(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
06 January 2022 (06.01.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/006232 A2

(51) International Patent Classification:
A61K 41/00 (2020.01)

(21) International Application Number:
PCT/US2021/039827

(22) International Filing Date:
30 June 2021 (30.06.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/047,078 01 July 2020 (01.07.2020) US

(71) Applicants: **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; Office of the General Counsel, Building 170, 3rd Floor, Main Quad, P.O. Box 20386, Stanford, California 94305 (US). **UNIVERSITY OF IOWA RESEARCH FOUNDATION** [US/US]; 200 S Capitol St, 2660 University Capitol Centre, Iowa City, Iowa 52242 (US).

(72) Inventors: **MAHAJAN, Vinit**; 1651 Page Mill Road, Palo Alto, California 94304 (US). **SUN, Young Joo**; 1651 Page Mill Road, Palo Alto, California 94304 (US). **VELEZ, Gabriel**; 1651 Page Mill Road, Palo Alto, California 94304 (US). **PARSONS, Dylan**; 3000 El Camino Real, Bldg. 5, Suite 300, Palo Alto, California 94306-2100 (US). **BASSUK, Alexander G**; 200 S Capitol St, 2660 University Capitol Centre, Iowa City, Iowa 52242 (US). **MCCRAY, Paul B**; 200 S Capitol St, 2660 University Capitol Centre, Iowa City, Iowa 52242 (US). **LI, Kun**; 200 S Capitol St, 2660 University Capitol Centre, Iowa City, Iowa 52242 (US). **ORTIZ, Miguel**; 200 S Capitol St, 2660 University Capitol Centre, Iowa City, Iowa 52242 (US).

(74) Agent: **BUCHBINDER, Jenny L.**; 201 Redwood Shores Pkwy, Suite 200, Redwood City, California 94065 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: DRUG REPURPOSING FOR TREATMENT OF VIRAL INFECTIONS

FIG. 1A

(57) Abstract: Methods of screening drugs and compounds, potentially useful in treating SARS-CoV-2, which have been approved previously by the U.S. Food and Drug Administration for other purposes, are provided. In particular, a computational protein structure-based phylogenetics approach was used to identify S1 -peptidases structurally related to transmembrane serine protease 2 (TMPRSS2), which is known to facilitate viral entry of SARS-CoV-2 into human cells. Selected inhibitors of these structurally related peptidases inhibit TMPRSS2 as well as SARS-CoV-2 infection, including the plasma kallikrein inhibitor, Avoralstat, the coagulation factor VII inhibitor, PCI-27483, and the trypsin inhibitor, soybean trypsin inhibitor (SBTI) and may be useful in treating SARS-CoV-2.

WO 2022/006232 A2

[Continued on next page]

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

DRUG REPURPOSING FOR TREATMENT OF VIRAL INFECTIONS

BACKGROUND OF THE INVENTION

[0001] The *Coronaviridae* family is responsible for several respiratory illnesses in humans. The severe acute respiratory syndrome coronavirus (SARS-CoV) caused SARS, an epidemic that originated in China in 2002 and ultimately resulted in more than 8,000 cases and 774 deaths across 26 countries (de Wit et al. (2016) Nat. Rev. Microbiol. 14(8):523-534). The current pandemic of coronavirus disease 19 (COVID-19) is caused by SARS-coronavirus 2 (SARS-CoV-2), a virus with high sequence similarity to SARS-CoV (Zhu et al. (2020) N. Engl. J. Med. 382(8):727-733). COVID-19, a viral pneumonia, was first identified in December 2019 in the Hubei province of China and has since spread to nearly 185 countries within the span of three months, causing an unprecedented global health crisis. The virus is alarmingly contagious, with a mean basic reproduction number (R_0) of 3.3, much higher than that of influenza (Zhu et al., supra), there were over 1,991,562 confirmed cases worldwide with 130,885 confirmed deaths (WHO, Coronavirus disease (COVID-19) Pandemic. [cited 2020 March 22] who.int/emergencies/diseases/novel-coronavirus-2019). Given that approximately 30% of carriers will be asymptomatic, the true number of cases is likely much higher and will continue to rapidly rise as more diagnostic capabilities are ramped up (Nishiura et al. (2020) J. Clin. Med. 9(2), Nishiura et al. (2020) Int J. Infect. Dis. 94:154-155). There is a critical need to find an effective treatment for COVID-19.

SUMMARY OF THE INVENTION

[0002] Methods of screening drugs and compounds, potentially useful in treating SARS-CoV-2, which have been approved previously by the U.S. Food and Drug Administration for other purposes, are provided. In particular, a computational protein structure-based phylogenetics approach was used to identify S1-peptidases structurally related to transmembrane serine protease 2 (TMPRSS2), which is known to facilitate viral entry of SARS-CoV-2 into human cells. Selected inhibitors of these structurally related peptidases were tested for their ability to inhibit TMPRSS2 as well as SARS-CoV-2 infection. In addition, methods of treating COVID-19 with peptidase inhibitors identified as having efficacy in inhibiting SARS-CoV-2 infection from this screening, including the plasma kallikrein inhibitor, Avoralstat, the coagulation factor VII inhibitor, PCI-27483, and the trypsin inhibitor, soybean trypsin inhibitor (SBTI), are also provided.

[0003] In one aspect, a method of treating a subject for an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is provided, the method comprising

administering to the subject a therapeutically effective amount of a composition comprising a plasma kallikrein inhibitor, a coagulation factor VII inhibitor, or a trypsin inhibitor.

[0004] In one embodiment, the plasma kallikrein inhibitor is Avoralstat.

[0005] In another embodiment, the coagulation factor VII inhibitor is PCI-27483.

[0006] In another embodiment, the trypsin inhibitor is soybean trypsin inhibitor (SBTI).

[0007] In certain embodiments, the composition comprising the plasma kallikrein inhibitor, coagulation factor VII inhibitor, or trypsin inhibitor is administered prophylactically (i.e., to prevent or lower the risk of an infection by SARS-CoV-2). Such prophylactic uses will be of particular value for subjects who are at increased risk for severe illness from COVID-19, including, without limitation, subjects who are immunodeficient, patients who have been treated with immunosuppressive agents, subjects who are 65 years of age or older, or subjects who have a genetic predisposition, condition, or disease, including, without limitation, acquired immunodeficiency syndrome (AIDS), cancer, diabetes (type 1 or type 2), chronic obstructive pulmonary disease (COPD), chronic kidney disease, obesity, an organ transplant, a heart condition, such as heart failure, coronary artery disease, or cardiomyopathies, sickle cell disease, cystic fibrosis, pregnancy, liver disease, asthma, or hypertension, or any other condition that makes them prone to developing infections.

[0008] In certain embodiments, multiple cycles of treatment are administered to the subject for a time period sufficient to eradicate the infection by SARS-CoV-2. For example, the composition can be administered according to a daily dosing regimen or intermittently.

[0009] The composition can be administered according to any suitable mode of administration. In certain embodiments, the composition is administered orally, intravenously, or by pulmonary inhalation. In some embodiments, the composition is administered by pulmonary inhalation using, for example, without limitation, a nebulizer, a metered dose inhaler (MDI), or a dry powder inhaler (DPI).

[0010] In certain embodiments, the method further comprises administering additional anti-viral therapy.

[0011] In certain embodiments, the subject is human.

[0012] In certain embodiments, the composition is administered in an amount sufficient to reduce viral entry or replication of SARS-CoV-2 in the subject.

[0013] In another aspect, a composition comprising a plasma kallikrein inhibitor, a coagulation factor VII inhibitor, or a trypsin inhibitor for use in the treatment of an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is provided. In certain embodiments, the composition further comprises a pharmaceutically acceptable excipient. In certain embodiments, the composition comprises Avoralstat, PCI-27483, or soybean trypsin inhibitor (SBTI).

[0014] In another aspect, a method of screening for an S1 peptidase inhibitor that inhibits an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is provided, the method comprising: a) modeling a TMPRSS2 S1 peptidase domain by using a structure of hepsin as a template; b) calculating a TMPRSS2 structural similarity score for a plurality of S1 peptidases, wherein S1 peptidases with the highest structural similarity scores are prioritized for further *in silico* analysis; c) modeling a TMPRSS2 binding pocket by comparison to a structure of a hepsin binding pocket; d) docking a plurality of the inhibitors of the prioritized S1 peptidases into the modeled TMPRSS2 binding pocket, wherein the inhibitors are prioritized for further screening based on their docking scores, wherein higher docking scores indicate more favorable binding interactions in the TMPRSS2 binding pocket; e) measuring inhibition of TMPRSS2 activity by one or more of the inhibitors selected for further screening, wherein inhibitors are prioritized according to how strongly they inhibit TMPRSS2; and f) measuring inhibition of an infection by the SARS-CoV-2 in a host cell or subject by one or more of the inhibitors that inhibited TMPRSS2 to identify an inhibitor that inhibits the infection by the SARS-CoV-2.

[0015] In certain embodiments, the inhibitor has been approved by the U.S. Food and Drug Administration previously for treatment of a disease or condition other than a SARS-CoV-2 infection.

[0016] In another aspect, an inhibitor identified by a screening method described herein for the treatment of a SARS-CoV-2 infection is provided.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] The invention is best understood from the following detailed description when read in conjunction with the accompanying drawings. It is emphasized that, according to common practice, the various features of the drawings are not to-scale. On the contrary, the dimensions of the various features are arbitrarily expanded or reduced for clarity. Included in the drawings are the following figures.

[0018] **FIGS. 1A-1C.** Sequence-based phylogenetic analysis of the S1-peptidase superfamily reveals the evolutionary relatedness of TMPRSS2 to other proteases: **(FIG. 1A)** The reconstructed S1-peptidases superfamily tree representing 600 sequences across all species. Evolutionary distances were inferred using the maximum likelihood method with 1,000 bootstrap replicas. The unrooted tree is represented with the main clades highlighted in different colors. Bootstrap values are projected onto the tree topology. The positions of TMPRSS2 and representative serine proteases are denoted by black squares in each clade for reference. **(FIG. 1B)** Schematic representation of TMPRSS/Hepsin domain architecture. **(FIG. 1C)** Ribbon tracing diagram representing the S1-peptidase domain of human TMPRSS2

colored by the ConSurf evolutionary conservation scale. Modeling is based on the structure of human Hepsin (PDB 1Z8G). The catalytic triad residues (H296, D345, and S441) and active site groove are highly conserved, while the surface-exposed regions are variable.

[0019] **FIGS. 2A-2C.** Structure-based phylogenetic analysis identifies sequentially divergent serine proteases with similar folds to TMPRSS2: **(FIG. 2A)** Bioinformatics workflow. Conventional sequence-based phylogenetic analysis relies of alignment of the primary sequence. 3DPhyloFold calculates a structural dissimilarity matrix (SDM) based on the overlay of 75 representative mammalian S1-peptidase domains with high-resolution structures. **(FIG. 2B)** Sequence-based phylogenetic tree of 75 representative peptidase domains. Evolutionary distances were inferred using the maximum likelihood method with 1,000 bootstrap replicas. The unrooted tree is represented with the main clusters highlighted in different colors. Branches are labeled according to the corresponding structures with respective PDB IDs in parentheses. **(FIG. 2C)** Structure-based phylogenetic tree of representative mammalian S1-peptidase domain structures. The evolutionary distance was inferred using the UPGMA method. The tree is drawn to scale, with branch lengths corresponding to the Ca RMSD (**FIG. 2**Å) of the pairwise structural alignments. The proteases with the highest structural similarity were: Hepsin, Acrosin, Trypsin (-1, -2, -B1, and -B2), Plasma Kallikrein, Coagulation Factors (-VII and -XI), and Enteropeptidase.

[0020] **FIGS. 3A-3E.** *In silico* docking of protease inhibitors to TMPRSS2 reveals possible new drug targets: **(FIG. 3A)** Serine protease domain structures were overlaid onto the TMPRSS2 homology model by minimum Ca RMSD algorithm in PyMOL. Each structure was manually inspected and the protease residues forming the sub-pockets (e.g. S1', S1, and such) that accommodate the different part of ligands were curated. The TMPRSS2 structural model (surface model) overlaid onto the inhibitor-bound (stick model) structure of human Hepsin was used to represent the S-1 peptidase domains' sub-pockets and their ligand binding site (PDB 1Z8G). Orange; S1' sub-pocket, yellow; S1 sub-pocket, green; S2 sub-pocket, cyan; S3 sub-pocket, slate; S4 sub-pocket, and violet; S5 sub-pocket. **(FIG. 3B)** Graphical representation of docking scores shows a good correlation between Glide and HADDOCK docking scores, with potential inhibitors clustering around the natural S2' peptide motif (Dashed circle). Interactions between the TMPRSS2 structural model and the docked chemical structures of **(FIG. 3C)** Camostat, **(FIG. 3D)** Avoralstat, and **(FIG. 3E)** PCI-27483.

[0021] **FIGS. 4A-4C.** Compounds targeting trypsin-like peptidases inhibit TMPRSS2 activity *in vitro*: **(FIG. 4A)** Michaelis-Menten analysis of Cbz-GGR-AMC hydrolysis by 250 nM TMPRSS2 in the presence of 5% DMSO. Data is displayed as mean ± SEM (n=3) and fit to the Michaelis-Menten equation (R^2 = 0.88). **(FIG. 4B)** Inhibition of Cbz-GGR-AMC (50 μM) hydrolysis by 250 nM TMPRSS2 in the presence of 10-500 μM Camostat (black), Avoralstat

(magenta), PCI-27483 (blue), or Antipain (green). The initial velocity for each condition was normalized to percent activity and plotted against the inhibitor concentration. Data represent the mean ± SEM of three technical replicates and are fit to the Hill equation (Camostat R^2 = 0.86; Avoralstat R^2 = 0.95; PCI-27483 R^2 = 0.96; Antipain R^2 = 0.81). **(FIG. 4C)** IC50 values derived from the data in Panel C. Data represent the mean ± SEM of three technical replicates.

[0022] **FIGS. 5A-5C.** Soybean Trypsin Inhibitor (SBTI) inhibits TMPRSS2 activity: Structure of TMPRSS2 in complex with SBTI. **(FIG. 5A)** Model of TMPRSS2 in complex with SBTI generated by HADDOCK. Presented structure is one of the best structures in the biggest size and the best Z-score cluster. HADDOCK score of the cluster is shown in parentheses. **(FIG. 5B)** SBTI Residues at the TMPRSS2/SBTI Interface are shown in magenta stick model (residues 561-566; PYRIRF, and 616-617; GW). Catalytic residues are shown in green, and R563 is the P1 residue. **(FIG. 5C)** Inhibition of Cbz-GGR-AMC (50 μM) hydrolysis by 250 nM TMPRSS2 in the presence of 2-150 μM SBTI (R^2 = 0.76). The initial velocity for each condition was plotted against the inhibitor concentration. Data represent the mean ± SEM of three technical replicates and are fit to the Hill equation.

[0023] **FIGS. 6A-6B.** Avoralstat inhibits cellular TMPRSS2 autoproteolysis: **(FIG. 6A)** HEK 293T cells were treated with 100 μM Camostat or Avoralstat 2 hours before transfection. Cells were transfected with 2 μg of TMPRSS2 (WT) plasmid and either control vector or TMPRSS2-S441A expressing vector which served as negative and positive controls, respectively. Cells were lysed at 24 hours post-transfection and were evaluated by immunoblotting with an anti-FLAG antibody. TMPRSS2 is expressed as a full-length and proteolytically cleaved to form ~53 KDa and 26 KDa fragments. The immunoblot shows concomitant increase in full-length TMPRSS2 levels after Camostat or Avoralstat treatment whereas its levels were significantly reduced in the vehicle (TMPRSS2 WT). Lower cleavage products around 26 KDa or less were also observed in vehicle and Avoralstat group. To confirm equal protein in each lane, membranes were re-probed with β-actin as a loading control. **(FIG. 6B)** Densitometry analysis. Data represent the mean ± SEM from two independent experiments were analyzed by 1-way ANOVA followed by Tukey's multiple comparisons test (**$p<0.01$; n = 6 for each group). NEG, negative control; S441A, positive control.

[0024] **FIGS. 7A-7F.** Avoralstat inhibits viral entry directed by SARS-CoV-2 spike proteins: Calu-3 cells were pre-incubated with the indicated concentrations of DMSO (negative control), Camostat, Avoralstat, PCI-27483, SBTI, and Antipain and subsequently inoculated with pseudovirus particles harboring **(FIG. 7A)** VSV-G or **(FIG. 7B)** SARS-CoV-2 Spike protein. Data represent the mean ± SEM and are fit to the Hill equation (Camostat R^2 = 0.90; Avoralstat R^2 = 0.75; PCI-27483 R^2 = 0.40; SBTI R^2 = 0.38; Antipain R^2 = 0.70). **(FIG. 7C)** Calu-3 cells were treated with medium containing 100 μM of DMSO (negative control), Camostat,

Avoralstat, PCI-27483, SBTI, and Antipain. Cells were then incubated with SARS-CoV-2 (multiplicity of infection, MOI = 0.1) in inhibitor-containing media. Cells were subsequently washed, and viral gRNA was determined by qRT-PCR. Data represent the mean ± SEM and were analyzed by 1-way ANOVA followed by Tukey's multiple comparisons test (****p<0.0001 compared to vehicle). **(FIG. 7D)** Calu-3 cells were pre-incubated with the indicated concentrations of Camostat and Avoralstat and subsequently inoculated with SARS-CoV-2 (multiplicity of infection, MOI = 0.1) in inhibitor-containing media. Cells were subsequently washed, and viral gRNA was determined by qRT-PCR. Data represent the normalized mean ± SEM and are fit to the Hill equation (Camostat R^2 = 0.92; Avoralstat R2 = 0.78). **(FIGS. 7E-7F)** Calu-3 cells were pre-incubated with the indicated concentrations of Camostat and Avoralstat and subsequently inoculated with **(FIG. 7E)** SARS-CoV-2 or **(FIG. 7F)** MERS-CoV (multiplicity of infection, MOI = 0.1) in inhibitor-containing media. Cells were subsequently washed, and viral gRNA was determined by qRT-PCR. Data represent the mean ± SEM were analyzed by 2-way ANOVA followed by Sidak's multiple comparisons test (*p<0.05, **p<0.01, ***p<0.001 compared to vehicle).

[0025] **FIGS. 8A-8B.** Intraperitoneal delivery of Avoralstat reduces SARS-CoV-2 infection in mice: Wild-type BALB/c mice transduced with Ad5-hACE2 were intranasally infected with **(FIG. 8A)** 3 x 10^3 or **(FIG. 8B)** 1 x 10^5 PFU of SARS-CoV-2. Mice were treated with Avoralstat, Camostat (30 mg/kg intraperitoneal injection), or vehicle (DMSO; negative control) four hours before and after being challenged by virus. Virus titers were measured in harvested lungs 1-day post infection. Data are represented as mean ± SEM and were analyzed by 1-way ANOVA followed by Tukey's multiple comparisons test (*p<0.05; **p<0.01).

[0026] **FIGS. 9A-9B.** Model structure of hTMPRSS2. **(FIG. 9A)** Schematic representation of TMPRSS/Hepsin domain architecture. **(FIG. 9B)** Ribbon tracing diagram representing the S1-peptidase domain of human TMPRSS2. The homology model was generated with three separate modeling programs (MODELLER, SWISS-Model, and Phyre2) using the structure of human Hepsin (PDB: 1Z8G) as a template. The catalytic triad residues (H296, D345, and S441) are represented by the yellow stick model.

[0027] **FIGS. 10A-10G.** Comparison of S1-peptidase structures: **(FIG. 10A)** Multiple sequence alignment of representative peptidase domains (SEQ ID NOS:2-7) from the structural phylogenetic analysis. The sequence alignment was generated in MAFFT and visualized using ESpript. **(FIG. 10B)** Overlay between hepsin (PDB 1O5E), **(FIG. 10C)** Factor VII (PDB 1CVW), **(FIG. 10D)** Factor XIV (1AUT), **(FIG. 10E)** CFAI (2XRC), **(FIG. 10F)** Haptoglobin (4F4O) and TMPRSS2 (cyan). The corresponding sequence identity and Ca RMSD is represented below the structural alignment. **(FIG. 10G)** A heatmap representing the sequence and structural similarity to hTMPRSS2 is displayed to the right of the tree. The first

column denotes the pairwise sequence identity (%) to hTMPRSS2 and the second column denotes the structural similarity (1/Ca RMSD) to the hTMPRSS2 model. The last column denotes the structural similarity 'score' which is calculated by dividing the pairwise sequence identity by the backbone RMSD.

[0028] **FIGS. 11A-11C**. *In silico* docking screen of serine protease inhibitors: **(FIG. 11A)** Docking scores of compounds curated based on our 3DPhyloFold Analysis. Interactions between the TMPRSS2 structural model and the docked chemical structures of **(FIG. 11B)** Spike protein S2' site and **(FIG. 11C)** Nafamostat.

[0029] **FIGS. 12A-12C.** Purification and *in vitro* characterization of TMPRSS2: **(FIG. 12A)** Coomassie-stained SDS-PAGE gel of purified recombinant TMPRSS2. The protein was purified using affinity (nickel-NTA) and size-exclusion chromatography. **(FIG. 12B)** Fluorescence tracing of 50 μM Cbz-GGR-AMC (grey) or suc-LY-AMC (blue) in the presence of 250 nM TMPRSS2. Data represent the mean ± SEM of three technical replicates. **(FIG. 12C)** Cbz-GGR-AMC (50 μM) hydrolysis by 250 nM TMPRSS2 in the presence of 10-500 μM Leupeptin (black), MDL-28170 (magenta) and Ritonavir (blue/green). The initial velocity for each condition was plotted against the inhibitor concentration. Data represent the mean ± SEM of three technical replicates.

[0030] **FIGS. 13A-13C.** Soybean trypsin inhibitor (SBTI) derivatives might inhibit TMPRSS2 activity: **(FIG. 13A)** Multiple sequence alignment of serine peptidase domains (SEQ ID NOS:8-11) that were used for the binding pocket analysis. The sequence alignment was generated in MAFFT and visualized using ESPript 3.0. Residues involved in the SBTI binding interface are highlighted in yellow. * - *indicates the residues forming catalytic triad.* **(FIG. 13B)** SBTI Residues at the TMPRSS2/SBTI Interface are shown in magenta stick model (residues 561-566; PYRIRF, and 616-617; GW). Catalytic residues are shown in green, and R563 is the P1 residue. **(FIG. 13C)** Model of porcine Trypsin in complex with SBTI generated by HADDOCK. HADDOCK score of the cluster is shown in parentheses.

DETAILED DESCRIPTION OF THE INVENTION

[0031] Before the present methods and compositions are described, it is to be understood that this invention is not limited to particular method or composition described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0032] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range

between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0033] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, some potential and preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. It is understood that the present disclosure supersedes any disclosure of an incorporated publication to the extent there is a contradiction.

[0034] As will be apparent to those of skill in the art upon reading this disclosure, each of the individual embodiments described and illustrated herein has discrete components and features which may be readily separated from or combined with the features of any of the other several embodiments without departing from the scope or spirit of the present invention. Any recited method can be carried out in the order of events recited or in any other order which is logically possible.

[0035] It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the inhibitor" includes reference to one or more inhibitors and equivalents thereof, e.g. compounds or drugs, known to those skilled in the art, and so forth.

[0036] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

[0037] The term "about", particularly in reference to a given quantity, is meant to encompass deviations of plus or minus five percent.

[0038] By “therapeutically effective dose or amount” of an S1-peptidase inhibitor such as a plasma kallikrein inhibitor (e.g., Avoralstat), a coagulation factor VII inhibitor (e.g., PCI-27483), or a trypsin inhibitor (e.g., SBTI) is intended an amount that, when the S1-peptidase inhibitor

is administered, as described herein, brings about a positive therapeutic response, such as inhibiting an infection by SARS-CoV-2. Additionally, an “effective amount” of an S1-peptidase inhibitor may inhibit TMPRSS2 activity, inhibit viral entry of SARS-CoV-2 into host cells, and/or inhibit replication of SARS-CoV-2.

[0039] "Pharmaceutically acceptable excipient or carrier" refers to an excipient that may optionally be included in compositions that causes no significant adverse toxicological effects to the patient.

[0040] "Pharmaceutically acceptable salt" includes, but is not limited to, amino acid salts, salts prepared with inorganic acids, such as chloride, sulfate, phosphate, diphosphate, bromide, and nitrate salts, or salts prepared from the corresponding inorganic acid form of any of the preceding, e.g., hydrochloride, etc., or salts prepared with an organic acid, such as malate, maleate, fumarate, tartrate, succinate, ethylsuccinate, citrate, acetate, lactate, methanesulfonate, benzoate, ascorbate, para-toluenesulfonate, palmoate, salicylate and stearate, as well as estolate, gluceptate and lactobionate salts. Similarly, salts containing pharmaceutically acceptable cations include, but are not limited to, sodium, potassium, calcium, aluminum, lithium, and ammonium (including substituted ammonium).

[0041] As used herein, the terms "treatment," "treating," and the like, refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. "Treatment," as used herein, covers any treatment of a disease in a mammal, particularly in a human, and includes: (a) increasing survival time; (b) decreasing the risk of death due to the disease; (c) preventing the disease from occurring in a subject which may be predisposed to the disease but has not yet been diagnosed as having it; (d) inhibiting the disease, i.e., arresting its development (e.g., reducing the rate of disease progression); and (e) relieving the disease, i.e., causing regression of the disease.

[0042] "Substantially purified" generally refers to isolation of a substance (e.g., compound, molecule, agent) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample, a substantially purified component comprises 50%, preferably 80%-85%, more preferably 90-95% of the sample.

[0043] The terms "subject," "individual," and "patient," are used interchangeably herein and refer to any mammalian subject for whom diagnosis, prognosis, treatment, or therapy is desired, particularly humans. "Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs, horses, cats, cows, sheep, goats, pigs, etc. In some cases, the methods of the invention find use in experimental animals, in veterinary application, and in the

development of animal models for disease, including, but not limited to, rodents including mice, rats, and hamsters; primates, and transgenic animals.

Methods

[0044] Methods for treating coronavirus disease 19 (COVID-19) with inhibitors of transmembrane serine protease 2 (TMPRSS2) are disclosed. Without being bound by a particular theory, TMPRSS2 is expressed in the host pulmonary, endocrine, gastrointestinal (GI), and genitourinary system and facilitates viral entry into cells by cleaving the spike glycoprotein, which is found on the surface of all coronaviruses. A structural phylogenetic analysis revealed that the peptidase catalytic sites of plasma kallikrein, coagulation factor XI, and trypsin share structural similarity with human TMPRSS2 (see Examples). Thus, inhibitors of plasma kallikrein, coagulation factor XI, and trypsin may be useful in treating COVID-19.

[0045] Inhibition of TMPRSS2 by inhibitors of plasma kallikrein, coagulation factor XI, and trypsin may be complete or partial (i.e., all activity, some activity, or most activity is blocked by an inhibitor). For example, an inhibitor may reduce the activity of TMPRSS2 by 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or any amount in between as compared to native or control levels. The plasma kallikrein inhibitor, Avoralstat, the coagulation factor VII inhibitor, PCI-27483, and the trypsin inhibitor SBTI have been shown to inhibit human TMPRSS2.

Formulations

[0046] Plasma kallikrein inhibitors (e.g., Avoralstat), coagulation factor VII inhibitors (e.g., PCI-27483), and trypsin inhibitors (e.g., SBTI) that inhibit TMPRSS2 can be incorporated into a variety of formulations. More particularly, these inhibitors may be formulated into pharmaceutical compositions by combination with appropriate pharmaceutically acceptable carriers or diluents. Pharmaceutical preparations are compositions that include one or more plasma kallikrein inhibitors (e.g., Avoralstat), coagulation factor VII inhibitors (e.g., PCI-27483), or trypsin inhibitors (e.g., SBTI) in a pharmaceutically acceptable vehicle. "Pharmaceutically acceptable vehicles" may be vehicles approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in mammals, such as humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is formulated for administration to a mammal. Such pharmaceutical vehicles can be lipids, e.g., liposomes, e.g., liposome dendrimers; liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like, saline; gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and

the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. Pharmaceutical compositions may be formulated into preparations in solid, semi-solid, liquid or gaseous forms, such as tablets, capsules, powders, granules, ointments, solutions, suppositories, injections, inhalants, gels, microspheres, and aerosols. As such, administration of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) can be achieved in various ways, including, without limitation, pulmonary, oral, intravenous, buccal, rectal, parenteral, intraperitoneal, transdermal, intradermal, or intracheal administration. The active agent may be systemic after administration or may be localized by the use of regional administration or use of an implant that acts to retain the active dose at the site of implantation. The active agent may be formulated for immediate activity or it may be formulated for sustained release.

[0047] For inclusion in a medicament, the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) may be obtained from a suitable commercial source. As a general proposition, the total pharmaceutically effective amount of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) administered parenterally per dose will be in a range that can be measured by a dose response curve.

[0048] For inhibitor-based therapies with a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), i.e., preparations to be used for therapeutic administration, may be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 μm membranes). Therapeutic compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle. The compositions comprising a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) may be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-mL vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous solution of compound, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized compound using bacteriostatic water-for-Injection. Alternatively, the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) may be formulated into lotions for topical administration.

[0049] Pharmaceutical compositions can include, depending on the formulation desired, pharmaceutically acceptable, non-toxic carriers of diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples

of such diluents are distilled water, buffered water, physiological saline, PBS, Ringer's solution, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation can include other carriers, adjuvants, or non-toxic, nontherapeutic, nonimmunogenic stabilizers, excipients and the like. The compositions can also include additional substances to approximate physiological conditions, such as pH adjusting and buffering agents, toxicity adjusting agents, wetting agents and detergents.

[0050] The composition can also include any of a variety of stabilizing agents, such as an antioxidant for example. When the pharmaceutical composition includes a polypeptide, the polypeptide can be complexed with various well-known compounds that enhance the *in vivo* stability of the polypeptide, or otherwise enhance its pharmacological properties (e.g., increase the half-life of the polypeptide, reduce its toxicity, enhance solubility or uptake). Examples of such modifications or complexing agents include sulfate, gluconate, citrate and phosphate. The nucleic acids or polypeptides of a composition can also be complexed with molecules that enhance their *in vivo* attributes. Such molecules include, for example, carbohydrates, polyamines, amino acids, other peptides, ions (e.g., sodium, potassium, calcium, magnesium, manganese), and lipids.

[0051] Further guidance regarding formulations that are suitable for various types of administration can be found in Remington's Pharmaceutical Sciences, Mace Publishing Company, Philadelphia, Pa., 17th ed. (1985). For a brief review of methods for drug delivery, see, Langer, Science 249:1527-1533 (1990).

[0052] The pharmaceutical compositions can be administered for prophylactic and/or therapeutic treatments. Toxicity and therapeutic efficacy of the active ingredient can be determined according to standard pharmaceutical procedures in cell cultures and/or experimental animals, including, for example, determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Therapies that exhibit large therapeutic indices are preferred.

[0053] The data obtained from cell culture and/or animal studies can be used in formulating a range of dosages for humans. The dosage of the active ingredient typically lines within a range of circulating concentrations that include the ED50 with low toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized.

[0054] The components used to formulate the pharmaceutical compositions are preferably of high purity and are substantially free of potentially harmful contaminants (e.g., at least National Food (NF) grade, generally at least analytical grade, and more typically at least

pharmaceutical grade). Moreover, compositions intended for *in vivo* use are usually sterile. To the extent that a given compound must be synthesized prior to use, the resulting product is typically substantially free of any potentially toxic agents, particularly any endotoxins, which may be present during the synthesis or purification process. Compositions for parental administration are also sterile, substantially isotonic and made under GMP conditions.

[0055] The plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) may be provided in addition to other agents. For example, in methods of treating an infection by SARS-CoV-2, a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) may be coadministered with other known antiviral therapies. For example, a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) may be administered in combination with monoclonal or polyclonal antibodies against the spike protein or a vaccine against SARS-CoV-2.

Administration

[0056] At least one therapeutically effective dose of a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and/or optionally other antiviral agents will be administered. By "therapeutically effective dose or amount" of a plasma kallikrein inhibitor (e.g., Avoralstat), a coagulation factor VII inhibitor (e.g., PCI-27483), or a trypsin inhibitor (e.g., SBTI) is intended an amount that, when the inhibitor is administered, as described herein, brings about a positive therapeutic response, such as inhibiting an infection by SARS-CoV-2. Additionally, an "effective amount" of an S1-peptidase inhibitor may inhibit TMPRSS2 activity, inhibit cleavage of the SARS-COV-2 spike protein, inhibit viral entry of SARS-CoV-2 into host cells, and/or inhibit replication of SARS-CoV-2.

[0057] In certain embodiments, multiple therapeutically effective doses of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and optionally other antiviral agents will be administered according to a daily dosing regimen, or intermittently. For example, a therapeutically effective dose can be administered, one day a week, two days a week, three days a week, four days a week, or five days a week, and so forth. By "intermittent" administration is intended the therapeutically effective dose can be administered, for example, every other day, every two days, every three days, and so forth. For example, in some embodiments, the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and/or optionally other antiviral agents will be administered twice-weekly or thrice-weekly for an extended period of time, such as for 1, 2, 3, 4, 5, 6, 7, 8...10...15…24 weeks, and so forth.

By “twice-weekly” or “two times per week” is intended that two therapeutically effective doses of the agent in question is administered to the subject within a 7 day period, beginning on day 1 of the first week of administration, with a minimum of 72 hours, between doses and a maximum of 96 hours between doses. By “thrice weekly” or “three times per week” is intended that three therapeutically effective doses are administered to the subject within a 7 day period, allowing for a minimum of 48 hours between doses and a maximum of 72 hours between doses. For purposes of the present invention, this type of dosing is referred to as “intermittent” therapy. In accordance with the methods of the present invention, a subject can receive intermittent therapy (i.e., twice-weekly or thrice-weekly administration of a therapeutically effective dose) for one or more weekly cycles until the desired therapeutic response is achieved. The agents can be administered by any acceptable route of administration as noted herein below.

[0058] In certain embodiments, combination therapy with a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), and/or trypsin inhibitor (e.g., SBTI), and optionally other antiviral agents is administered. The plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) can be administered prior to, concurrent with, or subsequent to an antiviral agent. If provided at the same time as the antiviral agent, the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), can be provided in the same or in a different composition. Thus, the two agents can be presented to the individual by way of concurrent therapy. By “concurrent therapy” is intended administration to a human subject such that the therapeutic effect of the combination of the substances is caused in the subject undergoing therapy. For example, concurrent therapy may be achieved by administering at least one therapeutically effective dose of a pharmaceutical composition comprising a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and at least one therapeutically effective dose of a pharmaceutical composition comprising at least one an antiviral agent according to a particular dosing regimen. Similarly, the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and optionally other antiviral agents, can be administered in at least one therapeutic dose. Administration of the separate pharmaceutical compositions can be at the same time (i.e., simultaneously) or at different times (i.e., sequentially, in either order, on the same day, or on different days), as long as the therapeutic effect of the combination of these substances is caused in the subject undergoing therapy.

[0059] In other embodiments, the pharmaceutical compositions comprising the agents, such as the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-

27483), or trypsin inhibitor (e.g., SBTI), and/or optionally other antiviral agents, is a sustained-release formulation, or a formulation that is administered using a sustained-release device. Such devices are well known in the art, and include, for example, transdermal patches, and miniature implantable pumps that can provide for drug delivery over time in a continuous, steady-state fashion at a variety of doses to achieve a sustained-release effect with a non-sustained-release pharmaceutical composition.

[0060] The pharmaceutical compositions comprising the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and optionally other anti-viral agents may be administered using the same or different routes of administration in accordance with any medically acceptable method known in the art. Suitable routes of administration include parenteral administration, such as subcutaneous (SC), intraperitoneal (IP), intramuscular (IM), intravenous (IV), or infusion, oral, pulmonary, nasal, topical, transdermal, and suppositories. Where the composition is administered via pulmonary delivery, the therapeutically effective dose is adjusted such that the soluble level of the agent, such as the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) in the bloodstream, is equivalent to that obtained with a therapeutically effective dose that is administered parenterally, for example SC, IP, IM, or IV. In some embodiments, pharmaceutical compositions comprising the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and optionally other antiviral agents are administered locally to the lungs. In some embodiments, the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and optionally other antiviral agents are administered topically such as on a patch or in a gel.

[0061] In some embodiments, the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and optionally other antiviral agents are administered by infusion or by local injection, e.g. by infusion at a rate of about 50 mg/h to about 400 mg/h, including about 75 mg/h to about 375 mg/h, about 100 mg/h to about 350 mg/h, about 150 mg/h to about 350 mg/h, about 200 mg/h to about 300 mg/h, about 225 mg/h to about 275 mg/h. Exemplary rates of infusion can achieve a desired therapeutic dose of, for example, about 0.5 mg/m^2/day to about 10 mg/m^2/day, including about 1 mg/m^2/day to about 9 mg/m^2/day, about 2 mg/m^2/day to about 8 mg/m^2/day, about 3 mg/m^2/day to about 7 mg/m^2/day, about 4 mg/m^2/day to about 6 mg/m^2/day, about 4.5 mg/m^2/day to about 5.5 mg/m^2/day. Administration (e.g., by infusion) can be repeated over a desired period, e.g., repeated over a period of about 1 day to about 5 days or once every several days, for example, about five days, over about 1 month, about 2 months, etc. The plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483),

or trypsin inhibitor (e.g., SBTI) also can be administered prior, at the time of, or after other therapeutic interventions, such as administration of antibodies against the SARS-COV-2 spike protein. The plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), can also be administered as part of a combination therapy, in which at least one antiviral agent is administered to the subject.

[0062] Factors influencing the respective amount of the various compositions to be administered include, but are not limited to, the mode of administration, the frequency of administration (i.e., daily, or intermittent administration, such as twice- or thrice-weekly), the particular disease undergoing therapy, the severity of the disease, the history of the disease, whether the individual is undergoing concurrent therapy with another therapeutic agent, and the age, height, weight, health, and physical condition of the individual undergoing therapy. Generally, a higher dosage of this agent is preferred with increasing weight of the subject undergoing therapy.

[0063] Individual doses of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and optionally other antiviral agents are typically not less than an amount required to produce a measurable effect on the subject, and may be determined based on the pharmacokinetics and pharmacology for absorption, distribution, metabolism, and excretion ("ADME") of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and optionally other antiviral agents and their by-products, and thus based on the disposition of the compositions within the subject. This includes consideration of the route of administration as well as dosage amount, which can be adjusted for topical (applied directly where action is desired for mainly a local effect), enteral (applied via digestive tract for systemic or local effects when retained in part of the digestive tract), or parenteral (applied by routes other than the digestive tract for systemic or local effects) applications. For instance, administration of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) may be orally or via injection, e.g., intravenous, or intramuscular, or by pulmonary delivery (e.g., with a nebulizer, metered dose inhaler (MDI), or dry powder inhaler (DPI)), or a combination thereof.

[0064] Disposition of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) and its corresponding biological activity within a subject is typically gauged against the fraction of the inhibitor present at a target of interest. For example, a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) once administered can accumulate with a glycoconjugate or other biological target that concentrates the material in infected lung cells and lung tissue. Thus, dosing regimens in which the plasma kallikrein

inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) is administered so as to accumulate in a target of interest over time can be part of a strategy to allow for lower individual doses. This can also mean that, for example, the dose of a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI)that is cleared more slowly *in vivo* can be lowered relative to the effective concentration calculated from *in vitro* assays (e.g., effective amount *in vitro* approximates mM concentration, versus less than mM concentrations *in vivo*).

[0065] As an example, the effective amount of a dose or dosing regimen can be gauged from the IC_{50} of a given plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) for inhibiting TMPRSS2 activity. By "IC_{50}" is intended the concentration of a drug required for 50% inhibition *in vitro*. Alternatively, the effective amount can be gauged from the EC_{50} of a given plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) concentration. By "EC_{50}" is intended the plasma concentration required for obtaining 50% of a maximum effect *in vivo*. In related embodiments, dosage may also be determined based on ED_{50} (effective dosage).

[0066] In general, an effective amount is usually not more than 200X the calculated IC_{50}. Typically, the amount of a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) that is administered is less than about 200X, less than about 150X, less than about 100X and many embodiments less than about 75X, less than about 60X, 50X, 45X, 40X, 35X, 30X, 25X, 20X, 15X, 10X and even less than about 8X or 2X than the calculated IC_{50}. In one embodiment, the effective amount is about 1X to 50X of the calculated IC_{50}, and sometimes about 2X to 40X, about 3X to 30X or about 4X to 20X of the calculated IC_{50}. In other embodiments, the effective amount is the same as the calculated IC_{50}, and in certain embodiments the effective amount is an amount that is more than the calculated IC_{50}.

[0067] An effect amount will typically not be more than 100X the calculated EC_{50}. For instance, the amount of an inhibitor that is administered is less than about 100X, less than about 50X, less than about 40X, 35X, 30X, or 25X and many embodiments less than about 20X, less than about 15X and even less than about 10X, 9X, 9X, 7X, 6X, 5X, 4X, 3X, 2X or 1X than the calculated EC_{50}. The effective amount may be about 1X to 30X of the calculated EC_{50}, and sometimes about 1X to 20X, or about 1X to 10X of the calculated EC_{50}. The effective amount may also be the same as the calculated EC_{50} or more than the calculated EC_{50}. The IC_{50} can be calculated by inhibiting TMPRSS2 protease activity and/or cleavage of the SARS-CoV-2 spike protein.

[0068] In order to achieve efficacy, the level of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) must be above a specific level for a specific time. Efficacy is dose dependent and higher levels of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) contribute to greater antiviral effects. In order to minimize toxicity, the level of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) may be maintained below a certain level within a specific time and for a specific time (a "rest period" allows clearance of the inhibitor). That is, the drug is kept below a certain level by a certain time before the next dose is given. Shorter rests between doses contribute to greater toxicity.

[0069] In certain embodiments, the method of treatment of a patient having an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) comprises a treatment cycle with a plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI), and/or optionally other antiviral agents followed by a rest period in which no plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) is administered to allow the patient to "recover" from the undesirable effects of the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI). Multiple doses of plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), or trypsin inhibitor (e.g., SBTI) can be administered according to a daily dosing regimen or intermittently, followed by a rest period.

Kits

[0070] Kits are provided comprising one or more containers holding compositions comprising at least one plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), and/or trypsin inhibitor (e.g., SBTI), and/or optionally one or more other antiviral agents for treating an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Compositions can be in liquid form or can be lyophilized. Suitable containers for the compositions include, for example, bottles, vials, syringes, and test tubes. Containers can be formed from a variety of materials, including glass or plastic. A container may have a sterile access port (for example, the container may be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle).

[0071] The kit can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as phosphate-buffered saline, Ringer's solution, or dextrose solution. It can also contain other materials useful to the end-user, including other pharmaceutically acceptable formulating solutions such as buffers, diluents, filters, needles, syringes, devices

for pulmonary delivery (nebulizer, metered dose inhaler (MDI), or dry powder inhaler), or other delivery devices. The delivery device may be pre-filled with the compositions.

[0072] The kit can also comprise a package insert containing written instructions for methods of using the compositions comprising the plasma kallikrein inhibitor (e.g., Avoralstat), coagulation factor VII inhibitor (e.g., PCI-27483), and/or trypsin inhibitor (e.g., SBTI) for treating a subject for an infection by SARS-CoV-2. The package insert can be an unapproved draft package insert or can be a package insert approved by the Food and Drug Administration (FDA) or other regulatory body. Alternatively, instructions may be provided on a computer readable medium, e.g., diskette, CD, DVD, flash drive, etc., on which the information has been recorded, or the instructions may be presented at a website address, which may be used via the internet to access the information at a removed site. Any convenient means for providing instructions for treating a subject for an infection by SARS-CoV-2 may be present in the kits.

Examples of Non-Limiting Aspects of the Disclosure

[0073] Aspects, including embodiments, of the present subject matter described above may be beneficial alone or in combination, with one or more other aspects or embodiments. Without limiting the foregoing description, certain non-limiting aspects of the disclosure numbered 1-43 are provided below. As will be apparent to those of skill in the art upon reading this disclosure, each of the individually numbered aspects may be used or combined with any of the preceding or following individually numbered aspects. This is intended to provide support for all such combinations of aspects and is not limited to combinations of aspects explicitly provided below:

1. A method of treating a subject for an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the method comprising administering to the subject a therapeutically effective amount of a composition comprising a plasma kallikrein inhibitor, a coagulation factor VII inhibitor, or a trypsin inhibitor.

2. The method of aspect 1, wherein the plasma kallikrein inhibitor is Avoralstat.

3. The method of aspect 1, wherein the coagulation factor VII inhibitor is PCI-27483.

4. The method of aspect 1, wherein the trypsin inhibitor is soybean trypsin inhibitor (SBTI).

5. The method of any one of aspects 1 to 4, wherein the composition is administered prophylactically or therapeutically.

6. The method of any one of aspects 1 to 5, wherein multiple cycles of treatment are administered to the subject for a time period sufficient to eradicate the infection by the SARS-CoV-2.

7. The method of any one of aspects 1 to 6, wherein the composition is administered according to a daily dosing regimen or intermittently.

8. The method of any one of aspects 1 to 7, wherein the composition is administered orally, intravenously, or by pulmonary inhalation.

9. The method of aspect 8, wherein pulmonary inhalation is performed with a nebulizer, a metered dose inhaler (MDI), and a dry powder inhaler (DPI).

10. The method of any one of aspects 1 to 9, further comprising administering additional anti-viral therapy.

11. The method of any one of aspects 1 to 10, wherein the subject is mammalian.

12. The method of aspect 11, wherein the subject is human.

13. The method of any one of aspects 1 to 12, wherein the composition is administered in an amount sufficient to reduce viral entry or replication of the SARS-CoV-2 in the subject.

14. A composition comprising a plasma kallikrein inhibitor, a coagulation factor VII inhibitor, or a trypsin inhibitor for use in the treatment of an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

15. The composition of aspect 14, further comprising a pharmaceutically acceptable excipient.

16. The composition of aspect 14 or 15, wherein the plasma kallikrein inhibitor is Avoralstat.

17. The composition of aspect 14 or 15, wherein the coagulation factor VII inhibitor is PCI-27483.

18. The composition of aspect 14 or 15, wherein the trypsin inhibitor is soybean trypsin inhibitor (SBTI).

19. A method of screening for an S1 peptidase inhibitor that inhibits an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the method comprising:

a) modeling a TMPRSS2 S1 peptidase domain by using a structure of hepsin as a template;
b) calculating a TMPRSS2 structural similarity score for a plurality of S1 peptidases, wherein S1 peptidases with the highest structural similarity scores are prioritized for further *in silico* analysis;
c) modeling a TMPRSS2 binding pocket by comparison to a structure of a hepsin binding pocket;
d) docking a plurality of inhibitors of the prioritized S1 peptidases into the modeled TMPRSS2 binding pocket, wherein the inhibitors are prioritized for further screening based on their docking scores, wherein higher docking scores indicate more favorable binding interactions in the TMPRSS2 binding pocket;
e) measuring inhibition of TMPRSS2 activity by one or more of the inhibitors selected for further screening, wherein inhibitors are prioritized according to how strongly they inhibit TMPRSS2; and
f) measuring inhibition of an infection by the SARS-CoV-2 in a host cell or subject by one or more of the inhibitors that inhibited TMPRSS2 to identify an inhibitor that inhibits the infection by the SARS-CoV-2.

20. The method of aspect 19, wherein the inhibitor has been previously approved by the U.S. Food and Drug Administration for treatment of a disease or condition other than the SARS-CoV-2 infection.

21. An inhibitor identified by the method of aspect 19 or 20 for the treatment of a SARS-CoV-2 infection.

EXAMPLES

[0074] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

[0075] General methods in molecular and cellular biochemistry can be found in such standard textbooks as Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook et al., HaRBor Laboratory Press 2001); Short Protocols in Molecular Biology, 4th Ed. (Ausubel et al. eds., John Wiley & Sons 1999); Protein Methods (Bollag et al., John Wiley & Sons 1996); Nonviral Vectors for Gene Therapy (Wagner et al. eds., Academic Press 1999); Viral Vectors (Kaplift & Loewy eds., Academic Press 1995); Immunology Methods Manual (I. Lefkovits ed., Academic Press 1997); and Cell and Tissue Culture: Laboratory Procedures in Biotechnology (Doyle & Griffiths, John Wiley & Sons 1998), the disclosures of which are incorporated herein by reference.

Example 1

Structural phylogenetic analysis of the TMPRSS2 protease predicts novel therapeutic strategies for SARS-CoV-2

Introduction

[0076] There is a critical need for COVID-19 prophylaxis and treatment in the early stages of disease for high-risk patients, front-line workers, and patients showing symptoms for more than a month (a.k.a “long haulers”) as the outbreak continues and resurges. Although vaccines are in development, it’s not clear they will be sufficiently effective (Lv et al., 2020). Adjunctive therapies that mitigate viral entry or replication in humans have extraordinary value to attenuate viral spread by asymptomatic or early-stage patients, and act as a safety net for vulnerable populations and primary workers at high-risk for infection. In response to the urgent need for therapeutics, there is active investigation into repurposing existing drugs. Many potential repurposed treatments target viral proteins directly, such as the main proteinase (3CLpro), papain-like protease (PLpro), and RNA-dependent RNA polymerase (RdRp), and include Lopinavir, Ritonavir, Remdesivir, and Ribavirin, which inhibit SARS-CoV-2 *in vitro*

(Baron et al., 2020; Gao et al., 2020; Zhou et al., 2020a). A less explored alternative strategy is to target human host proteins utilized by viruses. Even with a vaccine or a specific antiviral SARS-CoV-2 drug, combination therapies could be synergistic. With the genetic variation in SARS-CoV-2 strains and genetic evolution of corona and other viruses, targeting the key host molecules required for infection could be a more durable, long-term strategy for SARS-CoV prophylaxis, rather than targeting individual viral strains.

[0077] Transmembrane serine protease 2 (TMPRSS2), Angiotensin-converting enzyme 2 (ACE2), and Furin are the three host proteases that facilitate SARS-CoV-2 entry into host pneumocytes, explaining the predominantly respiratory symptoms caused by the virus (Morse et al., 2020; Zhou et al., 2020b). TMPRSS2, a member of the TMPRSS/Hepsin subfamily of serine proteases, contains a catalytically active extracellular domain and is expressed in the host pulmonary, endocrine, gastrointestinal (GI), genitourinary system (Hoffmann et al., 2020; Lin et al., 1999; Park, 2010; Uhlen et al., 2015). It facilitates viral entry into the host cell by serving as a priming protease for the spike (S) glycoprotein, targeting a specific peptide motif (termed the S2' site), which is found on the surface of all coronaviruses (Iwata-Yoshikawa et al., 2019; Kawase et al., 2012). The fact that TMPRSS2 is essential for viral infection, but unlike the other two host proteases, TMPRSS2 is not required for development or homeostasis, makes it an attractive drug target (Kim et al., 2006). Viral use of TMPRSS2 seems to vary by strain, and it has 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 being evaluated for COVID-19 treatment in humans (Shirato et al., 2013; Søgaard, 2020) (Clinical trial NCT04321096). Camostat's eight-minute plasma half-life and poor bioavailability necessitate intravenous administration, thereby preventing its widespread prophylactic use, and the safety and efficacy of Camostat for COVID-19 treatment is not known (Hoffmann et al., 2020; Shen et al., 2017). Thus, identification of additional serine protease inhibitors with improved pharmocokinetic properties that target TMPRSS2 remains imperative.

[0078] Conventional methods for identifying candidates for drug repurposing typically employ high-throughput screening (HTS) or *in silico* screening using libraries of compounds previously tested in humans (Talevi and Bellera, 2020). Although HTS methods can rapidly screen thousands of compounds, there are certain limitations. HTS methods utilize only a few, generalized experimental conditions with technical limitations, such as narrow dose range and experimental conditions, which may not account for the unique features and needs of each compound. This can lead to false positives and negatives. While false positives are filtered out in subsequent experiments, false negatives may overlook valuable inhibitory compounds. Because HTS uses a shotgun rather than a hypothesis driven approach, it may difficult to

ascertain the mechanism-of-action, and this may slow the downstream development of candidate drugs into human therapies. Thus, hypothesis-driven screening methods, utilizing protein structures and a limited number of compounds, remains a valuable and complementary strategy for drug-repurposing. Since there are limited biochemical assays for TMPRSS2, and there is no high-resolution crystal structure of TMPRSS2, it has not undergone extensive HTS screening.

[0079] A rational-based drug repurposing approach is to identify similar target proteins with preexisting drugs. Similar proteins are usually identified using primary sequence homology, but it is well-known that with this criteria, similar proteins can behave very differently. This is because their three-dimensional (3D) shape can vary significantly, since seemingly minor amino acid differences can confer dramatic functional changes in protein folding. While highly informative, primary sequence alignments often miss subtle but critical differences between protein structures due to the complexity of modeling interactions between distant residues in the primary structure that may be close in 3D space. In the case of serine proteases, despite their high sequence diversity, they display similarity in the 3D structure of their S1-peptidase domains (Page and Di Cera, 2008). This structural similarity has presented challenges in the identification of selective therapeutic inhibitors but sets up an opportunity for drug repurposing. We hypothesized that repurposed drugs selected from proteins related by their tertiary structure, rather than their primary structure, were likely to interact with TMPRSS2. We created a structure-based phylogenetic computational tool we call 3DPhyloFold to examine the functional diversity among protease structures and to systematically compare large numbers of protein structures (Velez et al., 2020). Here, we utilized this computational approach to identify structurally similar serine proteases with known therapeutic inhibitors, selected and tested drug repositioning candidates for TMPRSS2, and demonstrated effective inhibition of SARS-CoV-2 infection *in vitro* and *in vivo*. Our findings point to Avoralstat as a potential oral early or prophylactic therapy for COVID-19.

Results

Structure-based phylogeny identifies proteases that are structurally similar to TMPRSS2

[0080] To identify S1 peptidase domains related to TMPRSS2 by sequence homology, we aligned a seed group of manually curated sequences from transmembrane serine proteases that share the S1-peptidase fold and produced a profile hidden Markov model (HMM) (Finn et al., 2011). The profile HMM was then used in a HMMER search against the UniProt database to broaden the pool of S1-peptidase sequences to six hundred, which were then aligned. The

enriched multiple sequence alignment (MSA) subsequently underwent rigorous phylogenetic reconstruction (Nguyen et al., 2015). The reconstructed phylogenetic tree divided the six-hundred member S1-peptidase superfamily into seven main clades with high bootstrap values (**FIG. 1A**). TMPRSS and Hepsin sequences fell into the largest clade, which could be further divided into four separate clades. The TMPRSS/Hepsin clade (bootstrap 82%; 19 sequences) contained canonical TMPRSS members as well as proteases outside of the TMPRSS subfamily: Coagulation Factor XI and Plasma Kallikrein (Kallikrein-related B1 [KLKB1]; **FIGS. 1A-1B**; **FIG. 9A**). These results suggest that these proteins would also be structurally similar to TMPRSS2. A homology-based model of the human TMPRSS2 S1-peptidase domain was generated based off the structure of human Hepsin (PDB 1Z8G) using three different modeling programs (**FIG. 9**) (Eswar et al., 2006; Kelley et al., 2015; Schwede et al., 2003). This template was chosen because it was the closest in terms of sequence homology to TMPRSS2. The sequence alignment data was projected onto the TMPRSS2 model using ConSurf (Armon et al., 2001; Ashkenazy et al., 2016; Glaser et al., 2003). As expected, residues comprising TMPRSS2 hydrophobic core and active site displayed high evolutionary conservation scores and were among the 35 residues that were highly conserved in the MSA (**FIG. 1C**; **Table 1**).

[0081] The ability to determine subtle structural similarities between S1-peptidases would aid in the identification of TMPRSS2 inhibitors. So, we applied a structure-based phylogenetic approach (3DPhyloFold) to examine the 3D relationship among S1-peptidase structures. We first searched the Pfam database for structures of mammalian peptidases and selected 74 representative structures (one structure per unique protein; **Table 2**) (Finn et al., 2014). Structures (with reflection data deposited in the PDB) were evaluated by their reported global validation metrics in PDB-REDO (Joosten et al., 2009) and re-refined structural models were used for further analysis. We compared the primary and tertiary structures of 74 representative S1-peptidase domains to TMPRSS2 (**FIG. 2A**) (Finn et al., 2014). The 75 sequences (primary structures) were aligned and underwent conventional phylogenetic analysis as described above. TMPRSS2 clustered closely with Plasma Kallikrein (Kallikrein-related B1 [KLKB1]), Coagulation Factor XI (Factor XI), and Complement Factor I (CFAI), for example (**FIG. 2B**). The Kallikrein- and Trypsin-like clades clustered further away, suggesting that TMPRSS2 is sequentially divergent from these subfamilies of proteases (**FIG. 2B**). In 3DPhyloFold, pairwise structural comparisons of the TMPRSS2 model and the 74 representative structures were used to calculate a structural dissimilarity matrix (SDM) based on the root mean square deviation (RMSD) between protein alpha carbon (Ca) atoms (**FIG. 2A**). A structure-based phylogenetic tree was then generated from this SDM matrix using the UPGMA method (Sneath and Sokal, 1973). Interestingly, while some proteins remained closely related, the clustering of the structure-based tree was distinct to that of the sequence-based tree.

Previously close proteins CFAI and CTRB1, for example, were much further in the structure-tree (Ca RMSD = 1.38 Å and 0.77 Å, respectively; **FIG. 10**). Although distant in the sequence-based tree, the Trypsin-like clade clustered and Factor XII, for example, moved much closer to TMPRSS2 in our structure-based tree (**FIG. 2C**). This suggests that, while divergent in sequence, TMPRSS2 adopts a closer three-dimensional fold to Trypsin. The proteases with the highest structural similarity were Hepsin (Ca RMSD = 0.24 Å), Acrosin (Ca RMSD = 0.48 Å), Trypsin (-1, - 2, -B1, and -B2; Ca RMSD = 0.63 – 0.64 Å), Coagulation Factor-VII (Ca RMSD = 0.64 Å) and Factor -XI (Ca RMSD = 0.63 Å), and Plasma Kallikrein (Ca RMSD = 0.63 Å). The 3DPhyloFold computational approach identified two sequentially divergent proteases not considered by primary sequence methods, and prioritized proteins with 3D similarity to TMPRSS2 for further *in silico* analysis.

Evaluation of repositioned protease inhibitors based on similar structures

[0082] We searched the literature for known small-molecule and peptidomimetic inhibitors of Hepsin, Acrosin, Trypsin, Factor VII, Factor XI, and Plasma Kallikrein (see Methods). It was previously reported that S1-peptidases can be inhibited by compounds containing a 4-amindinobenzylamide moiety that mimics the chemical properties of basic amino acids (Schweinitz et al., 2004). We therefore focused our search on inhibitors containing an amidine, guanidine or structurally related groups. A key biochemical feature for the substrate specificity of the S1-peptidase domain is a basic amino acid (i.e. arginine or lysine) at the P1 residue (position 1; the first N-terminal residue from the cleavage site) of natural substrates. The basic residue forms strong hydrogen bonding interactions with the aspartate carboxylate that resides within the sub-pocket (S1 pocket; **FIG. 3A**). This search resulted in a curated list of ninety-four inhibitors developed for Hepsin, Acrosin, Trypsin, Factor VII, Factor XI, and Plasma Kallikrein, including four small molecules tested in human clinical trials and ninety experimental compounds only tested *in vitro* (**Table 3**).

[0083] To determine the interaction of the potentially repurposable inhibitors with TMPRSS2, we performed *in silico* docking experiments. first benchmarked using *in silico* docking calculations. To benchmark the docking calculations, the peptide sequence at the S2' site (KPSKRSF) within the SARS-CoV-2 spike glycoprotein (SARS-CoV-2 S) that is cleaved by TMPRSS2 was analyzed using the Glide Docking Tool in Schrödinger (Repasky et al., 2007) and HADDOCK (de Vries et al., 2010). Lower docking scores from this analysis (i.e. Glide Score and HADDOCK Score) indicate a more favorable 'ligand-receptor' binding interaction (Friesner et al., 2004; Friesner et al., 2006; Halgren et al., 2004). The SARS-CoV-2 S2' peptide docked well to our TMPRSS2 model and occupied the binding pocket (Glide score = -9.7; HADDOCK score = -97; **FIG. 3B**; **FIG. 11**; **Table 4**), providing a semi-quantitative benchmark

for favorable TMPRSS2 binding. In addition to the SARS-CoV-2 S2' spike protein motif, we docked the known serine protease inhibitor, Camostat, that is being tested for efficacy against COVID-19 (Clinical Trial NCT04353284). Camostat showed a less favorable interactions with TMPRSS2 (Glide score = -4.7; HADDOCK score = -52; **FIGS. 3B-3C**; **Table 4**).

[0084] We next evaluated the other compounds and observed a wide range of docking scores (0 to -11.66) (**FIG. 11A**), yet some clinically tested compounds showed favorable docking scores exceeding the Camostat score (**FIG. 3B**). Among seven known Plasma Kallikrein (KLKB1) inhibitors (**Table 4**), the top scoring inhibitor was Avoralstat (Glide score = -8.4; HADDOCK score = -71; **FIGS. 3B, FIG. 3D**), which completed a Phase 3 clinical trial for hereditary angioedema (HAE) (Cornpropst et al., 2016). We examined 20 Factor VII inhibitors. One Factor VII inhibitor, PCI-27483, docked well to TMPRSS2 (Glide score = -8.8; HADDOCK score = -106; **FIGS. 3B, 3E**). and completed a Phase 2 clinical trial as a combination treatment with Gemcitabine for advanced pancreatic cancer (Ramanathan et al., 2019). We also evaluated Antipain, a broad protease inhibitor reported to inhibit Trypsin and Coagulation Factors *in vitro* (Suda et al., 1972). Antipain displayed a favorable docking score (Glide score = -9.8; HADDOCK score = -75; **FIG. 3B**; **Table 4**) comparable to other top scoring compounds. Although Antipain is not directly amenable to repurposing because of its toxicity in humans, its high docking score validated the bioinformatic pipeline and suggests it could be used to investigate TMPRSS2 *in vitro* or serve as a modifiable lead compound.

[0085] To validate *in silico* docking, we created an *in vitro* biochemical assay to measure TMPRSS2 proteolytic activity and focused on repurposable human drugs. The recombinant peptidase domain of TMRPSS2 (residues 252 - 489) was expressed and purified from *E. coli*. TMPRSS2 activity was tested using fluorescence resonance energy transfer (FRET) assays measuring the hydrolysis of synthetic oligopeptides. When assayed with the serine protease urokinase substrate Cbz-GGR-AMC (Cbz-Gly-Gly-L-Arg-4-methylcoumaryl-7-amide), TMPRSS2 was catalytically active (**FIG. 12A**). There was no activity as expected with the cysteine protease substrate sLY-AMC (N-succinyl-Leu-Tyr-7-amido-4-methylcoumarin; negative control). An enzyme titration in the presence of 50 μM Cbz-GGR-AMC revealed that maximal activity occurred at high nanomolar (250 – 500 nM) protein concentrations. The kinetic parameters of Cbz-GGR-AMC hydrolysis by TMPRSS2 (Km = 30.1 ± 9.6 μM; $Vmax$ = 15.8 ± 2.1 RFU/sec) provided the appropriate conditions for inhibition assays (**FIG. 4A**).

[0086] Three commercially available inhibitors, Avoralstat, PCI-27483, and Antipain, with strong docking scores were selected to test their potency in a biochemical inhibition assay and compared to Camostat. We also tested three negative controls: MDL-28170, a proteasome inhibitor that does not target Trypsin-like proteases but is known to inhibit MERS-CoV infection (Shirato et al., 2013), Ritonavir, a human immunodeficiency virus (HIV) protease inhibitor that

has been tested and showed no effect on SARS-CoV-2 infection (Cao et al., 2020), and Leupeptin, a naturally occurring general protease inhibitor that performed poorly in our docking calculations. As expected, TMPRSS2 activity was not inhibited by MDL-28170, Ritonavir, or Leupeptin (**FIG. 12B**). TMPRSS2 inhibition by Camostat was previously reported using cellular assays but has not been demonstrated using in biochemical assays. In our system, TMPRSS2 activity was inhibited by Camostat (IC50 = 159 ± 34 μM; **FIGS. 4B-4C**). Avoralstat, PCI-27483, and Antipain similarly displayed inhibition in the micromolar range (IC50 = 69.4 ± 5.8 μM, 70.8 ± 3.1 μM, and 75.2 ± 6.7 μM respectively; **FIGS. 4B-4C**), which correlated well with their high docking scores. An inhibition assay using recombinant human HtrA1 (High Temperature Requirement Protein A1; residues 161 - 480; catalytic domain - PDZ domain) protein, which has P1 valine substrate specificity, was performed to determine whether the compounds might induce non-specific inhibition or protein aggregation effects at micro-molar concentrations, and there was no significant inhibition (data not shown). Taken together, these results suggested that these inhibitors might prevent viral entry directed by TMPRSS2-mediated cleavage of SARS-CoV-2 spike proteins.

Evaluation of a natural source inhibitor based on structure similarity

[0087] The 3DPhyloFold analysis also revealed that the TMPRSS2 3D structure was very similar to Trypsin structures (**FIG. 2**). Interestingly, one high-resolution Trypsin structure was porcine (*Sus scrofa*) Trypsin in complex with the Soybean (*Glycine max*) Trypsin Inhibitor (SBTI; PDB 1AVW). Despite only an overall 38% sequence identity to TMPRSS2, twenty-one of thirty-one residues in the porcine Trypsin binding pocket were conserved with TMPRSS2 (**FIG. 13A**). SBTI has been widely used as a cheap and natural source inhibitor for Trypsin in biomedical research (Song and Suh, 1998). We therefore hypothesized that SBTI might bind and inhibit TMPRSS2. We modeled the TMPRSS2-SBTI complex (HADDOCK score = -155; **FIG. 5A**) and identified the conserved inhibitory motif (PYRIRF; **FIG. 13B**), which docked well to the TMPRSS2 model (Glide score = -9.2, HADDOCK score = -77; **FIG. 5B**; **Table 5**). To validate this interaction, we tested SBTI's inhibitory potential in the *in vitro* recombinant TMPRSS2 inhibition assay. SBTI inhibited TMPRSS2 activity in the micromolar range (IC50 = 66.4 ± 19 μM; **FIG. 5C**).

Inhibition of TMPRSS2 autoproteolysis in cells

[0088] We further tested whether Camostat, Avoralstat, PCI-27483, Antipain, and SBTI could inhibit proteolysis of the full-length (FL) TMPRSS2 in cells. FL-TMPRSS2 contains an autoproteolysis motif (RQSR^IV (SEQ ID NO:1; ^ indicates expected cut site; residues 252 - 257) at the beginning of its S1-peptidase domain (residues 256 - 492). This autoproteolysis

site is subject to cleavage by TMPRSS2 itself and by other S1-peptidase domains. We therefore utilized this autoproteolytic process to probe the activity of TMPRSS2 in cells (Shulla et al., 2011). Human embryonic kidney (HEK) 293T cells were transfected with FL-TMPRSS2 expression plasmids with a C-terminal FLAG-tag. Upon autoproteolysis, the FL-TMPRSS2 construct (~54 kDa) liberates its FLAG-tagged S1-pepidase domain (~26 kDa). Western blotting of protein extracts from cells transfected with either wild-type (WT) or TMPRSS2-S441A mutant (S441A; loss-of-function mutant), using anti-FLAG antibody, showed the expression of TMPRSS2 with an approximate molecular weight of 54 kDa. There was high expression of the catalytically inactive TMPRSS2-S441A mutant. There was reduced expression of the catalytically active WT TMPRSS2, as previously reported, likely due to a high rate of auto-proteolysis resulting in enzyme's degradation (Shulla et al., 2011). Lower molecular weight bands (26 kDa), were observed only in the WT and not in the S441A mutant or GFP (negative control) transfected cells, confirming the bands were produced by the active TMPRSS2, and not by endogenous proteases present in 293T cells (**FIG. 6A**). Inhibitor compound treatment (100 μM dose) prevented FL-TMPRSS2 autoproteolysis and significantly increased the FL-TMPRSS2 band intensity (**FIGS. 6A-6B**). Some cleaved fragments were detected in Avoralstat but not after Camostat treatment. These results suggested that the candidate compounds could inhibit TMPRSS2 proteolytic activity in cells.

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

[0089] Cell infection assays were performed to test the inhibitory potential of the compounds. We first tested the inhibitors against pseudovirus cell entry using replication-defective vesicular stomatitis virus particles (VSV-G) or pseudovirus bearing the SARS-CoV-2 Spike glycoprotein. Lung epithelial Calu-3 2B4 cells were incubated with Camostat, Avoralstat, PCI-27483, SBTI, and Antipain in the 100 nM to 100 μM dose range. Particles harboring the pantropic VSV-G were used as a comparative control since they infect cells independent of TMPRSS2. Indeed, no inhibitor prevented VSV-G pseudovirus entry. However, Camostat inhibited SARS-CoV-2 pseudoviral entry (EC50 = 0.96 ± 0.5 μM **FIGS. 7A-7B**). Avoralstat displayed similar inhibition of SARS-CoV-2 pseudoviral entry (EC50 = 6.62 ± 2.2 μM), while SBTI was inhibited at higher concentrations (EC50 = 59.7 ± 26 μM; **FIG. 7B**). PCI-27483 and Antipain displayed modest inhibition at high micromolar concentrations but were too weak to determine reliable EC50 values in the tested dose range (**FIG. 7B**).

[0090] Next, inhibition of viral entry, replication, and reinfection was tested against authentic SARS-CoV-2 infection in Calu-3 2B4 lung cells. A 100 μM dose of Camostat and Avoralstat led to a significant reduction in SARS-CoV-2 infection (86% and 91%, respectively; $p<0.0001$; **FIG. 7C**), while PCI-27483, SBTI, and Antipain reduced SARS-CoV-2 infection by 55%, 86%,

and 34%, respectively ($p<0.0001$; **FIG. 7C**). A dose-response of Camostat and Avoralstat (100 pM to 10 μM dose range) displayed EC50 values lower than 100 nM (Camostat: 18.2 ± 2.9 nM; Avoralstat; 52.3 ± 16 nM; **FIGS. 7D-7F**), several orders of magnitude less than observed in the pseudovirus cell entry assay. Interestingly, SARS-CoV-2 showed more sensitively to the inhibitors than MERS strain of coronavirus (**FIGS. 7E-7F**). These results suggest that Avoralstat and Camostat can effectively reduce the SARS-CoV-2 entry and the effect of attenuated viral entry was amplified by reducing the subsequent bursts of virus replication.

Avoralstat reduces SARS-CoV-2 infection in mice

[0091] Other than convalescent serum, no therapeutic has been validated in an *in vivo* model of COVID-19 (Sun et al., 2020). Using a recently created mouse model of human SARS-CoV-2 lung infection, we compared the efficacy of Avoralstat and Camostat for preventing and reducing SARS-CoV-2 infection. As previously described, wild-type BALB/c mice were transduced with Ad5-hACE2, then intranasally infected with either 3×10^3 or 1×10^5 PFU of SARS-CoV-2. Mice were treated with Avoralstat, Camostat (30 mg/kg intraperitoneal injection), or vehicle (DMSO; negative control) four hours before and four hours after being challenged by virus. Virus titers were measured in lungs harvested 1-day after infection. Compared to vehicle, both Avoralstat and Camostat significantly reduced the lung viral titer by more than 70% in both cohorts (challenged with 1×10^5 PFU: Vehicle-Camostat $p = 0.013$, Vehicle-Avoralstat $p = 0.012$; challenged with 3×10^3 PFU: Vehicle-Camostat $p = 0.007$, Vehicle-Avoralstat $p = 0.005$; **FIG. 8**). Avoralstat-injected mice showed curly/ruffled fur but were otherwise healthy and displayed normal behavior. This strongly suggests that Avoralstat was as effective as Camostat in preventing of SARS-CoV-2 entry and replication *in vivo*. Importantly, it validates TMPRSS2 *in vivo* as a bonafide host target for antiviral SARS-CoV-2 therapy.

[0092] ***Modeling of the TMPRSS2 binding pocket*** – In order to conduct docking studies on human TMPRSS2, we first needed to determine its putative binding pocket. Proteases contain several sub-binding pockets (e.g. S1', S1, S2, etc.) that participate in the recognition of specific amino acid sidechains around the substrate scissile bond (Perona and Craik, 1995). These binding pockets are important for conferring protease substrate specificity and can give important insights into inhibitor design and the chemical modifications required to maximize specificity (Perona and Craik, 1995). We modeled the putative TMPRSS2 peptidase domain binding pocket and its conservation based on the structure of human Hepsin bound to an inhibitor (PDB 1Z8G; **FIG. 2A**; **FIG. 9B**). The primary specificity pocket (S1) of Hepsin contains an aspartate (Asp347) residue which confers a preference for arginine sidechains in the P1

position of its substrates. This S1 position is conserved in TMPRSS2 as well as 4 other Trypsin-like serine proteases from our structural phylogenetics analysis (**Table 1**). Core residues in the Hepsin S1' (e.g. Asp352 and Ser 353) and S2 (e.g. His203, Asp257, and Val375) were highly conserved. In contrast, the solvent exposed residues of S3, S4, and S5 subsites were more variable among the analyzed proteases. These results suggest that inhibitor design for TMPRSS2 should take advantage of the conserved core subsites for potency and the surface-exposed sites to achieve specificity.

[0093] ***In silico evaluation of experimental protease inhibitors*** – Acrosin, a Trypsin-like serine protease found in spermatozoa, shared a similar predicted fold to TMPRSS2 (sequence identity = 39%; Ca RMSD = 0.49 Å). Acrosin plays several important functions during fertilization (e.g. hydrolysis of the zona pellucida) and total Acrosin activity correlates with fertility potential in humans, making it an attractive target for male contraceptives (Baba et al., 1994; Flechon, 2016; Ning et al., 2013; Tummon et al., 1991; Welker et al., 1988; Yeste et al., 2017). We examined 16 known Acrosin inhibitors and assessed their inhibitory potential using docking calculations (**Table 3**). The top scoring inhibitors were A8, A9, and A16. Hepsin is a member of the TMPRSS/Hepsin subfamily of S1 peptidases and shares a similar predicted fold to TMPRSS2 (sequence identity = 41%; Ca RMSD = 0.24 Å). Hepsin processes a number of growth factors and enzymes such as Hepatocyte Growth Factor (HGF), Macrophage Stimulating Protein (MSP), and ACE2 (Ganesan et al., 2011; Herter et al., 2005; Heurich et al., 2014). Increased *HEPS* gene expression is associated with prostate cancer progression and metastasis (Kwon et al., 2017; Tang et al., 2014; Willbold et al., 2019). We therefore examined 13 known Hepsin inhibitors and assessed their inhibitory potential using docking (**Table 3**). The top scoring inhibitors were H1, H6, and H12. Docking comparisons of all the inhibitors against TMPRSS2 peptidase domain revealed that, compared to other known serine protease drugs, the current clinical trial candidates do not perform the best. Inhibitors such as PCI-27483, Avoralstat, Antipain, and Leupeptin displayed higher docking scores against closely related serine proteases, based on the modeling.

[0094] *Soybean trypsin inhibitor might bind to TMPRSS2* – SBTI or the soybean Kunitz-type trypsin inhibitor is a 181-residue protein known to inhibit Trypsin, Plasma Kallikrein, Factor Xa, and Plasmin in serum-free cell culture media (Borodin et al., 2013; Song and Suh, 1998). It does not inhibit metalloproteases, cysteine proteases, aspartic proteases, or tissue Kallikrein (Borodin et al., 2013). We therefore docked the SBTI protein to our TMPRSS2 model using HADDOCK (de Vries et al., 2010) (**FIG. 5A**), since the SBTI protein was too large to be handled in Glide. The SBTI protein was also docked onto Trypsin, Plasma Kallikrein (KLKB1),

and Factor VII structures for comparison (**FIG. 13**; **Table 5**). HADDOCK generated 200 TMPRSS2/SBTI docking structures, 188 of which fell into 6 clusters. The best cluster size was 116 with a HADDOCK score of -154.6 ± 7.5 and Z-score of -1.8 (**Table 5**). As expected, porcine Trypsin yielded the best HADDOCK score, since the Trypsin and SBTI structures used for docking were derived from their complex structure. The experimentally determined and HADDOCK-derived structures of Trypsin/SBTI were nearly identical (**FIG. 13C**). TMPRSS2 showed comparable HADDOCK score to those of known target proteases of SBTI (Trypsin and KLKB1), suggesting that SBTI might inhibit TMPRSS2 activity as well (**Table 5**).

[0095] We performed structural analysis on the HADDOCK-derived complex to identify the key residues involved in the TMPRSS2/SBTI interaction. A six amino acid peptide in the Trypsin-inhibitory loop of SBTI (residues 561 - 566; PYRIRF) occupied the active site of TMPRSS2, while two 2 residues at the C-terminal beta-strand of SBTI interacted with a grove adjacent to the active site (**FIG. 5B**). We docked the PYRIRF peptide to TMPRSS2, which resulted in a relatively high score (Glide Score = -9.2; HADDOCK Score = -73; **FIG. 5B**; **Table 4**). The binding mode of the peptide was similar to the HADDOCK-derived TMPRSS2/SBTI complex with the exception of the last residues at the N- and C-termini of the peptide. This likely due to the lack of structural restraints from the SBTI core structure leaving the PYRIRF peptide more flexible. The PYRIRF GlideScore is comparable to the compounds listed in **Table 4**, suggesting that this peptide motif might be sufficient to bind TMPRSS2. Taken together this suggests that the 8 residues of SBTI have potential to be optimized to increase the potency and selectivity of SBTI against TMPRSS2 or be used as a template for peptidomimetic inhibitors to target this host protease.

Discussion

[0096] The emerging SARS-CoV-2 pandemic presents a global health emergency. SARS-CoV-2's long incubation (asymptomatic) period has contributed to its widespread transmission worldwide (Perlman, 2020). Symptomatic patients, after a mean incubation time of 5 days, typically begin experiencing fever, cough, nasal congestion, fatigue, or other signs of upper respiratory tract infections (Velavan and Meyer, 2020). A small minority of patients also developed gastrointestinal symptoms, such as nausea, vomiting, or diarrhea (Guan et al., 2020). Although an estimated 80% of COVID-19 patients are mild or self-limited, the illness can progress to pneumonia, acute respiratory distress syndrome (ARDS), or shock, and leave lasting pulmonary changes (Cowling and Aiello, 2020; Wang et al., 2020). These mild or asymptomatic carriers can transmit disease to others. Around 5% of all infected patients were admitted to the intensive care unit and approximately half need invasive mechanical ventilation (Guan et al., 2020). The case fatality rate is around 2-3% although likely higher in the elderly

population and those with comorbidities (Bedford et al., 2020). Approximately 80% of all deaths occurred among adults greater than 60 years old (Team, 2020). A variety of approaches are being investigated, but it is likely that a combination of therapies may be useful at different stages of exposure and disease. Therapies that reduce the viral burden early in disease could reduce transmission to others, even once a vaccine is available.

[0097] Drug repurposing is an important strategy to address human disease, especially in the urgent setting of a global viral pandemic. A structure-based phylogenetics approach using 3DPhyloFold can identify closely related proteins missed by primary sequence comparisons, and this can lead to a mechanism-based, hypothesis driven selection of curated inhibitors candidates. Interestingly, Avoralstat was represented in high throughput screens but was likely missed due to the lack of sufficient testing conditions (e.g. dosage) (Bakowski et al., 2020). Regardless of the initial approach, candidate validation in biologically relevant cells and animal models using the specific virus is important in the downstream screening process.

[0098] Our studies show that targeting the host protease TMPRSS2 is a tenable strategy. It may be especially important if current therapeutic strategies targeting viral proteins do not confer long-lasting immunity in different genetic strains of corona and other viruses. Recently, it was demonstrated that TMPRSS2 is required for SARS-CoV-2 to enter host cells, as Camostat mesylate, a previously studied TMPRSS2 inhibitor, can block coronavirus infection *in vitro* (Hoffmann et al., 2020; Shen et al., 2017). Engineered VeroE6 cells with high TMPRSS2 expression have shown to be highly susceptible for SARS-CoV-2 infection, with a ten-fold greater number of infected cells compared to the original VeroE6 cells (Matsuyama et al., 2020). In addition, our SARS-CoV-2 neutralization assay in Calu-3 2B4 cells showed that SARS-CoV-2 entry can be more sensitively inhibited by Camostat and Avoralstat than MERS-CoV. This suggests that SARS-CoV-2 may have evolved to better utilize host proteases. In addition, TMPRSS2 is not only responsible for the SARS-CoV-2 viral entry by cleaving the S2' site of SARS-CoV-2 spike protein that leads to viral fusion, but it is also implicated in the cleavage of viral fusion proteins of many other corona viruses and influenza viruses: SARS-CoV; MERS-CoV; HCoV -229E, -OC43, -HKU1, and -NL63; Influenza virus -A (LPAIV, human viruses) and -B; Prainfluenza virus -1, -4a, -4b; and human Metapneumovirus (Shen et al., 2017). These findings suggest that targeting the host machinery mediating viral entry (i.e. ACE2, Furin, and TMPRSS2) could be long-term strategy for SARS-CoV intervention and prophylaxis. In addition, there are increasing number of reports that combination therapy, which is a cocktail use of drugs that target different molecules or pathways or that target the same molecule or pathway to reduce the likelihood of resistance, can be more effective to treat COVID-19. Many of treatments in use or in development for COVID-19 are mainly targeting viral proteins (i.e. inhibitor for viral RNA polymerase

[remdesivir]; inhibitor for SARS-CoV-2 main protease; antibodies) or suppressing patient's inflammatory response (i.e. suppressors of cytokine signaling [SOCS]). Thus, inhibitors targeting the host (human) protein responsible for the SARS-CoV-2 entry will be good additions to combination therapy for COVID-19.

[0099] TMPRSS2 has been targeted in humans in past Camostat trials where it was classified the drug as clinically safe and effective in to treat indications unrelated to coronavirus infections, such as chronic pancreatitis (Dyall et al., 2017). Following the SARS-CoV and MERS-CoV outbreaks, renewed investigations involving animal models and human cell lines suggested the potential of Camostat as an antiviral therapeutic. Studies in mice demonstrated that Camostat inhibits *in vivo* viral spread of another corona virus, SARS-CoV-1, reducing mortality following infection by ~60% (Zhou et al., 2015). Camostat blocked SARS-CoV-2 infection in lung cells, demonstrating its potential for use as an off-label therapeutic (Hoffmann et al., 2020). Consequently, a randomized, placebo-controlled Phase IIa clinical trial investigating the impact of Camostat in SARS-CoV-2 is being setup in Denmark (clinical trial NCT04321096). However, Camostat has a half-life of less than one minute (Choi et al., 2015; Schneider et al., 1994), requiring intravenous delivery and making its widespread prophylactic use unfeasible.

[00100] Avoralstat has advantages over Camostat. Because of its oral delivery, it is simpler to prescribe for high-risk groups (e.g. people exposed to diagnosed, same household with diagnosed, patients in self-quarantine, elderly, primary/essential workers, or front-line healthcare workers). From prior published clinical data, Avoralstat possess a favorable plasma half-life of 12-31 hours with relatively minor and manageable side effect profiles. The short half-life of Camostat is due to the easily cleavable ester bond, which results in rapid plasma breakdown followed by excretion. Avoralstat does not contain these reactive motifs, which results in its much longer half-life. The longer lasting feature in plasma of Avoralstat is more favorable for drug dose, formulation, and delivery. In our studies, intraperitoneal (IP) administration of 30 mg/kg dose of Avoralstat to infected mice and observed significant decrease in viral titer in mouse lungs. The IP administration data suggests that the oral delivery of Avoralstat might be efficacious for treating COVID-19. During Phase 1 clinical trials of Avoralstat, multiple ascending doses up to 2400 mg (taken in 3, 800 mg doses) were well tolerated with no dose limiting toxicities (Cornpropst et al., 2016). Patient were separated into two cohorts, single dose and multiple dose groups. The single dose group reported one adverse event, syncope (1/31). The multiple dose cohort reported diarrhea (6/40), flatulence (5/40), back pain (4/40), dizziness (3/40) and headache (2/40) During phase 3 clinical trials, patients (74) were given either 300 mg 3 times per day (36) or 500 mg 3 times per day (38). No grade 3 adverse events were observed throughout the 72-week trial period (Riedl et al.,

2018). The most common side effect proved to be GI discomfort (16/40) which involved diarrhea (24/74) and flatulence (15/74). Other adverse events reported include nasopharyngitis (13/74) and headaches (11/74). Although Avoralstat was administered orally, it could also be reformulated for intravenous or pulmonary delivery.

[00101] In addition to the rapid identification of drug repositioning candidates, our 3DPhyloFold analysis pointed us to an interesting crystal structure: a porcine Trypsin in complex with a natural Trypsin inhibitor SBTI. Our *in silico* analysis suggested that the SBTI might inhibit TMPRSS2 activity by utilizing the nearly identical structures and biochemical properties near the respective active sites between TMPRSS2 and Trypsin. Our *in vitro* assays confirmed that the natural Trypsin inhibitor can inhibit TMPRSS2 activity and SARS-CoV-2 entry to human lung cells. Even though the efficacy of SBTI on SARS-CoV-2 entry was not as good as Avoralstat, it suggests that the natural Trypsin inhibitors might be modified to develop peptidomimetic inhibitors for SARS-CoV-2 or have potential to serve as a cheap natural source SARS-CoV-2 intervention supplement that can be mass-produced. This is because that they have been considered as potential treatments for chronic diseases associated to overweight and obesity; their biological activities and safety are well characterized; and their industrial methods of isolation and purification are well established. Before any of these compounds are used as therapy, they must be thoroughly investigated in well-designed human clinical trials.

Conclusion

[00102] In conclusion, our study suggests that the 3DPhyloFold drug repurposing approach is an effective, rational-based method; inhibiting host serine protease TMPRSS2 is a valid strategy for treating COVID-19 and other viral infections; a clinical-trialed inhibitor Avoralstat may serve as a candidate for prophylaxis and combination therapy for COVID-19.

STAR Methods

Experimental Model and Subject Details

[00103] Mice, virus, and cells – Specific pathogen-free 6-10-week-old male and female BALB/c and C57BL/6 mice and 5-6-month-old C57BL/6 mice were purchased from Charles River Laboratories and maintained in the Animal Care Facilities at the University of Iowa. All protocols were approved by the Institutional Animal Care and Use Committees of the University of Iowa. The SARS-CoV-2 strains used in this research were isolated from COVID-19 patients in Guangzhou and in Washington state (Accession numbers: MT123290, MN985325.1), and passaged on Vero E6 and Calu-3 2B4 cells. African Green monkey kidney-derived Vero E6 cells and 17CL-1 cells were grown in Dulbecco's modified Eagle's medium

(DMEM, GIBCO, Grand Island, NY) supplemented with 10% fetal bovine serum (FBS). Calu-3 2B4 cells were grown in MEM (GIBCO, Grand Island, NY) supplemented with 20% FBS. The human serotype 5 adenoviral vector expressing human ACE2 under the control of the CMV promoter was previously described.

Method Details

[00104] *Database search and multiple sequence alignment* – We first searched the UniProt database for reviewed entries denoted as transmembrane serine proteases (containing an S1 peptidase domain). This initial search yielded 9 manually curated sequences. A seed multiple sequence alignment (MSA) of S1 peptidase domains was then constructed using MAFFT v7 with default parameters (alignment strategy: FFT-NS-1) (Katoh and Standley, 2013). Using HMMER-3.1 and the seed alignment, we produced an HMM profile and used it to broaden the search against the UniProt database (search restricted to reviewed sequences) (Finn et al., 2011). This search yielded a total of 828 S1 peptidase sequences. We discarded fragmented sequences (<200 amino acids) that appeared too short to truly represent the S1 peptidase fold and redundant proteins were further filtered using CD-HIT v4 with a threshold of 100% sequence identity (Li and Godzik, 2006). This resulted in a pool of 742 proteins that were aligned using MAFFT v7 (alignment strategy FFT-NS-2) (Katoh and Standley, 2013). Taxa (sequences) producing many gaps in the alignment were removed using MaxAlign, resulting in 600 S1 peptidase sequences (Gouveia-Oliveira et al., 2007).

[00105] *Phylogenetic tree reconstruction* – We used the IQ-TREE-1.6.2 algorithm to generate a maximum likelihood tree of the 600 S1 peptidase sequences (Nguyen et al., 2015). The IQ-TREE model finder tool was used to determine the best substitution model to fit the data (Kalyaanamoorthy et al., 2017). The Whelan & Goldman (WAG) substitution model was determined to be the best fit to the data (Whelan and Goldman, 2001). Bootstrap analysis was performed using the 'ultra-fast' method in IQ-TREE-1.6.2 with 1,000 replicas. Bootstrap values were projected onto the tree topology inferred from the base MSA.

[00106] *Structural modeling and conservation analysis of TMPRSS2* – Briefly, a BLAST search of the human TMPRSS2 S1-peptidase domain against the Protein Data Bank (PDB) returned the structure of human Hepsin (PDB 1Z8G) as the top hit. Other close matches were human Plasma Kallikrein (PDB 6ESO), type II human Plasminogen (PDB 4DUR), and human Prostatin (PDB 3E16). A model of the TMPRSS2 S1-peptidase domain was generated with the Hepsin template (41% sequence identity; PDB 1Z8G) using three different modeling

programs: Phyre2, MODELLER, and SWISS-Model (Eswar et al., 2006; Kelley et al., 2015; Schwede et al., 2003). The models were in agreement and aligned well with minor variations in the surface-exposed loop regions (Herter et al., 2005; Kelley et al., 2015). The TMPRSS2 homology model was then used as input for analysis in the ConSurf server (Armon et al., 2001; Ashkenazy et al., 2016; Glaser et al., 2003). The 600 sequences from our sequence-based phylogenetic analysis underwent multiple sequence alignment using MAFFT and conservation scores were calculated using the Bayesian method option in ConSurf as previously described (Velez et al., 2020). The ConSurf scores were mapped onto the B-value column of the TMPRSS2 homology model and visualized using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC). The TMPRSS2 binding pocket was inferred by comparison to the structure of hepsin bound to a peptidomimetic inhibitor, N-acetyl-6-ammonio-L-norleucyl-L-glutaminyl-N-[(1S)-4-([amino(iminio)methyl]amino)-1-(chloroacetyl)butyl]-L-leucinamide (PDB 1Z8G) in PyMOL.

[00107] *Structure-based phylogenetic analysis of S1-peptidases* – There are over 2,000 structures of S1-peptidase domains represented in the PDB. We therefore searched the Pfam database for structures of mammalian peptidases and selected 74 representative structures (one structure per protein; **Table 2**) (Finn et al., 2014). Structures (with reflection data deposited in the PDB) were evaluated by their reported global validation metrics in PDB-REDO (Joosten et al., 2009). Re-refined structural models were used for further analysis. Structures were superimposed using PyMOL to calculate the pairwise root mean square deviation (RMSD) between protein alpha carbon atoms (Ca). A structural dissimilarity matrix (SDM) was constructed using the Ca RMSD values in order to generate a phylogenetic tree as previously described (Velez et al., 2020). To expedite the pairwise alignment process, we developed a Python-based script (named 3DPhyloFold) to perform the pairwise alignment of protein structures and generate an SDM. The phylogenetic tree was constructed using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method in MEGAX (Molecular Evolutionary Genetics Analysis) software (Kumar et al., 2016; Sneath and Sokal, 1973). This algorithm was chosen as it is a straightforward method to construct a phylogenetic tree from a pairwise distance matrix alone (without the need for sequence information). For comparison, the sequences from the corresponding structures were also analyzed by sequence-based phylogeny. Briefly, the 75 S1-peptidase sequences were aligned with MAFFT v7 (Katoh and Standley, 2013) and underwent phylogenetic reconstruction in IQ-TREE-1.6.2 (Nguyen et al., 2015). The Jones-Taylor-Thornton (JTT) substitution model was determined to be the best fit to the data (Jones et al., 1992). Bootstrap analysis was performed using the 'ultra-fast' method in IQ-TREE-1.6.2 with 1,000 replicas. A TMPRSS2 structure similarity score for each analyzed

protease was calculated by dividing the pairwise sequence identity (to TMPRSS2) by the Ca RMSD of the pairwise alignment.

[00108] *Database search for S1-peptidase inhibitors* – We first searched for inhibitors designed for the related proteins. We filtered for cases where a strong Structure-Activity-Relationship (SAR) between the ligand and protein was studied, where applicable. We discarded studies where the inhibitors displayed low potency, focusing on groups of inhibitors that displayed sub-micromolar inhibition for their intended protein. We focused on cases where the inhibitors studied contained a guanidine, or guanidine-like, functional group to interact with the S1 specificity pocket. This led to the selection of a single paper that investigated inhibitors for said protein. Due to the vast number of possible inhibitors for each protein, we chose to only investigate a very small sample set. The inhibitors were then prepared and docked against our TMPRSS2 model protein structure.

[00109] *In silico docking calculations* – Published crystal structures of inhibitor-bound Hepsin (PDB 1O5E), Trypsin-3 (PDB 1H4W), Kallikrein-6 (PDB 1LO6), ram acrosin (PDB 1FIW), and bovine endopeptidase (PDB 1EKB) were loaded into Maestro software (Schrödinger Release 2019-3). The TMPRSS2 model described above was used. To prepare the proteins for docking and simulations, the protein preparation wizard was used to assign bond orders, add hydrogens, create zero-order bonds to metals, create disulfide bonds, and fill in missing side chains and loops. The default parameters were used for the optimization of hydrogen-bond assignment (sampling of water orientations and use of pH 7.0). Waters molecules beyond 3 Å of heteroatoms or with fewer than three hydrogen bonds to non-waters were removed. Restrained energy minimization was applied using the OPLS3e force field (Harder et al., 2016). Prepared protein systems were further checked by Ramachandran plots, ensuring there were no steric clashes. To generate receptor grids for small molecule docking, the co-crystalized ligand was selected as the grid-defining ligand for each system. Default van der Waals radius scaling parameters were used (scaling factor of 1, partial charge cutoff of 0.25). To generate receptor grids for peptide docking, the co-crystalized ligand was selected as the grid-defining ligand for each system, and the grid size was made suitable for peptides to be docked. Default van der Waals radius scaling parameters were used (scaling factor of 1, partial charge cutoff of 0.25). For docking of the ligands into the various prepared proteins, the 3D structure was loaded into Maestro software (Schrödinger Release 2019-3). Ligprep was used to prepare the ligands (by generating possible states at pH 7.0 ± 2.0 and retaining the specified stereochemical properties). The prepared small molecule ligands were then docked using the most stringent docking mode (extra precision, "XP") of Glide, with the following parameters:

dock flexibly, perform post-docking minimization, and show only top 2 posed for each ligand (Friesner et al., 2006). The prepared peptide segments were docked using the standard precision more for peptides in Glide, with the following parameters: dock flexibly, perform post-docking minimization, and show the top 100 posed for each ligand.

[00110] *Docking Soybean Trypsin Inhibitor (SBTI) to TMPRSS2 and other Trypsin-like S1-peptidase domains* – The HADDOCK 2.4 online docking tool was used to generate TMPRSS2/SBTI complex structure model. The TMPRSS2 homology model and the SBTI structure (PDB 1AVW) were used for docking (de Vries et al., 2010). To define the potential interaction surface between TMPRSS2 and SBTI, the TMPRSS2 homology model was superimposed to the wild boar trypsin structure which is in complex with a SBTI (PDB 1AVW) using PyMOL. The following residues of SBTI were designated as active residues: 501-502, 510, 512- 514, 560-572, and 616-617. The overall Ca RMSD between the two models was 0.54 Å. SBTI was also docked to porcine Trypsin (PDB 1AVW), human Coagulation Factor VII (PDB 1W7X), and human Plasma Kallikrein (PDB 6O1S). The HADDOCK scores represent the average score of the best cluster. The parameters and output files for the HADDOCK run can be found in supplementary files.

[00111] *Recombinant TMPRSS2 expression and purification* – The human TMPRSS2 peptidase domain sequence (residues 252 - 489) was cloned into a pET28a vector with a N-terminal 6x-His tag. Plasmids were amplified and isolated from DH5a cells and then transformed into *E. coli* BL21 (DE3). BL21 cells expressing TMPRSS2 were grown in 1 L shake-flask cultures, at 250 rpm in lysogeny broth (LB), at 37 °C until an OD 600 of 0.6 and then induced with 0.5 mM IPTG. Cells were grown at 18 °C for 18 hours, harvested, and centrifuged at 4,000 rpm for 20 minutes and pellets (10 mL per Liter of cell culture) were re-suspended immediately or frozen at -20 °C for later use. Cell pellets were re-suspended in 35 to 50 mL of lysis buffer (50 mM Tris, 150 mM NaCl, 20 mM Imidazole pH 8.0, one tablet of EDTA-free protease inhibitor [Roche], DNaseI [Roche]) and lysed using a sonic dismembrator and centrifuged for 30 minutes at 18,000 x g at 4 °C. The supernatant was discarded, and the pellets were re-suspended for 30 minutes to an hour at room-temperature in 50 mL of denaturation buffer (50 mM Tris, 150 mM NaCl, 6 M Guanidinium Chloride, 1 M L-Arginine, 2 mM DTT pH 8.0). After re-suspension, the denatured sample was filtered with 0.22 µm filter. The equivalent volume of refolding buffer-1 (50 mM Tris, 150 mM NaCl, 2 M Guanidinium Chloride, 1 M L-Arginine pH 8.0) was applied to the filtered sample, and it was applied to SnakeSkin® Dialysis Tubing (10,000 MWCO, Thermo Scientific). The sample in the dialysis bag went through slow refolding by dialyzing in 2 L of refolding buffer-1 for over-night at 4 °C.

After the over-night refolding, the sample was filtered with 0.22 µm filter to remove aggregates and went through another step of dialysis in 2 L of refolding buffer-2 (50 mM Tris, 150 mM NaCl, 1 M L-Arginine pH 8.0) for 1.5 hours at room-temperature. Aggregates are removed with 0.22 µm filter. The sample was concentrated with a 10 kDa NMWL spin concentrator and passed over a Superdex 200 (GE) size-exclusion (SEC) column connected to an ÄTKA™ pure fast protein liquid chromatography (FPLC) system (GE Healthcare Inc.) to remove residual impurities and aggregated protein. The column was equilibrated with SEC buffer (50 mM Tris, 150 mM NaCl, pH 8.0). Purity was assessed at each step on Coomassie-stained SDS-PAGE gels, and the final purity of recombinant TMPRSS2 peptidase domain used for *in vitro* assays were >95% chromatography (**FIG. 13D**).

[00112] *Recombinant HtrA1 expression and purification* – We adapted the purification and expression strategy from a previously published paper on HtrA1 (Eigenbrot et al., 2012). The human HtrA1 catalytic domain-PDZ domain sequence (residues 161 - 480) was cloned into a pET21a vector with a C-terminal 6x-His tag. Plasmids were amplified and isolated from DH5a cells and then transformed into *E. coli* BL21 (DE3). BL21 cells expressing HtrA1 were grown in 1 L shake-flask cultures, at 250 rpm in LB, at 37 °C until an OD 600 of 0.6 - 0.8 and then induced with 0.4 mM IPTG. Cells were grown at 18 °C for 20 - 24 hours, harvested, and centrifuged at 4,000 rpm for 20 minutes and pellets (10 mL per Liter of cell culture) were re-suspended immediately or frozen at -20 °C for later use. Cell pellets were re-suspended in 35 to 50 mL of lysis buffer (50 mM Tris, 1 M NaCl, 20 mM Imidazole pH 8.0, one tablet of EDTA-free protease inhibitor [Roche], DNaseI [Roche]) and lysed using a sonic dismembrator and centrifuged for 30 minutes at 18,000 x g at 4 °C. Cell debris was discarded and the supernatant was loaded onto a column packed with 5.5 mL (1CV) cobalt resin. The column was washed with 5 to 10 CV of wash buffer (50 mM Tris, 1 M NaCl, 20 mM Imidazole pH 8.0) and eluted with 10 CV of elution buffer (50 mM Tris, 200 mM NaCl, 300 mM Imidazole, 10% glycerol, 0.25% CHAPS pH 8.0). Eluted fractions were pooled and concentrated with a 30 kDa NMWL spin concentrator and passed over a Superdex 200 (GE) SEC column connected to an ÄTKA™ pure FPLC system (GE Healthcare Inc.) to remove residual impurities and aggregated protein. The column was equilibrated with SEC buffer (50 mM Tris, 200 mM NaCl, 0.25% CHAPS pH 8.0). Purity was assessed at each step on Coomassie-stained SDS-PAGE gels, and the final purity of recombinant TMPRSS2 peptidase domain used for *in vitro* assays were >90%.

[00113] *Measurement of TMPRSS2 activity* – TMPRSS2 proteolytic activity was measured by hydrolysis of the synthetic urokinase substrate, Cbz-GGR-AMC (Cbz-Gly-Gly-L-Arg-4-

methylcoumaryl-7-amide; Echelon). Briefly, 250 nM of purified TMPRSS2 was added to a reaction buffer containing 50 mM Tris-HCl (pH 8.0), and 150 mM NaCl in black-bottom 96-well plates (100 µL per reaction). Inhibition experiments were carried out in the presence of 50 µM Cbz-GGR-AMC in the presence of Camostat, Avoralstat, PCI-27483, Antipain, Leupeptin, and MDL-28170 (10 to 500 µM) or 5% DMSO (as a negative control). DMSO caused SBTI to precipitate out of solution (unpublished observation). Inhibition experiments with SBTI (2 to 150 µM) were therefore performed in the absence of DMSO. Reactions were run at 37 °C for 30 minutes on a fluorimetric plate reader (Tecan Spark, Männedorf Switzerland). Proteolytic activity was measured as change in raw fluorescence units (ΔRFU; λexc = 373 nm, λem = 455 nm) at 30-second intervals. All experiments were performed in triplicate. The initial velocity (RFU/sec) of the reaction was measured by calculating the slope of the fluorescence data from the first three minutes. Kinetic parameters were then calculated by direct fitting to the Michaelis-Menten or Hill equation in GraphPad Prism 8.

[00114] *Measurement of HtrA1 activity* – HtrA1 proteolytic activity was measured by Ab112152 protease activity assay kit (Abcam). Briefly, 250 nM of purified HtrA1 was added to a reaction buffer containing 50 mM Tris (pH 8.0), 200 mM NaCl, and 0.25% CHAPS in black-bottom 96-well plates (100 µL per reaction). Inhibition experiments were carried out in the presence of 50 µL substrate (component A; Ab112152; Abcam), diluted in 2X assay buffer (Ab112152; Abcam) in the presence of Camostat, Avoralstat, PCI-27483, Leupeptin, and MDL-28170 (10 to 500 µM) or 5% DMSO (as a negative control). DMSO caused SBTI to precipitate out of solution (unpublished observation). Inhibition experiments with SBTI (2 to 150 µM) were therefore performed in the absence of DMSO. Reactions were run at 37 °C for 30 minutes on a fluorimetric plate reader (Tecan Spark, Männedorf Switzerland). Proteolytic activity was measured as change in raw fluorescence units (ΔRFU; λexc = 490 nm, λem = 525 nm) at 1-minute intervals. All experiments were performed in triplicate. The steady-state velocity (RFU/sec) of the reaction was measured by calculating the slope of the fluorescence data from 12 to 30 minutes. Kinetic parameters were then calculated by direct fitting to the Hill equation in GraphPad Prism 8.

[00115] *TMPRSS2 autoproteolysis assay* – HEK 293T cells were obtained from the Viral Vector Core Facility at the University of Iowa. Cells were grown in Dulbecco modified Eagle medium (DMEM) (Gibco) supplemented with 5% fetal bovine serum (Gibco), penicillin and streptomycin (Gibco, WT15140-122) and were maintained in a humidified atmosphere of 5% CO2 at 37 °C. Plasmid pEGFPN1 was obtained from Clontech. TMPRSS2 cDNA (obtained from Loyola University Medical Center, Illinois) were generated as previously described by

Shulla et al., (2011) (Shulla et al., 2011). Briefly, TMPRSS2 cDNA, containing anti-FLAG epitope tag at carboxy terminal, were amplified with PCR using pCMV-Sport6-TMPRSS2 template. The resulting amplificates were cloned into the pCAGGS.MCS vector via SacI and XhoI sites. The enzymatically inactive pCAGGS-TMPRSS2(S441A)FLAG mutant cDNA was generated using QuickChange Site-Directed Mutagenesis Kit protocol according to manufacturer's instructions (Agilent Technologies). Transient transfections of HEK 293T cells with plasmids were performed using PolyFect transfection reagent according to the manufacturer's protocol (Qiagen). Briefly, HEK cells were plated at $0.2X10^6$ cells/well in 6 well plates 1 day before transfection. At 60-70% confluency, 100 µM Camostat and Avoralstat were added to the respective wells and incubated at 37 °C for 2 hours prior to transfection. For transfection, 2 µg of each plasmid (GFP [served as negative control], WT [TMPRSS2] and S411A mutant) were dissolved in the serum free media to a total volume of 0.1 mL. After proper mixing, 20 µL of PolyFect transfection reagent was added to the DNA solution followed by 10 minutes incubation at room temperature to allow complex formation. 0.6 mL of growth media (with serum and antibiotic) was then added to the reaction tubes and the complete transfection mix was immediately added onto the cells in the well plate containing fresh growth media. 24 hours post-transfection, cell lysates were prepared and processed for immunoblotting. HEK 293T cell lysates were prepared using HNB buffer (0.5% NP40, 0.5% deoxycholate, 150 mM NaCl, 25 mM Na HEPES, 0.01% BSA) containing 0.1% protease inhibitor (Sigma, P2714). After 20 minutes of ice incubation, the cell lysates were clarified by centrifugation at 2,000 g for 10 minutes. Supernatants were collected and the protein concentration, from the cell lysates, was determined using DC protein assay reagent kit using microplate assay protocol (BioRad). The samples for loading were prepared using β-mercaptoethanol and lamelli buffer (BioRad) and were boiled at 95 °C for 5 minutes before loading. Equal amounts of protein (~25 µg) was loaded in each well and the samples were subjected to SDS-PAGE (4 to 12% Bis-Tris gradient gel). After separation, proteins were transferred to a PVDF membrane, and the nonspecific binding sites were blocked for 1 hour at room temperature using 5% nonfat dry milk in TBST. Membranes were probed with mouse monoclonal anti-Flag antibody (1:1000; Sigma-Aldrich, catalogue F3165) for 16 hours (overnight) at 4 °C. After overnight incubation, the blots were washed three times with TBST (10 minutes/wash), and subsequently incubated with immunoglobulin G labelled with horseradish peroxidase conjugated secondary anti-mouse antibody (1:5000; ThermoFisher Scientific, catalogue 31432). Proteins were visualized by a SuperSignal™ West Pico PLUS chemiluminescence detection reagents (ThermoFisher Scientific, catalogue 34580) using MyECL imager (ThermoFisher Scientific). To confirm equal protein in each lane, membranes were re-probed with β-actin (1:5000; Sigma, A2228), used as a loading control. All the data

were analyzed using GraphPad Prism 8.0 (GraphPad, San Diego, CA). Data were analyzed by 1-way ANOVA followed by Tukey's multiple comparisons test. Differences of $p \leq 0.05$ were considered statistically significant.

[00116] *Pseudovirus transduction assay* – Vesicular Stomatitis Virus (VSV) pseudovirus was produced following published protocols (Qing et al., 2020). Briefly, HEK 293T cells were transfected to express either the SARS-CoV-2 spike protein (with the cytoplasmic tail removed; residues 1 - 1255) or the full-length VSV-G protein. Then, these cells were transduced with a VSV vector expressing luciferase instead of the G protein (VSV-ΔG-Luc), pseudotyped with VSV-G. After 2 hours at 37 °C, the cells were washed 3 times to remove residual virus. Supernatant containing pseudovirus was harvested 3 times at 24-hour intervals and centrifuged to remove cellular debris. Pseudovirus from the 3 collections was pooled and ultracentrifuged with a 20% sucrose cushion for purification and concentration (100x). For the transduction assays, Calu-3 2B4 cells were grown in 96-well plates until confluent. Cells were incubated with the respective compounds for 1 hour at 37 °C. After 1 hour, cells were transduced with pseudovirus, maintaining the same concentration of compounds, and incubated overnight. Transduction efficiency was assessed by quantifying luciferase activity in cell lysates using a commercial kit (Luciferase Assay System, Promega) and a plate-reading luminometer (SpectraMax i3x, Molecular Devices). The 50% effective dose (EC50) was calculated using GraphPad Prism 8.

[00117] *Infectious SARS-CoV-2 neutralization assay* – Calu-3 2B4 cells were plated in 48 well plates. Cells were incubated with medium containing indicated compounds or vehicle for 1 hour at 37 °C. The medium was removed and SARS-CoV-2 (MOI=0.1) in medium containing indicated compounds were added into each well. The cells were incubated with viruses for 1 hour at 37 °C. Next, the viruses were removed, and cells were rinsed with PBS once to remove remaining viruses. After that, cells were incubated with medium containing indicating compounds overnight. Following day, supernatants were harvested for cellular total RNA (isolated by Trizol) and viral genomic RNA was quantified by qRT-PCR.

[00118] *Transduction and infection of mice* – Mice were lightly anesthetized with isoflurane and transduced intranasally with 2.5×10^8 FFU of Ad5-ACE2 or Ad5-Empty in 75 mL DMEM. Five days post transduction, mice were infected intranasally with SARS-CoV-2 (3×10^3 or 1×10^5 PFU) in a total volume of 50 mL DMEM. Mice were monitored and weighted daily. Infected mice were treated with Avoralstat, Camostat (30 mg/kg intraperitoneal injection), or vehicle

(DMSO; negative control) four hours before and after being challenged by virus. Virus titers were measured in harvested lungs by plaque assay 1-day post infection.

[00119] *SARS-CoV-2 plaque assay* – Lung homogenate supernatants were serially diluted in DMEM. Vero E6 cells in 12 well plates were inoculated at 37 ºC in 5% CO_2 for 1 hour with gentle rocking every 15 minutes. After removing the inocula, plates were overlaid with 1.2% agarose containing 4% FBS. After further incubation for 2 days, overlays were removed, and plaques were visualized by staining with 0.1% crystal violet. Viral titers were calculated as plaque forming units (PFU) per gram tissue. All work with SARS-CoV-2 was conducted in the Biosafety Level 3 (BSL3) Laboratories of the University of Iowa.

[00120] *Data and code availability* – The 3DPhyloFold code is open source and freely available at github.com/MahajanLab/3DPhyloFold. The user manual, implementation notes, and description of methodology and examples are available on the site.

References

[00121] Coronavirus disease (COVID-19) Pandemic (World Health Organization (WHO) (2020)).

[00122] Armon, A., Graur, D., and Ben-Tal, N. (2001). ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J Mol Biol 307, 447-463.

[00123] Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 44, W344-350.

[00124] Bakowski, M.A., Beutler, N., Chen, E., Nguyen, T.-T.H., Kirkpatrick, M.G., Parren, M., Yang, L., Ricketts, J., Gupta, A.K., Hull, M.V., et al. (2020). Oral drug repositioning candidates and synergistic remdesivir combinations for the prophylaxis and treatment of COVID-19. bioRxiv, 2020.2006.2016.153403.

[00125] Baron, S.A., Devaux, C., Colson, P., Raoult, D., and Rolain, J.M. (2020). Teicoplanin: an alternative drug for the treatment of coronavirus COVID-19? Int J Antimicrob Agents, 105944.

[00126] Bedford, J., Enria, D., Giesecke, J., Heymann, D.L., Ihekweazu, C., Kobinger, G., Lane, H.C., Memish, Z., Oh, M.D., Sall, A.A., et al. (2020). COVID-19: towards controlling of a pandemic. Lancet.

[00127] Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., Ruan, L., Song, B., Cai, Y., Wei, M., et al. (2020). A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-

19. N Engl J Med 382, 1787-1799. Choi, J.Y., Kang, Y.J., Jang, H.M., Jung, H.Y., Cho, J.H., Park, S.H., Kim, Y.L., and Kim, C.D. (2015). Nafamostat Mesilate as an Anticoagulant During Continuous Renal Replacement Therapy in Patients With High Bleeding Risk: A Randomized Clinical Trial. Medicine (Baltimore) 94, e2392.

[00128] Cornpropst, M., Collis, P., Collier, J., Babu, Y.S., Wilson, R., Zhang, J., Fang, L., Zong, J., and Sheridan, W.P. (2016). Safety, pharmacokinetics, and pharmacodynamics of avoralstat, an oral plasma kallikrein inhibitor: phase 1 study. Allergy 71, 1676-1683.

[00129] Cowling, B.J., and Aiello, A. (2020). Public health measures to slow community spread of COVID-19. J Infect Dis.

[00130] de Vries, S.J., van Dijk, M., and Bonvin, A.M. (2010). The HADDOCK web server for data-driven biomolecular docking. Nat Protoc 5, 883-897.

[00131] de Wit, E., van Doremalen, N., Falzarano, D., and Munster, V.J. (2016). SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 14, 523-534.

[00132] Dyall, J., Gross, R., Kindrachuk, J., Johnson, R.F., Olinger, G.G., Jr., Hensley, L.E., Frieman, M.B., and Jahrling, P.B. (2017). Middle East Respiratory Syndrome and Severe Acute Respiratory Syndrome: Current Therapeutic Options and Potential Targets for Novel Therapies. Drugs 77, 1935-1966.

[00133] Eigenbrot, C., Ultsch, M., Lipari, M.T., Moran, P., Lin, S.J., Ganesan, R., Quan, C., Tom, J., Sandoval, W., van Lookeren Campagne, M., et al. (2012). Structural and functional analysis of HtrA1 and its subdomains. Structure 20, 1040-1050.

[00134] Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics *Chapter 5*, Unit-5 6.

[00135] Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., *et al.* (2014). Pfam: the protein families database. Nucleic Acids Res *42*, D222-230. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res *39*, W29-37.

[00136] Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., *et al.* (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem *47*, 1739-1749.

[00137] Friesner, R.A., Murphy, R.B., Repasky, M.P., Frye, L.L., Greenwood, J.R., Halgren, T.A., Sanschagrin, P.C., and Mainz, D.T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem *49*, 6177-6196.

[00138] Gao, J., Tian, Z., and Yang, X. (2020). Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Biosci Trends *14*, 72-73.

[00139] Ghinai, I., McPherson, T.D., Hunter, J.C., Kirking, H.L., Christiansen, D., Joshi, K., Rubin, R., Morales-Estrada, S., Black, S.R., Pacilli, M., *et al.* (2020). First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. Lancet.

[00140] Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163-164.

[00141] Gouveia-Oliveira, R., Sackett, P.W., and Pedersen, A.G. (2007). MaxAlign: maximizing usable data in an alignment. BMC Bioinformatics *8*, 312.

[00142] Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C., *et al.* (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med.

[00143] Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T., and Banks, J.L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem *47*, 1750-1759.

[00144] Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., Wang, L., Lupyan, D., Dahlgren, M.K., Knight, J.L., *et al.* (2016). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput *12*, 281-296.

[00145] Herter, S., Piper, D.E., Aaron, W., Gabriele, T., Cutler, G., Cao, P., Bhatt, A.S., Choe, Y., Craik, C.S., Walker, N., *et al.* (2005). Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers. Biochem J *390*, 125-136.

[00146] Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell.

[00147] 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

[00148] Infection. J Virol 93.

[00149] Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8, 275-282.

[00150] Joosten, R.P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A.C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, A.L., Deleage, G., et al. (2009). PDB_REDO: automated re-refinement of X-ray structure models in the PDB. J Appl Crystallogr 42, 376-384.

[00151] Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587-589.

[00152] Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30, 772-780.

[00153] Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F., and Matsuyama, S. (2012). Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. J Virol 86, 6537-6545.

[00154] Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10, 845-858.

[00155] 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.

[00156] Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33, 1870-1874.

[00157] Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658-1659.

[00158] Lin, B., Ferguson, C., White, J.T., Wang, S., Vessella, R., True, L.D., Hood, L., and Nelson, P.S. (1999). Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 59, 4180-4184.

[00159] Liu, Y., Gayle, A.A., Wilder-Smith, A., and Rocklov, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 27.

[00160] Lv, H., Wu, N.C., Tsang, O.T., Yuan, M., Perera, R., Leung, W.S., So, R.T.Y., Chan, J.M.C., Yip, G.K., Chik, T.S.H., et al. (2020). Cross-reactive Antibody Response between SARS-CoV-2 and SARS-CoV Infections. Cell Rep 31, 107725.

[00161] Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., Nagata, N., Sekizuka, T., Katoh, H., Kato, F., *et al.* (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proc Natl Acad Sci U S A.

[00162] Morse, J.S., Lalonde, T., Xu, S., and Liu, W.R. (2020). Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. Chembiochem *21*, 730738.

[00163] Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol *32*, 268-274.

[00164] Nishiura, H., Jung, S.M., Linton, N.M., Kinoshita, R., Yang, Y., Hayashi, K., Kobayashi, T., Yuan, B., and Akhmetzhanov, A.R. (2020a). The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020. J Clin Med *9*.

[00165] Nishiura, H., Kobayashi, T., Suzuki, A., Jung, S.M., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A.R., Linton, N.M., *et al.* (2020b). Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). Int J Infect Dis.

[00166] Page, M.J., and Di Cera, E. (2008). Serine peptidases: classification, structure and function. Cell Mol Life Sci *65*, 1220-1236.

[00167] Park, Y. (2010). TMPRSS2 (transmembrane protease, serine 2) (Atlas of Genetics and Cytogenetics in Oncology and Haematology).

[00168] Perlman, S. (2020). Another Decade, Another Coronavirus. N Engl J Med *382*, 760-762.

[00169] Qing, E., Hantak, M.P., Galpalli, G.G., and Gallagher, T. (2020). Evaluating MERS-CoV Entry Pathways. Methods Mol Biol *2099*, 9-20.

[00170] Ramanathan, R.K., Thomas, G.W., Khorana, A.A., Shah, S., Zhou, C., Wong, S., Cole, G., Jr., James, D., and Gabrail, N.Y. (2019). A Phase 2 Study of PCI-27483, a Factor VIIa Inhibitor in Combination with Gemcitabine for Advanced Pancreatic Cancer. Oncology *96*, 217-222.

[00171] Repasky, M.P., Shelley, M., and Friesner, R.A. (2007). Flexible ligand docking with Glide. Curr Protoc Bioinformatics *Chapter 8*, Unit 8 12.

[00172] Riedl, M.A., Aygoren-Pursun, E., Baker, J., Farkas, H., Anderson, J., Bernstein, J.A., Bouillet, L., Busse, P., Manning, M., Magerl, M., *et al.* (2018). Evaluation of avoralstat, an oral kallikrein inhibitor, in a Phase 3 hereditary angioedema prophylaxis trial: The OPuS-2 study. Allergy *73*, 1871-1880.

[00173] Schneider, C.A., Voth, E., Theissen, P., Wienhard, K., Wagner, R., Baer, F.M., Sechtem, U., and Schicha, H. (1994). [Assessing myocardial viability in chronic myocardial infarct with 18F-fluoro-D-glucose positron emission tomography and 99mTc-MIBI SPECT]. Z Kardiol *83*, 124-131.

[00174] Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res *31*, 3381-3385.

[00175] Schweinitz, A., Steinmetzer, T., Banke, I.J., Arlt, M.J., Sturzebecher, A., Schuster, O., Geissler, A., Giersiefen, H., Zeslawska, E., Jacob, U., *et al.* (2004). Design of novel and

selective inhibitors of urokinase-type plasminogen activator with improved pharmacokinetic properties for use as antimetastatic agents. J Biol Chem 279, 33613-33622.

[00176] Shen, L.W., Mao, H.J., Wu, Y.L., Tanaka, Y., and Zhang, W. (2017). TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. Biochimie 142, 1-10.

[00177] Shirato, K., Kawase, M., and Matsuyama, S. (2013). Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. J Virol 87, 12552-12561.

[00178] Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S., and Gallagher, T. (2011). A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. J Virol 85, 873-882.

[00179] Sneath, P.H.A., and Sokal, R.R. (1973). Unweighted Pair Group Method with Arithmetic Mean. In Numerical taxonomy The principles and practice of numerical classification (USA: San Francisco, W.H. Freeman and Company.), pp. 230-234.

[00180] Søgaard, O.S. (2020). NCT04321096: The Impact of Camostat Mesilate on COVID-19 Infection: An Investigator-initiated Randomized, Placebo-controlled, Phase IIa Trial (ClinicalTrials.gov: U.S. National Library of Medicine).

[00181] Song, H.K., and Suh, S.W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. J Mol Biol 275, 347-363.

[00182] Suda, H., Aoyagi, T., Hamada, M., Takeuchi, T., and Umezawa, H. (1972). Antipain, a new protease inhibitor isolated from actinomycetes. J Antibiot (Tokyo) 25, 263-266.

[00183] Sun, J., Zhuang, Z., Zheng, J., Li, K., Wong, R.L.Y., Liu, D., Huang, J., He, J., Zhu, A., Zhao, J., et al. (2020). Generation of a Broadly Useful Model for COVID-19 Pathogenesis, Vaccination, and Treatment. Cell. Talevi, A., and Bellera, C.L. (2020). Challenges and opportunities with drug repurposing: finding strategies to find alternative uses of therapeutics. Expert Opin Drug Discov 15, 397-401.

[00184] Team, C.C.-R. (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. In Morbidity and Mortality Weekly Report (MMWR) (Centers for Disease Control and Prevention).

[00185] Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science 347, 1260419.

[00186] Velavan, T.P., and Meyer, C.G. (2020). The COVID-19 epidemic. Trop Med Int Health 25, 278-280.

[00187] Velez, G., Sun, Y.J., Khan, S., Yang, J., Herrmann, J., Chemudupati, T., MacLaren, R.E., Gakhar, L., Wakatsuki, S., Bassuk, A.G., et al. (2020). Structural Insights into the Unique Activation Mechanisms of a Nonclassical Calpain and Its Disease-Causing Variants. Cell Rep 30, 881-892 e885.

[00188] Wang, Y., Dong, C., Hu, Y., Li, C., Ren, Q., Zhang, X., Shi, H., and Zhou, M. (2020). Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. Radiology, 200843.

[00189] Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18, 691-699.

[00190] Zhou, D., Dai, S.M., and Tong, Q. (2020a). COVID-19: a recommendation to examine the effect of hydroxychloroquine in preventing infection and progression. J Antimicrob Chemother.

[00191] Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al. (2020b). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270273.

[00192] Zhou, Y., Vedantham, P., Lu, K., Agudelo, J., Carrion, R., Jr., Nunneley, J.W., Barnard, D., Pohlmann, S., McKerrow, J.H., Renslo, A.R., et al. (2015). Protease inhibitors targeting coronavirus and filovirus entry. Antiviral Res 116, 76-84.

[00193] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 382, 727-733.

Supplemental References

[00194] Avgeris, M., and Scorilas, A. (2016). Kallikrein-related peptidases (KLKs) as emerging therapeutic targets: focus on prostate cancer and skin pathologies. Expert Opin Ther Targets 20, 801-818.

[00195] Baba, T., Azuma, S., Kashiwabara, S., and Toyoda, Y. (1994). Sperm from mice carrying a targeted mutation of the acrosin gene can penetrate the oocyte zona pellucida and effect fertilization. J Biol Chem 269, 3184531849.

[00196] Borodin, E.A., Pamirsly, I.E., Shtarberg, M.A., Dorovskikh, V.A., Korotkikh, A.V., Tarumizu, C., Takamatsu, K., and Yamamoto, S. (2013). Effects of Soybean Trypsin Inhibitor on Hemostasis. In Soybean - Bio-Active Compounds, H.A. El-Shemy, ed. (IntechOpen).

[00197] Chevillet, J.R., Park, G.J., Bedalov, A., Simon, J.A., and Vasioukhin, V.I. (2008). Identification and characterization of small-molecule inhibitors of hepsin. Mol Cancer Ther 7, 3343-3351.

[00198] de Vries, S.J., van Dijk, M., and Bonvin, A.M. (2010). The HADDOCK web server for data-driven biomolecular docking. Nat Protoc 5, 883-897.

[00199] Fischer, P.M. (2018). Design of Small-Molecule Active-Site Inhibitors of the S1A Family Proteases as Procoagulant and Anticoagulant Drugs. J Med Chem 61, 3799-3822.

[00200] Flechon, J.E. (2016). The acrosome of eutherian mammals. Cell Tissue Res 363, 147-157.

[00201] Ganesan, R., Kolumam, G.A., Lin, S.J., Xie, M.H., Santell, L., Wu, T.D., Lazarus, R.A., Chaudhuri, A., and Kirchhofer, D. (2011). Proteolytic activation of pro-macrophage-stimulating protein by hepsin. Mol Cancer Res 9, 1175-1186.

[00202] Herter, S., Piper, D.E., Aaron, W., Gabriele, T., Cutler, G., Cao, P., Bhatt, A.S., Choe, Y., Craik, C.S., Walker, N., et al. (2005). Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers. Biochem J 390, 125-136.

[00203] Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., and Pohlmann, S. (2014). TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. J Virol 88, 1293-1307.

[00204] Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell.

[00205] Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F., and Matsuyama, S. (2012). Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. J Virol 86, 6537-6545.

[00206] Kwon, H., Han, J., Lee, K.Y., Son, S.H., and Byun, Y. (2017). Recent Advances of Hepsin-Targeted Inhibitors. Curr Med Chem 24, 2294-2311.

[00207] Lin, J., Deng, H., Jin, L., Pandey, P., Quinn, J., Cantin, S., Rynkiewicz, M.J., Gorga, J.C., Bibbins, F., Celatka, C.A., et al. (2006). Design, synthesis, and biological evaluation of peptidomimetic inhibitors of factor XIa as novel anticoagulants. J Med Chem 49, 7781-7791.

[00208] Neefjes, J.J., and Ploegh, H.L. (1992). Inhibition of endosomal proteolytic activity by leupeptin blocks surface expression of MHC class II molecules and their conversion to SDS resistance alpha beta heterodimers in endosomes. EMBO J *11*, 411-416.

[00209] Ning, W., Zhu, J., Zheng, C., Liu, X., Song, Y., Zhou, Y., Zhang, X., Zhang, L., Sheng, C., and Lv, J. (2013). Fragment-based design of novel quinazolinon derivatives as human acrosin inhibitors. Chem Biol Drug Des *81*, 437-441.

[00210] Paul, J.J., Kircus, S.R., Sorrell, T.N., Ropp, P.A., and Thorp, H.H. (2006). Effects of coordinating metal ions on the mediated inhibition of trypsin by bis(benzimidazoles) and related compounds. Inorg Chem *45*, 5126-5135. Perona, J.J., and Craik, C.S. (1995). Structural basis of substrate specificity in the serine proteases. Protein Sci *4*, 337-360.

[00211] Ramanathan, R.K., Thomas, G.W., Khorana, A.A., Shah, S., Zhou, C., Wong, S., Cole, G., Jr., James, D., and Gabrail, N.Y. (2019). A Phase 2 Study of PCI-27483, a Factor VIIa Inhibitor in Combination with Gemcitabine for Advanced Pancreatic Cancer. Oncology *96*, 217-222.

[00212] Riedl, M.A., Aygoren-Pursun, E., Baker, J., Farkas, H., Anderson, J., Bernstein, J.A., Bouillet, L., Busse, P., Manning, M., Magerl, M., *et al.* (2018). Evaluation of avoralstat, an oral kallikrein inhibitor, in a Phase 3 hereditary angioedema prophylaxis trial: The OPuS-2 study. Allergy *73*, 1871-1880.

[00213] Song, H.K., and Suh, S.W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. J Mol Biol *275*, 347-363.

[00214] Suda, H., Aoyagi, T., Hamada, M., Takeuchi, T., and Umezawa, H. (1972). Antipain, a new protease inhibitor isolated from actinomycetes. J Antibiot (Tokyo) *25*, 263-266.

[00215] Tang, X., Mahajan, S.S., Nguyen, L.T., Beliveau, F., Leduc, R., Simon, J.A., and Vasioukhin, V. (2014). Targeted inhibition of cell-surface serine protease Hepsin blocks prostate cancer bone metastasis. Oncotarget *5*, 1352-1362.

[00216] Tummon, I.S., Yuzpe, A.A., Daniel, S.A., and Deutsch, A. (1991). Total acrosin activity correlates with fertility potential after fertilization in vitro. Fertil Steril *56*, 933-938.

[00217] Welker, B., Bernstein, G.S., Diedrich, K., Nakamura, R.M., and Krebs, D. (1988). Acrosomal proteinase activity of human spermatozoa and relation of results to semen quality. Hum Reprod *3 Suppl 2*, 75-80.

[00218] Willbold, R., Wirth, K., Martini, T., Sultmann, H., Bolenz, C., and Wittig, R. (2019). Excess hepsin proteolytic activity limits oncogenic signaling and induces ER stress and autophagy in prostate cancer cells. Cell Death Dis *10*, 601.

[00219] Yamamoto, M., Matsuyama, S., Li, X., Takeda, M., Kawaguchi, Y., Inoue, J.I., and Matsuda, Z. (2016). Identification of Nafamostat as a Potent Inhibitor of Middle East Respiratory Syndrome Coronavirus S Protein-Mediated Membrane Fusion Using the Split-Protein-Based Cell-Cell Fusion Assay. Antimicrob Agents Chemother *60*, 6532-6539.

[00220] Yeste, M., Jones, C., Amdani, S.N., and Coward, K. (2017). Oocyte Activation and Fertilisation: Crucial Contributors from the Sperm and Oocyte. Results Probl Cell Differ *59*, 213-239.

[00221] Zhao, J., Tian, W., Qi, J., Lv, D., Liu, Y., Jiang, Y., Dong, G., Chen, Q., Zhou, Y., Zhu, J., et al. (2014). Design and synthesis of phenylisoxazole derivatives as novel human acrosin inhibitors. Bioorg Med Chem Lett 24, 2802-2806.

[00222] Zhou, Y., Vedantham, P., Lu, K., Agudelo, J., Carrion, R., Jr., Nunneley, J.W., Barnard, D., Pohlmann, S., McKerrow, J.H., Renslo, A.R., et al. (2015). Protease inhibitors targeting coronavirus and filovirus entry. Antiviral Res 116, 76-84.

Table 1. Conserved residues in S1-peptidases: Conserved residues in 600 S1 peptidases across all species. Amino acids are numbered according their position in human TMPRSS2.

Position	Amino Acid	MSA Conservation (%)	TMPRSS2 Function
259	G	99.3	-
270	Q	60.8	-
281	C	95.0	S1' sub-binding pocket
293	T	81.2	-
294	A	95.0	-
295	A	92.8	-
296	H	92.3	Catalytic His residue; Active site
297	C	95.8	S1' sub-binding pocket
313	G	97.5	-
345	D	99.0	Catalytic Asp residue; Active site
347	A	45.8	-
351	L	93.5	-
366	L	93.3	-
383	G	99.2	-
384	W	92.0	-
385	G	98.5	-
397	L	74.8	-
410	C	100	-
426	C	100	-
427	A	86.8	-
437	C	94.2	S1 sub-binding pocket
440	D	98.5	S1' sub-binding pocket
441	S	93.3	Catalytic Ser residue; Active

			site
442	G	99.3	-
443	G	94.8	-
444	P	94.8	-
445	L	78.7	-
457	G	99.3	-
460	S	92.0	S2 sub-binding pocket
465	C	99.7	S1 sub-binding pocket
471	P	84.5	-
474	Y	85.3	S1 sub-binding pocket
483	W	100	-
484	I	88.2	-

Table 2. Structures used in the structure-based phylogenetics analysis. ** - denotes structural models re-refined in PDB-REDO.*

Protein	Species	Residues	PDB	Chain	PDB residues	Resolution (Å)	Ligand?
Acrosin	Pig	40 - 283	1FIZ*	A	16 - 238	2.90	Yes
Acrosin	Sheep	40 - 283	1FIW*	A	16 - 238	2.10	Yes
Azurocidin	Human	27 - 239	1A7S*	A	1 - 213	1.12	No
Cathepsin G	Human	21 - 238	1AU8	A	16 - 238	1.90	Yes
Chymotrypsin-Like Elastase 1	Pig	27 - 259	1B0E*	A	16 - 238	1.80	Yes
Chymotrypsin-Like Elastase 2	Pig	29 - 262	1BRU	P	16 - 238	2.30	Yes
Complement Factor D	Human	26 - 248	1BIO*	A	16 - 238	1.50	Yes
Complement Factor D	Mouse	26 - 249	5FCR*	A	16 - 238	1.25	No
Complement Factor I	Human	340 - 569	2XRC*	B	322 - 551	2.69	No
Complement Factor 1R	Human	464 - 697	1GPZ*	A	447 - 680	2.90	No
Complement Factor 1S	Human	438 - 675	1ELV	A	423 - 660	1.70	No
Chymase	Human	22 - 240	1KLT*	A	16 - 238	1.90	Yes
Chymotrypsinogen B	Rat	169 - 256	1KDQ*	B	151 - 238	2.55	No
Chymotrypsin-C	Bovine	30 - 262	1PYT	D	716 - 938	2.35	No
Chymotrypsin-C	Human	30 - 262	4H4F*	A	16 - 238	1.90	Yes
Duodenase-1	Bovine	20 - 238	1EUF*	A	16 - 238	2.40	No
Enteropeptidase	Bovine	801 - 1030	1EKB	B	16 - 238	2.30	Yes
Coagulation Factor VII	Human	213 - 447	1CVW*	H	16 - 238	2.28	Yes
Coagulation Factor IX	Human	227 - 454	1RFN	A	16 - 238	2.80	Yes

Coagulation Factor IX	Pig	183 - 409	1PFX	C	16 - 237	3.00	Yes
Coagulation Factor X	Bovine	234 - 461	1KIG	H	16 - 238	3.00	Yes
Coagulation Factor X	Human	235 - 462	1C5M*	D	16 - 238	1.95	No
Coagulation Factor XI	Human	388 - 618	1XX9*	A	16 - 238	2.20	Yes
Coagulation Factor XII	Human	373 - 609	6B74*	B	16 - 238	2.32	Yes
Coagulation Factor XIV	Human	212 - 445	1AUT	C	16 - 238	2.80	Yes
Granzyme A	Human	29 - 254	1OP8*	A	16 - 238	2.50	No
Granzyme B	Human	21 - 240	1FQ3	A	16 - 238	3.10	No
Granzyme B	Rat	21 - 241	1FI8	A	16 - 238	2.20	Yes
Granzyme C	Mouse	21 - 241	3FZZ*	A	21 - 241	2.50	No
Granzyme H	Human	21 - 239	3TJU*	A	16 - 234	2.70	Yes
Granzyme K	Human	27 - 254	1MZA*	A	16 - 238	2.23	No
Granzyme M	Human	26 - 249	2ZGC*	A	1 - 224	1.96	No
Haptoglobin	Human	162 - 399	4WJG*	2	158 - 395	3.10	No
Haptoglobin	Pig	103 - 340	4F4O*	C	103 - 340	2.90	No
Hepatocyte Growth Factor	Human	495 - 716	1SHY*	A	495 - 716	3.22	No
Hepatocyte Growth Factor A	Human	408 - 641	1YBW*	A	408 - 641	2.70	No
Hepatocyte Growth Factor L	Human	484 - 704	2ASU*	B	484 - 704	1.85	No
Hepsin	Human	163 - 400	1O5E*	H	16 - 238	1.75	Yes
Kallikrein-1	Human	25 - 254	1SPJ*	A	16 - 238	1.70	No
Kallikrein-2	Horse	25 - 253	1GVZ	A	16 - 238	1.42	No
Kallikrein-2	Human	25 - 253	4NFE*	A	16 - 238	1.90	Yes
Kallikrein-2	Rat	25 - 251	1TON	A	16 - 238	1.80	No
Kallikrein-3	Human	25 - 253	2ZCH*	P	16 - 238	2.83	No
Kallikrein-4	Human	31 - 247	2BDG*	A	16 - 238	1.95	Yes
Kallikrein-5	Human	67 - 285	2PSX*	A	16 - 238	2.30	Yes
Kallikrein-6	Human	22 - 237	1GVL	A	16 - 238	1.80	No
Kallikrein-7	Human	30 - 245	2QXG*	A	16 - 238	2.60	Yes
Kallikrein-8	Human	33 - 252	5MS3*	A	16 - 238	2.30	No
Kallikrein-8	Mouse	33 - 252	1NPM	A	16 - 238	2.10	No
Kallikrein-10	Human	50 - 269	5LPE*	B	20 - 238	2.65	No
Kallikrein-13	Mouse	25 - 253	1AO5*	A	16 - 238	2.60	No
Kallikrein-Related B1 (KLKB1)	Human	391 - 621	2ANW*	A	16 - 238	1.85	Yes
Kallikrein-Related B1 (KLKB1)	Mouse	391 - 621	5GVT*	A	16 - 238	2.61	No
Kallikrein-Related B3 (KLKB3)	Mouse	25 - 253	1SGF*	G	16 - 238	3.15	No
Kallikrein-Related B4 (KLKB4)	Mouse	29 - 248	1SGF*	A	25 - 238	3.15	No
Mast Cell Protease 2	Rat	21 - 239	3RP2*	A	16 - 238	1.90	No

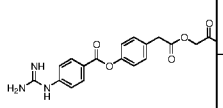
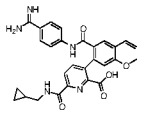
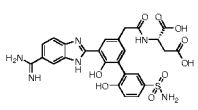
MBL Serine Protease 1	Human	449 - 691	3GOV*	B	449 - 691	2.55	No
MBL Serine Protease 2	Human	445 - 679	1Q3X*	A	445 - 679	2.23	No
Membrane-Type Serine Protease 1	Human	615 - 849	1EAW	A	16 - 238	2.93	Yes
Myeloblastin	Human	28 - 243	1FUJ*	A	16 - 238	2.20	No
Neutrophil Elastase	Human	30 - 242	1B0F*	A	16 - 238	3.00	Yes
Proproteinase-E	Bovine	12 - 246	1PYT	C	416 - 638	2.35	No
Prostatin	Human	45 - 281	3DFJ*	A	45 - 281	1.45	No
Serine Protease 57	Human	34 - 258	4Q7X*	A	16 - 238	2.55	No
Thrombin B	Human	364 - 430	2HNT	C	16 - 72	2.50	No
Thrombin B	Mouse	361 - 610	2OCV*	B	16 - 238	2.20	No
Tissue Plasminogen	Human	311 - 556	1A5H	A	16 - 238	2.90	Yes
Trypsin	Boar	9-231	1AVW*	A	9-231	1.75	Yes
Trypsin	Pig	134 - 224	1AKS*	B	146 - 238	1.80	No
Trypsin-1	Bovine	24 - 239	1AQ7*	A	16 - 238	2.20	Yes
Trypsin-2	Rat	24 - 239	1AMH*	A	16 - 238	2.50	No
Trypsin-3	Human	81 - 296	1H4W*	A	16 - 238	1.70	Yes
Trypsin B1	Human	31 - 267	1LTO	A	16 - 238	2.20	No
Trypsin B2	Human	31 - 267	1A0L	A	16 - 238	3.00	Yes
Urokinase	Human	179 - 419	1C5W*	B	16 - 238	1.94	Yes
Urokinase	Mouse	180 - 421	5LHN*	A	16 - 238	2.55	Yes

Table 3. Inhibitors targeting TMPRSS2 and structurally similar proteases

Inhibitor(s)	Disease/Indication	Target(s)	Phase	Reference(s)
Camostat mesylate	SARS-CoV-1 SARS-CoV-2	TMPRSS2	Preclinical	(Hoffmann et al., 2020; Kawase et al., 2012; Zhou et al., 2015)
Nafamostat mesylate	MERS-CoV SARS-CoV-2	TMPRSS2	Preclinical	(Yamamoto et al., 2016)
Avoralstat	Angioedema	Plasma Kallikrein	Phase 3	(Riedl et al., 2018)
PCI-27483	Pancreatic Cancer	Factor VII	Phase 2	(Ramanathan et al., 2019)
Antipain	Biochemical Assays	Trypsin/Papain	Experimental	(Suda et al., 1972)

Leupeptin	Biochemical Assays	Trypsin/Papain	Experimental	(Neefjes and Ploegh, 1992)
Hepsin Inhibitors (H1 – 12)	Prostate Cancer	Hepsin	Experimental	(Chevillet et al., 2008)
Trypsin Inhibitors (T1 – 6)	Biochemical Assays	Trypsin	Experimental	(Paul et al., 2006)
Factor VII Inhibitors (FVII1 – 19)	Coagulopathy	Factor VII	Experimental	(Fischer, 2018)
Factor XI Inhibitors (FXI1 – 32)	Coagulopathy	Factor XI	Experimental	(Lin et al., 2006)
Acrosin Inhibitors (A1 – 16)	Contraception	Acrosin	Experimental	(Zhao et al., 2014)
Plasma Kallikrein Inhibitors (K1 – 5)	Prostate Cancer	Plasma Kallikrein	Experimental	(Avgeris and Scorilas, 2016)

Table 4. Docking scores of protease inhibitors and substrates to TMPRSS2 and related peptidase domains ** - denotes the compound did not result docking score. Numbers in parentheses indicates STDEV.*

Inhibitor	**Structure**	**Docking Tool**	**S1-Peptidase Domain**			
			TMPRSS2	**KLKB1**	**Factor VII**	**Trypsin**
SARS-CoV-2 Spike S2'	$^{+}H_3N$-KPSKRSF-COO^{-}	Glide	-9.67	-10.68	-8.56	-10.02
		HADDOCK	-97.0 (2.7)	-113 (1.3)	-92.2 (2.2)	-79.3 (2.7)
Camostat		Glide	-4.68	0.45	-4.14	N.D.*
		HADDOCK	-51.6 (2.3)	-48.4 (2.0)	-42.4 (0.9)	-42.7 (0.2)
Avoralstat		Glide	-7.78	-5.59	-1.74	-4.22
		HADDOCK	-70.6 (0.7)	-78.1 (2.6)	-56.0 (1.1)	-56.5 (0.5)
PCI-27483		Glide	-8.78	-5.63	-12.7	-9.15
		HADDOCK	-106 (2.7)	-81.5 (1.2)	-84.7 (1.3)	-90.2 (5.2)
Antipain		Glide	-9.77	-6.71	-8.98	-8.98

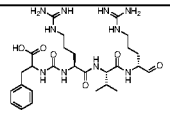
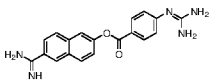
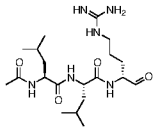
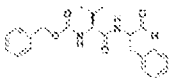
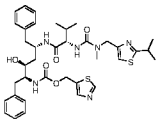
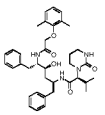
		HADDOCK	-75.1 (1.4)	-72.7 (4.4)	-61.2 (3.6)	-56.8 (2.2)
Nafamostat		Glide	-8.89	-3.12	-7.82	-6.77
		HADDOCK	-72.8 (4.0)	-60.4 (1.4)	-44.8 (1.8)	-56.0 (0.5)
Leupeptin		Glide	-7.08	-5.54	-8.22	-7.57
		HADDOCK	-49.6 (1.6)	-48.1 (2.6)	-35.3 (1.8)	-41.1 (2.6)
MDL-28170		Glide	-5.55	-7.68	-6.31	-5.41
		HADDOCK	-39.2 (1.3)	-40.8 (3.2)	-29.0 (1.2)	-37.0 (0.7)
Ritonavir		Glide	-6.02	-6.82	-5.10	-6.15
		HADDOCK	-53.5 (1.4)	-54.0 (1.6)	-44.4 (4.0)	-46.4 (2.6)
Lopinavir		Glide	-6.33	-8.13	-6.18	-6.75
		HADDOCK	-45.4 (1.1)	-42.6 (3.1)	-43.5 (2.1)	-50.5 (1.3)
SBTI Peptide	$^{+}H_3N$-PWRIRF-COO^{-}	Glide	-9.18	-9.35	-7.50	-7.97
		HADDOCK	-77.3 (0.5)	-82.5 (1.3)	-58.2 (2.5)	-67.5 (2.5)

Table 5. Docking scores of SBTI to TMPRSS2 and related S1-peptidase domains

S1-Peptidase Domain	**TMPRSS2**	**Trypsin**	**KLKB1**	**Factor VII**
PDB ID	Model	1AVW	6O1S	1W7X
HADDOCK Score (STDEV)	-154.6 (7.5)	-171.7 (1.3)	-140.1 (6.1)	-87.7 (2.0)
Z-Score	-1.8	-1.4	-1.7	-1.5

That which is claimed is:

1. A method of treating a subject for an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the method comprising administering to the subject a therapeutically effective amount of a composition comprising a plasma kallikrein inhibitor, a coagulation factor VII inhibitor, or a trypsin inhibitor.

2. The method of claim 1, wherein the plasma kallikrein inhibitor is Avoralstat.

3. The method of claim 1, wherein the coagulation factor VII inhibitor is PCI-27483.

4. The method of claim 1, wherein the trypsin inhibitor is soybean trypsin inhibitor (SBTI).

5. The method of any one of claims 1 to 4, wherein the composition is administered prophylactically or therapeutically.

6. The method of any one of claims 1 to 5, wherein multiple cycles of treatment are administered to the subject for a time period sufficient to eradicate the infection by the SARS-CoV-2.

7. The method of any one of claims 1 to 6, wherein the composition is administered according to a daily dosing regimen or intermittently.

8. The method of any one of claims 1 to 7, wherein the composition is administered orally, intravenously, or by pulmonary inhalation.

9. The method of claim 8, wherein pulmonary inhalation is performed with a nebulizer, a metered dose inhaler (MDI), and a dry powder inhaler (DPI).

10. The method of any one of claims 1 to 9, further comprising administering additional anti-viral therapy.

11. The method of any one of claims 1 to 10, wherein the subject is mammalian.

12. The method of claim 11, wherein the subject is human.

13. The method of any one of claims 1 to 12, wherein the composition is administered in an amount sufficient to reduce viral entry or replication of the SARS-CoV-2 in the subject.

14. A composition comprising a plasma kallikrein inhibitor, a coagulation factor VII inhibitor, or a trypsin inhibitor for use in the treatment of an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

15. The composition of claim 14, further comprising a pharmaceutically acceptable excipient.

16. The composition of claim 14 or 15, wherein the plasma kallikrein inhibitor is Avoralstat.

17. The composition of claim 14 or 15, wherein the coagulation factor VII inhibitor is PCI-27483.

18. The composition of claim 14 or 15, wherein the trypsin inhibitor is soybean trypsin inhibitor (SBTI).

19. A method of screening for an S1 peptidase inhibitor that inhibits an infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the method comprising:

a) modeling a TMPRSS2 S1 peptidase domain by using a structure of hepsin as a template;
b) calculating a TMPRSS2 structural similarity score for a plurality of S1 peptidases, wherein S1 peptidases with the highest structural similarity scores are prioritized for further *in silico* analysis;
c) modeling a TMPRSS2 binding pocket by comparison to a structure of a hepsin binding pocket;
d) docking a plurality of inhibitors of the prioritized S1 peptidases into the modeled TMPRSS2 binding pocket, wherein the inhibitors are prioritized for further screening based on their docking scores, wherein higher docking scores indicate more favorable binding interactions in the TMPRSS2 binding pocket;
e) measuring inhibition of TMPRSS2 activity by one or more of the inhibitors selected for further screening, wherein inhibitors are prioritized according to how strongly they inhibit TMPRSS2; and

f) measuring inhibition of an infection by the SARS-CoV-2 in a host cell or subject by one or more of the inhibitors that inhibited TMPRSS2 to identify an inhibitor that inhibits the infection by the SARS-CoV-2.

20. The method of claim 19, wherein the inhibitor has been previously approved by the U.S. Food and Drug Administration for treatment of a disease or condition other than the SARS-CoV-2 infection.

21. An inhibitor identified by the method of claim 19 or 20 for the treatment of a SARS-CoV-2 infection.


S1-Peptidase Superfamily
Kallikrein-1
Kallikrein-4
Snake Venom Serine Protease
Granzyme G
Trypsin-2
Azuricidin
Plasminogen
Chymotrypsin-Like
KLKB1
TMPRSS2
Hepsin
Factor XI
Factor VII
Haptoglobin
0.6 (a.u.)

FIG. 1A

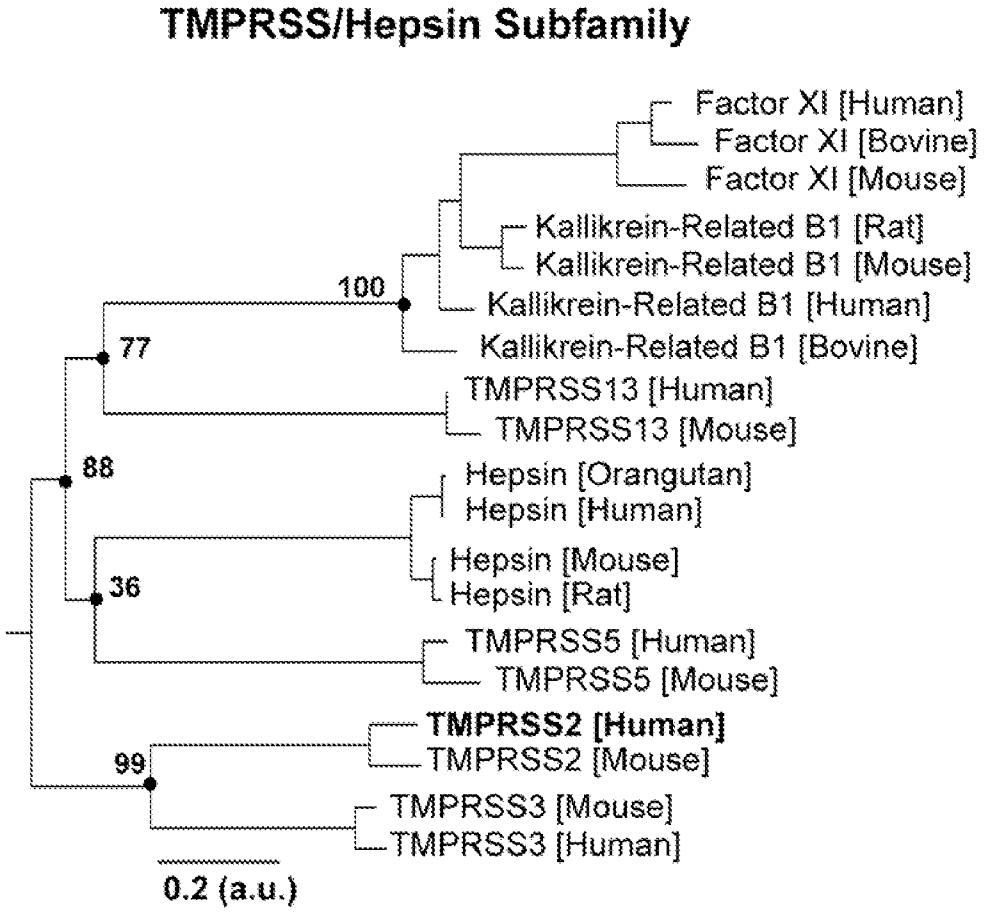

TMPRSS/Hepsin Subfamily
Factor XI [Human]
Factor XI [Bovine]
Factor XI [Mouse]
Kallikrein-Related B1 [Rat]
Kallikrein-Related B1 [Mouse]
Kallikrein-Related B1 [Human]
Kallikrein-Related B1 [Bovine]
TMPRSS13 [Human]
TMPRSS13 [Mouse]
Hepsin [Orangutan]
Hepsin [Human]
Hepsin [Mouse]
Hepsin [Rat]
TMPRSS5 [Human]
TMPRSS5 [Mouse]
TMPRSS2 [Human]
TMPRSS2 [Mouse]
TMPRSS3 [Mouse]
TMPRSS3 [Human]
100
77
88
36
99
0.2 (a.u.)

FIG. 1B

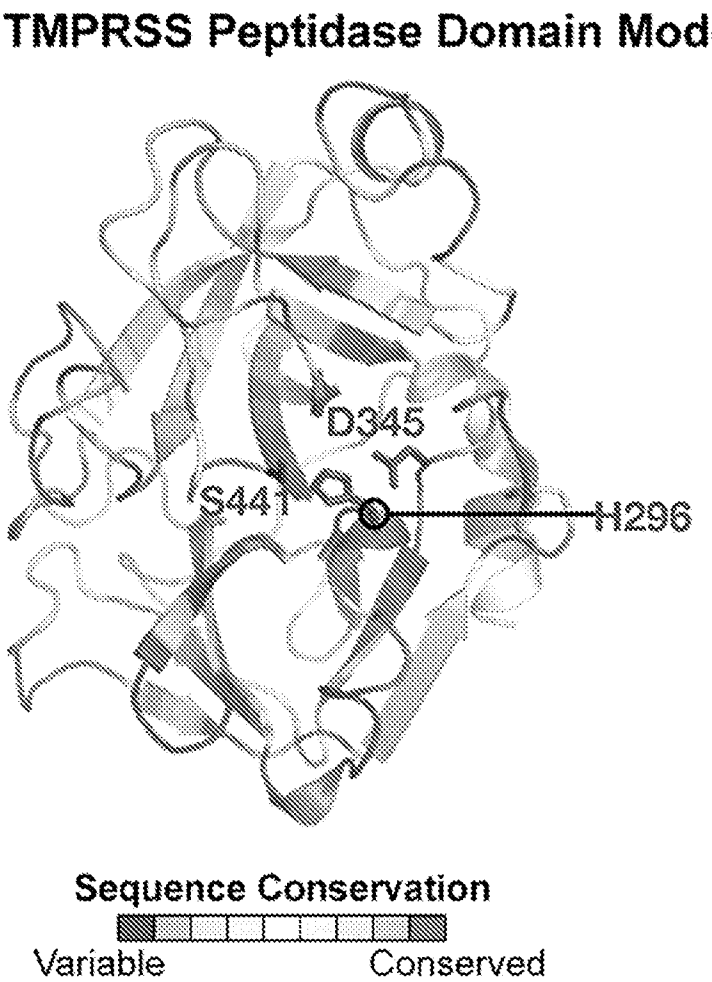

TMPRSS Peptidase Domain Model
D345
S441
H296
Sequence Conservation
Variable
Conserved

FIG. 1C

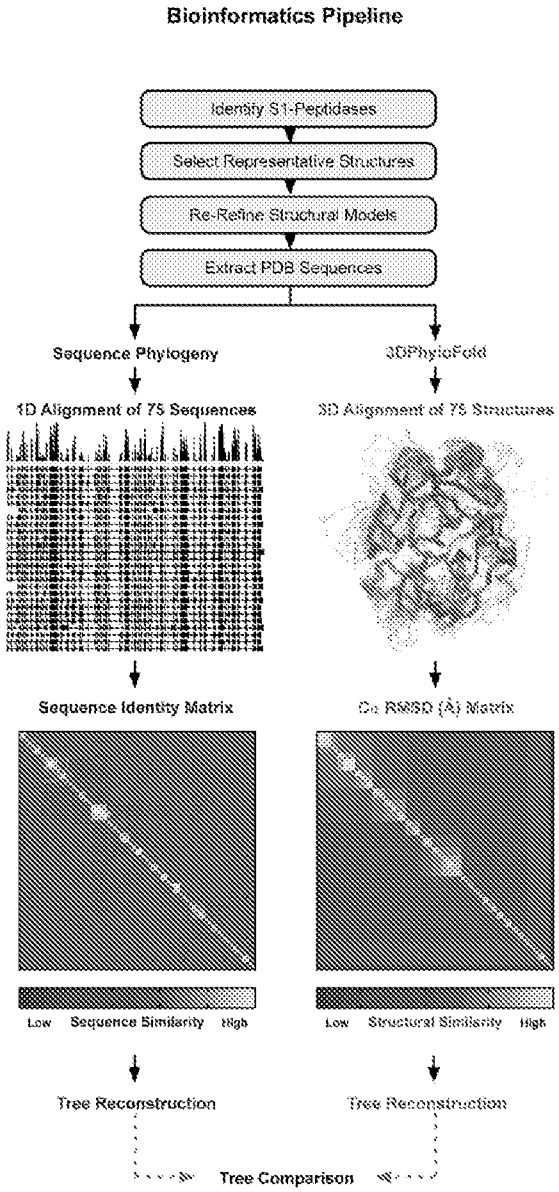

Bioinformatics Pipeline
Identify S1-Peptidases
Select Representative Structures
Re-Refine Structural Models
Extract PDB Sequences
Sequence Phylogeny
3DPhyloFold
1D Alignment of 75 Sequences
3D Alignment of 75 Structures
Sequence Identity Matrix
Cα RMSD (Å) Matrix
Low Sequence Similarity High
Low Structural Similarity High
Tree Reconstruction
Tree Reconstruction
Tree Comparison

FIG. 2A

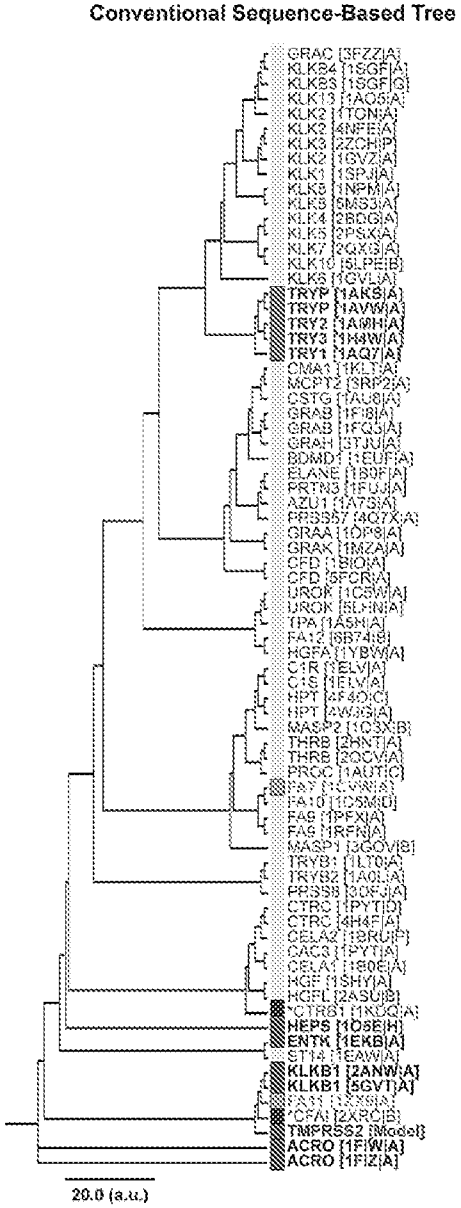

Conventional Sequence-Based Tree
GRAC [3FZZ|A]
KLKB4 [1SGF|A]
KLKB3 [1SGF|G]
KLK13 [1AO5|A]
KLK2 [1TON|A]
KLK2 [4NFE|A]
KLK3 [2ZCH|P]
KLK2 [1GVZ|A]
KLK1 [1SPJ|A]
KLK8 [1NPM|A]
KLK9 [5M53|A]
KLK4 [2BDG|A]
KLK5 [2PSX|A]
KLK7 [2QXG|A]
KLK10 [5LPE|B]
KLK6 [1GVL|A]
TRYP [1AKS|A]
TRYP [1AVW|A]
TRY2 [1A0H|A]
TRY3 [1H4W|A]
TRY1 [1AQ7|A]
CMA1 [1KLT|A]
MCPT2 [3RP2|A]
CSTG [1AU8|A]
GRAB [1FI8|A]
GRAB [1FQ3|A]
GRAH [3TJU|A]
BDMD1 [1EUF|A]
ELANE [1B0F|A]
PRTN3 [1FUJ|A]
AZU1 [1A7S|A]
PRSS57 [4Q7X|A]
GRAA [1DP8|A]
GRAK [1MZA|A]
CFD [1BIO|A]
CFD [5FCR|A]
UROK [1CBW|A]
UROK [5LHN|A]
TPA [1A5H|A]
FA12 [5B74|B]
HGFA [1YBW|A]
C1R [1ELV|A]
C1S [1ELV|A]
HPT [4F4D|C]
HPT [4W9S|A]
MASP2 [1C3X|B]
THRB [2HNT|A]
THRB [2OCV|A]
PROC [1AUT|C]
FA7 [1L2W|A]
FA10 [1CSM|D]
FA9 [1PFX|A]
FA9 [1RFN|A]
MASP1 [3GOV|B]
TRYB1 [1LTO|A]
TRYB2 [1A0L|A]
PRSS3 [3DFJ|A]
CTRC [1PYT|D]
CTRC [4H4F|A]
CELA2 [1BRU|P]
CAC3 [1PYT|A]
CELA1 [1B0E|A]
RGP [1SHY|A]
HGFL [2ASU|B]
*CTRB1 [1KDQ|A]
HEPS [1O5E|H]
ENTK [1EKB|A]
ST14 [1EAW|A]
KLKB1 [2ANW|A]
KLKB1 [5GVT|A]
FA11 [2XRB|A]
*CFAI [2XRC|B]
TMPRSS2 [Model]
ACRO [1F1W|A]
ACRO [1F1Z|A]
20.0 (a.u.)

FIG. 2B

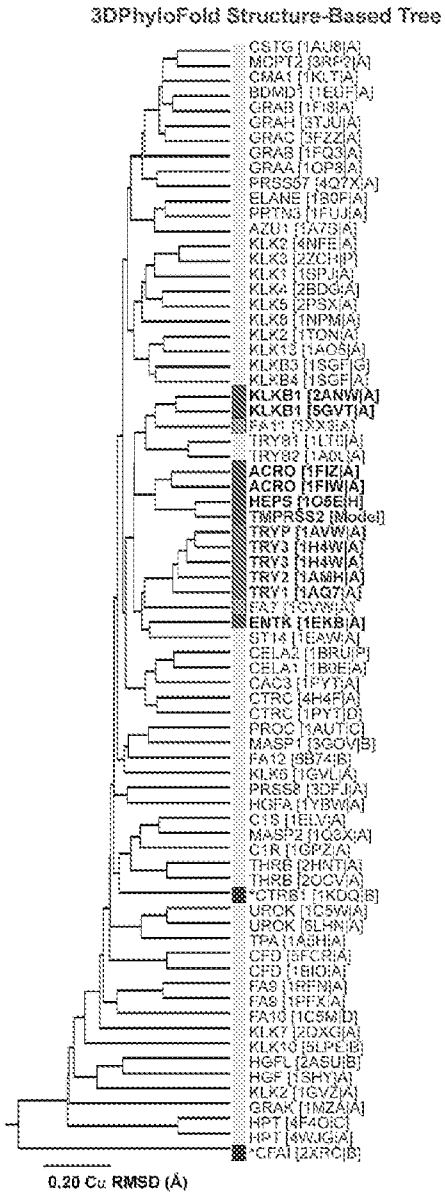

3DPhyloFold Structure-Based Tree
GSTG [1AUS|A]
MDPT2 [3RP3|A]
C5AA1 [1KLT|A]
BDMD1 [1EGF|A]
GRAB [1FI8|A]
GRAH [3TJU|A]
GRAC [3FZZ|A]
GRAB [1FQ3|A]
GRAA [1OP8|A]
PRSS57 [4Q7X|A]
ELANE [1B0F|A]
PRTN3 [1FUJ|A]
AZU1 [5A7S|A]
KLK2 [4NFE|A]
KLK3 [2ZCH|P]
KLK1 [1SPJ|A]
KLK4 [2BDG|A]
KLK5 [2PSX|A]
KLK6 [1NPM|A]
KLK2 [1TON|A]
KLK13 [1AO5|A]
KLKB3 [1SGF|G]
KLKB4 [1SGF|A]
KLKB1 [2ANW|A]
KLKB1 [5GVT|A]
FA11 [5XX3|A]
TRY81 [1LTI|A]
TRY82 [1A0L|A]
ACRO [1FIZ|A]
ACRO [1FIW|A]
HEPS [1O5E|H]
TMPRSS2 [Model]
TRYP [1AVW|A]
TRY3 [1H4W|A]
TRY3 [1H4W|A]
TRY2 [1AMH|A]
TRY1 [1AG7|A]
FA7 [1CVW|A]
ENTK [1EKB|A]
ST14 [1EAW|A]
CELA2 [1BRU|P]
CELA1 [1B0E|A]
CAC3 [1PYT|A]
CTRC [4H4F|A]
CTRC [1PYT|D]
PROC [1AUT|C]
MASP1 [3GOV|B]
FA12 [5B74|B]
KLK5 [1GVL|A]
PRSS8 [3DFJ|A]
HGFA [1YBW|A]
C1S [1ELV|A]
MASP2 [1C3X|A]
C1R [1GPZ|A]
THRB [2HNT|A]
THRB [2OCV|A]
*CTRB1 [1KDQ|B]
UROK [1C5W|A]
UROK [5LHN|A]
TPA [1A5H|A]
CFD [5FCR|A]
CFD [1BIO|A]
FA9 [1RFN|A]
FA8 [1PFX|A]
FA10 [1C5M|D]
KLK7 [2DXG|A]
KLK10 [5LPE|B]
HGFL [2ASU|B]
HGF [1SHY|A]
KLK2 [1GVZ|A]
GRAK [1MZA|A]
HPT [4F4O|C]
HPT [4WJG|A]
*CFAI [2XRC|B]
0.20 Cα RMSD (Å)

FIG. 2C

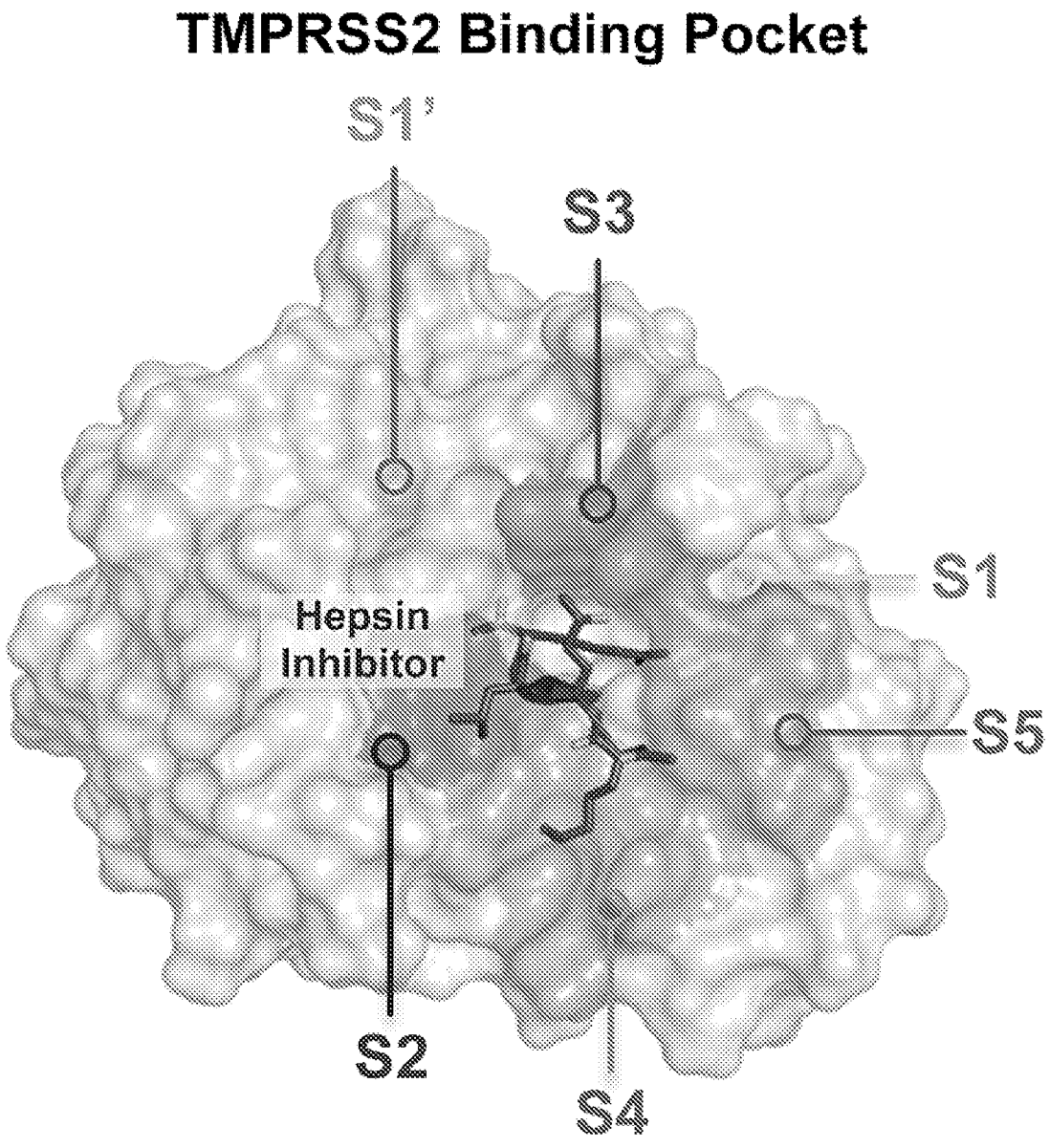


FIG. 3A

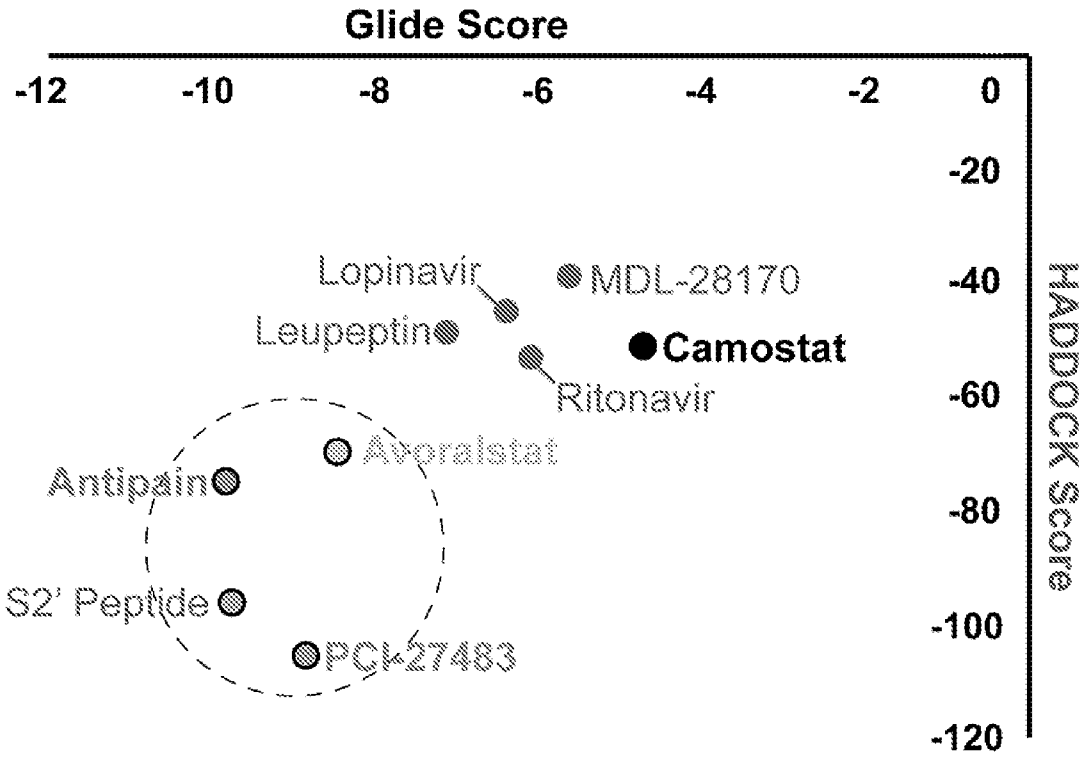

Glide Score
-12
-10
-8
-6
-4
-2
0
-20
-40
-60
-80
-100
-120
HADDOCK Score
Lopinavir
MDL-28170
Leupeptin
Camostat
Ritonavir
Aprotinin
Antipain
S2' Peptide
PCK27483

FIG. 3B

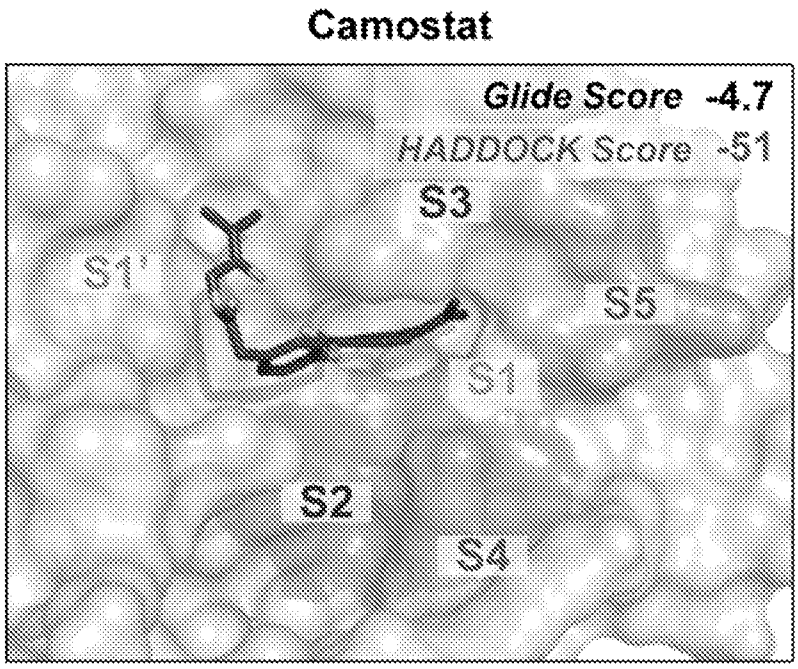


FIG. 3C

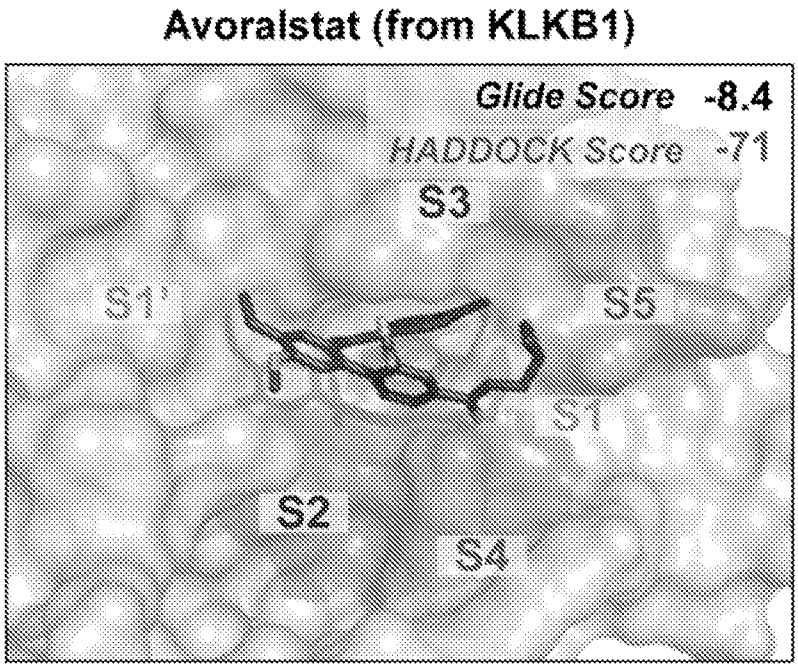

Avoralstat (from KLKB1)
Glide Score -8.4
HADDOCK Score -71
S3
S1'
S5
S1
S2
S4

FIG. 3D

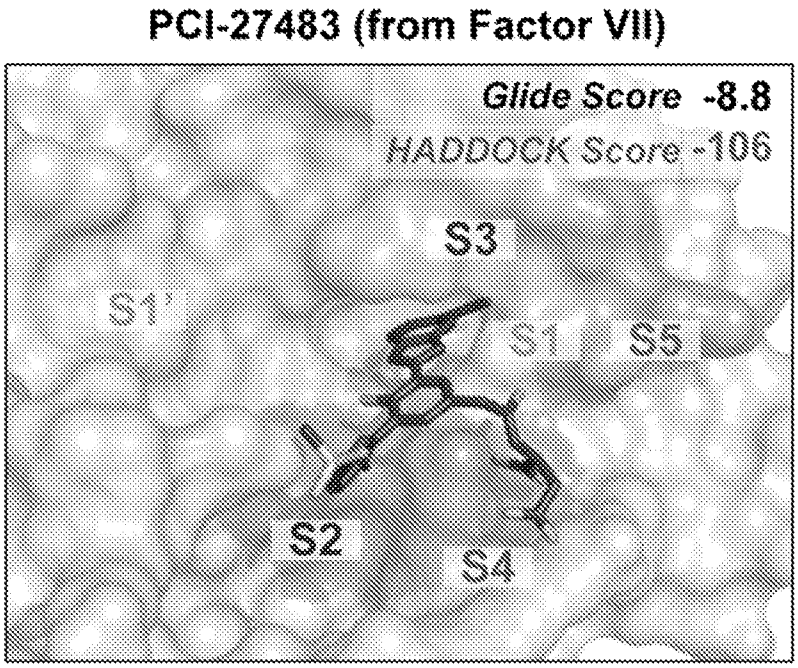

PCI-27483 (from Factor VII)
Glide Score -8.8
HADDOCK Score -106
S3
S1'
S1
S5
S2
S4

FIG. 3E

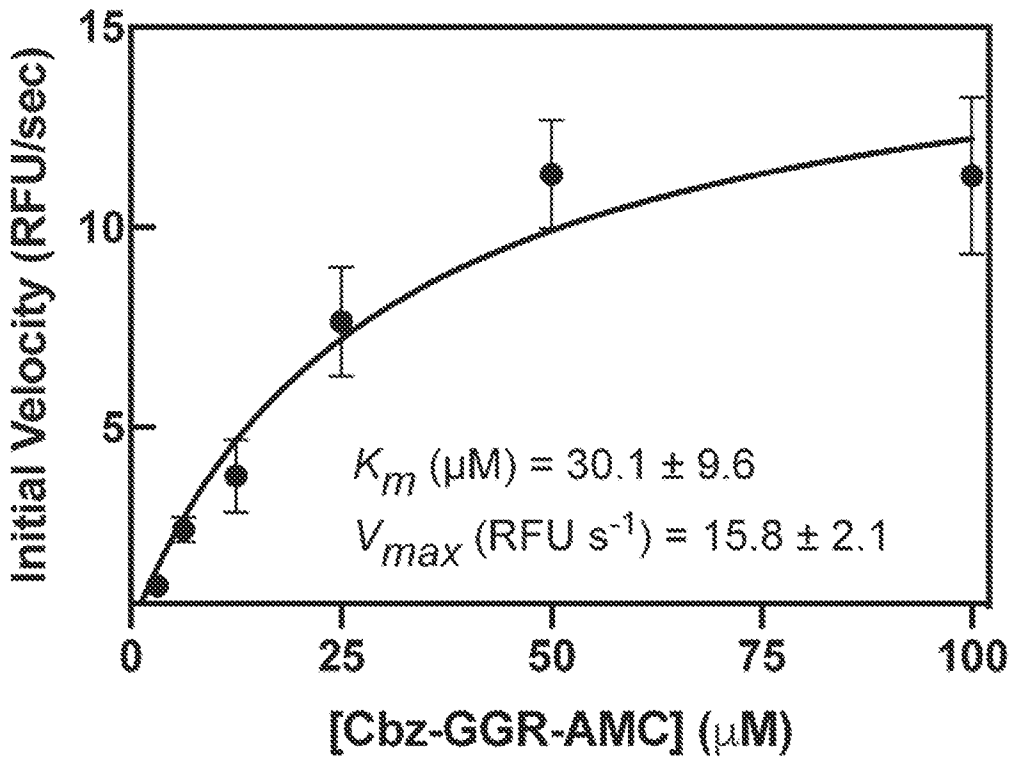

Initial Velocity (RFU/sec)
15
10
5
0
25
50
75
100
[Cbz-GGR-AMC] (μM)
K_m (μM) = 30.1 ± 9.6
V_{max} (RFU s^{-1}) = 15.8 ± 2.1

FIG. 4A

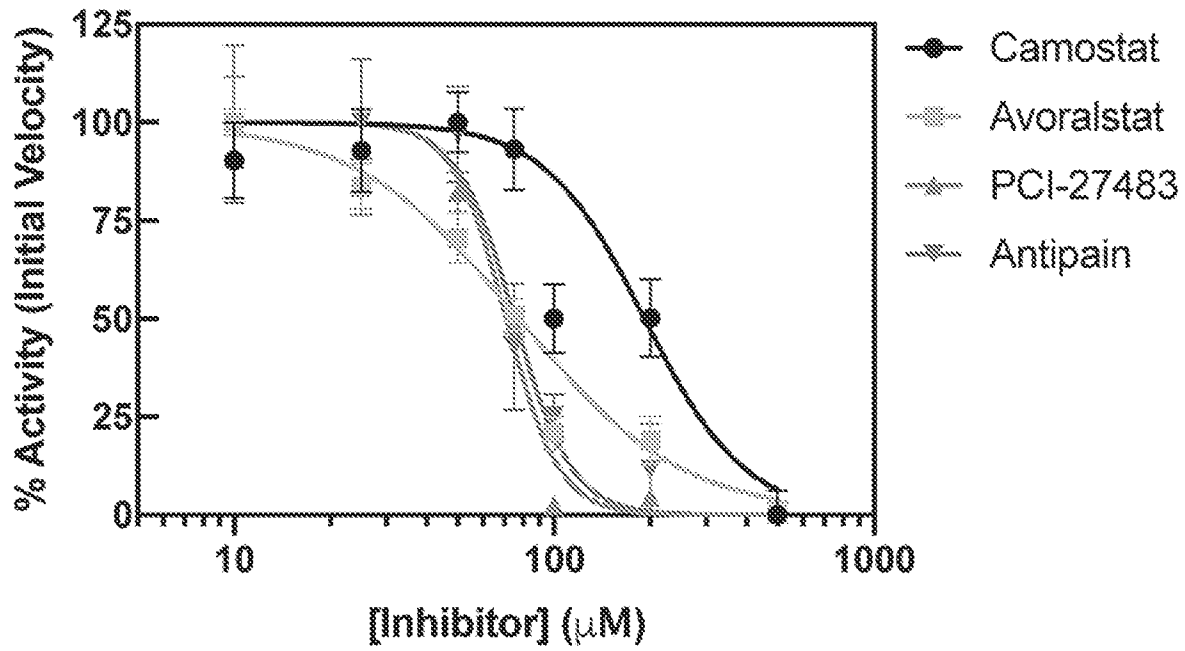

% Activity (Initial Velocity)
125
100
75
50
25
0
10
100
1000
[Inhibitor] (μM)
Camostat
Avoralstat
PCI-27483
Antipain

FIG. 4B

Inhibitor	IC_{50} (μM)
Camostat	159 ± 34
Avoralstat	69.4 ± 5.8
PCI-27483	70.8 ± 3.1
Antipain	75.2 ± 6.7

FIG. 4C

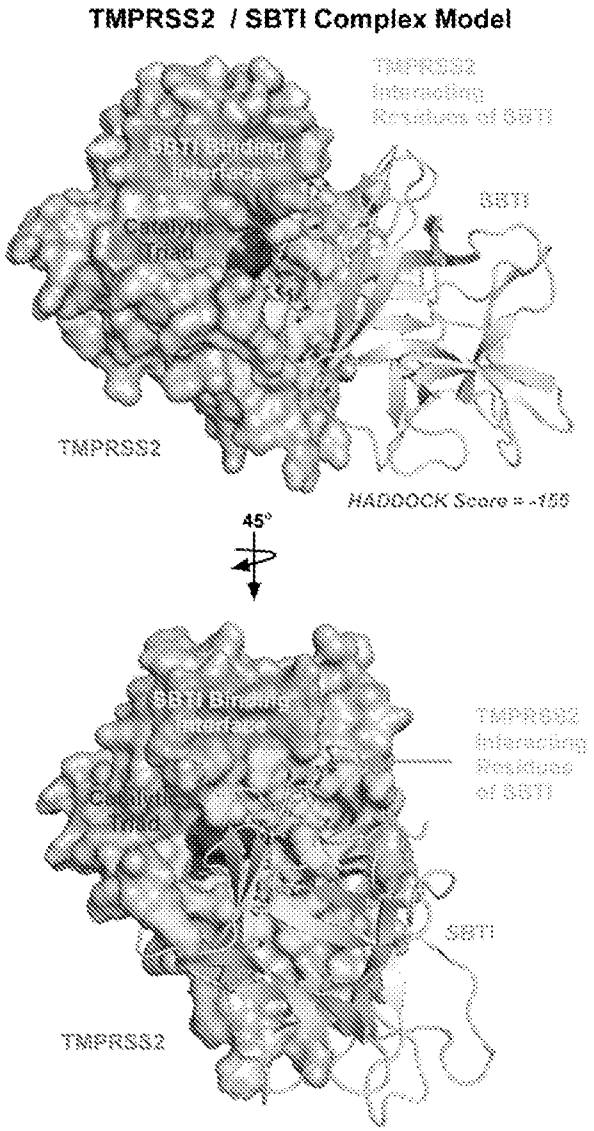

TMPRSS2 / SBTI Complex Model
TMPRSS2 Interacting Residues of SBTI
SBTI binding
SBTI
TMPRSS2
HADDOCK Score = -159
45°
SBTI binding
TMPRSS2 Interacting Residues of SBTI
SBTI
TMPRSS2

FIG. 5A

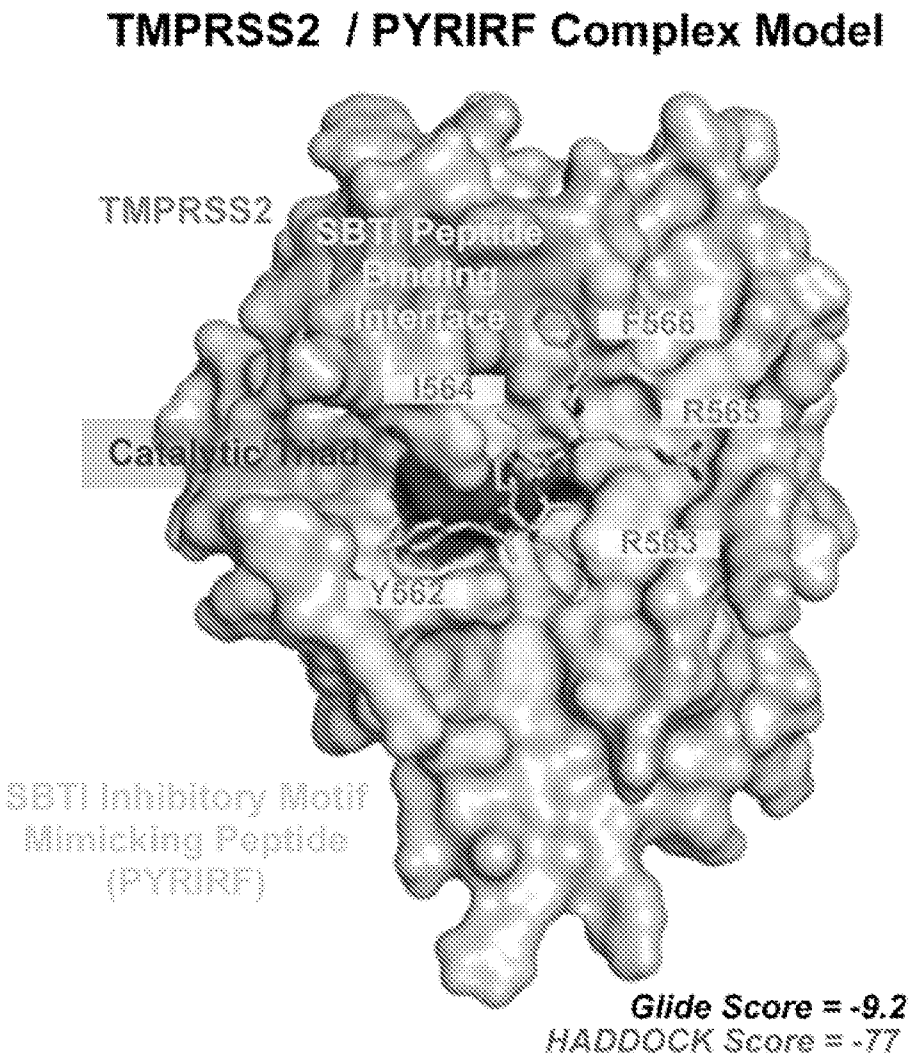

TMPRSS2 / PYRIRF Complex Model
TMPRSS2
SBTI Peptide
Binding
Interface
F566
I564
R565
Catalytic Triad
R563
Y662
SBTI Inhibitory Motif
Mimicking Peptide
(PYRIRF)
Glide Score = -9.2
HADDOCK Score = -77

FIG. 5B

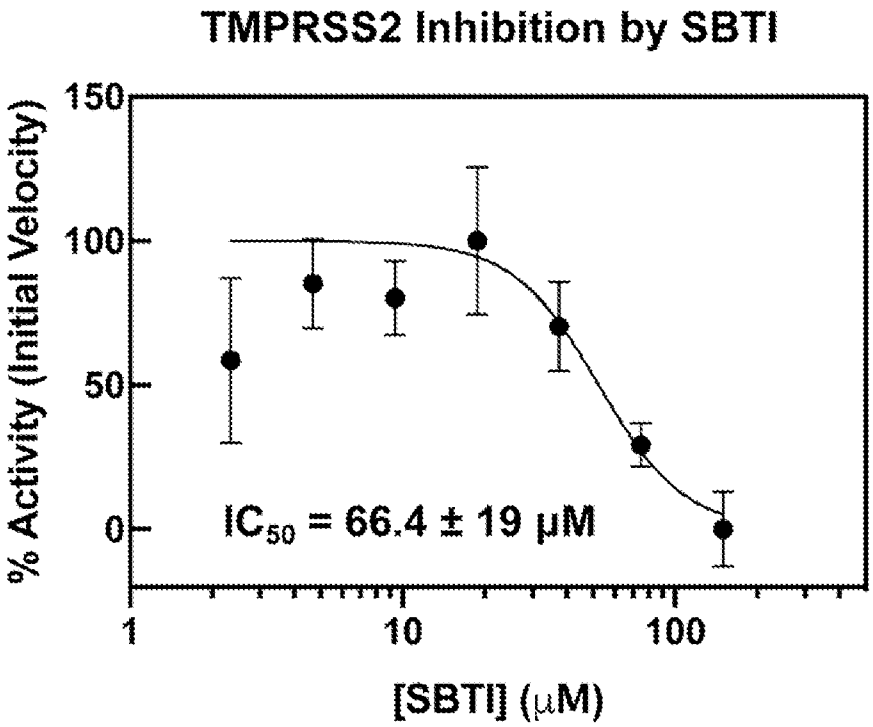

TMPRSS2 Inhibition by SBTI
% Activity (Initial Velocity)
150
100
50
0
IC50 = 66.4 ± 19 µM
1
10
100
[SBTI] (µM)

FIG. 5C

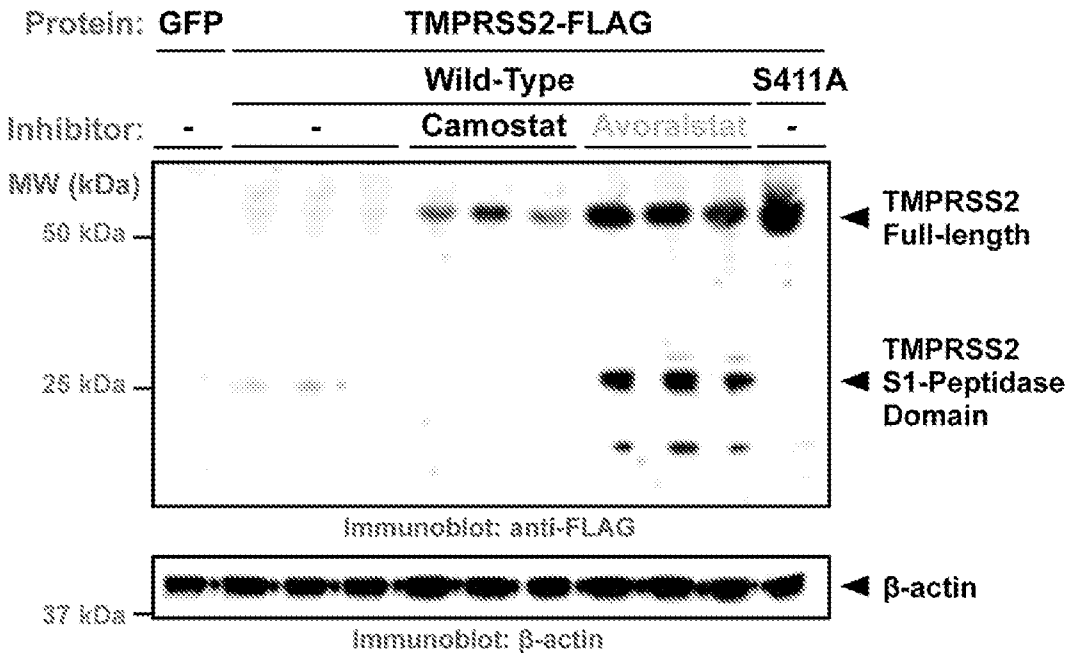

Protein:
GFP
TMPRSS2-FLAG
Wild-Type
S411A
Inhibitor:
-
-
Camostat
Avoralstat
-
MW (kDa)
50 kDa
25 kDa
TMPRSS2
Full-length
TMPRSS2
S1-Peptidase
Domain
Immunoblot: anti-FLAG
37 kDa
β-actin
Immunoblot: β-actin

FIG. 6A

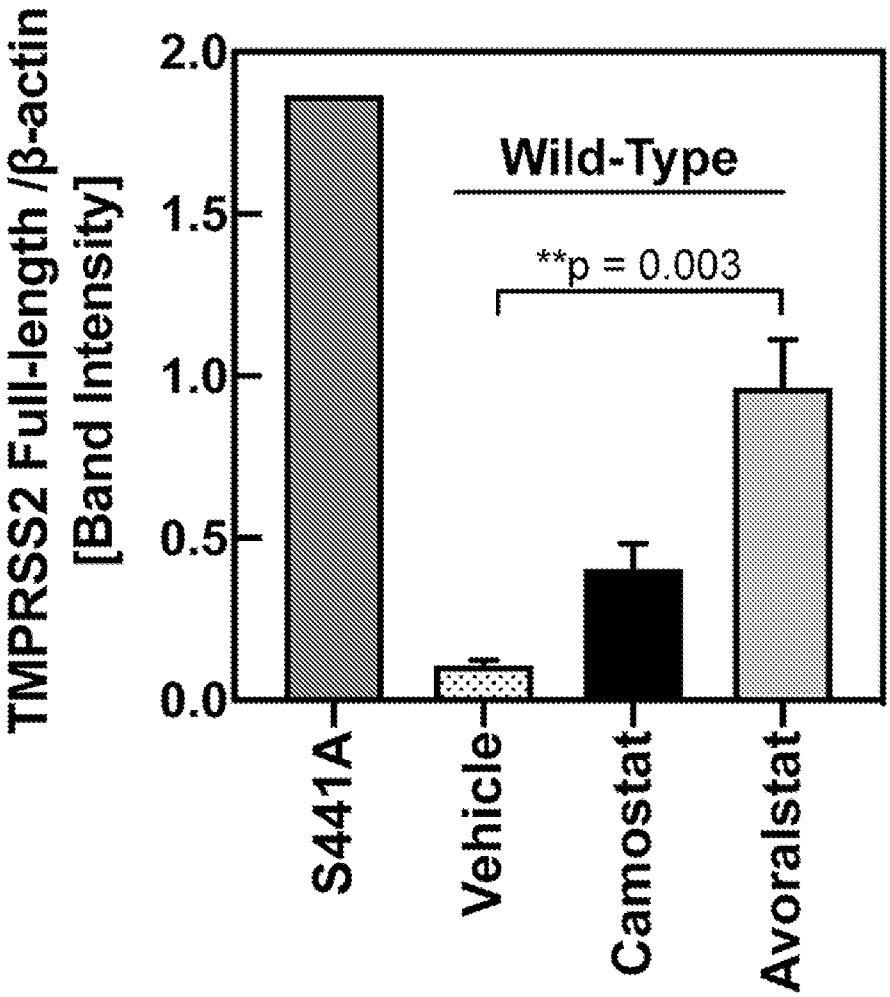

TMPRSS2 Full-length /β-actin
[Band Intensity]
2.0
1.5
1.0
0.5
0.0
Wild-Type
**p = 0.003
S441A
Vehicle
Camostat
Avoralstat

FIG. 6B

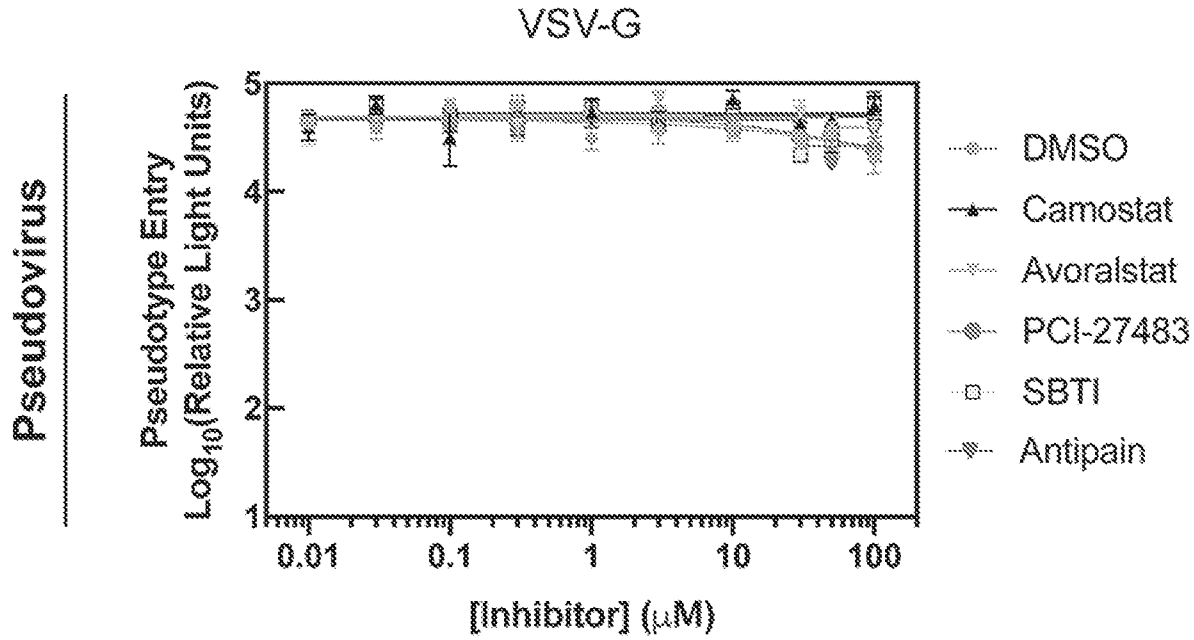

VSV-G
Pseudovirus
Pseudotype Entry
Log10(Relative Light Units)
5
4
3
2
1
0.01
0.1
1
10
100
[Inhibitor] (μM)
DMSO
Camostat
Avoralstat
PCI-27483
SBTI
Antipain

FIG. 7A

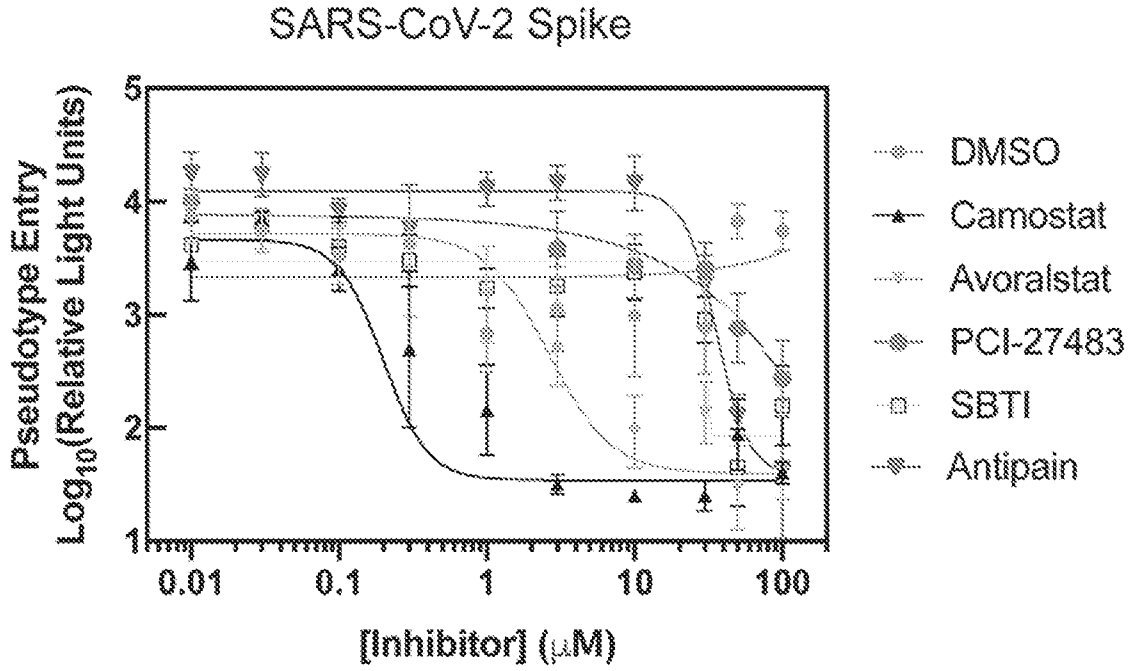

SARS-CoV-2 Spike
Pseudotype Entry
Log10(Relative Light Units)
5
4
3
2
1
0.01
0.1
1
10
100
[Inhibitor] (μM)
DMSO
Camostat
Avoralstat
PCI-27483
SBTI
Antipain

FIG. 7B

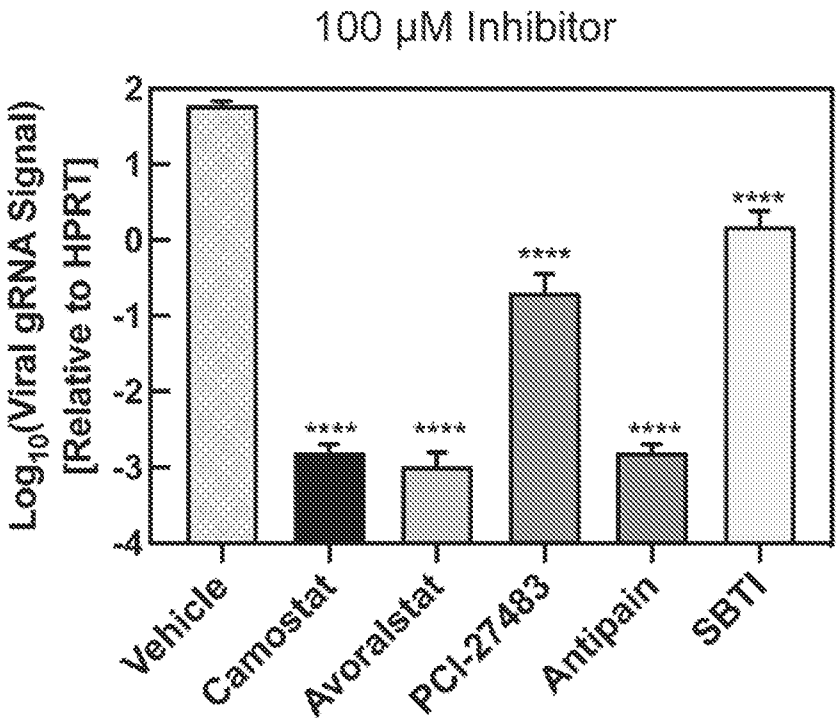

100 µM Inhibitor
Log10(Viral gRNA Signal)
[Relative to HPRT]
2
1
0
-1
-2
-3
-4
****
****
****
****
****
Vehicle
Camostat
Avoralstat
PCI-27483
Antipain
SBTI

FIG. 7C

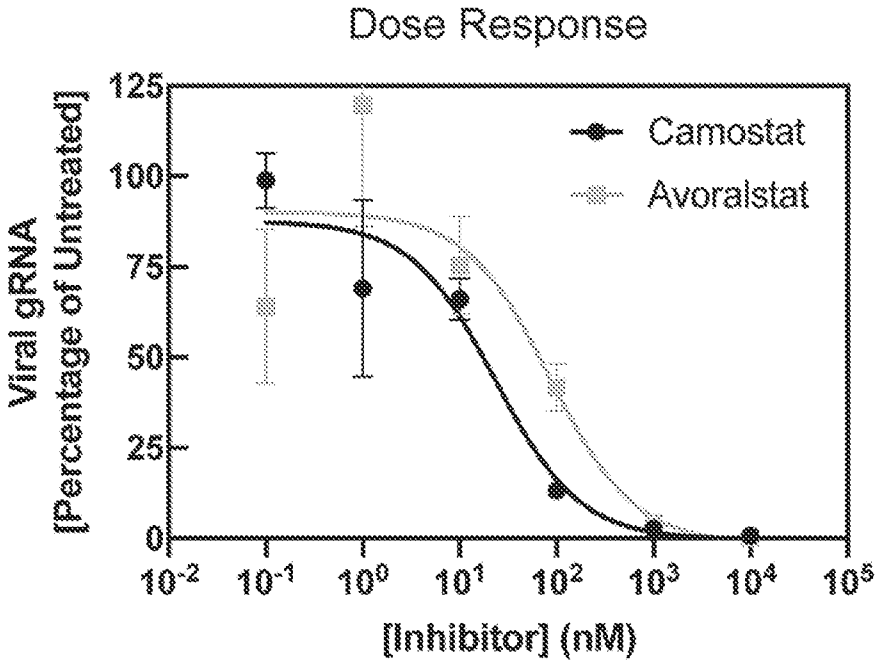

Dose Response
Viral gRNA
[Percentage of Untreated]
125
100
75
50
25
0
Camostat
Avoralstat
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
[Inhibitor] (nM)

FIG. 7D

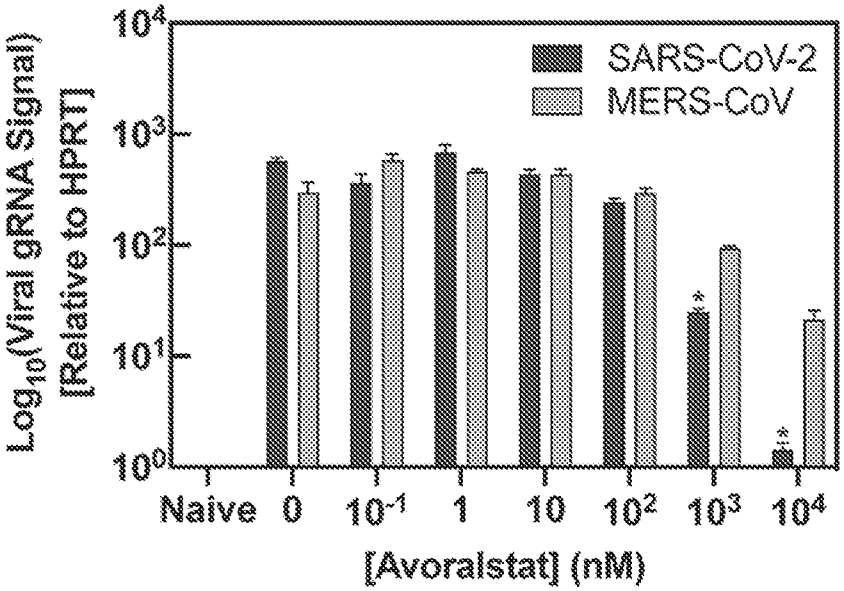

SARS-CoV-2
MERS-CoV
Log10(Viral gRNA Signal)
[Relative to HPRT]
10^4
10^3
10^2
10^1
10^0
Naive
0
10^-1
1
10
10^2
10^3
10^4
[Avoralstat] (nM)

FIG. 7E

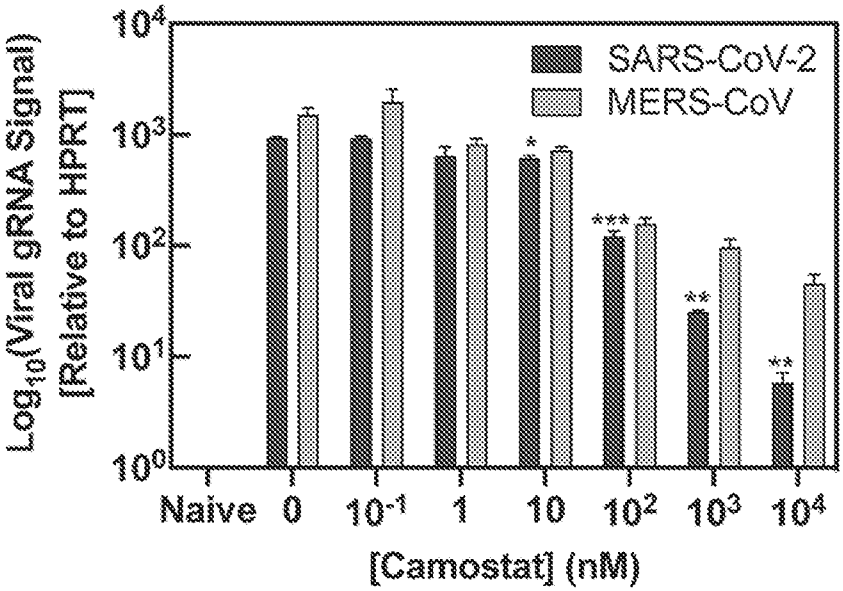

Log10(Viral gRNA Signal)
[Relative to HPRT]
10^4
10^3
10^2
10^1
10^0
SARS-CoV-2
MERS-CoV
*
***
**
**
Naive
0
10^-1
1
10
10^2
10^3
10^4
[Camostat] (nM)

FIG. 7F

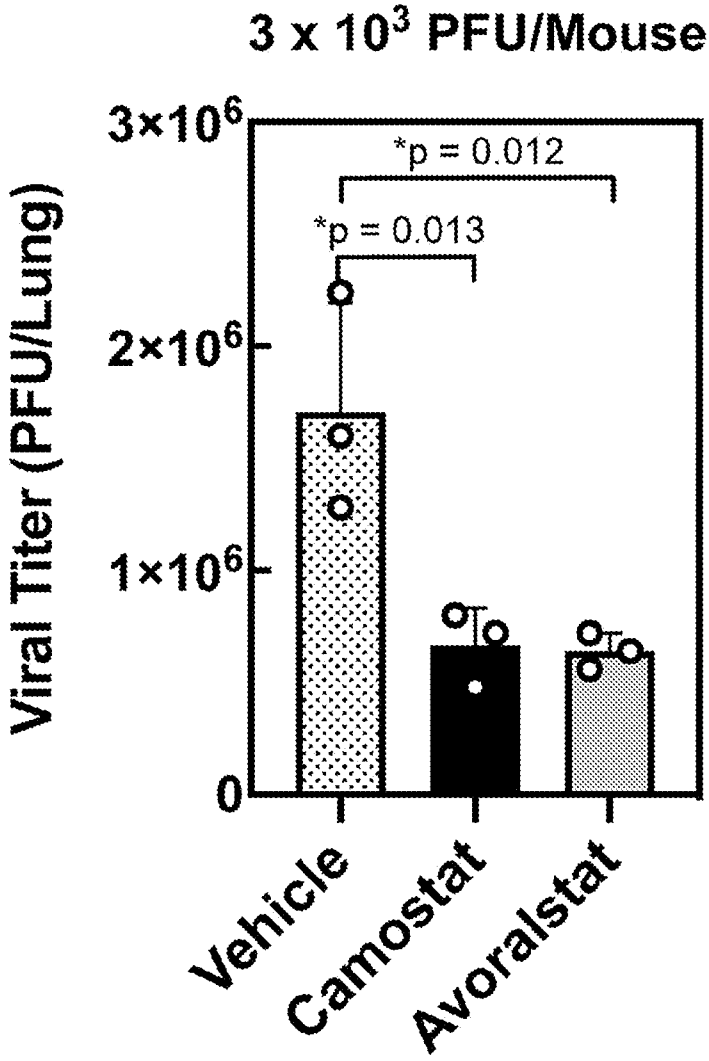

3 x 10³ PFU/Mouse
Viral Titer (PFU/Lung)
3×10⁶
2×10⁶
1×10⁶
0
*p = 0.012
*p = 0.013
Vehicle
Camostat
Avoralstat

FIG. 8A

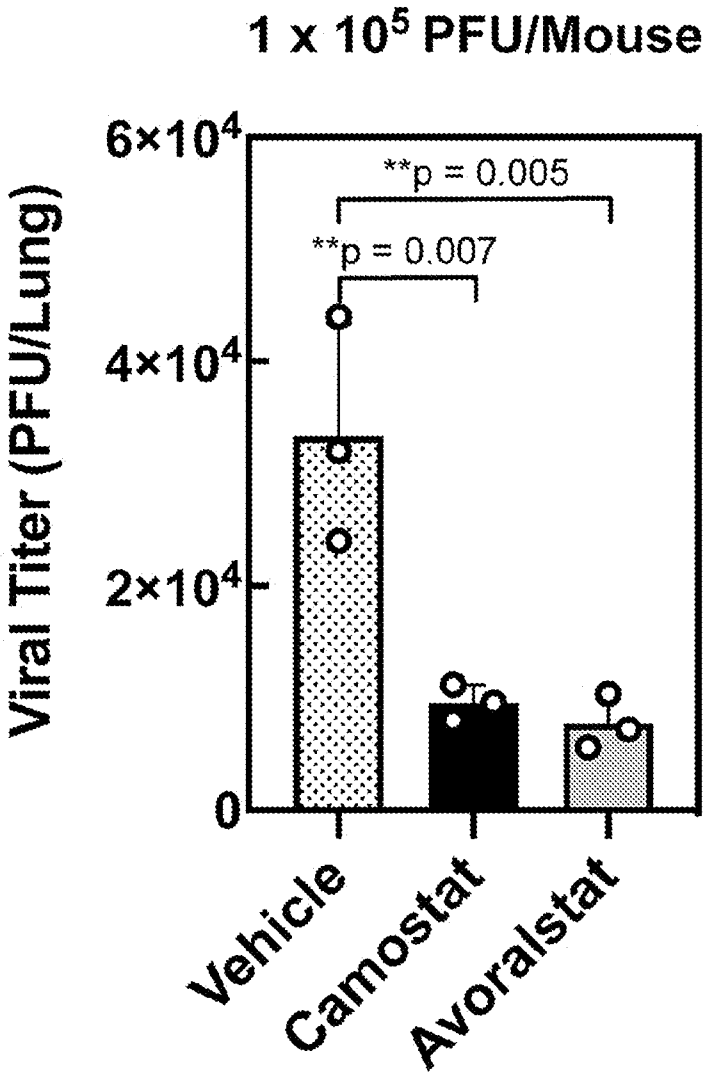

1 x 10^5 PFU/Mouse
Viral Titer (PFU/Lung)
6×10^4
4×10^4
2×10^4
0
**p = 0.005
**p = 0.007
Vehicle
Camostat
Avoralstat

FIG. 8B

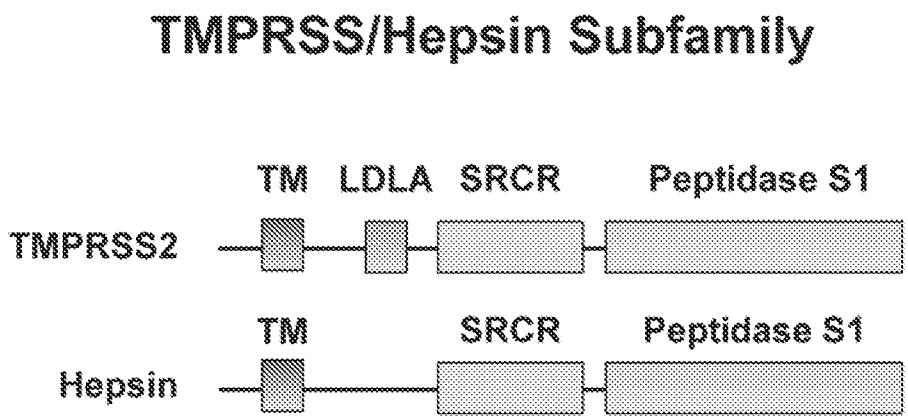

TMPRSS/Hepsin Subfamily
TM
LDLA
SRCR
Peptidase S1
TMPRSS2
TM
SRCR
Peptidase S1
Hepsin

FIG. 9A

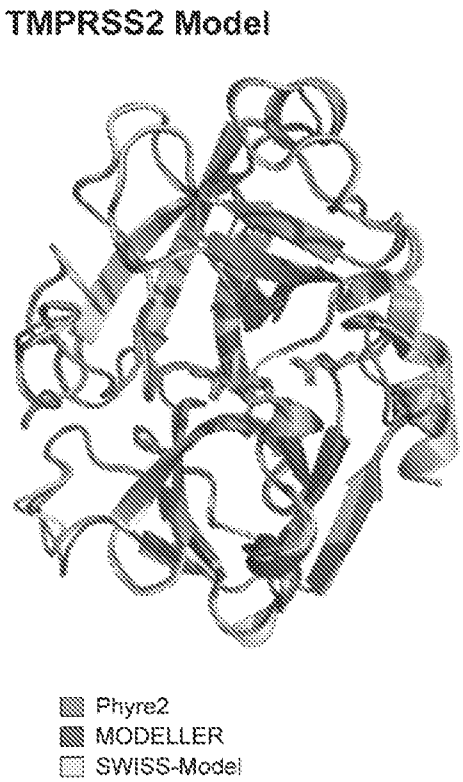

TMPRSS2 Model
Phyre2
MODELLER
SWISS-Model

FIG. 9B

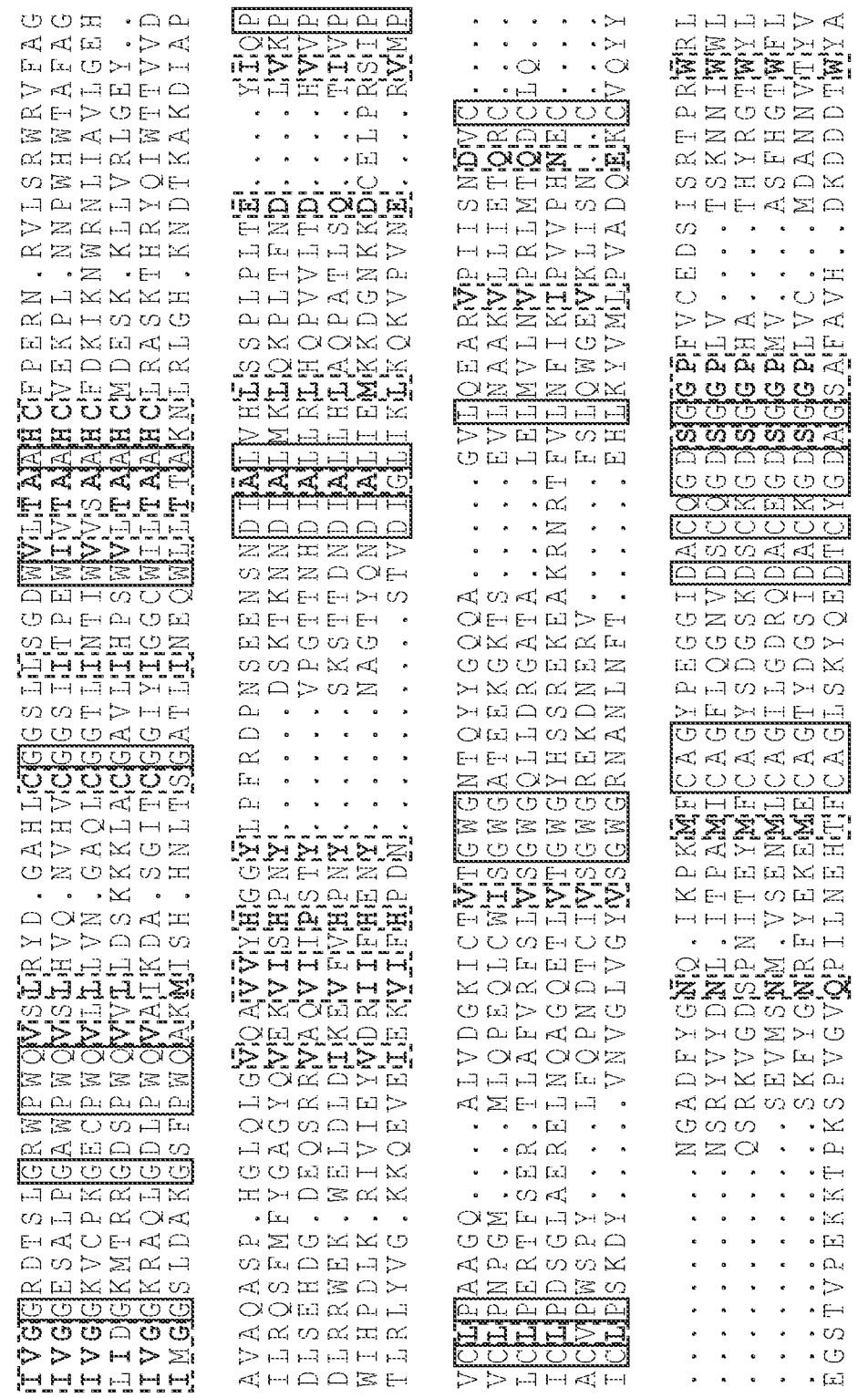

Hepsin [1O5E|A]
TMPRSS2
Factor VII [1CVW|A]
Factor XIV [1AUT|A]
CFAI [2XRC|B]
Haptoglobin [4F4O|C]

FIG. 10A

Hepsin

Seq. ID = 41%
Cα RMSD = 0.2 Å

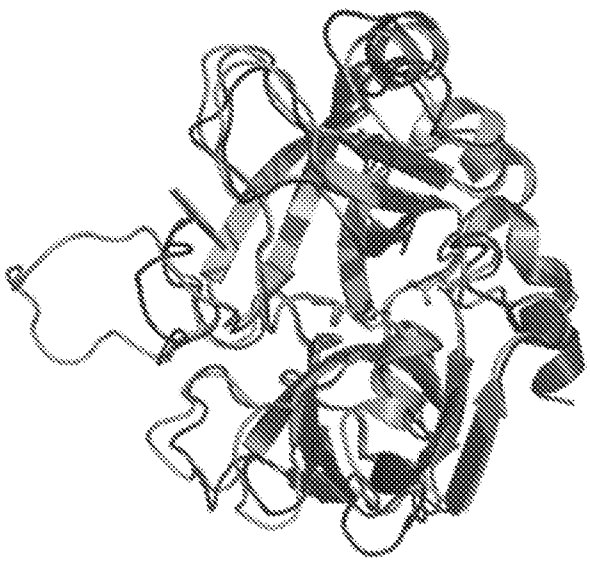
FIG. 10B

Factor VII

Seq. ID = 37%
Cα RMSD = 0.6 Å

FIG. 10C

Factor XIV



Seq. ID = 33%
Cα RMSD = 0.8 Å

FIG. 10D

CFAI

Seq. ID = 33%
Cα RMSD = 1.4 Å

FIG. 10E

Haptoglobin

Seq. ID = 24%
Cα RMSD = 1.1 Å

FIG. 10F

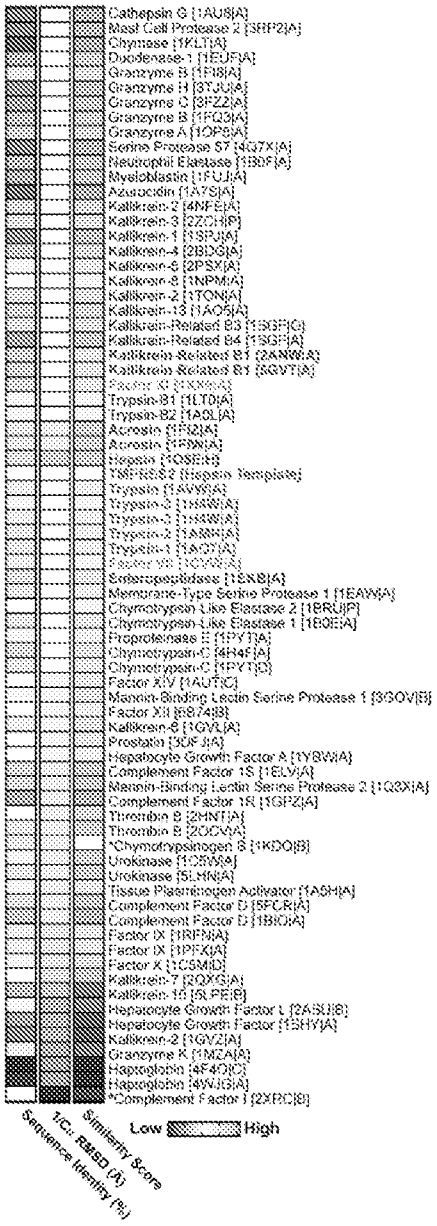

Cathepsin G [1AU8|A]
Mast Cell Protease 2 [3RP2|A]
Chymase [1KLT|A]
Duodenase-1 [1EUF|A]
Granzyme B [1FI8|A]
Granzyme H [3TJU|A]
Granzyme C [3FZ2|A]
Granzyme B [1FQ3|A]
Granzyme A [1OP5|A]
Serine Protease S7 [4G7X|A]
Neutrophil Elastase [1B0F|A]
Myeloblastin [1FUJ|A]
Azurocidin [1AZS|A]
Kallikrein-2 [4NFE|A]
Kallikrein-3 [2ZCH|P]
Kallikrein-1 [1SPJ|A]
Kallikrein-4 [2BDG|A]
Kallikrein-6 [2PSX|A]
Kallikrein-8 [1NPM|A]
Kallikrein-2 [1TON|A]
Kallikrein-13 [1AO5|A]
Kallikrein-Related B3 [1SGF|C]
Kallikrein-Related B4 [1SGF|A]
Kallikrein-Related B1 [2ANW|A]
Kallikrein-Related B1 [5GVT|A]
Trypsin-B1 [1LT0|A]
Trypsin-B2 [1A0L|A]
Acrosin [1PS2|A]
Acrosin [1PSW|A]
Hepsin [1OSE|B]
TMPRSS2 (Hepsin Template)
Trypsin [1AVW|A]
Trypsin-2 [1H4W|A]
Trypsin-3 [1H4W|A]
Trypsin-2 [1A8R|A]
Trypsin-1 [1AQT|A]
Enteropeptidase [1EKB|A]
Membrane-Type Serine Protease 1 [1EAW|A]
Chymotrypsin Like Elastase 2 [1BRU|P]
Chymotrypsin-Like Elastase 1 [1B0E|A]
Proprotelnase E [1PYT|A]
Chymotrypsin-C [4H4F|A]
Chymotrypsin-C [1PYT|D]
Factor XIV [1AUT|C]
Mannan-Binding Lectin Serine Protease 1 [3GOV|B]
Factor XII [6B74|B]
Kallikrein-6 [1GVL|A]
Prostatin [3DFJ|A]
Hepatocyte Growth Factor A [1YBW|A]
Complement Factor 1S [1ELV|A]
Mannan-Binding Lectin Serine Protease 2 [1Q3X|A]
Complement Factor 1R [1GPZ|A]
Thrombin B [2HNT|A]
Thrombin B [2OCV|A]
*Chymotrypsinogen B [1KDO|B]
Urokinase [1O8W|A]
Urokinase [5LHN|A]
Tissue Plasminogen Activator [1A5H|A]
Complement Factor D [5FCR|A]
Complement Factor D [1BIO|A]
Factor IX [1RFN|A]
Factor IX [1PFX|A]
Factor X [1C5M|D]
Kallikrein-7 [2QXG|A]
Kallikrein-10 [5LPE|B]
Hepatocyte Growth Factor L [2A5U|B]
Hepatocyte Growth Factor [1SHY|A]
Kallikrein-2 [1GVZ|A]
Granzyme K [1MZA|A]
Haptoglobin [4F4O|C]
Haptoglobin [4WJG|A]
*Complement Factor I [2XRC|B]
Low
High
Similarity Score
1/C$_\alpha$ RMSD (Å)
Sequence Identity (%)

FIG. 10G

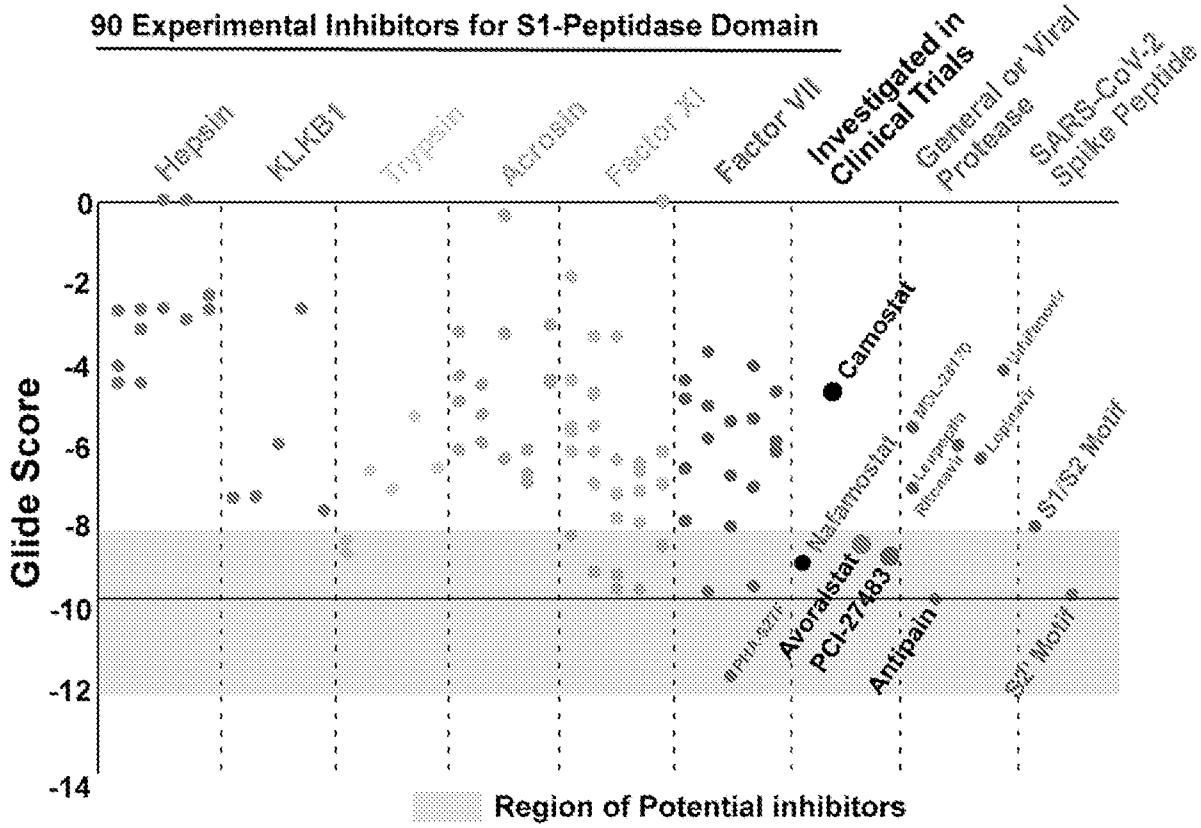

90 Experimental Inhibitors for S1-Peptidase Domain
Hepsin
KLKB1
Trypsin
Acrosin
Factor XI
Factor VII
Investigated in Clinical Trials
General or Viral Protease
SARS-CoV-2 Spike Peptide
Glide Score
0
-2
-4
-6
-8
-10
-12
-14
Camostat
Nafamostat
Avoralstat
PCI-27483
Antipain
S1/S2 Motif
S2' Motif
Region of Potential inhibitors

FIG. 11A

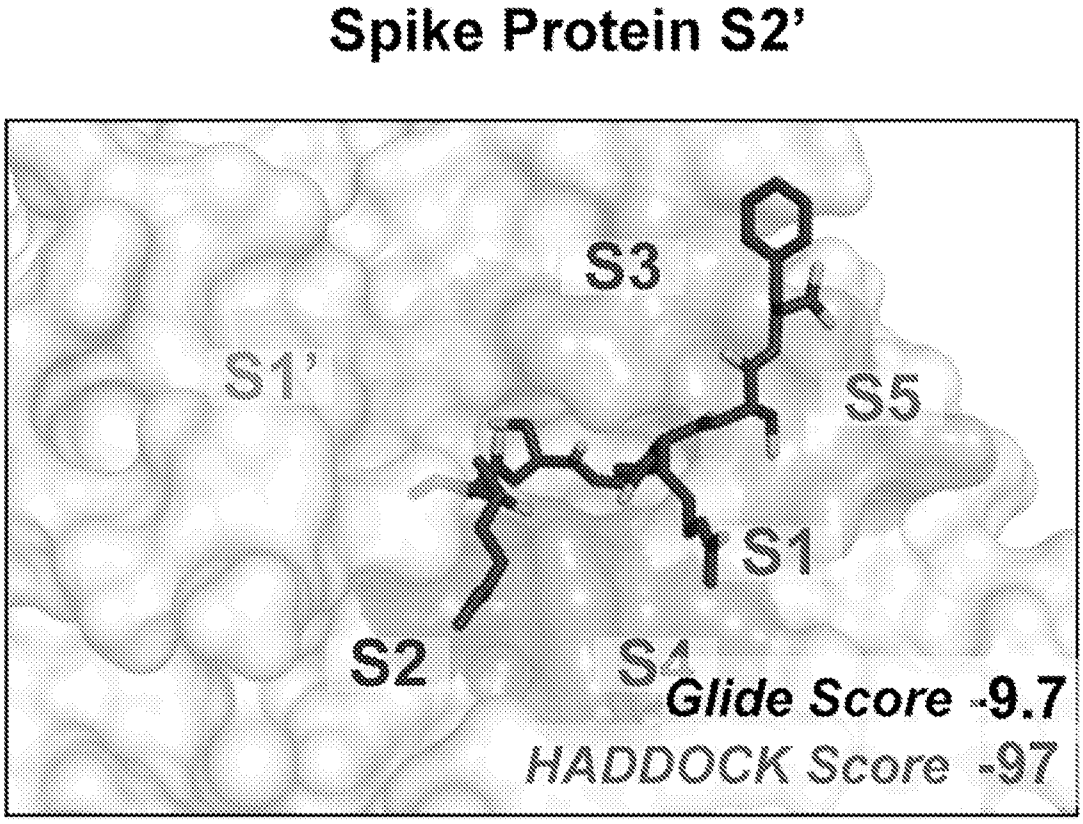

Spike Protein S2'
S3
S1'
S5
S1
S2
S4
Glide Score -9.7
HADDOCK Score -97

FIG. 11B

Nafamostat

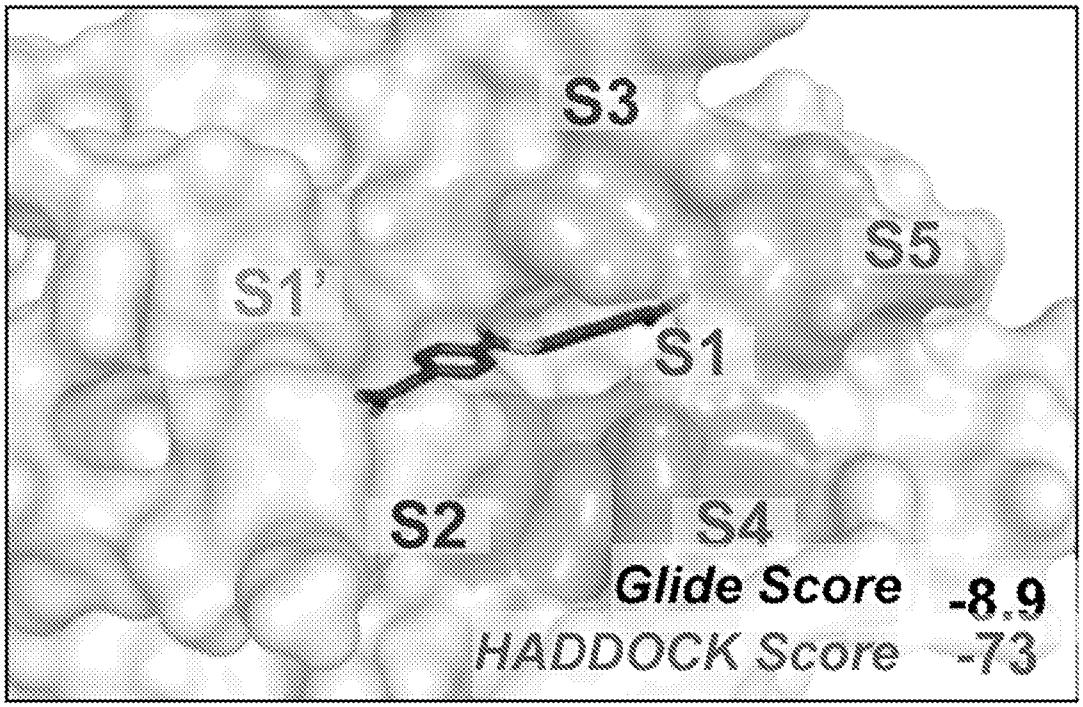


FIG. 11C

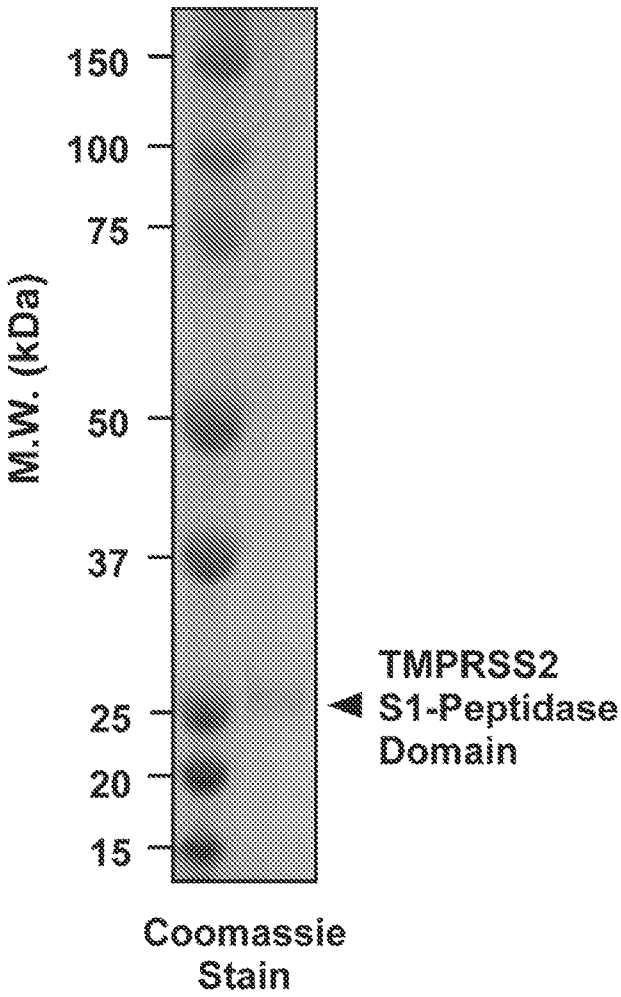

150
100
75
50
37
25
20
15
M.W. (kDa)
TMPRSS2
S1-Peptidase
Domain
Coomassie
Stain

FIG. 12A

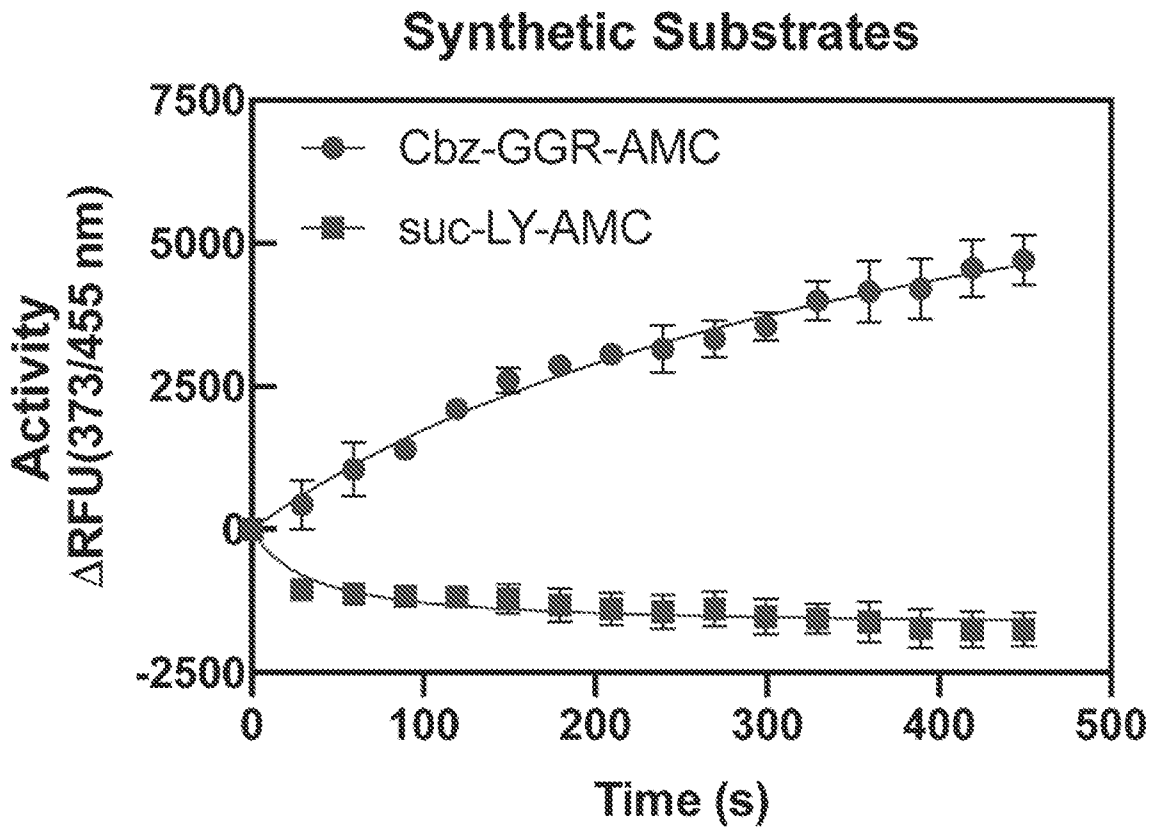

Synthetic Substrates
Cbz-GGR-AMC
suc-LY-AMC
Activity
ΔRFU(373/455 nm)
7500
5000
2500
0
-2500
0
100
200
300
400
500
Time (s)

FIG. 12B

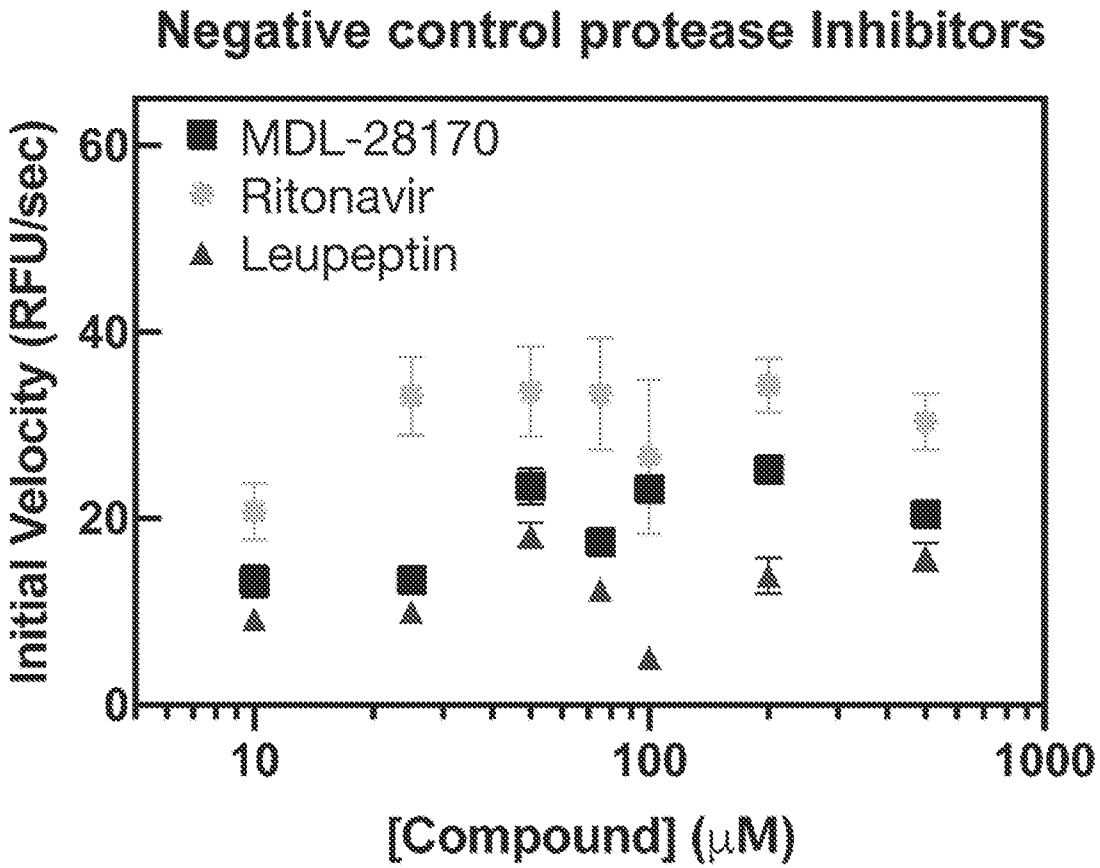

Negative control protease Inhibitors
MDL-28170
Ritonavir
Leupeptin
Initial Velocity (RFU/sec)
60
40
20
0
10
100
1000
[Compound] (μM)

FIG. 12C

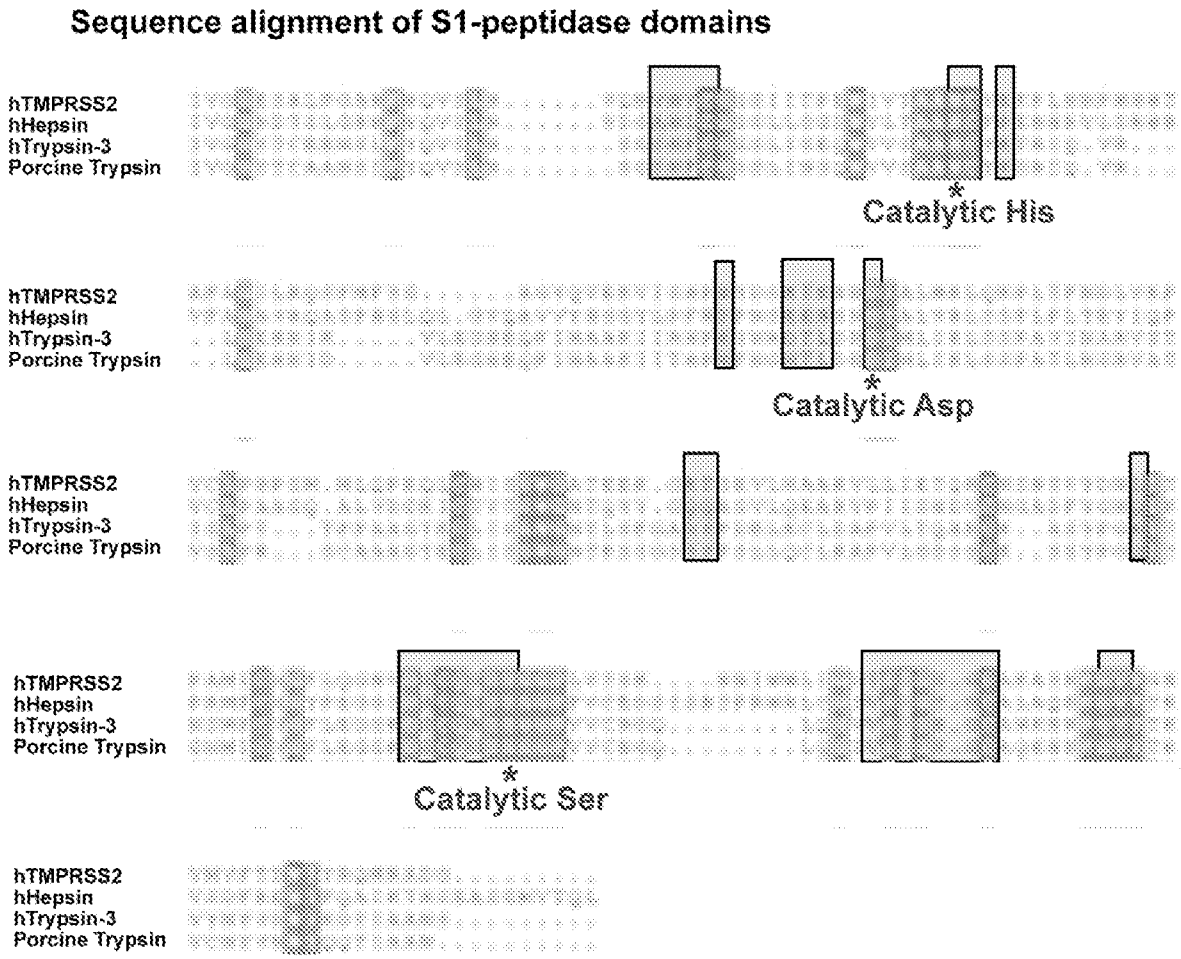

Sequence alignment of S1-peptidase domains
hTMPRSS2
hHepsin
hTrypsin-3
Porcine Trypsin
Catalytic His
Catalytic Asp
Catalytic Ser

FIG. 13A

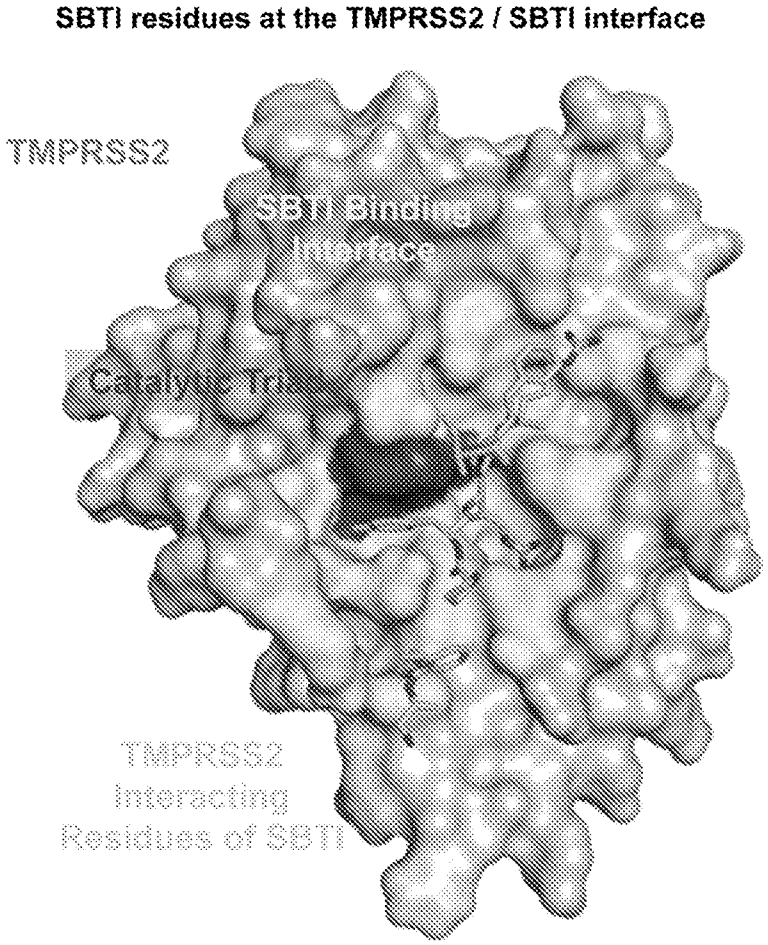
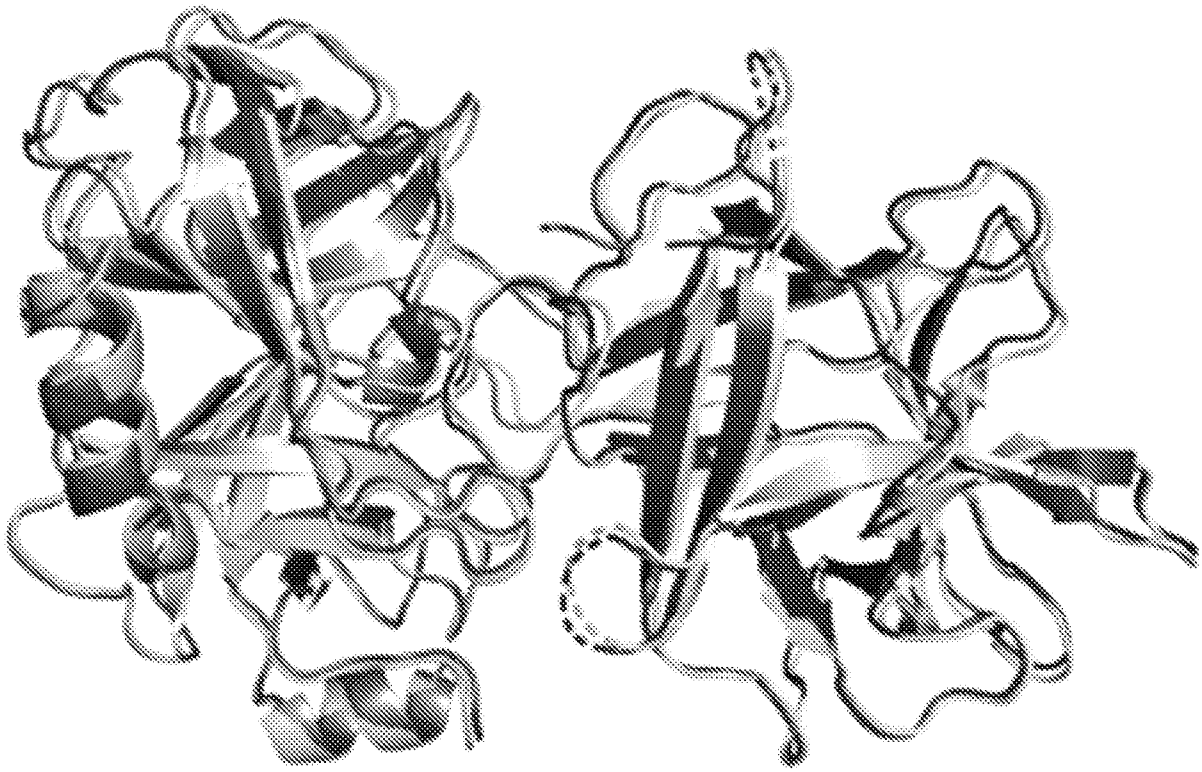

SBTI residues at the TMPRSS2 / SBTI interface
TMPRSS2
SBTI Binding Interface
Catalytic Tri
TMPRSS2 Interacting Residues of SBTI

FIG. 13B

Porcine Trypsin / SBTI complex (HADDOCK score = -172)



Porcine Trypsin
(HADDOCK / 1AVW)

SBTI
(HADDOCK / 1AVW)

FIG. 13C