

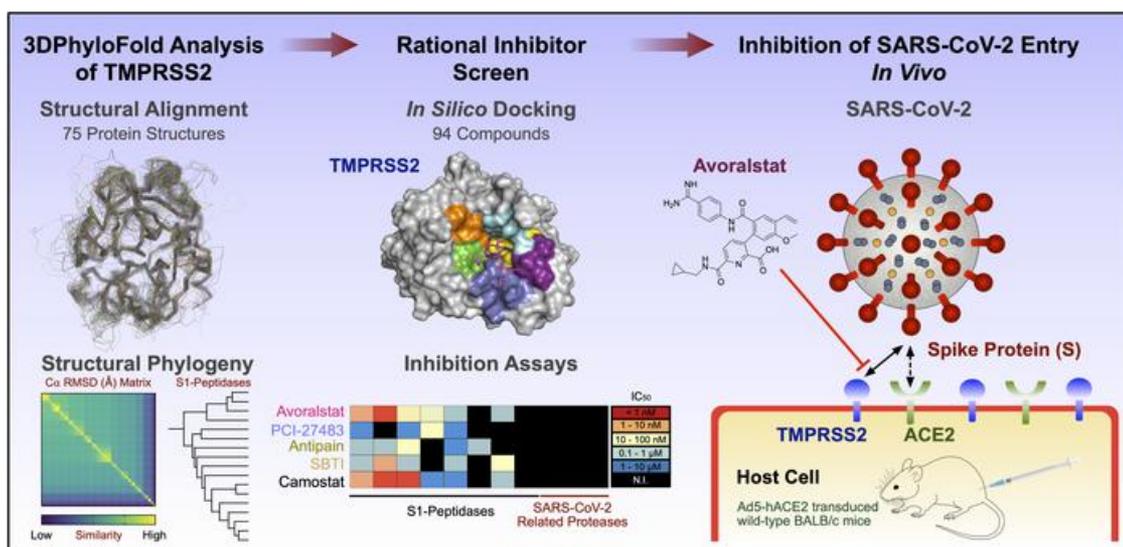
# Structure-based phylogeny identifies Avoralstat as a TMPRSS2 inhibitor that prevents SARS-CoV-2 infection in mice

Young Joo Sun, ... , Alexander G. Bassuk, Vinit B. Mahajan

*J Clin Invest.* 2021. <https://doi.org/10.1172/JCI147973>.

Research In-Press Preview COVID-19 Infectious disease

## Graphical abstract



Find the latest version:

<https://jci.me/147973/pdf>



1 **Structure-based phylogeny identifies Avoralstat as a TMPRSS2 inhibitor that prevents SARS-CoV-2**  
2 **infection in mice**

3  
4 Young Joo Sun<sup>1</sup>, Gabriel Velez<sup>1,2</sup>, Dylan Parsons<sup>1,3</sup>, Kun Li<sup>4</sup>, Miguel Ortiz<sup>4</sup>, Shaunik Sharma<sup>4</sup>, Paul B. McCray  
5 Jr. #<sup>4,5</sup>, Alexander G. Bassuk #<sup>4,6,7</sup>, Vinit B. Mahajan #<sup>1,8</sup>  
6

7 **Affiliations:**

8 <sup>1</sup>Molecular Surgery Lab, Byers Eye Institute, Department of Ophthalmology, Stanford University, Palo Alto, CA,  
9 USA, 94304

10 <sup>2</sup>Medical Scientist Training Program, University of Iowa, Iowa City, IA, USA, 52242

11 <sup>3</sup>Stanford ChEM-H Medicinal Chemistry Knowledge Center, Stanford University, Palo Alto, CA, USA, 94305

12 <sup>4</sup>Department of Pediatrics, University of Iowa, Iowa City, IA, USA, 52242

13 <sup>5</sup>Department of Microbiology and Immunology, University of Iowa, Iowa City, IA, USA, 52242

14 <sup>6</sup>Department of Neurology, University of Iowa, Iowa City, IA, USA, 52242

15 <sup>7</sup>Iowa Neuroscience Institute, University of Iowa, Iowa City, IA, USA, 52242

16 <sup>8</sup>Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA, 94304  
17

18 **# Corresponding authors:**

19 **Vinit B. Mahajan**, Byers Eye Institute, Department of Ophthalmology, Stanford University, Palo Alto, CA, USA,  
20 94304.

21 Mailing address: 1651 Page Mill Rd. - Room #2113, Palo Alto, CA, USA, 94304.

22 Phone: 650-723-6995. Fax: 650-498-1528. E-mail: [vinit.mahajan@stanford.edu](mailto:vinit.mahajan@stanford.edu)

23 **Alexander G. Bassuk**, Department of Pediatrics, Neurology, and The Iowa Neuroscience Institute (INI),  
24 University of Iowa, Iowa City, IA, USA, 52242.

25 Mailing address: 25 S. Grand Ave, Med Labs Building - Room 2040a, Iowa City, IA, USA, 52242.

26 Phone: 773-744-6338. E-mail: [alexander-bassuk@uiowa.edu](mailto:alexander-bassuk@uiowa.edu)

27 **Paul B. McCray Jr.**, Department of Pediatrics, University of Iowa, Iowa City, IA, USA, 52242.

28 Mailing address: 6320 PBDB, 169 Newton Road, Iowa City, IA, USA, 52242.

1 Phone: 319-335-6844. E-mail: [paul-mccray@uiowa.edu](mailto:paul-mccray@uiowa.edu)

2

3 The authors have declared that no conflict of interest exists.

4

## 1 **Abstract**

2 Drugs targeting host proteins can act prophylactically to reduce viral burden early in disease and limit  
3 morbidity, even with antivirals and vaccination. Transmembrane serine protease 2 (TMPRSS2) is a human  
4 protease required for SARS-CoV-2 viral entry and may represent such a target. We hypothesized that drugs  
5 selected from proteins related by their tertiary structure, rather than their primary structure, were likely to  
6 interact with TMPRSS2. We created a structure-based phylogenetic computational tool named 3DPhyloFold to  
7 systematically identify structurally similar serine proteases with known therapeutic inhibitors and demonstrated  
8 effective inhibition of SARS-CoV-2 infection in vitro and in vivo. Several candidate compounds, Avoralstat, PCI-  
9 27483, Antipain, and Soybean-Trypsin-Inhibitor, inhibited TMPRSS2 in biochemical and cell infection assays.  
10 Avoralstat, a clinically tested Kallikrein-related B1 inhibitor, inhibited SARS-CoV-2 entry and replication in  
11 human airway epithelial cells. In an in vivo proof of principle, Avoralstat significantly reduced lung tissue titers  
12 and mitigated weight-loss when administered prophylactically to SARS-CoV-2 susceptible mice indicating its  
13 potential to be repositioned for COVID-19 prophylaxis in humans.

## **Introduction**

The coronavirus disease 2019 (COVID-19), caused by SARS-coronavirus 2 (SARS-CoV-2), has spread globally causing over 2,540,000 deaths (World Health Organization, WHO). Prophylactic and early-stage therapies are needed for high-risk populations. Even with vaccines, adjunctive therapies that mitigate viral entry or replication may attenuate disease severity and reduce viral spread by asymptomatic and early-stage patients. In response to the urgent need for therapeutics, there is investigation into repositioning existing drugs towards viral proteins (e.g., Remdesivir). An alternative strategy is to target human proteins utilized by viruses with small molecules. This approach can work synergistically with vaccination and may be especially important for individuals where vaccination is contraindicated or deferred, to prevent viral transmission that may occur after vaccination, for front-line workers exposed to repeated high viral load, in countries where sophisticated vaccine delivery and storage is unavailable, and potentially for protection from viral mutation and other viruses using similar host mechanisms.

Transmembrane serine protease 2 (TMPRSS2) is a human serine protease that is a priming protease for the spike glycoprotein found on the surface of all coronaviruses (1, 2). TMPRSS2's S1-peptidase domain is required for SARS-CoV-2 entry into host epithelial cells in the upper- and lower-respiratory tract (3, 4), but it is not necessary for development or homeostasis in mice, making it an attractive drug target (5). It is yet to be determined whether TMPRSS2 inhibition mitigates SARS-CoV-2 infection in vivo. Camostat, a serine protease inhibitor originally developed for acute pancreatitis, inhibits TMPRSS2 in vitro and is in clinical trial for the treatment of COVID-19 (NCT04321096) (6, 7). However, Camostat's plasma half-life is less than one minute, and its efficacy for COVID-19 is yet to be determined (4, 8, 9). Thus, identification of inhibitors targeting TMPRSS2 with improved pharmacokinetic properties remains important.

Conventional methods for identifying drug candidates typically employ high-throughput screening (HTS) or in silico screening using compound libraries previously tested in humans (10). In silico screening for TMPRSS2 has been challenging as there is no high-resolution TMPRSS2 molecular structure. Although HTS methods can rapidly screen thousands of compounds, there are certain limitations. HTS methods utilize only a few,

1 generalized experimental parameters with technical limitations, such as narrow dose range and experimental  
2 conditions, which may not account for the unique features of each compound. This can lead to false positives  
3 and negatives. While false positives are filtered out in subsequent experiments, false negatives may overlook  
4 valuable compounds. Because HTS uses shotgun rather than hypothesis-driven approaches, it may be difficult  
5 to ascertain the mechanism-of-action, and this may slow the downstream development of candidate drugs into  
6 human therapies. Hypothesis-driven screening methods, utilizing protein structures and a limited number of  
7 compounds, remains a valuable and complementary strategy for drug-repositioning. One approach to rational-  
8 based drug repositioning is to identify proteins with preexisting drugs that are similar to the target protein.

## 1 **Results**

2 ***In silico drug repositioning by 3DPhyloFold.*** To identify drug repositioning candidates, we created a  
3 computational/hypothesis-driven drug repurposing method called 3DPhyloFold that identifies structurally  
4 similar proteins to rationally select candidate inhibitors. A comprehensive phylogenetic analysis of 600 S1-  
5 peptidases sequentially related to the TMPRSS2 S1-peptidase domain (TMPRSS2-S1P) showed TMPRSS2-  
6 S1P clustered closely to canonical TMPRSS family members, like Hepsin, as well as proteases outside of the  
7 TMPRSS subfamily: Coagulation Factor XI and Kallikrein-related B1 (KLKB1; Supplemental Figure 1A-B).  
8 TMPRSS2-S1P was closest to Hepsin, and a homology-based model was generated (Supplemental Figure  
9 1C). Next, 3DPhyloFold determined the 3D relationship of TMPRSS2-S1P to other S1-peptidase structures  
10 (Figure 1; Supplemental Table 2). Using structural quality metrics (see Methods), 74 S1-peptidases and  
11 TMPRSS2-S1P were aligned by conventional sequence phylogenetic analysis. TMPRSS2-S1P clustered  
12 closely with KLKB1, Factor XI, and Complement Factor I (CFAI), while the Kallikrein- and Trypsin-like clades  
13 clustered further away (Figure 2A). In 3DPhyloFold, pairwise structural comparisons of the representative  
14 tertiary structures were used to calculate a structural dissimilarity matrix (SDM) based on the root mean square  
15 deviation between protein alpha-carbons ( $C\alpha$  RMSD; Figure 1). A structure-based phylogenetic tree was then  
16 generated from the SDM (Figure 2B). Clustering of the structure-based tree was distinct to that of the  
17 sequence-based tree. Proteins close in the primary sequence analysis (e.g., CFAI and CTRB1) were much  
18 farther away in the 3DPhyloFold structure-tree (Supplemental Figure 2G). Although distant in the sequence-  
19 based tree, the Trypsin-like clade and Factor VII moved much closer to TMPRSS2-S1P in the structure-based  
20 tree (Figure 2B). This suggested that, while divergent in sequence, TMPRSS2-S1P adopts a three-dimensional  
21 fold closer to Trypsin and Factor VII. We prioritized the six S1-peptidases with the highest structural similarity  
22 to TMPRSS2-S1P: Hepsin, Acrosin, Trypsin, Factor VII, Factor XI, and KLKB1.

23  
24 Using these six proteases, we sought known small-molecules and peptidomimetic inhibitors containing a  
25 guanidine, or structurally related groups (see Methods), since S1-peptidases are inhibited by compounds  
26 containing a 4-amindinobenzylamide moiety. The 4-amindinobenzylamide moiety mimics the key specificity  
27 feature of S1-peptidase substrates: an arginine residue at the first N-terminal residue of the substrates'

1 cleavage site (P1) that forms a strong interaction with an aspartate in the corresponding S1 subpocket of the  
2 peptidase (Figure 3A) (11). This search curated ninety experimental compounds and four small molecules  
3 previously tested in human clinical trials, which docked well to TMPRSS2-S1P (Figure 3B-C; Supplemental  
4 Table 3-4; Supplemental Figure 3). In addition, 3DPhyloFold analysis revealed a natural Trypsin-inhibiting  
5 protein based on the structure of porcine-Trypsin with Soybean-Trypsin-Inhibitor (SBTI; PDBID 1AVW; Figure  
6 3C). Since the porcine Trypsin binding pocket was similar to that of TMPRSS2-S1P (~68% sequence identity),  
7 we modeled the TMPRSS2-S1P/SBTI complex and identified a conserved inhibitory motif (PYRIRF) with  
8 favorable docking interactions, suggesting SBTI might bind and inhibit TMPRSS2-S1P (Supplemental Figure 4;  
9 Supplemental Table 4-5).

10  
11 **Biochemical evaluation of 3DPhyloFold inhibitors.** We focused on the inhibitory potential of human drugs  
12 available for repositioning, including Avoralstat, PCI-27483, and Antipain, along with SBTI. A biochemical  
13 inhibition assay using the extracellular region of purified recombinant TMPRSS2 (residues 106 – 492) was  
14 utilized to test inhibition by these compounds (7). The rank order potency against TMPRSS2 was Avoralstat  
15 ( $IC_{50} = 2.73 \pm 0.19$  nM), SBTI ( $IC_{50} = 121 \pm 4$  nM), Antipain ( $IC_{50} = 748 \pm 63$  nM), and PCI-27483 ( $IC_{50} = 1.41 \pm$   
16  $0.04$   $\mu$ M; Figure 3D). Inhibition by Avoralstat was as potent as Camostat ( $IC_{50} = 1.01 \pm 0.10$  nM), which targets  
17 TMPRSS2 and is currently under clinical investigation for SARS-CoV-2 treatment (7). We further explored the  
18 selectivity profile of the four 3DPhyloFold inhibitors and Camostat as a positive control by testing them against  
19 six S1-peptidases identified in 3DPhyloFold (i.e., KLKB1, Trypsin, Factor VIIa, Factor Xa, KLK1, and KLK7),  
20 three other proteases involved in SARS-CoV-2 infection (i.e., Furin, Mpro, and PLpro), and a negative control  
21 Papain. As expected, each 3DPhyloFold compound displayed potent inhibition of its original target-proteases,  
22 and there was no inhibition of non-S1-proteases. Strikingly, Avoralstat was more than 18-fold more selective  
23 towards TMPRSS2 than other S1-proteases (Figure 3E; Supplemental Table 6). Camostat was not as  
24 selective as Avoralstat. We further characterized Avoralstat's specificity by expanding the protease screen to  
25 include an additional 60 structurally distant proteases, including MMPs, Caspases, Cathepsins, and cysteine-  
26 or aspartyl-proteases (4, 12-14). Avoralstat displayed potent inhibition of other S1-proteases consistent with  
27 their proximity to TMPRSS2 in the 3DPhyloFold tree, while displaying no inhibition of non-S1-proteases (Figure

3F; Supplemental Table 7), suggesting inhibition was specific and not due to protein aggregation effects. Notably, Avoralstat inhibited several proteins that were structurally similar to TMPRSS2, including Factor VIIa and Tryptase b2, despite being distant in primary sequence (Figure 4A). Conversely, Avoralstat was less effective at inhibiting proteases that clustered further from TMPRSS2 on the 3DPhyloFold tree: Chymotrypsin ( $IC_{50} = >1 \mu M$ ) and Elastase, ( $IC_{50} = >1 \mu M$ ; Figure 4B), despite their proximity in the primary sequence phylogenetic tree. To further confirm that the compounds target the protease domain of TMPRSS2, we tested inhibition using catalytically-active recombinant S1P domain (residues 252 – 489; Supplemental Figure 5) and found similar inhibition trends (Figure 4C). Taken together, these results suggested Avoralstat was highly selective for TMPRSS2, consistent with the predictions by structural phylogenetic analysis.

**Cellular evaluation of 3DPhyloFold inhibitors.** Inhibition of full-length TMPRSS2 (TMPRSS2-FL) proteolytic activity was then tested in cells. TMPRSS2-FL contains an autoproteolysis motif (residues 252-257), which is subject to cleavage and can be used to probe the activity of TMPRSS2 in cells (15). Cells were transfected with either wild-type (WT) or loss of function TMPRSS2-S441A mutant (Figure 5A-B; Supplemental Figure 6). Compared to the inactive S441A mutant, TMPRSS2-WT showed reduced signal by immunoblot as previously reported (Figure 5B) (15). Inhibitor treatment prevented TMPRSS2-FL autoproteolysis and significantly increased the TMPRSS2-FL band intensity (Figure 5B).

To test if the compounds specifically inhibited viral entry, we used vesicular stomatitis virus (VSV)-based pseudovirions bearing the SARS-CoV-2 spike glycoprotein and a luciferase reporter system. Human Calu-3 2B4 airway cells were incubated with Camostat, Avoralstat, PCI-27483, Antipain, and SBTI. Pseudovirions harboring the pantropic VSV glycoprotein (VSV-G) served as controls since they transduce cells independent of TMPRSS2 (16). Indeed, no compounds prevented VSV-G pseudovirus entry, since the luciferase signal remained constant (Figure 5C). Camostat inhibited SARS-CoV-2 pseudovirus entry ( $EC_{50} = 0.7 \pm 0.2 \mu M$ ), and Avoralstat displayed similar inhibition ( $EC_{50} = 2.8 \pm 0.7 \mu M$ ). PCI-27483, Antipain, and SBTI displayed modest inhibition but were too weak to determine reliable  $EC_{50}$  values (Figure 5D). Next, inhibition of authentic SARS-CoV-2 was tested in Calu-3 2B4 cells by measuring viral genomes. Camostat, Avoralstat, and Antipain

1 significantly reduced SARS-CoV-2 replication (the amount of nucleocapsid gene [viral RNA] compared to  
2 vehicle, respectively;  $p < 0.0001$ ). PCI-27483 and SBTI showed less inhibition (Figure 5E). A dose-response of  
3 Camostat and Avoralstat displayed significant reduction in SARS-CoV-2 infection beginning at 100 nM.  
4 Camostat and Avoralstat caused more than a ten-fold decrease in viral RNA signal with a 1  $\mu$ M dose (Figure  
5 5F). SARS-CoV-2 showed more sensitivity to Avoralstat and Camostat than MERS-CoV, another coronavirus  
6 that also uses TMPRSS2 to facilitate entry (Supplemental Figure 7) (6).

7  
8 ***Avoralstat inhibits SARS-CoV-2 entry in vivo.*** No therapy targeting host sensitizing proteases has been  
9 validated in an in vivo model of COVID-19. There is no known viral infection dose or animal model that fully  
10 recapitulates human disease, so the critical in vivo measure for testing prophylactic efficacy is the reduction of  
11 viral load. Using a mouse model of SARS-CoV-2 lung infection (Ad5-hACE2 transduced wild-type BALB/c  
12 mice) (17), we compared the efficacy of Avoralstat and Camostat in modifying SARS-CoV-2 infection. Cohorts  
13 of mice were infected intranasally with either  $3 \times 10^3$  or  $1 \times 10^5$  PFU of SARS-CoV-2, respectively. Mice were  
14 treated with Avoralstat, Camostat (30 mg/kg intraperitoneal injection), or vehicle (DMSO). Lungs were  
15 harvested 1 day after infection and viral titers measured by plaque assay. Both Avoralstat and Camostat  
16 significantly reduced the lung tissue titers in both cohorts (Figure 6A-B). In a third cohort of mice, Avoralstat or  
17 Camostat was administered 4 hours prior and 4 hours after a  $1 \times 10^5$  PFU of SARS-CoV-2 intranasal  
18 challenge. Mice were given twice daily drug doses for three days post infection (dpi). Lungs harvested at 5-dpi  
19 demonstrated significant reduction in viral titers by both drugs. Strikingly, the lung tissue virus titers were below  
20 the limit of detection in 3 of 4 Avoralstat-treated mice (Figure 6C). Changes in weight, indicating the severity of  
21 illness, was monitored. Beginning at 4-dpi, there was significant weight loss in the vehicle- and Camostat-  
22 treated mice, while the weight of the Avoralstat-treated mice remained relatively constant, suggesting a  
23 significant protective effect (Figure 6D). Although there was significant weight-loss in Camostat group, an  
24 Avoralstat therapeutic effect was observed later at 7-dpi compared to the vehicle-treated groups (Figure 6D). In  
25 a fourth cohort of mice, a biological dose-response was strongly supported after we further increased the  
26 SARS-CoV-2 challenge dose to  $1 \times 10^6$  PFU. Avoralstat or Camostat were administered 4 hours prior and 4  
27 hours after a SARS-CoV-2 intranasal challenge. Mice were then given two drug doses daily for 3-dpi. At the

1 higher challenge dose, an early viral titer reduction was not observed as seen in lower titers (i.e.,  $3 \times 10^3$  or  $1 \times$   
2  $10^5$  PFU). Yet a significant decrease of viral titer was observed at 4-dpi for both Avoralstat- and Camostat-  
3 treated groups (Figure 6E). Moreover, Avoralstat still showed a significant weight rescue effect beginning from  
4 7-dpi while Camostat did not show any rescue effect compared to the vehicle-treated group (Figure 6F). Thus,  
5 the inhibitory effect of Avoralstat observed in biochemical and cell assays, extended to prophylactic treatment  
6 of mice infected with escalating doses of SARS-CoV-2.

## 1 Discussion

2 Drug repositioning is an important strategy to address human disease at a faster pace than conventional drug  
3 development, especially in the setting of a global viral pandemic. Avoralstat, a clinically tested oral KLKB1  
4 inhibitor evaluated for the treatment of hereditary angioedema, successfully inhibited SARS-CoV-2 infection  
5 and illness in mice. Avoralstat is orally bioavailable, which could facilitate prophylactic administration to people  
6 at high risk for COVID-19, particularly where specialized vaccine transport, cold storage, and medically skilled  
7 delivery staff are not available. Avoralstat possess a favorable plasma half-life of 12-31 hours, compared to the  
8 short half-life of Camostat due to an easily cleavable ester bond, giving it a terminal half-life of roughly 1 hour  
9 (18-21). It is possible that the observed efficacy of Avoralstat compared to Camostat in our *in vivo* study is due  
10 to its longer plasma half-life. If Avoralstat administered intraperitoneally to mice was 100% available in plasma,  
11 a 30 mg/kg of dose is roughly 80 nM in the water volume of a mouse (average volume percentile of water in  
12 mice is 73%). This molar concentration is about 29-fold higher compared to the *in vitro* IC<sub>50</sub>. Although  
13 bioavailability was not directly measured, the mouse Avoralstat dose of 30 mg/kg translates to a Maximum  
14 Recommended Starting Dose (MRSD) of 0.224 mg/kg in humans according to FDA guidelines, which is  
15 markedly lower than the dose previously administered in clinical trials (up to 1500 mg per day). Our  
16 biochemical assay data for Avoralstat's protease selectivity showed that there are 10 proteases that have  
17 similar or lower Avoralstat IC<sub>50</sub> values than for TMPRSS2 (Figure 4A), yet Avoralstat showed relatively minor  
18 and manageable side effects in humans. No grade 3 adverse events were observed in phase 1 through phase  
19 3 clinical trials for Avoralstat and serious adverse events were no more prevalent in treatment compared to  
20 placebo groups (20, 21). It is possible that the observed efficacy of Avoralstat compared to Camostat in our *in*  
21 *vivo* study is due to its longer plasma half-life. Avoralstat has relatively minor and manageable side effects. The  
22 doses we tested in mice through intraperitoneal administration correlate with a substantially lower dose than  
23 previous oral-administration to humans in clinical trials (22), suggesting an appropriate dose of Avoralstat for  
24 treating COVID-19 may be achievable with reasonable safety.

25  
26 The application of a targeted structure-based phylogeny approach allowed us to identify and rationally prioritize  
27 several candidate TMPRSS2 inhibitors not considered by other drug repositioning strategies: 3DPhyloFold

1 pointed to closely related proteins missed by primary-sequence comparisons, supporting a mechanism-based,  
2 hypothesis-driven selection of curated inhibitor candidates. Many of the small molecules tested in this study  
3 could be further developed for alternative routes of administration and for more potency and selectivity against  
4 TMPRSS2. The SBTI protein might serve as a cheap and natural source TMPRSS2 inhibitor, since it has been  
5 widely used in biomedical research (23). Interestingly, Avoralstat and PCI-27483 were both represented in high  
6 throughput screens but may have been missed due to the lack of sufficient testing conditions (e.g., dosage or  
7 cell line) (24).

8  
9 Our in vivo studies underscore that targeting TMPRSS2 is a tenable strategy for COVID-19 treatment. A  
10 reduction of viral load achieved by an alternative mechanism to that of vaccination could act synergistically to  
11 reduce illness and transmission. In addition, TMPRSS2 is implicated in the cleavage of the envelope-  
12 glycoproteins of many other viruses, including SARS-CoV, MERS-CoV, HCoV-229E, HCoV-OC43, HCoV-  
13 HKU1, and HCoV-NL63; Influenza viruses; Parainfluenza viruses; and human Metapneumovirus (8). Thus,  
14 targeting this host machinery could be applied as a long-term strategy for future zoonotic coronaviruses and  
15 other respiratory viruses. This may be especially important if targeting viral proteins are only partially effective,  
16 natural infection does not confer long-lasting immunity, and combination therapies are needed to reduce the  
17 likelihood of resistance (25).

## 1 **Methods**

2 ***Experimental model and subject details: mice, virus, and cells.*** Specific pathogen-free 6-week-old male  
3 and female BALB/c mice and were purchased from Envigo and maintained in the Animal Care Facilities at the  
4 University of Iowa. The human serotype 5 adenoviral vector expressing human ACE2 under the control of the  
5 CMV promoter was previously described (VVC-McCray-7580; University of Iowa Viral Vector Core) (17). The  
6 SARS-CoV-2 strains (SARS-Related Coronavirus 2 Isolate USA-WA1/2020) were obtained from BEI (Cat. #  
7 NR-52281) and Calu-3 2B4 cells (obtained from the Perlman Laboratory, University of Iowa). HEK-293T cells  
8 (ATCC® Cat. # CRL-3216) were obtained from the Viral Vector Core Facility at the University of Iowa. pVSV-  
9  $\Delta$ G-Luc was previously described (16). Calu-3 2B4 cells were grown in MEM (GIBCO, Grand Island, NY)  
10 supplemented with 20% FBS.

11  
12 ***Database search and sequence alignment.*** We first searched the UniProt database for reviewed entries  
13 denoted as transmembrane serine proteases (containing an S1-peptidase domain). This initial search yielded 9  
14 manually curated sequences. A seed multiple sequence alignment (MSA) of S1-peptidase domains was then  
15 constructed using MAFFT v7 (alignment strategy: FFT-NS-1) (26). Using HMMER-3.1 and the seed alignment,  
16 we produced an HMM profile and used it to broaden the search against the UniProt database (search  
17 restricted to reviewed sequences) (27). This search yielded a total of 828 S1-peptidase sequences. We  
18 discarded fragmented sequences (<200 amino acids) that appeared too short to truly represent the S1-  
19 peptidase fold and redundant proteins were further filtered using CD-HIT v4 (100% threshold) (28). This  
20 resulted in a pool of 742 proteins that were aligned using MAFFT v7 (alignment strategy FFT-NS-2) (26).  
21 Sequences producing many gaps in the alignment were removed using MaxAlign, resulting in 600 S1-  
22 peptidase sequences (29).

23  
24 ***Phylogenetic tree reconstruction.*** We used the IQ-TREE-1.6.2 algorithm to generate a maximum likelihood  
25 tree of the 600 S1-peptidase sequences (30). The IQ-TREE model finder tool was used to determine the best  
26 substitution model to fit the data. The Whelan & Goldman (WAG) substitution model was determined to be the

1 best fit to the data. Bootstrap analysis was performed using the 'ultra-fast' method in IQ-TREE-1.6.2 with 1,000  
2 replicas.

3  
4 **Structural modeling of TMPRSS2-S1P.** Briefly, a BLAST search of human TMPRSS2-S1P against the  
5 Protein Data Bank (PDB) returned the structure of human Hepsin (PDB 1Z8G) as the top hit. Other close  
6 matches were KLKB1 (PDB 6ESO), Plasminogen (PDB 4DUR), and Prostatin (PDB 3E16). A TMPRSS2-S1P  
7 model was generated with the Hepsin template (41% sequence identity) using Phyre2, MODELLER, and  
8 SWISS-Model. The models were in agreement and aligned well with minor variations in surface-exposed loop  
9 regions. The TMPRSS2-S1P model was then analyzed by ConSurf as previously described (31). The 600  
10 sequences from our sequence-based phylogenetic analysis underwent MSA using MAFFT and conservation  
11 scores were calculated using the Bayesian method option in ConSurf. The TMPRSS2-S1P binding pocket was  
12 inferred by comparison to the structure of Hepsin bound to a peptidomimetic inhibitor (PDB 1Z8G) in PyMOL  
13 (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

14  
15 **Structure-based phylogenetic analysis.** There are over 2,000 structures of S1-peptidase domains  
16 represented in the PDB. We therefore searched the Pfam database for structures of mammalian peptidases  
17 and selected 74 representative structures (representing the wild-type protein) with an atomic resolution 3.2 Å  
18 or better (Supplemental Table 2) (32). One structure per unique protein, fitting the above criteria, was selected.  
19 Structures (with reflection data deposited in the PDB) were evaluated by their reported global validation metrics  
20 in PDB-REDO (33). Re-refined structural models were used for further analysis. Structures were superimposed  
21 using PyMOL to calculate the pairwise root mean square deviation (RMSD) between protein alpha carbon  
22 atoms ( $C\alpha$ ). A structural dissimilarity matrix (SDM) was constructed using the  $C\alpha$  RMSD values in order to  
23 generate a phylogenetic tree as previously described (31). To expedite the pairwise alignment process, we  
24 developed a Python-based script (named 3DPhyloFold) to perform the pairwise alignment of protein structures  
25 and generate an SDM. The phylogenetic tree was constructed using the UPGMA (Unweighted Pair Group  
26 Method with Arithmetic Mean) method in MEGAX software as previously described (31). For comparison, the  
27 sequences from the corresponding structures were also analyzed by sequence-based phylogeny. The 75 S1-

1 peptidase sequences were aligned with MAFFT v7 (26) and analyzed in IQ-TREE-1.6.2. (30). The Jones-  
2 Taylor-Thornton (JTT) substitution model was determined to be the best fit to the data. Bootstrap analysis was  
3 performed in IQ-TREE-1.6.2 (1,000 replicas). A TMPRSS2-S1P structure similarity score for each analyzed  
4 protease was calculated by dividing the pairwise sequence identity (to TMPRSS2-S1P) by the C $\alpha$  RMSD of the  
5 pairwise alignment (Supplemental Figure 2).

6  
7 **Database search for S1-peptidase inhibitors.** We first searched for inhibitors designed for the related  
8 proteins. We filtered for cases where a strong Structure-Activity-Relationship (SAR) between the ligand and  
9 protein was studied, where applicable. We discarded studies where the inhibitors displayed low potency,  
10 focusing on groups of inhibitors that displayed sub-micromolar inhibition for their intended protein. We focused  
11 on cases where the inhibitors studied contained a guanidine, or guanidine-like, functional group to interact with  
12 the S1 specificity pocket. Inhibitors were then prepared and docked against our TMPRSS2-S1P model.

13  
14 **In silico docking calculations.** Published crystal structures of inhibitor-bound Trypsin-3 (PDB 1H4W), KLKB1  
15 (PDB 6O1S), and Factor VII (PDB 1W7X) were loaded into Maestro software (Schrödinger Release 2019-3).  
16 The TMPRSS2-S1P model described above was used. The protein preparation wizard was used to prepare  
17 the proteins for docking and simulations. The default parameters were used for the optimization of hydrogen-  
18 bond assignment (sampling of water orientations and use of pH 7.0). Waters molecules beyond 3 Å of  
19 heteroatoms or with fewer than three hydrogen bonds to non-waters were removed. Restrained energy  
20 minimization was applied using the OPLS3e force field. Prepared protein systems were further checked by  
21 Ramachandran plots, ensuring there were no steric clashes. To generate receptor grids for small molecule  
22 docking, the co-crystallized ligand was selected as the grid-defining ligand for each system. Default van der  
23 Waals radius scaling parameters were used (scaling factor of 1, partial charge cutoff of 0.25). For peptides, the  
24 grid size was made suitable for peptides to be docked. Default van der Waals radius scaling parameters were  
25 used (scaling factor of 1, partial charge cutoff of 0.25). For docking of the ligands into the various prepared  
26 proteins, the 3D structure was loaded into Maestro. Ligprep was used to prepare the ligands (by generating  
27 possible states at pH 7.0  $\pm$  2.0 and retaining the specified stereochemical properties). The prepared small

1 molecule ligands and peptide fragments were then docked using the most stringent docking mode (extra  
2 precision, “XP”) of Glide. Parameters and output files for the Glide runs can be found in Mendeley Data under  
3 the dataset identifier (DOI): 10.17632/h3pmycddwc.1.

4  
5 **Docking Soybean Trypsin Inhibitor to TMPRSS2.** The HADDOCK 2.4 online docking tool was used to  
6 generate TMPRSS2-S1P/SBTI complex structure model (34). The TMPRSS2-S1P homology model and the  
7 SBTI structure (PDB 1AVW) were used for docking. To define the potential interaction surface between  
8 TMPRSS2 and SBTI, the TMPRSS2-S1P homology model was superimposed to the wild boar trypsin structure  
9 in complex with SBTI (PDB 1AVW) using PyMOL. The following residues of SBTI were designated as active  
10 residues: 501-502, 510, 512- 514, 560-572, and 616-617. The overall C $\alpha$  RMSD between the two models was  
11 0.54 Å. SBTI was also docked to porcine Trypsin (PDB 1AVW), human Factor VII (PDB 1W7X), and human  
12 KLKB1 (PDB 6O1S). The HADDOCK scores represent the average score of the best cluster. The parameters  
13 and output files for the HADDOCK run can be found in Mendeley Data under the dataset identifier (DOI):  
14 10.17632/h3pmycddwc.1.

15  
16 **Protease activity array.** Avoralstat, SBTI, PCI-27483, and Antipain were assessed for inhibition against  
17 TMPRSS2 and a panel of recombinant proteases by commercial services from Reaction Biology Corp. The  
18 Reaction Biology Corp profile tested in a 10-dose IC<sub>50</sub> with, in triplicate, a 4-fold serial dilution starting at 10  $\mu$ M  
19 against 11 proteases in Figure 3E and a 3-fold serial dilution starting at 10  $\mu$ M against 70 proteases in Figure  
20 4A. Compounds exhibit no fluorescent background that could interfere with the assay. The protease activities  
21 were monitored as a time-course measurement of the increase in fluorescence signal from fluorescently  
22 labeled peptide substrate, and initial linear portion of slope (signal/min) was analyzed.

23  
24 **TMPRSS2-S1P expression and purification.** The human TMPRSS2-S1P sequence (residues 252 - 489) was  
25 cloned into a pET28a vector with a N-terminal 6x-His tag. Plasmids were amplified and isolated from DH5 $\alpha$   
26 cells and transformed into *E. coli* BL21 (DE3). BL21 cells expressing TMPRSS2-S1P were induced with 0.5  
27 mM IPTG. Cell pellets were resuspended in 35 to 50 mL of lysis buffer (50 mM Tris, 150 mM NaCl, 20 mM

1 Imidazole pH 8.0, one tablet of EDTA-free protease inhibitor [Roche; Product # COEDTAF-RO], DNaseI  
2 [Roche; Product #11284932001]) and lysed and centrifuged for 30 minutes at 18,000 x g at 4 °C. Pellets were  
3 denatured (50 mM Tris, 150 mM NaCl, 6 M Guanidinium Chloride, 1 M L-Arginine, 2 mM DTT pH 8.0),  
4 resuspended, and filtered with 0.22 µm filter. Refolding buffer-1 (50 mM Tris, 150 mM NaCl, 2 M Guanidinium  
5 Chloride, 1 M L-Arginine pH 8.0) was applied to SnakeSkin Dialysis Tubing (10,000 MWCO; Thermo  
6 Scientific™) and underwent refolding by dialyzing in 2 L of refolding buffer-1 at 4 °C. After the over-night  
7 refolding, the sample was filtered with 0.22 µm filter to remove aggregates and went through another step of  
8 dialysis in 2 L of refolding buffer-2 (50 mM Tris, 150 mM NaCl, 250 mM L-Arginine pH 8.0) for 1.5 hours at  
9 room-temperature. Sample was concentrated with a 10 kDa NMWL spin concentrator and passed over a  
10 HiLoad® 16/600 Superdex® 200 pg (GE Healthcare, Cat. # 28-9893-35) size-exclusion (SEC) column  
11 connected to an ÄTKA™ pure fast protein liquid chromatography (FPLC) system (GE Healthcare Inc.). The  
12 column was equilibrated with SEC buffer (50 mM Tris, 150 mM NaCl, pH 8.0). The final purity of recombinant  
13 TMPRSS2-S1P used for in vitro assays were >95% (Supplemental Figure 5A).

14  
15 **Measurement of TMPRSS2-S1P activity.** TMPRSS2-S1P proteolytic activity was confirmed by hydrolysis of  
16 the synthetic urokinase substrate, Cbz-GGR-AMC (Echelon Biosciences; Product #869-25). An enzyme  
17 titration in the presence of 50 µM Cbz-GGR-AMC revealed that maximal TMPRSS2-S1P activity occurred at  
18 high nanomolar (250 – 500 nM) protein concentrations (data not shown). The remaining assays were  
19 performed as followed: Briefly, 250 nM of purified TMPRSS2-S1P was added to a reaction buffer containing 50  
20 mM Tris-HCl (pH 8.0), and 150 mM NaCl in black-bottom 96-well plates (100 µL per reaction). Inhibition  
21 experiments were carried out in the presence of 50 µM Cbz-GGR-AMC in the presence 10 to 500 µM  
22 compound: Camostat (Sigma-Aldrich; Cat. # SML0057), Avoralstat (MedChemExpress; Cat. # HY-16735),  
23 PCI-27483 (Cayman Chemical; Item #21334), Antipain (Sigma-Aldrich; Product #A6191), Leupeptin (Sigma-  
24 Aldrich; Product #L2884), MDL-28170 (Sigma-Aldrich; Product #M6690), Ritonavir (Sigma-Aldrich; Product  
25 #SML0491), or 5% DMSO (as a negative control). DMSO caused SBTI (Roche; Product #10109886001) to  
26 precipitate out of solution (unpublished observation). Inhibition experiments with SBTI (2 to 150 µM) were  
27 therefore performed in the absence of DMSO. Reactions were run at 37 °C for 30 minutes on a fluorimetric

1 plate reader (Tecan Spark, Männedorf Switzerland). Proteolytic activity was measured as change in raw  
2 fluorescence units ( $\Delta$ RFU;  $\lambda_{exc} = 373$  nm,  $\lambda_{em} = 455$  nm) at 30-second intervals. All experiments were  
3 performed in triplicate. The initial velocity (RFU/sec) of the reaction was measured by calculating the slope of  
4 the fluorescence data from the first three minutes. Kinetic parameters were then calculated by direct fitting to  
5 the Michaelis-Menten or Hill equation in GraphPad Prism 8 (GraphPad, San Diego, CA). There was no activity  
6 as expected with the cysteine protease substrate sLY-AMC (Bachem; Product #4002047; negative control;  
7 Supplemental Figure 5).

8  
9 ***TMPRSS2 autoproteolysis assay.*** HEK-293T cells (ATCC<sup>®</sup> Cat. # CRL-3216) were obtained from the Viral  
10 Vector Core Facility at the University of Iowa. Cells were grown in Dulbecco modified Eagle medium (DMEM)  
11 supplemented with 5% fetal bovine serum (Gibco), penicillin and streptomycin (Gibco, WT15140-122) and  
12 were maintained in a humidified atmosphere of 5% CO<sub>2</sub> at 37 °C. Plasmid pEGFPN1 was obtained from  
13 Clontech. TMPRSS2-FL cDNA (pcDNA3.1-SARS-2-S-C9; obtained from the Gallagher Laboratory, Loyola  
14 University Medical Center, Illinois). Briefly, TMPRSS2-FL cDNA, containing a C-terminal anti-FLAG epitope  
15 tag, were amplified with PCR using pCMV-Sport6-TMPRSS2 template. The amplicates were cloned into  
16 pCAGGS.MCS via SacI and XhoI sites. The enzymatically-inactive pCAGGS-TMPRSS2(S441A)FLAG mutant  
17 cDNA was generated using QuickChange Site-Directed Mutagenesis Kit per manufacturer instructions (Agilent  
18 Technologies). Transient transfections of HEK-293T cells were performed using PolyFect transfection reagent  
19 per manufacturer instructions (Qiagen). For transfection, 2  $\mu$ g of each plasmid (GFP [served as negative  
20 control], TMPRSS2 WT and S411A mutant) were dissolved in serum free media. PolyFect (20  $\mu$ L) was added  
21 to the DNA solution followed by 10-minute incubation at room temperature. Growth media (0.6 mL) was then  
22 added to the reaction tubes and the transfection mix was immediately added onto the cells. 24 hours post-  
23 transfection, cell lysates were prepared using HNB buffer containing 0.1% protease inhibitor (Sigma-Aldrich,  
24 #P2714), incubated on ice for 20 minutes and centrifuged at 2,000 x g for 10 minutes. Supernatants were  
25 collected and protein concentration determined by DC protein assay reagent kit (BioRad). After separation by  
26 SDS-PAGE (4 to 12% Bis-Tris gradient gel), proteins were transferred to a PVDF membrane and blocked for  
27 1-hr at room temperature using 5% nonfat dry milk in TBST. Membranes were probed with mouse monoclonal

1 anti-Flag antibody (1:1,000; Sigma-Aldrich; Cat. #F3165) for 16 hours at 4° C. Blots were then washed three  
2 times with TBST (10 minutes/wash) and subsequently incubated with immunoglobulin-G labelled with  
3 horseradish peroxidase conjugated secondary anti-mouse antibody (1:5,000; Thermo Scientific™; Cat.  
4 #31432). Proteins were visualized by SuperSignal™ West Pico PLUS chemiluminescence reagent on a  
5 MyECL imager (Thermo Scientific™). Membranes were re-probed with  $\beta$ -actin (1:5000; Sigma-Aldrich, Cat.  
6 #A2228) as a loading control. The TMPRSS2-FL band intensity of each lane was normalized using the band  
7 intensity of corresponding  $\beta$ -actin loading control. Then, the normalized intensity of each lane was converted to  
8 the relative band intensity by comparison to the normalized band intensity of TMPRSS2-S441A in the same  
9 gel.

11 ***Pseudovirus transduction assay.*** HEK-293T cells were transfected to express either the SARS-CoV-2 spike  
12 protein (with the cytoplasmic tail removed; residues 1 - 1255) or the full-length Vesicular Stomatitis Virus  
13 (VSV)-G protein. Then, these cells were transduced with a VSV vector expressing luciferase (VSV- $\Delta$ G-Luc),  
14 and pseudotyped with SARS-CoV-2 spike protein or VSV-G. After 2 hours at 37° C, the cells were washed 3  
15 times to remove residual virus. Supernatant containing pseudovirus was harvested 3 times at 24-hour intervals  
16 and centrifuged to remove cellular debris. Pseudovirus from the 3 collections was pooled and ultracentrifuged  
17 through a 20% sucrose cushion for purification and concentration (100x). For the transduction assays, Calu-3  
18 2B4 cells were grown in 96-well plates until confluent. Cells were incubated with the respective compounds for  
19 1 hour at 37° C. After 1 hour, cells were transduced with pseudovirus, maintaining the same concentration of  
20 compounds, and incubated overnight. Transduction efficiency was assessed by quantifying luciferase activity  
21 in cell lysates using a commercial kit (Luciferase Assay System, Promega, Cat. #E1500) and a plate-reading  
22 luminometer (SpectraMax i3x, Molecular Devices).

24 ***Infectious SARS-CoV-2 neutralization assay.*** The 2019n-CoV/USA-WA1/2019 strain of SARS-CoV-2  
25 (Accession number: MT985325.1) used in these studies was passaged on Calu-3 2B4 cells and sequence  
26 verified. Calu-3 2B4 cells were plated in 48 well plates. Cells were incubated with medium containing indicated  
27 compounds or vehicle for 1 hour at 37° C. The medium was removed and SARS-CoV-2 (MOI=0.1) in medium

1 containing indicated compounds were added into each well. The cells were incubated with viruses for 1 hour at  
2 37 °C. Next, the viruses were removed, and cells were rinsed with PBS once to remove remaining viruses.  
3 After that, cells were incubated with medium containing indicating compounds overnight. Following day, Total  
4 cellular RNA was isolated using Directzol RNA MiniPrep kit (Zymo Research, Cat. # R2052) from TRIzol  
5 (Invitrogen; Cat. #15596018). A DNase treatment step was included. Total RNA (500 ng) was used for cDNA  
6 syntheses by High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems; Cat. # 4368814). Realtime  
7 PCR was applied to quantify viral genomic RNA and HPRT mRNA levels (SARS-2-N1-F primer:  
8 GACCCCAAATCAGCGAAAT; SARS-2-N1-R primer: TCTGGTACTGCCAGTTGAATCTG; Human HPRT-F  
9 primer: AGGATTTGGAAAGGGTGTTC; Human HPRT-R primer: CAGAGGGCTACAATGTGATGG;  
10 Integrated DNA Technologies). The relative abundance of viral genomic RNA normalized to HPRT was  
11 calculated and presented as  $2^{-\Delta CT}$ .

12  
13 **Transduction and infection of mice.** Mice were anesthetized with ketamine/xylazine (87.5 mg/kg  
14 ketamine/12.5 mg/kg xylazine) and transduced intranasally with  $2.5 \times 10^8$  FFU of Ad5-ACE2 in 75 µL DMEM.  
15 Five days post transduction, mice were infected intranasally with SARS-CoV-2 ( $3 \times 10^3$  or  $1 \times 10^5$  PFU) in a  
16 total volume of 50 µL DMEM. Infected mice were treated with Avoralstat, Camostat (30 mg/kg intraperitoneal  
17 injection), or vehicle (DMSO; negative control) either four hours before and after being challenged by virus, or  
18 two doses per day (8 to 9 hours apart) for three days post infection. Virus titers were measured in harvested  
19 lungs by plaque assay 1-, 2-, 4-, or 5-days post infection. The weight was monitored for 10- or 12-days post  
20 infection.

21  
22 **SARS-CoV-2 plaque assay.** Lung homogenate supernatants were serially diluted in DMEM. Vero E6 cells in  
23 12 well plates were inoculated at 37 °C in 5% CO<sub>2</sub> for 1 hour with gentle rocking every 15 minutes. After  
24 removing the inocula, plates were overlaid with 1.2% agarose containing 10% FBS. After further incubation for  
25 3 days, overlays were removed, and plaques were visualized by staining with 0.1% crystal violet. Viral titers  
26 were calculated as plaque forming units (PFU) per lung. All work with SARS-CoV-2 was conducted in the

1 Biosafety Level 3 (BSL3) Laboratories of the University of Iowa. These studies were approved by the  
2 University of Iowa Institutional Animal Care and Use Committee.

### 3 4 **Statistics.**

5 **Kinetic parameters of biochemical proteases assays:** Data is displayed as mean  $\pm$  SEM (n = 3; number of  
6 reactions) and fit to the Hill equation (Figure 3D, 4B-C) or to the Michaelis-Menten equation (Supplemental  
7 Figure 5C) in GraphPad Prism 8 (GraphPad, San Diego, CA).

8 **TMPRSS2 autoproteolysis assay** (Figure 5B): Densitometry analysis. The analyzed densitometry data from  
9 the total of 5 gel runs in Figure 5A and Supplemental Figure 6 were combined (n=14 for vehicle and Camostat,  
10 respectively; n=5 for each group of GFP, Avoralstat, PCI-27483, Antipain, and SBTI). Data represent the mean  
11  $\pm$  SEM from two independent experiments analyzed by 1-way ANOVA followed by Dunnett's multiple  
12 comparisons test using GraphPad Prism 8.0. Differences of  $p < 0.0332$  were considered statistically significant  
13 (\* $p < 0.0332$ , \*\* $p < 0.0021$ , \*\*\* $p < 0.0002$ ).

14 **Pseudovirus transduction assay** (Figure 5C-D): Data represent the mean  $\pm$  SEM (n = 6; number of technical  
15 replicates) and are fit to the Hill equation (Figure 5C: Camostat  $R^2 = 0.71$ ; Avoralstat  $R^2 = 0.74$ ; PCI-27483  $R^2$   
16 = 0.20; SBTI  $R^2 = 0.49$ ; Antipain  $R^2 = 0.48$ ).

17 **Infectious SARS-CoV-2 neutralization assay** (Figure 5E): SARS-CoV-2 viral gRNA in the presence of 100  
18  $\mu$ M of DMSO (vehicle; negative control) or inhibitor. Data represent the mean  $\pm$  SEM (n = 3; number of  
19 technical replicates) and were analyzed by 1-way ANOVA followed by Tukey's multiple comparisons test  
20 (\*\*\*\* $p < 0.0001$  compared to vehicle).

21 **Infectious SARS-CoV-2 neutralization assay** (Figure 5F; Supplemental Figure 7): SARS-CoV-2 or MERS-  
22 CoV viral gRNA as a function of Camostat or Avoralstat concentration. Data represent the mean  $\pm$  SEM (n = 3;  
23 number of technical replicates) were analyzed by 2-way ANOVA followed by Dunnett's multiple comparisons  
24 test (\* $p < 0.0332$ , \*\*\*\* $p < 0.0001$  compared to vehicle).

25 **Viral titers in transduction and infection of mice** (Figure 6A-B, E): Data are represented as mean  $\pm$  SEM (n  
26 = 3; number of mice) and were analyzed by 1-way ANOVA followed by Tukey's multiple comparisons test  
27 (\* $p < 0.0332$ ; \*\* $p < 0.0021$ , \*\*\* $p < 0.0002$ , \*\*\*\* $p < 0.0001$  compared to vehicle).

1 ***Viral titers in transduction and infection of mice*** (Figure 6C): Data are represented as mean  $\pm$  SEM (n = 4  
2 for each group) and were analyzed by 1-way ANOVA followed by Tukey's multiple comparisons test  
3 (\*p<0.0332; \*\*p<0.0021, \*\*\*p<0.0002, \*\*\*\*p<0.0001 compared to vehicle).

4 ***Weights in transduction and infection of mice*** (Figure 6D, F): Data are represented as mean  $\pm$  SEM (n = 6;  
5 number of mice) and were analyzed by 2-way ANOVA followed by Dunnett's multiple comparisons test  
6 (\*p<0.0332; \*\*p<0.0021, \*\*\*p<0.0002, \*\*\*\*p<0.0001 compared to vehicle).

7  
8 ***Study approval.***

9 All animal studies were approved by the Institutional Animal Care and Use Committees of the University of  
10 Iowa.

## **Author Contributions**

Study concept and design: AGB and VBM. Acquisition of data: YJS, GV, DP, KL, MO, SS. Data analysis and interpretation: YJS, GV, DP, KL, MO, SS, PBM, AGB, VBM. Drafting of the manuscript: YJS, GV, DP, PBM, AGB, VBM. Critical revision of the manuscript: PBM, AGB, VBM. Obtained funding: PBM, AGB, VBM. Administrative, technical, and material support: VBM, AGB, PBM. Study supervision: VBM, AGB, PBM. YJS, GV, DP, and KL are co-first authors. The order of the co-first authors was assigned on the basis of their relative contribution to the study. PBM, AGB, and VBM jointly supervised the work.

## **Acknowledgments**

The following reagent was deposited by the Centers for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate USA-WA1/2020, NR-52281. U.S. At Stanford University, we thank Jing Yang, Angela S. Li, and Teja Chemudupati in the Department of Ophthalmology for technical assistance; Dr. Jeffrey L. Goldberg (Chair of Ophthalmology) for supporting our research during the COVID-19 campus lockdown; and Dr. Corey Liu and Dr. Daniel Fernandez in the Stanford ChEM-H Macromolecular Structure Knowledge Center for supporting protein production along with Dr. Mark Smith (Director) in the Stanford ChEM-H Medicinal Chemistry Knowledge Center for helpful discussions. From the University of Iowa, we thank Shu Wu in the Department of Pediatrics for technical assistance; and we thank Dr. Stanley Perlman for helpful discussions. From Loyola University Chicago, we thank Dr. Thomas Gallagher for sharing TMPRSS2 plasmids.

**Funding information.** VBM and AGB are supported by NIH grants [R01EY 030151, R01 EY031952, R01NS98950, R01EY024665, and P30EY026877], and Research to Prevent Blindness, New York, New York. The study was also supported by Stanford ChEM-H/IMA awarded to VBM. GV is supported by NIH grants [F30EYE027986 and T32GM007337]. DP is supported by a NEI/NIH T32EY027816 training grant. The study was also supported by P01AI060699 (PBM); and the Pathology Core, which is partially supported by the Center for Gene Therapy for Cystic Fibrosis (NIH Grant P30DK-54759), and the Cystic Fibrosis Foundation. PBM is supported by the Roy J. Carver Charitable Trust. MO is supported by a NHLBI/NIH T32HL007638 training grant.

**Data and Materials Availability.** Correspondence and requests for materials should be addressed to Vinit B. Mahajan ([vinit.mahajan@stanford.edu](mailto:vinit.mahajan@stanford.edu)). Reagents are available with a Materials Transfer Agreement. 3DPhyloFold is open source and available at Mendeley Data under the dataset identifier (DOI): 10.17632/kk3gkzdsbf.2. The implementation notes, code, and description of methodology are available on the site. The raw docking data and parameters have been deposited to Mendeley Data with the dataset identifier (DOI): 10.17632/h3pmycddwc.1.

## 1 References

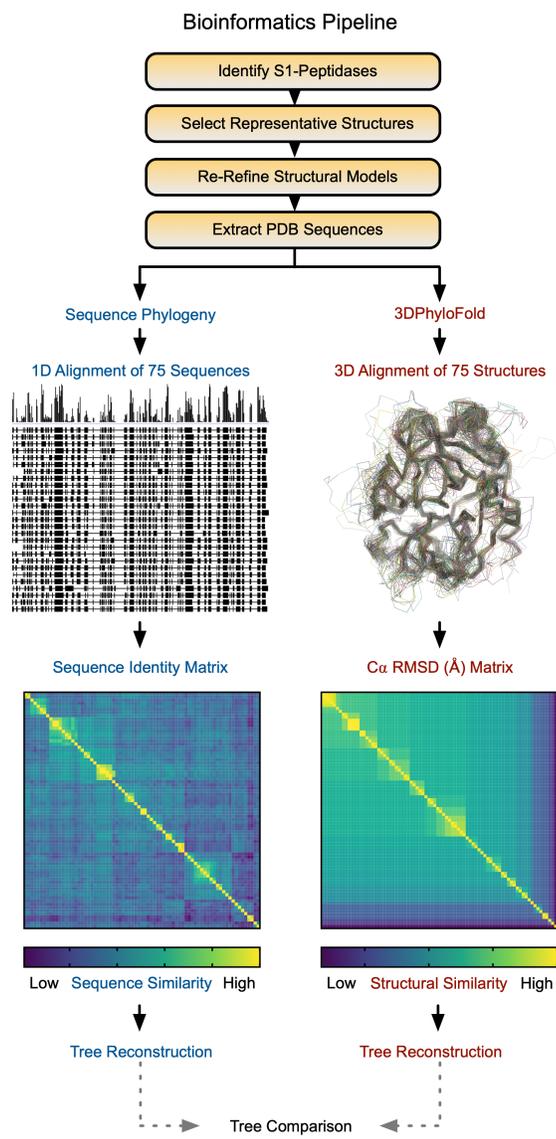
- 2 1. Kawase M, Shirato K, van der Hoek L, Taguchi F, and Matsuyama S. Simultaneous treatment of  
3 human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute  
4 respiratory syndrome coronavirus entry. *J Virol.* 2012;86(12):6537-45.
- 5 2. Iwata-Yoshikawa N, Okamura T, Shimizu Y, Hasegawa H, Takeda M, and Nagata N. TMPRSS2  
6 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus  
7 Infection. *J Virol.* 2019;93(6).
- 8 3. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a  
9 new coronavirus of probable bat origin. *Nature.* 2020;579(7798):270-3.
- 10 4. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell  
11 Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.*  
12 2020.
- 13 5. Kim TS, Heinlein C, Hackman RC, and Nelson PS. Phenotypic analysis of mice lacking the *Tmprss2*-  
14 encoded protease. *Mol Cell Biol.* 2006;26(3):965-75.
- 15 6. Shirato K, Kawase M, and Matsuyama S. Middle East respiratory syndrome coronavirus infection  
16 mediated by the transmembrane serine protease TMPRSS2. *J Virol.* 2013;87(23):12552-61.
- 17 7. Shrimp JH, Kales SC, Sanderson PE, Simeonov A, Shen M, and Hall MD. An Enzymatic TMPRSS2  
18 Assay for Assessment of Clinical Candidates and Discovery of Inhibitors as Potential Treatment of  
19 COVID-19. *ACS Pharmacol Transl Sci.* 2020;3(5):997-1007.
- 20 8. Shen LW, Mao HJ, Wu YL, Tanaka Y, and Zhang W. TMPRSS2: A potential target for treatment of  
21 influenza virus and coronavirus infections. *Biochimie.* 2017;142:1-10.
- 22 9. Midgley I, Hood AJ, Proctor P, Chasseaud LF, Irons SR, Cheng KN, et al. Metabolic fate of 14C-  
23 camostat mesylate in man, rat and dog after intravenous administration. *Xenobiotica.* 1994;24(1):79-92.
- 24 10. Talevi A, and Bellera CL. Challenges and opportunities with drug repurposing: finding strategies to find  
25 alternative uses of therapeutics. *Expert Opin Drug Discov.* 2020;15(4):397-401.

- 1 11. Schweinitz A, Steinmetzer T, Banke IJ, Arlt MJ, Sturzebecher A, Schuster O, et al. Design of novel and  
2 selective inhibitors of urokinase-type plasminogen activator with improved pharmacokinetic properties  
3 for use as antimetastatic agents. *J Biol Chem*. 2004;279(32):33613-22.
- 4 12. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of M(pro) from SARS-CoV-2 and discovery of  
5 its inhibitors. *Nature*. 2020;582(7811):289-93.
- 6 13. Shin D, Mukherjee R, Grewe D, Bojkova D, Baek K, Bhattacharya A, et al. Papain-like protease  
7 regulates SARS-CoV-2 viral spread and innate immunity. *Nature*. 2020;587(7835):657-62.
- 8 14. Johnson BA, Xie X, Kalveram B, Lokugamage KG, Muruato A, Zou J, et al. Furin Cleavage Site Is Key  
9 to SARS-CoV-2 Pathogenesis. *bioRxiv*. 2020.
- 10 15. Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, and Gallagher T. A transmembrane  
11 serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates  
12 virus entry. *J Virol*. 2011;85(2):873-82.
- 13 16. Whitt MA. Generation of VSV pseudotypes using recombinant DeltaG-VSV for studies on virus entry,  
14 identification of entry inhibitors, and immune responses to vaccines. *J Virol Methods*. 2010;169(2):365-  
15 74.
- 16 17. Sun J, Zhuang Z, Zheng J, Li K, Wong RLY, Liu D, et al. Generation of a Broadly Useful Model for  
17 COVID-19 Pathogenesis, Vaccination, and Treatment. *Cell*. 2020.
- 18 18. Schneider CA, Voth E, Theissen P, Wienhard K, Wagner R, Baer FM, et al. [Assessing myocardial  
19 viability in chronic myocardial infarct with 18F-fluoro-D-glucose positron emission tomography and  
20 99mTc-MIBI SPECT]. *Z Kardiol*. 1994;83(2):124-31.
- 21 19. Choi JY, Kang YJ, Jang HM, Jung HY, Cho JH, Park SH, et al. Nafamostat Mesilate as an  
22 Anticoagulant During Continuous Renal Replacement Therapy in Patients With High Bleeding Risk: A  
23 Randomized Clinical Trial. *Medicine (Baltimore)*. 2015;94(52):e2392.
- 24 20. Cornpropst M, Collis P, Collier J, Babu YS, Wilson R, Zhang J, et al. Safety, pharmacokinetics, and  
25 pharmacodynamics of avoralstat, an oral plasma kallikrein inhibitor: phase 1 study. *Allergy*.  
26 2016;71(12):1676-83.

- 1 21. Riedl MA, Aygoren-Pursun E, Baker J, Farkas H, Anderson J, Bernstein JA, et al. Evaluation of  
2 avoralstat, an oral kallikrein inhibitor, in a Phase 3 hereditary angioedema prophylaxis trial: The OPuS-  
3 2 study. *Allergy*. 2018;73(9):1871-80.
- 4 22. . Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy  
5 Volunteers [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/estimating-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/estimating-maximum-safe-starting-dose-initial-clinical-trials-therapeutics-adult-healthy-volunteers)  
6 [maximum-safe-starting-dose-initial-clinical-trials-therapeutics-adult-healthy-volunteers](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/estimating-maximum-safe-starting-dose-initial-clinical-trials-therapeutics-adult-healthy-volunteers). Accessed July  
7 2, 2020.
- 8 23. Song HK, and Suh SW. Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex  
9 with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from  
10 *Erythrina caffra* and tissue-type plasminogen activator. *J Mol Biol*. 1998;275(2):347-63.
- 11 24. Bakowski MA, Beutler N, Chen E, Nguyen T-TH, Kirkpatrick MG, Parren M, et al. Oral drug  
12 repositioning candidates and synergistic remdesivir combinations for the prophylaxis and treatment of  
13 COVID-19. *bioRxiv*. 2020:2020.06.16.153403.
- 14 25. Fragkou PC, Belhadi D, Peiffer-Smadja N, Moschopoulos CD, Lescure FX, Janocha H, et al. Review of  
15 trials currently testing treatment and prevention of COVID-19. *Clin Microbiol Infect*. 2020.
- 16 26. Katoh K, and Standley DM. MAFFT multiple sequence alignment software version 7: improvements in  
17 performance and usability. *Mol Biol Evol*. 2013;30(4):772-80.
- 18 27. Finn RD, Clements J, and Eddy SR. HMMER web server: interactive sequence similarity searching.  
19 *Nucleic Acids Res*. 2011;39(Web Server issue):W29-37.
- 20 28. Li W, and Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or  
21 nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-9.
- 22 29. Gouveia-Oliveira R, Sackett PW, and Pedersen AG. MaxAlign: maximizing usable data in an alignment.  
23 *BMC Bioinformatics*. 2007;8:312.
- 24 30. Nguyen LT, Schmidt HA, von Haeseler A, and Minh BQ. IQ-TREE: a fast and effective stochastic  
25 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268-74.

- 1 31. Velez G, Sun YJ, Khan S, Yang J, Herrmann J, Chemudupati T, et al. Structural Insights into the  
2 Unique Activation Mechanisms of a Non-classical Calpain and Its Disease-Causing Variants. *Cell Rep.*  
3 2020;30(3):881-92 e5.
- 4 32. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families  
5 database. *Nucleic Acids Res.* 2014;42(Database issue):D222-30.
- 6 33. Joosten RP, Salzemann J, Bloch V, Stockinger H, Berglund AC, Blanchet C, et al. PDB\_REDO:  
7 automated re-refinement of X-ray structure models in the PDB. *J Appl Crystallogr.* 2009;42(Pt 3):376-  
8 84.
- 9 34. de Vries SJ, van Dijk M, and Bonvin AM. The HADDOCK web server for data-driven biomolecular  
10 docking. *Nat Protoc.* 2010;5(5):883-97.
- 11

1 **Figures and figure legends**

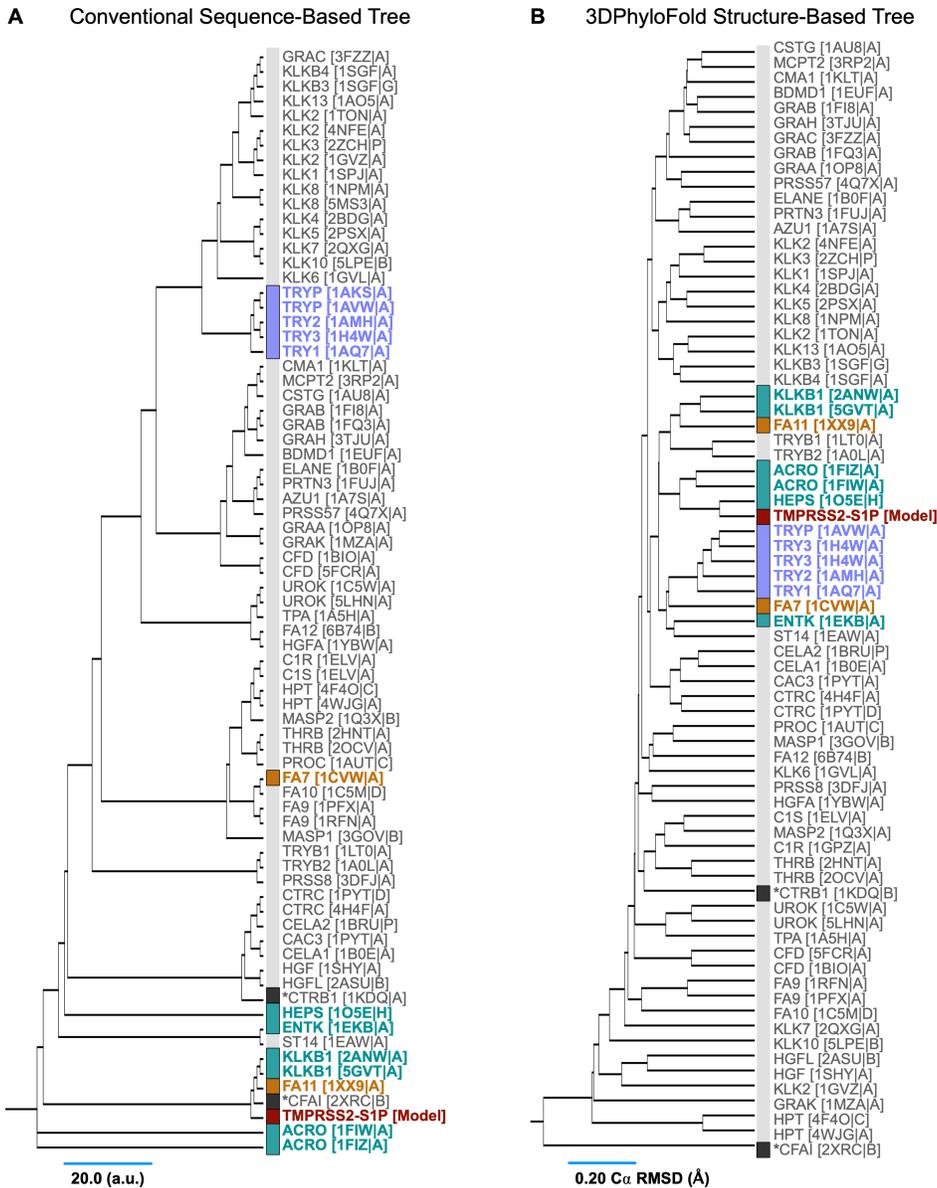


2

3 **Figure 1. Bioinformatics workflow for sequence-based phylogenetic analysis and structure-based**  
4 **phylogenetic analysis**

5 Conventional sequence-based phylogenetic analysis relies on alignment of the primary sequence.  
6 3DPhyloFold calculates a structural dissimilarity matrix (SDM) based on the overlay of 75 representative  
7 mammalian S1-peptidase domains with high-resolution structures. The workflow can determine the differences  
8 and similarities between the sequence-based phylogeny and structure-based phylogeny of a domain family.

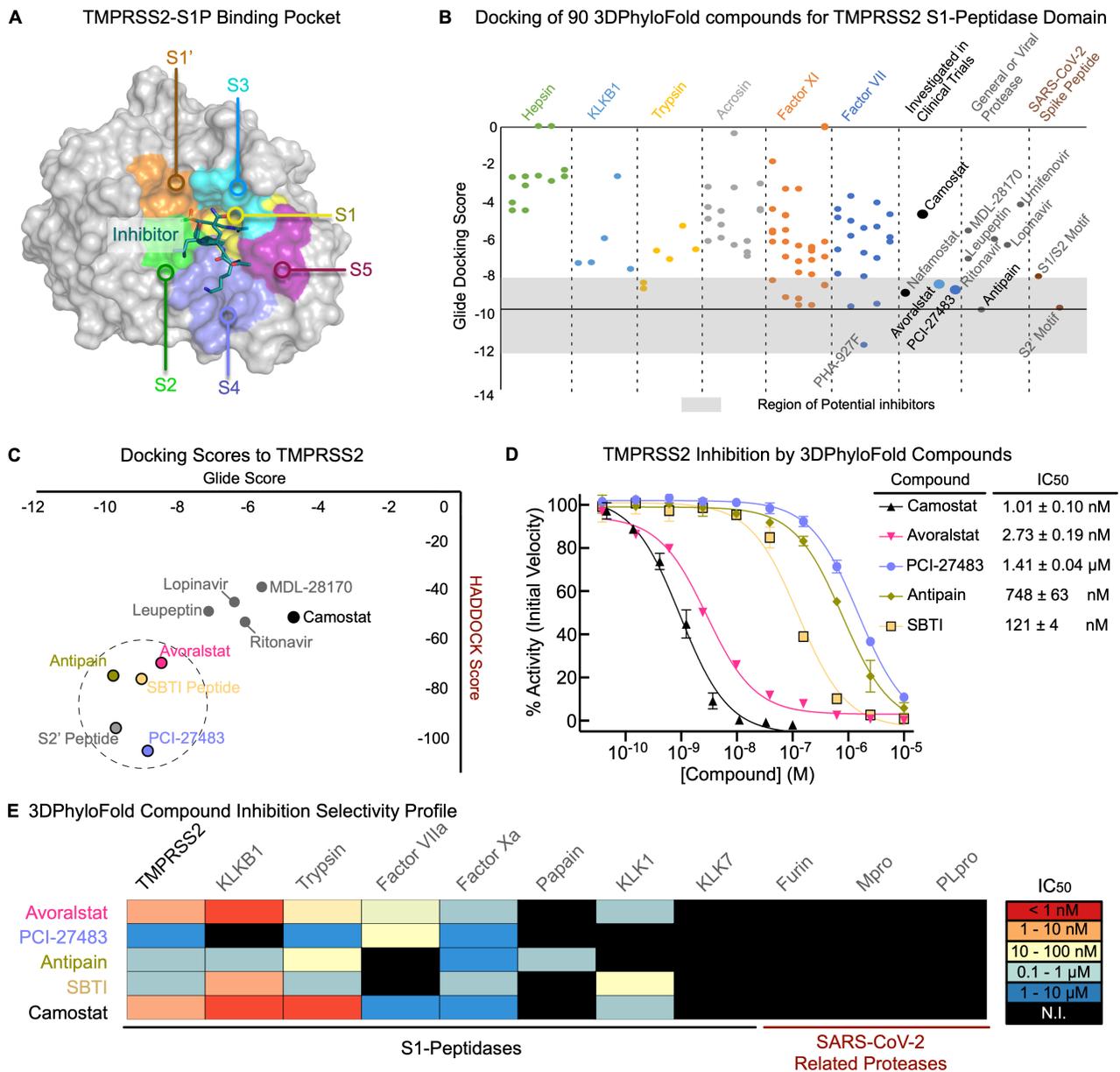
9



**Figure 2. Structure-based phylogenetic analysis identifies sequentially divergent serine proteases with similar folds to TMPRSS2**

**(A)** Sequence-based phylogenetic tree of 75 representative peptidase domains with the main clusters highlighted in different colors. Branches are labeled according to the corresponding structures and PDB IDs.

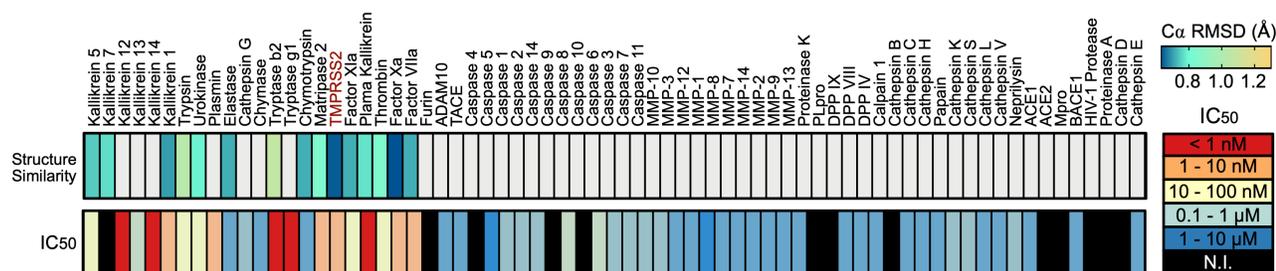
**(B)** Structure-based phylogenetic tree of mammalian S1-peptidase domain structures. Evolutionary distance was inferred using the UPGMA method. Branch lengths correspond to the C $\alpha$  RMSD (Å) of the pairwise-structural-alignments calculated in 3DPhyloFold. The proteases with the highest structural similarity to TMPRSS2-S1P were Hepsin, Acrosin, Trypsin, Factor VII, Factor XI, and KLKB1.



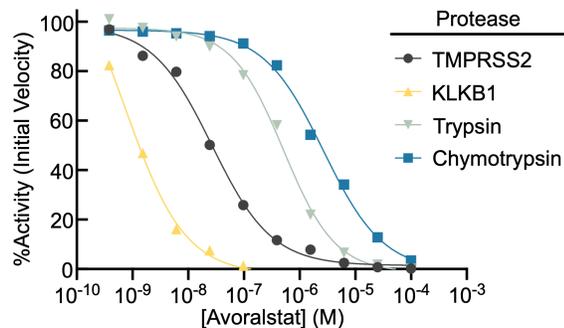
**Figure 3. Compounds derived from 3DPhyloFold in silico analysis inhibit TMPRSS2 activity in vitro**

**(A)** TMPRSS2-S1P structural model ligand-binding sub-pockets (S1', S1, S2, S3, S4, S5). **(B)** Docking scores of compounds curated from 3DPhyloFold and **(C)** correlation between algorithms. Potential inhibitors clustered around the natural S2'-peptide motif. **(D)** TMPRSS2 inhibition by 3DPhyloFold compounds. IC<sub>50</sub> data represent mean ± SEM; n=3; calculated from the Hill equation. **(E)** IC<sub>50</sub> value selectivity profile against S1- and SARS-CoV-2-related proteases. IC<sub>50</sub> data was calculated from the Hill equation (mean ± SEM; n=3).

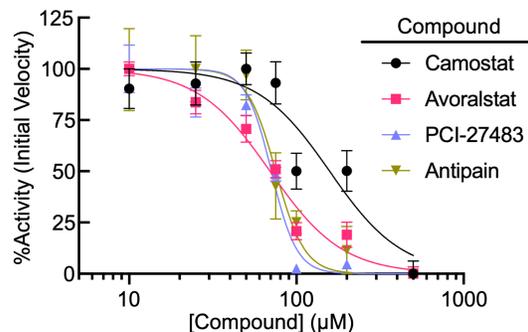
### A Avoralstat Protease Inhibition Selectivity



### B 3DPhyloFold Protease Inhibition by Avoralstat Comparison



### C TMPRSS2-S1P Domain Inhibition



**Figure 4. Avoralstat inhibits TMPRSS2 activity in vitro**

(A) Avoralstat selectivity against 70 proteases. Top row -  $C\alpha$ -RMSD of pairwise-structural-alignments to TMPRSS2. (B) Protease inhibition by Avoralstat correlates with 3DPhyloFold prediction. (C) TMPRSS2-S1P domain inhibition.  $IC_{50}$  data represent mean  $\pm$  SEM;  $n=3$ ; calculated from the Hill equation.

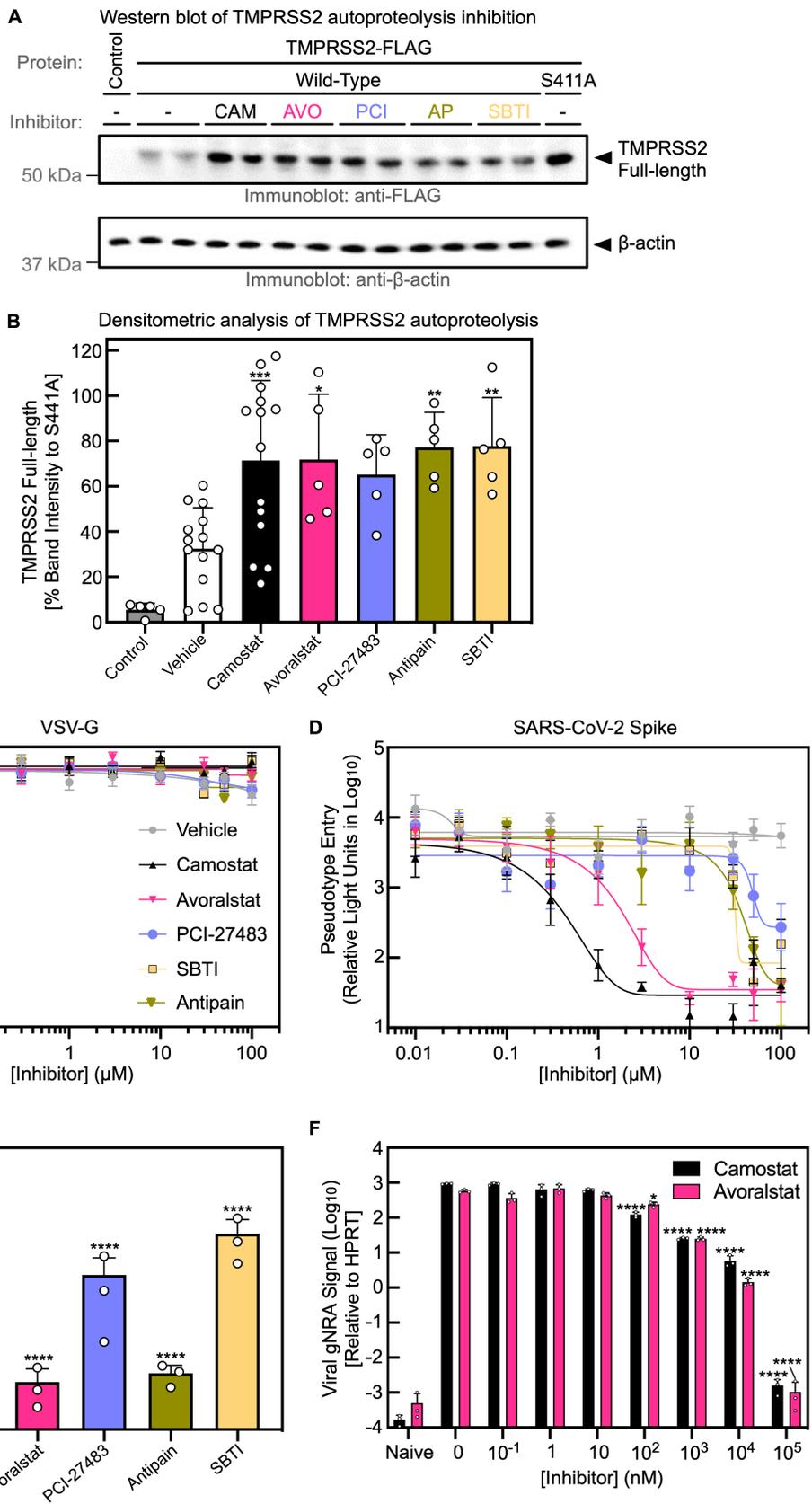
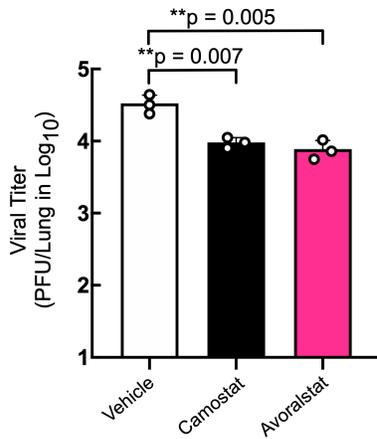


Figure 5. Avoralstat inhibits viral entry directed by SARS-CoV-2 spike proteins

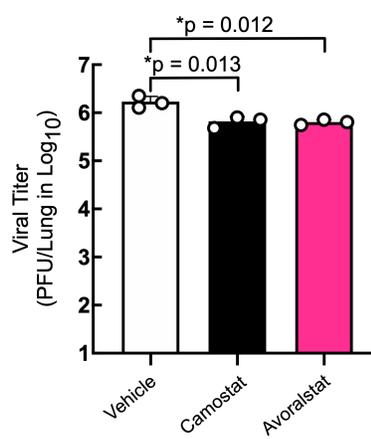
1 **(A-B)** HEK-cells treated with Camostat, Avoralstat, PCI-27483, Antipain, or SBTI 2-hr before TMPRSS2  
2 transfection. TMPRSS2 reduced autoproteolysis (increased TMPRSS2-FL signal; 1-way ANOVA followed by  
3 Dunnett's multiple comparisons test; \* $p < 0.0332$ , \*\* $p < 0.0021$ , \*\*\* $p < 0.0002$  compared to vehicle;  $n = 5$  for each  
4 group except vehicle and Camostat [ $n = 14$ ]). Calu-3-cells were treated with compounds and inoculated with  
5 pseudovirions harboring **(C)** VSV-G or **(D)** SARS-CoV-2-spike-protein ( $n = 6$ ; fit to the Hill equation). Calu-3  
6 cells were treated with **(E)** 100  $\mu\text{M}$  or **(F)** indicated concentrations of each compound, then incubated with  
7 SARS-CoV-2. Viral gRNA was measured after 24-hrs. Data represent mean  $\pm$  SEM;  $n = 3$ . **(E)** Compounds  
8 reduced viral signal at 100  $\mu\text{M}$  (1-way ANOVA followed by Tukey's multiple comparisons test; mean  $\pm$  SEM;  $n$   
9 = 3; \*\*\*\* $p < 0.0001$  compared to vehicle). **(F)** Viral signal was reduced beginning from 100 nM (2-way ANOVA  
10 followed by Dunnett's multiple comparisons test; mean  $\pm$  SEM;  $n = 3$ ; \* $p < 0.0332$ , \*\*\*\* $p < 0.0001$  compared to  
11 vehicle).

**A** 1-day post infection viral titer

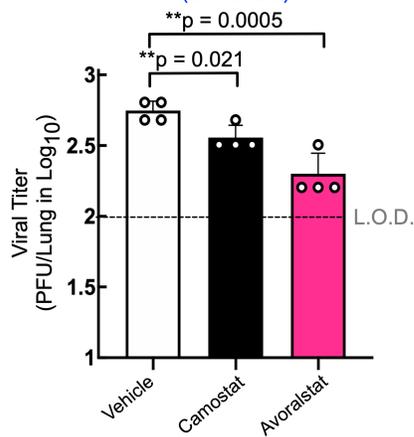
Infection dose:  $3 \times 10^3$  PFU/Mouse  
 Drug dose: 1 dose (at -4 hrs), 1 dose (at +4 hrs)

**B** 1-day post infection viral titer

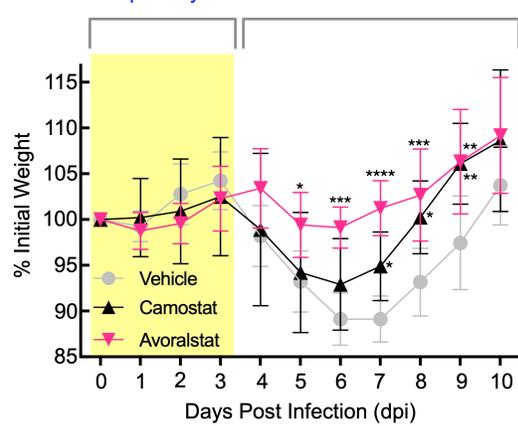
Infection dose:  $1 \times 10^5$  PFU/Mouse  
 Drug dose: 1 dose (at -4 hrs), 1 dose (at +4 hrs)

**C** 5-day post infection viral titer

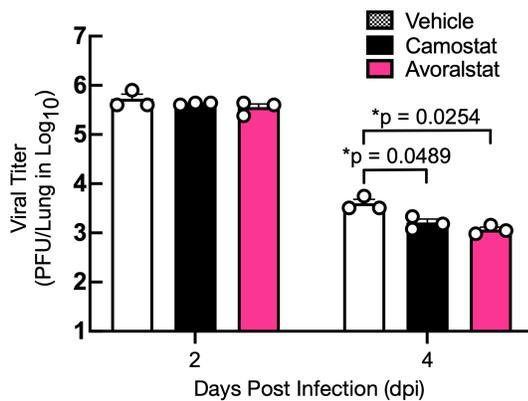
Infection dose:  $1 \times 10^5$  PFU/Mouse  
 Drug dose: 2 drug doses per day (DPI 0 to 3)

**D** Mouse Average Weight Change

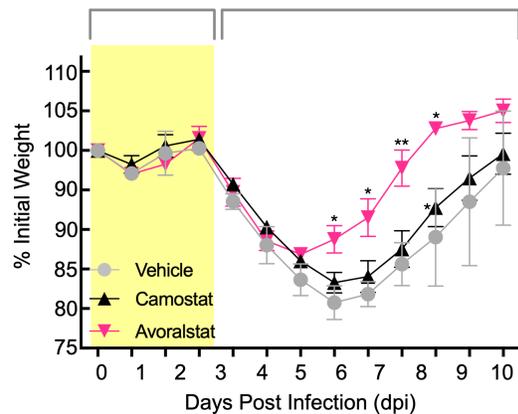
Infection dose:  $1 \times 10^5$  PFU/Mouse  
 Drug dose: 2 drug doses per day (DPI 0 to 3), Observe

**E** Post infection viral titer

Infection dose:  $1 \times 10^6$  PFU/Mouse  
 Drug dose: 2 drug doses per day (DPI 0 to 3)

**F** Mouse Average Weight Change

Infection dose:  $1 \times 10^6$  PFU/Mouse  
 Drug dose: 2 drug doses per day (DPI 0 to 3), Observe



1 **Figure 6. Avoralstat reduces SARS-CoV-2 infection in mice**

2 **(A-C)** Ad5-hACE2-transduced BALB/c mice treated with Avoralstat, Camostat, or DMSO 4-hrs before and after  
3 SARS-CoV-2 viral challenge showed significantly reduced lung viral titer (1-way ANOVA followed by Tukey's  
4 multiple comparisons test; n = 3 or 4; L.O.D., limit of detection). **(D)** Avoralstat protected mice from weight-loss  
5 better than Camostat, n = 6 for each group (2-way ANOVA followed by Dunnett's multiple comparisons test;  
6 \*p<0.0332, \*\*p<0.0021, \*\*\*p<0.0002, \*\*\*\*p<0.0001 compared to vehicle; mean ± SEM). **(E)** Avoralstat  
7 significantly reduced lung viral titer, n = 3 for each group (1-way ANOVA followed by Tukey's multiple  
8 comparisons test). **(F)** At the highest dose of SARS-CoV-2, mouse weight-loss was better mitigated by  
9 Avoralstat compared to vehicle and Camostat, n = 6 for each group (2-way ANOVA followed by Dunnett's  
10 multiple comparisons test; \*p<0.0332, \*\*p<0.0021 compared to vehicle; mean ± SEM).

11