

Article

Molecular Insights into Human Transmembrane Protease Serine-2 (TMPS2) Inhibitors against SARS-CoV2: Homology Modelling, Molecular Dynamics, and Docking Studies

Safaa M. Kishk ^{1,*}, Rania M. Kishk ², Asmaa S. A. Yassen ³, Mohamed S. Nafie ⁴, Nader A. Nemr ⁵, Gamal ElMasry ^{6,7}, Salim Al-Rejaie ⁷ and Claire Simons ⁸

- ¹ Pharmaceutical Medicinal Chemistry Department, Faculty of Pharmacy, Suez Canal University, Ismailia 41522, Egypt
 - ² Microbiology and Immunology Department, Faculty of Medicine, Suez Canal University, Ismailia 41522, Egypt; raniakishk@med.suez.edu.eg
 - ³ Pharmaceutical Organic Chemistry Department, Faculty of Pharmacy, Suez Canal University, Ismailia 41522, Egypt; asmaa_yaseeen@pharm.suez.edu.eg
 - ⁴ Chemistry Department, Faculty of Science, Suez Canal University, Ismailia 41522, Egypt; Mohamed_nafie@science.suez.edu.eg
 - ⁵ Endemic and Infectious Diseases Department, Faculty of Medicine, Suez Canal University, Ismailia 41522, Egypt; nadernemr@med.suez.edu.eg
 - ⁶ Faculty of Agriculture, Suez Canal University, Ismailia 41522, Egypt; gamal.elmasry@agr.suez.edu.eg
 - ⁷ Department of Pharmacology & Toxicology, College of Pharmacy, King Saud University, Riyadh 11564, Saudi Arabia; rejaie@ksu.edu.sa
 - ⁸ School of Pharmacy & Pharmaceutical Sciences, Cardiff University, King Edward VII Avenue, Cardiff CF103NB, UK; simonsc@cardiff.ac.uk
- * Correspondence: safaa_keshk@pharm.suez.edu.eg

Received: 3 October 2020; Accepted: 24 October 2020; Published: 29 October 2020



Abstract: The severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), which caused novel corona virus disease-2019 (COVID-19) pandemic, necessitated a global demand for studies related to genes and enzymes of SARS-CoV2. SARS-CoV2 infection depends on the host cell Angiotensin-Converting Enzyme-2 (ACE2) and Transmembrane Serine Protease-2 (TMPRSS2), where the virus uses ACE2 for entry and TMPRSS2 for S protein priming. The TMPRSS2 gene encodes a Transmembrane Protease Serine-2 protein (TMPS2) that belongs to the serine protease family. There is no crystal structure available for TMPS2, therefore, a homology model was required to establish a putative 3D structure for the enzyme. A homology model was constructed using SWISS-MODEL and evaluations were performed through Ramachandran plots, Verify 3D and Protein Statistical Analysis (ProSA). Molecular dynamics simulations were employed to investigate the stability of the constructed model. Docking of TMPS2 inhibitors, camostat, nafamostat, gabexate, and sivelestat, using Molecular Operating Environment (MOE) software, into the constructed model was performed and the protein-ligand complexes were subjected to MD simulations and computational binding affinity calculations. These in silico studies determined the tertiary structure of TMPS2 amino acid sequence and predicted how ligands bind to the model, which is important for drug development for the prevention and treatment of COVID-19.

Keywords: TMPS2; SARS-CoV2; homology model; molecular dynamics

1. Introduction

Emerging respiratory diseases such as severe acute respiratory syndrome coronavirus (SARS) is a vital threat to public health. New emerging diseases remain a cause of outbreaks, epidemics, and pandemics [1,2]. A new fatal respiratory virus in Wuhan, China, was recently identified in mid-December 2019. The RNA virus, isolated from the patients' bronchoalveolar fluid, was investigated by Metagenomic RNA Sequencing. This RNA virus was identified by the phylogenetic study of the complete viral genome to be a member of the Coronaviridae family; the SARS-like coronavirus community (Betacoronavirus genus, Sarbecovirus subgenus) [3–5]. The new emerging respiratory virus was named Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV2) by the World Health Organization (WHO) and the disease is now referred to as coronavirus-19 (COVID-19).

Coronaviruses are species of Nidoviral viruses, including Coronaviridae, Arteriviridae, Roniviridae, and Mesoniviridae [6,7]. The Coronaviridae virus family is divided into two subfamilies: four genera coronavirinae (Alpha, Beta, Gamma, and Delta Coronavirus), and torovirinae [8]. The genera of the Beta coronavirus include respiratory viruses such as SARS, MERS coV, and SARS-CoV2.

Coronaviruses encode four structural proteins in their genomes: Spike (S), Glycoprotein (E), Nucleocapsid (N), and Envelope (E) proteins. The viral attachment protein, which may bind the virion to a host receptor, is considered to be the (S) protein [8]. Angiotensin Converting Enzyme 2 (ACE2) is part of the renin-angiotensin system and its expression protects against acute respiratory distress [9,10]. SARS-CoV employs ACE2 as a host cell entry receptor [11,12]. ACE2, with some mutations in key residues of the SARS-CoV2 (S) protein, was also recognized as the entry receptor for SARS-CoV2 [13,14]. There are, however, indications that the SARS-CoV2 (S) protein binds to the receptor with an affinity much higher than that observed for the SARS-CoV (S) protein, which may play a major role in the rapid spread of COVID-19 worldwide [15]. ACE2 had been considered as the entry receptor [11,16], while (S) protein priming is performed by the transmembrane serin proteases type II TMPRSS2 [17–19]. However, the viral spread and pathogenesis of the infected host was shown to involve TMPRSS2 activities [20–22] (Figure 1). The TMPRSS2 gene codes for the TMPS2 protein, which is present in all possible targets of SARS-CoV-2 infection, including airway epithelial cells, cardiac endothelium, microvascular endothelial cells, kidney, and digestive tract [23]. Therefore, targeting TMPS2 might represent a suitable approach for both preventing viral infection and treatment of severely ill patients [24].

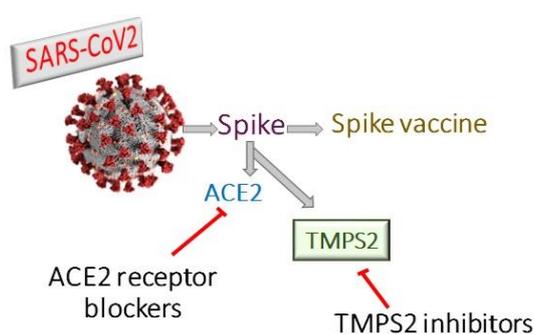


Figure 1. Therapeutic approaches for treatment of covid-19 either by ACE2 receptor blockers, TMPS2 inhibitors or spike vaccine.

The TMPS2 inhibitors; camostat mesylate and nafamostat mesylate were found to slow down cell infection by SARS-CoV in preclinical models [25–27]. Nafamostat and camostat have anti-inflammatory activity in the airways by reducing inflammatory cytokine production in a chronic allergen-induced asthma model [28,29]. Furthermore, nafamostat can inhibit the coagulation and fibrinolytic systems, the kallikrein-kinin system and activate protease-activated receptors [30].

Consequently, the anti-inflammatory, anti-coagulant, and fibrinolytic properties of camostat and nafamostat might contribute to the reduction of symptoms and complications occurring in COVID-19 patients. Both agents are approved in Japan for treatment of pancreatitis [31] and nafamostat is also used as an anticoagulant in extracorporeal circulation [32]. The introduction of nafamostat in addition to antivirals such as lopinavir and hydroxychloroquine in the treatment of elderly COVID-19 patients with pneumonia, showed clinical and radiological improvement without significant adverse effects [33].

Gabexate and sivelestat are serine protease inhibitors marketed in Italy and Japan for the treatment of pancreatitis and have a mechanism similar to nafamostat and camostat (i.e., inhibition of TMPS2).

To our knowledge, there is no crystal structure for TMPS2 available in the Protein Data Bank (PDB), which necessitates the construction of a three-dimensional (3D) structure model to determine the protein architecture and binding interactions of TMPS2 inhibitors to facilitate the improvement of drug design research against COVID-19.

2. Results and Discussion

2.1. Homology Searching and Template Identification

The ExPASy proteomics server was used to obtain the amino acid sequence of the target protein TMPS2. Using the UniProtKB database from the ExPASy server, the amino acid sequence of TMPS2 Homo sapiens in FASTA format was obtained (Figure 2). This protein has a UniProt identifier O15393-1 [34] and is composed of 492 amino acid residues.

```
>sp|O15393|TMPS2_HUMAN Transmembrane protease serine 2 OS=Homo sapiens OX=9606 GN=TMPS2 PE=1 SV=3
MALNSGSPPAIGPYENHGYQENPYPAQPTVVPTVYEVHPAQYYPSPVPQYAPRVLTQASNPVVTQPKSPSGTVCTSKTKKALCITLTLGTFVLG
AALAAGLLWKFMGSKCSNSGIECDSSGTCINPSNWCDGVSHCPGGEDENRCVRLYGNPNFILQVYSSQRKSWHPVCQDDWNNENYGRAACRDM
GYKNNFYSSQGIVDDSGSTFSMKLNTSAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLNSSRQSRIVGGESALPGAWPWQVSLHVQNVHVCG
GSIITPEWIVTAAHCVEKPLNPNWHWTFAGILRQSFMYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPMMMLQPE
QLCWISGWGATEEKGKTSEVLNAAKVLLETQRCNSRYVDNLTIPAMICAGFLQGNVDSQCQGDSSGGPLVTSKNNIWWLIGDTSWVSGCAKAYR
PGVYGNVMVFTDWYRQMRADG
```

Figure 2. The amino acid sequence of TMPS2 in FASTA format showing the scavenger receptor cysteine-rich (SRCR) domain highlighted in yellow (amino acid sequence 150–242) and the serine peptidase domain highlighted in cyan (amino acid sequence 265–489).

Initial screening for suitable templates was carried out by BLAST analyses against the protein data bank (PDB) using both SWISS-MODEL [35] and NCBI blastp [36] (Table 1). SWISS-MODEL analysis revealed 5CE1 and 1Z8G as optimal templates, both of which are crystal structures of transmembrane protease serine 1 with 33.61 and 33.52% identity, respectively. NCBI blastp analysis identified 6KD5_B, transmembrane protease serine 13 (peptidase domain) as the first hit with 45.19% identity, followed by a group of human plasma kallikrein crystal structures (e.g., 6O1G_A) with ~42% identity.

Table 1. Initial screening for possible templates using BLAST and SWISS-MODEL analyses against the PDB resolved structures.

PDB Code	Protein	Query Cover	E-Value	Percent Identity
5CE1_A	TMPS1	94%	7e−67	33.61
1Z8G_A	TMPS1	69%	4e−62	33.52
6KD5_B	TMPS13	47%	3e−64	45.19
6O1G_A	Plasma kallikrein	51%	2e−61	42.21

2.2. Homology Modelling

Homology models for TMPS2 were generated using SWISS-MODEL from the four identified templates, which were subsequently evaluated using MolProbability [37], ProSA [38] and Verify3D [39] to assess the quality of the models, the data of which is provided in Table 2.

Table 2. Validation data for the four TMPS2 homology models.

Model	Ramachandran Outliers	ProSA Z-Score	Verify3D
Model 1 (5CE1_A template)	3 (Ser208, Ala216, Arg255)	−8.67	95.38%
Model 2 (1Z8G_A template)	3 (Ser215, Leu248, Val415)	−8.76	91.59%
Model 3 (6KD5_B template)	2 (Pro301, Asn433)	−6.80	97.02%
Model 4 (6O1G_A template)	5 (Asn303, Asn304, Phe321, Pro305, Pro369)	−6.57	91.56%

Overlap of the four models (Figure 3) showed a high degree of homology between all four models in the serine peptidase domain, however the greatest coverage was with the models generated from TMPS1 crystal structures 5CE1 and 1Z8G, which have both the scavenger receptor cysteine rich (SRCR) and serine peptidase domains.

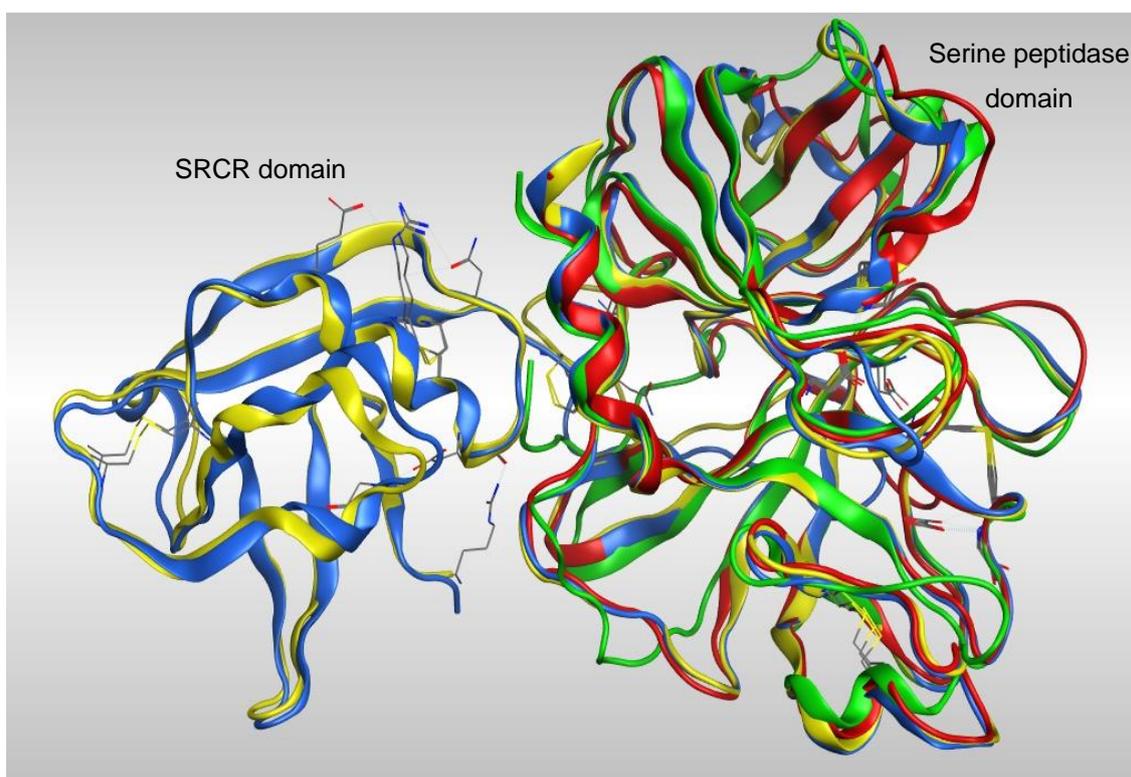


Figure 3. Overlap of the four TMPS2 models generated from crystal structures 5CE1_A (blue), Please ensure that intended meaning has been retained. 1Z8G_A (yellow), 6KD5_B (red) and 6O1G_A (green).

TMPS2 models 1 (5CE1_A) and 2 (1Z8G_A) were comparable using Ramachandran plot, Verify3D and ProSA (Tables S1 and S2 and Figures S1–S6 in Supplementary Materials). To evaluate further, both models were subjected to 150 ns molecular dynamics (MD) simulations using the Desmond programme of Maestro [40]. Model 1 started from the Root Mean Square Deviation (RMSD) 1.40 Å and rapidly reached an equilibrium plateau with a final RMSD of 2.86 Å at 150 ns. Model 2 started with a higher RMSD of 2.10 Å but showed more fluctuation and only began to plateau at around 90 ns

(RMSD 4.63 Å) with a final RMSD of 4.81 Å at 150 ns (Figure 4). Overlap of the proteins after MD simulation diverged in the SRCR domain but showed better alignment in the serine peptidase domain (Figure S7 in Supplementary Materials). From MD simulations, model 1 would appear to be the more stable system and so was chosen for further studies.

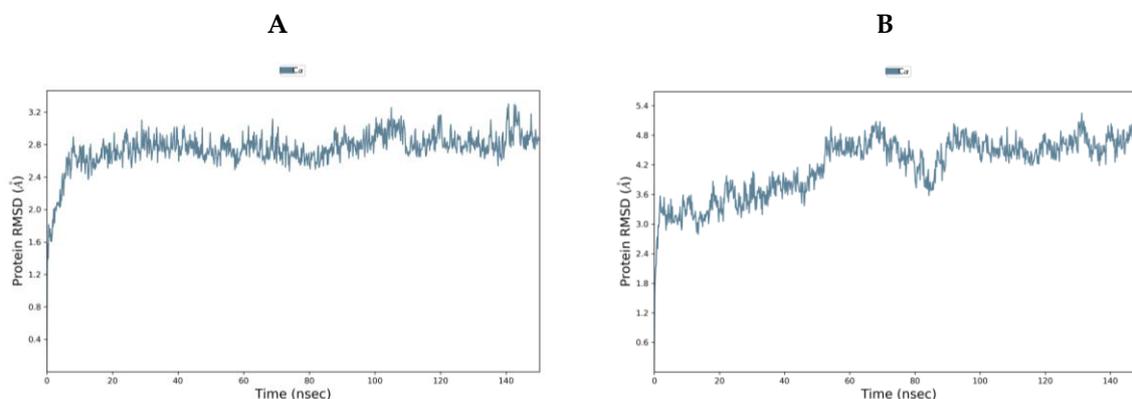


Figure 4. RMSD [Å] plot with respect to time in nanoseconds during 150 ns MD stimulation of (A) TMPS2 model 1 (5CE1_A template) and (B) TMPS2 model 2 (1Z8G_A template).

2.3. TMPS2 Structure Evaluation

A blastp analysis was performed using the NCBI server, query transmembrane protease serine against human sequences in the swissprot database [41]. Details for TMPS2 (O1593) and the closest five matches were obtained from the Uniprot server and are presented in Table 3.

Table 3. Active site and domains of the most homologous TMPS proteins.

Protein	Uniprot ID	Active Site His/Asp/Ser	Cleavage Site	Domains			X-ray Structure
				LDL-Receptor Class A	SRCR	Serine Peptidase	
TMPS1	P05981	203/257/353	162–163	-	54–151	163–405	5CE1, 1Z8G, 1P57
TMPS2	O15393	296/345/441	255–256	112–149	150–242	256–489	-
TMPS3	P57727	257/304/401	216–217	72–108	109–205	217–449	-
TMPS4	Q9NRS4	245/290/387	204–205	61–93	94–204	205–434	-
TMPS5	Q9H353	258/308/405	217–218	-	112–207	218–453	-
TMPS13	Q9BYE2	366/414/511	325–326	204–226	195–325	326–559	6KD5

SRCR = scavenger receptor cysteine rich; LDL = low-density lipoprotein.

A Clustal Omega [42] alignment was performed on the TMPS proteins. The three amino acids (His, Asp, Ser) that form the catalytic triad required for peptide bond hydrolysis, the cleavage site (Arg-Ile/Val-Val-Gly-Gly/Ile) responsible for proteolytic activation, and the disulphide forming cysteine residues are highly conserved (Figure S8 in Supplementary Materials). TMPS2 most closely aligns with TMPS3 (41.95%) as illustrated from the percent identity matrix and the cladogram (Figure 5).

Percent Identity Matrix – created by Clustal2.1

1:	sp P05981 TMPS1_HUMAN	100.00	33.95	34.20	34.03	33.78	36.44
2:	sp Q9NRS4 TMPS4_HUMAN	33.95	100.00	29.26	35.17	36.12	34.49
3:	sp Q9H3S3 TMPS5_HUMAN	34.20	29.26	100.00	33.18	32.94	34.89
4:	sp Q9BYE2 TMPS13_HUMAN	34.03	35.17	33.18	100.00	33.47	36.36
5:	sp O15393 TMPS2_HUMAN	33.78	36.12	32.94	33.47	100.00	41.95
6:	sp P57727 TMPS3_HUMAN	36.44	34.49	34.89	36.36	41.95	100.00



Figure 5. Percent identity matrix and phylogenetic cladogram generated from the Clustal alignment of TMPS proteins.

The protein structure is stabilised by five disulphide bonds, which are found in the SRCR non-catalytic domain [43], formed by Cys172–Cys231 and Cys185–Cys241, and in the serine peptidase domain by Cys281–Cys297, Cys410–Cys426 and Cys437–Cys465. The catalytic triad of His296, Asp345 and Ser441 spans the active site with Ser441 on one side and Asp345 and His296 on the other side of the S1 pocket (Figures 6 and 7). Asp435 sits at the base of the S1 pocket and Ser436 and Asp440 form the right wall. Asp435, Gly462 and Gly472 create a negatively charged S1 site and the combination of Ser441, Gly462 and Gly472 form a deep hydrophobic pocket to accommodate hydrophobic amino acids of a peptide substrate. The characteristic oxyanion hole of chymotrypsin family serine protease enzymes is formed by the backbone of Gly439 and Ser441. The polypeptide binding site is formed by residues 460–462 (Ser460, Trp461 and Gly462), which would be expected to form an antiparallel β -sheet with the backbone of the P1-P3 residues of a peptide substrate (Figure 7, Table S3 and Figure S9 in Supplementary Materials).

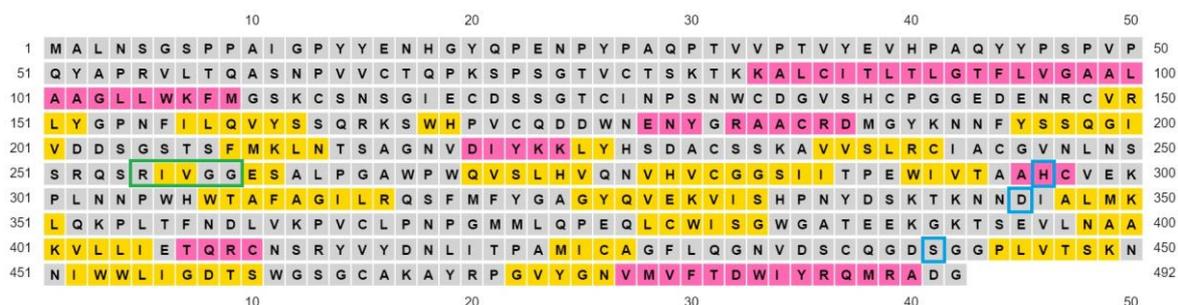


Figure 6. Amino acid sequence of TMPS2 showing α -helices coloured in pink, β -sheets coloured in yellow, loops coloured in grey, the catalytic triad residues are surrounded in cyan and the cleavage site is surrounded in green.

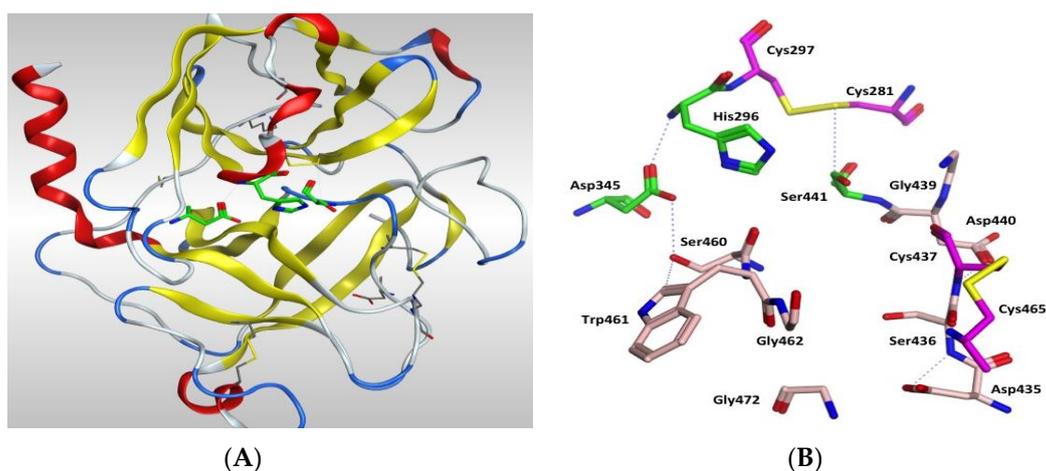


Figure 7. (A) 3D image of TMPS2 with the position of the catalytic triad (green) in the serine peptidase domain shown. The two six-stranded β -barrels (yellow) are positioned either side of the catalytic triad. (B) Serine peptidase active site showing the catalytic triad (green), the amino acids that form the S1 pocket (light pink) and the two disulfides Cys281-Cys297 and Cys437-Cys465 (magenta).

Having established the binding site and amino acids within the S1 pocket, docking was performed with ligands to generate TMPS2-ligand complexes.

2.4. Refinement of TMPS2 Homology Model after MDS and Docking Studies

The complexes were prepared by docking a database of ligands in the TMPS2-5CE1 homology model using MOE [44]. Using site finder, the active site chosen contained Ser441 and the majority of the S1 pocket amino acids: Glu389, Tyr416, Asp435, Ser436, Cys437, Gln438, Ser441, Thr459, Ser460, Trp461, Gly462, Ser463, Gly464, Cys465, Ala466, Arg470, Pro471, Gly472 and Val473.

The ligand-protein complexes were subjected to 200 ns molecular dynamics (MD) simulations using the Desmond programme of Maestro (Figure 8).

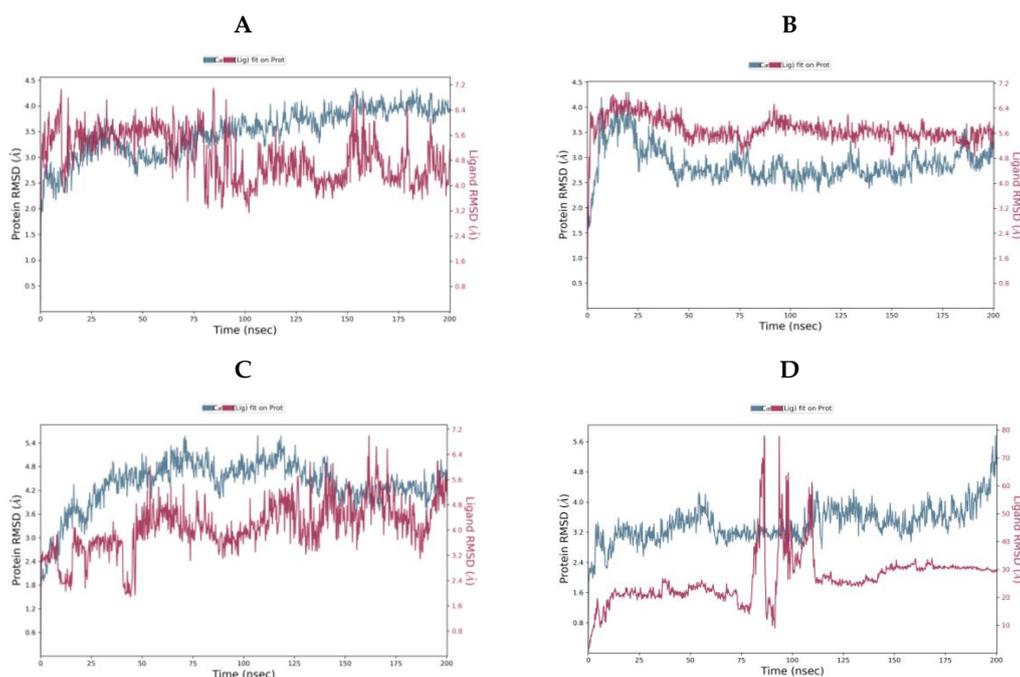


Figure 8. Protein-Ligand RMSD over the 200 ns MD simulation for (A) camostat (B) nafamostat (C) gabexate (D) sivelestat.

For camostat, nafamostat and gabexate, the protein equilibrates. However, for sivelestat, the RMSD is still increasing indicating the protein is undergoing a significant conformational change at the end of the simulation. Likewise, the ligand RMSD for sivelestat is significantly larger than the RMSD of the protein indicating that the ligand has diffused away from the initial binding site. The most stable system is the nafamostat-TMPS2 complex, which equilibrates rapidly, while there are more fluctuations with camostat and gabexate (Table 4).

Table 4. Protein-ligand RMSD (Å) at 0 and 200 ns.

Ligand-Complex	RMSD (Å) 0 ns		RMSD (Å) 200 ns	
	Protein	Ligand	Protein	Ligand
Camostat	2.18	5.32	3.90	5.05
Nafamostat	1.83	4.85	3.21	5.83
Gabexate	2.20	2.98	4.35	5.87
Sivelestat	2.25	2.21	5.07	29.42

Nafamostat was optimal with respect to binding, with the phenyl guanidine group situated within the S1 pocket through a salt bridge with Asp435 and H-bonding interactions with Asp435, Ser436, Ser463 and Gly462 (Figure 9). The phenyl ring sits in the hydrophobic pocket composed of Thr459, Ser460, Trp461 and Val473. The carbonyl oxygen of the ester linkage forms a water mediated H-bonding interactions with Ser441 and the amidine group forms direct and water mediated H-bonding interactions with Glu299 (Figure 9). Visualization using MOE also identified an aryl-H binding interaction between the phenyl ring of nafamostat and the backbone of Cys437 (Figure 9B). As previously reported [45], the catalytic steps for the cleavage of peptide-like bonds involve an acyl-enzyme intermediate formation between the substrate and Ser441 followed by the hydrolysis of the acyl-enzyme intermediate, releasing the cleaved substrate and restoring the active form of the enzyme. Nafamostat with its guanidinobenzoate moiety can inhibit TMPS2 by mimicking the natural substrates, where the ester group, resembling a peptide bond, can react with the catalytic serine forming the acyl-enzyme intermediate while the guanidinobenzoate group becomes linked to the catalytic serine Ser441, rendering nafamostat an effective chemical inhibitor, which meets our findings after MDS.

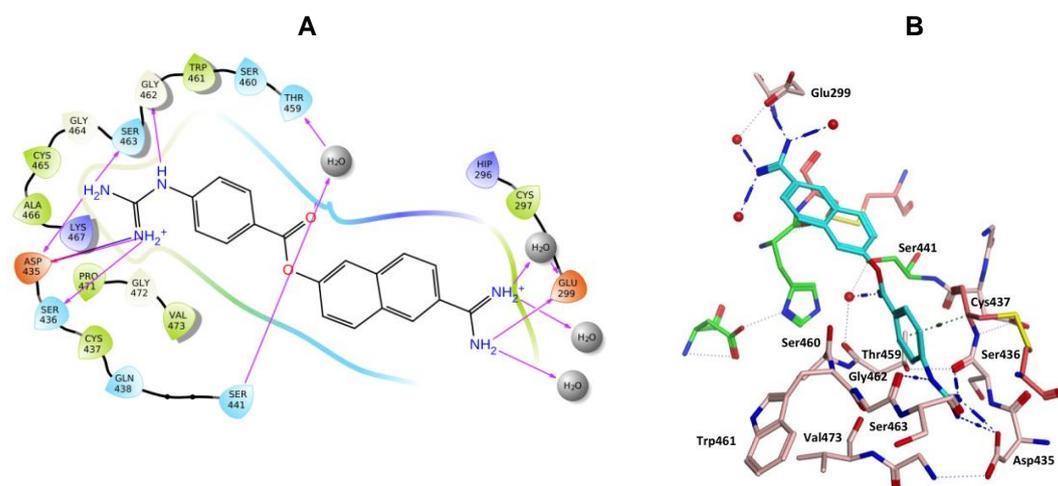


Figure 9. (A) 2D Ligand interactions and (B) 3D visualization of nafamostat (cyan).

Camostat and gabexate were similar to nafamostat in the binding profiles of the guanidine group with the salt bridge formed with Asp435 and H-bonding interactions, direct or water mediated with Asp435, Ser436, Ser463 and Gly464 for camostat (Figure 10), and with Asp435, Gly464 and Arg470 for gabexate (Figure 11). Camostat formed H-bonding interaction between the ester and amide carbonyl

oxygen atoms on the opposite side with water solvation molecules, while gabexate formed H-bonding interaction between the two ester carbonyl oxygens and water of solvation and the terminal ester also formed a H-bonding interaction with His296 (Figures 10 and 11).

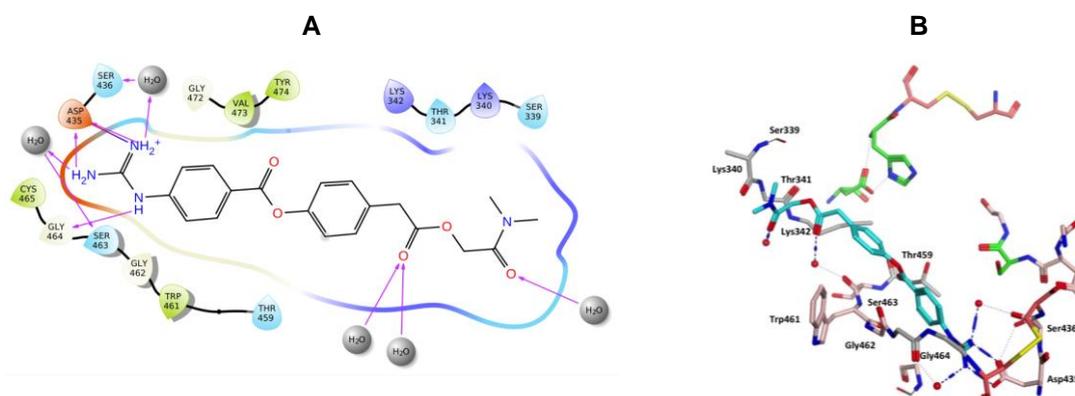


Figure 10. (A) 2D Ligand interactions and (B) 3D visualization of camostat (cyan).

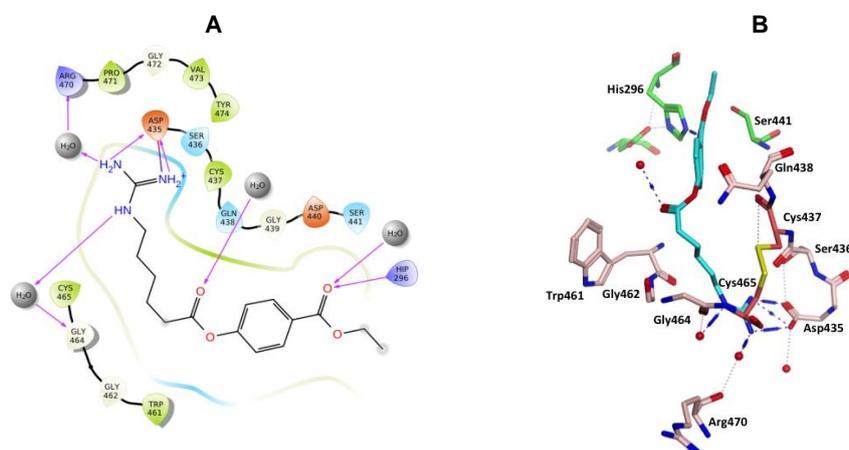


Figure 11. (A) 2D Ligand interactions and (B) 3D visualization of gabexate.

Sivelestat did not bind in the S1 binding site and, as indicated from the protein-ligand RMSD (Figure 8D), had moved away from the original site. Figure 12 shows the original position of sivelestat (cyan) compared with the final position (orange) after 200 ns MD simulation.

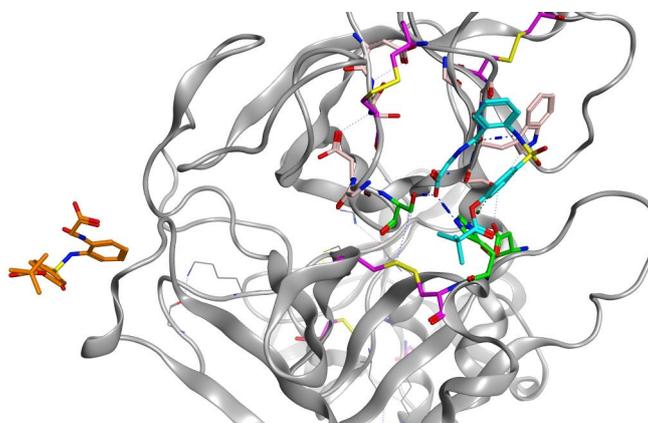


Figure 12. 3D visualization of sivelestat showing the original position (cyan) in the S1 binding pocket and after 200 ns MD simulation (orange).

For the ligand-TMPS2 complexes of camostat, nafamostat and gabexate complexes, the mean ΔG (bind) was calculated [46] over two time frames, from 100–200 ns and the final 10 ns (190–200 ns) of the MD simulation (Table 5). The ΔG values indicate positioning within the TMPS2 protein was optimal for nafamostat, with ΔG values of -60.99 ± 4.27 and -61.47 ± 4.44 kcal/mol, respectively over the two timeframes. Thus, nafamostat would appear to have better binding affinity, which supports the observations from protein-ligand RMSD (Figure 8B) and interaction figures (Figure 9).

Table 5. Computational binding affinity (ΔG) of the ligand-protein complexes.

Complex	ΔG (kcal/mol) 100–200 ns	ΔG (kcal/mol) 190–200 ns
Camostat-TMPS2	-48.87 ± 5.07	-46.24 ± 4.42
Nafamostat-TMPS2	-60.99 ± 4.27	-61.47 ± 4.44
Gabexate-TMPS2	-54.37 ± 4.86	-51.91 ± 4.86

Protein interactions with the ligand could be categorized by type and summarized, as shown in the plots below (Figure 13), where the protein-ligand interactions (or ‘contacts’) are categorized into four types: hydrogen bonds, hydrophobic, ionic and water bridges. This can also be visualised from the ligand-protein contacts 2D summary over the course of the 200 ns simulation (Figure S10 in Supplementary Materials).

Limitations of this study:

Although MD is a robust computational method for the study of protein-ligand complexes, the results described here are theoretical, therefore experimental enzyme inhibition (IC_{50}) assays, with the appropriate replicas, against TMPS2 are required to validate the theoretical findings.

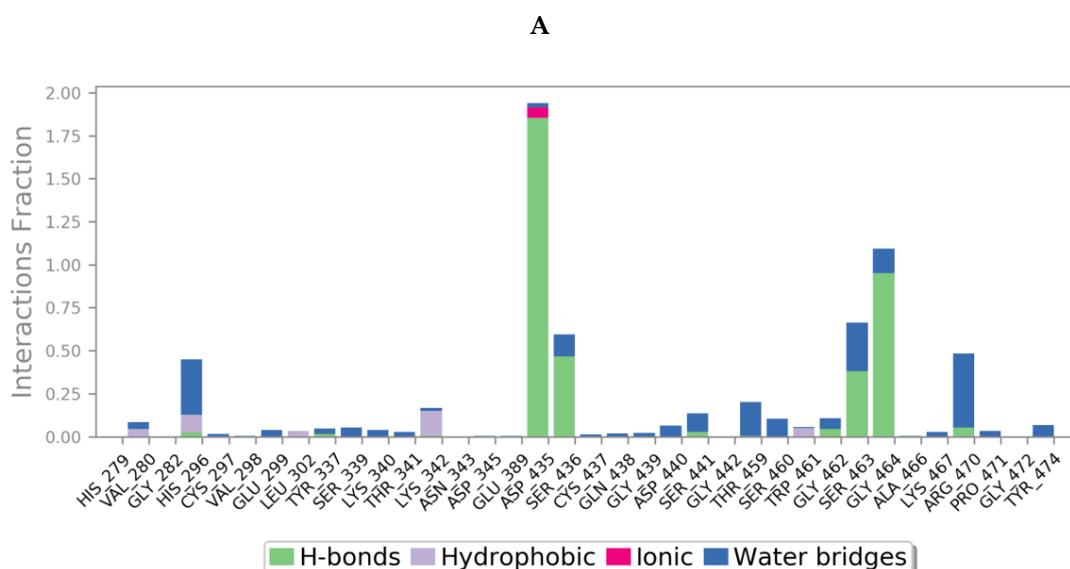


Figure 13. Cont.

3. Materials and Methods

3.1. Transmembrane Protease Serine 2 (TMPS2_HUMAN) Molecular Modelling

The homologous sequences for building the 3D structure were searched using NCBI-BlastP (basic local alignment search tool). For alignments, BLOSUM62 substitution matrix was used as a comparison matrix. Clustal Omega from the European Bioinformatics Institute (EBI) was used to determine the conserved areas between the selected proteins and target protein TMPS2. SWISS-MODEL server accessible via ExPASy web server was used in comparative modelling of TMPS2. The constructed models were then evaluated.

3.2. Refinement and Evaluation of the Model

The Ramachandran plot was provided by the MolProbity server at Duke Biochemistry, US [37]. The PDB files of the homology models were uploaded. H-atoms were added, and the gaps were filled in the protein backbone with JiffiLoop (beta-test). The results were displayed and subdivided into regions of favoured, allowed, and disallowed [47]. The output data demonstrated a comparison between atom contacts, protein geometry, and peptide omegas in tables. A file of the best model in a PDB format was uploaded on ProSA (Protein Structure Analysis) [38] and the Z-score was calculated. Three parameters for each amino acid; secondary structure, degree of buried surface area, and fraction of side chain area that was covered by polar atoms were evaluated by Verify 3D service [39]. For each residue in the structure, these three parameters were evaluated, and a correlation was determined between the observed parameters and the ideal ones to which they were assigned.

3.3. Molecular Dynamic Simulation (MDS) and Docking Studies

Molecular docking and molecular dynamics simulations were performed as previously described [48]. Briefly docking studies using the TMPS2 homology model to generate PDB files of the TMPS2-ligand complexes were performed using MOE [44] until an RMSD gradient of 0.01 Kcal/mol/Å with the MMFF94 forcefield (ligands) and partial charges were automatically calculated. Docking was performed using the Alpha Triangle placement to determine the poses, refinement of the results was done using the MMFF94 forcefield, and rescoring of the refined results using the London ΔG scoring function was applied. The output database dock file was created with different poses for each ligand and arranged according to the final score function (S), which is the score of the last stage that was not set to zero.

Molecular dynamics simulations were run on the TMPS2-ligand complexes with the PDB files first optimised with protein preparation wizard in Maestro by assigning bond orders, adding hydrogen, and correcting incorrect bond types. A default quick relaxation protocol was used to minimise the MD systems with the Desmond programme. The orthorhombic water box allowed for a 10 Å buffer region between protein atoms and box sides. Overlapping water molecules were deleted, and the systems were neutralised with Na⁺ ions and salt concentration 0.15 M. Force-field parameters for the complexes were assigned using the OPLS_2005 forcefield, that is, a 200 ns molecular dynamic run in the NPT ensemble (T = 300 K) at a constant pressure of 1 bar. Energy and trajectory atomic coordinate data were recorded at each 1.2 ns. Prime/MMGBAS, available in Schrödinger Prime suite, was used to calculate the binding free energy of the ligands with TMPS2.

$$\Delta G(\text{bind}) = E_{\text{complex}}(\text{minimised}) - (E_{\text{ligand}}(\text{minimised}) + E_{\text{receptor}}(\text{minimised}))$$

Two mean $\Delta G(\text{bind})$ values were calculated from (1) each 5 frames of the final 100 ns, and (2) each frame of the final 10 ns of the MD simulation. The average generated ΔG was from each energy minimised frame using the equation shown above.

4. Conclusions

In this study, a high-quality homology model of the 3D human TMPS2 was constructed to enable investigation of TMPS2 as a promising drug target for the development of therapeutics in the treatment

of SARS-CoV2. Molecular dynamics simulations were performed, and the active site conserved amino acids were selected as a docking site for camostat, nafamostat, gabexate, and sivelestat. Quite clearly nafamostat has optimal binding owing to its ability to anchor both ends of its structure through a salt bridge with Asp345 and the guanidine group, and H-bonding with Glu299 and the amidine group. Camostat and gabexate are able to anchor the guanidine group effectively via salt bridge formation with Asp345, however the dimethylamide of camostat is not able to form an anchor with the protein, while gabexate, which is a more flexible molecule, was able to form a H-bonding interaction with His296. Computational binding affinity (ΔG) studies support the findings of the MD studies and the requirement of the basic guanidine moiety, or equivalent functional group, in the salt bridge formation with Asp345. Sivelestat, which lacks a basic functional group, is unable to form a tether with Asp435 and rather diffuses from the S1 site to a position where the acidic carboxylate group of sivelestat can form a salt bridge with a basic amino acid, Arg316 (Figure S11 in Supplementary Materials).

The studies also indicate the preference for a basic moiety at the other end of the inhibitor to form an ionic bond with an acidic amino acid and secure the inhibitor within the S1 binding pocket as observed with nafamostat with an amidine group. The ability of gabexate to tether the other end of its structure through a H-bonding interaction provides some scope to investigate additional H-bond forming functional groups. In addition, the optimal length and flexibility of a TMPS2 inhibitor ligand to fit and bind within the S1 pocket would warrant further investigation through inhibitor design.

Supplementary Materials: Table S1: Stereochemical quality assessment of TMPS2 model 1 (5CE1_A template), Table S2: Stereochemical quality assessment of TMPS2 model 2 (1Z8G_A template), Table S3: Comparison between TMPS2 and the template 5CE1_A, Figure S1: Ramachandran plot for TMPS2 model 1 (5CE1_A template); 92.2% (317/344) of all residues were in favoured (98%) regions, 99.1% (341/344) of all residues were in allowed (>99.8%) regions. There were 3 outliers (ϕ , ψ): 208 Ser (−58.3, −174.6), 216 Ala (−25.8, −45.7) and 255 Arg (88.3, 25.7), Figure S2: Ramachandran plot for TMPS2 model 2 (1Z8G_A template); 94.2% (323/343) of all residues were in favoured (98%) regions, 99.1% (340/343) of all residues were in allowed (>99.8%) regions. There were 3 outliers (ϕ , ψ): 215 Ser (−56.0, 2.4), 248 Leu (−50.9, −74.7) and 415 Val (−104.8, −94.8), Figure S3: Z-scores of TMPS2 model 1 (5CE1_A template) determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue), and local model quality by plotting energies as a function of amino acid sequence position, a positive value indicates erroneous parts; most of the amino acids gave negative values, Figure S4: Z-scores of TMPS2 model 2 (1Z8G_A template) determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue), and local model quality by plotting energies as a function of amino acid sequence position, a positive value indicates erroneous parts; most of the amino acids gave negative values, Figure S5: Verify 3D results for model 1 (5CE1_A template) showing that the model passes in the 3D/1D profile, Figure S6: Verify 3D results for model 2 (1Z8G_A template) showing that the model passes in the 3D/1D profile, Figure S7: Overlap of 5CE1_A (blue) and 1Z8G_A (yellow) after MD simulation, Figure S8: Sequence alignment of TMPS enzymes using Clustal O in which "*" means that the residues are identical, ":" means that conserved substitutions have been observed, "." means that semi-conserved substitutions are observed. The SRCR domain is indicated in blue lettering, cysteine residues involved in disulphide bond formation are highlighted in yellow, the catalytic triad residues are highlighted in cyan and the cleavage site is highlighted in pink, Figure S9: Comparison between the secondary structure of TMPS2 and the template 5CE1_A after alignment and superimposition using MOE software, showing α -helices as red lines, β -sheets as yellow arrows and loops as blue lines, Figure S10: A schematic of detailed ligand atom interactions of (A) camostat, (B) nafamostat, (C) gabexate and (D) sivelestat with the protein residues of TMPS2 protein. Interactions that occur more than 30.0% of the simulation time in the selected trajectory (0 through 200 ns) are shown, Figure S11: 2D Ligand interactions of sivelestat in TMPS2 protein.

Author Contributions: Conceptualization, S.M.K., C.S., R.M.K., A.S.A.Y., N.A.N., and M.S.N.; methodology, C.S., and S.M.K.; software, C.S., and S.M.K.; validation, S.M.K., C.S., R.M.K. and N.A.N.; formal analysis, S.M.K., C.S., R.M.K., and G.E.; investigation, S.M.K., C.S., A.S.A.Y., M.S.N., S.A.-R., and N.A.N.; writing—original draft preparation, S.M.K., C.S., R.M.K., N.A.N., M.S.N., and A.S.A.Y.; writing—review and editing, C.S., S.M.K., and A.S.A.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Researchers Supporting Project Number (RSP-2020/120) King Saud University, Riyadh, Saudi Arabia for financial support.

Acknowledgments: Molecular dynamics simulations were undertaken using the supercomputing facilities at Cardiff University operated by Advanced Research Computing at Cardiff (ARCCA) on behalf of the Cardiff Supercomputing Facility and the HPC Wales and Supercomputing Wales (SCW) projects. We acknowledge support of the latter, which is part-funded by the European Regional Development Fund (ERDF) via the Welsh Government. The authors extend their appreciation to the Researchers Supporting Project Number (RSP-2020/120) King Saud University, Riyadh, Saudi Arabia for financial support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Drosten, C.; Günther, S.; Preiser, W.; Werf, S.; Brodt, H. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1967–1976. [[CrossRef](#)] [[PubMed](#)]
2. Barile, E.; Baggio, C.; Gambini, L.; Shiryayev, S.A.; Strongin, A.Y.; Pellicchia, M. Potential Therapeutic Targeting of Coronavirus Spike Glycoprotein Priming. *Molecules* **2020**, *25*, 2424. [[CrossRef](#)] [[PubMed](#)]
3. Shi, M.; Lin, X.; Tian, J.; Chen, L.; Chen, X. Redefining the Invertebrate RNA Virosphere. *Nature* **2016**, *540*, 539–543. [[CrossRef](#)]
4. Hu, D.; Zhu, C.; Ai, L.; Ting, H.; Wang, Y. Genomic Characterization and Infectivity of a Novel SARS-like Coronavirus in Chinese Bats. *Emerg. Microbes Infect.* **2018**, *7*, 1–10. [[CrossRef](#)]
5. Tao, Z.; Tian, J.; Pei, Y.; Yuan, M.; Zhang, Y.; Dai, F. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579*, 265–269.
6. Zirkel, F.; Kurth, A.; Quan, P.; Briese, T.; Ellerbrok, H. An Insect Nidovirus Emerging from a Primary Tropical Rainforest. *mBio* **2011**, *2*, 21–23. [[CrossRef](#)]
7. Gorbalenya, A.E.; Enjuanes, L.; Ziebuhr, J.; Snijder, E.J. Nidovirales: Evolving the Largest RNA Virus Genome. *Virus Res.* **2006**, *117*, 17–37. [[CrossRef](#)]
8. Lewicki, D.N.; Gallagher, T.M. Quaternary Structure of Coronavirus Spikes in Complex with Carcinoembryonic Antigen-related Cell Adhesion Molecule Cellular Receptors. *J. Biol. Chem.* **2002**, *277*, 19727–19734. [[CrossRef](#)]
9. Imai, Y.; Kuba, K.; Rao, S.; Huan, Y.; Guo, F. Angiotensin-converting Enzyme 2 Protects from Severe Acute Lung Failure. *Nature* **2005**, *436*, 112–116. [[CrossRef](#)]
10. Kalra, R.S.; Tomar, D.; Meena, A.S.; Kandimalla, R. SARS-CoV-2, ACE2, and Hydroxychloroquine: Cardiovascular Complications, Therapeutics, and Clinical Readouts in the Current Settings. *Pathogens* **2020**, *9*, 546. [[CrossRef](#)] [[PubMed](#)]
11. Li, W.; Moore, M.J.; Vasilieva, N.; Sui, J. Angiotensin-converting Enzyme 2 is a Functional Receptor for the SARS Coronavirus. *Nature* **2003**, *426*, 450–454. [[CrossRef](#)] [[PubMed](#)]
12. Hofmann, H.; Pyrc, K.; Hoek, L.; Geier, M.; Berkhout, B. Human Coronavirus NL63 Employs the Severe Acute Respiratory Syndrome Coronavirus Receptor. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7988–7993. [[CrossRef](#)] [[PubMed](#)]
13. Yushun, W.; Jian, S.; Rachel, G.; Ralph, S.; Baric, F.L. Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of. *J. Virol.* **2020**, *94*, 1–9.
14. Lu, R.; Zhao, X.; Li, J. Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *Lancet* **2020**, *395*, 565–574. [[CrossRef](#)]
15. Wrapp, D.; Wang, N.; Corbett, K.S. Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science* **2020**, *367*, 1260–1263. [[CrossRef](#)] [[PubMed](#)]
16. Di Paola, L.; Hadi-Alijanvand, H.; Song, X.; Hu, G.; Giuliani, A. The Discovery of a Putative Allosteric Site in the SARS-CoV-2 Spike Protein Using an Integrated Structural/Dynamic Approach. *J. Proteome Res.* **2020**. [[CrossRef](#)]
17. Glowacka, I.; Bertram, S.; Müller, M.; Allen, P.; Soilleux, E. Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *J. Virol.* **2011**, *85*, 4122–4134. [[CrossRef](#)]
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*, 12658–12664. [[CrossRef](#)]
19. Shulla, A.; Heald-Sargent, T.; Subramanya, G.; Zhao, J.; Perlman, S.; Gallagher, T. A Transmembrane Serine Protease is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. *J. Virol.* **2011**, *85*, 873–882. [[CrossRef](#)]
20. Iwata-yoshikawa, N.; Okamura, T.; Shimizu, Y.; Hasegawa, H. TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *J. Virol.* **2019**, *93*. [[CrossRef](#)]

21. Shirato, K.; Kawase, M.; Matsuyama, S. Wild-type Human Coronaviruses Prefer Cell-surface TMPRSS2 to Endosomal Cathepsins for Cell Entry. *J. Virol.* **2020**, *517*, 9–15. [[CrossRef](#)] [[PubMed](#)]
22. Eleftheriou, P.; Amanatidou, D.; Petrou, A.; Geronikaki, A. In Silico Evaluation of the Effectivity of Approved Protease Inhibitors against the Main Protease of the Novel SARS-CoV-2 Virus. *Molecules* **2020**, *25*, 2529. [[CrossRef](#)] [[PubMed](#)]
23. Dong, M.; Zhang, J.; Ma, X. ACE2, TMPRSS2 distribution and extrapulmonary organ injury in patients with COVID-19. *Biomed. Pharmacother.* **2020**, *131*, 110678. [[CrossRef](#)]
24. Strobe, J.D.; Figg, W.D. TMPRSS2: Potential Biomarker for COVID-19 Outcomes. *J. Clin. Pharmacol.* **2020**, *60*, 801–807. [[CrossRef](#)] [[PubMed](#)]
25. Kawase, M.; Shirato, K.; Van der Hoek, L. Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *J. Virol.* **2012**, *86*, 6537–6545. [[CrossRef](#)] [[PubMed](#)]
26. Zhou, Y.; Vedantham, P.; Lu, K. Protease inhibitors targeting coronavirus and filovirus entry. *Antivir. Res.* **2015**, *116*, 76–84. [[CrossRef](#)]
27. Yamamoto, M.; Matsuyama, S.; Li, X. 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.* **2016**, *60*, 6532–6539. [[CrossRef](#)]
28. Lin, C.C.; Lin, L.J.; Wang, S.D. The effect of serine protease inhibitors on airway inflammation in a chronic allergen-induced asthma mouse model. *Mediators Inflamm.* **2014**, 879326. [[CrossRef](#)]
29. Yamaya, M.; Shimotai, Y.; Hatachi, Y. The serine protease inhibitor camostat inhibits influenza virus replication and cytokine production in primary cultures of human tracheal epithelial cells. *Pulm. Pharmacol. Ther.* **2015**, *33*, 66–74. [[CrossRef](#)]
30. Kim, H.S.; Lee, K.E.; Oh, J.H. Cardiac arrest caused by nafamostat mesylate. *Kidney Res. Clin. Pract.* **2016**, *35*, 187–189. [[CrossRef](#)]
31. Chang, J.; Lee, S.; Kim, H.; Cho, Y. Nafamostat for Prophylaxis against Post-Endoscopic Retrograde Cholangiopancreatography Pancreatitis Compared with Gabexate. *Gut Liver* **2009**, *3*, 205–210. [[CrossRef](#)]
32. Han, S.; Han, W.; Song, H. Validation of Nafamostat Mesilate as an Anticoagulant in Extracorporeal Membrane Oxygenation: A Large-Animal Experiment. *Korean J. Thorac. Cardiovasc. Surg.* **2018**, *51*, 114–121. [[CrossRef](#)] [[PubMed](#)]
33. Jang, S.; Rhee, J.Y. Three cases of treatment with nafamostat in elderly patients with COVID-19 pneumonia who need oxygen therapy. *Int. J. Infect Dis.* **2020**. [[CrossRef](#)] [[PubMed](#)]
34. Protein knowledgebase (UniProtKB). Available online: <https://www.uniprot.org/uniprot/O15393> (accessed on 27 May 2020).
35. Swiss Institute of Bioinformatics, University of Basel. Available online: <https://swissmodel.expasy.org/> (accessed on 3 June 2020).
36. National Center for Biotechnology Information (NCBI). Available online: <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins> (accessed on 5 June 2020).
37. Williams et al. (2018) MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Science* **27**: 293–315. Available online: <http://molprobity.biochem.duke.edu/> (accessed on 7 June 2020).
38. Protein Structure Analysis. Available online: <https://prosa.services.came.sbg.ac.at/prosa.php> (accessed on 7 June 2020).
39. Verify 3D Standalone Server. Available online: <https://servicesn.mbi.ucla.edu/Verify3d/> (accessed on 8 June 2020).
40. *Schrödinger Release 2020-1: Desmond Molecular Dynamics System, D. E.*; Shaw Research: New York, NY, USA; Maestro-Desmond Interoperability Tools, Schrödinger: New York, NY, USA, 2020.
41. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2008**, *10*, 421. [[CrossRef](#)]
42. EMBL-EBI Search and Sequence Analysis Tools APIs. Available online: <https://www.ebi.ac.uk/Tools/msa/clustalo/> (accessed on 8 June 2020).
43. Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162. [[CrossRef](#)]
44. *Molecular Operating Environment (MOE)*; Chemical Computing Group ULC: Montreal, QC, Canada, 2019.

45. Hempel, T.; Raich, L.; Olsson, S.; Azouz, N.; Klingler, L.; Rothenberg, M.; Noé, F. Molecular mechanism of SARS-CoV-2 cell entry inhibition via TMPRSS2 by Camostat and Nafamostat mesylate. *bioRxiv* **2020**. [[CrossRef](#)]
46. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Mod.* **2011**, *51*, 69–82. [[CrossRef](#)]
47. Morris, A.L.; MacArthur, M.W.; Hutchinson, E.G.; Thornton, J.M. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, *12*, 345–364. [[CrossRef](#)] [[PubMed](#)]
48. Binjubair, F.A.; Parker, J.E.; Warrilow, A.G. Small molecule inhibitors targeting sterol 14 α -Demethylase (CYP51): Synthesis, Molecular Modelling and Evaluation against *Candida albicans*. *ChemMedChem* **2020**, *15*, 1294–1309. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).