In silico docking screen identifies airway host protease targets for human SERPINs

Joaquin Rodriguez Galvan[1], Maren de Vries[1], Ashley Fisher[1], Rachel A Prescott[1], Keaton M Crosse[1], Ralf Duerr[1] and Meike Dittmann[1]*

[1] New York University Grossman School of Medicine, New York, NY

* Corresponding author

Key words: influenza A virus, SARS-CoV-2, rhinovirus, adenovirus, parainfluenza virus

Abstract

Proteases play key roles in viral replication cycles. They can provide cleavage maturation of viral glycoproteins, processing of viral polyproteins, or disassembly of viral capsids. Thus, proteases constitute ideal targets for antiviral intervention – pharmaceutically, by small molecule inhibitors, or naturally, by host immune responses. Indeed, we and others have shown that individual members of the Serine protease inhibitor (SERPIN) family have specific antiviral function by blocking proteolytic steps inherent to viral replication cycles. Whether additional members of the large SERPIN family possess antiviral activity and whether SERPINs function as part of the antiviral cell-intrinsic immune response, is currently unknown. Here, we found that specific SERPINs are produced upon infection with clinically relevant respiratory viruses *in vitro* and *in vivo*, and in concert with classical interferon-stimulated genes. We next developed a structure-based *in silico* screen to uncover non-canonical SERPIN-protease pairs. We identified several SERPINs with potential antiviral function, including: SERPINE1 targeting cathepsin L, required for SARS-CoV-2 entry; SERPINB8 targeting furin, required for glycoprotein maturation cleavage of numerous viruses; and SERPINB2 targeting adenovirus protease, which suggests the first direct-acting antiviral SERPIN. Our study demonstrates how proteolysis is modulated for antiviral defense and how this process could inform antiviral targets against clinically relevant respiratory pathogens.

Introduction

Respiratory viruses constitute a serious risk to human health. For example, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and influenza A virus (IAV) have originated major pandemics leading to severe respiratory impairment and death for millions of people and have caused considerable collateral damage to the economy and society [1]. Respiratory viruses encompass viruses from diverse families that show inherent tropism for the respiratory tract. Despite this commonality, and due to their considerable genetic and structural diversity, they can cause different clinical outcomes [2]. In particular, infections with IAV and SARS-CoV-2 are associated with a more severe prognosis in adults, whereas infections with other viruses, such as adenovirus or parainfluenza virus, tend to be milder in this age group [2,3]. These differences in pathogenesis may be attributed to a virus' ability to evade or downmodulate host immune responses or to viral tropism to the upper or lower respiratory tract.

Although respiratory viruses are highly heterogeneous in their replication strategies and lifecycles within a host cell, they all rely on a critical group of enzymes: proteases [4]. The origin of these proteases may be host or viral, and the roles they play in viral replication differ for each virus [4,5]. For example, most enveloped viruses rely on cleavage of their fusogenic glycoprotein for entry via fusion with host cell membranes, as is the case for parainfluenza viruses, influenza viruses, and coronaviruses, among others. The proteases executing viral glycoprotein cleavage are typically host-encoded, and are present at the plasma membrane (either intra- or extracellularly) or within endocytic or secretory pathways, aligning with the compartment in which the cleavage

event occurs [3,6-9]. Another protease function is for viral polyprotein processing, which is the cleavage of viral polyprotein precursors into individual functional subunits upon translation [9,10].This is the case for 3C-like and papain-like proteases ($3CL^{pro}$ and PL^{pro}), which are viral proteases encoded within the SARS-CoV-2 genome [9]. Other proteases play roles in assembly/disassembly of viral particles inside the cell, as is the case for the adenovirus-encoded protease (Ad^{pro}), which allows for assembly of the protein capsid by cleaving capsid subunits, among other functions [11,12].

The multifaceted and integral role proteases play within the life cycles of many viruses has led to their targeting for antiviral therapeutics. In fact, multiple antiviral FDA-approved antiviral drugs are in fact small molecule protease inhibitors. These drugs target the viral proteases of HIV-1, HCV and most recently SARS-CoV-2. The targeted viral proteases all cleave large poly- chains into functional subunits, which are an essential steps of these viral life cycles [13-15]. The clinical success of these antivirals further illustrates the importance of proteases within viral lifecycles and highlights the potential of protease inhibition for successful antiviral intervention.

Similar to pharmacological inhibition, humans encode genes to regulate and inhibit proteolytic activity [16]. In this study, we focus on the largest family of human-encoded protease inhibitors: serine protease inhibitors (SERPINs) [16]. SERPINs are structurally and mechanistically conserved, consisting of a globular domain with critical central beta sheets and a reactive center loop (RCL) that protrudes from the main SERPIN body [16]. The SERPIN mechanism of action is well-established at the molecular level. As a

SERPIN's RCL inserts into the protease active site, mimicking a substrate, the protease catalytic triad residues initiate the proteolysis mechanism, resulting in a covalent bond between the protease main active residue (nucleophile) and the P1 residue of the SERPIN RCL [16]. Subsequently, the SERPIN loop triggers its insertion into the central beta sheets, thereby trapping the protease in a covalent, suicide-like inhibitory complex [16]. Specificity for a given SERPIN-mediated protease inhibition stems mainly from appropriate fitting of a SERPIN's RCL into the protease active site. Other factors determining SERPIN-protease specificity are actions that stabilize the SERPIN-protease complex, which is intrinsically unstable. Stabilization is achieved through the action of co-factors, or through characteristics inherent to the SERPIN loop structure itself, both of which can prevent SERPIN collapse, aggregation, or cleavage [16-18].

Given the protease requirement for most respiratory viruses to successfully complete their replication cycles, it seems intuitive that SERPINs may elicit antiviral activity. Indeed, our group was the first to describe that SERPINE1 (plasminogen activator inhibitor 1, PAI-1) is antiviral against influenza A H1 and H3 viruses, by preventing the proteolytic maturation of those low pathogenic hemagglutinins [19]. Moreover, we found that *SERPINE1* is an interferon-stimulated gene, indicating its role in the innate immune response against viruses [19]. Since then, other groups have described the antiviral effect of SERPINA1 against SARS-CoV-2 by inhibiting TMPRSS2, a protease required for the plasma membrane-mediated route of entry of SARS-CoV-2 [20,21]. However, beyond these two studies, little is known about other SERPIN family members in the context of viral infection. Therefore, it would not be surprising to identify additional

SERPINs with different protease targets that may elicit antiviral activity against diverse viruses in the human airways. Here, we investigate the following questions: which SERPINs are expressed in the human airways, either at steady-state or upon viral infection? What are their target proteases beyond those previously identified? Lastly, based on this information, could they have antiviral potential against clinically relevant respiratory viruses?

We found that multiple SERPINs that are differentially expressed to higher levels in COVID-19 patient-derived bronchoalveolar lavage fluid (BALF) via readily available scRNA-seq data. Complimentarily, we found 10 SERPINs upregulated in a model of the human airway epithelium upon infection with a diverse panel of respiratory viruses. With this information we then established a protein-protein docking screen that enabled us to identify novel SERPIN-protease pairs based on structure. More specifically, based on the fit of the SERPIN RCL into the protease active site, we predicted interaction and potential inhibition of the protease by a given SERPIN. Using this approach, we found several predicted SERPIN-protease pairs that may be key to understanding unexplored aspects of the innate antiviral response against respiratory viruses.

Results

SERPINs are differentially expressed in airway epithelial cells of COVID-19 patients.

SERPINs have been shown to play defined roles during inflammation, cancer metastasis, angiogenesis, and neurodegenerative disease [22]. We were the first to report that *SERPINE1* expression is enhanced on the mRNA level by interferon signaling in human cell lines and primary human airway epithelial cultures [19]. While most SERPIN functions to date have been characterized at steady-state, select SERPINs are upregulated upon specific stimuli, including interferons, tumor necrosis factors, or tumor growth factors [23-25]. All these stimuli can be produced during viral infections [2,26,27].

We first asked whether, in addition to *SERPINE1*, other members of the SERPIN family are also upregulated in response to viral infections. To begin answering this question, we mined a published single cell RNA-seq (scRNA-seq) dataset of bronchoalveolar lavage fluids (BALF) from patients with mild or severe COVID-19, or uninfected control patients [28] (Figure 1A). This dataset allowed us to investigate mRNA expression levels of individual *SERPIN*s in different cell types at the site of and during an acute viral infection in humans. We first focused on epithelial cells, as they are the primary replication site for acute SARS-CoV-2 infection and the main site of action for direct-acting antiviral genes (Figure 1B). We found that *IRF1*, a bona fide interferon-stimulated gene with antiviral function [29], was highly expressed in epithelial cells from patients

with mild and severe COVID-19, serving as positive control. A number of *SERPIN*s were expressed at low levels in epithelial cells of uninfected patients, and at higher levels in patients with mild or severe COVID-19, most notably *SERPINB2*, *SERPINB5*, *SERPINB11* and *SERPINF1*. Others were already robustly present in epithelial cells of uninfected patients and further upregulated in COVID-19 patients, namely *SERPINB1*, *SERPINB3*, *SERPINB6*, and *SERPING1*. *SERPINA1*, somewhat counterintuitively, followed an inverse pattern and was an exception, as it was highly expressed in non-infected patient epithelial cells but downregulated during SARS-CoV-2 infections. Expression of *SERPINA3*, *SERPINA5*, *SERPINA9*, *SERPINA11*, *SERPINB7*, *SERPINC1*, *SERPIND1*, *SERPINE2*, *SERPINE3* and *SERPINI1* changed less than 2-fold in COVID-19 patients as compared to control patients (Figure 1B). Correlation analysis revealed that *SERPIN* mRNA expression levels correlate with those of canonical TNF-alpha and interferon-stimulated genes [29] (Supplementary Figure 1), suggesting that *SERPINs* are expressed during viral-induced inflammation. Taken together, our analyses show that a subset of *SERPIN*s are expressed high mRNA levels during SARS-CoV-2 infection in airway epithelial cells.

Next, we asked whether this upregulation of *SERPIN*s in COVID-19 patients could also be observed in other BALF cell types, and we determined log2-fold mRNA expression changes compared to uninfected patients. Monocytes, innate lymphoid cells, neutrophils, and dendritic cells, while readily detected in COVID-19 patients, were not detected in uninfected patients, and thus -fold changes could not be determined for these cell types. We found a greater number of highly-expressed *SERPINs* in epithelial

cells in comparison to immune cells such as macrophages, T cells, B cells, and NK cells (Figure 1 C). In addition, the magnitude of *SERPIN* mRNA expression was higher overall in epithelial cells from both mild and severe COVID-19 patients. Given that major role of SERPINs is to inhibit proteases, we next analyzed expression of proteases in epithelial cells. While we found only minimal changes in mRNA expression of airway proteases upon SARS-CoV-2 infection, the protease family with the most robust expression at baseline was cathepsins (Supplementary Figure 2). A notable exception was *TMPRSS4*, which was >10-fold upregulated in COVID-19 patients as compared to uninfected patients. Interestingly, TMPRSS4 has been shown to process SARS-CoV-2 Spike protein [7,30].

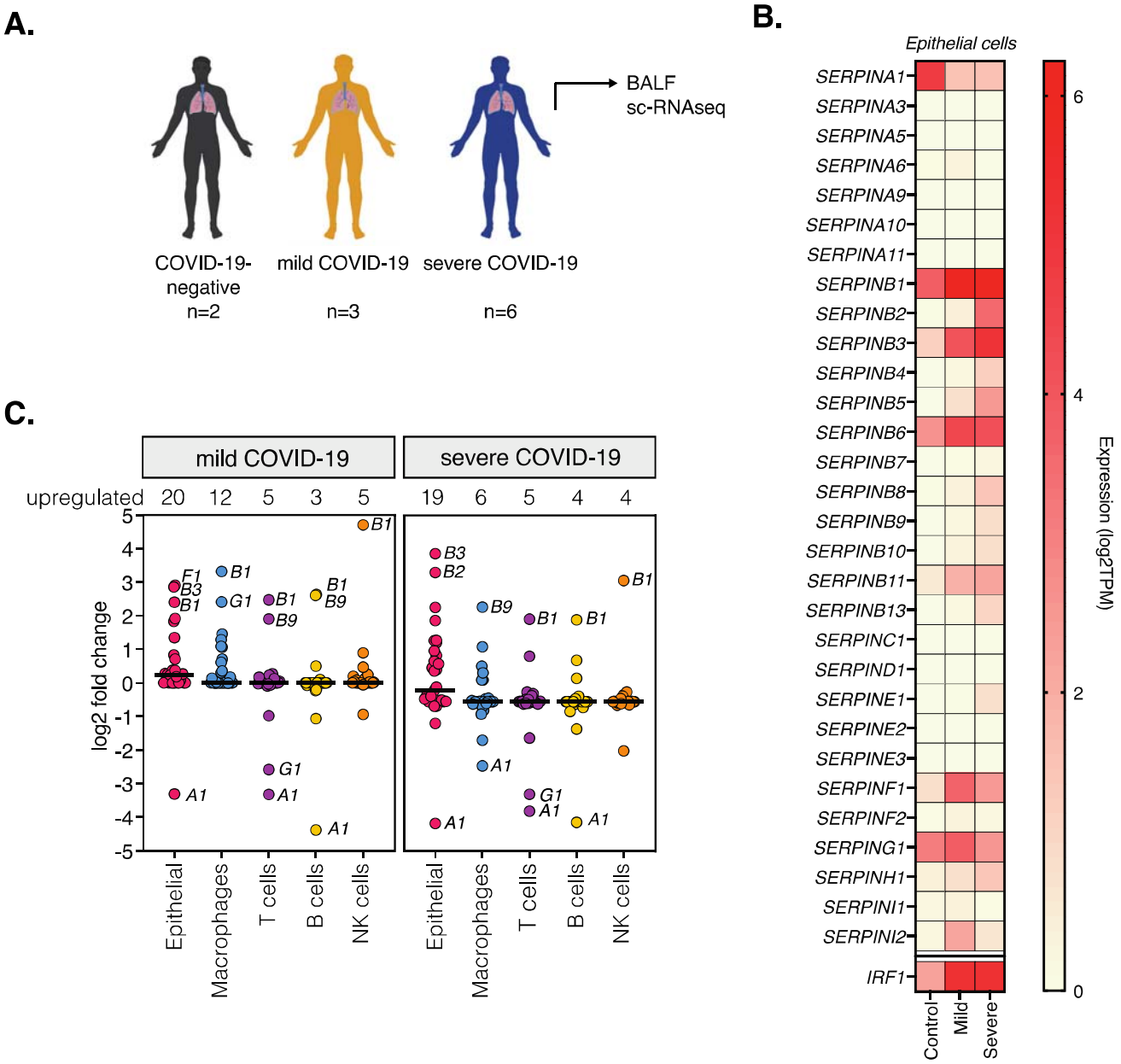

Figure 1. Integrated data from GEO accession number GSE145926 analyzed for *SERPIN* mRNA expression. A. Single-cell RNAseq (scRNA-seq) was performed on bronchioalveolar lavage fluids (BALF) of uninfected patients (n=2), and COVID-19 patients with mild (n=3) or severe (n=6) symptoms. **B.** Expression of *SERPIN*s and of interferon-stimulated gene *IRF1* in epithelial cells. TPM, reads per million. **C.** Fold-change of *SERPIN* expression across cell types relative to control patients. Number of upregulated (>2 fold) SERPINs is annotated above, and names of most up- or downregulated *SERPIN*s are annotated in the graph.

Together, our analyses suggest that *SERPIN*s are expressed to high levels in patients with COVID-19, predominantly in epithelial cells, and that *SERPIN* expression correlates with markers of the antiviral and inflammatory immune response.

SERPINs are expressed in response to respiratory virus infection in human airway epithelial cultures

While the COVID-19 pandemic allowed us to obtain unique samples such as BALF from acutely SARS-CoV-2-infected patients to study *SERPIN* and protease mRNA expression, such specimens are not readily available for other respiratory virus infections. Therefore, we next analyzed SERPIN expression *in vitro*, using a human airway epithelial culture (HAEC) model, which closely mimics the human bronchial epithelium in terms of cell composition (basal, ciliated, goblet and club cells), architecture, and secreted extracellular environment, including mucus and proteases [15]. We infected HAEC with a panel of clinically relevant respiratory viruses, including adenovirus 5 (AdV5), human reovirus (ReoV), human rhinovirus A (RhV), human parainfluenzavirus 3 (HPIV3), influenza A/California/2009 H1N1 virus (IAV) and SARS-CoV-2 ancestral isolate WA-1. At 24 or 72 h post infection, we harvested the cultures, isolated RNA, and quantified specific transcripts via RT-qPCR (Fig 2A).

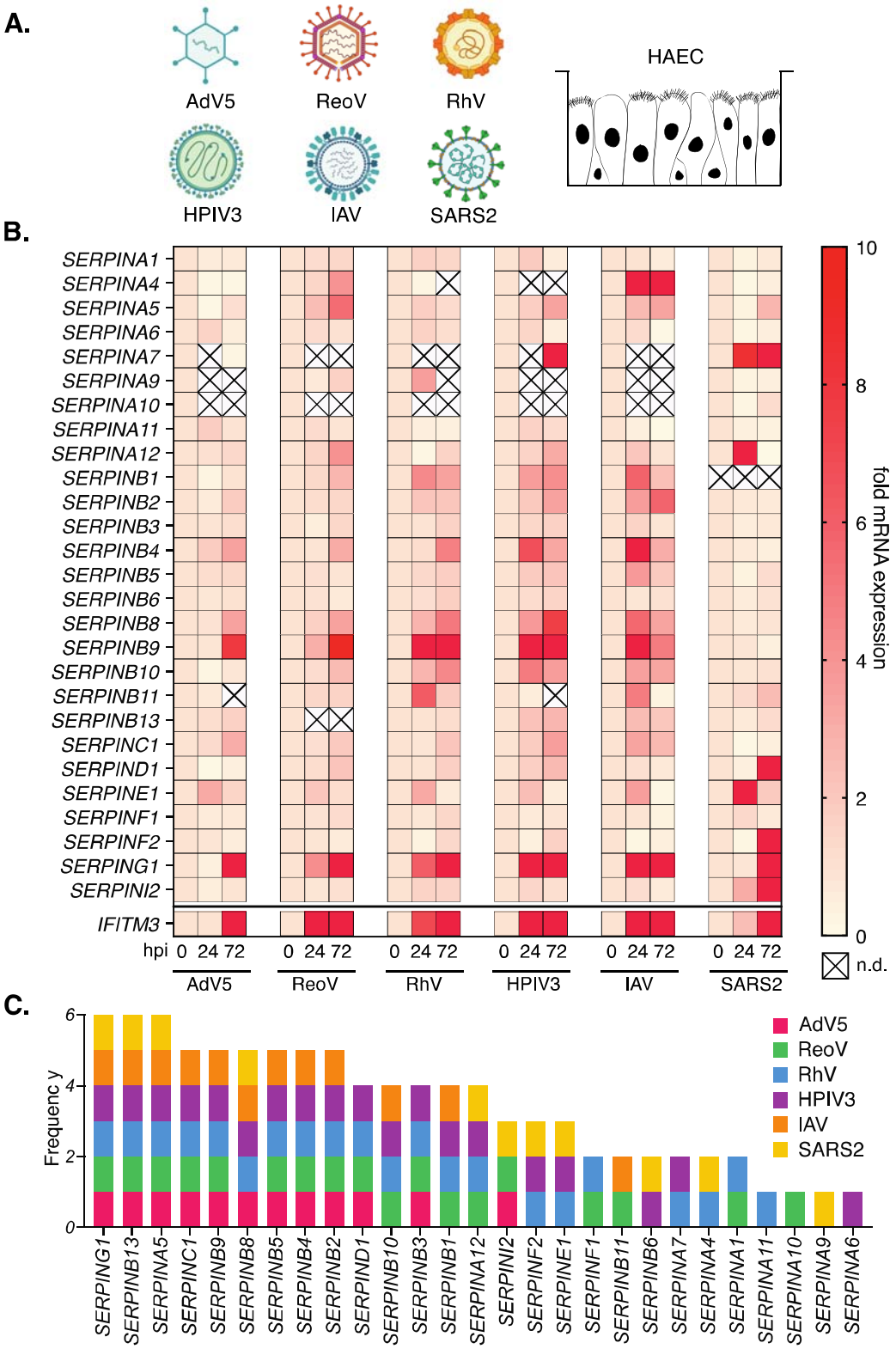

Figure 2. *SERPIN* mRNA expression upon viral infection of human airway epithelial cultures (HAEC). A. Schematic representation of a panel of respiratory viruses and of a trans-well containing a

polarized human airway epithelial culture (HAEC) growing in air-liquid interface. HAEC were infected apically with human rhinovirus A (RhV), influenza A/Californa/09 H1N1 virus (IAV), human parainfluenzavirus 3 (HPIV3), or SARS-CoV-2 WA-1 (SARS2), or basolaterally with human adenovirus 5 (AdV5) or human reovirus (ReoV). Cultures were lysed at indicated time points post-infection, total RNA isolated and transcripts quantified by RT-qPCR. Prototype interferon-stimulated gene *IFITM3* mRNA levels shown as -fold change over uninfected cultures over time. **B.** *SERPIN* and *IFITM3* mRNA, shown as -fold change over uninfected cultures at indicated time points. n.d., not detected. **C.** Frequency of upregulated (2-fold) SERPINs across viruses.

We first determined expression of the interferon-stimulated gene *IFITM3* (10) upon infection with the different viruses. As expected, we found *IFITM3* upregulated upon infection with all viruses tested (Figure 2B), validating the induction of an antiviral host response in HAEC upon infection. Of note, robust (>8-fold) *IFITM3* upregulation for both AdV5 and SARS-CoV-2 occurred later than for the other viruses, at the 72 h timepoint, which may reflect different viral replication rates, differential viral immune escape mechanisms, or both. Next, we examined *SERPIN* expression. Across all viral infections, we found 10 *SERPIN*s differentially expressed to higher levels (Figure 2B): *SERPINB1, SERPINB2, SERPINB4, SERPINB8, SERPINB9, SERPINB10, SERPINB13, SERPINC1, SERPINE1,* and *SERPING1.* Others, for example *SERPINF1*, are not differentially expressed upon infection with any virus, suggesting that they may not be part of the induced antiviral response in airway epithelial cells. Interestingly, and in agreement with the scRNA-seq data from COVID-19 patients (Figure 1), some *SERPIN*s are downregulated, as is the case for *SERPINA1*. Furthermore, by analyzing the frequencies of *SERPIN* expression across our panel of viruses, we observe that a subset of *SERPIN*s that are upregulated upon infection with each of the 6 viruses (e.g.

SERPING1), which may indicate a broad antiviral function of these *SERPINs* (Figure 2C). In contrast, other *SERPIN*s are differentially expressed upon infection by select viruses (e.g. *SERPINA1*), suggesting a different regulatory mechanism, and potentially select specificity against viruses.

Table 1. SERPINs and published targets

SERPIN	**protease(s)**	**target(s) involved in the lifecycle of***
A1	elastase, thrombin, trypsin, chymotrypsin, uPA	picornaviruses, RhV, HSV-1, HSV-2
A4	tissue kallikrein	IAV
A5	protein C, uPA, kallikrein	IAV
A10	factor Xa	HIV, SARS1, IAV, IBV, HSV, PIV-1, AdV, CMV, RSV, DENV
B1	protease 3, cathepsin G, elastase, Cathepsin L	RPV, HIV1, IAV, SARS1, SARS2, RhV
B2	uPA, tPA	HPV, IAV
B3	cathepsin G, cathepsin L, cathepsin K	SARS1, SARS2
B4	cathepsin G, chymase	IAV
B6	cathepsin G, thrombin, chymotrypsin, plasmin	HIV-1, HSV-2, IAV, HEV, HCV
B8	furin	EBOV, SINV, CoV, RSV, IAV, SFV, BDV, PIV, HSV-1, HSV-2, SARS2
B9	granzyme B, subtilisin A	HSV-1, HSV-2, IAV
B10	trypsin	IAV
B13	cathepsin L, cathepsin K	SARS1, SARS2
C1	thrombin, factor Xa	HIV-1, HSV-2, PIV, IAV, IBV, RSV, AdV, CMV
D1	thrombin	HSV-1, HSV-2
E1	uPA/tPA, TMPRSS2, HAT, trypsin	DENV, IAV, IBV, SeV, HCV, HSV-1, HSV-2
F2	plasmin	IAV
G1	complement C1 esterase	IAV
I1	uPA, urokinase, plasmin	IAV

Abbreviations: AdV, adenovirus; BDV, borna disease virus; CMV, cytomegalovirus; CoV, coronavirus; DENV, dengue virus; HCV, hepatitis C virus; HEV, hepatitis E virus; HSV, herpes sim- plex virus; IAV, influenza A virus; IBV, influenza B virus; PIV, parainfluenzavirus; RPV, rabbitpox virus; RSV, respiratory syncytial virus; RV, rotavirus; SARS1, SARS-CoV; SARS2, SARS-CoV-2; SFV, semliki forest virus; tPA, tissue plasminogen activator; uPA, urokinase plasminogen activator. * assembled from [19,31-62]

With this, our data indicate that a subgroup of *SERPIN*s are upregulated in response to viral infection, which correlates with the induction of a reference

ISG, thus indicating that some *SERPIN* family members may be part of the cell-intrinsic antiviral response in the human airway epithelial epithelium.

Development of an *in silico* protein-protein docking screen to identify SERPIN-protease pairs

So far, we have shown that *SERPIN*s such as *SERPINB1, SERPINB2, SERPINB4, SERPINB8, SERPINB9, SERPINB10, SERPINB13, SERPINC1, SERPINE1,* and *SERPING1* are upregulated upon diverse viral infections, suggesting that they may be part of the cell-intrinsic antiviral response. Due to their function in inhibiting proteases, we hypothesize that these SERPINs inhibit viruses by hindering proteases that are required in their respective viral life cycles. Although SERPINs are not recognized broadly as antiviral proteins, remarkably, many canonical SERPINs target proteases are identical to those required during viral life cycles (Table 1). However, the list of experimentally identified SERPIN targets may not be complete. Indeed, we previously found that SERPINE1, well-characterized to inhibit uPA and tPA [16], also inhibits human airway trypsin and TMPRSS2 [19]. Subsequently, it is feasible that we may yet discover SERPINs with previously unappreciated target proteases that are critical in viral infections.

To identify novel SERPIN-protease pairs, we developed an *in silico* screen based on 3D protein structures and leveraging our molecular knowledge of the SERPIN mechanism of action. A SERPIN's specificity for a given target protease is predominantly determined by the fit of the protruding SERPIN RCL into the protease active site [16,62]. A good fit increases the likelihood for the formation of a SERPIN-protease Michaelis

complex [18], followed by a nucleophilic attack by the protease active residue on the SERPIN RCL, and, ultimately, for the formation of a covalent link resulting in irreversible protease inhibition. Therefore, structural modeling of the fit between a SERPIN's RCL and a protease active site can be used to predict a favorable interaction, and thus identify novel SERPIN targets.

Given that our previous data has focused on SERPIN expression in airway epithelial cells, and that respiratory viruses are a clinically important group, we focused our search on host proteases expressed in the human airways, and proteases encoded by respiratory viruses. We first curated a database of 52 proteases expressed in the respiratory tract, based on the LungMAP human expression database (Figure 3A) [63]. The final database contained members of the cysteine, serine, aspartic and threonine protease families. Next, we obtained, if publicly available, the 3D structures of each individual protease and each of the 10 SERPIN candidates. If not publicly available, we homology-modeled the 3D structure using Swiss Modeler 2.0 [64-66] (Figure 3A). Next, we mapped all protease active sites and all SERPIN RCL by assessing individual entries on UniProt [67], with the goal of performing protein-protein docking of each individual protease with each individual SERPIN via the HADDOCK 2.4 server [68,69]. HADDOCK is an informed protein-protein docking server, which docks regions of interest and predicts many structural conformations of the complex, then ranks them by score (Figure 3A). The calculated HADDOCK score is proportional to the Van der Waals, electrostatic, desolvation, and restraint energy of the complex, therefore allowing for the identification of complexes with favorable interaction energy [70]. Since these

energy values are of negative nature, a lower HADDOCK score corresponds to a higher likelihood of a favorable interaction. We used this information to guide our docking screen. While specifying the eight central active residues of each SERPIN (P4-P4') that dictate specificity [16,71], we docked the entire SERPIN RCL to the complete protease active site, and identified putative binding pairs based on low HADDOCK scores.

To test the feasibility of our strategy, we docked known target and known non-target proteases with SERPINB1 and determined their HADDOCK scores. For example, SERPINB1 is a known inhibitor of Cathepsin G, but does not inhibit Granzyme A [72,73]. As expected, HADDOCK identified that the SERPINB1-Cathepsin A complex represents a good fit of the RCL into the protease active site. In contrast, the SERPINB1 RCL fails to fit into the active site of Granzyme A (Figure 3B). Indeed, the distance between active residues of Cathepsin G and SERPINB1 is <5A, whereas the distance between active residues of Granzyme A and SERPINB1 is multiple times broader (Figure 3B), suggesting that the positive control pair exhibits better positioning of active residues for initiation of the nucleophilic attack by the protease and furthering the SERPIN inhibitory mechanism [16].

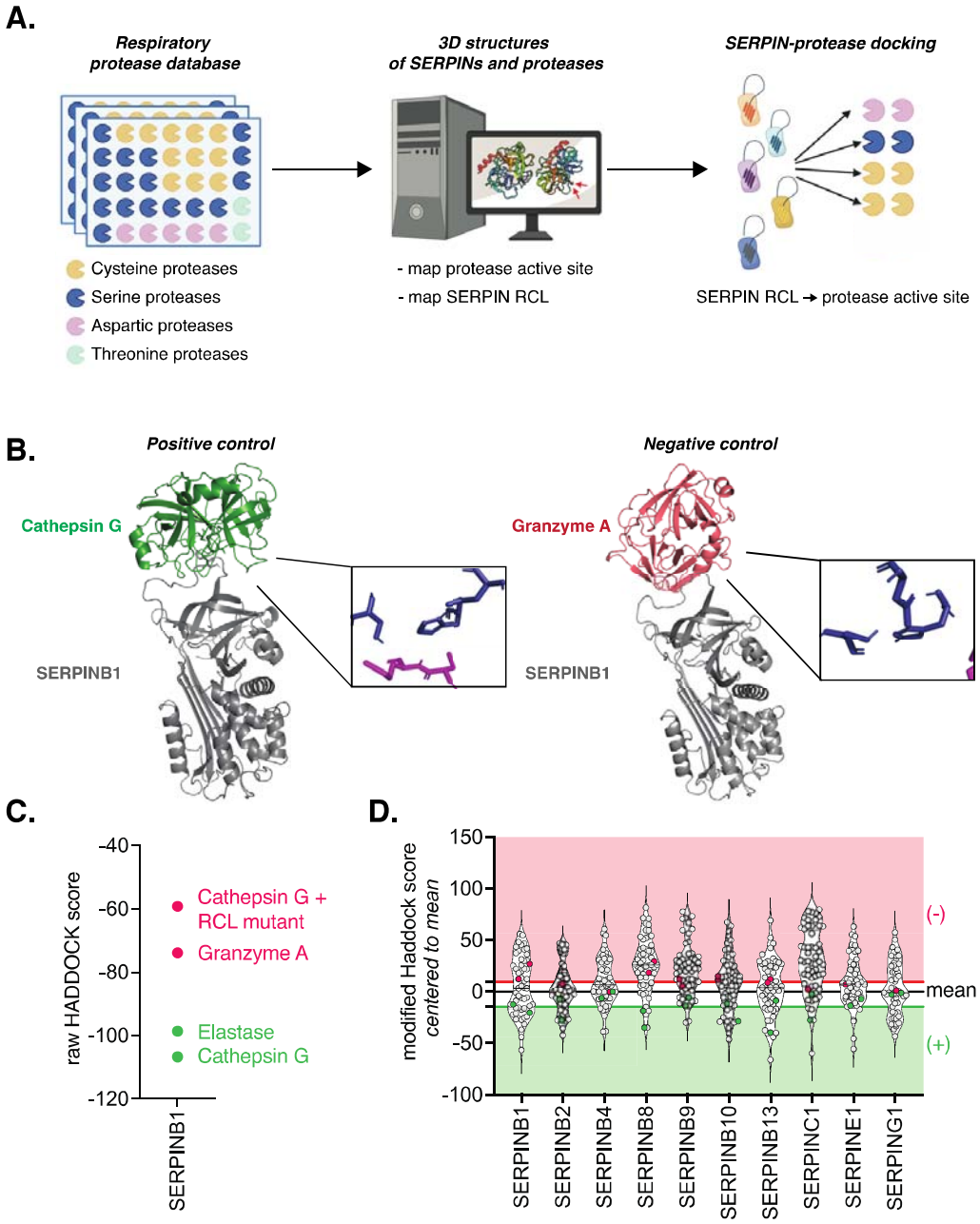

Figure 3. Establishing the *in silico* screen to identify SERPIN-protease pairs.

Figure 3. Establishing the *in silico* screen for identification of SERPIN-protease pairs. A. Workflow of the *in silico* docking screen. **B.** Validation of docking using as an example the structures of SERPINB1 with positive (Cathepsin G) or negative (Granzyme A) controls. **C.** Validation of raw scoring of known positive and negative hits. Example of raw HADDOCK scores for SERPINB1 positive controls (Cathepsin G, Elastase) or negative controls (RCL mutant + elastase, Granzyme A). **D.** Normalized scoring to

facilitate the detection of new SERPIN-protease interaction pairs. Violin plots of HADDOCK scores for all SERPIN-protease pairs. Red line, mean negative control; shaded red, non-hit range; green line, mean positive control, shaded green, hit range; red data points, negative controls; green data points, positive controls. n=10 SERPINsm n=52 proteases.

We further validated our strategy by docking two known positive and two known negative controls for each SERPIN and quantifying their interaction via HADDOCK. If no negative controls were known, we generated SERPIN RCL *in silico* mutants by converting P4-P4' residues to alanines, determining the structure of the mutated SERPIN using Swiss Modeler 2.0 [64-66], and docking the mutant structure to a positive control protease. For example, SERPINB1 docked to the positive controls Cathepsin G and Elastase yielded more negative HADDOCK scores, while SERPINB1 docked to the negative control Granzyme A or the SERPINB1 RCL mutant docked to Elastase yielded more positive scores (Figure 3C). These results validated our strategy to assess interactions between individual SERPINs and proteases.

Upon validation of our strategy, we performed docking for each of the 52 individual proteases, with each of the 10 individual SERPINs or RCL mutant SERPINs, and calculated a raw HADDOCK score for each of the pairs. We then adjusted our dataset for SERPIN-SERPIN variation and mathematically discriminated pairs based on positive and negative controls. To be able to define a positive (likely inhibitory) threshold common to all SERPINs in this study, we created a normalized Z-score centered to the mean of both positive and negative controls for each SERPIN (Figure 3D). Using this modified HADDOCK score (Figure 3D) combined with visual assessment of the

SERPIN RCL fit into the protease active site (Figure 3B), we proceeded to screen our panel of respiratory protease-SERPIN pairs in greater detail one by one (Figure 4A).

In silico screen identifies potential SERPIN-protease pairs relevant to viral life cycles.

To visualize novel SERPIN targets that may have antiviral activity in the airways, we plotted modified HADDOCK scores for each SERPIN-protease pair in a heatmap (Figure 4A). Proteases are listed on the x-axis and include 52 host airway proteases and 3 virally encoded proteases. The 10 SERPINs that we previously identified to be expressed in HAEC upon viral infections (Figure 2) are listed on the y-axis.

First, we observe that known SERPIN-protease pairs display a low HADDOCK score (shown in dark red; i.e. SERPINB1 and Cathepsin A) and known non-binders display a high HADDOCK score (shown in dark grey; i.e. SERPINC1 and TMPRSS11C), as expected. Second, we observe that individual SERPINs, such as SERPINC1, SERPINB8 and SERPINB9, display low HADDOCK scores with only few host proteases, suggesting that these SERPINs have high specificity. In contrast, other SERPINs, such as SERPINB1 or SERPINB13, display low HADDOCK scores with several host proteases along with proteases from different families, suggesting that these SERPINs may be more promiscuous. Third, we observe that some host proteases, notably most TMPRSS family members and Cathepsin E, do not show negative HADDOCK scores with most SERPINs, suggesting that these proteases may be generally less sensitive to the action of SERPINs. This may be particularly relevant

for viruses that rely on these proteases (particularly TMPRSS2 and TMPRSS4) to complete their life cycles, such as SARS-CoV-2 and IAV [7,30]. In contrast, other host proteases, most notably trypsin, plasma kallikrein and cathepsin A show low HADDOCK scores with almost all SERPINs, suggesting that they may be targeted by multiple SERPINs redundantly. Finally, we examined relevant viral proteases such as SARS-CoV-2 $3CL^{pro}$ and PL^{pro}, as well as Ad^{pro}. We observe a strikingly low HADDOCK score for AdV5 protease with SERPINB2. Indeed, the docked structure of the Ad^{pro}-SERPINB2 complex shows a good fit of the SERPINB2 RCL into the Ad^{pro} active site (Figure 4B). Moreover, the structure reveals the presence of exosites, which are defined as residues outside the active regions in both SERPINs and proteases that further stabilize the interaction [62,71]. When looking at the protease active site in detail, we observe that the active cysteine is positioned in proximity for a nucleophilic attack to the P1-P1' central active residues of SERPINB2 (Figure 4C). This gives us confidence of a good fit and potential subsequent completion of the SERPIN inhibitory mechanism. To our knowledge, this is the first indication of a human SERPIN potentially targeting a protease of viral origin.

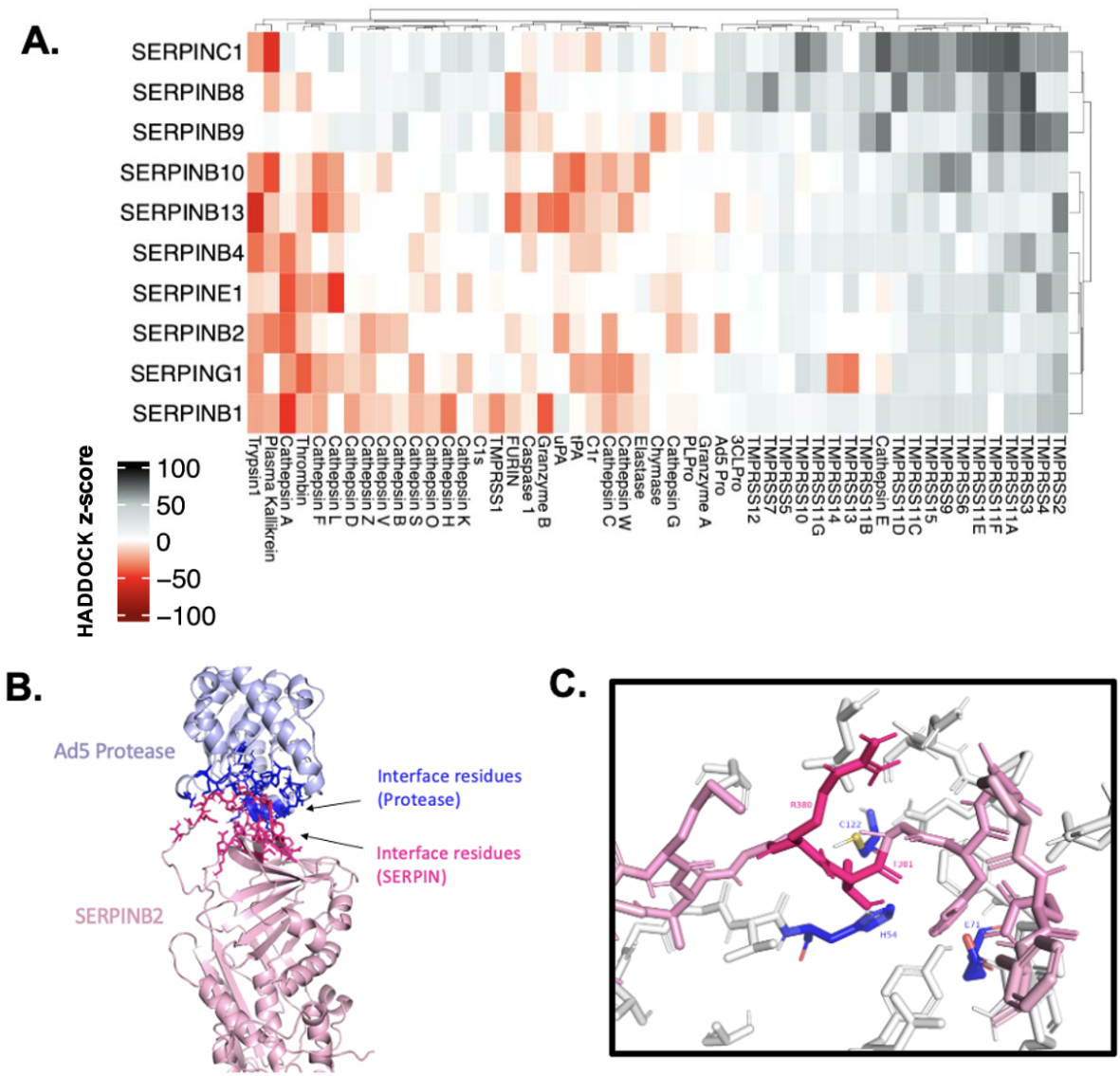

Figure 4. *In silico* screen of 10 SERPINs against a panel of host respiratory or viral proteases. A. Heatmap-formatted results of docking screen, where Z-scores centered to Haddock mean of control SERPIN-protease pairs are shown in a color scale. Negative Z-scores (red) are identified hits, positive Z-scores (grey) are non-hits. B. HADDOCK v2.4 complex of Adenovirus 5 Protease docked to SERPINB2, showing appropriate Michaelis SERPIN-protease fitting and exosites. Interacting residues are highlighted in sticks; visualized via Pymol v2.0. C. Zoomed-in view of the residues within the catalytic triad of Adenovirus 5 Protease. Blue, nitrogen; red, oxygen; yellow, sulfur. Visualized on Pymol v2.0.

Discussion

Our group was the first to show that a SERPIN could target viral proteolysis and cause an antiviral effect, particularly by inhibiting IAV H1 cleavage [19]. Moreover, we found that SERPINE1 was able to inhibit proteases beyond its 'canonical' target uPA [19]. However, several questions remained: Is the antiviral phenotype SERPIN- and virus-specific? Are other SERPINs involved in responding to respiratory viral infection? Which other respiratory proteases are targeted by SERPINs? We show here that select SERPINs are upregulated upon SARS-CoV-2 infection in humans, particularly in SARS-CoV-2-permissive airway epithelial cells (Figure 1). Using a human airway epithelial culture *in vitro* model, we further identified SERPINs that are differentially expressed upon infection with other respiratory viruses (Fig 2). We developed and executed an *in silico* docking screen which identified novel SERPIN target proteases with relevance in respiratory virus lifecycles (Fig 4). In all, our study shows that SERPINs are upregulated upon viral infection and may target several proteases involved in the life cycles of respiratory viruses.

SERPINs have been studied in many contexts of human disease and sequencing efforts have shown tissue specificities in their expression [16,74]. Increasing data shows that SERPINs can be expressed in response to inflammatory cytokines [19,23-25,75]. In agreement with these data, we show that *SERPINs* are expressed in response to inflammation both in COVID-19 patients and in virus-infected human airway epithelial cultures. Interestingly, seven clade B *SERPINs* are differentially expressed to higher levels in our *in vitro* HAEC model; namely, *SERPINB1*, *SERPINB2*, *SERPINB4*,

SERPINB8, *SERPINB9*, *SERPINB10*, and *SERPINB13*. This suggests that they are produced in response to respiratory viral infection; however, the mechanism of SERPIN gene regulation remains unclear. Moreover, clade B SERPINs have been shown to emerge through gene duplications, indicative of host-pathogen mediated evolution. Interestingly, we observed a difference between our *in vitro* HAEC model and data obtained from BALF of SARS-CoV-2 infected patients, as we see additional SERPINs upregulated *in vivo*. This may be due to the presence of activated immune cells in the respiratory tract, which are not present in the HAEC model, because, as previously shown, specific SERPIN levels can be stimulated by macrophage-derived proinflammatory cytokines [76]. This represents a limitation of the HAEC approach, as it lacks non-epithelial sources of signaling molecules. However, our *in vitro* model still captures intra-epithelial signals, making it highly informative for respiratory virus biology at the host-pathogen interface. Our scRNA-seq data identified cathepsins as the most robustly expressed family of proteases, even in uninfected control patients, indicating their relevance in respiratory virus infections. Altogether, we show that SERPINs are upregulated upon infection with diverse respiratory viruses, suggesting that they may be part of the respiratory antiviral arsenal.

Molecular docking has been widely used in drug discovery, for instance to visualize and predict the fit of a small molecule inhibitor inside the catalytic site of an enzyme [77]. Additionally, docking screens are used to screen small molecule compounds to inhibit or enhance a relevant enzymatic reaction [77]. In contrast to this vast body of studies using docking to predict protein-drug interactions, there is limited data docking protein-protein

interactions. The HADDOCK platform allows for informed protein-protein docking, optimizing the docked complex by “informing” the process; namely, using only the regions of interests on both molecules to guide the interaction [68]. This greatly benefits our study as it is known that the SERPIN mechanism of protease inhibition is specific to the RCL reaching into the active site. However, to validate each score that HADDOCK generates, we must visually assess whether the fit is appropriate, which is feasible for the scale of this investigation, but may be a limitation for high-throughput use of this method. Future optimization of this screen may focus on quantifying the fit-specific interaction energies to help select appropriate complexes.

The HADDOCK platform provides a score of the putative complexes, which allows us to quantitatively assess a SERPIN’s ability to fit into the protease active site and predict candidates for inhibition based on an informed protein-protein docking approach. This is particularly important as a single amino acid change in a SERPIN RCL can drastically shift protease targets [71]. Thus, structural modeling and docking is a powerful tool to narrow down candidates when characterizing SERPIN activity, and our approach provides a wide lens for screening candidate binding pairs. However, some limitations are inherent to our strategy. For example, is does not consider the action of cofactors on SERPIN specificity, as cofactors may not be co-crystalized with SERPINs or proteases. One such example is SERPINC1, which requires the cofactor heparin to mediate the interaction in a bridge-like manner between SERPINC1 and its canonical target thrombin [16]. While SERPINC1-heparin-thrombin complex has been crystallized and can be used as a template for modeling, other cofactors may not be known or have

been co-crystallized with any of the relevant molecules [78]. It is also important to note that our screen does not take into account intracellular localization of SERPIN and target protease. Both molecules need to meet and locate within the same subcellular compartment or extracellular space for inhibition to occur. For instance, SERPINB8 targets furin, which localizes in the golgi apparatus [52,79,80]. When expressed, SERPINB8 is located along the secretory pathway and primarily in the golgi apparatus, therefore facilitating productive inhibition [79]. Despite this limitation of our approach, the localization of our ten SERPINs of interest is described in the literature, as well as most proteases. Moreover, some of the large protease families are localized in the same compartment, such as TMPRSSs and cathepsins, which localize to the plasma membrane and endosomes, respectively [30,81,82]. We used this information, when available, to prioritize hits that we expect to colocalize or temporarily meet throughout the same pathway, based on literature.

Viruses can utilize redundant host proteolytic activities; thus, a combination of SERPINs may be required for an antiviral effect. Redundant proteolytic pathways have been shown in SARS-CoV-2 and IAV, as their fusion glycoproteins can be processed by a wide range of proteases [7,30,39,81]. This is facilitated by the presence of a polybasic cleavage site in their fusogenic glycoproteins, which renders them accessibility to a broad range of host respiratory proteases for activation [83]. From this, we can infer that a combination of SERPINs may be required for a measurable antiviral effect. For these reasons, hits from our screen will require future experimental validation.

Our *in silico* screen revealed that out of our panel of ten SERPINs upregulated during viral infections, all were predicted to inhibit more than one protease, but some were predicted to be more promiscuous than others. Particularly promiscuous is SERPINB1, which is the most ubiquitously expressed SERPIN of the B clade [74] and has been shown to block cathepsin proteases such as cathepsin G [16]. In contrast, other SERPINs are more specific, as is the case for SERPINC1, which does not bind favorably with many respiratory proteases *in silico*, likely because its targets are more focused to blood clotting and coagulation pathways [16].

We identified several previously unknown SERPIN-protease pairs with potential function in viral restriction. First, we identified a human SERPIN potentially targeting a viral protease: SERPINB2 and Ad^{pro}. If confirmed experimentally, this would be the first report of a direct-acting antiviral SERPIN, which not only represents cross-species modulation of enzymes at the host-pathogen interface but could inform a novel antiviral target. Second, we identified SERPINs potentially targeting host proteases that process surface glycoproteins of multiple clinically impactful respiratory viruses. This includes SERPINB8 inhibiting furin, which serves to mature glycoproteins of high-pathogenic IAV, HPIV3 and SARS-CoV-2 [5,7,81,84-86]. Third, we identified additional putative inhibitors of the SARS-CoV-2 lifecycle. This is of particular interest, as the current COVID-19 pandemic has highlighted the need for tools that have the potential to inform novel therapeutics. SARS-CoV-2 encodes two proteases: $3CL^{pro}$ PL^{pro}, which both cleave polyprotein intermediates [9]. Inhibition of such proteases indeed stops viral replication at a single cycle time frame[15]. We show that clade B SERPINs such as

SERPINB1, SERPINB2 and SERPINB13 may inhibit these proteases. Further, SARS-CoV-2 relies on host proteases to mediate fusion with the plasma or endosomal membranes and enter cells by cleaving the Spike (S) protein and triggering its fusion peptide [9]. A key protease mediating this endosomal entry is cathepsin L, however other cathepsin family members such as cathepsins V, B, K and S have also been shown to cleave S [39]. The omicron lineage has been described to use the cathepsin/endosomal pathway of entry as opposed to the plasma membrane, and omicron mutations in S have been associated with this phenotype [87]. Interestingly, we found cathepsins to be more prone to interaction with our panel of SERPINs. This adds relevance towards describing all putative inhibitors of cathepsin L as a way to target the currently predominant omicron clade.

We believe these studies to be an important initial step in characterizing the broad and important role of SERPINs as an antiviral mechanism. Future directions include testing for physical inhibition of the select proteases by their SERPIN hits. Furthermore, we will characterize the mechanisms of the regulation of these SERPINs, and whether they are part of particular antiviral signaling pathways. Lastly, we will perform antiviral assays in overexpression models to assess the magnitude of antiviral activity elicited by SERPINs.

Material and Methods

COVID19 Patient scRNA-seq data acquisition and analysis

Publicly available COVID-19 patient scRNA-seq data were obtained from the Gene Expression Omnibus database under the accession code GSE145926 [28]. Aligned and normalized transcript per million (TPM) values were obtained from the integrated ToppCell web server [63]. Frequencies and log2 fold changes were calculated from the normalized TPM values by the healthy controls. Upregulated SERPINs were defined as those expressed to >2-fold higher levels in COVID-19 patients compared to uninfected controls. Plots were generated in GraphPad Prism v9.

Differentiation of HAEC

Differentiation of hTert-immortalized bronchial basal progenitor cells into stratified human airway epithelial cultures was performed as stated in the following publication [15].

Infection of HAEC with respiratory virus panel

Bci-NS1.1 cells (kind gift of Dr. Ronald Crystal) were seeded into Pneumacult™-Ex Plus Medium (StemCell) and passaged at least two times before plating ($7x10^4$ cells/well) on rat-tail collagen type-1 coated permeable transwell membrane supports (6.5mm; Corning Inc) to generate human airway epithelial cultures (HAE). HAE were maintained in Pneumacult™-Ex Plus Medium until confluent, then grown at an air-liquid interface with Pneumacult™-ALI Medium in the basal chamber for approximately 3 weeks to form

differentiated, polarized cultures that resemble *in vivo* pseudostratified mucociliary epithelium. For infection with influenza A/California/07/2009 virus (BEI resources, NR-13663), HPIV3-GFP (kind gift of Dr. Peter Collins), and Rhinovirus A16 (kind gift of Dr. Ann Palmenberg), cultures were washed twice with room-temperature PBS, then virus was added apically ($5x10^6$ PFU for IAV, $1.12x10^7$ FFU for PIV, and 0.39 $TCID_{50}$ for Rhinovirus) in 50 µl of PBS and incubated for one hour at 37°C for IAV and PIV, and at 4°C for 90 minutes for Rhinovirus. Viruses were aspirated and HAE cultures were washed apically once with room-temperature PBS, then incubated at 37°C until endpoint. For infection with human Reovirus T3 SA (kind gift of Dr. Terrence Dermody) and Adenovirus 5-GFP (kind gift of Dr. Patrick Hearing), transwells were placed inverted in a sterile 5cm petri dish bottom and all remaining basal media was aspirated from the basal side of the transwells. 25 µl of each virus was added to the basal side of the transwells ($1.3x10^8$ PFU for Reovirus and $8.85x10^5$ PFU for Adenovirus) and incubated for one hour at 37°C for Adenovirus and room temperature for Reovirus. Viruses were aspirated from the basal side of the transwells, and transwells were placed back into ALI media and incubated at 37°C until endpoint. At each endpoint, transwell membranes were submerged in 375 µl Buffer RLT (Qiagen) and cell lysate was frozen at -80°C for future RNA extraction.

RT-qPCR of virus-infected HAEC

Infected HAEC were lysed as described above, and RNA was isolated using the Qiagen RNeasy kit following manufacturer's instructions. Extracted RNA was then quantified via NanoDrop and diluted to correct difference in yield. Relative mRNA levels, normalized

to housekeeping gene *RPS11*, were determined by RT-qPCR (SuperScript III First Strand Synthesis System, Life Technologies and PowerUP SYBR Green Master Mix, Thermo Fisher Scientific). Primers used are listed in Supplementary Table 1.

Protease and SERPIN database & structure modeling

Protein structures were obtained directly from the Protein Data Bank (PDB) or modeled by homology using Swiss PDB Modeler [64-66]. If obtained from the PDB, structures were refined by removing solvents, co-factors, small molecules, and co-crystallized atoms. Selection of proteases for later docking was based on the LungMAP database [63] of human airway expressed proteases and protease families, mainly prioritizing those expressed to >10 Log2TPM on epithelial cells of donors.

Active sites and SERPIN RCLs were mapped on to refined structures using the UniProt server [67].

Protein-protein docking

Docking of complexes was performed using the HADDOCK v2.4 server [68,69]. Selection of primary residues to perform docking was based on the residues specified on UniProt as the RCL's 8 critical residues found to dictate SERPIN specificity - P4-P4' for SERPINs and the catalytic triad within the protease active site for proteases. Secondary residues were defined on a range of 13A radius from the primary residues, this ensures to consider the protease pockets that may fit RCL sidechains and SERPIN-protease exosites that may stabilize interactions and complex formation. Furthermore,

the full RCL was given full flexibility, starting at the poly-alanine segment and ending at the PF loop end motif. Upon submission and finish of an entry on the HADDOCK v2.4 server, we evaluated the resulting complex position and specifically selected complexes that had RCLs fit inside the protease active site. If no such complex was found, the protease was removed from the screen. The score of the visually fitting complexes was then used to guide the initial optimization and calculate thresholds for the Z-scores. Structures of the complexes were then saved and visualized on PyMoL v2.0 for distance calculation and deeper examination of the positioned residues (Figure 4) [88].

Statistical analysis and docking screen scoring

To define the thresholds of SERPIN-protease docking, Haddock scores were determined from at least two positive and two negative control proteases for each SERPIN, known to bind or not bind SERPINs *in vitro,* respectively. For each SERPIN, average HADDOCK scores were calculated separately for the positive and negative controls, and the mean of both averaged scores served as the in-silico threshold of SERPIN-protease interaction. The SERPIN-specific HADDOCK thresholds were used to normalize the full data set of HADDOCK scores and to generate Z-scores, i.e., the thresholds were set to zero. Violin plots were generated using GraphPad Prism v.9, and clustered heatmaps of the normalized, Z-scored data were generated using the ComplexHeatmap package in program R v.4.1.2 [89].

Acknowledgements

We thank all members of the Dittmann lab for their critical input on data and the manuscript. We also thank Dr. Ronald Crystal for the kind gift of Bci-NS1.1 cells, Dr. Terrence Dermody for Reovirus T3 SA, Dr. Ann Palmenberg for Rhinovirus A16, Dr. Peter Collins for PIV3-GFP, and Dr. Patrick Hearing for Adenovirus 5-GFP. Research was supported by the following grants from NIH/NIAID: R01AI143639 and R21AI139374 to M.D., as well as T32AI100853 to J.R.G. Work was further supported by The Vilcek Institute of Graduate Biomedical Sciences, and by NYU Grossman School of Medicine Startup funds.

References

1. Organization, W.H. WHO COVID-19 Dashboard. **2020**.
2. Weston, S.; Frieman, M.B. Respiratory Viruses. *Encyclopedia of Microbiology* **2019**, 85-101, doi:doi:10.1016/B978-0-12-801238-3.66161-5.
3. Henrickson, K.J. Parainfluenza viruses. *Clin Microbiol Rev* **2003**, *16*, 242-264, doi:10.1128/cmr.16.2.242-264.2003.
4. Kemp, G.; Webster, A.; Russell, W.C. Proteolysis is a key process in virus replication. *Essays Biochem* **1992**, *27*, 1-16.
5. Aguilar, H.C.; Buchholz, D.W. Finding proteases that make cells go viral. *J Biol Chem* **2020**, *295*, 11408-11409, doi:10.1074/jbc.H120.015153.
6. Forsythe, H.M.; Galvan, J.R.; Yu, Z.; Pinckney, S.; Reardon, P.; Cooley, R.B.; Zhu, P.; Rolland, A.D.; Prell, J.S.; Barbar, E. Multivalent binding of the partially disordered SARS-CoV-2 nucleocapsid phosphoprotein dimer to RNA. *Biophys J* **2021**, doi:10.1016/j.bpj.2021.03.023.
7. Bertram, S.; Glowacka, I.; Steffen, I.; Kühl, A.; Pöhlmann, S. Novel insights into proteolytic cleavage of influenza virus hemagglutinin. *Rev Med Virol* **2010**, *20*, 298-310, doi:10.1002/rmv.657.
8. Chang, A.; Dutch, R.E. Paramyxovirus fusion and entry: multiple paths to a common end. *Viruses* **2012**, *4*, 613-636, doi:10.3390/v4040613.
9. Masters, P.S. The molecular biology of coronaviruses. *Adv Virus Res* **2006**, *66*, 193-292, doi:10.1016/s0065-3527(06)66005-3.
10. Jensen, L.M.; Walker, E.J.; Jans, D.A.; Ghildyal, R. Proteases of human rhinovirus: role in infection. *Methods Mol Biol* **2015**, *1221*, 129-141, doi:10.1007/978-1-4939-1571-2_10.
11. Kulanayake, S.; Tikoo, S.K. Adenovirus Core Proteins: Structure and Function. *Viruses* **2021**, *13*, doi:10.3390/v13030388.
12. Greber, U.F.; Webster, P.; Weber, J.; Helenius, A. The role of the adenovirus protease on virus entry into cells. *EMBO J* **1996**, *15*, 1766-1777.
13. de Leuw, P.; Stephan, C. Protease inhibitors for the treatment of hepatitis C virus infection. *GMS Infect Dis* **2017**, *5*, Doc08, doi:10.3205/id000034.
14. Lv, Z.; Chu, Y.; Wang, Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)* **2015**, *7*, 95-104, doi:10.2147/HIV.S79956.
15. de Vries, M.; Mohamed, A.S.; Prescott, R.A.; Valero-Jimenez, A.M.; Desvignes, L.; O'Connor, R.; Steppan, C.; Devlin, J.C.; Ivanova, E.; Herrera, A.; et al. A comparative analysis of SARS-CoV-2 antivirals characterizes 3CL. *J Virol* **2021**, doi:10.1128/JVI.01819-20.
16. Law, R.H.; Zhang, Q.; McGowan, S.; Buckle, A.M.; Silverman, G.A.; Wong, W.; Rosado, C.J.; Langendorf, C.G.; Pike, R.N.; Bird, P.I.; et al. An overview of the serpin superfamily. *Genome Biol* **2006**, *7*, 216, doi:10.1186/gb-2006-7-5-216.
17. Sanrattana, W.; Maas, C.; de Maat, S. SERPINs-From Trap to Treatment. *Front Med (Lausanne)* **2019**, *6*, 25, doi:10.3389/fmed.2019.00025.
18. Ye, S.; Cech, A.L.; Belmares, R.; Bergstrom, R.C.; Tong, Y.; Corey, D.R.; Kanost, M.R.; Goldsmith, E.J. The structure of a Michaelis serpin-protease complex. *Nat Struct Biol* **2001**, *8*, 979-983, doi:10.1038/nsb1101-979.

680 19. Dittmann, M.; Hoffmann, H.H.; Scull, M.A.; Gilmore, R.H.; Bell, K.L.; Ciancanelli, M.;
681 Wilson, S.J.; Crotta, S.; Yu, Y.; Flatley, B.; et al. A serpin shapes the extracellular
682 environment to prevent influenza A virus maturation. *Cell* **2015**, *160*, 631-643,
683 doi:10.1016/j.cell.2015.01.040.
684 20. Rosendal, E.; Mihai, I.S.; Becker, M.; Das, D.; Frängsmyr, L.; Persson, B.D.; Rankin, G.D.;
685 Gröning, R.; Trygg, J.; Forsell, M.; et al. Serine Protease Inhibitors Restrict Host
686 Susceptibility to SARS-CoV-2 Infections. *mBio* **2022**, *13*, e0089222,
687 doi:10.1128/mbio.00892-22.
688 21. Azouz, N.P.; Klingler, A.M.; Callahan, V.; Akhrymuk, I.V.; Elez, K.; Raich, L.; Henry, B.M.;
689 Benoit, J.L.; Benoit, S.W.; Noé, F.; et al. Alpha 1 Antitrypsin is an Inhibitor of the SARS-
690 CoV-2-Priming Protease TMPRSS2. *Pathog Immun* **2021**, *6*, 55-74,
691 doi:10.20411/pai.v6i1.408.
692 22. Heit, C.; Jackson, B.C.; McAndrews, M.; Wright, M.W.; Thompson, D.C.; Silverman, G.A.;
693 Nebert, D.W.; Vasiliou, V. Update of the human and mouse SERPIN gene superfamily.
694 *Hum Genomics* **2013**, *7*, 22, doi:10.1186/1479-7364-7-22.
695 23. Prada, A.E.; Zahedi, K.; Davis, A.E. Regulation of C1 inhibitor synthesis. *Immunobiology*
696 **1998**, *199*, 377-388, doi:10.1016/S0171-2985(98)80042-9.
697 24. He, W.; Tan, R.; Dai, C.; Li, Y.; Wang, D.; Hao, S.; Kahn, M.; Liu, Y. Plasminogen activator
698 inhibitor-1 is a transcriptional target of the canonical pathway of Wnt/beta-catenin
699 signaling. *J Biol Chem* **2010**, *285*, 24665-24675, doi:10.1074/jbc.M109.091256.
700 25. Dong, C.; Zhu, S.; Wang, T.; Yoon, W.; Goldschmidt-Clermont, P.J. Upregulation of PAI-1
701 is mediated through TGF-beta/Smad pathway in transplant arteriopathy. *J Heart Lung*
702 *Transplant* **2002**, *21*, 999-1008, doi:10.1016/s1053-2498(02)00403-5.
703 26. Lee, A.J.; Ashkar, A.A. The Dual Nature of Type I and Type II Interferons. *Front Immunol*
704 **2018**, *9*, 2061, doi:10.3389/fimmu.2018.02061.
705 27. Liu, T.; Zhang, L.; Joo, D.; Sun, S.C. NF-κB signaling in inflammation. *Signal Transduct*
706 *Target Ther* **2017**, *2*, doi:10.1038/sigtrans.2017.23.
707 28. Liao, M.; Liu, Y.; Yuan, J.; Wen, Y.; Xu, G.; Zhao, J.; Cheng, L.; Li, J.; Wang, X.; Wang, F.; et
708 al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat*
709 *Med* **2020**, *26*, 842-844, doi:10.1038/s41591-020-0901-9.
710 29. Wang, W.; Xu, L.; Brandsma, J.H.; Wang, Y.; Hakim, M.S.; Zhou, X.; Yin, Y.; Fuhler, G.M.;
711 van der Laan, L.J.; van der Woude, C.J.; et al. Convergent Transcription of Interferon-
712 stimulated Genes by TNF-α and IFN-α Augments Antiviral Activity against HCV and HEV.
713 *Sci Rep* **2016**, *6*, 25482, doi:10.1038/srep25482.
714 30. Zang, R.; Gomez Castro, M.F.; McCune, B.T.; Zeng, Q.; Rothlauf, P.W.; Sonnek, N.M.; Liu,
715 Z.; Brulois, K.F.; Wang, X.; Greenberg, H.B.; et al. TMPRSS2 and TMPRSS4 promote SARS-
716 CoV-2 infection of human small intestinal enterocytes. *Sci Immunol* **2020**, *5*,
717 doi:10.1126/sciimmunol.abc3582.
718 31. Walters, M.S.; Gomi, K.; Ashbridge, B.; Moore, M.A.; Arbelaez, V.; Heldrich, J.; Ding, B.S.;
719 Rafii, S.; Staudt, M.R.; Crystal, R.G. Generation of a human airway epithelium derived
720 basal cell line with multipotent differentiation capacity. *Respir Res* **2013**, *14*, 135,
721 doi:10.1186/1465-9921-14-135.

722 32. Lazarowitz, S.G.; Choppin, P.W. Enhancement of the infectivity of influenza A and B
723 viruses by proteolytic cleavage of the hemagglutinin polypeptide. *Virology* **1975**, *68*,
724 440-454, doi:10.1016/0042-6822(75)90285-8.
725 33. Klenk, H.D.; Rott, R.; Orlich, M.; Blödorn, J. Activation of influenza A viruses by trypsin
726 treatment. *Virology* **1975**, *68*, 426-439, doi:10.1016/0042-6822(75)90284-6.
727 34. Hamilton, B.S.; Whittaker, G.R. Cleavage activation of human-adapted influenza virus
728 subtypes by kallikrein-related peptidases 5 and 12. *J Biol Chem* **2013**, *288*, 17399-17407,
729 doi:10.1074/jbc.M112.440362.
730 35. Golden, J.W.; Schiff, L.A. Neutrophil elastase, an acid-independent serine protease,
731 facilitates reovirus uncoating and infection in U937 promonocyte cells. *Virol J* **2005**, *2*,
732 48, doi:10.1186/1743-422X-2-48.
733 36. Molla, A.; Hellen, C.U.; Wimmer, E. Inhibition of proteolytic activity of poliovirus and
734 rhinovirus 2A proteinases by elastase-specific inhibitors. *J Virol* **1993**, *67*, 4688-4695,
735 doi:10.1128/JVI.67.8.4688-4695.1993.
736 37. Gong, D.; Farley, K.; White, M.; Hartshorn, K.L.; Benarafa, C.; Remold-O'Donnell, E.
737 Critical role of serpinB1 in regulating inflammatory responses in pulmonary influenza
738 infection. *J Infect Dis* **2011**, *204*, 592-600, doi:10.1093/infdis/jir352.
739 38. Huang, I.C.; Bosch, B.J.; Li, F.; Li, W.; Lee, K.H.; Ghiran, S.; Vasilieva, N.; Dermody, T.S.;
740 Harrison, S.C.; Dormitzer, P.R.; et al. SARS coronavirus, but not human coronavirus NL63,
741 utilizes cathepsin L to infect ACE2-expressing cells. *J Biol Chem* **2006**, *281*, 3198-3203,
742 doi:10.1074/jbc.M508381200.
743 39. Bollavaram, K.; Leeman, T.H.; Lee, M.W.; Kulkarni, A.; Upshaw, S.G.; Yang, J.; Song, H.;
744 Platt, M.O. Multiple sites on SARS-CoV-2 spike protein are susceptible to proteolysis by
745 cathepsins B, K, L, S, and V. *Protein Sci* **2021**, *30*, 1131-1143, doi:10.1002/pro.4073.
746 40. Avril, L.E.; Di Martino-Ferrer, M.; Pignede, G.; Séman, M.; Gauthier, F. Identification of
747 the U-937 membrane-associated proteinase interacting with the V3 loop of HIV-1 gp120
748 as cathepsin G. *FEBS Lett* **1994**, *345*, 81-86, doi:10.1016/0014-5793(94)00410-2.
749 41. Zhirnov, O.P.; Golyando, P.B.; Ovcharenko, A.V. Replication of influenza B virus in
750 chicken embryos is suppressed by exogenous aprotinin. *Arch Virol* **1994**, *135*, 209-216,
751 doi:10.1007/BF01309780.
752 42. Visseren, F.L.; Bouwman, J.J.; Bouter, K.P.; Diepersloot, R.J.; de Groot, P.H.; Erkelens,
753 D.W. Procoagulant activity of endothelial cells after infection with respiratory viruses.
754 *Thromb Haemost* **2000**, *84*, 319-324.
755 43. Du, L.; Kao, R.Y.; Zhou, Y.; He, Y.; Zhao, G.; Wong, C.; Jiang, S.; Yuen, K.Y.; Jin, D.Y.;
756 Zheng, B.J. Cleavage of spike protein of SARS coronavirus by protease factor Xa is
757 associated with viral infectivity. *Biochem Biophys Res Commun* **2007**, *359*, 174-179,
758 doi:10.1016/j.bbrc.2007.05.092.
759 44. Sutherland, M.R.; Friedman, H.M.; Pryzdial, E.L. Thrombin enhances herpes simplex
760 virus infection of cells involving protease-activated receptor 1. *J Thromb Haemost* **2007**,
761 *5*, 1055-1061, doi:10.1111/j.1538-7836.2007.02441.x.
762 45. Knabb, M.T.; Danielsen, C.A.; McShane-Kay, K.; Mbuy, G.K.; Woodruff, R.I. Herpes
763 simplex virus-type 2 infectivity and agents that block gap junctional intercellular
764 communication. *Virus Res* **2007**, *124*, 212-219, doi:10.1016/j.virusres.2006.11.006.

765 46. Stokol, T.; Yeo, W.M.; Burnett, D.; DeAngelis, N.; Huang, T.; Osterrieder, N.; Catalfamo, J.
766 Equid herpesvirus type 1 activates platelets. *PLoS One* **2015**, *10*, e0122640,
767 doi:10.1371/journal.pone.0122640.
768 47. Stieneke-Gröber, A.; Vey, M.; Angliker, H.; Shaw, E.; Thomas, G.; Roberts, C.; Klenk, H.D.;
769 Garten, W. Influenza virus hemagglutinin with multibasic cleavage site is activated by
770 furin, a subtilisin-like endoprotease. *EMBO J* **1992**, *11*, 2407-2414, doi:10.1002/j.1460-
771 2075.1992.tb05305.x.
772 48. Esumi, M.; Ishibashi, M.; Yamaguchi, H.; Nakajima, S.; Tai, Y.; Kikuta, S.; Sugitani, M.;
773 Takayama, T.; Tahara, M.; Takeda, M.; et al. Transmembrane serine protease TMPRSS2
774 activates hepatitis C virus infection. *Hepatology* **2015**, *61*, 437-446,
775 doi:10.1002/hep.27426.
776 49. Tang, T.; Jaimes, J.A.; Bidon, M.K.; Straus, M.R.; Daniel, S.; Whittaker, G.R. Proteolytic
777 Activation of SARS-CoV-2 Spike at the S1/S2 Boundary: Potential Role of Proteases
778 beyond Furin. *ACS Infect Dis* **2021**, *7*, 264-272, doi:10.1021/acsinfecdis.0c00701.
779 50. Baron, J.; Tarnow, C.; Mayoli-Nüssle, D.; Schilling, E.; Meyer, D.; Hammami, M.;
780 Schwalm, F.; Steinmetzer, T.; Guan, Y.; Garten, W.; et al. Matriptase, HAT, and TMPRSS2
781 activate the hemagglutinin of H9N2 influenza A viruses. *J Virol* **2013**, *87*, 1811-1820,
782 doi:10.1128/JVI.02320-12.
783 51. Zhang, X.; Fugère, M.; Day, R.; Kielian, M. Furin processing and proteolytic activation of
784 Semliki Forest virus. *J Virol* **2003**, *77*, 2981-2989, doi:10.1128/jvi.77.5.2981-2989.2003.
785 52. Brennan, S.O.; Nakayama, K. Furin has the proalbumin substrate specificity and serpin
786 inhibitory properties of an in situ hepatic convertase. *FEBS Lett* **1994**, *338*, 147-151,
787 doi:10.1016/0014-5793(94)80353-6.
788 53. Follis, K.E.; York, J.; Nunberg, J.H. Furin cleavage of the SARS coronavirus spike
789 glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology* **2006**,
790 *350*, 358-369, doi:10.1016/j.virol.2006.02.003.
791 54. Goto, H.; Kawaoka, Y. A novel mechanism for the acquisition of virulence by a human
792 influenza A virus. *Proc Natl Acad Sci U S A* **1998**, *95*, 10224-10228,
793 doi:10.1073/pnas.95.17.10224.
794 55. Moon, K.B.; Turner, P.C.; Moyer, R.W. SPI-1-dependent host range of rabbitpox virus
795 and complex formation with cathepsin G is associated with serpin motifs. *J Virol* **1999**,
796 *73*, 8999-9010, doi:10.1128/JVI.73.11.8999-9010.1999.
797 56. Choi, H.K.; Tong, L.; Minor, W.; Dumas, P.; Boege, U.; Rossmann, M.G.; Wengler, G.
798 Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and
799 the organization of the virion. *Nature* **1991**, *354*, 37-43, doi:10.1038/354037a0.
800 57. Nava, V.E.; Rosen, A.; Veliuona, M.A.; Clem, R.J.; Levine, B.; Hardwick, J.M. Sindbis virus
801 induces apoptosis through a caspase-dependent, CrmA-sensitive pathway. *J Virol* **1998**,
802 *72*, 452-459, doi:10.1128/JVI.72.1.452-459.1998.
803 58. Sugrue, R.J.; Brown, C.; Brown, G.; Aitken, J.; McL Rixon, H.W. Furin cleavage of the
804 respiratory syncytial virus fusion protein is not a requirement for its transport to the
805 surface of virus-infected cells. *J Gen Virol* **2001**, *82*, 1375-1386, doi:10.1099/0022-1317-
806 82-6-1375.
807 59. Ortmann, D.; Ohuchi, M.; Angliker, H.; Shaw, E.; Garten, W.; Klenk, H.D. Proteolytic
808 cleavage of wild type and mutants of the F protein of human parainfluenza virus type 3

809 by two subtilisin-like endoproteases, furin and Kex2. *J Virol* **1994**, *68*, 2772-2776,
810 doi:10.1128/JVI.68.4.2772-2776.1994.
811 60. Richt, J.A.; Fürbringer, T.; Koch, A.; Pfeuffer, I.; Herden, C.; Bause-Niedrig, I.; Garten, W.
812 Processing of the Borna disease virus glycoprotein gp94 by the subtilisin-like
813 endoprotease furin. *J Virol* **1998**, *72*, 4528-4533, doi:10.1128/JVI.72.5.4528-4533.1998.
814 61. Hahm, B.; Han, D.S.; Back, S.H.; Song, O.K.; Cho, M.J.; Kim, C.J.; Shimotohno, K.; Jang,
815 S.K. NS3-4A of hepatitis C virus is a chymotrypsin-like protease. *J Virol* **1995**, *69*, 2534-
816 2539, doi:10.1128/JVI.69.4.2534-2539.1995.
817 62. Gettins, P.G.; Olson, S.T. Inhibitory serpins. New insights into their folding,
818 polymerization, regulation and clearance. *Biochem J* **2016**, *473*, 2273-2293,
819 doi:10.1042/BCJ20160014.
820 63. Ardini-Poleske, M.E.; Clark, R.F.; Ansong, C.; Carson, J.P.; Corley, R.A.; Deutsch, G.H.;
821 Hagood, J.S.; Kaminski, N.; Mariani, T.J.; Potter, S.S.; et al. LungMAP: The Molecular
822 Atlas of Lung Development Program. *Am J Physiol Lung Cell Mol Physiol* **2017**, *313*, L733-
823 L740, doi:10.1152/ajplung.00139.2017.
824 64. Guex, N.; Peitsch, M.C.; Schwede, T. Automated comparative protein structure modeling
825 with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **2009**,
826 *30 Suppl 1*, S162-173, doi:10.1002/elps.200900140.
827 65. Bienert, S.; Waterhouse, A.; de Beer, T.A.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede,
828 T. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res* **2017**,
829 *45*, D313-D319, doi:10.1093/nar/gkw1132.
830 66. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer,
831 F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling
832 of protein structures and complexes. *Nucleic Acids Res* **2018**, *46*, W296-W303,
833 doi:10.1093/nar/gky427.
834 67. Consortium, U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*
835 **2021**, *49*, D480-D489, doi:10.1093/nar/gkaa1100.
836 68. de Vries, S.J.; van Dijk, M.; Bonvin, A.M. The HADDOCK web server for data-driven
837 biomolecular docking. *Nat Protoc* **2010**, *5*, 883-897, doi:10.1038/nprot.2010.32.
838 69. Honorato, R.V.; Koukos, P.I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.;
839 Giachetti, A.; Rosato, A.; Bonvin, A.M.J.J. Structural Biology in the Clouds: The WeNMR-
840 EOSC Ecosystem. *Front Mol Biosci* **2021**, *8*, 729513, doi:10.3389/fmolb.2021.729513.
841 70. Siebenmorgen, T.; Zacharias, M. Evaluation of Predicted Protein-Protein Complexes by
842 Binding Free Energy Simulations. *J Chem Theory Comput* **2019**, *15*, 2071-2086,
843 doi:10.1021/acs.jctc.8b01022.
844 71. Marijanovic, E.M.; Fodor, J.; Riley, B.T.; Porebski, B.T.; Costa, M.G.S.; Kass, I.; Hoke, D.E.;
845 McGowan, S.; Buckle, A.M. Reactive centre loop dynamics and serpin specificity. *Sci Rep*
846 **2019**, *9*, 3870, doi:10.1038/s41598-019-40432-w.
847 72. van der Burgh, R.; Meeldijk, J.; Jongeneel, L.; Frenkel, J.; Bovenschen, N.; van Gijn, M.;
848 Boes, M. Reduced serpinB9-mediated caspase-1 inhibition can contribute to
849 autoinflammatory disease. *Oncotarget* **2016**, *7*, 19265-19271,
850 doi:10.18632/oncotarget.8086.
851 73. Cooley, J.; Takayama, T.K.; Shapiro, S.D.; Schechter, N.M.; Remold-O'Donnell, E. The
852 serpin MNEI inhibits elastase-like and chymotrypsin-like serine proteases through

853 efficient reactions at two active sites. *Biochemistry* **2001**, *40*, 15762-15770,
854 doi:10.1021/bi0113925.
855 74. Silverman, G.A.; Whisstock, J.C.; Askew, D.J.; Pak, S.C.; Luke, C.J.; Cataltepe, S.; Irving,
856 J.A.; Bird, P.I. Human clade B serpins (ov-serpins) belong to a cohort of evolutionarily
857 dispersed intracellular proteinase inhibitor clades that protect cells from promiscuous
858 proteolysis. *Cell Mol Life Sci* **2004**, *61*, 301-325, doi:10.1007/s00018-003-3240-3.
859 75. Lappin, D.; Whaley, K. Regulation of C1-inhibitor synthesis by interferons and other
860 agents. *Behring Inst Mitt* **1989**, 180-192.
861 76. Schroder, W.A.; Hirata, T.D.; Le, T.T.; Gardner, J.; Boyle, G.M.; Ellis, J.; Nakayama, E.;
862 Pathirana, D.; Nakaya, H.I.; Suhrbier, A. SerpinB2 inhibits migration and promotes a
863 resolution phase signature in large peritoneal macrophages. *Sci Rep* **2019**, *9*, 12421,
864 doi:10.1038/s41598-019-48741-w.
865 77. Meng, X.Y.; Zhang, H.X.; Mezei, M.; Cui, M. Molecular docking: a powerful approach for
866 structure-based drug discovery. *Curr Comput Aided Drug Des* **2011**, *7*, 146-157,
867 doi:10.2174/157340911795677602.
868 78. Jin, L.; Abrahams, J.P.; Skinner, R.; Petitou, M.; Pike, R.N.; Carrell, R.W. The anticoagulant
869 activation of antithrombin by heparin. *Proc Natl Acad Sci U S A* **1997**, *94*, 14683-14688,
870 doi:10.1073/pnas.94.26.14683.
871 79. Leblond, J.; Laprise, M.H.; Gaudreau, S.; Grondin, F.; Kisiel, W.; Dubois, C.M. The serpin
872 proteinase inhibitor 8: an endogenous furin inhibitor released from human platelets.
873 *Thromb Haemost* **2006**, *95*, 243-252, doi:10.1160/TH05-08-0561.
874 80. Izaguirre, G.; Arciniega, M.; Quezada, A.G. Specific and Selective Inhibitors of Proprotein
875 Convertases Engineered by Transferring Serpin B8 Reactive-Site and Exosite
876 Determinants of Reactivity to the Serpin α1PDX. *Biochemistry* **2019**, *58*, 1679-1688,
877 doi:10.1021/acs.biochem.8b01295.
878 81. Zhao, M.M.; Yang, W.L.; Yang, F.Y.; Zhang, L.; Huang, W.J.; Hou, W.; Fan, C.F.; Jin, R.H.;
879 Feng, Y.M.; Wang, Y.C.; et al. Cathepsin L plays a key role in SARS-CoV-2 infection in
880 humans and humanized mice and is a promising target for new drug development.
881 *Signal Transduct Target Ther* **2021**, *6*, 134, doi:10.1038/s41392-021-00558-8.
882 82. Pišlar, A.; Mitrović, A.; Sabotič, J.; Pečar Fonović, U.; Perišić Nanut, M.; Jakoš, T.; Senjor,
883 E.; Kos, J. The role of cysteine peptidases in coronavirus cell entry and replication: The
884 therapeutic potential of cathepsin inhibitors. *PLoS Pathog* **2020**, *16*, e1009013,
885 doi:10.1371/journal.ppat.1009013.
886 83. Peacock, T.P.; Goldhill, D.H.; Zhou, J.; Baillon, L.; Frise, R.; Swann, O.C.; Kugathasan, R.;
887 Penn, R.; Brown, J.C.; Sanchez-David, R.Y.; et al. The furin cleavage site in the SARS-CoV-
888 2 spike protein is required for transmission in ferrets. *Nat Microbiol* **2021**, *6*, 899-909,
889 doi:10.1038/s41564-021-00908-w.
890 84. Gamblin, S.J.; Haire, L.F.; Russell, R.J.; Stevens, D.J.; Xiao, B.; Ha, Y.; Vasisht, N.;
891 Steinhauer, D.A.; Daniels, R.S.; Elliot, A.; et al. The structure and receptor binding
892 properties of the 1918 influenza hemagglutinin. *Science* **2004**, *303*, 1838-1842,
893 doi:10.1126/science.1093155.
894 85. Yin, H.S.; Paterson, R.G.; Wen, X.; Lamb, R.A.; Jardetzky, T.S. Structure of the uncleaved
895 ectodomain of the paramyxovirus (hPIV3) fusion protein. *Proc Natl Acad Sci U S A* **2005**,
896 *102*, 9288-9293, doi:10.1073/pnas.0503989102.

86. Porotto, M.; Murrell, M.; Greengard, O.; Moscona, A. Triggering of human parainfluenza virus 3 fusion protein (F) by the hemagglutinin-neuraminidase (HN) protein: an HN mutation diminishes the rate of F activation and fusion. *J Virol* **2003**, *77*, 3647-3654, doi:10.1128/jvi.77.6.3647-3654.2003.
87. Du, X.; Tang, H.; Gao, L.; Wu, Z.; Meng, F.; Yan, R.; Qiao, S.; An, J.; Wang, C.; Qin, F.X. Omicron adopts a different strategy from Delta and other variants to adapt to host. *Signal Transduct Target Ther* **2022**, *7*, 45, doi:10.1038/s41392-022-00903-5.
88. pymol.bib<source-app name="BibDesk" version="1.5.2">BibDesk</source-app><rec-number>1</rec-number><ref-type>34</ref-type>Schrodinger, LLC<secondary-authors><subsidiary-authors></subsidiary-authors></secondary-authors>2015<pub-dates>November</pub-dates>AxPyMOL<pdf-urls><related-urls></related-urls></pdf-urls>unpublished19/08/2010<modified-date>22/12/2015</modified-date>
pymol.bib<source-app name="BibDesk" version="1.5.2">BibDesk</source-app><rec-number>2</rec-number><ref-type>34</ref-type>Schrodinger,, LLC<secondary-authors><subsidiary-authors></subsidiary-authors></secondary-authors>2015<pub-dates>November</pub-dates>JyMOL<pdf-urls><related-urls></related-urls></pdf-urls>unpublished19/08/2010<modified-date>22/12/2015</modified-date>
pymol.bib<source-app name="BibDesk" version="1.5.2">BibDesk</source-app><rec-number>3</rec-number><ref-type>34</ref-type>Schrodinger,, LLC<secondary-authors><subsidiary-authors></subsidiary-authors></secondary-authors>2015<pub-dates>November</pub-dates>PyMOL<pdf-urls><related-urls></related-urls></pdf-urls><research-notes>PyMOL The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC.</research-notes>unpublished19/08/2010<modified-date>22/12/2015</modified-date>.
89. Team, R.C. A Language and Environment for Statistical Computing. **2014**.