2021

SARS-CoV-2 Host Surface Protease Purification and Inhibitor Identification

Nemo (Zi Hao) Liu

SARS-CoV-2 Host Surface Protease Purification and Inhibitor Identification

April 5th, 2021

Abstract

With the increasing appearance of highly infectious SARS-CoV-2 variants, mass vaccination becomes a priority for governments around the world. Even if vulnerable populations are vaccinated, antivirals are still required to treat those severely affected by the virus. In the past year, clinical trials have been performed on Camostat mesylate, an inhibitor for the host surface serine protease TMPRSS2, as an antiviral to reduce the severity of SARS-CoV-2 infections. As a general inhibitor for serine proteases, long-term use of Camostat mesylate can lead to the suppression of the patients' adaptive immune system through off-target inhibition of pro-inflammatory cytokines. In this study, we outline the optimization of expressing and purifying active hTMPRSS2 in *E. coli* (previously thought to be unfeasible), as well as formulating an *in silico* and *in vitro* hybrid HTS pipeline to identify small molecule inhibitors that have higher specificity to the active site of TMPRSS2. After experimenting with different fusion expression constructs and Ni-IMAC purification protocols, the autocatalytic conditions to yield active TMPRSS2 were found to require low salt (150 mM NaCl 50 mM Tris pH 8.0) and given at least 72 hours at 4 ^{o}C. Following the successful generation of a quality homology model, successful *in silico* screening results yielded compounds with a maximum of 15-fold higher binding affinity to the active site of TMPRSS2 compared to Camostat. With an *in vitro* hit rate of 10% after enriching a commercially available compound library (N = 264, 158 compounds), a patented compound (CIVICYXXDIQUPZ-UHFFFAOYSA-N) was found to inhibit TMPRSS2 to the same degree as recognized antiviral Camostat mesylate with higher binding specificity.

Introduction

Historically, little attention was given to the annual common cold, usually caused by the coronaviridae family of viruses. In the past year, however, a new variant of SARS-CoV, namely SARS-CoV-2 (SARS-2), was found the be the culprit of the global pandemic we are experiencing to date. Like the common cold, its primary site of infection is the upper respiratory tract and mainly affects the respiratory mucosa of the nose, throat, sinuses, and larynx (1). One of the more dangerous symptoms of SARS-2 includes shortness of breath, and/or heavy breathing that is often treated by using a ventilator to pump oxygen into the lungs of those affected (2). Furthermore, in the long-term, SARS-2 has been identified to cause long-standing damage to the alveoli in the lungs and ultimately can lead to breathing problems (2).

As we learn more about the targeted cells, previous research has identified that SARS-2 preferentially infects the epithelial cells on the lung and bronchi through the same mechanism as SARS-CoV and one of the mechanisms of MERS-CoV (1). To enter the patient's lung cells, the Spike protein on the surface of SARS-2 must be primed by a host cell transmembrane serine 2 protease (TMPRSS2) through cleavage at the S1/S2 sites for the virus to be recognized and bind tightly to the angiotensin-converting enzyme 2 (ACE2) receptor, also found on the host's epithelial cell surface (3, 4).

The cross-disciplinary initiative to combat the COVID-19 pandemic has led to the identification of TMPRSS2 as a promising molecular target for SARS-CoV-2 antiviral therapy (4). Unlike MERS-CoV, proteolytic processing of the SARS-2 Spike protein by TMPRSS2 is required for cell entry, thus, there exists continued efforts to screen and test possible drug compounds that can inhibit TMPRSS2 to ultimately block SARS-2 proliferation (4). Although there are other host-proteases that SARS-2 require for infection, TMPRSS2 is an ideal molecular target since it has no vital function in the host cell (5), and thus, inactivation of TMPRSS2 would have little detrimental effect on the host organism, as shown previously in knock-out mouse models (6). Furthermore, pre-existing drugs to treat other TMPRSS2

mediated diseases such as prostate cancer, have been used in clinical trials and the results indicate that a Camostat mesylate dose of 600 mg/day will result in antiviral activity (7).

Although the pre-existing clinically approved drug Camostat mesylate appears to be the solution, it is only a temporary one. Long-term studies have shown that Camostat mesylate should only be taken under extreme conditions as the drug shows off-target effects that can weaken the immune system (7, 8). Immunological studies indicate that Camostat mesylate inhibits pro-inflammatory cytokines such as IL-6 and TNF-alpha, which may result in the suppression of immune responses giving leeway to opportunistic pathogens and pre-existing infections (8). SARS-2 clinical trial studies with Camostat mesylate also indicate that for a long-term solution to COVID-19, inhibitors of TMPRSS2 should have increased potency and specificity (9). TMPRSS2 remains a target of choice as Iwata-Yoshikawa *et al*. suggest that TMPRSS2 is a pan target for all coronaviruses due to its promiscuous catalytic activity (10). For example, entry promotion by TMPRSS2 has even been observed in other respiratory viruses such as influenza A (11). Furthermore, TMPRSS2 overexpression is linked to certain prostate cancers (1, 12). Ultimately, TMPRSS2 is a clinically important molecular target and there is a significant need for a TMPRSS2-specific inhibitor.

We hypothesize that the toxicity and off-target effects of current TMPRSS2 inhibitors can be reduced by identifying a compound with improved target specificity and affinity. Using an *in vitro* fluorogenic assay for TMPRSS2 activity (13), we aim to identify a clinically useful inhibitor by screening a custom, enriched compound library. The compounds in the library will be selected using *in silico* docking methods to enrich a large bio-active library for successful inhibitors. Since there are currently no available structures of TMPRSS2 in the PDB, a co-crystal structure will not only facilitate downstream drug optimizations but will enable the development of novel antiviral therapies.

To screen TMPRSS2 inhibitors *in vitro*, pure and active TMPRSS2 is required. *In vivo,* TMPRSS2 activation is achieved by autocatalytic cleavage of the inactive, full-length TMPRSS2 (residues 1-492) between residues R255 and I256 (14). Autoproteolysis releases the active protease (residues 256-492)

from the rest of the protein (14). *In vitro*, this requirement for TMPRSS2 autoproteolysis can be exploited during the protein expression and purification process to select for properly folded and active $TMPRSS2_{265\text{-}492}$. Previous work has demonstrated that residues 108-492 of TMPRSS2 are sufficient for autoproteolysis (13). Removal of the first 107 residues from full-length TMPRS2 simplifies the protein production process as there appears to be a hydrophobic transmembrane domain (TM) encoded at residues 84-106 that is responsible for poor solubility. Previous work by Shrimp *et al.* have shown that active protease can be recovered from recombinant TMPRSS2 108-492 expressed in yeast. Although bacterial expression of this TMPRSS2 construct has never yielded active protease in the literature (13), we believe that the advantages of using a bacterial expression system merit further attempts to isolate active TMPRSS2 from *E. coli*.

Results

Bacterial expression of TMPRSS2$_{108-492}$ yields insoluble protein

Three different fusion constructs of TMPRSS2 108-492 were designed for *E. coli* expression (Figure 1). In the simplest case, a TEV protease-cleavable His_6 tag was fused to the N-terminus of TMPRSS2 108-492 to generate the construct His-TMPRSS2 (Figure 1A). His-TMPRSS2 was successfully expressed in *E. coli* as indicated by a species appearing in the proteome at ~40 kDa, corresponding to the full, uncleaved His-TMPRSS2 construct (expected molecular weight: 42.6 kDa). Unfortunately, no detectable TMPRSS2 was isolated following enrichment of His-tagged proteins by Ni-IMAC. Using SDS-PAGE analysis we concluded that the recombinant protein was entirely lost to precipitation, as the His-TMPRSS2 was only found in the insoluble lysate pellet following lysate clarification (Supplemental Figure 1).

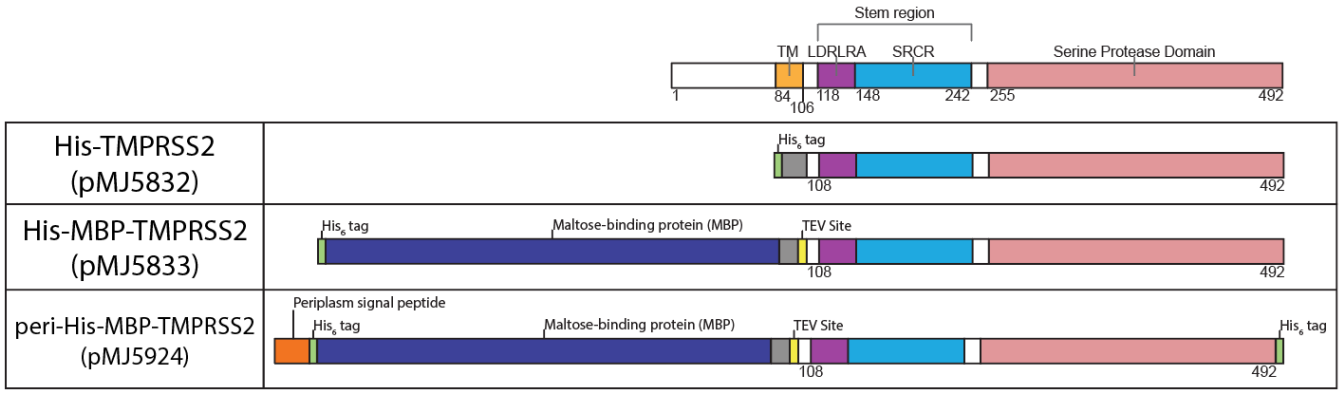


Figure 1. Expression constructs of TMPRSS2. A. pMJ5832 expression construct of TMPRSS2 with only an N-terminal His_6 tag. **B. pMJ5833** expression construct was designed with a single N-terminal His_6 + MBP tag. **C. pMJ5924** expression construct was designed later with an additional C-terminal His_6 tag as well as a periplasm signal peptide.

An N-terminal MBP fusion improves solubility

To address the issue of poor solubility, we added an additional solubility-promoter into the TMPRSS2 108-492 fusion construct: Maltose-binding protein (MBP). MBP was fused between the flexible linker and the N-terminal His_6-tag to generate His-MBP-TMPRSS2 (Figure 1B). Following the same expression and Ni-IMAC purification strategy as for His-TMPRSS2, a large amount of soluble protein was isolated. Specifically, two species were isolated in Ni-IMAC elution: a large species with an apparent MW between 76 and 100 kDa and a smaller species with an apparent MW between 37 and 50 kDa (Supplemental Figure 2A). The two recovered Ni-IMAC species were separated using Gel filtration chromatography (Supplemental Figure 2B), and the proteolytic activities of both proteins were tested using a previously reported fluorogenic assay for TMPRSS2 activity (13). Neither the large ~80 kDa nor the small ~40 kDa species showed any detectable activity, while 15 µM of purified Chymotrypsin showed a 4.9-fold increase of signal from the background controls (Figure 2). We can reasonably conclude that the larger protein corresponds to full-length, uncleaved His-MBP-TMPRSS2 (expected MW: 86.4 kDa), while the smaller protein corresponds to the MBP fusion that was separated from the rest of the construct either by TMPRSS2 autocatalysis (expected MW: 60.2 kDa) or by proteolytic cleavage of the TEV protease site (expected MW: 43.8 kDa). We would not expect either of the two recovered species to harbor any TMPRSS2 activity since the protease domain is either inactivated, as in the case of the full-length construct, or missing as in the case of the recovered MBP fusion. Therefore, it appears that the protease domain is being removed at some point during protein extraction and purification (the addition of a C-terminal His_6-tag fusion may remedy this issue). We also hypothesize that some of the TMPRSS2 that was expressed may be misfolding during protein expression since a large portion of the soluble protein recovered was full-length, uncleaved, and non-functional.

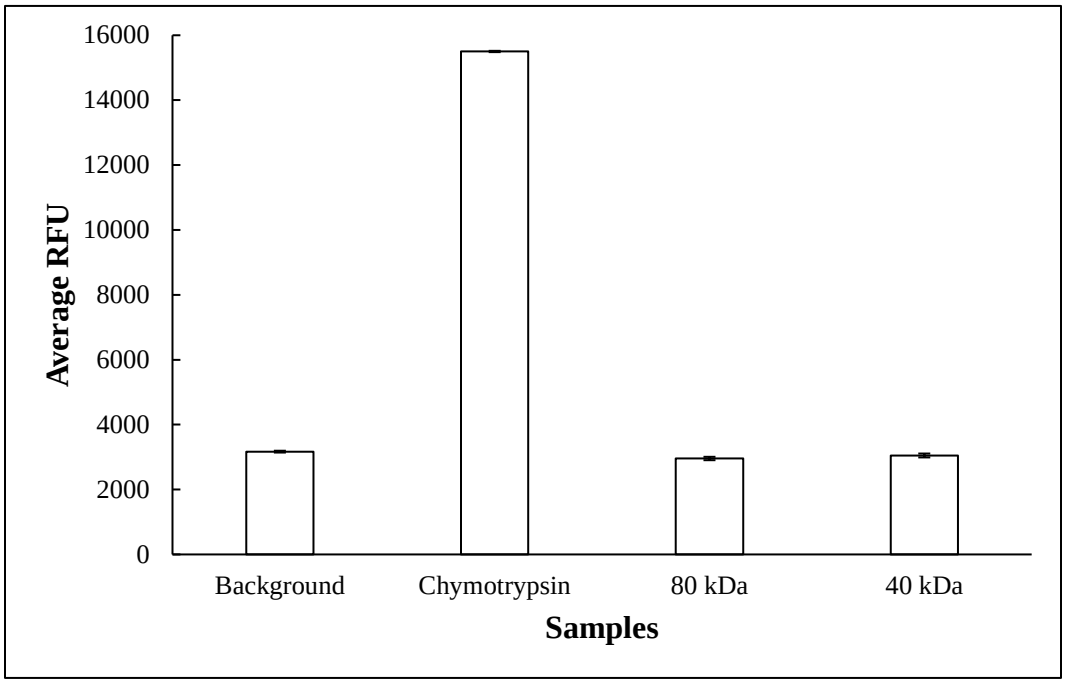


Figure 2. Fluorescence activity assay for the proteins purified from pMJ5833. Relative fluorescence units (RFU) were measured using the Synergy H1 plate reader. Samples in triplicate were run for 60 minutes with measurements of emission = 340 nm, excitation = 440 nm every minute at 37 ºC. Maximal RFU readings across samples were used to generate this bar graph. All samples contained 20 µM of fluorescence peptide. Background sample contained activity buffer only (150 mM NaCl, 50 mM Tris pH 8.0). The concentration of Chymotrypsin was 15 µM for positive control. 80 kDa and 40 kDa species were at 2 µM. No statistical increase relative to background activity was observed for the 80 kDa and 40 kDa species. The assay layout can be found in Supplemental Table 1.

A C-terminal His_6 fusion allows for recovery of active protease

To address the loss of the TMPRSS2 protease domain during extraction and purification of His-MBP-TMPRSS2, an additional His_6 tag was added to the C-terminus of the construct (Figure 1C). In addition, a signal peptide was added to the N-terminus of the construct to promote proper protein folding by directing the translated protein to the periplasmic space of the bacterium (15). The resulting construct, peri-His-MBP-TMPRSS2-His, was expressed from *E. coli* and purified by Ni-IMAC (Figure 3A).

During purification, a species eluted off the Ni-IMAC resin with an apparent MW of ~25 kDa, during 36.0 mM Imidazole wash. This corresponds to the expected MW of the proteolytic domain of TMPRSS2 following autocatalysis (27.0 kDa). However, if these species were the cleaved TMPRSS2 protease domain, we would expect the presence of a His_6 tag, which would not allow this protein to elute from the Ni-IMAC resin at such a low concentration of Imidazole. The absence of a His_6 tag was successfully demonstrated by Western blot using an Anti-His_6 primary antibody (Figure 3B). Thus, it can be concluded that the 25 kDa band corresponds to a bacterial contaminate, rather than a fragment of the peri-His-MBP-TMPRSS2-His construct.

At 300 mM Imidazole, a single species eluted from the Ni-IMAC resin at an apparent MW of ~75 kDa. This protein corresponds to the full-length uncleaved peri-His-MBP-TMPRSS2-His construct. The identity of this protein was confirmed by Western Blot to be the entire, intact peri-His-MBP-TMPRSS2-His construct (Figure 3B). Western blot analysis also revealed the presence of two bands at ~25kDa and ~50kDa that were not visible under Coomassie stain. These bands are consistent with TMPRSS2 autocatalysis, as we would expect two polypeptide products of size 62.2 and 27.0 kDa. Given the sensitivity required to visualize these two bands, the relative ratio of cleaved to uncleaved TMPRSS2 construct is very low. We hypothesize that this is due to the high saline conditions present in the protein sample that are not conducive to protease activity. To test this hypothesis, we exchanged the buffered conditions of the recovered TMPRSS2 construct from 1 M NaCl, 300 mM Imidazole, 20 mM Tris, pH

8.0, 15% Glycerol into 150 mM NaCl, 50 mM Tris, pH 8.0 by buffer exchange chromatography, and allowed the protein to incubate in these buffer conditions at 4 °C for 72 hours. Following this incubation period, the protein banding pattern was analyzed by SDS-PAGE (Figure 3C). It became clear that the full TMPRSS2 construct has undergone autoproteolytic cleavage as two new species were observed that are consistent with our expected products of TMPRSS2 autocatalysis. Although the cleavage efficiency was much less than 100%, the serine protease activity of the sample improved dramatically. The Ni-IMAC elution did not harbor any detectable serine protease activity, but then surpassed the activity of 1 µM Chymotrypsin following the 72-hour incubation (Figure 3D). This sample of active TMPRSS2 was then used for all downstream inhibitor screening.

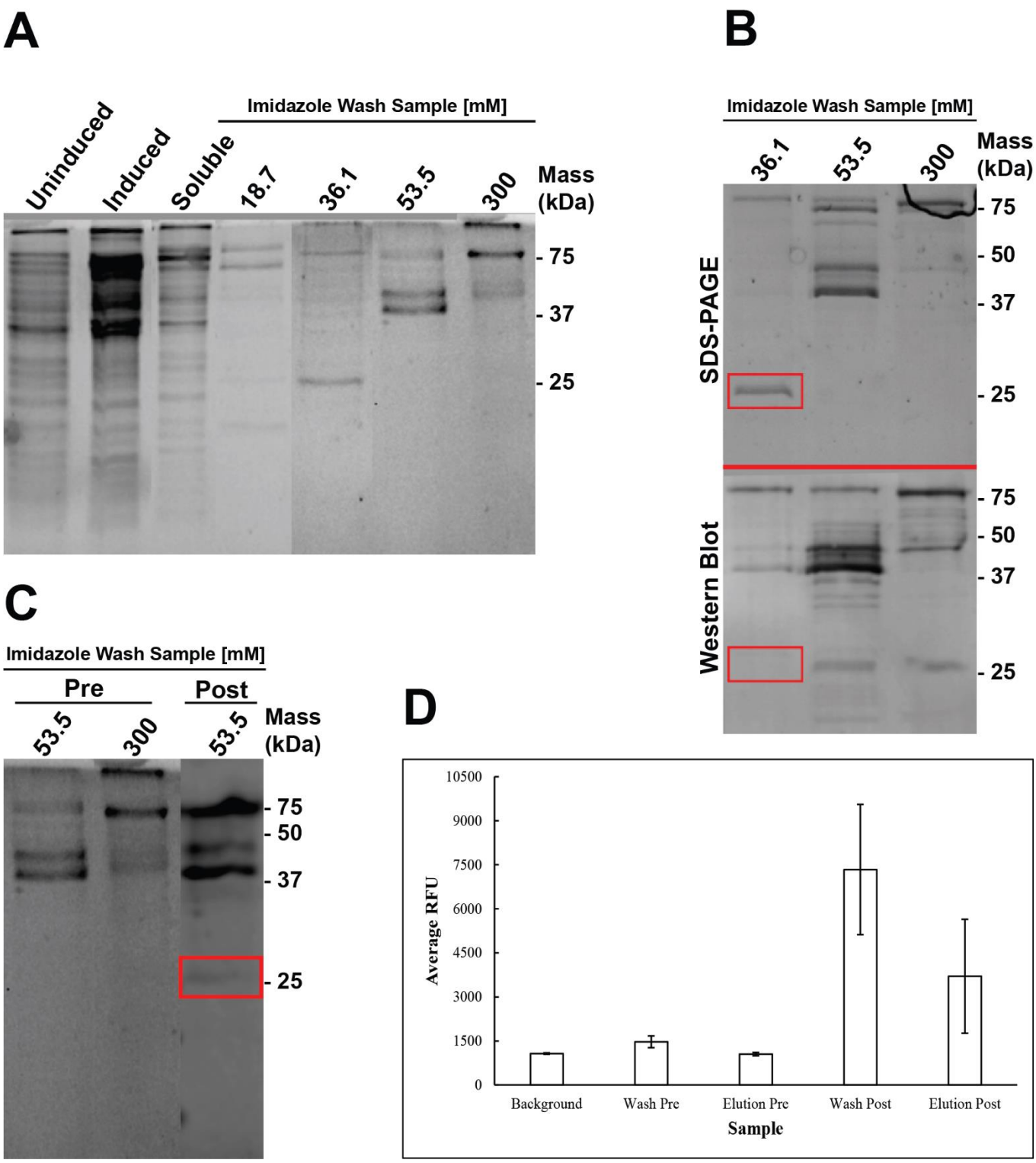

A
Imidazole Wash Sample [mM]
Uninduced
Induced
Soluble
18.7
36.1
53.5
300
Mass (kDa)
- 75
- 37
- 25
B
Imidazole Wash Sample [mM]
36.1
53.5
300
Mass (kDa)
SDS-PAGE
Western Blot
- 75
- 50
- 37
- 25
C
Imidazole Wash Sample [mM]
Pre
Post
53.5
300
53.5
Mass (kDa)
- 75
- 50
- 37
- 25
D
Average RFU
10500
9000
7500
6000
4500
3000
1500
0
Background
Wash Pre
Elution Pre
Wash Post
Elution Post
Sample

Figure 3. Ni-IMAC Purification of pMJ5924 and verification of active TMPRSS2. A. Full-length protein of interest (~75 kDa) eluted in the 300 mM Imidazole wash. SDS-PAGE was run for 60 mins at 165 V. **B.** ~25 kDa species in the 18.7 mM Imidazole does not have a His 6x tag confirmed with a Western Blot with mAbs against His 6x. Slight amounts of autocleaved active TMPRSS2 (27.0 kDa) in the 53.5 mM and 300 mM Imidazole wash samples can be detected in the Western Blot due to their C-terminal His 6x tag. **C.** SDS-PAGE of 53.5 mM fraction before and after transfer into low salt (150 mM NaCl, 50 mM Tris pH 8.0) activity buffer indicating presence of autocleaved active TMPRSS2 over time. No post 300 mM Imidazole data was available due to lack of protein sample after running the activity assay. SDS-PAGE was run for 60 mins at 165 V. **D.** Activity assay with 10 µM of fluorescence peptide verifying active protease after transfer to low salt buffer and given 72 hrs for autocatalytic events to occur. Wash samples indicate 53.5 mM species and Elution samples indicate 300 mM species. Pre samples indicate TMPRSS2 before buffer exchange and given the time for autocleavage. Post samples indicate TMPRSS2 after buffer exchange and given the time for autocleavage into active moiety. Samples were performed in triplicate and fluorescence was measured using the Synergy H1 plate reader (excitation = 340 nm, emission = 440 nm) at 37 ^{o}C every minute for 60 minutes. An average of the maximum RFU measurements from each sample was used to generate a bar graph. The assay layout can be found in Supplemental Table 2.

Generation and Validation of a TMPRSS2 Homology Model

In order to screen potential inhibitors of TMPRSS2 *in silico*, an accurate 3D model of the protein is needed. Currently, there are no experimental structures of TMPRSS2 deposited in the PDB, however, TMPRSS2 is a member of the serine protease family of proteins, of which there are many structures available. Using the Rosetta-based web-server application, ROBETTA, the amino acid sequence of hTMPRSS2 was uploaded and a homology model was generated by Rosetta comparative modeling (16, 17). The homology model was built from a deposited crystal structure of human Hepsin (PDB accession 1Z8G), a Serine protease that covers 69% of the TMPRSS2 sequence with a 33.52% residue identity (18). Although other crystal structures exist with a higher residue identity, 1Z8G was chosen as the template model as it contains the highest sequence coverage to TMPRSS2 of all available structures (Figure 4A). A sequence identity of 33.5% is more than sufficient to build an accurate homology model. The resulting TMPRSS2 homology model covers residues 146-492, which includes the full serine protease domain and part of the stem region (Figure 4B).

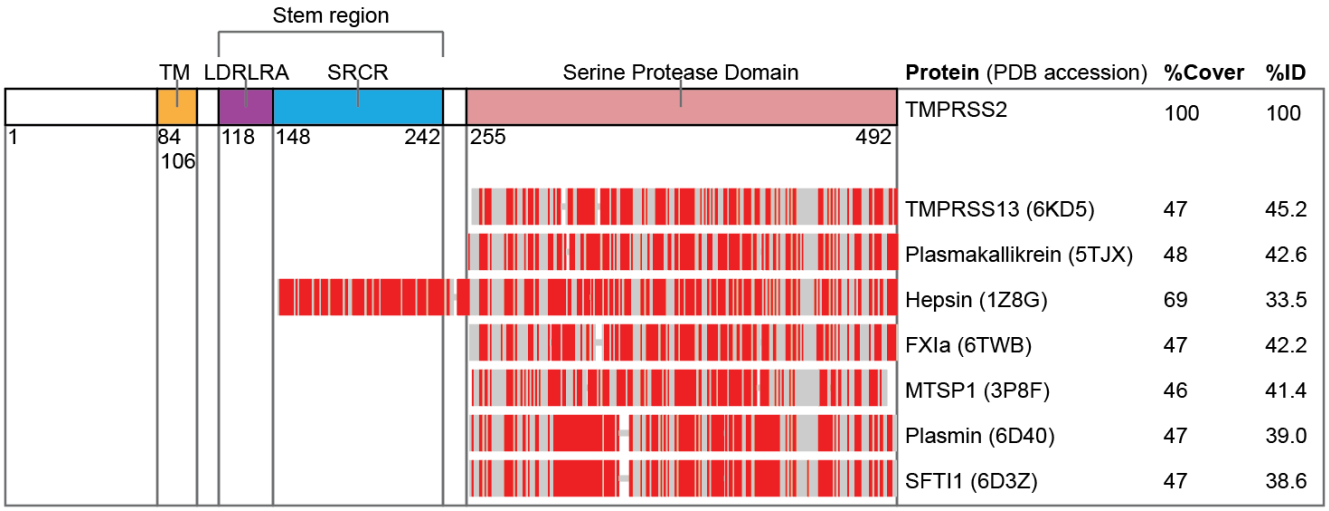


Figure 4A. Available homology template structures in the PDB. Schematic representing a sequence alignment of select serine protease crystal structures against full-length TMPRSS2. Sequence differences between the crystal structure and TMPRSS2 are highlighted in red.

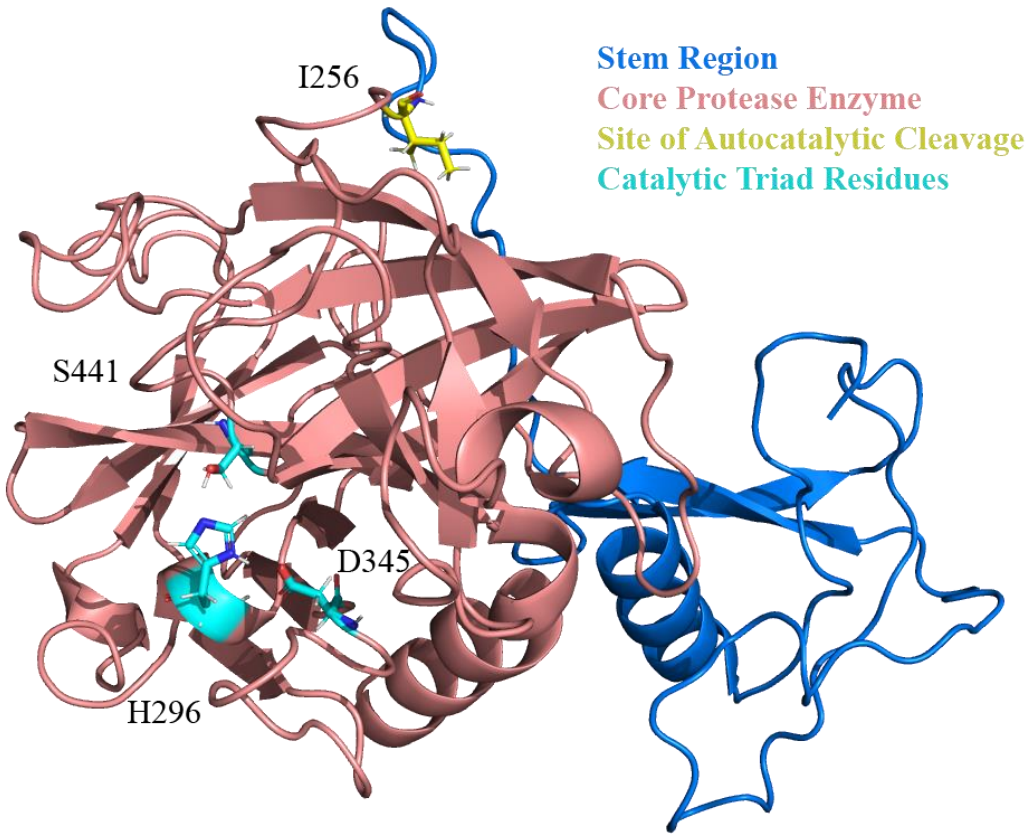


Figure 4B. Rosetta Homology Model of TMPRSS2. Modeling was built from the Hepsin crystal structure 1Z8G. The dark blue region represents the stem and inactive portion of TMPRSS2, while the salmon region represents the active protease after auto-catalytic cleavage of the peptide bond between R255 and I256 (highlighted in yellow). Active site residues have been highlighted in cyan.

The fidelity of the Rosetta homology model was then assessed by various model quality indicators using the SAVESv6.0 webserver. The ERRAT function, which analyzes characteristic atomic interactions to identify model building errors, output a score of 95.0673, indicating that the model lies within a 95% confidence interval of predicted fidelity (19). Verify3D indicated that 99.58% of residues passed the 3D-1D profile calculation (20). The PROVE algorithm concluded that the volumes of only 3.8% of the atoms in the model deviate significantly from expected values (21). The WHATCHECK and PROCHECK programs both asses model quality using an amalgamation of various metrics. WHATCHECK yielded positive Z-scores and PROCHECK indicated that 98.8% of residues were within acceptable working limits, indicating a high-quality 3D model (22, 23).

A commercial compound library was enriched for potential TMPRSS2 inhibitors in-silico

The McMaster CMCB HTS compound library is a diverse compound library of 264,158 commercially available synthetic small molecules. The diversity of the CMCB HTS makes it an excellent source of new compound leads. Unfortunately, screening this entire library for TMPRSS2 inhibitors *in vitro* would not be feasible given the resources allocated to this project. It is feasible however to screen the entire library *in silico* using high-throughput computational docking to predict which of the 264,158 compounds may be high-affinity binders. Computational docking of all compounds in the library was performed against our TMPRSS2 homology model using MCULE, a high-throughput docking server based on the AutoDock (Vina) algorithm (24, 25). The compound docking site was anchored to a single atom within the active site on TMPRSS2: the τ-Nitrogen of His-296. The *in silico* screen identified known inhibitors Camostat and Nafamostat as potential binders with MCLUE docking scores of -8.3 and -7.0 respectively. The top 50 scoring compounds that were identified from the CMCB HTS library had a score of at least -8.7 or lower (Table 1), which corresponds to a predicted binding affinity that is at least 2-fold greater than Nafamostat and 7-fold greater than Camostat. These top 50 compounds effectively make up a small custom compound library that is heavily enriched for potential TMPRSS2 inhibitors. This custom library was then procured from the McMaster CMCB for *in vitro* screening.

InChlKey	MACID	Docking Score	Number of Bioassays
ZQMARHJJCQQSGC-UHFFFAOYSA-N	MAC-0664895	-9.9	13
WFUCZQLZSCDWHG-UHFFFAOYSA-N	MAC-0491518	-9.7	1
CIVICYXXDIQUPZ-UHFFFAOYSA-N	MAC-0666566	-9.7	17
MWTMRUSFOBKPJE-UHFFFAOYSA-N	MAC-0686256	-9.6	6
GJXAYAYGGVUMHZ-UHFFFAOYSA-N	MAC-0717961	-9.5	917
IXJJCXATJSXNJM-UHFFFAOYSA-N	MAC-0482043	-9.5	4
QJVLKEOJQZVOPT-UHFFFAOYSA-N	MAC-0681929	-9.3	677
GGPVZQRCWVNSKB-UHFFFAOYSA-N	MAC-0494039	-9.3	0
KCEIUECOYSVCTL-UHFFFAOYSA-N	MAC-0477808	-9.3	2
INSKMBIFNOPTGQ-UHFFFAOYSA-N	MAC-0708880	-9.2	4
QJIYHKNUDOFFGS-UHFFFAOYSA-N	MAC-0657349	-9.2	738
NPAPYOPZEVKCJN-UHFFFAOYSA-N	MAC-0699313	-9.2	1
CGTVHWBQCJZXHU-UHFFFAOYSA-N	MAC-0666551	-9.2	12
QGVPODQDMHDEBY-UHFFFAOYSA-N	MAC-0696562	-9.1	1
JMMMWRQOEVDNTQ-UHFFFAOYSA-N	MAC-0561118	-9.1	3
ATWXEPRXKWCISQ-UHFFFAOYSA-N	MAC-0499031	-9.1	0
QSWYNUPJWYWAER-UHFFFAOYSA-N	MAC-0508230	-9.1	6
PSVYYCUSDUTFLF-UHFFFAOYSA-N	MAC-0474599	-9.1	2
DBUMOEMDCLFJDD-UHFFFAOYSA-N	MAC-0692057	-9	1
NQTHCFBKVMEVAF-UHFFFAOYSA-N	MAC-0719469	-9	705
GSNSZQXXSDUZND-UHFFFAOYSA-N	MAC-0716305	-9	759
NSTUWDWGGWTMOI-UHFFFAOYSA-N	MAC-0467823	-9	7
XHPCBNTZWHRVDF-UHFFFAOYSA-N	MAC-0706811	-9	771
XZEBJOUVYWWRDA-UHFFFAOYSA-N	MAC-0714063	-9	4
BSNNGWKKNANMFS-UHFFFAOYSA-N	MAC-0704220	-9	4
BJIPVHLRWSDKOS-UHFFFAOYSA-N	MAC-0644991	-8.9	801
PYMLELABQHUYOZ-UHFFFAOYSA-N	MAC-0505637	-8.9	758
MXXQURDFCZJTSX-UHFFFAOYSA-N	MAC-0481271	-8.9	60
AXGDASCARFOXEO-UHFFFAOYSA-N	MAC-0659864	-8.9	24
WONWATWGCIUMMQ-UHFFFAOYSA-N	MAC-0710498	-8.9	569
AOFAEARYDJCPGH-UHFFFAOYSA-N	MAC-0480442	-8.9	2
WQYLZKUBEXOBMX-UHFFFAOYSA-N	MAC-0713877	-8.9	731
HKJUBTSNKWVAPN-UHFFFAOYSA-N	MAC-0492780	-8.9	2
KKJZHOSFNMOHED-UHFFFAOYSA-N	MAC-0690918	-8.9	730
XZXLTQQMDYGJAL-UHFFFAOYSA-N	MAC-0444040	-8.9	26
CPEFTPDZUABFGW-UHFFFAOYSA-N	MAC-0693767	-8.9	3
WFWZYEFQAXVZPG-UHFFFAOYSA-N	MAC-0514282	-8.9	3
AKNOFIBDFSVONX-UHFFFAOYSA-N	MAC-0697660	-8.9	6
NJKMQAYNFHYEKB-UHFFFAOYSA-N	MAC-0667688	-8.9	9
HZEMZJZBXOLPIC-UHFFFAOYSA-N	MAC-0705954	-8.9	562
OMWGRZZRTCGAIT-UHFFFAOYSA-N	MAC-0701746	-8.8	777
OFRKZFRUMGVTMH-UHFFFAOYSA-N	MAC-0662547	-8.8	10
MPXINFNIWPYTSO-UHFFFAOYSA-N	MAC-0487532	-8.8	3

MXQNUEGNMIHJLT-UHFFFAOYSA-N	MAC-0443014	-8.8	29
HJLUHGHXBHRWLU-UHFFFAOYSA-N	MAC-0715720	-8.8	1
DKCNXHOCGXTEGV-UHFFFAOYSA-N	MAC-0711971	-8.7	775
IXEMIDNPWNMESH-UHFFFAOYSA-N	MAC-0674310	-8.7	10
RMRUYSVGZWIDJH-UHFFFAOYSA-N	MAC-0713553	-8.7	1
QZBJNXWXBXTYAO-UHFFFAOYSA-N	MAC-0502204	-8.7	644
CJLIRSYAWQBVKI-UHFFFAOYSA-N	MAC-0634718	-8.7	452

Table 1. A custom compound library that is heavily enriched for potential TMPRSS2 inhibitors. MACIDs represent the proprietary identification number for the McMaster CMCD. Compounds are ranked by pose score (more negative = higher binding affinity). The number of biological assays and appearances in literature (according to PubChem) have been highlighted in the rightmost column.

A successful TMPRSS2 inhibitor was identified in vitro

The active TMPRSS2 that was purified from *E. coli* was used to screen for TMPRSS2 inhibitors in our custom 50-compound library. Due to limitations in protein yield, only the top 10 scoring compounds in the library were assayed for inhibition in triplicate along with the two validated inhibitors Camostat mesylate and Nafamostat mesylate. Compounds were assayed for their effect on TMPRSS2 activity at a constant molar concentration, 25 µM. Camosat and Nafamostat both inhibited TMPRSS2 activity to 91.0±3.4% and 92.0±2.2% of basal activity levels respectively. Most compounds tested showed no significant decrease in TMPRSS2 activity, except compound C3, which inhibited activity to 92.6±5.0% (Figure 5). This result suggests that C3 inhibits TMPRSS2 to a similar degree as Camostat mesylate and Nafamostat mesylate, both potent inhibitors of TMPRSS2.

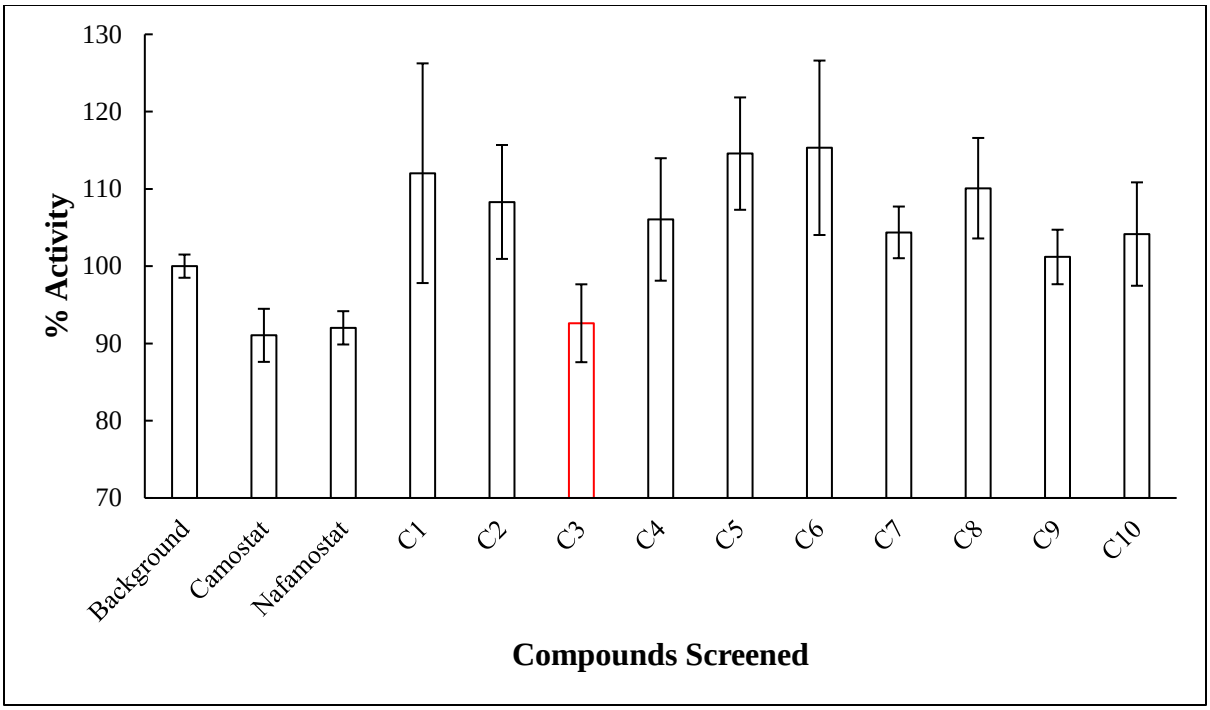


Figure 5. *In vitro* compound screening for inhibitors against TMPRSS2. The top 10 compounds from the top 50 list were chosen for an *in vitro* screen against TMPRSS2. Synergy H1 plate reader was used to measure relative fluorescence units every minute for 4 hours at 37 ºC on samples in triplicate with 25 µM compound, 10 µM peptide, and 0.6 µM of active TMPRSS2. Background readings were taken from the DMSO control well. Average RFU was normalized against background readings to achieve % activity. Camosat and Nafamostat both inhibited TMPRSS2 activity to 91.0±3.4% and 92.0±2.2%. C3 highlighted in red, inhibited activity to 92.6±5.0%. The assay layout can be found in Supplemental Table 3.

Compound C3 shows higher specificity for active site binding compared to Camostat and Nafamostat

MCULE's C3 docking model (docking score (DS) = -9.7) was used to compare to the docking models of Camostat (DS = -7.0) and Nafamostat (DS = -8.3). Using the ratio of 1 DS = 5.4-fold difference in binding affinity, C3 had a predicted 14.58-fold and 7.56-fold higher binding affinity to TMPRSS2 compared to Camostat and Nafamostat. Furthermore, analyzing the electrostatic surface model, C3 better fits into the binding pockets surrounding the active site region compared to Camostat and Nafamostat (Figure 6). From the electrostatic surface model, docking into the α pocket is exhibited by all compounds while binding to the β pocket is only coordinated by C3.

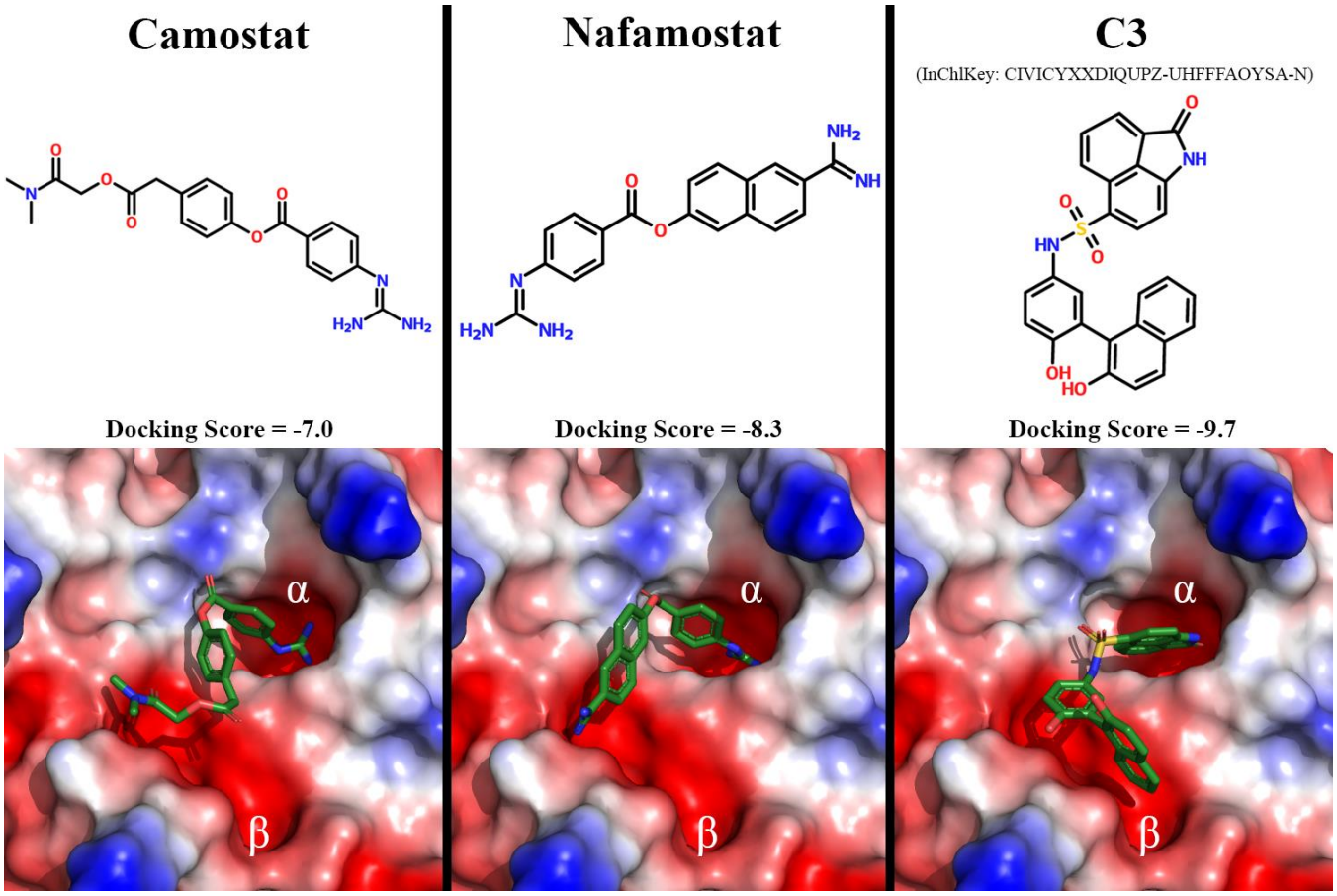


Figure 6. Electrostatic surface models of Camostat, Nafamostat, and C3 binding to the TMPRSS2 active site. All species can be seen docking to the α binding pocket. A flexible linker region allows the compounds to dock to the β binding pockets. The multi-ring structure of C3 allows better docking to the lower binding pockets and fills in the larger binding pocket as well. Compound structures are drawn above the docking images.

Methods

Homology Modeling and Verification

Due to the lack of any digital structure of TMPRSS2 in the Protein Databank, an accurate homology model was required to initialize the *in-silico* screening process for potential inhibitors of TMPRSS2. Initially, a highly cited and widely accepted homology modeling algorithm, Phyre2.0, was used. After sub-par results and an unusual structure due to overlapping atoms, The Robetta webserver was used instead (17). The primary sequence of full-length TMPRSS2 (108-492) was uploaded to the Robetta webserver which uses the Rosetta comparative modeling algorithm. A homology model based on Hepsin was generated and later, its active site residues and spatial position were verified by aligning the Robetta structure with a solved X-ray crystal model of one of our positive controls, Trypsin.

To ensure the workability and accuracy of our homology model, protein structure verification was performed by uploading the homology model to a well-documented webserver (SAVESv6) to ensure no unusual residues/linkages were detected and that our protease is stable.

In Silico Compound Screening

High throughput screening was performed using the MCULE webserver. From McMaster's compound library of 264, 158 molecules, the top-scoring 250 molecules were kept to be filtered later on. After our TMPRSS2 model was uploaded to MCULE, the τ-Nitrogen atom was selected on His296 to be a target for docking. The first basic property filter used the default settings apart from a maximum of 1 rule-of-five violation, and a minimum logP of -0.4. A REOS filter was applied thereafter followed by a sampler filter with size 100000 (randomized). Lastly, a diversity threshold of 0.85 between compounds was selected and the docking algorithm Vina (from AutoDock) was used.

After the initial screen, 250 compounds ranked by docking score (more negative indicates higher binding affinity) were systematically cross-referenced with the PubChem database for any previous use in biological assays for inhibitors and patents. Only the top 50 compounds with a binding affinity of at

least a factor of 1.5 higher than previously documented inhibitors (Camostat and Nafamostat) were selected to be ordered from McMaster's compound database.

Expression Construct Design

Given the cysteine-rich composition, auto-catalytic nature, and expression profile of TMPRSS2 (Genscript: NM_005656.4), three fusion protein constructs were designed to be expressed and purified from *E. coli* cells. Design and modifications were performed on SnapGene. The first fusion construct had only an N-terminal 6x His tag followed by the full length 108-492 TMPRSS2 sequence. The second construct (pMJ5833) included an N-terminal 6x His-MBP followed by the full-length sequence. Lastly, the third construct had an N-terminal periplasm signaling peptide-6x His-MBP tag followed by the full length 108-492 (pMJ5924).

Plasmid construction

Full-length TMPRSS2 (GenBank cDNA accession: NM_005656.4) in plasmid pcDNA3.1 was obtained from GenScript (Piscataway, NJ) and was given the plasmid accession number pMJ5825. Codons corresponding to residues 108-492 were amplified by PCR from pMJ5825 using primers MJ8249 (GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAAACCTGTATTTTCAGGGCTTCATGGGCAGCAAGTGCTCC) and MJ8245 (GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGCCGTCTGCCCTCATTTGT), to incorporate an N-terminal TEV-protease site and Gateway-compatible attB sites that flank the coding sequence. The PCR product was cloned into pDONR-201 by Gateway BP cloning to generate the Gateway entry clone pMJ5827. To generate the *E. coli* expression plasmids for His-TMPRSS2 and His-MBP-TMPRSS2, $TMPRSS2_{108\text{-}492}$ was cloned from pMJ5827 into either pDEST527 or pDEST566 via Gateway LR cloning to incorporate an N-terminal His_6- or His_6-MBP- fusion respectively. To generate the expression

plasmid for peri-His-MBP-TMPRSS2-His, $TMPRSS2_{108-492}$ was PCR amplified from pMJ5825 using primers MJ8249 and MJ8285 (GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAatggtggtgatgatggtgGCCGTCTGCCCTCATTTGT). This produced the same PCR product as before, but with an additional His_6 tag fusion on the C-terminus of TMPRSS2. The PCR product was cloned into pDONR-201 by Gateway BP cloning to generate the Gateway entry clone pMJ5923. Finally, TMPRSS2 from pMJ5923 was cloned into pDEST-periHisMBP to add an N-terminal peri-His_6-MBP fusion, yielding the *E. coli* expression vector pMJ5924. The sequence of the TMPRSS2 insert in all Gateway entry clone plasmids were verified by two Sanger sequencing reactions using either primers MJ3863 (TTAACGCTAGCATGGATCT) or MJ3864 (AACATCAGAGATTTTGAGACAC).

Growth of E. coli and Expression of TMPRSS2

After obtaining the final plasmid constructs from the LR reaction using the presto mini plasmid kit, 3 µL of the plasmid was transformed into an aliquot of BL21 (DE3) T1R competent cells using the same protocol described above. Instead of a single-colony liquid culture, 3 colonies from the BL21 plate were used for a 50 mL liquid culture. During this time, 2x 1 L of autoclaved LB was prepared for inoculation. 1 mL of 1000x Kan was added to the 1 L autoclaved and warmed up LB. 10 mL of the 50 mL liquid culture was used to inoculate each of the 1 L flasks of LB growth media. Both flasks of 1 L culture were grown at 37 ^{o}C on a shaker until its OD_{600} reached a target of 0.3-0.5. OD_{600} measurements were taken every 45 minutes.

After the OD_{600} has reached our target, 1 mL of IPTG was added to the flasks (final concentration 1 mM) and the cells were grown overnight at 30 ^{o}C. The total 2 L sample of liquid culture was centrifuged, and four pellets were collected (500 mL culture in each large centrifuge tube).

Purification of TMPRSS2

Before injecting the contents into our 5 mL Ni-column, the column was cleaned to strip impurities and regenerate the column with fresh NiCl (100 mM) solution. 10 column volumes (CV) of mili-Q water was run through the column at 2 mL/min using a peristaltic pump, followed by 10 CV of 0.5 M NaOH. Another 10 CV of water was injected before and after washing the column with 20 CV 2M NaCl + 50 mM EDTA at 2 mL/min. Finally, 3 CV of 100 mM NiCl was used to recharge the column at 0.5 mL/min. Before the initial preparation of 10 CV of Ni-A solution (20 mM Tris pH 8.5, 1 M NaCl, 10% (v/v) Glycerol, 10 mM imidazole, 2 mM BME), 10 CV of water was injected again.

After the column was equilibrated with Ni-A, a cell pellet obtained from the previous step was suspended in Ni-A buffer. The homogenous solution of cells was lysed using the French press and the lysate was spun down at 20, 000 RPM for 50 minutes to separate the soluble fraction of cell contents. After centrifugation, the soluble fraction was injected into our equilibrated column to be purified using the AKTA Start.

For the entire wash and elution process, Ni-B solution (20 mM Tris pH 8.5, 1 M NaCl, 10% (v/v) Glycerol, 300 mM imidazole, 2 mM BME) was used in incremental amounts. Initially 50-100 mL of 0% Ni-B wash was collected. After the 0% wash, 15 mL increments of Ni-B percentage were collected (e.g. 0-1.5%, 1.5-3%, 3-4.5% 4.5-6%, 6-7.5%, 7.5-9%, 9-10.5%, 10.5-12%, 12-13.5%, 13.5-15%) totaling 150 mL of washed samples. After the washes, 100% Ni-B was used to collect 30 mL of the elutant.

For the modified second purification of pMJ5924 and mock purification of pMJ5899, the soluble lysate was cycled within the Ni column for 2 hours at 0.5 mL/min using the peristaltic pump before 15 mL of the 0% wash was taken. Further changes included taking two samples of the 10% wash after 15 mL 1.5% increments of Ni-B.

After the Ni IMAC purification process, an SDS-PAGE was run (165 V for 60 mins) with all the wash and elution fractions to measure protein length. A Bradford assay was performed to calculate protein

concentration before and after concentrating the protein using a 10 kDa centrifugal porous membrane concentrator.

Two methods to exchange the buffer that TMPRSS2 was in were used. The first purification of pMJ5924 underwent a buffer exchange by Desalting chromatography. The second purification of pMJ5924 underwent overnight dialysis after being concentrated to ~0.8 mL using a 10 kDa centrifugal porous membrane concentrator.

Assay Design and Optimization

The fluorescence activity assay was developed from a pre-existing assay for the activity of TMPRSS2 (13). The peptide (Boc-Gln-Ala-Arg-AMC) used for this assay was ordered from BACHEM and 10 µL aliquots of 100 µM stock were prepared in DMSO. The volumes, final concentrations, and the order of addition of peptide, TMPRSS2, control proteases, and compounds can be found in the supplemental tables 1 to 3. Fluorescence readings were captured every minute using the Synergy H1 96-well plate reader for a minimum of 60 minutes at 37 ^{o}C with an excitation wavelength of 340 nm and emission detection of 440 nm. For the compound screen, a total read of 4 hours was implemented so fluorescent peptide degradation can also be observed.

For the compound screen, the plate with TMPRSS2, activity buffer, and compounds were orbitally shaken at 282 CPM at 37 ^{o}C in the Synergy H1 plate reader for 10 minutes before adding peptide. For all the activity screens, the fluorescent peptide was added last to initiate the reaction.

Discussion

As of the end of this project (April 2nd, 2021), a crystal structure of hTMPRSS2 still has not been deposited into the protein databank. Not only would a high-resolution crystal structure of TMPRSS2 enhance the accuracy of *in silico* compound screening, but it would also open up opportunities for more accurate molecular dynamics simulations and research elucidating the autocatalytic mechanism and Spike priming abilities (26). However, sufficient active protease is required for downstream studies including the structural characterization by X-ray crystallography. Our study has outlined a protein expression and purification method to isolate active TMPRSS2 from *E. coli*. To date, an abundant amount of active TMPRSS2 has not yet been produced from *E. coli* by recombinant expression (13). Previous purifications of TMPRSS2 used monoclonal antibodies to purify the extracellular domains of TMPRSS2 from HEK293T cells (27), as well as expressing and purifying TMPRSS2 from yeast. Not only do HEK293T cell lines increase the demand for resources in the case of protein purification, but commercially purchasable TMPRSS2 expressed by yeast is very costly. We hope that future studies can use the data from our purification trials to optimize an efficient and economically friendly protocol to purify large amounts of active TMPRSS2 from *E. coli*.

The lack of enough protein to screen all 50 compounds of interest was from the many challenges surrounding the purification of TMPRSS2. Firstly, we were aware of its intrinsic autocatalytic activity, but we did not know under what conditions TMPRSS2 would cleave itself after R255 (14). Although there were no previous studies regarding the timing and conditions of the autocatalytic cleavage event, we initially hypothesized that due to the spatial proximity of active site and target residues, autocatalysis would occur immediately. After subsequent purifications and activity assays, we have concluded that autocatalysis takes at least 72 hours to occur and does so in low-salt concentrations (150 mM NaCl). Secondly, we could not add irreversible protease inhibitors into our Ni IMAC protocol due to TMPRSS2 being a serine protease itself. The lack of protease inhibitors could be a limitation during purification due to possible bacterial protease contaminants but given the mechanism of Ni-IMAC, we're confident that

the eluted proteins would have to be our protein of interest due to their His 6x tag specificities as well as thorough gradient washes of increasing Imidazole concentration to wash off rouge proteases (28). In the future, a site-directed mutagenesis control purification using the same expression and purification protocols could be performed to increase the statistical confidence that the elutant contains our protease of interest.

Another note to make for future purifications would be to have a His 6x tag only at the C-terminal. After our initial purification, we would give our full-length protein the proper time and environment where autocatalysis could occur and perform another Ni-IMAC with our autocleaved-active TMPRSS2. Due to the presence of a C-terminal His 6x tag, our purity would increase if we ran our autocleaved sample through the Ni-column again. Furthermore, after cross-referencing with different purification protocols, a single poly-histidine on the C-terminus would have been a better option as our active proteolytic core resides on the C-terminal after auto-catalytic cleavage (28). Additionally, to optimize the time, temperature, buffer, and salt concentrations where most of the full-length species would undergo autocatalytic events, multiple activity assays would have to be performed. For example, after coming off the first Ni-column, samples could be divided up and stored in the -80 ^{o}C immediately to maximize optimization trials over time. Samples then can be dialyzed into variable salt concentration buffers overnight as an example of testing the salt conditions. Furthermore, the ratio of full-length to cleaved-active protease can be monitored daily, considering the time variable. Ultimately, not only can more compound screens be performed with purer protease, the activity assay can be further optimized to use as little protein as possible for a noticeable activity window. During the *in vitro* compound screen, a dose-dependent experiment could also be carried out with a low, medium, and high concentration of compound required to inhibit protease function.

From our *in vitro* compound screen, we have identified one compound of interest (Compound 3 92.6±5.0% activity) that had matching inhibitory effects with our clinically approved control compound Camostat mesylate and its analog Nafamostat mesylate. Given only top 10 compounds were screened, our

current hit rate presents itself at 10%. On average, typical high throughput screens for drug-like compounds can have a hit rate between 0.01% and 0.14% (29). Not only did our preemptive *in silico* screen using an enriched compound library improve the average hit rate by 100-fold, but it also fits well within the range of expected hit rates of *in vitro* screened compounds selected from a virtual screen (literature hit rates between 1% and 40%) (29). Furthermore, we have demonstrated that computational screening methods can narrow down compounds of interest to save money and the time it would have taken to screen all 264, 158 compounds in an *in vitro* setting. In addition, *in silico* technologies have been highly praised in the past century as an efficient means to screen drug-like compounds, but also the algorithms to predict structures and off-target interactions become increasingly more accurate (30). Our results ultimately can serve as a proof-of-concept regarding *in silico* and *in vitro* hybrid drug-screening techniques. However, the key to an accurate *in silico* screen is the structural quality of the molecular target of interest (29).

To date, homology models have been praised as good starting points for structural and functional analysis, however, for a model to be employed in high throughput *in silico* drug screen, a high-resolution model would greatly improve the accuracy of compound hits and the reliability of the docking scores (30). Given the inherent high conservation of the active site in serine proteases (31), the high-resolution structure of Hepsin (1Z8G) (18), upon which our homology model was built, provided a quality homology model of TMPRSS2. However, even if the scaffold structure has a high resolution, digital structure validation screens would have to be initiated due to uncontrollable variables inherent to a homology model (32). After generating the structure from Rosetta, the model was run through the Structural Analysis and Verification Server (SAVESv6.0) to verify our model to assess the quality and usability. SAVESv6.0 was chosen due to its extensive suite of validation algorithms and popularity in the field of computational biochemistry. The results calculated from ERRAT, VERIFY3D, PROVE, WHATCHECK, and PROCHECK gave us confidence in our homology model. As ERRAT's overall quality factor looks for incorrectly built regions in protein models, our homology model scored a 95% confidence interval of

predicted fidelity (19). VERIFY3D determines the correctness of a model by comparing its 3D profile against solved high-resolution protein structures that matches the input sequence, the 99.58% of resides that passed indicates our model has correct 3D protein folding properties given its sequence (20). After calculating the atomic volumes and calculating Z-scores of individual atoms, PROVE only reported 3.8% of atoms deviate significantly from expected values, where the reported usability of a model falls after 30% (21). As a control, the crystal structure of Hepsin (1Z8G) was processed through SAVESv6.0 and the results were only slightly higher than the predicted structure of TMPRSS2. Although other verification methods such as SwissViewProt and running molecular dynamics simulations exist, SAVESv6.0 was not only time-friendly but utilized a wide range of different algorithms to calculate the quality of our model (32).

As accurate as the algorithms are for *in silico* experiments, they do not give us the full picture of a protein-compound interaction. Based on a multitude of studies questioning the legitimacy and statistically significance of the results generated from *in silico* docking of compounds, it has been proven that *in silico* compound screening is not only a good starting point to save resources, but as mentioned before, *in vitro* hit rates increases after using *in silico* algorithms to enrich a compound library (33). MCULE was chosen for this endeavor as its primary docking algorithm uses AutoDock (Vina) and produces docking scores that are highly representative of *in vitro* binding studies. In a comparison of docking approaches study, AutoDock (Vina) was compared to Glide, GOLD, LeDock, rDOCk, and a slew of other commercial applications (34). AutoDock (Vina) was found to have the highest correlations between experimental binding affinities and docking scores within the study (34). From McMaster's 264,158 compounds, only the top 250 scoring compounds were chosen to be saved after the library was run through MCULE's structure-based virtual screen workflow. During the workflow, drug-like filters were added to ensure that the top 250 hits could all be used for a drug-like compound screen. The first filter dictated that there could be only a maximum of 1 rule-of-five (RO5) violation. The RO5 filter was put in place because although Lipinski's rules for what has to be considered an orally active drug, there

have been exceptions (35), and allowing one exception to any of the five rules will increase the sampling size of our compounds. The second filter that was changed was the minimum value of log(P); setting the minimum value of log(P) = -0.4 initializes the parameters for this drug-like filter based on hydrophobicity (36). Lastly, a diversity selection of a maximum of 0.85 similarity was put in place to avoid compounds that would look too similar structurally. Not only would the diversity parameter increase our chances of finding unique compounds, but it would also allow for an increased sampling size based on structural differences (37). Although all 250 hits had a greater binding affinity to the H296 residue on TMPRSS2 than Camostat (-7.0 docking score), only 50 were selected for *in vitro* assay screening. The method for selecting 50 out of 250 was primarily based on the docking score. Secondary considerations were made after searching for each compound's presence in previous drug screens and/or patents.

After identifying C3 (CIVICYXXDIQUPZ-UHFFFAOYSA-N) to have inhibited TMPRSS2 to the same degree as Camostat mesylate and Nafamostat mesylate, its docked structure was extracted from MCULE and analyzed using PyMOL's electrostatic surface model function. When placed side-by-side with the Camostat and Nafamostat docked structure, distinct differences can be seen with how the triple-ring structure linked directly to the Sulfone functional group fills the upper binding pocket and makes non-polar interactions using its benzyl ring structures. Furthermore, since all compounds fit into the α pocket, it can be deduced that binding to the α pocket is a requirement for protease inhibition. C3 shows uniqueness where it can also coordinate with the β pocket. Due to the differences in the docking scores and models, it can be hypothesized that coordinating the β pocket introduces further specificity for the active site on TMPRSS2. The middle region linking the two-pocket binding functional groups appears to be flexible, however, the sulfone functional group specializes in coordinating the two bulky compound rings to stabilize at an angle to better fit the active site. For future *in silico* screens, C3 can be used as a starting structure in MCULE's 1-click-scaffold-hop program for lead optimization. Further research into C3 also reveals its patents, especially as a documented small molecule inhibitor against West Nile virus

replication (38). Ultimately, even though only 10 compounds were screened, C3 seems promising with its highly specific characteristics as well as previous documentation in clinical trials for other viral infections.

Due to the COVID-19 pandemic, an uncontrollable limitation for this project included shipping cancelations from unexpected industry lab shutdown and the limited time thereafter to work on the project. However, invaluable research experience was gained from designing fusion expression constructs, performing gateway cloning protocols, transforming custom expression vectors into capable cells, and protein purification methodologies. To gain more insight into the inhibitory roles of compounds, an IC50 curve from the measured RFU data would provide a greater foundation to compare the efficacy of the selected compounds (13). Furthermore, to improve the generalizability of our compounds, an *in vivo* lung-cell-based infectivity assay would need to be developed to further assess the efficacy of compounds. To our advantage, Dr. XXXXX at XXXXX University currently has the cells and protocols required for an *in vivo* infectivity screen (39). After developing and optimizing activity and infectivity assays, efforts should be focused on generating enough protein to perform crystallization trials to generate a high-resolution digital structure via X-ray crystallography. A new high throughput compound screen would be required as the new model would now be the gold standard in terms of representativeness of hTMPRSS2.

Ultimately, this project has demonstrated that even though antiviral drugs may already exist, computational high throughput compound screening was beneficial for detecting compounds with higher binding specificity. Although most of the time was spent on troubleshooting the purification protocol, valuable information regarding the autocatalytic conditions, fluorescence assay sensitivity, and the positive correlation between *in silico* versus *in vitro* results.

References

1. Mollica, V., Rizzo, A., and Massari, F. (2020) The pivotal role of TMPRSS2 in coronavirus disease 2019 and prostate cancer. *Future Oncology*. **16**, 2029–2033

2. Ye, Z., Zhang, Y., Wang, Y., Huang, Z., and Song, B. (2020) Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. *Eur Radiol*. **30**, 4381–4389

3. Hussain, M., Jabeen, N., Amanullah, A., Baig, A. A., Aziz, B., Shabbir, S., and Raza, F. (2020) Structural Basis of SARS-CoV-2 Spike Protein Priming by TMPRSS2. *bioRxiv*. 10.1101/2020.04.21.052639

4. 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., and Pöhlmann, S. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. **181**, 271-280.e8

5. Tarnow, C., Engels, G., Arendt, A., Schwalm, F., Sediri, H., Preuss, A., Nelson, P. S., Garten, W., Klenk, H.-D., Gabriel, G., and Böttcher-Friebertshäuser, E. (2014) TMPRSS2 Is a Host Factor That Is Essential for Pneumotropism and Pathogenicity of H7N9 Influenza A Virus in Mice. *Journal of Virology*. **88**, 4744–4751

6. Kim, T. S., Heinlein, C., Hackman, R. C., and Nelson, P. S. (2006) Phenotypic Analysis of Mice Lacking the Tmprss2-Encoded Protease. *Mol Cell Biol*. **26**, 965–975

7. Hoffmann, M., Hofmann-Winkler, H., Smith, J. C., Krüger, N., Sørensen, L. K., Søgaard, O. S., Hasselstrøm, J. B., Winkler, M., Hempel, T., Raich, L., Olsson, S., Yamazoe, T., Yamatsuta, K., Mizuno, H., Ludwig, S., Noé, F., Sheltzer, J. M., Kjolby, M., and Pöhlmann, S. (2020) Camostat mesylate inhibits SARS-CoV-2 activation by TMPRSS2-related proteases and its metabolite GBPA exerts antiviral activity. *bioRxiv*. 10.1101/2020.08.05.237651

8. Talukdar, R., and Tandon, R. K. (2008) Pancreatic stellate cells: new target in the treatment of chronic pancreatitis. *J Gastroenterol Hepatol*. **23**, 34–41

9. Hempel, T., Raich, L., Olsson, S., P. Azouz, N., M. Klingler, A., Hoffmann, M., Pöhlmann, S., E. Rothenberg, M., and Noé, F. (2021) Molecular mechanism of inhibiting the SARS-CoV-2 cell entry facilitator TMPRSS2 with camostat and nafamostat. *Chemical Science*. **12**, 983–992

10. Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., and Nagata, N. (2019) TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *Journal of Virology*. 10.1128/JVI.01815-18

11. Shen, L.-W., Qian, M.-Q., Yu, K., Narva, S., Yu, F., Wu, Y.-L., and Zhang, W. (2020) Inhibition of Influenza A virus propagation by benzoselenoxanthenes stabilizing TMPRSS2 Gene G-quadruplex and hence down-regulating TMPRSS2 expression. *Scientific Reports*. **10**, 7635

12. WILSON, S., GREER, B., HOOPER, J., ZIJLSTRA, A., WALKER, B., QUIGLEY, J., and HAWTHORNE, S. (2005) The membrane-anchored serine protease, TMPRSS2, activates PAR-2 in prostate cancer cells. *Biochemical Journal*. **388**, 967–972

13. Shrimp, J. H., Kales, S. C., Sanderson, P. E., Simeonov, A., Shen, M., and Hall, M. D. (2020) An Enzymatic TMPRSS2 Assay for Assessment of Clinical Candidates and Discovery of Inhibitors as Potential Treatment of COVID-19. *ACS Pharmacol Transl Sci*. 10.1021/acsptsci.0c00106

14. Afar, D. E., Vivanco, I., Hubert, R. S., Kuo, J., Chen, E., Saffran, D. C., Raitano, A. B., and Jakobovits, A. (2001) Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia. *Cancer Res*. **61**, 1686–1692

15. Nallamsetty, S., Austin, B. P., Penrose, K. J., and Waugh, D. S. (2005) Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli. *Protein Science*. **14**, 2964–2971

16. Song, Y., DiMaio, F., Wang, R. Y.-R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013) High resolution comparative modeling with RosettaCM. *Structure*. 10.1016/j.str.2013.08.005

17. Hiranuma, N., Park, H., Baek, M., Anishchanka, I., Dauparas, J., and Baker, D. (2020) Improved protein structure refinement guided by deep learning based accuracy estimation. *bioRxiv*. 10.1101/2020.07.17.209643

18. Herter, S., Piper, D. E., Aaron, W., Gabriele, T., Cutler, G., Cao, P., Bhatt, A. S., Choe, Y., Craik, C. S., Walker, N., Meininger, D., Hoey, T., and Austin, R. J. (2005) Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers. *Biochemical Journal*. **390**, 125–136

19. Colovos, C., and Yeates, T. O. (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. **2**, 1511–1519

20. Lüthy, R., Bowie, J. U., and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*. **356**, 83–85

21. Pontius, J., Richelle, J., and Wodak, S. J. (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol*. **264**, 121–136

22. Hooft, R. W. W., Vriend, G., Sander, C., and Abola, E. E. (1996) Errors in protein structures. *Nature*. **381**, 272–272

23. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*. **26**, 283–291

24. Kiss, R., Sandor, M., and Szalai, F. A. (2012) http://Mcule.com: a public web service for drug discovery. *J Cheminform*. **4**, P17

25. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem*. **31**, 455–461

26. Markosian, C., Di Costanzo, L., Sekharan, M., Shao, C., Burley, S. K., and Zardecki, C. (2018) Analysis of impact metrics for the Protein Data Bank. *Scientific Data*. **5**, 180212

27. Chen, X., Xu, Z., Zeng, S., Wang, X., Liu, W., Qian, L., Wei, J., Yang, X., Shen, Q., Gong, Z., and Yan, Y. (2019) The Molecular Aspect of Antitumor Effects of Protease Inhibitor Nafamostat Mesylate and Its Role in Potential Clinical Applications. *Front. Oncol.* 10.3389/fonc.2019.00852

28. Bornhorst, J. A., and Falke, J. J. (2000) [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymol.* **326**, 245–254

29. Zhu, T., Cao, S., Su, P.-C., Patel, R., Shah, D., Chokshi, H. B., Szukala, R., Johnson, M. E., and Hevener, K. E. (2013) Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based Upon a Critical Literature Analysis. *J Med Chem.* **56**, 6560–6572

30. Zloh, M., and Kirton, S. B. (2018) The benefits of in silico modeling to identify possible small-molecule drugs and their off-target interactions. *Future Medicinal Chemistry*. **10**, 423–432

31. Krem, M. M., and Di Cera, E. (2001) Molecular markers of serine protease evolution. *EMBO J.* **20**, 3036–3045

32. Haddad, Y., Adam, V., and Heger, Z. (2020) Ten quick tips for homology modeling of high-resolution protein 3D structures. *PLOS Computational Biology*. **16**, e1007449

33. T G, N., Venkatachalam, A., and Sugathan, S. (2017) High-Throughput and In Silico Screening in Drug Discovery. in *Bioresources and Bioprocess in Biotechnology*, 10.1007/978-981-10-3573-9_11

34. Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., Tian, S., and Hou, T. (2016) Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **18**, 12964–12975

35. Doak, B. C., Over, B., Giordanetto, F., and Kihlberg, J. (2014) Oral Druggable Space beyond the Rule of 5: Insights from Drugs and Clinical Candidates. *Chemistry & Biology*. **21**, 1115–1142

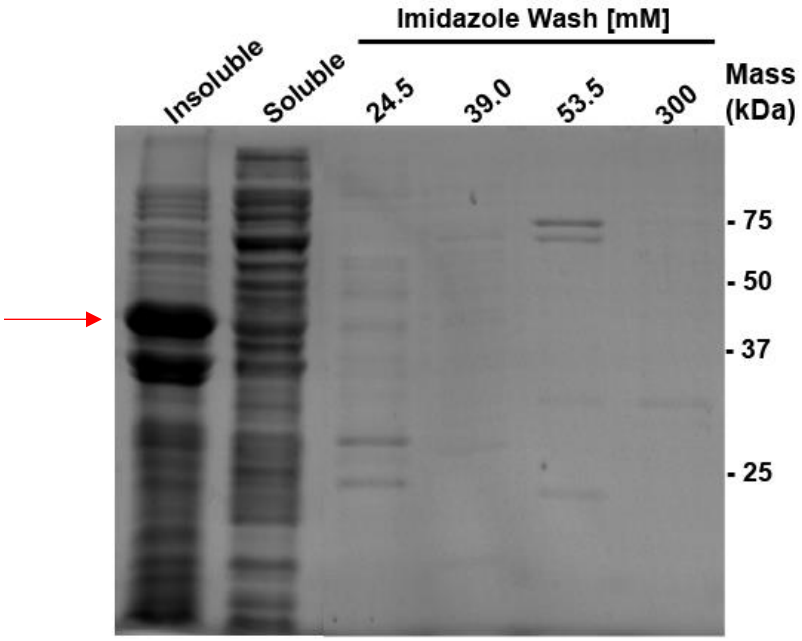
36. Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. (1999) A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1**, 55–68

37. Huggins, D. J., Venkitaraman, A. R., and Spring, D. R. (2011) Rational Methods for the Selection of Diverse Screening Compounds. *ACS Chem Biol*. **6**, 208–217

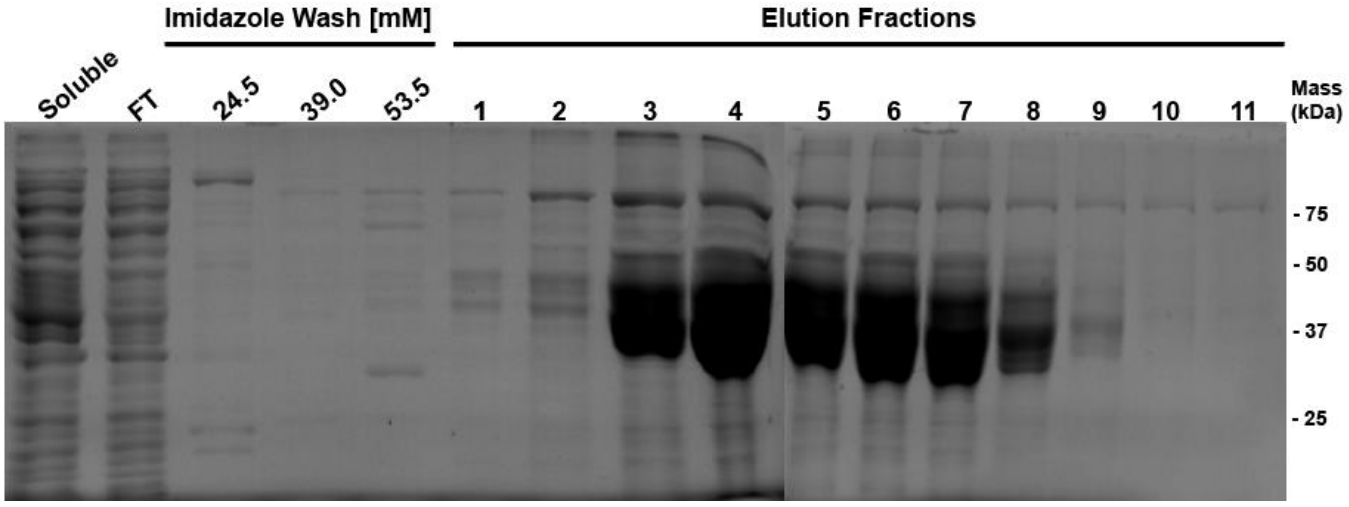
38. Gu, B., Block, T., and Cuconati, A. (2007) Small Molecule Inhibitors Against West Nile Virus Replication. [online] https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2007005541 (Accessed April 4, 2021)

39. Zheng, M., Zhao, X., Zheng, S., Chen, D., Du, P., Li, X., Jiang, D., Guo, J.-T., Zeng, H., and Lin, H. Bat SARS-Like WIV1 coronavirus uses the ACE2 of multiple animal species as receptor and evades IFITM3 restriction via TMPRSS2 activation of membrane fusion. *Emerg Microbes Infect*. **9**, 1567–1579

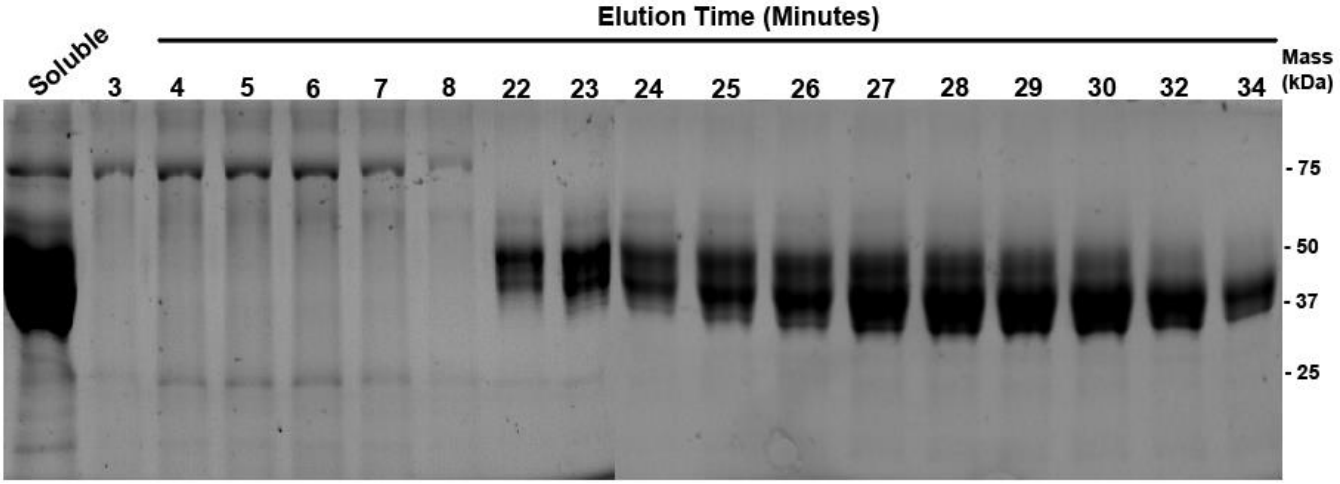
Supplemental Data



Supplemental Figure 1. Ni-IMAC Purification results from pMJ5832. Protein of interest, red arrow indicates expected molecular weight = 42.6 kDa, was not eluted at 300 mM Imidazole wash. Variable imidazole concentration washes were collected at 15 mL intervals. The majority of the protein can be seen in the insoluble lane slightly above 37 kDa. SDS-PAGE was run for 45 minutes at 165 V.



Supplemental Figure 2A. Ni-IMAC Purification results from pMJ5833. Larger species with apparent MW between 76 and 100 kDa eluted between fractions 3 and 8. Smaller species with apparent MW between 37 and 50 kDa eluted between fractions 9 and 11. Variable imidazole concentration washes were collected at 15 mL intervals. SDS-PAGE was run for 45 minutes at 165 V.



Supplemental Figure 2B. Gel filtration chromatography separation of pMJ5833 isolated by Ni-IMAC. Smaller species around 76 – 100 kDa were separated from 3 – 8 minutes. Larger species around 37 – 50 kDa were separated from 22 – 34 minutes. SDS-PAGE was run for 45 minutes at 165 V.

Well ID	Buffer (µL)	80 kDa Species (µL)	Fluorescence Peptide (µL)	80 kDa Species [µM]	Fluorescence Peptide [µL]
A1+A5+A9	43.42	6.58	0	2	0
A2+A6+A10	38.42	6.58	5	2	5
A3+A7+A11	33.42	6.58	10	2	10
A4+A8+A12	23.42	6.58	20	2	20
B1+B5+B9	46.71	3.29	0	1	0
B2+B6+B10	41.71	3.29	5	1	5
B3+B7+B11	36.71	3.29	10	1	10
B4+B8+B12	26.71	3.29	20	1	20
C1+C5+C9	48.36	1.64	0	0.5	0
C2+C6+C10	43.36	1.64	5	0.5	5
C3+C7+C11	38.36	1.64	10	0.5	10
C4+C8+C12	28.36	1.64	20	0.5	20
D1+D5+D9	50	0	0	0	0
D2+D6+D10	45	0	5	0	5
D3+D7+D11	40	0	10	0	10
D4+D8+D12	30	0	20	0	20

Supplemental Table 1A. 96-Well setup for activity assay of the 80 kDa species from the pMJ5833 purification. Activity buffer consisted of 150 mM NaCl, 50 mM Tris pH 8.0. In the same 96-Well plate, the 40 kDa species were also measured for possible protease activity. Various protein and peptide concentrations were utilized for optimization purposes. 2 µL of concentrated Chymotrypsin was added to wells D4, D8, D12 after no activity was detected (final concentration 15.38 µM). Buffer was added first, and fluorescent peptide was added last to initiate the reaction.

Well ID	Buffer (µL)	40 kDa Species (µL)	Fluorescence Peptide (µL)	40 kDa Species [µM]	Fluorescence Peptide [µL]
E1+E5+E9	47.08	2.92	0	2	0
E2+E6+E10	42.08	2.92	5	2	5
E3+E7+E11	37.08	2.92	10	2	10
E4+E8+E12	27.08	2.92	20	2	20
F1+F5+F9	48.54	1.46	0	1	0
F2+F6+F10	43.54	1.46	5	1	5
F3+F7+F11	38.54	1.46	10	1	10
F4+F8+F12	28.54	1.46	20	1	20
G1+G5+G9	49.27	0.73	0	0.5	0
G2+G6+G10	44.27	0.73	5	0.5	5
G3+G7+G11	39.27	0.73	10	0.5	10
G4+G8+G12	29.27	0.73	20	0.5	20
H1+H5+H9	50	0	0	0	0
H2+H6+H10	45	0	5	0	5
H3+H7+H11	40	0	10	0	10
H4+H8+H12	30	0	20	0	20

Supplemental Table 1B. 96-Well setup for activity assay of the 40 kDa species from the pMJ5833 purification. Activity buffer consisted of 150 mM NaCl, 50 mM Tris pH 8.0. In the same 96-Well plate, the 80 kDa species were also measured for possible protease activity. Various protein and peptide concentrations were utilized for optimization purposes. Buffer was added first, and fluorescent peptide was added last to initiate the reaction.

Stock Solutions				
Compound	MW (Da)	Stock Concentration (mg/mL)	(µM)	Dissolved In
Peptide	667.16	0.066716	100	1x Buffer
PMSF	174.2	1.742	10000	1x Buffer
Lysozyme	14300	0.715	50	1x Buffer
Trypsin	23300	1.165	50	1x Buffer
Chymotrypsin	25000	1.25	50	1x Buffer
Fraction 1.5-3	27000	0.4365	16.16667	1x Buffer
Fraction 7.5-9	27000	0.063333	2.345679	1x Buffer
Fraction 13.5-15	27000	0.076667	2.839506	1x Buffer
Elution	27000	0.07	2.592593	1x Buffer

Well	Protease	Protease Conc.	PMSF	Vol Buffer (µL)	Vol Peptide (µL)	Vol PMSF (µL)	Vol Protease (µL)
A1	None	Low	-	90	10	0	0
A2	None	High	-	90	10	0	0
A3	None	Low	+	80	10	10	0
A4	None	High	+	80	10	10	0
A5	None	Low	-	90	10	0	0
A6	None	High	-	90	10	0	0
A7	None	Low	+	80	10	10	0
A8	None	High	+	80	10	10	0
A9	None	Low	-	90	10	0	0
A10	None	High	-	90	10	0	0
A11	None	Low	+	80	10	10	0
A12	None	High	+	80	10	10	0
B1	Lysozyme	Low	-	88	10	0	2
B2	Lysozyme	High	-	60	10	0	30
B3	Lysozyme	Low	+	78	10	10	2
B4	Lysozyme	High	+	50	10	10	30
B5	Lysozyme	Low	-	88	10	0	2
B6	Lysozyme	High	-	60	10	0	30
B7	Lysozyme	Low	+	78	10	10	2
B8	Lysozyme	High	+	50	10	10	30
B9	Lysozyme	Low	-	88	10	0	2
B10	Lysozyme	High	-	60	10	0	30
B11	Lysozyme	Low	+	78	10	10	2
B12	Lysozyme	High	+	50	10	10	30
C1	Trypsin	Low	-	88	10	0	2
C2	Trypsin	High	-	60	10	0	30
C3	Trypsin	Low	+	78	10	10	2
C4	Trypsin	High	+	50	10	10	30
C5	Trypsin	Low	-	88	10	0	2
C6	Trypsin	High	-	60	10	0	30
C7	Trypsin	Low	+	78	10	10	2
C8	Trypsin	High	+	50	10	10	30
C9	Trypsin	Low	-	88	10	0	2
C10	Trypsin	High	-	60	10	0	30
C11	Trypsin	Low	+	78	10	10	2
C12	Trypsin	High	+	50	10	10	30
D1	Chymotrypsin	Low	-	88	10	0	2
D2	Chymotrypsin	High	-	60	10	0	30
D3	Chymotrypsin	Low	+	78	10	10	2
D4	Chymotrypsin	High	+	50	10	10	30
D5	Chymotrypsin	Low	-	88	10	0	2
D6	Chymotrypsin	High	-	60	10	0	30

D7	Chymotrypsin	Low	+	78	10	10	2
D8	Chymotrypsin	High	+	50	10	10	30
D9	Chymotrypsin	Low	-	88	10	0	2
D10	Chymotrypsin	High	-	60	10	0	30
D11	Chymotrypsin	Low	+	78	10	10	2
D12	Chymotrypsin	High	+	50	10	10	30
E1	18.7 mM IMD	Low	-	80	10	0	10
E2	18.7 mM IMD	High	-	40	10	0	50
E3	18.7 mM IMD	Low	+	70	10	10	10
E4	18.7 mM IMD	High	+	30	10	10	50
E5	18.7 mM IMD	Low	-	80	10	0	10
E6	18.7 mM IMD	High	-	40	10	0	50
E7	18.7 mM IMD	Low	+	70	10	10	10
E8	18.7 mM IMD	High	+	30	10	10	50
E9	18.7 mM IMD	Low	-	80	10	0	10
E10	18.7 mM IMD	High	-	40	10	0	50
E11	18.7 mM IMD	Low	+	70	10	10	10
E12	18.7 mM IMD	High	+	30	10	10	50
F1	36.1 mM IMD	Low	-	80	10	0	10
F2	36.1 mM IMD	High	-	40	10	0	50
F3	36.1 mM IMD	Low	+	70	10	10	10
F4	36.1 mM IMD	High	+	30	10	10	50
F5	36.1 mM IMD	Low	-	80	10	0	10
F6	36.1 mM IMD	High	-	40	10	0	50
F7	36.1 mM IMD	Low	+	70	10	10	10
F8	36.1 mM IMD	High	+	30	10	10	50
F9	36.1 mM IMD	Low	-	80	10	0	10
F10	36.1 mM IMD	High	-	40	10	0	50
F11	36.1 mM IMD	Low	+	70	10	10	10
F12	36.1 mM IMD	High	+	30	10	10	50
G1	53.5 mM IMD	Low	-	80	10	0	10
G2	53.5 mM IMD	High	-	40	10	0	50
G3	53.5 mM IMD	Low	+	70	10	10	10
G4	53.5 mM IMD	High	+	30	10	10	50
G5	53.5 mM IMD	Low	-	80	10	0	10
G6	53.5 mM IMD	High	-	40	10	0	50
G7	53.5 mM IMD	Low	+	70	10	10	10
G8	53.5 mM IMD	High	+	30	10	10	50
G9	53.5 mM IMD	Low	-	80	10	0	10
G10	53.5 mM IMD	High	-	40	10	0	50
G11	53.5 mM IMD	Low	+	70	10	10	10
G12	53.5 mM IMD	High	+	30	10	10	50
H1	300 mM IMD	Low	-	80	10	0	10
H2	300 mM IMD	High	-	40	10	0	50
H3	300 mM IMD	Low	+	70	10	10	10

H4	300 mM IMD	High	+	30	10	10	50
H5	300 mM IMD	Low	-	80	10	0	10
H6	300 mM IMD	High	-	40	10	0	50
H7	300 mM IMD	Low	+	70	10	10	10
H8	300 mM IMD	High	+	30	10	10	50
H9	300 mM IMD	Low	-	80	10	0	10
H10	300 mM IMD	High	-	40	10	0	50
H11	300 mM IMD	Low	+	70	10	10	10
H12	300 mM IMD	High	+	30	10	10	50

	Technical replicate:	Replicate 1				Replicate 2				Replicate 3			
	PMSF (1mM):	-	-	+	+	-	-	+	+	-	-	+	+
	Protein Concentration:	[Low]	[High]	[Low]	[High]	[Low]	[High]	[Low]	[High]	[Low]	[High]	[Low]	[High]
Sample	Buffer only	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
	Lysozyme	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
	Trypsin	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
	Chymotrypsin	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
	18.7 mM IMD	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
	36.1 mM IMD	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
	53.5 mM IMD	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
	300 mM IMD	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

Supplemental Table 2A. 96-Well setup for activity assay of the Imidazole washes of interest from the pMJ5924 purification. Activity buffer consisted of 150 mM NaCl, 50 mM Tris pH 8.0. Various protein and peptide concentrations were utilized for optimization purposes. Buffer was added first, and fluorescent peptide was added last to initiate the reaction. Total reaction volume = 100 µL. Final concentrations of: PMSF = 1000 µM, fluorescence peptide = 10 µM, max concentration of low protease = 1 µM, max concentration of high protease = 15 µM. IMD represents Imidazole.

	Label	Sample
Controls	/	Activity Buffer Only
	C	Chymotrypsin (15 µM) in Activity Buffer
	C0	Chymotrypsin (15 µM) in 10 mM IMD
	C10	Chymotrypsin (15 µM) in 39 mM IMD
	C100	Chymotrypsin (15 µM) in 300 mM IMD
pMJ5924	**1.5**	14.35 mM IMD Wash
	3.0	18.70 mM IMD Wash
	4.5	23.05 mM IMD Wash
	6.0	27.40 mM IMD Wash
	7.5	31.75 mM IMD Wash
	9.0	36.10 mM IMD Wash
	10.5	40.45 mM IMD Wash
	12.0	44.80 mM IMD Wash
	13.5	49.15 mM IMD Wash
	15.0	53.50 mM IMD Wash
	100	300.00 mM IMD Wash

	Replicate 1				Replicate 2				Replicate 3			
A	/		6.0		/		6.0		/		6.0	
B	C		7.5		C		7.5		C		7.5	
C	C0		9.0		C0		9.0		C0		9.0	
D	C10		10.5		C10		10.5		C10		10.5	
E	C100		12.0		C100		12.0		C100		12.0	
F	1.5		13.5		1.5		13.5		1.5		13.5	
G	3.0		15.0		3.0		15.0		3.0		15.0	
H	4.5		100		4.5		100		4.5		100	
	1	2	3	4	5	6	7	8	9	10	11	12

Supplemental Table 2B. 96-Well setup for activity assay of the Imidazole washes of all fractions from the pMJ5924 purification after autocatalysis. Activity buffer consisted of 150 mM NaCl, 50 mM Tris pH 8.0. Buffer was added first, and fluorescent peptide was added last to initiate the reaction. 90 µL of each sample was added to each well (samples right off the Ni-IMAC column were used). 10 µL of fluorescent peptide (final concentration 10 µM) was added to achieve a total reaction volume of 100 µL. IMD represents Imidazole.

Label	Compound (25 µM)	Protein (0.6 µM)
--	DMSO	-
-+	DMSO	+
C+	Camostat mesylate	+
N+	Nafamostat mesylate	+
C1	Compound 1	+
C2	Compound 2	+
C3	Compound 3	+
C4	Compound 4	+
C5	Compound 5	+
C6	Compound 6	+
C7	Compound 7	+
C8	Compound 8	+
C9	Compound 9	+
C10	Compound 10	+

	Replicate 1				Replicate 2				Replicate 3			
A	--	-+	C+	N+	--	-+	C+	N+	--	-+	C+	N+
B	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
C	C5	C6	C7	C8	C5	C6	C7	C8	C5	C6	C7	C8
D	C9	C10			C9	C10			C9	C10		
	1	2	3	4	5	6	7	8	9	10	11	12

Supplemental Table 3. 48-Well setup for *in vitro* compound screen. Total well volumes were 50 µL. Final concentrations of compound and protein are given in the table above. The final concentration of fluorescent peptide used was 10 µM. Activity buffer (150 mM NaCl 50 mM Tris pH 8.0) was used to top up the volume to 50 µL. A protein-in-buffer stock was made using 400 µL of pMJ5924 53.5 mM Imidazole + 1150 µL activity buffer. 38.75 µL of Buffer only was added to "--" well. 38.75 µL of protein-in-buffer stock was added to other wells. 1.25 µL of DMSO or 1.25 µL of 1 mM compound in DMSO was subsequently added. The plate was shaken at 37 ^{o}C for 10 mins before adding 10 µL of 100 µM stock peptides. To account for different reaction timings, fluorescent peptide was added starting with C10, doing 3 replicates, wait until 30 sec has passed since starting C10, then move to C9…