The spike protein, however, must be processed before fulfilling its task. This occurs by involving the S1 and S2 subunits, which become the site of cleavages.[8] The first proteolytic cleavage is made at the border between S1 and S2 (S1/S2 site, Figure 1A) and involves a conformational change in the S2 domain separating the two subunits. This operation is made by furin during virus maturation. Another cleavage site, termed S2′, is located in the S2 domain (Figure 1A) and allows the fusion of the viral and cellular membranes to facilitate the release of the N-coated RNA genome into the cytoplasm.[8] Once the spike protein is activated, it interacts with the cellular receptors of the target cell, specifically the angiotensin-converting enzyme 2 (ACE2), which is a carboxypeptidase of 805 amino acids that hydrolyses angiotensin II.[8] There are two ways for the virus to enter the cell, and depending on the route taken, the cleavage on S2′ is made by a different protease. In one of these cases, the transmembrane prosthesis, serine 2 (TMPRSS2) protein is involved, and the cleavage occurs at the cell surface (Figure 1B).[8] The cleavage causes the division of S1 and S2 and the exposure of the fusion peptide, which will be pushed toward the target membrane. The fusion of the two membranes, viral and cellular, leads to the formation of a fusion pore through which the RNA is released into the cytoplasm.[8] TMPRSS2 belongs to the type 2 transmembrane serine protease (TTSP) family that is constituted by 19 surface-expressed trypsin-like serine proteases.[9] These proteins participate in the remodeling of the extracellular matrix,[10] in the proteolytic activation of

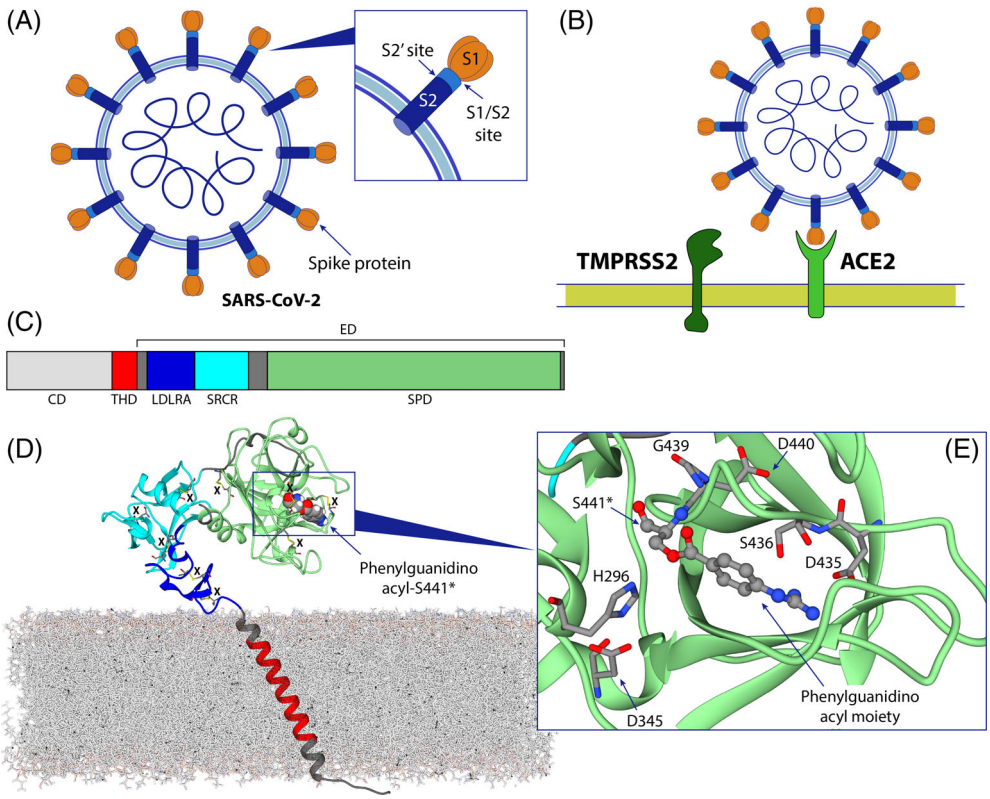


FIGURE 1 (A) SARS-CoV-2 schematic representation and detail of the spike protein. The proteolytic S1/S2 and S2′ cleavage sites are highlighted. (B) Schematic representation of SARS-CoV-2 binding to the human ACE2 receptor and position in the membrane of TMPRSS2. (C) Domain organization in the TMPRSS2 sequence. (D) Starting TMPRSS2-membrane assembly. The protein skeleton is shown as a ribbon. The membrane is in light gray, while the protein ribbons are in red, blue, cyan and light green for THD, LDLRA, SRCR and SPD, respectively. The acylated Ser441* residue and the phenylguanidino acyl moiety are reported as spheres colored according the the atom type, while the nine structure stabilizing disulphide bonds are in sticks highlighted with a "X" character. (E) Detail of the phenylguanidino acyl moiety covalently bound to Ser441* (ball-and-stick) along with residues involved in the catalytic triad (Asp345 and His296) and interacting residues within 3.5 Å. Hydrogen atoms are ignored for clarity.

membrane proteins[11] and in other epithelial homeostasis roles.[12] All the members of the TTSP family are disulphide-rich, experience complex post-translational regulation and proteolytic cleavage activation.[9] Cancers and tumor cell proliferation, invasiveness and metastasis causes a dysregulation in TTSP activity,[13–15] while basal activity levels of TMPRSS2 in normal tissues can be utilized by several viruses (such as influenza A and B viruses[16,17] or coronaviruses[18]) for infection at the host cell membrane. Consequently, TMPRSS2 represents an interesting pharmacological target to prevent SARS-CoV-2 from infecting the human cells.[19,20]

Human TMPRSS2 is composed of three domains (Figure 1C): the cytoplasmic domain (CD, residues 1–84), a transmembrane helix domain (THD, 85–105) and an extracellular domain (ED, 106–492). The ED can be further divided into a low-density lipo-protein receptor domain class A domain (LDLRA, 112–149), a scavenger receptor cysteine-rich domain (SRCR, 150–242) and serine protease domain (SPD, 256–489) containing the catalytic site.

Despite the importance of this protein in combating SARS-CoV-2 infection, no experimental structure of the protein was available until 2022. In 2021, a molecular dynamics (MD) study by Hempel et al. based on a model of TMPRSS2 SPD (residues 256–491)[21] in complex with the hydrolytic product of two known protease inhibitors (namely Nafamostat and Camostat) was published.[22] The simulations, conducted in the tens of–hundreds of microsecond time scale, suggested that in the design of new covalent TMPRSS2 inhibitors one should consider the position of the catalytic serine (Ser441) with respect to Asp435. It was also noted that the phenylguanidino acyl moiety resulting from the inhibitors hydrolysis is slightly shorter than the size of the TMPRSS2 modeled cavity. It was also suggested that a putative new inhibitor should be size-compatible to a hydrophobic patch found in the catalytic pocket.[22]

The x-ray crystal structure of the human TMPRSS2 ED has been recently released (PDB ID 7MEQ, residues 148–491, solved at 1.95 Å resolution) (Figure 1D).[23] The structure reports the TMPRSS2 structure after its reaction with the potent nonselective trypsin-like serine protease inhibitor Nafamostat.[23] This inhibitor reacts with Ser441 and is cleaved into a phenylguanidino acyl moiety thar remains covalently bound to TMPRSS2 Ser441 (Ser441* hereafter) (Figure 1E) and a molecule of 6-amidino-2-napthol. The phenylguanidino acyl moiety also forms a salt-bridge with Asp435 and H-bonds with Ser436 and Gly439 by using its guanidium group. The tertiary structure of the ED is also stabilized by nine disulphide bridges located in the ED (cysteine pairs 113–126, 120–139, 133–148, 172–231, 185–241, 244–365, 281–297, 410–426, and 437–465). Notably, in the TMPRSS2 model used in the 2021 MD study[21,22] there is no indication that the disulfide bonds should be considered in order to achieve a reliable model. Here we took advantage of the TMPRSS2 experimental x-ray structure, while the unsolved parts of TMPRSS2 were modeled through AlphaFold2 and made available through the AlphaFold Protein Structure Database.[24] AlphaFold2 is a recent machine learning algorithm that incorporates physical and biological knowledge about protein structure for the calculation of reliable structural models.[25] Despite the extremely good results obtained in the recent 14th Critical Assessment of protein Structure Prediction (CASP14),[26] AlphaFold2 still shows some problems in the modeling of membrane proteins such as the one at the center of this study.[27] In fact, while TMPRSS2 ED and THD are modeled with good accuracy, the CD is unfolded and localized in regions where the membrane should be or even outside the cell. For this reason, our study focuses only on THD and ED (residues 71–492) (Figure 1). Here we report the results of atomistic molecular dynamics simulations in explicit solvent and membrane performed on the phenylguanidino acyl-bound TMPRSS2 (holo form hereafter) and the inhibitor-deprived protein (apo form). For each system, two 2 μs-long sets of simulations were run. Comparisons between the dynamics of the protein and the active site in the presence or absence of the inhibitor are then discussed with a view to developing new TMPRSS2-specific drugs.

2 | RESULTS AND DISCUSSION

2.1 | General behavior of the apo and holo TMPRSS2

Visual inspection of the trajectories revealed a large mobility of the THD in the membrane with respect to the ED. This fact may be due to the physiological role of the protein of to the absence of the CD and reflects both in the root mean square deviation (RMSD) behavior of the whole protein (Figures S1 and S2) and in the root mean square fluctuations (Figure 2A). Indeed, in both the apo and holo simulations the N-terminal portion, comprising residues 71–115, fluctuates a lot with respect to the other regions of the protein although it remains folded as an α-helix. The other domains are more rigid, and experience mean fluctuations in the order of 1–2 Å. In order to remove the bias on the RMSD caused by the large fluctuations of the THD, the weighted RMSD (wRMSD) was estimated using a Gaussian weighting scheme (Figures 2B and S3).[28] This procedure not only reduced the average RMSD value, but also showed that the wRMSD behavior of the two trajectories run for each system fluctuate around an average value of ca. 1.2–1.3 Å and are essentially superimposable. In other words, there is an agreement between the two replicas of each system in the degree of conformational changes with respect to the reference structure (i.e., the minimized structure). Excluding the THD region, the only region showing differences in the RMSF trend between the apo and holo system is in the range of residues 432–444. This region comprises a long loop delimiting the bottom of the ligand binding pocket which also comprises Ser441* in the holo structure (Figure 1E). As one can expect, in the holo structure, the fluctuations of the 432–444 loop are smaller than in the apo simulations. An analysis of the residues in contact with Ser441* (Figure S4) shows that in the apo form Ser441 forms transient contacts with Gly282, Gly385, Ala386, and Ser463; while in the holo form Ser441* establishes stable interactions mainly with Cys437, Gln438, Gly439, Asp440, and Thr459.

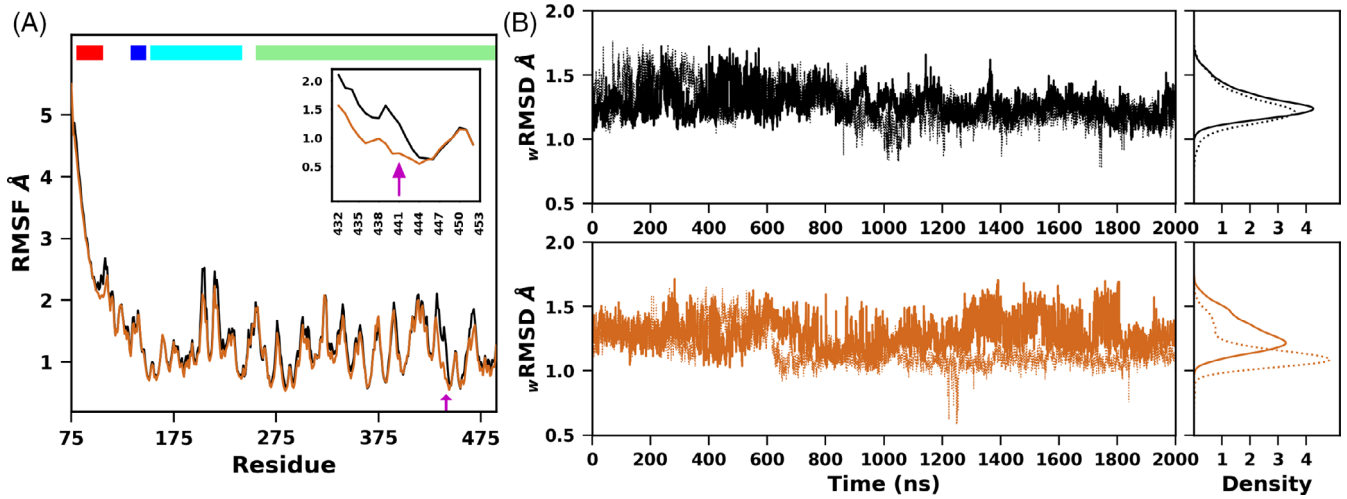
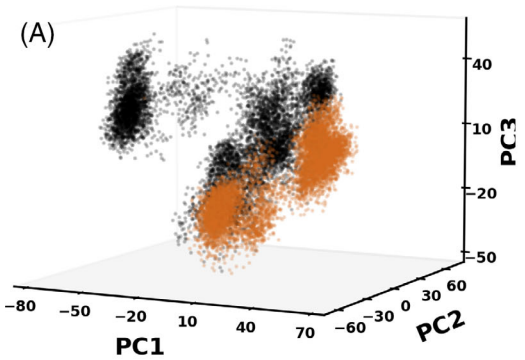


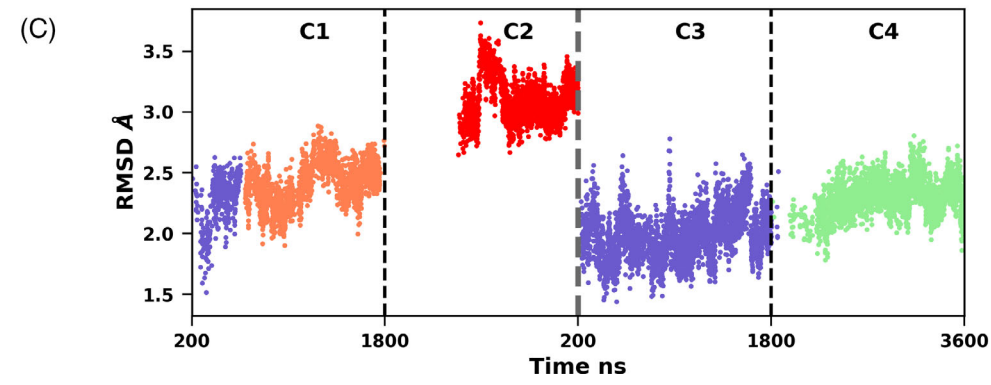
FIGURE 2 Calculated residue averaged Cα root mean square fluctuations (RMSF, A) and time evolution of the weighted Cα root mean square deviations (wRMSD, B) of TMPRSS2 in the apo (black) and holo (orange) form. The wRMSD distributions over the last 1.8 μs are also reported. Dashed lines represent data for replica simulations. In panel A the first and the last three residues were excluded from the RMSF calculation due to high fluctuations, while the inset shows the region (±10 residues) flanking Ser441* (arrow in magenta); while horizontal bars report the position of each domain in the sequence and are colored according to the domain coloration in Figure 1C.

FIGURE 3 Principal component and clustering analyses on TMPRSS2 ED. (A) Projections of apo (black) and holo (orange) snapshots along the first three principal components. (B) Summary of the population of the four main clusters (C1–C4) obtained by clustering on the principal component space defined by the first five components. (C) ED domain Cα RMSD calculated on last 1.8 μs of simulation time. Only frames belonging to the four main populated clusters are reported and colored in orange, red, purple and light green for C1, C2, C3, and C4, respectively. Dashed gray line separates frames belonging to apo and holo systems.



(B)

Cluster	%Fraction	%Apo	%Holo
C1	0.12	1.00	0.00
C2	0.13	1.00	0.00
C3	0.25	0.10	0.90
C4	0.18	0.00	1.00



2.2 | Cumulative motions in the apo and holo form

Motion correlations between various sub-parts of TMPRSS2 were characterized by a calculation of the covariance matrices of the Cα displacements and a subsequent principal component analysis (PCA). The resulting plot of cumulative variance vs. eigenvector index (Figures S5 and S6) and visual inspection of the corresponding projection of protein motion along the principal vectors revealed that for both apo and holo system the largest protein motions are dominated by the THD displacement in different directions. Interestingly, in the apo system the first eigenvector is also characterized by significative motions also occurring in the active site region and, in particular, in the already mentioned 432–444 loop. This result agrees with the previous fluctuations analysis. To better characterize the fluctuations in the active site region, we repeated the PCA analysis on the ED alone. To do so, the last 1.8 μs of each replica of each system were joined together and superimposed before being analyzed in order to distinguish between different conformational states. Figure 3 shows the

results of the PCA and clustering analyses on the ED. The conformational space explored by the first three covariance eigenvectors was inspected. The resulting scatter plots (Figure 3A) were used to search for significative substates of conformational populations in the two systems. From a qualitative examination, the holo system appears to be limited to mainly two states, while the apo ED is less stiff. Indeed, in the apo protein it is possible to observe a wide number of conformations that are only in part in the same conformational region of the holo system. A cluster analysis of points on the PCA subspace revealed that 68% of all the conformations observed in the joint trajectory are ascribable to one of the four most populated clusters (C1-C4, Figure 3B,C), while the remaining frames are part of minor clusters or cannot be included in a single cluster. Interestingly, the largest proportion of the conformations not included in the four main clusters is attributable to the apo system, confirming the larger conformational space of the protein when it is not involved in ligand binding. Of the main clusters, C1 and C2 represent two conformations observed only in the apo form, while C4 is typical of the holo system. C3 is observed mainly in the holo system but also in the initial stages of the first replica of the apo protein. This is probably because the apo simulations started from the holo system depleted of the phenylguanidino acyl moiety. Then, the apo system took some time to relax to conformations typical of the real apo state and, at the beginning of one of the two simulations, it retained the initial holo conformation.

In order to highlight the differences between each cluster of the ED obtained through PCA, the centroid of each cluster was determined. Subsequently, the 1000 snapshots closest to the centroid structure were chosen for each cluster and the RMSF per residue was calculated on these four structural subsets (Figure 4). In general, in the clusters observed during simulations on the holo system, greater rigidity is observed compared to clusters referring to the apo system, together with a greater similarity between the cluster centroid and its immediate neighbors. Also, based on previous analyses, it is not surprising to note that in apo system there is greater conformational variability in the active site region. However, in C1 there is considerable flexibility in the LDLRA region. In particular, the LDLRA and the SRCR domains move in anticorrelated motion with respect to the SPD domain. The consequence of this movement is the opening or closing of a pocket between the LDLRA/SRCR and the SPD domains. Other regions with notable movements include the 386–401 loop bordering the active site on one side and the 251–261 loop adjacent to the latter. Another very flexible region is the 201–206 loop, that is in the SRCR domain and on the opposite side of the protein with respect to the active site. However, the movements described above do not affect the overall domain architecture and the interdomains residue contact map. In both the apo and holo system the number of interdomain contacts remain the same for the largest part of the simulation time (Figure S7).

2.3 | Volumetric and geometric analyses

As a next step, attention was paid to the consequences of the motions determined through the PCA of the ED. The cavities on the surface of the ED with a lifetime of at least 60% of the simulation time were determined and the results are shown in Figure 5A. In the apo form there is a pocket in the correspondence of the active site, as expected. However, this pocket appears to be more flexible than other clefts on

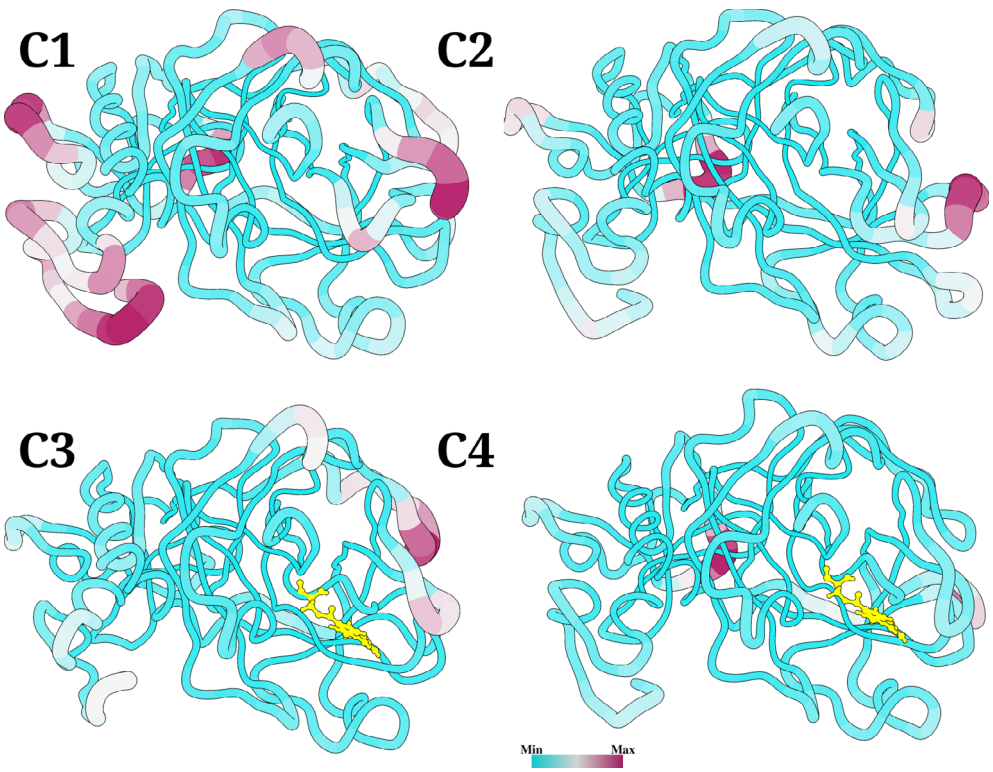


FIGURE 4 Structural mapping of residue level fluctuations. The representative snapshot of each of the four ED most populated clusters is depicted in worm representation whose thickness is proportional to RMSF values estimated over 1000 closest snapshots to the respective centroids. The phenylguanidino acyl moiety is highlighted in yellow for cluster 3 and 4 that are mainly constituted by frames of holo system.

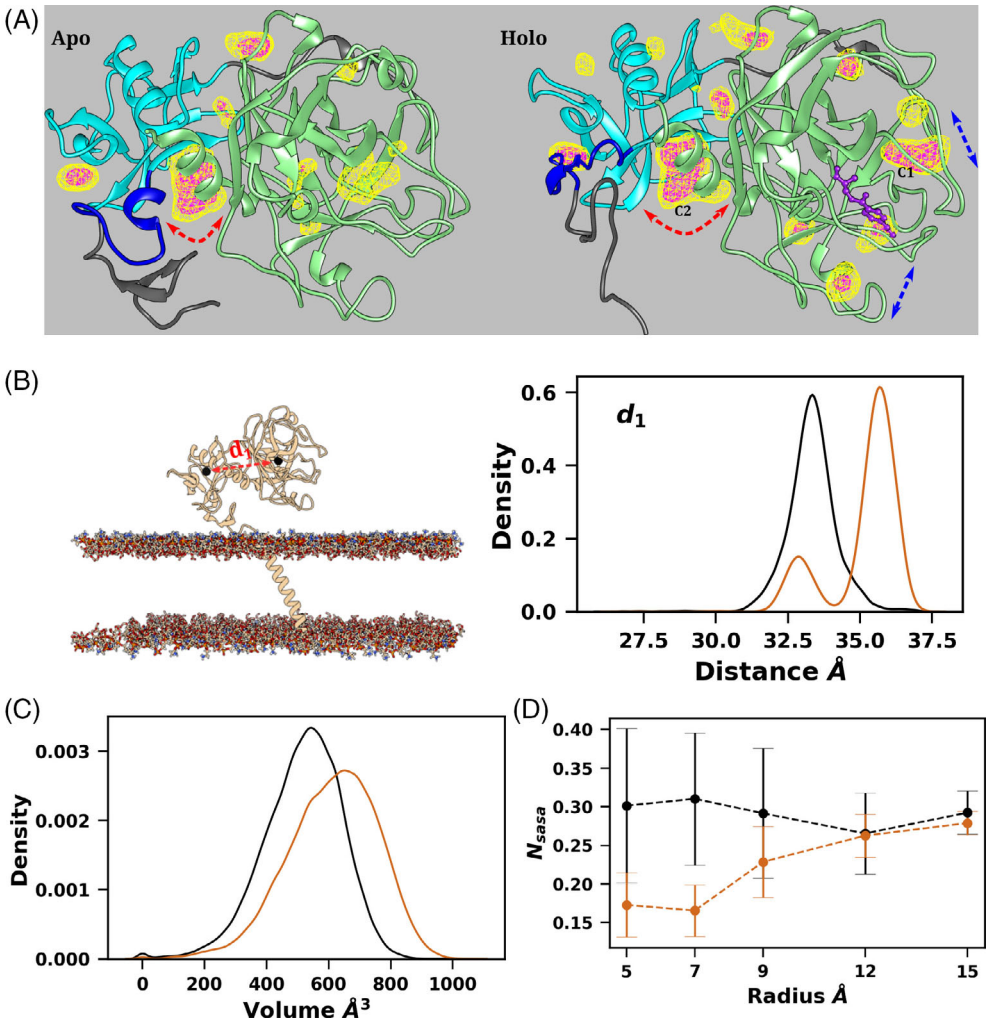


FIGURE 5 Volumetric and geometric analyses. Formation of clefts in protein was identified by volumetric analysis (A). Iso surfaces of clefts sustaining for 60 (yellow) to 75 (magenta) percent of the simulation time are depicted for apo (left) and holo (right) systems. In the holo system, the ligand is shown in purple and dashed blue arrows highlight additional clefts visible in the SPD (the largest cavity, C1, is discussed in the text). Red arrows indicate the elongation of the cleft between LDLRA/SRCR (cyan) and SPD (light green) domains (cavity C2), which is further characterized by distance between respective domain centroids (B) and cleft volume (C). (D) shows the normalized solvent accessibility of residues identified within varying radius from Ser441* for the apo (black) and holo (orange) systems.

the ED surface, that are stable for 75% of the simulation time, suggesting a large flexibility in the active site, which is in agreement with the previous analysis of the trajectories and a phenomenon also observed for other proteases.[29–31] From the cavity analysis, it is also possible to notice the formation of some small cavities in the holo form of the protein that are not present in the apo form plus a large cavity overhanging the active site and close to the phenylguanidino acyl moiety (cavity C1 in Figure 5A). The latter is formed by residues in the 251–261 loop (namely, Val257, Gly258, Glu260, and Ser261) as well as in 386–401 loop (Glu388, Glu398, Gly391, Lys392, Thr393, Ser394, Glu395, Leu397, and Asn398) as reported in Figure S8. In other words, the presence of the ligand appears to stabilize the binding pocket and consequently to release the movement of the 386–401 and 251–261 loops. Interestingly, cavity C1 is present in both replicas (see Figure S9), suggesting that this structure is very stable in the microsecond time scale. A second large cavity is present in both the apo and holo form but appears to be larger in the holo form (cavity C2 in Figure 5). This second cavity is located between the LDLRA/SRCR and the SPD domains. In order to get an idea of the size of the cavity opening, the distance between selected residues in the center of the SRCR and the SPD domains was measured as a function of the simulation time (Figure 5B). In the apo system, the two domains remain much closer (average distance ca. 33 Å) than in the holo system, where the average distance is ca. 35 Å. Interestingly, in the apo form the domains composing the ED appears to assume a single conformation, while in the holo form there are two possible SRCR-SPD arrangements, one similar to that observed in the apo form and another more open. The analysis of cavity volumes between the LDLRA/SRCR and the SPD confirms this trend (Figure 5C). Indeed, such a cavity in the holo form is larger (average volume ca. 450 Å^3) than in the apo form (ca. 350 Å^3), suggesting that the presence of a ligand in the active site increases the conformational variability of the LDLRA/SRCR domains with respect to the SPD.

Finally, the correct characterization of the residues involved in the active site cavity is made complex due to conformational variability of the cavity itself. To compare the active site in the apo and holo form of TMPRSS2, the normalized solvent accessibility of the residues (N_{SASA}) identified within varying radius from the active site was calculated. Such an analysis considers the solvent exposed surface of all residues within an increasing cut-off value and the resulting area is then normalized over the number of residues within the cut-off. In the absence of a cavity, the N_{SASA} value is constant, while it tends to grow

with a trend that increases to an asymptote in the presence of a pocket. Such an analysis (Figures 5D and S9) shows that, in agreement with the previous analysis, the solvent exposed surface of the protein in the apo form is larger and more variable than in the holo form, where the presence of the phenylguanidino acyl moiety tightens and stabilizes the active site region.

3 | CONCLUSIONS

In conclusion, in this work we performed atomistic MD simulations on the most complete TMPRSS2 structure used so far. The inclusion of the THD also allows to study the motion of the extracellular parts of the proteins with respect to the membrane. The latter domain was absent in the already cited study by Hempel and co-workers performed prior to the release of the experimental structure of the protein[22] in which similar simulations were conducted on a model of TMPRSS2 ED without the remaining domains. As expected, our simulations showed that the catalytic site of apo TMPRSS2 is more flexible than in the holo form. Indeed, the phenylguanidino acyl moiety bound to Ser441 stabilizes the active site through the formation of several intermolecular interactions. On the other hand, and less expected, the simulations on the holo enzyme revealed the formation of a stable cavity in the proximity of the phenylguanidino acyl moiety. Interestingly, even if the initial model used by Hempel et al. differs from the recent experimental crystal structure (Figure S11), this cavity is present also in their simulations, yet smaller and present for a shorter period of time (see Figure S12). This result could be used in designing new specific TMPRSS2 inhibitors. Indeed, one of the biggest challenges in designing new and specific protease inhibitors is the high similarity among the active sites of this family of proteins. Because of this, inhibitors are often able to inhibit the activity of more than a single protease. Indeed, it has been reported that the 115 annotated human protease inhibitors are responsible for regulating the activity of the 612 known human proteases.[32] It is likely that these numbers will change as protease and inhibitor families become more refined. However, the ratio of approximately one protease inhibitor to five proteases is likely to remain constant. One potential solution for optimizing the specificity of protease inhibitors, which has been explored extensively in recent years, is to identify new sites in the protein that can be used as potential targets for the designed inhibitors, such as the one reported here. The simulations reported in this work therefore proved to be useful in showing features of TMPRSS2 that had not previously been highlighted by studying the available structures or models. The cavity close to the active site identified in the simulation of the holo form may provide some docking points for the development of new inhibitors with greater specificity. Furthermore, the inclusion of the THD revealed a second cavity located between the LDLRA/SRCR and the SPD domains that allows for more flexibility in the development of new inhibitors. The function of the latter cavity is yet to be demonstrated, but if it has a functional role, it is certainly an excellent starting point for the development of new and more specific inhibitors.

4 | MATERIALS AND METHODS

4.1 | Modeling and simulation

The x-ray structure of human TMPRSS2 (PDB ID: 7MEQ)[23] is resolved only for the extracellular region (residues 148–491) in complex with a covalently bound inhibitor fragment (phenylguanidino acyl deriving from Nafamostat hydrolysis) and contains several unsolved segments. The starting structure for the present simulations was obtained from the AlphaFold (AF) Structure Database (entry O15393)[24] and terminal regions with low prediction accuracy were removed. Thus, the starting model included residues Ser71-Gly492 (Figure 1D). For the extracellular region, the root mean square deviation (RMSD) between the x-ray and AlphaFold2 models was 1.4 Å over 325 Cα atom pairs, indicating that the latter model closely represents the holo state. In particular, the model of the SPD is in excellent agreement with the experimental structure (Cα RMSD ca. 0.4 Å, Figure S13). The phenylguanidino acyl moiety was covalently incorporated into the starting model after being superimposed on the x-ray structure to maintain its bound conformation and chemical environment (Figure 1E), while Ser441 was left unmodified in the apo system. The starting models were embedded in a membrane bilayer of 600 lipid molecules consisting of POPE and POPG in a 7:3 ratio using the CHARMM-GUI server.[33–37] The assembly was immersed in a TIP3P[38] water box extending up to 15 Å from the protein surface. To achieve electroneutrality, the system was neutralized, and physiological ionic strength was mimicked through the addition of 0.1 M KCl. Ligand partial charges were obtained by fitting the electrostatic potential derived at the HF/6-31G* theory of level with Gaussian16[39] via the restrained electrostatic potential (RESP) procedure[40] using R.E.D. tools.[41] Ligand parameters were adapted from the Amber ff10 forcefield following best practices explained elsewhere.[42,43]

The minimization, equilibration and production runs were performed using GROMACS software (v2021.5),[44–46] with the Amber ff14SB force field for the protein[47] and Lipid14 for the lipids.[48] Systems were energy-minimized using the steepest-descent algorithm for 5000 steps and then equilibrated with an NVT simulation using the v-rescale thermostat[49] at 303.15 K for 100 ps. This was followed by an NPT equilibration at 1 atm pressure using the Parrinello-Rahman barostat[50] and at 303.15 K using the Nose-Hoover thermostat[51] for a further 1 ns. Finally, 2 μs NPT production runs were performed for both systems, including two replicate runs which were initiated from the same equilibration configurations but with different random velocity distributions. In all simulations, a time step of 2 fs was employed and the LINCS algorithm[52] was used to impose constraints on the hydrogen-containing bonds. Periodic boundary conditions (PBC) were applied, and the Particle Mesh Ewald (PME) method[53] was used to calculate long-range electrostatic interactions. The equilibration and production runs were executed on the CINECA MARCONI100 and the ENI HPC5 clusters. These supercomputers have similar architectures consisting of nodes equipped with 4× NVIDIA V100 GPUs per node so the optimal setup for both machines, with this version of GROMACS, was to use 1 node per run, resulting in

performances of about 170 ns/day for the holo systems and 190 ns/day for the apo runs.

4.2 | Trajectory analysis

The simulated trajectories were analyzed with the MDAnalysis,[54] MDTraj 1.9.8[55] and mdciao 0.0.5[56] Python packages, using in-house Jupyter notebooks.[57] Plots were generated with matplotlib and seaborn libraries. UCSF Chimera 1.16[58] was used for molecular figures. Analyses requiring structural alignments were performed based on Cα coordinates using equilibrated snapshots as the reference structure. All analyses except RMSD were performed on a single trajectory obtained by combining the final 1.8 μs sampling of replicate runs, allowing statistics of analyzed observables to be obtained over 3.6 μs of data. Weighted RMSD (wRMSD) was estimated using a Gaussian weighting scheme.[28] The method involves an iterative fit, where the weights for the atoms used in the alignment (we used Cα atoms) are assigned with a Gaussian function, defined as:

$$W_i = e^{-(r_i^2)/C} \quad (1)$$

where r_i is the distance of the ith atom between the two aligned conformations, and C is an arbitrary scaling factor. At each iteration, the weights are updated and the fitting proceeds until convergence in ΔwRMSD is achieved. A convergence threshold of 1.0×10^{-4} Å and a scaling factor of 5 Å^2 were used for the estimates. To obtain appropriate initial weights, the two conformations are nearly aligned via a standard RMSD fit. The root mean square fluctuation (RMSF) was estimated in windows of 100 ns. Reported values are the average over all windows. Principal component analysis (PCA) was performed on snapshots extracted at 500 ps intervals. Clustering analysis was performed based on the density of points on the PCA subspace using the HDBSCAN algorithm.[59,60] The distance between SRCR and SPD domains was calculated based on the coordinates of the closest Cα atom to the respective domain centroids. The pockets on the TMPRSS2 surface were identified using MDpocket.[61] The solvent accessible surface area (SASA) was calculated using MDTraj with a probe radius of 1.4 Å. The resulting value was normalized by the number of residues identified within spherical zones of different radii around the geometric center of the ligand site.

AUTHOR CONTRIBUTIONS

Giorgia Frumenzio: Conceptualization; investigation; methodology; validation; visualization; writing – review and editing; software; formal analysis; data curation; resources; funding acquisition; writing – original draft. **Balasubramanian Chandramouli:** Conceptualization; investigation; writing – original draft; methodology; validation; visualization; writing – review and editing; software; formal analysis; project administration; supervision; resources; funding acquisition. **Neva Besker:** Writing – review and editing; writing – original draft; resources; supervision. **Alessandro Grottesi:** Writing – original draft; writing – review and editing; resources; supervision; methodology. **Carmine Talarico:** Writing – original draft; writing – review and editing; supervision; resources. **Francesco Frigerio:** Conceptualization; writing – original draft; writing – review and editing; resources; data curation; investigation; supervision; methodology; visualization. **Andrew Emerson:** Conceptualization; investigation; writing – original draft; methodology; writing – review and editing; software; formal analysis; project administration; data curation; supervision; resources. **Francesco Musiani:** Supervision; formal analysis; project administration; conceptualization; writing – original draft; writing – review and editing; methodology; validation; visualization; software; funding acquisition; resources; investigation.

ACKNOWLEDGMENTS

We acknowledge CINECA for allocating computer time on the MARCONI100 and ENI for time on the HPC5 cluster. F.M. was supported by Ministero dell'Istruzione, dell'Università e della Ricerca (RFO grant 2020) and by Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding authors upon reasonable request. The representative structures of the clusters obtained from the simulation trajectories are available at the address: https://github.com/HPC-MD-GROUP/TMPRSS.

ORCID

Francesco Musiani https://orcid.org/0000-0003-0200-1712

REFERENCES

1. Hartenian E, Nandakumar D, Lari A, Ly M, Tucker JM, Glaunsinger BA. The molecular virology of coronaviruses. *J Biol Chem*. 2020;295(37):12910-12934.
2. Tang G, Liu Z, Chen D. Human coronaviruses: origin, host and receptor. *J Clin Virol*. 2022;155:105246.
3. Coronaviridae Study Group of the International Committee on Taxonomy of V. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;5(4):536-544.
4. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021;19(3):141-154.
5. WHO. WHO coronavirus (COVID-19) dashboard | WHO coronavirus (COVID-19) dashboard with vaccination data. 2022; https://covid19.who.int/
6. Malone B, Urakova N, Snijder EJ, Campbell EA. Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat Rev Mol Cell Biol*. 2022;23(1):21-39.
7. Kung YA, Lee KM, Chiang HJ, Huang SY, Wu CJ, Shih SR. Molecular virology of SARS-CoV-2 and related coronaviruses. *Microbiol Mol Biol Rev*. 2022;86(2):e0002621.
8. Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol*. 2022;23(1):3-20.
9. Bugge TH, Antalis TM, Wu Q. Type II transmembrane serine proteases. *J Biol Chem*. 2009;284(35):23177-23181.

10. Mukai S, Yorita K, Kawagoe Y, et al. Matriptase and MET are prominently expressed at the site of bone metastasis in renal cell carcinoma: immunohistochemical analysis. *Hum Cell*. 2015;28(1):44-50.
11. Heurich A, Hofmann-Winkler H, Gierer S, Liepold T, Jahn O, Pohlmann S. TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *J Virol*. 2014;88(2):1293-1307.
12. List K, Kosa P, Szabo R, et al. Epithelial integrity is maintained by a matriptase-dependent proteolytic pathway. *Am J Pathol*. 2009;175(4):1453-1463.
13. Larzabal L, Nguewa PA, Pio R, et al. Overexpression of TMPRSS4 in non-small cell lung cancer is associated with poor prognosis in patients with squamous histology. *Br J Cancer*. 2011;105(10):1608-1614.
14. Webb SL, Sanders AJ, Mason MD, Jiang WG. Type II transmembrane serine protease (TTSP) deregulation in cancer. *Front Biosci (Landmark ed)*. 2011;16(2):539-552.
15. Ko CJ, Huang CC, Lin HY, et al. Androgen-induced TMPRSS2 activates matriptase and promotes extracellular matrix degradation, prostate cancer cell invasion, tumor growth, and metastasis. *Cancer Res*. 2015;75(14):2949-2960.
16. Bottcher E, Matrosovich T, Beyerle M, Klenk HD, Garten W, Matrosovich M. Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and HAT from human airway epithelium. *J Virol*. 2006;80(19):9896-9898.
17. Limburg H, Harbig A, Bestle D, et al. TMPRSS2 is the major activating protease of influenza a virus in primary human airway cells and influenza B virus in human type II pneumocytes. *J Virol*. 2019;93(21):e00649–19.
18. Matsuyama S, Nagata N, Shirato K, Kawase M, Takeda M, Taguchi F. Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J Virol*. 2010;84(24):12658-12664.
19. McKee DL, Sternberg A, Stange U, Laufer S, Naujokat C. Candidate drugs against SARS-CoV-2 and COVID-19. *Pharmacol Res*. 2020;157:104859.
20. Ragia G, Manolopoulos VG. Inhibition of SARS-CoV-2 entry through the ACE2/TMPRSS2 pathway: a promising approach for uncovering early COVID-19 drug therapies. *Eur J Clin Pharmacol*. 2020;76(12):1623-1630.
21. Rensi S, Altman RB, Liu T, et al. Homology modeling of TMPRSS2 yields candidate drugs that may inhibit entry of SARS-CoV-2 into human cells. *ChemRxiv*. 2020. doi:10.26434/chemrxiv.12009582.v1
22. Hempel T, Raich L, Olsson S, et al. Molecular mechanism of inhibiting the SARS-CoV-2 cell entry facilitator TMPRSS2 with camostat and nafamostat. *Chem Sci*. 2021;12(3):983-992.
23. Fraser BJ, Beldar S, Seitova A, et al. Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. *Nat Chem Biol*. 2022;18(9):963-971.
24. Varadi M, Anyango S, Deshpande M, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*. 2022;50(D1):D439-D444.
25. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589.
26. Jumper J, Evans R, Pritzel A, et al. Applying and improving AlphaFold at CASP14. *Proteins*. 2021;89(12):1711-1721.
27. Camporesi G, Minzoni A, Morasso L, Ciurli S, Musiani F. Nickel import and export in the human pathogen helicobacter pylori, perspectives from molecular modelling. *Metallomics*. 2021;13(12):1–14.
28. Damm KL, Carlson HA. Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures. *Biophys J*. 2006;90(12):4558-4573.
29. Ha Y. Structure and mechanism of intramembrane protease. *Semin Cell Dev Biol*. 2009;20(2):240-250.
30. Hilgenfeld R, Lei J, Zhang L. The structure of the zika virus protease, NS2B/NS3(pro). *Adv Exp Med Biol*. 2018;1062:131-145.
31. Gossen J, Albani S, Hanke A, et al. A blueprint for high affinity SARS-CoV-2 Mpro inhibitors from activity-based compound library screening guided by analysis of protein dynamics. *ACS Pharmacol Transl Sci*. 2021;4(3):1079-1095.
32. Farady CJ, Craik CS. Mechanisms of macromolecular protease inhibitors. *Chembiochem*. 2010;11(17):2341-2346.
33. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, et al. CHARMM: the biomolecular simulation program. *J Comput Chem*. 2009;30(10):1545-1614.
34. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*. 2008;29(11):1859-1865.
35. Lee J, Cheng X, Swails JM, et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J Chem Theory Comput*. 2016;12(1):405-413.
36. Lee J, Hitzenberger M, Rieger M, Kern NR, Zacharias M, Im W. CHARMM-GUI supports the Amber force fields. *J Chem Phys*. 2020;153(3):035103.
37. Wu EL, Cheng X, Jo S, et al. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J Comput Chem*. 2014;35(27):1997-2004.
38. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926-935.
39. *Gaussian 16 Rev. C.01* [computer program].
40. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem*. 1993;97(40):10269-10280.
41. Dupradeau FY, Pigache A, Zaffran T, et al. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys*. 2010;12(28):7821-7839.
42. Chandramouli B, Del Galdo S, Mancini G, Tasinato N, Barone V. Tailor-made computational protocols for precise characterization of small biological building blocks using QM and MM approaches. *Biopolymers*. 2018;109(10):e23109.
43. Chandramouli B, Del Galdo S, Mancini G, Barone V. Mechanistic insights into metal ions transit through threefold ferritin channel. *Biochim Biophys Acta Gen Subj*. 2019;1863(2):472-480.
44. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun*. 1995;91(1):43-56.
45. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Mol Model Annu*. 2001;7(8):306-317.
46. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701-1718.
47. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 2006;65(3):712-725.
48. Dickson CJ, Madej BD, Skjevik AA, et al. Lipid14: the Amber lipid force field. *J Chem Theory Comput*. 2014;10(2):865-879.
49. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007;126(1):014101.
50. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys*. 1981;52(12):7182-7190.
51. Evans DJ, Holian BL. The nose–hoover thermostat. *J Chem Phys*. 1985;83(8):4069-4074.

52. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*. 1997;18(12):1463-1472.
53. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103(19):8577-8593.
54. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. 2011;32(10):2319-2327.
55. McGibbon RT, Beauchamp KA, Harrigan MP, et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys J*. 2015;109(8):1528-1532.
56. *gph82/mdciao: First Stable Release Candidate [Computer Program]. Version 0.0.4*. Zenodo; 2021.
57. Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks ? A Publishing Format for Reproducible Computational Workflows. Paper presented at: *20th International Conference on Electronic Publishing* (01/01/16); 2016.
58. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612.
59. McInnes L, Healy J, Astels S. hdbscan: hierarchical density based clustering. *J Open Source Softw*. 2017;2(11):205.
60. Chandramouli B, Melino G, Chillemi G. Smyd2 conformational changes in response to p53 binding: role of the C-terminal domain. *Mol Oncol*. 2019;13(6):1450-1461.
61. Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*. 2011;27(23):3276-3285.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Frumenzio G, Chandramouli B, Besker N, et al. Conformational response to ligand binding of TMPRSS2, a protease involved in SARS-CoV-2 infection: Insights through computational modeling. *Proteins*. 2023;91(9):1288-1297. doi:10.1002/prot.26548