Structural Modeling of the TMPRSS Subfamily of Host Cell Proteases Reveals Potential Binding Sites

Diego E. Escalante,[1] Austin Wang[1] and David M. Ferguson[1,2]

[1] Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455

[2] Center for Drug Design, University of Minnesota, Minneapolis, MN 55455

Corresponding Author: ferguson@umn.edu

This file includes

- Main text
- Figures 1-6
- Supplementary Information

Abstract

The transmembrane protease serine subfamily (TMPRSS) has at least eight members with known protein sequence: TMPRSS2, TMPRRS3, TMPRSS4, TMPRSS5, TMPRSS6, TMPRSS7, TMPRSS9, TMPRSS11, TMPRSS12 and TMPRSS13. A majority of these TMPRSS proteins have key roles in human hemostasis as well as promoting certain pathologies, including several types of cancer. In addition, TMPRSS proteins have been shown to facilitate the entrance of respiratory viruses into human cells, most notably TMPRSS2 and TMPRSS4 activate the spike protein of the SARS-CoV-2 virus. Despite the wide range of functions that these proteins have in the human body, none of them have been successfully crystallized. The lack of structural data has significantly hindered any efforts to identify potential drug candidates with high selectivity to these proteins. In this study, we present homology models for all members of the TMPRSS family including any known isoform (the homology model of TMPRSS2 is not included in this study as it has been previously published). The atomic coordinates for all homology models have been refined and equilibrated through molecular dynamic simulations. The structural data revealed potential binding sites for all TMPRSS as well as key amino acids that can be targeted for drug selectivity.

Introduction

The Type II Transmembrane Serine Protease (TTSP) family consists of 4 different subfamilies that are differentiated by specific domains: the HAT/DESC subfamily, the Hepsin/TMPRSS subfamily, the Matriptase subfamily, and the Corin subfamily.[1] The proteins are created as single-chain, inactive proenzymes that are activated by cleaving the basic amino acid residue, arginine or lysine, within the conserved activation motif between the pro-domain and catalytic domain.[1, 2] Once activated, they remain associated with the membrane through disulfide bonds between the catalytic and transmembrane domains.[1] There are currently 14 different TMPRSS proteins: TMPRSS2, TMPRSS3, TMPRSS4, TMPRSS5, TMPRSS6, TMPRSS7, TMPRSS9, TMPRSS11A, TMPRSS11B, TMPRSS11D, TMPRSS11E, TMPRSS11F, TMPRSS12, and TMPRSS13.[3] These proteases have been found to be in multiple organs and tissues throughout the body.[4-10] This subfamily has a diverse set of physiological functions like homeostasis and proteolytic cascades, and while not all functions have been found, their involvement in various pathogenicities is becoming apparent as shown in Table 1 below.[11] Several inhibitors have also been identified including aprotinin, camostat, leupeptin, AEBSF, nafamostat, and gabexate.[12]

Table 1 Known roles of TMPRSS proteins.

Protein Name	Role in the Body
TMPRSS 2	essential role in the pathogenesis of certain viruses like SARS-CoV-2 and MERS-CoV [7].
TMPRSS 3 & TMPRSS 5	Expressed in the inner ear as part of the cochlear hair cells where mutations result in hearing loss [1][4].
TMPRSS 4	Overexpressed in pancreatic carcinomas [5].
TMPRSS 6	Regulates iron homeostasis through controlling hepcidin levels, and mutations cause an iron-refractory iron deficiency anemia (IRIDA) [6].
TMPRSS 9	Improves the invasive capabilities of the progression of pancreatic cancer [8].
TMPRSS 11A	Down-regulated in esophagus cancer [9].
TMPRSS 11B	Promotes lung cancer through glycolytic metabolism and increased lactate export [10].
TMPRSS 12	Commonly expressed in colorectal cancer [11].

Recently, the Hepsin/TMPRSS subfamily has recently received considerable attention not only due to the potential role these enzymes play in the development of cancers but to the function TMPRSS2 plays in mediating the virulence of SARS-CoV-2. The TMPRSS2 protein and ACE2 receptor are key elements of the main pathway by which SARS-CoV-2 (and related CoV pathogens) are internalized by lung cells.[12] The virus uses TMPRSS2 to cleave the Spike protein at a specific site (defined by Arg255 and Ile256) to activate membrane insertion.[12] ACE2 recruits the Spike protein to the host cell surface through specific interaction with the receptor binding domain of the Spike protein.[12] The Spike protein contains two functional subunits: the S1 subunit that allows binding of the virus to the host cell surface receptor and the S2 subunit that allows the fusion of the viral and cellular membranes.[13, 14] In vivo studies have shown TMPRSS2-knockout mice with down-regulated TMPRSS2 show less severe lung pathology as compared to controls when exposed to SARS-CoV-2.[15, 16] Additional work has shown that TMPRSS2 is the main processing enzyme for virus entry in lung cells.[17]

TMPRSS2 has therefore emerged as a primary target for the design and discovery of drugs for treating SARS-CoV-2 infections. An excellent starting point in the search for potent drugs is camostat. This compound is clinically approved for use in treating pancreatitis in Japan and is a known inhibitor of TMPRSS2. While camostat has shown efficacy in preventing mice infected with SARS-CoV from dying,[12] it is a pan-trypsin-like serine protease inhibitor and is not highly selective for TMPRSS2. Similarly, there are no known or reported small molecules that are highly selective for any of the other member of the TMPRSS family. The main reason for this is the lack of crystal structures of the TMPRSS protein subfamily. Any structural information on these proteases could lead to the development of drugs more selective than the current alternatives. This study extends prior work on camostat bound to TMPRSS2 to identify common elements of recognition across the TMPRSS sub-family.

Computational Methods

Homology Modeling. The 14 proteins (TMPRSS2, TMPRSS3, TMPRSS4, TMPRSS5, TMPRSS6, TMPRSS7, TMPRSS9, TMPRSS11A, TMPRSS11B, TMPRSS11D, TMPRSS11E, TMPRSS11F, TMPRSS12, TMPRSS13) amino acid sequence was obtained from the UnitProt database (Gene ID: O15393, P57727, Q9NRS4, Q9H3S3, Q9DBI0, Q7RTY8, Q7Z410, Q6ZMR5, Q86T26, O60235, Q9UL52, Q6ZWK6, Q86WS5, Q9BYE2, respectively), and the crystal structures with high sequence identity, available in the Protein Databank (PDB), were retrieved through the BLASTp algorithm. The differences between proteins will be discussed in the results section. Across the 14 receptors, 25 crystal structures were used as templates to build the TMPRSS protein models. Of the 25 crystal structures, the 3 notable crystals structures were a urokinase-type plasminogen activator with a position 190 alanine mutation (1O5E), DESC1, part of the TTSP family, (2OQ5), and a human plasma kallikrein (6O1G) each with a known ligand included with each crystal structure. The 3 ligands included were 6-Chloro-2-(2-Hydroxy-Biphenyl-3-yl)-1H-indole-5-carboxamidine, benzamide, and 7SD from 1O5E, 2OQ5, and 6O1G, respectively, and were kept in the homology models created. The Pfam database was used to obtain the amino acid region corresponding to the trypsin domain for each respective TMPRSS protein. Found through the Pfam database, all the crystal structures were truncated to only their trypsin domains. The Schrodinger prime homology model suite program was used to align each sequence and identify any conserved secondary structure assignments. To construct a homology model for each TMPRSS protein, each protein had at least 5 crystal structures chosen which are listed in Supplementary Table 1. Once completed, a final consensus model was constructed by using each homology model that was built specifically for the respective protein. The final model for each TMPRSS protein is labeled as TMPRSSxhm with x being the specific name of the protein. The models were refined through a process of short molecular dynamics (MD) simulations that will be described below.

Molecular Dynamics. All of the molecular dynamic simulation stages were completed by using the SANDER.MPI function of the AMBER 18 software package unless otherwise stated[18]. The ligand structures were pre-processed with the Antechamber package to assign AM1-BCC partial charges. The ligand and enzyme structures were processed using LEaP to assign ff15ipq[19] or gaff[20] and ff14SB[21] force field parameters for the enzyme and ligand, respectively. All the complexes were submerged within a periodic box of TIP3P water with a 10Å buffer region. Each molecule was initially minimized by using the steepest descent method for 100 steps followed by 9900 steps of conjugate gradient minimization. Then, a stepwise heating procedure occurred where the system temperature was slowly ramped from 0K to 300K over 15,000 steps and then relaxed over 5,000 steps to where the average temperature was kept constant at 300k using a weak-coupling algorithm. For both heating stages, the position of all the non-solvent atoms were restrained with a harmonic potential with a force constant of 25 kcal/mol-Å. After that, a two-step procedure was performed in which the average pressure was maintained at 1bar for 20,000 steps through pressure relaxation time of 0.2ps and the Berendsen barostat. A harmonic potential with a force constant of 5 kcal/mol-Å restrained the position of all non-solvent atoms. A relaxation stage of 20,000 steps followed in which the pressure relaxation time was increased to 2ps, and the only position of the receptor Cα was restrained by a harmonic potential with a force constant of 0.5 kcal/mol-Å. Lastly, all the production runs were carried out without any restraints and were kept at a consistent 300K temperature and 1bar pressure.

Results

All fifteen sequences have a high enough degree of identity to template models (>40%), as shown in Figure 1. This allowed us to build homology models for all members of the TMPRSS family. All of the homology models were subject to loop refinement using the Schrodinger refinement package. The refined structures were equilibrated in a water box at 300K and 1atm as described in the methods section. All the TMPRSS structures rapidly equilibrated and their root-mean-squared-deviation plateaued at an average of

1.5-1.6Å. The equilibrated atomic coordinates are provided in the supplementary information section. A structural alignment of all equilibrated TMPRSS' is shown in Figure 2. Most of the tertiary structure is conserved by all members of this family (i.e. the alpha-helices and beta-sheets). However, the loop found between residues 50-70 is very dissimilar, which is a result of the low sequence identity in this region. Furthermore, all TMPRSS proteins share a conserved histidine, aspartic acid, and serine that form the catalytic triad as shown in Supplementary Table 2. While the catalytic triad appears to be in different points across the amino acid sequences, the histidine, aspartic acid, and serine are all located at similar points in structural space within the trypsin region. From the structural alignment it is observed that the catalytic triad is also conserved in geometry, as shown in Figure 3.

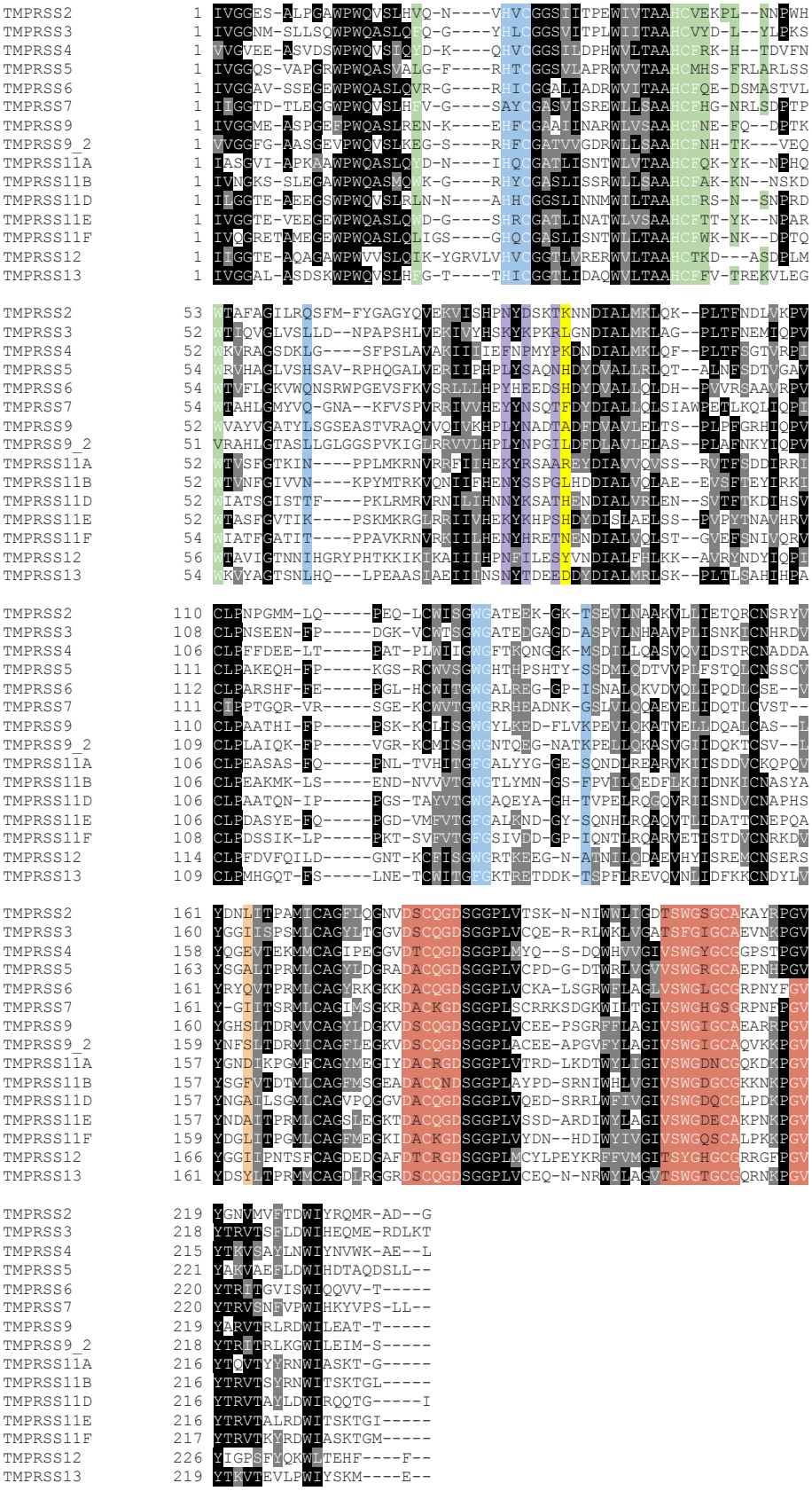

Figure 1. Sequence alignment of all TMPRSS family members. Residues conserved in >50% of all TMPRSS sequences are highlighted in black. From previous analysis of TMPRSS2, there are 5 potential pockets (A-E) that play a role in guiding the ligands to bind properly near the active site of TMPRSS2.[22] Each of these components have been highlighted in Figure 1: position 87 in yellow, pocket A in red, pocket B in blue, pocket C in green, pocket D in purple, and pocket E in orange.

Figure 2. Structural alignment of refined and equilibrated TMPRSS homology models.

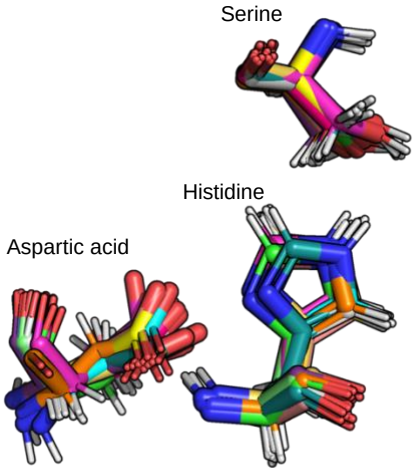

Figure 3. Structural alignment of the catalytic triad in all TMPRSS homology models.

Drug Binding Site. Literature reports have shown that camostat is hydrolyzed by carboxylesterases to form the active metabolite 4-(4-guanidinobenzoyloxy) phenylacetate (GBPA), as shown in figure 4. The active metabolite GBPA has a negatively charged carboxylate group which is stabilized through the formation of a salt bridge with the lysine residue in position 87 of TMPRSS2 and simultaneously anchored via the guanidino group as shown in Figure 5.[22] The salt bridge anchors the drug to position the scissile bond of camostat at the optimal distance for attack by the catalytic serine. The cleavage of the scissile bond by the catalytic serine results in the acylation of the enzyme, thus inhibiting TMPRSS2.

Since K87 (K342 in Escalante et. al.[22]) has been shown to play a role in stabilizing the active metabolite of camostat,[22] examining the other TMPRSS proteins for the amino acid in this position is crucial as it can elucidate similar stabilization mechanisms. From Figure 1, TMPRSS4 is the only other protein that conserves lysine. TMPRSS5, TMPRSS6, and TMPRSS11A maintain a positively charged amino acid through histidine and arginine, respectively. On the other hand, TMPRSS13 carries a negative charge through aspartic acid. TMPRSS11F contains the uncharged polar amino acid asparagine. TMPRSS7 and TMPRSS12 have aromatic hydrophobic side chains phenylalanine and tyrosine, respectively. All the other TMPRSS proteins contain chain hydrophobic amino acids.

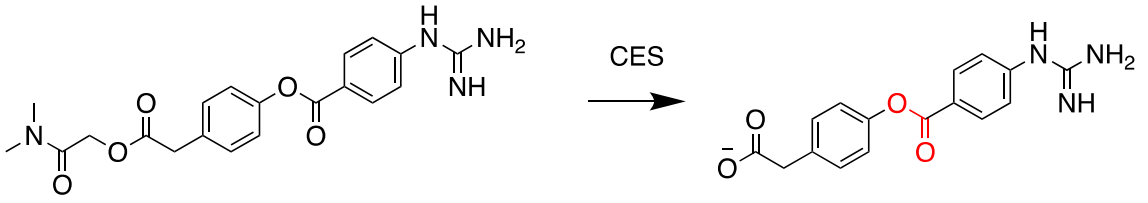

Figure 4. Camostat is hydrolized to the active metabolite 4-(4-guanidinobenzoyloxy) phenylacetate (GBPA) by carboxylesterases (CES). The scissile ester bond attacked by the catalytic serine of TMPRSS2 is highlighted in red.

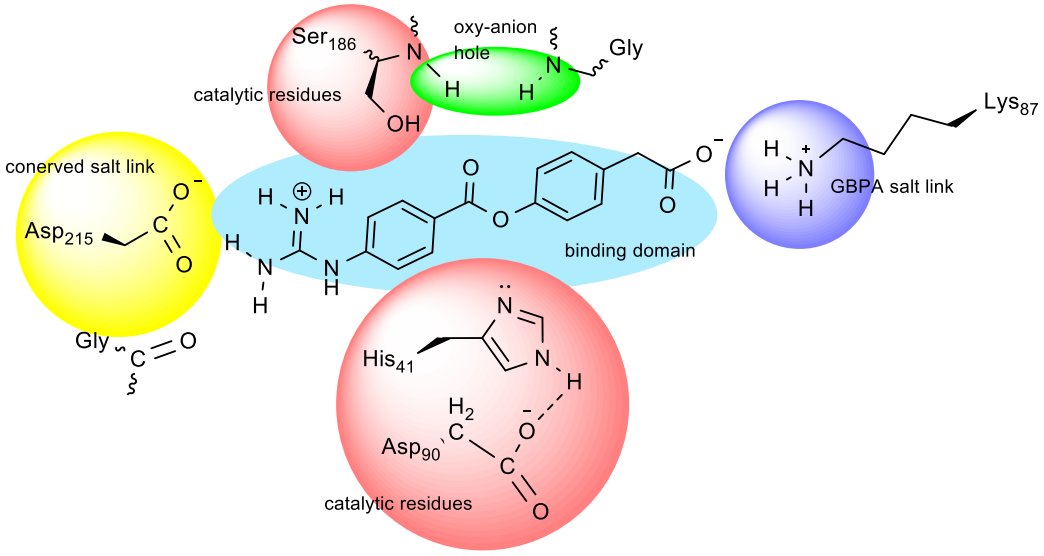

Figure 5 Contact interaction diagram between GBPA and TMPRSS2 active site. The anchoring residues near the active site are highlighted in yellow, the catalytic triad is highlighted in red and the lysine distal anchor is highlighted in blue.

Analysis of Pockets A-E. The location of the five binding pockets identified in the TMPRSS family is shown in Figure 6. Pocket A has two critical positions that vary between TMPRSS proteins. An analysis of Figure 1 shows TMPRSS2 has an uncharged, polar amino acid residue in position 181, which is maintained in TMPRSS3, TMPRSS4, TMPRSS9, TMPRSS12, and TMPRSS13. The specific two amino acids used are serine or threonine. Other TMPRSS proteins have an alanine in this position. The second important location in pocket A is position 204 that is an uncharged, polar amino acid residue (threonine) in TMPRSS2, TMPRSS3, TMPRSS12, and TMPRSS13 while all other TMPRSS proteins have a hydrophobic side chain amino acid (e.g. valine) in that position. Pocket B shows some possible discriminant residues for ligand recognition including differences at position 25. This position is valine or isoleucine in TMPRSS2, TMPRSS3, TMPRSS4, TMPRSS6, TMPRSS12, and TMPRSS13. TMPRSS7, TMPRSS9, and TMPRSS11B contain an aromatic hydrophobic side chain. Uncharged, polar amino acids threonine and glutamine are in this position for TMPRSS5, TMPRSS11A, and TMPRSS11F whereas TMPRSS11D and TMPRSS11E have a histidine and arginine in this position, respectively. Pocket C

shows significant variance across the family. At position 47, TMPRSS2 and TMPRSS3 contain a hydrophobic amino acid (i.e. leucine). For this pocket, the two isoforms of TMPRSS9 contain different amino acids at this position. TMPRSS5, one isoform of TMPRSS9, TMPRSS11A, and TMPRSS11E have aromatic hydrophobic amino acids phenylalanine and tyrosine. TMPRSS4 and TMPRSS11B have positively charged amino acids, histidine and lysine, respectively. TMPRSS6 contains the negatively charged amino acid, aspartate. TMPRSS7, the second isoform of TMPRSS9, TMPRSS11D, TMPRSS11F, and TMPRSS13 have uncharged, polar amino acids of asparagine and threonine. Lastly, TMPRSS12 has no amino acid in this matching location. Pocket D is of significant interest due to the potential interaction of GBPA and lys87 (as highlighted above). The amide of the prodrug and carboxylate of the active metabolite bind in this domain. The lysine, however, is conserved in only two proteins in this family, TMPRSS2 and TMPRSS4. While pocket E shows differences between the amino acids, interactions in this domain do not appear to play a role in ligand binding due to their locations relative to the binding pocket.

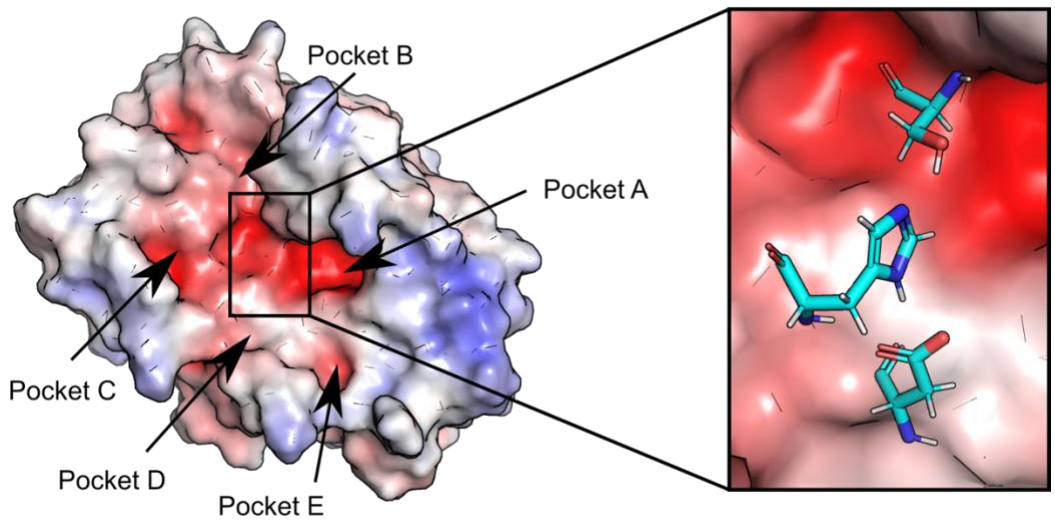

Figure 6 *(left)* Five binding pockets (A-E) in TMPRSS2 have been identified to be located near the active site. Pocket A corresponds to the anchor residue highlighted in yellow in **Figure 5**. The location of the active site catalytic triad is highlighted by the black rectangle. *(right)* A zoomed in view of the catalytic triad residues: serine, histidine, arginine (from top to bottom).

Discussion

The alignments and structural analyses have shown there is significant variation across the TMPRSS family of protein to exploit in the design of selective inhibitors. Of the potential pockets and positions that influence binding, pockets A, B, C, and D have differences in the primary sequence that may drive ligand binding and selectivity. From these specific regions, differences in position 87 in pocket D may be important for ligand binding. Prior work has shown that lys87 may be a key anchor point for the carboxylate group of GBPA (the active metabolite of camostat). Only TMPRSS2 and TMPRSS4 share the conserved lysine at position 87. This is interesting because the Spike protein is known to be cleaved by both TMPRSS2 and TMPRSS4 to gain cell entry. Targeting this position may therefore provide an approach to selectively inhibit these two enzyme and reduce off target binding of camostat. Pocket A also contains an area of dense negative charges through its primary sequence. This allows for a strong ionic interaction between ligands containing positively charged amino acids and TMPRSS proteins. Because of this interaction, current ligands with large bulky positively charged regions, like nafamostat, can form strong ionic interactions with pocket A.

Conclusion

Within the TTSP family, members of the TMPRSS subfamily have been implicated in disease progression involving cancer and infectious diseases such as COVID-19. The connection between Spike protein processing, disease progression, and TMPRSS2 activity has sparked great interest in the repurposing of drugs such as camostat and nafamostat for COVID-19 drug therapy. The alignments presented and analysis performed has shown there are significant differences in amino acid residues within the ligand binding pockets of the TMPRSS family of enzymes. Since drug selectivity and off target binding can have dramatic consequences on adverse side effects of drugs the data given here is designed to help identify unique epitopes in TMPRSS2 that can be exploited in the design of more selective agents. One of the more interesting results of this study is the identification of lys87 as a potential selectivity

element for inhibiting TMPRSS2 and TMPRSS4. This lysine is only conserved in these two family members which are both enzymes that have been shown to process the SARS-CoV Spike protein. The ability to selectively block TMPRSS2,4 and avoid off target binding to other TMPRSS proteins would provide significant advantages in the development of better therapeutic agents that block protease-mediate cleavage of the Spike protein. One problem that is not easily addressed is the conserved nature of the polar-anionc region in pocket A. Drugs like nafamostat that contain a bulky cationic group are recognized across all members that display this motif. This, may in part explain the promiscuous binding of compounds showing pan-trypsin protease activity. These results suggest it may not be possible to improve the selectivity of these agents for repurposing in the design of COVID-19 agents.

References

1. Bugge, T. H.; Antalis, T. M.; Wu, Q., Type Ii Transmembrane Serine Proteases. *Journal of Biological Chemistry* **2009**, 284, 23177-23181.

2. Antalis, T. M.; Bugge, T. H.; Wu, Q., Membrane-Anchored Serine Proteases in Health and Disease. *Progress in molecular biology and translational science* **2011**, 99, 1-50.

3. Consortium, U., Uniprot: A Hub for Protein Information. *Nucleic acids research* **2015**, 43, D204-D212.

4. Cal, S.; Quesada, V.; Garabaya, C.; Lopez-Otin, C., Polyserase-I, a Human Polyprotease with the Ability to Generate Independent Serine Protease Domains from a Single Translation Product. *Proc Natl Acad Sci U S A* **2003**, 100, 9185-90.

5. Vaarala, M. H.; Porvari, K. S.; Kellokumpu, S.; Kyllonen, A. P.; Vihko, P. T., Expression of Transmembrane Serine Protease Tmprss2 in Mouse and Human Tissues. *J Pathol* **2001**, 193, 134-40.

6. Chen, Y. W.; Lee, M. S.; Lucht, A.; Chou, F. P.; Huang, W.; Havighurst, T. C.; Kim, K.; Wang, J. K.; Antalis, T. M.; Johnson, M. D.; Lin, C. Y., Tmprss2, a Serine Protease Expressed in the Prostate on the Apical Surface of Luminal Epithelial Cells and Released into Semen in Prostasomes, Is Misregulated in Prostate Cancer Cells. *Am J Pathol* **2010**, 176, 2986-96.

7. Glowacka, I.; Bertram, S.; Müller, M. A.; Allen, P.; Soilleux, E.; Pfefferle, S.; Steffen, I.; Tsegaye, T. S.; He, Y.; Gnirss, K., Evidence That Tmprss2 Activates the Severe Acute Respiratory

Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *Journal of virology* **2011**, 85, 4122-4134.

8. Ziegler, C. G.; Allon, S. J.; Nyquist, S. K.; Mbano, I. M.; Miao, V. N.; Tzouanas, C. N.; Cao, Y.; Yousif, A. S.; Bals, J.; Hauser, B. M., Sars-Cov-2 Receptor Ace2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* **2020**, 181, 1016-1035. e19.

9. Zang, R.; Gomez Castro, M. F.; McCune, B. T.; Zeng, Q.; Rothlauf, P. W.; Sonnek, N. M.; Liu, Z.; Brulois, K. F.; Wang, X.; Greenberg, H. B.; Diamond, M. S.; Ciorba, M. A.; Whelan, S. P. J.; Ding, S., Tmprss2 and Tmprss4 Promote Sars-Cov-2 Infection of Human Small Intestinal Enterocytes. *Sci Immunol* **2020**, 5.

10. Wallrapp, C.; Hahnel, S.; Muller-Pillasch, F.; Burghardt, B.; Iwamura, T.; Ruthenburger, M.; Lerch, M. M.; Adler, G.; Gress, T. M., A Novel Transmembrane Serine Protease (Tmprss3) Overexpressed in Pancreatic Cancer. *Cancer Res* **2000**, 60, 2602-6.

11. Webb, S. L.; Sanders, A. J.; Mason, M. D.; Jiang, W. G., Type Ii Transmembrane Serine Protease (Ttsp) Deregulation in Cancer. *Front Biosci* **2011**, 16, 539-552.

12. Shen, L. W.; Mao, H. J.; Wu, Y. L.; Tanaka, Y.; Zhang, W., Tmprss2: A Potential Target for Treatment of Influenza Virus and Coronavirus Infections. *Biochimie* **2017**, 142, 1-10.

13. Liu, S.; Xiao, G.; Chen, Y.; He, Y.; Niu, J.; Escalante, C. R.; Xiong, H.; Farmar, J.; Debnath, A. K.; Tien, P.; Jiang, S., Interaction between Heptad Repeat 1 and 2 Regions in Spike Protein of Sars-Associated Coronavirus: Implications for Virus Fusogenic Mechanism and Identification of Fusion Inhibitors. *The Lancet* **2004**, 363, 938-947.

14. Fung, T. S.; Liu, D. X., Human Coronavirus: Host-Pathogen Interaction. *Annual Review of Microbiology* **2019**, 73, 529-557.

15. Hatesuer, B.; Bertram, S.; Mehnert, N.; Bahgat, M. M.; Nelson, P. S.; Pöhlman, S.; Schughart, K., Tmprss2 Is Essential for Influenza H1n1 Virus Pathogenesis in Mice. *PLoS Pathology* **2013**, 9, e1003774.

16. Iwata-Yoshikawa, N.; Okamura, T.; Shimizu, Y.; Hasegawa, H.; Takeda, M.; Nagata, N., Tmprss2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *Journal of Virology* **2019**, 93.

17. Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S., Sars-Cov-2 Cell Entry Depends on Ace2 and Tmprss2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, 10.1016/j.cell.2020.02.052.

18. Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M., Amber 18; 2018. *University of California, San Francisco*.

19. Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T., Further Along the Road Less Traveled: Amber Ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *Journal of Chemical Theory and Computation* **2016**, 12, 3926-3947.

20. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics and Modelling* **2006**, 25, 247-260.

21. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., Ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. *Journal of Chemical Theory and Computation* **2015**, 11, 3696-3713.

22. Escalante, D. E.; Ferguson, D. M., Structural Modeling and Analysis of the Sars-Cov-2 Cell Entry Inhibitor Camostat Bound to the Trypsin-Like Protease Tmprss2. *Med Chem Res* **2021**, 10.1007/s00044-021-02708-7, 1-11.

Supplementary Appendix

Supplemental Figure 1 TMPRSS2 Potential Pockets for ligand binding.

Supplementary Table 1 PDB codes selected with each TMPRSS enzyme.

Protein Name	Organism	Homologic Protein	Identity (%)	Accession Codes
TMPRSS 3	Homo Sapien	Serine Protease Hepsin	45.08	1O5E
	Bos Taurus	Enteropeptidase	44.64	1EKB
	Homo Sapien	Enteropeptidase Catalytic Light Chain	42.92	4DGJ
	Mus Musculus	Plasma Kallikrein	42.5	5GVT
	Homo Sapien	Plasma Kallikrein	42.44	6O1G
	Homo Sapien	TMPRSS 11E	42.24	2OQ5
	Astacus leptodactylus	Hepatopancreas Trypsin	42.19	2F91
	Homo Sapien	Plasminogen	42.17	1QRZ
	Homo Sapien	Tryptase alpha/beta-1	41.13	6O1F
TMPRSS 4	Mus Musculus	Plasma Kallikrein	42.8	6A8O
	Sus scrofa	BETA-ACROSIN HEAVY CHAIN	42.04	1FIZ
	Homo Sapien	Plasma Kallikrein	41.03	2ANY
	Homo Sapien	Prostasin	40.93	3DFJ
	Bos Taurus	Thrombin Heavy Chain	40.42	1DLK
	Homo Sapien	Serine Protease Hepsin	40.24	1O5E
TMPRSS 5	Salmo Salar	Trypsin I	44.02	1UTK
	Homo Sapien	Serine Protease Hepsin	42.51	1O5E
	Homo Sapien	TMPRSS 11E	41.63	2OQ5
	Homo Sapien	ST14 Protein	41.15	3P8F
	Homo Sapien	Alpha I Tryptase	41.15	2F9N
	Homo Sapien	Plasma Kallikrein	40.5	6O1G

TMPRSS 6	Homo Sapien	TMPRSS 11E	46.12	2OQ5
	Homo Sapien	ST14 Protein	44.73	3P8F
	Homo Sapien	Suppressor of Tumorigenicity 14	44.73	1EAW
	Homo Sapien	Membrane-type Serine Protease 1	44.3	3BN9
	Homo Sapien	Plasma Kallikrein	43.4	6O1G
	Homo Sapien	Serine Protease Hepsin	41.7	1O5E
	Homo Sapien	Coagulation Factor Xi	41.1	1ZHP
TMPRSS 7	Homo Sapien	Suppressor of Tumorigenicity 14	46.89	1EAW
	Homo Sapien	ST14 Protein	46.89	3P8F
	Homo Sapien	Membrane-type Serine Protease 1	46.47	3BN9
	Homo Sapien	TMPRSS 11E	45.06	2OQ5
	Homo Sapien	Prostasin	42.46	3GYL
	Homo Sapien	Coagulation Factor Xi	42.44	1XX9
TMPRSS 9 Peptidase S1	Homo Sapien	TMPRSS 11E	45.53	2OQ5
	Homo Sapien	Suppressor of Tumorigenicity 14	44.17	1EAW
	Homo Sapien	ST14 Protein	44.17	3P8F
	Homo Sapien	Membrane-type Serine Protease 1	43.75	3BN9
	Homo Sapien	Enteropeptidase Catalytic Light Chain	41.81	4DGJ
	Homo Sapien	Serine Protease Hepsin	40.98	1O5E
	Homo Sapien	Coagulation Factor Xa-trypsin Chimera	40.52	1FXY
TMPRSS 9 Peptidase S2	Homo Sapien	Suppressor of Tumorigenicity 14	44.54	1EAW
	Homo Sapien	ST14 Protein	44.54	3P8F
	Homo Sapien	Membrane-type Serine Protease 1	44.12	4IS5
	Homo Sapien	TMPRSS 11E	44.59	2OQ5
	Sus scrofa	Trypsin	40.34	3MYW
TMPRSS 11A	Homo Sapien	TMPRSS 11E	55.84	2OQ5
	Homo Sapien	Serine Protease Hepsin	43.61	1O5E
	Homo Sapien	Prostasin	39.09	3DFJ
	Ovis aries	BETA-ACROSIN HEAVY CHAIN	39.6	1FIW
	Homo Sapien	Coagulation Factor Xi	37.39	1XX9
TMPRSS 11B	Homo Sapien	TMPRSS 11E	49.14	2OQ5
	Homo Sapien	Serine Protease Hepsin	42.02	1O5E
	Homo Sapien	Prostasin	41.91	3DFJ
	Homo Sapien	Plasma Kallikrein	40.34	6O1G
	Ovis aries	BETA-ACROSIN HEAVY CHAIN	40.16	1FIW
TMPRSS 11D	Homo Sapien	TMPRSS 11E	50.43	2OQ5
	Homo Sapien	Serine Protease Hepsin	46.25	1O5E
	Ovis aries	BETA-ACROSIN HEAVY CHAIN	41.83	1FIW
	Homo Sapien	Tryptase alpha/beta-1	40.57	5WI6
	Homo Sapien	Coagulation Factor Xi	40.08	1XX9
TMPRSS 11E	Homo Sapien	Suppressor of Tumorigenicity 14	46.94	1EAW
	Homo Sapien	ST14 Protein	46.94	3P8F
	Homo Sapien	Membrane-type Serine Protease 1	46.53	3BN9
	Homo Sapien	Serine Protease Hepsin	41.18	1O5E
	Homo Sapien	Plasma Kallikrein	40.34	6O1G
TMPRSS 11F	Homo Sapien	TMPRSS 11E	58.12	2OQ5
	Homo Sapien	Serine Protease Hepsin	44.58	1O5E
	Homo Sapien	Plasma Kallikrein	42.36	6O1G
	Homo Sapien	Suppressor of Tumorigenicity 14	41.98	1EAW
	Homo Sapien	ST14 Protein	41.98	3P8F
	Homo Sapien	Membrane-type Serine Protease 1	41.56	3BN9
	Homo Sapien	Coagulation Factor Xi	40.08	1ZHM
TMPRSS 12	Mus Musculus	Plasma Kallikrein	40.08	6A8O
	Homo Sapien	Enteropeptidase Catalytic Light Chain	40.76	4DGJ
	Homo Sapien	Plasma Kallikrein	40.42	6O1G
	Bos Taurus	Enteropeptidase	38.66	1EKB
	Ovis aries	BETA-ACROSIN HEAVY CHAIN	39.11	1FIW

TMPRSS 13	Homo Sapien	Plasma Kallikrein	44.58	6O1G
	Homo Sapien	Plasma Kallikrein Light Chain	44.17	5F8T
	Homo Sapien	Plasminogen	43.29	1QRZ
	Homo Sapien	TMPRSS 11E	42.92	2OQ5
	Homo Sapien	Serine Protease Hepsin	42.51	1O5E
	Homo Sapien	PLASMIN	42.37	1BML
	Homo Sapien	Coagulation Factor Xi	41.91	1XX9
	Homo Sapien	Suppressor of Tumorigenicity 14	40.76	1EAW
	Homo Sapien	ST14 Protein	40.76	3P8F
	Homo Sapien	Prostasin	40.33	3GYL
	Homo Sapien	Tryptase Alpha-1	40.25	2F9O

Supplemental Table 2 Comparison of TMPRSS Proteins showing the catalytic triad residues, peptidase and trypsin regions. The percent identity and percent positives was calculated relative to TMPRSS2.

Protein Name	Organism	Catalytic Triads	Peptidase Domain	Trypsin Region	Identities (%)	Positives (%)
TMPRSS2	Homo Sapien	H296, D345, S441	256-489	256-485	100	100
TMPRSS3	Homo Sapien	H257, D304, S401	217-449	217-445	53	65
TMPRSS4	Homo Sapien	H245, D290, S387	205-434	205-430	43	59
TMPRSS5	Homo Sapien	H258, D308, S405	218-453	218-449	45	58
TMPRSS6	Homo Sapien	H617, D668, S762	577-811	577-807	40	57
TMPRSS7	Homo Sapien	H646, D694, S790	606-840	606-836	40	56
TMPRSS9	Homo Sapien	H243, D292, S387	203-436 (S1)	203-432	43	57
	Homo Sapien	H544, D592, S687	504-736 (S2)	504-732	38	54
TMPRSS11A	Homo Sapien	H230, D275, S371	190-420	190-416	39	54
TMPRSS11B	Homo Sapien	H225, D270, S366	185-415	185-411	39	57
TMPRSS11D	Homo Sapien	H227, D272, S368	187-417	187-413	43	57
TMPRSS11E	Homo Sapien	H232, D277, S373	192-422	192-418	42	57
TMPRSS11F	Homo Sapien	H248, D293, S389	206-437	206-433	42	60
TMPRSS12	Homo Sapien	H122, D171, S268	78-318	78-314	38	57
TMPRSS13	Homo Sapien	H366, D414, S511	326-559	326-555	45	59

The atomic coordinates for all the generated homology models can be found as PDB files in the supplementary .zip folder.