ACS Infectious Diseases

pubs.acs.org/journal/aidcbc

Article

The SARS-CoV-2 SSHHPS Recognized by the Papain-like Protease

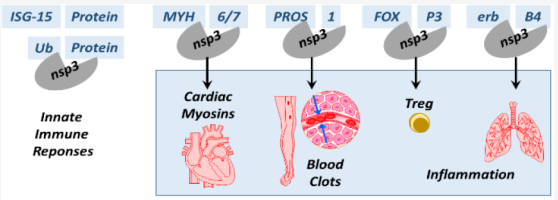
Nathanael D. Reynolds, Nathalie M. Aceves, Jinny L. Liu, Jaimee R. Compton, Dagmar H. Leary, Brendan T. Freitas, Scott D. Pegan, Katarina Z. Doctor, Fred Y. Wu, Xin Hu, and Patricia M. Legler*

Cite This: https://doi.org/10.1021/acsinfecdis.0c00866

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Viral proteases are highly specific and recognize conserved cleavage site sequences of ~6−8 amino acids. Short stretches of homologous host−pathogen sequences (SSHHPS) can be found spanning the viral protease cleavage sites. We hypothesized that these sequences corresponded to specific host protein targets since >40 host proteins have been shown to be cleaved by Group IV viral proteases and one Group VI viral protease. Using PHI-BLAST and the viral protease cleavage site sequences, we searched the human proteome for host targets and analyzed the hit results. Although the polyprotein and host proteins related to the suppression of the innate immune responses may be the primary targets of these viral proteases, we identified other cleavable host proteins. These proteins appear to be related to the virus-induced phenotype associated with Group IV viruses, suggesting that information about viral pathogenesis may be extractable directly from the viral genome sequence. Here we identify sequences cleaved by the SARS-CoV-2 papain-like protease (PLpro) *in vitro* within human MYH7 and MYH6 (two cardiac myosins linked to several cardiomyopathies), FOXP3 (an X-linked T_{reg} cell transcription factor), ErbB4 (HER4), and vitamin-K-dependent plasma protein S (PROS1), an anticoagulation protein that prevents blood clots. Zinc inhibited the cleavage of these host sequences *in vitro*. Other patterns emerged from multispecies sequence alignments of the cleavage sites, which may have implications for the selection of animal models and zoonosis. SSHHPS/nsP is an example of a sequence-specific post-translational silencing mechanism.



Cleavage of Host Protein Sequences by the SARS-CoV-2 PLpro & Post-translational Silencing

KEYWORDS: *SSHHPS, FOXP3, MYH6, MYH7, PROS1, ErbB4(HER4)*

The cleavage of host proteins by the viral proteases of Group IV (+)ssRNA viruses has been known for almost four decades.[1] One of the earliest examples of host protein cleavage was that of histone H3 by the foot-and-mouth viral protease.[2−4] Histone H3 may not be a surprising target, as interferon (IFN) is known to stimulate the transcription of hundreds of genes in order to attain an antiviral state within the cell and around it. At present there are >40 examples of host proteins that have been shown to be cleaved by viral proteases (Table 1), suggesting that the cleavage site sequences (~6−8 amino acids) within the viral polyprotein are host or hostlike sequences depending on their sequence identity.[1] The protease cleavage site sequences in the structural proteins recognized by the host proteases furin and TMPRSS2 also could be considered to be host or host-like protein sequences encoded by the viral genome. The vast majority of the viral proteases that cleave host proteins belong to Group IV (only one virus, HIV, was from Group VI[5]). While the primary targets of viral proteases are likely the viral polyprotein and the host's innate immune response proteins that antagonize viral replication, there appear to be off-target cleavages of other host proteins. Some of these host proteins have no obvious role in the generation of immune responses; instead they appear to be related to the virus-induced symptoms or causes of death, or both.[1,6] For Cas enzymes, on- and off-target cleavages are well-known;[7,8] but for the viral proteases, they are less well characterized. In our earlier publications,[1,9] we identified a host protein called TRIM14 using BLAST that contained a Group IV viral protease cleavage site sequence and found that the protein was cleavable both *in vitro* and in virus-infected cells; its cleavage could be inhibited by treating the cells with a cysteine protease inhibitor.[1] TRIM14 has been proposed to be a component of the mitochondrial antiviral signalosome (MAVS).[10] Here we refine and apply our methods to identify the potential targets of viral proteases from two Group IV viruses with distinct phenotypes, Zika and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). We hypothesize that there may be a correlation between the protease's targets and the virus-induced phenotypes.

Many Group IV viral enzymes are produced compactly on a single nonstructural (ns) polyprotein, which is proteolyzed to release the active enzymes. Each ns protein (nsP) can contain

Special Issue: Antiviral Therapeutics

Received: December 11, 2020

Table 1. Host Protein Sequences That Have Been Reported to Be Cleaved by Group IV and VI Viral Proteases

Virus	Family	Viral Protease	Host Protein Substrate	Cleavage site in Host protein	Reference
			Group IV		
Poliovirus	*Picornaviridae*	3Cpro	**RIG-1**	LKKFPQ↓GQKGKV	Barral, et al.[85]
			TATA-binding Protein	QGLASPQ↓GAMTPG	Das, et al.[86]
			TATA-binding Protein	AAAVQQ↓STSQQA	Kundu, et al. [87]
			Poly(A)-binding protein (PABP)	VHVQ↓GQ	Kuyumcu-Martinez, et al.[88]
			eIF5B	VMEQ↓G	de Breyne, et al.[89]
		2Apro	**TATA-binding Protein**	MMPY↓GTGLTP	Yalamanchili, et al.[90]
Rhinovirus type 1a		3Cpro	**NF-κB**	LLNQ↓GIP	Neznanov, et al.[91]
			eIF5B	VMEQ↓G	de Breyne, et al.[89]
Echovirus type 1			**NF-κB**	LLNQ↓GIP	Neznanov, et al.[91]
Coxsackie B virus		2Apro	**Dystrophin**	KIHTT↓GFKDQ, PGLTTI↓GASPT	Badorff, et al. [6]
		3Cpro	**eIF5B**	VMEQ↓G	de Breyne, et al. [89]
Foot and Mouth disease Virus (FMDV)		3Cpro	**NEMO**	LALPSQ↓RRSPPE	Wang, et al. [92]
			eIF4A	TNVRAE↓VQKLQM	Li, et al.[3]
			Histone H3	PRKQL↓ATKAA	Falk, et al.[2]
		Leader protease	**eIF4G**	SFANLG↓RTTLST	Foeger, et al.[93]
Hepatitis A Virus		3Cpro	**NEMO**	PVLKAQ↓ADIYK	Wang, et al.[92]
		3ABC	**MAVS**	LASQ↓VDSP	Yang, et al.[94]
		3CD	**TRIF**	DWSQ↓GCSL	Qu, et al.[95]
			TRIF	IREQSQ↓HLDG	Qu, et al.[95]
Enterovirus 68 (EV68)		3Cpro	**TRIF**	GSAGPQ↓SL, PAAFPQ↓SL	Xiang, Z, et al.[96]
			IRF-7	AGLQ↓A, AVQQ↓S	Xiang, Z, et al.
Enterovirus 71 (EV71)		3Cpro	**TRIF**	AGPQ↓S	Lei, X., et al. [97]
			IRF-7	AVQQ↓S	Lei, X., et al. [98]
Norovirus MD145-12	*Calciviridae*	3CLpro	**PABP**	VHVQ↓GQN, AIPQ↓TQE	Kuyumcu-Martinez, et al. [99]
Feline Calicivirus (Vesivirus)			**PABP**	WTAQ↓GAR	Kuyumcu-Martinez, et al. [99]
Dengue 1 and 2	*Flaviviridae*	ns2B/ns3	**STING (MITA)**	VRACLGCPLRR↓GALLLLSIY	Yu, et al.[100], Stabell, et al.[76]
West Nile Virus			**STING (MITA)**		Ding, et al [101]
Japanese Encephalitis Virus			**STING (MITA)**		Ding, et al [101]
Zika Virus		ns2B/ns3	**STING (MITA)**	HIHSRYR↓GSYWRTVR	Ding, et al [101]
			SFRP1	SEGGRR↓GAALGVLL	Morazanni, et al [1]
			Gs alpha	QVAGRR↓GAALPCSL	
			NT5M	VPAGRR↓GAAGGLGL	
			Forkhead box protein G1	KLAFKR↓GARLTSTG	
			Septin-2	SERLKR↓GGRK	Li ,H, et al.[38]
Hepatitis C Virus		ns3/4A	**MAVS**	EREVPC↓HRPS	Meylan, et al.[102], Bellecave, et al.[103]
			TRIF	PPPPPSSTPC↓SAHLTPSSLE	Li, K., et al. [104]
VEEV	*Togaviridae*	nsP2	**TRIM14**	DCFATGRHYWEVDVQEAGA↓GWWVGA	Morazanni, et al [1]
SARS-CoV-2	*Coronaviridae*	nsP3 Plpro	MYH6, MYH7, **FOXP3,** PROS1, ErbB4(HER4)	IALKGG↓K, RDLRGG↓A, LLIALRGG↓KI, LTEILNGG↓V	Reynolds, et al (this paper)
			ULK1	RKLSLGGG↓RPY	Mohamud, et al [105]
			IRF-3	VLSCLGGG↓LALWR	Moustaqil, et al [106]
			Ubiquitin, ISG15	RLRGG↓	Shin, et al.[107], Freitas, et al.[22]
		3CL Main Protease	**NLRP12**	GKLFQ↓GR, SVVLQ↓AN	Moustaqil, et al [106]
			TAB1	ASLQ↓S, LTLQ↓S	
			CTBP1, **IRAK1**	GTRVQ↓SVEQI, STLQ↓AGLAA	Miczi, M, et al.[27]
			Group VI		
HIV	*Retroviridae*	HIV-1 protease	**eIF4GI**	KIIA↓TVL, TVL↓MTE, FSA↓LQQAVP	Ventoso, et al.[5]

multiple enzymes and domains strung together, and thus, a localization signal in one domain can enable a linked enzyme to enter into an unexpected location such as the nucleus.[11] In coronaviruses, the papain-like protease (PLpro) of nsP3 is

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

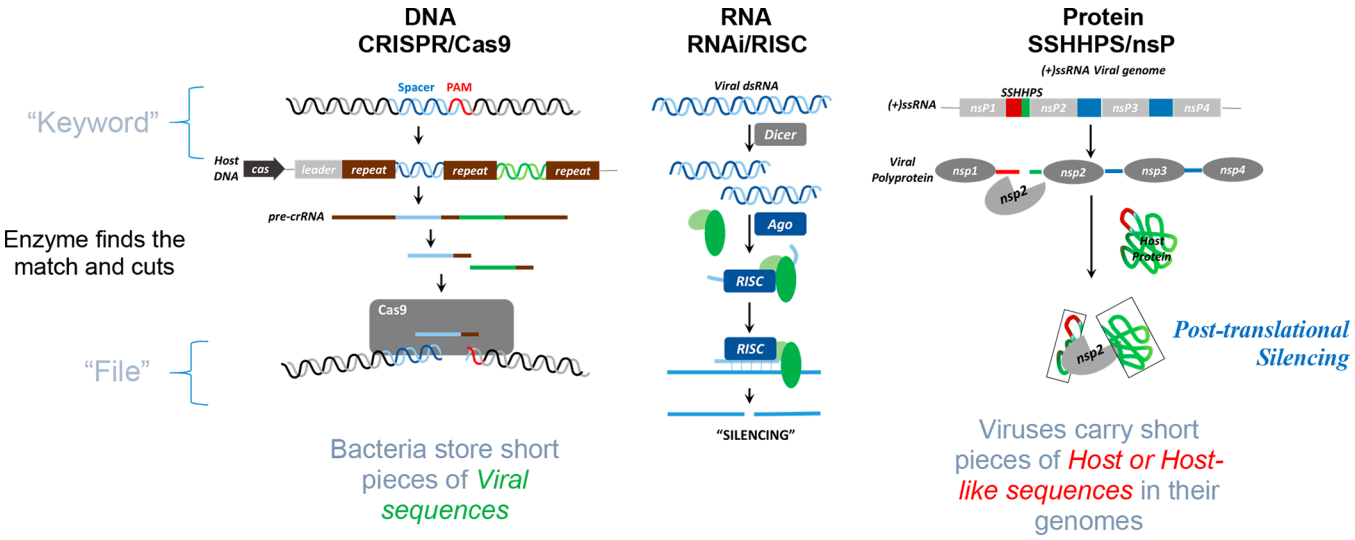


Figure 1. Sequence-directed targeted cleavages. Silencing can occur at the level of DNA, RNA, or protein. Each of these silencing mechanisms utilizes short sequences to direct cleavage of a larger target(s). In CRISPR/Cas systems, short pieces of viral sequences are embedded in the host genome, whereas in SSHHPS/nsP, short pieces of host or hostlike sequences can be found in the viral protease cleavage sites. These antagonistic mechanisms can be seen as "search" programs, in which the short sequence functions as a keyword and the target is a file (containing the search term) that is cut. Each has an enzyme that searches for the match between the short sequence and the sequence in the file and then cleaves the target. SSHHPS/nsP is an example of a sequence-specific post-translational silencing mechanism that is transient.[1] The CRISPR/Cas9 and RNAi figures have been adapted from refs 107 and 108.

anchored to the endoplasmic reticulum (ER) membrane on the cytoplasmic face, and cleavage of host proteins may occur co-translationally or post-translationally; nsP3 is also involved in inducing double membrane vesicles.[12]

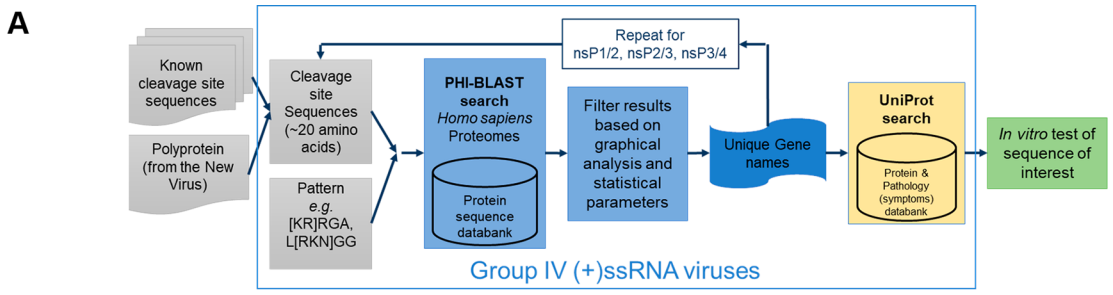
Short stretches of homologous host-pathogen sequences (SSHHPS) can be found at the viral protease cleavage sites.[1] Group IV viruses appear to mount a targeted attack on the host by destroying strategic proteins, enabling them to transiently impair the host's immune responses to enable viral replication (Table 1). Notably, some of the host proteins are targeted by more than one viral protease, suggesting that they are not randomly selected sequences. Several of the proteins listed are components of the MAVS (Figure S1) and Toll-like receptor pathways; cleavage of various components may short-circuit these cascades at different points. CRISPR/Cas systems embed short viral sequences in a host's genome to protect a host from a virus, whereas SSHHPS/nsP systems appear to embed host or host-like sequences in the viral polyprotein in order to protect the virus from the host. Both are defense systems and can be viewed as sequence-specific silencing mechanisms[1] (Figure 1).

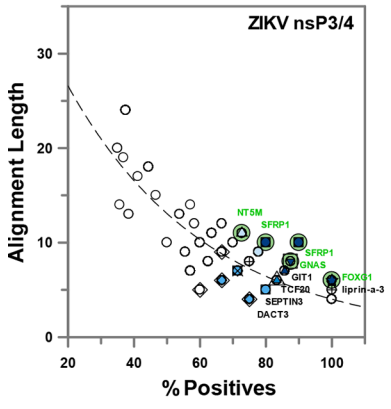
Sequence-specific events imply information storage. Information can be stored directly in a linear sequence such as a CRISPR spacer sequence but also in the complementary subsites of a viral protease. The incorporation of host sequences into viral genomes has been noted for several (+)ssRNA viruses (Group IV) and retroviruses (Group VI) and may be the result of RNA recombination events[13−15] (Figure S2). In coronaviruses and picornaviruses, recombination hot spots are non-randomly distributed in the RNA genome. Polymerase "pausing" during the synthesis of viral RNA may be due to stable RNA secondary structures. These pauses may allow templates to dissociate and rebind the polymerase (template switching).[16] In poliovirus, viral RNA recombination events occur concomitantly with RNA synthesis as the polymerase jumps from one template to another.[17] There is also evidence of a polymerase-independent mechanism of RNA recombination.[18,19] Coronaviruses and picornaviruses, in particular, are known to undergo both homologous and non-homologous recombination events at relatively high frequencies in animals and in cultured cells; however, the recombination mechanisms are not well-delineated.[15] As noted earlier, the production of innate immune response mRNAs can be induced by IFN, and these transcripts may be present during infection, potentiating their incorporation. An example of the incorporation of the host's cellular RNA into Group IV viruses[13,15] is that of the pestivirus bovine viral diarrhea virus (Flaviviridae). This virus occasionally incorporates cellular genes and fragments into its genome, and the ubiquitin (Ub) sequence is frequently incorporated by this virus.[15,18] With regard to the viral protease cleavage sites (*i.e.*, SSHHPS), the viral RNA sequences that encode them share sequence identity with several bat genomic sequences, from which they may have originated, while the resulting protein sequences are similar to those found in the proteins of bats, humans, and other species. The incorporation of new host sequences or the mutation and evolution of these sequences may enable these viruses to propagate in more than one species.

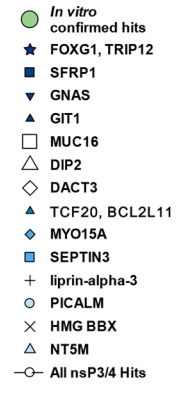
SARS-CoV-2, the betacoronavirus that causes COVID-19, shares the majority of its genome with the SARS-CoV virus that emerged in 2002.[20] Both have two cysteine proteases: the PLpro and a main protease (3CL).[12,21] In addition to its role in polyprotein processing, the PLpro can cleave Ub-like interferon-stimulated gene 15 (ISG15) and Ub. These two proteins function in signaling and mounting host innate immune responses against viral infections.[12,22,23] By acting as an antagonist toward ISG15- and Ub-dependent cellular processes, the PLpro can block cytokine production, which in turn down-regulates the interferon response and consequently minimizes the immune response to enable replication.[12,22,24,25]

The (+)ssRNA viruses are known to elicit diverse yet relatively consistent sets of symptoms in humans. The symptoms of many of these viral infections have remained

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX







C

% Pos	Alignment Length	Gene Name	Acc. No.	Bit Score	Role or Relationship
90%	**10**	**SFRP1**	**NP_003003.3**	12.4	**Brain Dev., Apoptosis**
88%	**8**	**GSA**	**CAA39484.1**	9.3	Eye, Bone Dev.
73%	**11**	**NT5M**	**NP_064586.1**	**7.5**	**Smith-Magenis Syndr.**
86%	7	DUSP28	NP_001028747.1	6.7	Cell growth
86%	**7**	**GIT1**	**AAH48196.1**	6.7	**Microcephaly-like**
86%	7	GXYLT2	NP_001073862.1	6.7	Glycosyltransf, Notch
86%	7	SLC47A2	NP_001243592.1	6.7	Brain, BBB, renal
100%	**6**	**FOXG1**	**CAA52240.1**	5	**Microcephaly**
83%	6	NDST2	NP_003626.1	5	Heparin biosyn.
88%	8	MUC16	AAK74120.3	4.7	Mucin
100%	**6**	TRIP12	NP_004229.1	4.7	ASD, Brain Dev.
83%	6	SRSF1	BAD92795.1	4.7	Innate Immun. Related
83%	6	INSR	NP_000199.2	4.7	Diabetes
83%	6	DIP2A	EAX09275.1	4.7	Axon Patterning
83%	**6**	TCF20	NP_001365347.1	4.7	ASD
83%	**6**	BCL2L11	XP_005263607.1	4.7	Apoptosis
83%	6	CGREF1	EAX00636.1	4.7	Cell growth regulator
83%	6	KCTD17	XP_011528676.1	4.7	Movement disorders
78%	**9**	PICALM	AAB07762.1	4.3	Aβ clearance
80%	5	SEPTIN3	EAW60478.1	4.7	Cytoskel. GTPase, Brain

D

% Positives	Alignment Length	Predicted Host Protein Targets									
35%	20	CTTN									
37%	19	PLEKHO1									
38%	24	WDR46									
38%	13	BAC85254.1									
44%	18	CTNNA3	BCAP29								
47%	15	EAW73648.1									
50%	10	MYO1B									
56%	9	ACE1	LONRF1								
57%	7	TRIP12	ULF								
57%	14	Fab fragment									
58%	12	HOXA10									
60%	5	DACT3									
60%	10	PLD2	Tcell receptor								
63%	8	PPCS	APBA3								
64%	11	PER3	SUMF1	HAUS5	KIAA1549	TNFRSF14					
67%	6	DACT3	CRLF1	EMILIN1	EAW75658.1	EAW60506.1	XP_024309045.1				
67%	9	EAW81861.1	KIF26A	BAZ1A							
67%	9	BAC85254.1	RAP1GAP	MYO15A							
67%	12	ACCN4	ASIC4	LONRF1	AGN04211.1	BCL2L11	DACT3				
70%	10	BTNL9	NPAS1	ERCC6	PGBD3	EAW93099.1	CDKN2A	EAW81860.1			
71%	7	CRLF1	RHOT	TDRD6	FLJ00328						
71%	7	DPY19L2	TK1	EAW81930.1	EAW63421.1	EAW99549.1	GSC2	TRABD2A	LOC107984862	TMEM121	TSPAN10	EAX02218.1
73%	11	NT5M	PHF1	PP1597	SEL1L3	C20orf204	LOC114841035	INSM2	LINC02538	ALYREF	SEPT3	TIKI1
75%	4	FOXG1	ABCA7	RHOT2	FAM169B	BBX	MAP3K1	MYO15A	TRABD2A	FAM169B		
75%	8	Liprin-alpha-3	MAXlikeX									
78%	9	RNApol2	CALM	ACE1	ACAP1	DACT3	MAX-likeX	PER3	RAP1GAP	ARMC6	MYO15A	
80%	5	CGREF1	SEPT3	ARMC6	AAG48268.1	BAD93077.1	ANKLE1	DHX30	MTR3	EAW92235.1		
83%	6	NDST2	CRTAP	7tm_1								
83%	6	ZNF354B	SP5									
86%	7	SNRPA1	SLC47A2	BCL2L11	KCTD17	CGREF1	PCCA	PABPN1L	TF20			
86%	7	DUSP28	GIT1	SRSF1	7tm_1	DACT3	DIP2	INSR				
88%	8	GSA	Mucin-16	FLJ31196	PLECKHA4							
90%	10	SFRP1	GXYLT2	ACAP1								
100%	4	TRIP12	EAW94576.1									
100%	5	BAZ1A	Liprin-alpha-3									
100%	6	TRIP12	FOXG1									

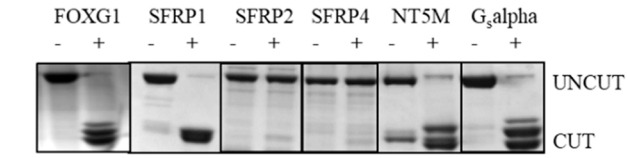


Figure 2. SSHHP sequences for the Zika virus ns2B/3 protease. (A) Diagram showing the workflow for identifying cleavable sequences in host proteins and their significance. (B) Plot of alignment length *vs* % positives from the PHI-BLAST HitTable.csv output file. The proteins that were tested for cleavage *in vitro* are shown in green. (C) The HitTable.csv file was re-sorted by percent positives, alignment lengths, and bit score. The protein's function, phenotype of knockout mice, or hereditary human diseases associated with mutations in the gene are shown in the last column and were taken from PubMed, UniProt, GeneCards, or MGI.[109] (D) Distribution of the proteins in the graph plotted in (B). Each point in the graph represents one or more proteins. (E) *In vitro* protease assays using CFP-YFP substrates from Morazzani *et al.*[1]

unchanged; for others significant increases in virulence (*e.g.*, MERS *vs* SARS) have occurred. Much like computer programs, viruses carry a code and output a characteristic set of symptoms; the genome (the input) determines the output. Constants and variables are defined in every program, and this information must be stored, but not necessarily in a linear sequence or with high accuracy.

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

In 2004, Kiemer *et al.*[26] created a neural network for prediction of SARS 3CL main protease cleavage sites in host proteins. Their top predicted cleavage sites were in the cystic fibrosis transmembrane conductance regulator (CFTR), transcription factors (CREB-RP and OCT-1), and components of the Ub pathway. While their predictions were unconfirmed biochemically, they also hypothesized that these sequences could be related to viral pathogenesis. More recently, Miczi *et al.*[27] demonstrated that the SARS-CoV-2 3CL main protease could cleave a sequence in human C-terminal binding protein 1 (CTBP1) and interleukin-1 receptor associated kinase 1 (IRAK1), a substrate predicted by Kiemer *et al.* Earlier, Blom *et al.*[28] developed a neural network for picornaviral proteases, in which one of the predicted host protein substrates was dystrophin. Badorff *et al.*[6,29] demonstrated that Coxsackievirus protease 2A could cut dystrophin both *in vitro* and in cells and proposed that cleavage of dystrophin initiated a cascade of events that led to dilated cardiomyopathy.[6,29] This was the first host protein identified by bioinformatic methods that was shown to be cut in cells.[6] A cleavage-resistant dystrophin knock-in mouse was later shown to have a decrease in sarcolemmal disruption and cardiac virus titer following Coxsackievirus infection,[30] linking the 2A proteolytic cleavage of dystrophin to the development of cardiomyopathy.

Stabell *et al.*[31] also acknowledged relationships between the cleavability of a host's STING protein and the appearance of symptoms in dengue-2-infected nonhuman primates (NHPs).[31] Lack of cleavability trended with a lack of symptoms and low titers. Notably, single amino acid substitutions in STING proteins from other species were capable of making these proteins cleavable *in vitro*.

Neural network algorithms are trained with sets of cleavable and uncleavable sequences and are therefore specific to the protease, which limits their use. The low throughput of these programs also makes analysis of an entire proteome time-consuming. Since the programming and training of a neural network is not easily performed by the average bench scientist, we developed methods that utilize web applications such as PHI-BLAST and spreadsheet and graphing programs like Microsoft Excel. Visualization of trends and relationship(s) between the proteins containing the cleavage sites was also difficult to deduce from lists of largely untested hits. Here we describe some interesting trends that can be extracted from plots of the PHI-BLAST hits and *in vitro* data. We also have added the related disease states to our tables and show that each protease produces a list with a general theme that appears related to viral pathogenesis. Knowing a virus' host targets can yield information about its strategy of attack. These host or hostlike sequences may provide some of the first sequence-based correlates for the prediction of symptoms from viral genome sequence.

In this paper, we describe our approach for searching, plotting, and identifying host targets of a viral protease. This method generally can be applied to other Group IV viruses for hypothesis generation and subsequent testing in cell-based assays. Our results show that the PLpro of SARS-CoV-2 is capable of cleaving sequences in MYH6, MYH7, FOXP3, ErbB4(HER4), and plasma Protein S (PROS1) *in vitro* and that these cleavages may be related to viral pathogenesis.

RESULTS

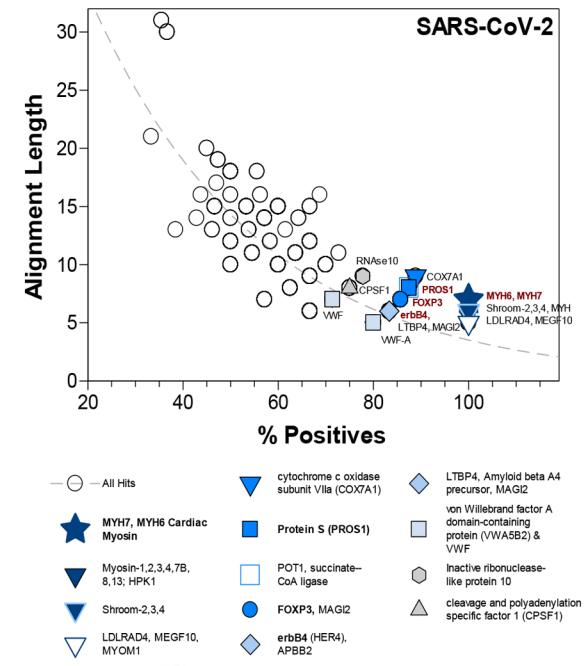
In our earlier work,[1,9,32] we selected potential host protein substrates using the viral protease cleavage sites (~20- to 25-amino acids) and BLAST to search the human proteome. The likelihood of cleavage was assessed using empirical rules: (1) for cysteine proteases the P2 residue is typically highly conserved,[33] while for serine proteases P1 is typically conserved (Schechter and Berger nomenclature is used throughout), and (2) the sequence identity or similarity to a conserved cleavage site motif sequence spans more than four amino acids.[34] BLAST searches produced ~1500 hits for three cleavage sites. Many of the alignments contained spurious matches to the short sequences, and only a few captured the conserved cleavage site motif. As a consequence, only a few hits were identified (~1 to 4 hits per virus). Using just the above criteria, we were still able to identify cleavable sequences with a high success rate using our *in vitro* and cell-based assays; however, the search process was slow and inefficient, and proteins with cleavable sequences were likely missed. We refined the search method and improved the identification of hits by plotting the search results and visualizing the predicted hits so trends could be identified.

Zika Viral ns2B/3 Protease Cleavage Sites in the Human Proteome. First, we used PHI-BLAST to search for Zika viral (ZIKV) protease cleavage sites within the human proteome (Figure 2A). We used a short-sequence-length search with a pattern deduced from three polyprotein cleavage sites (e.g., [KR]RG) and searched the human proteome for host targets. The graphical plots of the parameters generated by -BLAST, such as percent identity, alignment length, mismatches, gap opens, query start, query end, sequence start, sequence end, expectation value (E-value), bit score, and percent positives, were further analyzed to reduce the ~500 hits to those that are most significant (~20). These parameters were output by BLAST in a file called "HitTable.csv". The graph in Figure 2B shows that the alignment lengths relate to the % positive matches. The % positives value allows conservative substitutions to have positive scores in the scoring matrix. Short alignment lengths tended to have a higher percentage of matches. Plotting the previously analyzed ZIKV protease cleavage sites (Figure 2E)[1] onto the graph showed a trend. From the graph the probability of cleavage was determined from the sequence's similarity to a known cleavage site. Notably, the ~500 hits appear as less than 50 points on these graphs, and each point contains several different targets (Figure 2D). Redundant hits (*e.g.*, isoforms) and hits with similar scores could be effectively binned together. Fewer redundant hits were found when the UniProtKB/Swiss-Prot database was selected in BLAST.

The number of unique proteins and their E-values and bit scores varied within each point in the graph. The E-value is the number of expected hits of similar score (quality) that could be found just by chance, and therefore, the smaller the E-value, the more significant the score and the alignment. Although the common cutoff number for E-values is <0.02, when the queried sequence is short (<50, and in our case ≤25), the E-values are much higher. The bit score in the BLAST hit results is derived from the raw alignment score taking the statistical properties of the scoring system into account. The bit scores do not depend on the database size; and since they are normalized with respect to the scoring system, they can be used to compare alignment scores from different searches.[35] The higher the bit score, the higher is the sequence similarity. The hits could be sorted by gap opens, bit score, % positives, and alignment length in Excel (Figure 2C). The rightmost point in the graph

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

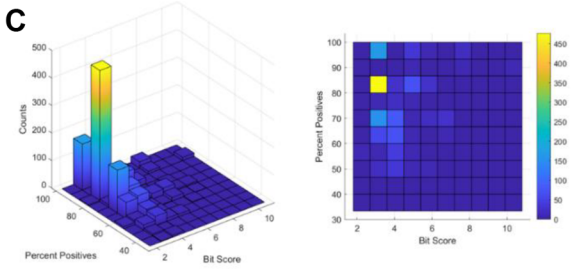
A



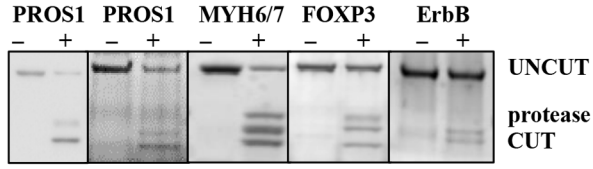
B

Accession number	% Pos	No. Align	Bit Score	Label	Role or Relationship
NP_001301006.1	**88**	**8**	**10.8**	**PROS1**	**Blood clots**
NP_000248.2	**100**	**7**	7.8	**MYH7**	**Cardiomyopathies**
NP_002462.2	**100**	**7**	**7.8**	**MYH6**	**Cardiomyopathies**
NP_056265.2	88	8	7.5	POT1	Telomere stability
NP_001107849.1	**86**	**7**	**6**	**FOXP3**	**Treg Cells**
NP_065768.2	100	6	6	SHRM2,3,4	Cell shape
NP_001267468.1	100	6	6	RNF19A	E3 Ub ligase
NP_001036009.1	83	6	6	LTBP4	Urban Rifkin Disease
AAB21058.1	89	9	5.7	COX7A1	mitochondria
EAW92976.1	83	6	5.7	APBB2	Amyloid binding protein
NP_006276.2	86	6	5.3	TESK1	Testis-specific Kinase
NP_036433.2	83	6	5.3	MAGI2	Nephrotic Syndrome
AAH22499.1	83	6	5.3	EFNB3	Nipah virus receptor
XP_016869336.1	83	6	5.3	CSPP1	Cytokinesis
NP_004446.2	83	6	5.3	EXTL1	Heparan biosyn.
NP_056025.2	83	6	5.3	MTCL1	Microtubule Crosslink.
XP_024305638.1	86	7	5	LCTL	Lactase-like, eye
NP_005954.3	100	6	5	MYH1	Skeletal myosin
NP_001093582.1	100	6	5	MYH2	Skeletal myosin
NP_002461.2	100	6	5	MYH3	Skeletal myosin
NP_060003.2	100	6	5	MYH4	Skeletal myosin
NP_002463.2	100	6	5	MYH8	Skeletal myosin
NP_003793.2	100	6	5	MYH13	Fast Twitch Myosin
NP_001609.2	100	6	5	PARP1	ADP-ribo., Inflam, IFNγ
XP_016883475.1	83	6	5	MYH7B	Slow ATPase Myosin
NP_003841.1	88	8	4.7	SUCLA2	Succinate CoA Ligase
XP_016874971.1	86	7	4.7	TPCN1	Calcium Channel
NP_055418.2	83	6	4	CDR2L	Cerebellar degen.
NP_001186788.1	100	5	4	ANGPT1	Vascular growth factor
XP_016859069.1	**83**	**6**	**3.4**	**ERBB4**	**ALS, BPD**
NP_003794.3	100	5	3.1	MYOM1	Muscle, sarcomere
NP_001243474.1	100	5	2.8	MEGF10	EMARDD

C



D



E

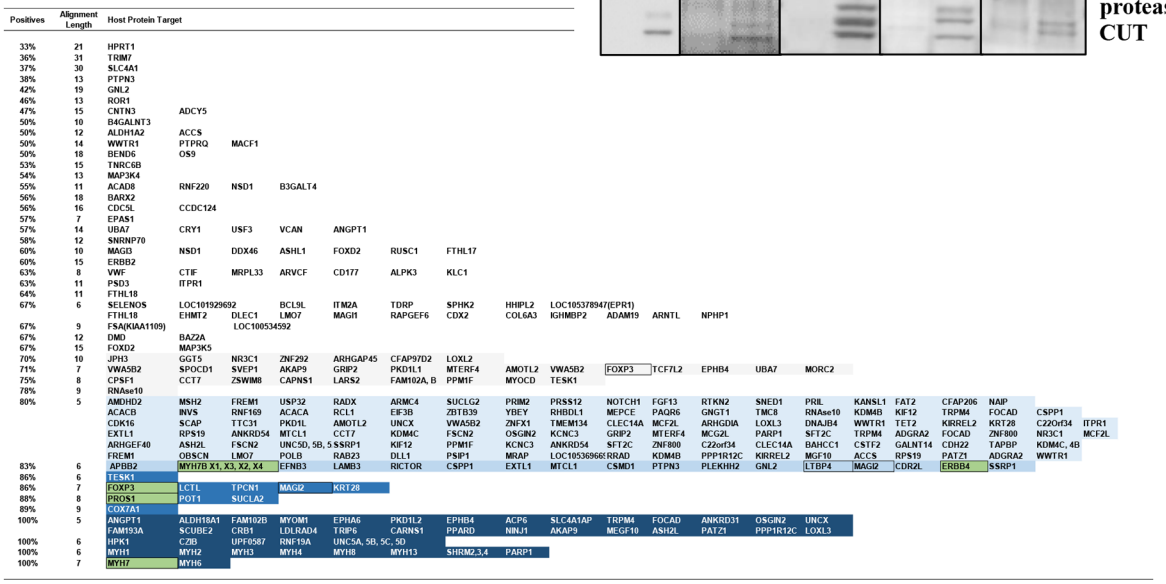


Figure 3. Analysis of the COVID-19 SARS-CoV-2 SSHHP sequences. (A) Hits from three PHI-BLAST searches. The shades of blue correspond to the distribution in (E). (B) Partial list of predicted hits sorted by their bit scores, alignment lengths, and % positives. The cleavage sequences tested *in vitro* are highlighted in green. (C) Three-dimensional graph showing the relationships between alignment length, % positives, and bit score. The histogram shows the counts of the PHI-BLAST hit results compiled from the cleavage sites. The most populated point was 80% positives over five amino acids. The hits at this point have low bit scores. (D) Summary of the *in vitro* cleavage assays. From left to right: CFP−PROS1−YFP immunoblot, CFP−PROS1−YFP, CFP−MYH6/7−YFP, CFP−FOXP3−YFP, and CFP−ErbB4−YFP SDS-PAGE gels. All of the substrates showed cleavage, and the preferred substrate was MYH6/7. (E) Distribution of the proteins in the graph in (A), colored in shades of blue to show where they can be found in the graph in (A).

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

Table 2. Predicted Host Protein Substrates of SARS-CoV-2 PLpro from PHI-BLAST Searches and Related Disease States or *In Vivo* Functions (from PubMed, UniProt and GeneCards)

Alignment Length	% Positives	SARS-CoV-2 PLpro SGVTRELMRELNGG↓AVTRYV NMMVTNNTFTLKGG↓APTKVT VVNVVTTKIALKGG↓KIVNNW	*In vitro* Cleavage	Host Protein Containing the SSHHPS	Related Disease States or Role in Cell
8	88	LLIALRGG|KIEVQL	Y	**Vitamin K-dep. Protein S (PROS1)**	Anticoagulant plasma protein, deficiencies result in **blood clots**, Thrombophilia
7	100	EAEQIALKGG|KKQLQK	Y	**MYH7, cardiac myosin**	**Cardiomyopathy,** familial hypertrophic 1 (CMH1), ventricular hypertrophy; Myopathy, myosin storage, autosomal dominant (MSMA); Scapuloperoneal myopathy MYH7-related (SPMM); Cardiomyopathy, dilated 1S (CMD1S); Myopathy, distal, 1 (MPD1); Left ventricular non-compaction 5 (LVNC5)
7	100	EAEQIALKGG|KKQLQK	Y	**MYH6, cardiac myosin**	**Cardiomyopathy**, familial hypertrophic 14 (CMH14), Atrial septal defect 3 (ASD3), Cardiomyopathy, dilated 1EE (CMD1EE), Sick sinus syndrome 3 (SSS3)
8	88	PLNQLKGG|TIVNVY	ND	Protection of telomeres 1 (POT1)	Melanoma, cutaneous malignant, Glioma 9 (GLM9)
7	86	FQGRDLRGG|AHASSS	Y	**FOXP3**	T_{reg} Cell Transcription Factor crucial for the development and inhibitory function of **T_{reg} cells**, can repress the expression of IL2 and IFN-gamma; Immunodeficiency polyendocrinopathy, enteropathy, X-linked syndrome (Neonatal onset insulin-dep. Diabetes, infections, diarrhea, thrombocytopenia, anemia, and eczema, usually lethal in infancy).
6	100	PWGFTLKGG|REHGEP	ND	Shroom2	Ocular albinism
6	100	PWGFTLKGG|LEHGEP	ND	Shroom3	Neutrophil Actin Dysfunction
6	100	PWGFTLKGG|LEHCEP	ND	Shroom4	May regulate cellular and cytoskeletal architecture by modulating the spatial distribution of myosin II, Cognitive disabilities, X-linked
6	100	RKSRELNGG|VDGIAS	ND	RNF19A E3 Ub-ligase	Ubiquitinates pathogenic SOD1 variants
6	83	GYCFRELRGG|ECASPL	ND	LTBP4	Urban-Rifkin-Davis syndrome (URDS), Duchenne muscular dystrophy (DMD),
5	80	RASGGLGLRGG|APARRS	ND	CSPP1	Joubert syndrome 21 (JBTS21)
5	80	FLPEAHPLRGG|APGQLR	ND	EXTL1	Involved in heparan sulfate biosynthesis
5	80	LHDAGLRGG|APLPGP	ND	MTCL1	Epithelial cell polarization
6	83	PEGQLGFELKGG|AENGQF	ND	**MAGI2**	**Nephrotic syndrome 15 (NPHS15)**
5	100	LTCQCLNGG|ACNTLD	ND	**MEGF10**	Myopathy, early-onset, areflexia, **respiratory distress,** and dysphagia (EMARDD). Death from respiratory failure in childhood occurs in about half of patients, ventilator-dependent
8	75	ESLMSSLNGG|SVPSEL	ND	Myocardin (MYOCD)	Transcription factor that functions primarily in cardiac muscle and smooth muscle, plays a role in cardiogenesis, urinary bladder development, and myogenesis
8	63	ACTIQLRGG|QIMTLK	ND	von Willebrand factor (VWF)	Bleeding disorders, deficiencies lead to impaired platelet aggregation, Von Willebrand disease 1 & 2
7	71	NEQLALRGG|AETTAD	ND	von Willebrand factor A domain containing 5B2 (VWA5B2)	Unknown function
6	100	AEQLALKGG|KKQIQ	ND	Myosin-1, -2, -3, -4, -8, -13	Myopathy, proximal, and ophthalmoplegia (MYPOP), Arthrogryposis, distal, 2A (DA2A), Carney complex, Arthrogryposis, distal, 7 (DA7)
7	86	MPQLNGG|GGDDLG	ND	Transcription factor 7-like 2 isoform 2 (TCF7L2)	Wnt signaling, human variants have been associated with Diabetes
9	78	IACELKGG|KCHKSS	ND	Inactive RNAse10	Hypotrichosis (fragile hairs that break easily, leaving short, sparse scalp hairs)
6	100	VAAVALKGG|YNELCD	ND	MAP4K1, Hematopoietic progenitor kinase 1 (HPK1)	HPK1 negatively regulates TCR signaling and T cell-mediated immune responses to dampen the T-cell response
6	83	NLTEILNGG|VYVDQN	Y	**ErbB4 (HER4)**	ErbB4 deletion leads to changes in lung function and structure similar to bronchopulmonary dysplasia, downregulation of surfactant protein D expression and an underlying **chronic inflammation in the lungs,** ALS
6	100	CTNPAPLNGG|AFCEGM	ND	Netrin receptor (UNC5A,B,C,D)	Axon guidance
6	83	SDFAVLNGG|KINSDI	ND	APBB2	Associated with Alzheimer's and insulin-producing beta cell function
6	83	ELGIALRGG|GAADPPF	ND	Ephrin-B3	Interacts with Nipah virus and Hendra virus glycoproteins
6	67	EGDGDLRGG|RILLGH	ND	Histone-lysine N-methyltransferase (EHMT2)	Primary hyperoxaluria; Carbohydrate metabolic disorder; Histone MTase that mono- and dimethylates 'Lys-9' of histone H3
5	100	GNLNGG|ATFGGG	ND	Set1/Ash2 Histone Mtase subunit (ASH2L)	Component of the Set1/Ash2 histone methyltransferase (HMT) complex, a complex that specifically methylates 'Lys-4' of histone H3, but not if the neighboring 'Lys-9' residue is already methylated
5	100	TKEVLLKGG|KREEE	ND	Angiopoietin 1	Angioedema
5	100	AALESLKGG|KKAK	ND	Myomesin-1	Hypertrophic Cardiomyopathy and Dilated Cardiomyopathy, Major component of the vertebrate myofibrillar M band. Binds myosin, titin, and light meromyosin.
5	100	REAAELNGG|LSRECS	ND	LDL receptor class A domain containing protein 4	Negative regulator of TGF-beta signaling
5	80	HPELLLRGG|AAAAGD	ND	FOXD2	Probable transcription factor involved in embryogenesis and somatogenesis.

corresponds to the longest sequence with 100% sequence similarity within the data set.

Unique hits within each point in the graph produced a skewed distribution (Figure 2D). The *in vitro* hits for ZIKV (NT4M, SFRP1, a $G_{s,alpha}$ protein, FOXG1) had bit scores greater than 3 and were mostly in the upper region of the distribution. Two proteins associated with microcephaly, FOXG1[36] and GIT1,[37] were found above the mode of the distribution. FOXG1 syndrome has been classified as an autism spectrum disorder (ASD). Transcription factor 20 (TCF20, UniProt Q9UGU0) and TRIP12 (UniProt Q13669) have been linked to ASDs and were in a similar region in the graph. In Figure 2B, TRIP12 overlapped with FOXG1 and TCF20 with bcl-2-like protein 11, a protein involved in apoptosis and anoikis. A general theme was noted, as several hits were related to brain development or had roles in neurons, while others were related to innate immune response. Interestingly, septin-2 recently was shown to be cleaved by the ZIKV protease in virus-infected cells; it is thought to be cut at R306 (SERLKR↓GGRKVENE) on the basis of mass spectrometry (MS) data.[38] Septin-2 was found in the BLAST search using the capsid/intracellular capsid (C/Ci) protein cleavage site sequence. Septin-3 was also among our ns3/4A predictions. The location of septin-2 on the graph is shown in Figure S3.

PLpro Cleavage Sites of SARS-CoV-2 within the Human Proteome. Next, the three PLpro cleavage sites of SARS-CoV-2 were predicted from a sequence alignment to the SARS and MERS CoV polyproteins using their known PLpro cleavage sites[12] (Figure 3). Since the CoV PLpro enzymes are protease−deubiquitinase (PRO-DUB) enzymes,[22,25,39] the sequences of Ub and ISG15 were included in the subsite tolerances of the PHI-BLAST pattern. The PHI-BLAST input consisted of three SARS-CoV-2 cleavage site sequences (14 residues N-terminal to the scissile bond and six residues C-terminal to the scissile bond) and the L[RKN]GG pattern.

The hits were plotted for all three cleavage sites and produced 290 unique hits (Figure 3A). The most populated point (80% positives, five-residue alignment) contained 115 different proteins. To estimate the noise, we plotted the immunoglobulins that came up as "hits". These 90 "hits" were distributed throughout the data: 39 of 90 were localized to one point at 80% positives and five aligned residues (*i.e.*, one mismatch), but fewer were within the 70−100th percentile region. The immunoglobulin hits were removed from the following analysis. A total of 70 different gene products were in the region above the most populated point. The plots were used for selection of host targets with the highest probability of cleavage based on their similarity to a known cleavable sequence found in the virus.

From our analyses, the host proteins with the highest sequence identity to a SARS-CoV-2 PLpro viral protease cleavage site were PROS1, FOXP3, MYH6, and MYH7. MYH6 and MYH7 are both cardiac myosins and contained a similar cleavage site sequence. There are nine sarcomeric

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

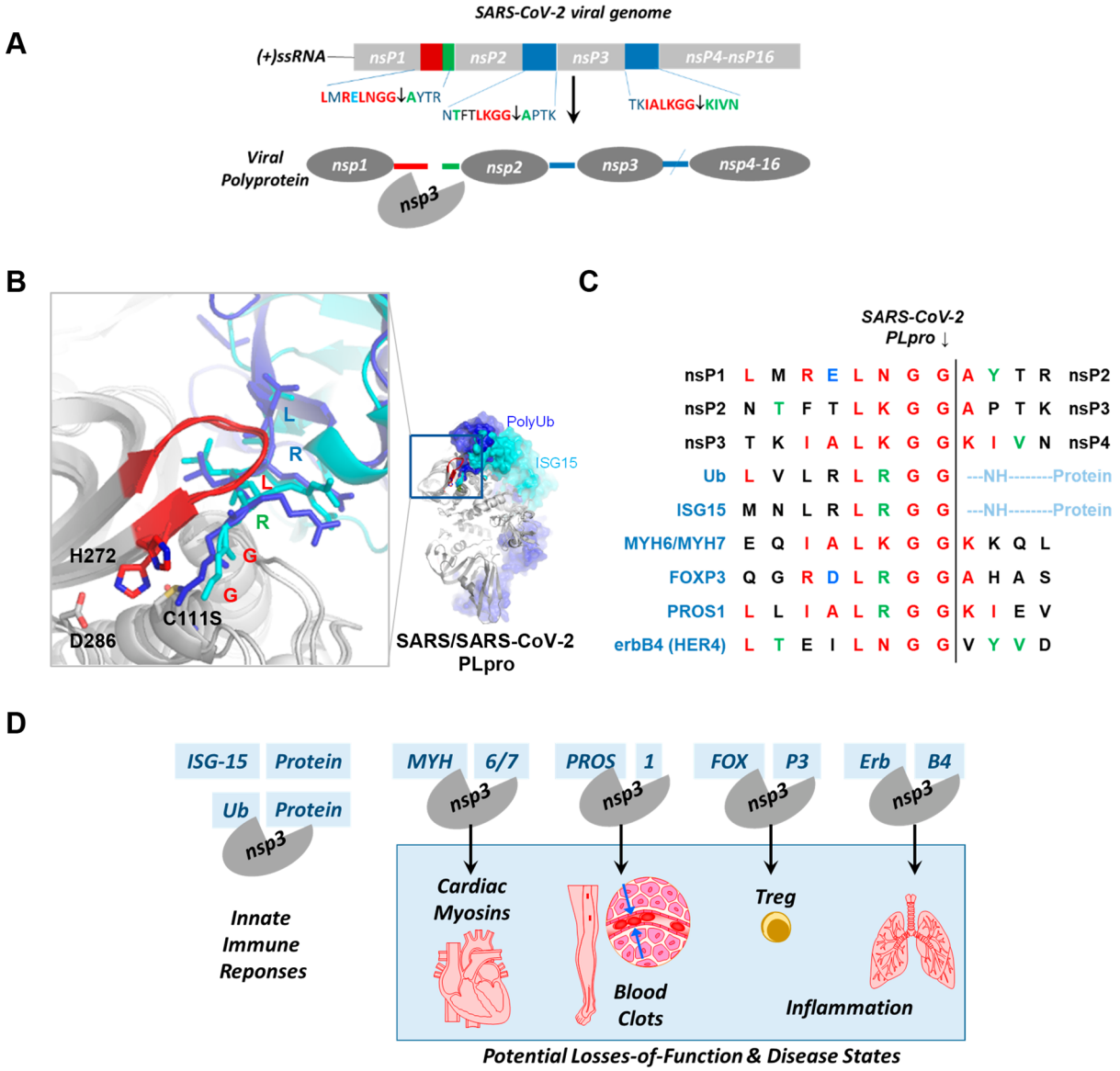


Figure 4. SSHHP sequences in SARS-CoV-2. (A) Three PLpro cleavage sites in the viral polyprotein are recognized by the SARS-CoV-2 PLpro. (B) The CoV PLpro enzymes also have deubquitinase and de-ISGylating activity. The SARS-CoV PLpro (white ribbon, PDB entry 5E6J[110]) bound to a diubiquitin activity-based probe (dark blue) is overlaid with the SARS-CoV-2 PLpro C111S variant (white ribbon, PDB entry 6YVA) bound to murine ISG-15 (cyan). In the inset is the structure of two molecules of the PLpro, showing the overlay of the Ub and ISG-15 proteins along the protease. The active-site residues of SARS-CoV-2 are labeled in black. (C) Sequences similar to the viral protease cleavage site sequences can be found in other host proteins, including ISG-15, Ub, MYH6, MYH7, FOXP3, PROS1, and ErbB4(HER4). (D) The host proteins containing the cleavage site sequences have loss-of-function phenotypes that are similar to the observed symptoms and causes of death of COVID-19, *i.e.*, the virus-induced phenotype.

muscle genes in mammals, three nonmuscle genes, and one smooth muscle gene (13 in total).[40] All three cardiac sarcomeric myosins (MYH6, MYH7, and MYH7B) and the six skeletal myosins (MYH1, MYH2, MYH3, MYH4, MYH8, and MYH13) carried a sequence similar to the nsP3/4 cleavage site and were predicted for SARS-CoV-2; these proteins were not highly predicted for MERS. Notably, the theme of the list changed. The SARS-CoV-2 hits (Table 2) were cardiovascular, inflammation, kidney, respiratory, or blood-related proteins (Figure 4). Sequences from PROS1, FOXP3, MYH6/7, and ErbB4 (HER4) were selected for *in vitro* assays. The ErbB4 sequence appears further down the list, while the point in the graph is near the FOXP3 point (Figure 3A). The method enabled analysis of the entire human proteome (~20 000 sequences) regardless of the protein's tissue- or cell-specific expression.

Confirmation of Protease Cleavage *In Vitro* and Significance of the Hits. The most significant substrate alignments that we identified from the PHI-BLAST search results were tested with *in vitro* protease assays using His-tagged cyan and yellow fluorescent protein constructs (CFP−cleavage site−YFP), recombinant proteins, human serum or bovine heart lysates, and mass spectrometry (Figures 5− 7; also see the Supporting Information). Several hits appeared to have relationships to the observed symptoms or causes of death for COVID-19 (Figure 4). Symptoms reported for COVID-19 include high fever, dry cough, fatigue, difficulty breathing, excessive inflammatory responses in the lungs,

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

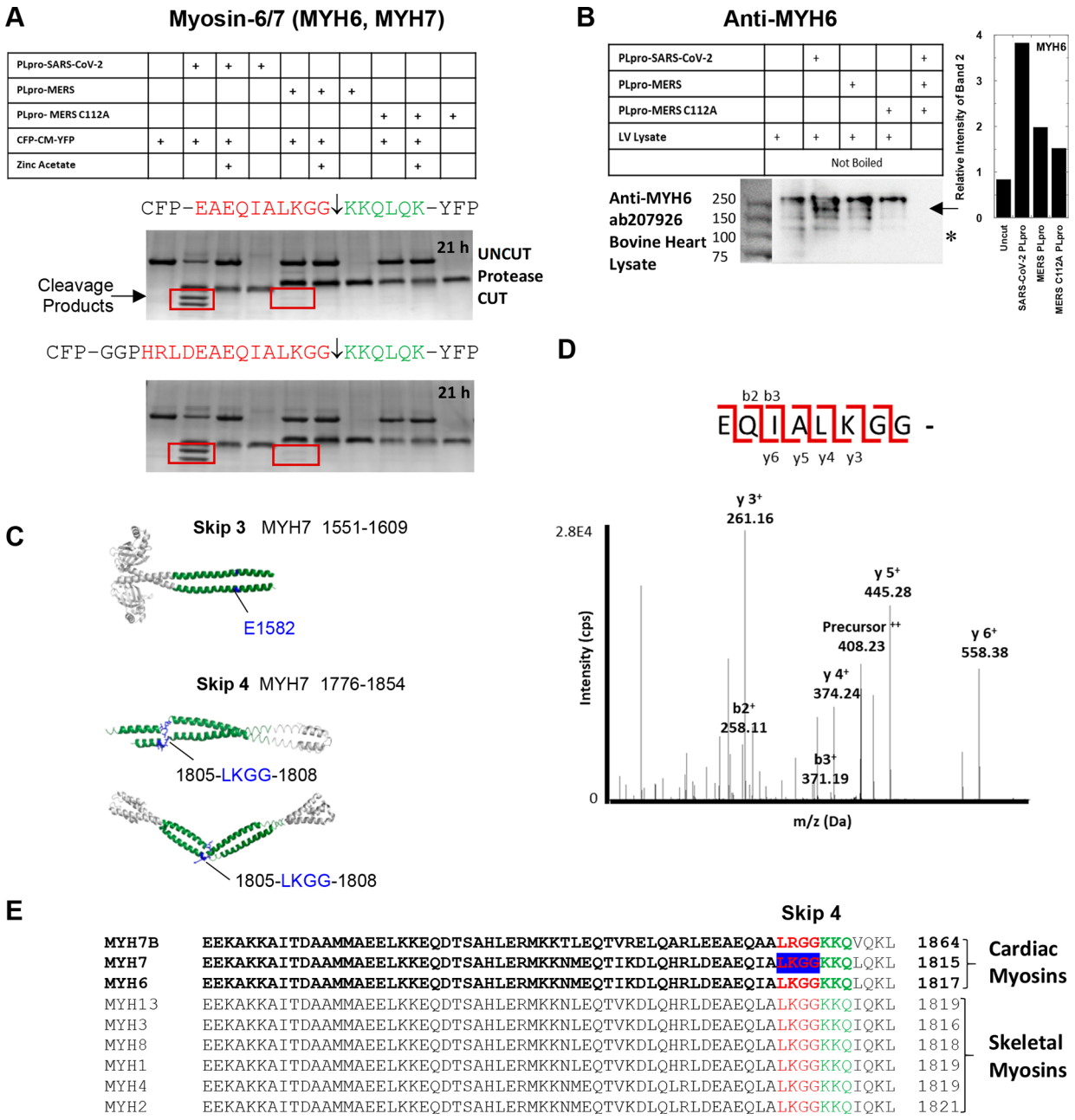


Figure 5. *In vitro* cleavage of myosin-6/7 (MYH6, MYH7) sequences by the SARS-CoV-2 PLpro and MERS PLpro. (A) Two CFP−MYH6/7−YFP substrates were tested; one containing 16-amino acids and a second containing 20-amino acids. The cleavage products are boxed in red. (B) Bovine heart lysates made from tissue taken from the bottom of the left ventricle were treated with the purified SARS-CoV-2 PLpro, MERS PLpro, or MERS C112A PLpro *in vitro*. The arrow points to the new band that appeared in immunoblots using an anti-MYH6 antibody (ab207926). MYH7 immunoblots can be found in the Supporting Information. (C) The LKGG cleavage site sequence in MYH6 and MYH7 corresponds to a helix-breaking sequence called a skip. Regions of myosin containing Skip 3 (PDB entry 4XA4) and Skip 4 (PDB entry 4XA6) are shown. The skip residues are shown in blue. (D) Mass spectra confirming the cleavage site in the CFP−MYH6/7−YFP substrate. (E) The LKGG or LRGG sequence can be found in the three cardiac myosins as well as the six skeletal myosins.

pneumonia, hypoxia, acute respiratory distress syndrome (ARDS), muscle and body aches, headache, loss of appetite, taste or smell, diarrhea,[41] blood clots,[42] heart damage,[43,44] and kidney damage.[45]

Cardiac Myosin Is Readily Cut by the SARS-CoV-2 PLpro and Slowly Cut by the MERS PLpro. MYH6 and MYH7 encode cardiac myosin heavy chains and are highly conserved. Myosins form the regular bipolar thick filaments of sarcomeres. The SARS-CoV-2 PLpro readily cut the cardiac myosin sequence EAEQIALKGG↓KKQLQK in the CFP/YFP substrate and produced heavy product bands (Figure 5A), while the MERS PLpro produced only weak bands after 24 h at room temperature (23 ± 5 °C). A longer substrate, CFP−HRLDEAEQIALKGG↓KKQLQK−YFP, was also tested and found to be cleavable. Bovine heart tissue lysates were prepared in MPER from the right atrium (RA), bottom of the left ventricle (LVB), and anterior face of the left ventricle (LVF). In immunoblots probed with an anti-MYH6 antibody (ab207926), LVB lysates treated with the CoV PLpro proteases produced a more intense band at the expected

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

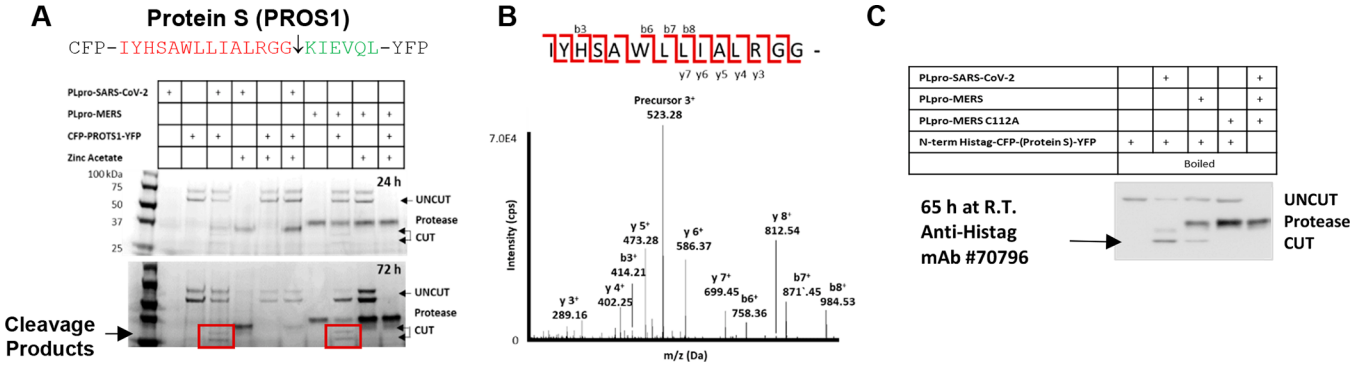


Figure 6. *In vitro* cleavage of the PROS1 sequence by the SARS-CoV-2 PLpro and MERS PLpro. (A) A 20-amino acid Protein S (PROS1) sequence embeded between cyan and yellow fluorescent protein was slowly cleaved by the SARS-CoV-2 PLpro and MERS PLpro enzymes over the course of 72 h at room temperature. The cleavage products were visible in SDS-PAGE gels. (B) Cleavage of the sequence by the MERS PLpro was confirmed by mass spectrometry. The annotated Lys-C fragment is shown. (C) Immunoblots using an anti-His-tag antibody to detect the uncut and cut His-tagged CFP−PROS1−YFP substrate. Cleavage of the His-tagged CFP−PROS1−YFP substrate by both the SARS-CoV-2 PLpro and MERS PLpro was detected after 65 h of incubation at room temperature. The arrow points to the cleavage products.

molecular weight (MW) (the MW of full-length MYH6 is 223.7 kDa, and that of the cleaved protein is 208.5 kDa) (Figure 5B); a blot with actin-loading controls is shown in Figure S4. The upper band was faint or not visible in untreated lysates and was intense in the PLpro protease-treated lysates (saturated signal). The lower band also intensified in CoV PLpro protease-treated LVF lysates when probed with the anti-MYH7 antibody (ab11083). The tissue from the bottom of the heart (LVB) produced three bands with the anti-MYH7 antibody, one of which disappeared after 5 days of cleavage by the CoV PLpro protease at room temperature (Figure S5). RA lysates showed little to no change in banding patterns when probed with the anti-MYH7 antibody (Figures S5) and a 4-fold increase in intensity of the lower band for the SARS-CoV-2 PLpro-treated lysate when probed with the anti-MYH6 antibody (Figure S4). MYH6 and MYH7 are 93% identical and 97.8% similar in sequence.

The PLpro cleavage site in MYH6/MYH7 is in the fourth skip region of the myosin (Figure 5C); the ^{1806}KGGKK^{1810} sequence is a helix-breaking sequence. Skip regions create discontinuities in the coiled-coil structure of myosin.[46] The fourth skip region creates a highly flexible molecular hinge and provides C-terminal rod flexibility, and it is functionally important in thick filament assembly in cardiomyocytes.

A distinct pattern of myofibrillar fragmentation has been reported in SARS-CoV-2-infected human cardiomyocytes from cell culture and autopsy[44] that is consistent with periodic proteolytic cleavage of the myofibrils into individual sarcomeric units. Perez-Bermejo *et al.* showed that the long striated myofibrils characteristic of healthy cardiomyocytes are diced up into small fragments after SARS-CoV-2 infection.[44] Cultured cardiomyocytes infected with SARS-CoV-2 display cytopathic effects and beating halts within 72 h postinfection.[47] Viral particles have been observed by electron microscopy, and the SARS-CoV-2 RNA was detected by RT-PCR in cardiac tissue taken from an infected female child post-mortem.[48] In a German study, virus in cardiac tissue was detected in 62% of the lethal COVID-19 cases.[43] Several human mutations associated with cardiomyopathies have been identified in MYH7 and MYH6. Heart damage has been noted in approximately 20−35% of the hospitalized COVID-19 patients,[43] but a connection between the cardiac myosins and the SARS-CoV-2 PLpro had not been previously established. Another predicted hit was myomesin (Table 2); dystrophin also contains an LKGG sequence. The results suggest that the PLpro may be involved in the virus-induced myofibril damage; however, additional cell-based experiments are still needed to confirm this.

A Sequence within Vitamin K-Dependent Plasma Protein S (PROS1) Is Cut by the SARS-CoV-2 PLpro and MERS PLpro. Protein S is an anticoagulation protein produced in the liver; human hepatocytes show evidence of SARS-CoV-2 infection with viral particles in the cytoplasm, mitochondrial swelling, and hepatic apoptosis.[49] Protein S acts as a cofactor to Protein C, a serine protease that inactivates Factors Va and VIIIa by cleavage to terminate the coagulation reaction. Hereditary deficiencies of Protein S in humans lead to pulmonary embolisms, recurrent venous blood clots, recurrent pregnancy clots, childhood stroke, and purpura fulminans (severe clotting throughout the body). Protein S is also vitamin K (phylloquinone)-dependent, and dietary deficiencies (*e.g.*, green leafy vegetables) can affect its activity. Dofferhoff *et al.* recently showed that extrahepatic vitamin K insufficiency was related to poor COVID-19 outcome.[50] Vitamin K activates hepatic coagulation factors and the extrahepatic anticoagulant Protein S.[50] Hypercoagulability of the blood of COVID-19 patients and multiple thrombi have been noted at autopsy.[51,52] Blood clots have been reported in cases of COVID-19 and SARS; however, data for the few cases of MERS are scarce.[53] Lemke *et al.*,[54] Zhang *et al.*,[59] and Stoichitoiu *et al.*[55] have postulated that depletion of Protein S may be related to the observed hypercoagulability of the blood. Lemke *et al.*[54] and Stoichitoiu *et al.*[55] also proposed that PROS1 deficiency may play a role in the immune hyper-reactions observed in humans with COVID-19 since PROS1 is involved in the activation of immunosuppressive TAM (Tyro3, Axl, and Mer) receptors that prevent hyperinflammation and lung injury. PROS1 is one of two activating ligands of the TAM family of receptor tyrosine kinases.[54]

Most PROS1 in the serum is bound to complement inhibitor C4BP, and only 30−40% of the PROS1 is in the free form.[56,57] Free Protein S can be measured using functional assays (activity), and total Protein S (free and bound) can be measured by antigenic (immunologic) assays. Stoichitoiu *et al.*,[55] Martin-Rojas *et al.*,[58] and Zhang *et al.*[59] used functional assays to measure free PROS1 and found that the activity of

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

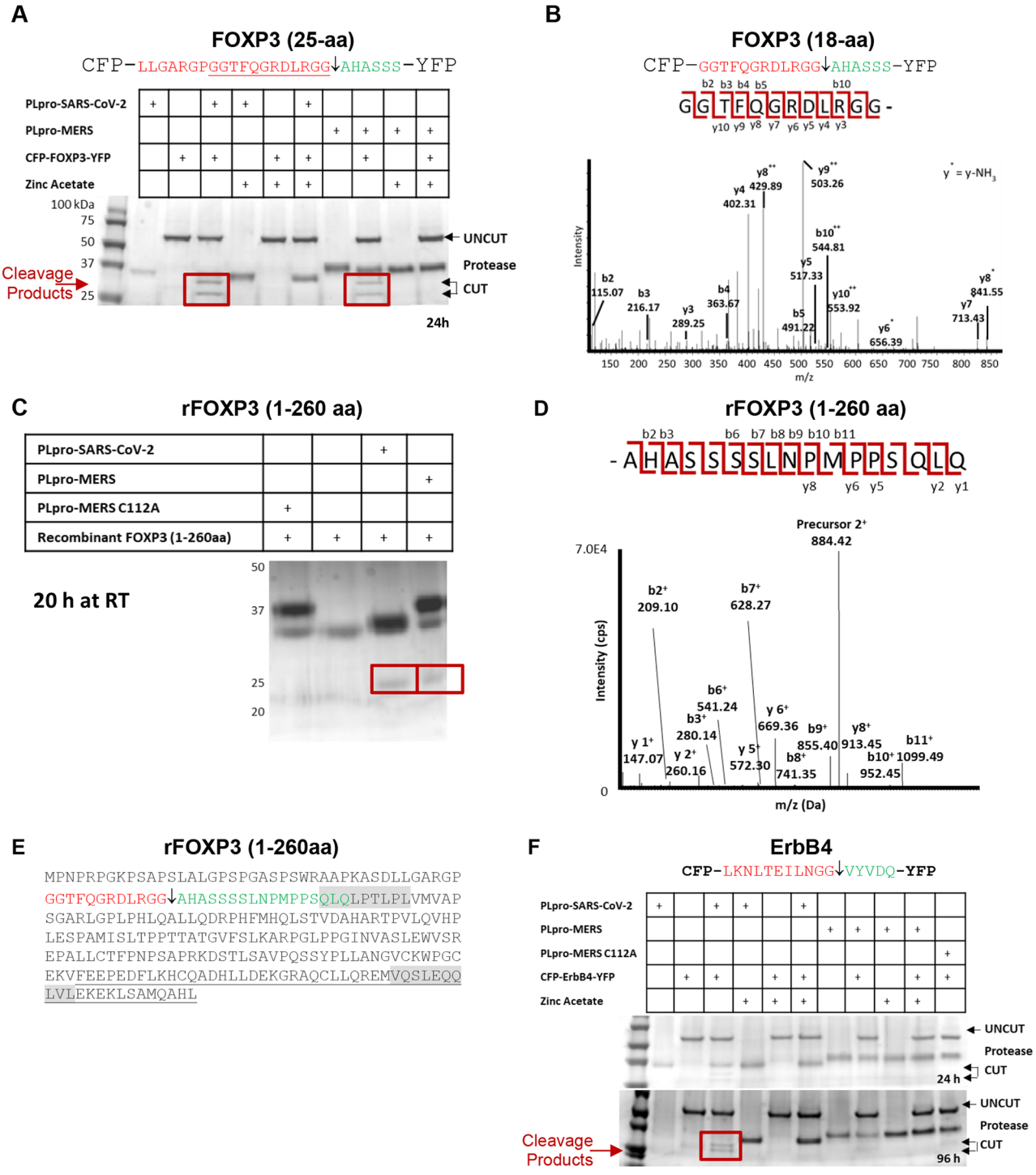


Figure 7. *In vitro* cleavage of FOXP3 sequences by the SARS-CoV-2 PLpro and MERS PLpro and cleavage of the ErbB4 (HER4) sequence by the SARS-CoV-2 PLpro. (A) SDS-PAGE gel showing cleavage of the CFP/YFP substrate containing a 25-amino acid (aa) sequence from FOXP3 by both the SARS-CoV-2 PLpro and MERS PLpro enzymes after 24 h of incubation at room temperature. Cleavage was inhibited by zinc acetate. (B) Cleavage of a CFP/YFP substrate containing 18-amino acids of the FOXP3 sequence was confirmed by MS using LysC. (C) Recombinant FOXP3 (ab226445) amino acids 1−260 (Q9BZS1) with an N-terminal His tag has a calculated MW of 32 kDa. The cleavage products have calculated MWs of 9.5 and 22.5 kDa for the N- and C-terminal fragments, respectively. Bands near the 25 kDa MW marker were observed in the samples treated with the SARS-CoV-2 PLpro and MERS PLpro. The SARS-CoV-2 product band (red box) was cut from the gel for MS analysis. (D) The cleavage site for the SARS-CoV-2 product from the reaction (red box) in (C) was confirmed by MS. The methionine was oxidized in this peptide. (E) The cleavage site in recombinant FOXP3 protein is N-terminal to the zinc finger and leucine zipper region of the protein (underlined) and the DNA-binding domain (sequence not shown). The nuclear export sequences are highlighted in gray. (F) Cleavage of the 16 aa ErbB4 (HER4) sequence embedded in the CFP/YFP substrate. Product bands were observed for the SARS-CoV-2 PLpro after 96 h of incubation at room temperature.

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

free PROS1 was below the normal range in humans with COVID-19. Martin-Rojas *et al.*[58] showed that it was significantly lower in COVID-19 patients receiving low-flow oxygen therapy.[58] Zhang *et al.*[59] also showed that the Factor VIII activity was above the normal range in all of their COVID-19 patients and that the Factor V activity was lower in near-terminal-stage patients. Factor V synergistically stimulates PROS1 activity.[57] Stoichitoiu *et al.*[55] showed a correlation between PROS1 activity and clinical severity and lung damage.[55] All three studies showed that PROS1 activity (free Protein S) was below the normal range in COVID-19 patients.

In vitro, the SARS-CoV-2 PLpro and MERS PLpro enzymes were both able to cleave the Protein S sequence in between the CFP and YFP proteins but at a low rate, and cleavage products were visible in SDS-PAGE gels after 1−3 days of room-temperature incubation and in immunoblots (Figure 6). PROS1 is abundant in human serum and was readily detectable in immunoblots; however, the protein was difficult to quantify because the bands were diffuse at high concentrations, and polyclonal rabbit antibodies did not detect a cleaved product in the protease-treated diluted serum (only uncut PROS1 could be detected). Subtle changes in intensity were observed in pooled human serum treated with the SARS-CoV-2 PLpro for 16−18 h at room temperature but were not significant (Figure S6). If only free PROS1 is cleavable, a 30−40% reduction in band intensity (*i.e.*, a less than 2-fold change) would be expected if all of the free protein were cleaved. During longer incubations, the PROS1 in the untreated serum samples (uncut control) and MERS PLpro C112A control reaction began to degrade. The cleavage of the 20-amino acid CFP−PROS1−YFP substrate by the CoV PLpro enzymes required long incubations (65−72 h at room temperature). We confirmed cleavage of the PROS1 sequence in the CFP/YFP substrate using MS and identified the IYHSAWLLIALRGG↓ peptide (Figure 6); the variant IDHSAWLLIALRGG has also been found in humans. Since PROS1 is a secreted protein, the nascent polypeptide may be cut or sequestered co-translationally at the ER by the CoV proteases intracellularly, as opposed to extracellularly in the serum. The cleavage sequence is present near the center of the protein (the MW of full-length PROS1 is ∼75 kDa, and the cleaved products have calculated MWs of 39.9 and 35.3 kDa). The decrease in free PROS1 activity in COVID-19 patient blood samples[55,58,59] is consistent with the loss of function (*i.e.*, inability to bind and activate Protein C) or reduced secretion of active PROS1 during SARS-CoV-2 infection.

Forkhead Box P3 (FOXP3) Is Cut by the SARS-CoV-2 PLpro and MERS PLpro. T_{reg} cells are present at barrier sites such as the gastrointestinal tract, lung, and skin and can limit excessive immune responses and prevent inflammatory bowel disease (IBS) and autoimmune gastritis.[60] The anti-inflammatory roles of glucocorticoids such as dexamethasone are also mediated by FOXP3+ T_{reg} cells.[61] While T_{reg} cells can suppress tissue damage mediated by virus-specific effector T cells, they can also inhibit host immunity and prevent viral clearance, leading to a persistent infection.[62] Lymphopenia has been observed in MERS, SARS, and SARS-CoV-2 infections,[63,64] and MERS and SARS have been shown to infect T cells.[65] From autopsy samples of SARS-infected humans, *in situ* hybridization and electron microscopy show virus in the circulating blood, spleen (white pulp), lymph nodes, and lymphoid tissue.[65] The aforementioned TAM receptor tyrosine kinases are also present on T_{reg} cells.[66] FOXP3 is an X-linked transcription factor whose expression has been used as a cell marker of CD4(+) CD25(+) T_{reg} cells.[60] FOXP3 is the master regulator of T_{reg} cell development and function, and it acts as both a transcriptional activator and repressor.

FOXP3 was among the top proteins predicted to be cleaved by the SARS-CoV-2 PLpro. Both the SARS-CoV-2 PLpro and MERS PLpro enzymes were able to cleave 18- and 25- amino acid FOXP3 sequences (Figure 7A). Mass spectrometry was also used to confirm the cleavage site (Figure 7B). The SARS-CoV-2 PLpro and MERS PLpro enzymes also were able to cut a recombinant FOXP3 (rFOXP3) protein containing residues 1−260 (32 kDa). This construct lacks the C-terminal DNA binding domain but includes the repressor domain, two nuclear export signals (NESs), and the zinc finger and leucine zipper motifs. The rFOXP3 protein was treated with the SARS-CoV-2 PLpro or MERS PLpro enzyme. A band at the expected MW (∼23 kDa) was observed after treatment with either protease; the smaller N-terminal fragment (∼9.5 kDa) was not observed in gels. The ∼23 kDa band was excised from the gel and analyzed by MS (Figure 7C,D). The AHASSSSLNPMPPSQLQ peptide corresponds to the C-terminal fragment and was identified from this band, consistent with PLpro cleavage after the LRGG sequence. The leucine zipper is a helical dimerization motif (Figure 7E), and the N-terminal domain of FOXP3 containing the cleavage site is the repressor domain. It contains two nuclear export sequences (^{68}QLQLPTLPL and ^{239}VQSLEQQLVL); these allow the protein to move into the cytoplasm. The nuclear localization signal (NLS) sequence ^{414}RKKR^{417} is located at the end of the DNA-binding domain. The CoV PLpro enzymes were able to cut the short 18- to 25- amino acid FOXP3 sequences in the CFP/YFP substrates as well as the longer 1−260 amino acid fragment of FOXP3.

The subcellular localization of FOXP3 in T_{reg} and non-T_{reg} cells has been studied.[67] Upon activation of CD4(+)CD25(−) T cells, FOXP3 is expressed in the cytoplasm, while in CD4(+)CD25(+) T_{reg} cells, the protein is predominantly in the nucleus. Proteolytic cleavage of FOXP3 may impair activation of T_{reg} cells, but the induction of apoptosis and subsequent lymphopenia may be due to the virus' structural proteins since apoptosis can be induced by expression of the SARS E protein in Jurkat T-cells.[68]

A subpopulation of T_{reg} cells can stop expressing FOXP3 and can begin to express inflammatory cytokines.[60] Deletion of the *foxp3* gene leads to the loss of functional T_{reg} cells and can promote the development of autoimmune diseases and inflammatory syndromes. FOXP3 activity is regulated by T-cell receptor signaling, inflammatory cytokines, and danger signals.[60] T_{reg} cells are induced by specific antigens, and populations of inducible T_{reg} cells decrease with age[69] and depend on gender.[70] In a transcriptomics study of T cells from SARS-CoV-2-infected humans, hospitalized patients had a reduced proportion of SARS-CoV-2-reactive T_{reg} cells compared with non-hospitalized patients.[71] Genetic mutations in the *foxp3* gene are always accompanied by the lack of functional T_{reg} cells,[60] and thus, the post- or cotranslational cleavage of FOXP3 may account for the reduced proportion of reactive T_{reg} cells. Cleavage of FOXP3 by the CoV PLpro may contribute to the excessive inflammation observed in COVID-19.

A Sequence within ErbB4 (HER4) Is Cut by the SARS-CoV-2 PLpro. ErbB4 is a receptor tyrosine kinase and a member of the epidermal growth factor receptor family. ErbB4 knockout

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

mice die by embryonic day 11 because of defective heart development.[72] However, mice expressing ErbB4 under a cardiac-specific myosin promoter can be rescued. These mice have aberrant cranial nerve architecture and mammary gland defects as well as alveolar hypoplasia and hyper-reactive airways, and they display changes similar to those seen in bronchopulmonary dysplasia (BPD). The lungs of these mice show signs of chronic inflammation.[72,73] The ErbB4 sequence in the CFP/YFP substrate was cut by the SARS-CoV-2 protease (Figure 7F), but no significant cleavage was observed with the MERS PLpro. This protein was located near FOXP3 in the graph (Figure 3A).

Other Predicted Hits: Respiratory- and Kidney-Related. MEGF10 is a transmembrane protein of the multiple epidermal growth factor family that is thought to mediate cell−cell adhesion, act as an engulfment receptor of apoptotic cells, and regulate muscle development and repair; it was predicted but had a low bit score. In adult mice, MEGF10 is expressed in the diaphragm.[74] Children born with MEGF10 mutations have early-onset myopathy areflexia respiratory distress and dysphagia syndrome (EMARDD) and become ventilator-dependent. Another predicted hit identified for both SARS-CoV-2 and MERS was membrane-associated guanylate kinase WW and PDZ domain containing protein 2 (MAGI2). MAGI2 mutations have been associated with nephrotic syndrome 15. Amyloid beta (A4) precursor protein-binding family B member 2 (APBB2) was also predicted; *Apbb2*$^{-/-}$ knockout mice have significant glucose intolerance, and their pancreatic islet cells display blunted insulin secretion after high glucose stimulation.[75]

DISCUSSION

One long-standing challenge in viral bioinformatics is linking sequence with symptoms. Sequence-based correlates for symptoms and virulence have practical utility in biosurveillance. Viral genomes are known to predictably cause a set of symptoms, but the information that drives these changes in the host has not been systematically mined from the viral genome. While the viral PLpro must retain its specificity for Ub- and ISG15-containing substrates, it also has acquired substrate specificity for other host proteins. In the lock-and-key analogy, these enzymes are relatively "loose" locks that recognize several substrates. These protease subsites in a sense store more information than a lock recognizing only one key. Here, we established a general method for predicting the host targets of Group IV viral proteases to enable testing and validation. We and others[1,26,29] hypothesized that the cleavage of host proteins by viral proteases may contribute to viral pathogenesis. We noted that the loss-of-function mutations in the genes of the corresponding target proteins led to symptoms that were similar to the symptoms caused by the viruses. Moreover, as we changed the cleavage site sequences, the theme of the targeted proteins within the list correspondingly changed. For example, the list of hits for the Zika viral protease contained several proteins that had been associated with neurological development and disorders, while the list for SARS-CoV-2 contained hits that were cardiovascular-, inflammation-, or blood-related (Figures 2 and 3). Both lists contained proteins involved in generating immune responses. While more data still are needed to confirm the trends, our *in vitro* data and predictions suggest an interesting correlation between the protease cleavage sites and the virus-induced phenotype.

Notably, our method is not enzyme-specific and can be applied generally to other Group IV viral proteases. It enabled us to identify cleavable sequences to test that were from the entire host proteome, including proteins that are expressed in a cell-type-specific manner that might be missed by other omic methods. Cleavable protein sequences could be found in the same general region of the graph and distribution. Product release is an essential step in catalysis, and thus, pull-down experiments can miss these important host substrates. In cells, degradation of cut (*i.e.*, damaged) proteins also occurs, complicating their identification. The methods described here may be useful for hypothesis generation.

Although there are no defined "rules" for cleavage site analysis, we tentatively note some general trends:[1,9,26] (1) cleavability tends to trend with symptoms of illness; (2) uncleavable sequences trended with a lack of symptoms or poorer viral replication, or both; and (3) host targets were immune-response-related or symptom-related, or both. It should be noted that cleavage of a protein can, in some cases, lead to activation (*e.g.*, zymogen) and may not necessarily lead to a loss of function.

Enzymes and proteins involved in generating the innate immune responses such as components of the MAVS and STING were common targets, as were subunits of mitochondrial cytochrome *c* oxidase.[76] Forkhead box (FOX) proteins were also common among the hits. While proteins that generate innate immune responses may be the primary targets of these proteases, we and others[1,26,28,29,38] have shown that cleavable short sequences recognized by the viral proteases also can be found in proteins that are not known to be involved in generating immune responses (*e.g.*, dystrophin,[29] septin-2,[38] and cardiac myosins). There were also noticeable trends within families of viruses (Table 1). For example, proteases from Picornaviridae and Calciviridae tended to recognize substrates with QG sequences, and several from Flaviviridae recognized RG substrates, whereas those from Coronaviridae recognized GG substrates (PLpro) and QG/QA/QS substrates (3CL).

In our cleavage assays, the band intensities of the cleavage products were more prominent for the MYH6/MYH7 substrate than any of the other tested substrates. MYH6/7 was the rightmost point in the graph. The accuracy of the predictions was highly dependent upon the pattern used in PHI-BLAST and the number of residues that matched known cleavage sites. In the future, it may be possible to incorporate kinetic parameters to further classify the hits. The junctional sequences in the polyproteins of alphaviruses are cut in a particular order and act as an internal timer during infection. Correlations between rates of cleavage and the occurrence of symptoms (early or late in infection) may be possible.

Sequence-based correlates of virulence are still being sought. Virulence may be related to the severity of the diseases associated with the protein being silenced post-translationally or the virus' ability to antagonize the innate immune responses and replicate, or both. Mutations in the E2 structural protein and 5′ UTR of alphaviruses are also known to attenuate these viruses.[77,78]

The proteases of Group IV viruses are validated drug targets (validated by mutation and deletion).[79,80] While in some cases we can link the viral proteases to the cleavage of host proteins and viral pathogenesis,[6] it is currently unclear whether these viral proteases outlive other viral enzymes and continue to damage host tissues even after viral replication has stopped. Foreign proteases such as those of botulinum neurotoxins are

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

SSHHPS Alignments COVID-19 nsP3/4

Protease	nsP3/4 Plasma Protein S	nsP3/4 MYH7	nsP3/4 ISG15
COVID-19 nsP3/4	VVNVVTTKIALKGG↓KIVNNW	VVNVVTTKIALKGG↓KIVNNW	VVNVVTTKIALKGG↓KIVNNW
MERS nsP3/4	ILSVRFTANKIVGG|APTWFN	ILSVRFTANKIVGG|APTWFN	ILSVRFTANKIVGG|APTWFN
SARS nsP3/4	VVNVITTKISLKGG|KIVSTW	VVNVITTKISLKGG|KIVSTW	VVNVITTKISLKGG|KIVSTW
Ubiquitin	KESTLHLVLRLRGG|	KESTLHLVLRLRGG|	KESTLHLVLRLRGG|
Organism	↓	↓	↓
Homo sapiens	IYHSAWLLIALRGG|KIEVQL	HRLDEAEQIALKGG|KKQLQK	PLSTVFMNLRLRGG|GTEPGG
Pan troglodytes	IDHSAWLLIALRGG|KIEVQL	HRLDEAEQIALKGG|KKQLQK	PLSTVFMNLRLRGG|GTEPGG
Macaca mulatta	IDHSAWLLISLRGG|KIEVQL	HRLDEAEQIALKGG|KKQLQK	PLSTVFMNLRLRGG|GTEPGG
Bos Taurus	SDHSAWFLIALREG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	KGCTVFMNLRLRGG|------
Camelus dromedarius	SDHSAWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	ARCTVYMNLRLRGG|GAGPGG
Camelus bactrianus	SDHSAWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	ARCTVYMNLRLRGG|GAGPGG
Cricetulus griseus	LDHSNWLLIALRDG|KIEVQF	HRLDEAEQIALKGG|KKQLQK	PQCIVFMNLRLRGG|GVRRVC
Mesocricetus auratus	LDHSNWLLIALRDG|KIEVQF	HRLDEAEQIALKGG|KKQLQK	PQCTVFMNLRLRGG|GGRGVC
Mus musculus	LDHSNWLLIALRDG|KIEVQF	HRLDEAEQIALKGG|KKQLQK	PQCTVI**K**HLRLRGG|GGDQCA
Rattus norvegicus	LDHSNWLLIALREG|KIEVQF	HRLDEAEQIALKGG|KKQLQK	PHCTVIMNLRLRGG|GDYYF-
Mustela putorius furo	LDHSAWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	PQCTVFMNLRLRGG|EGCAGG
Mustela ermina	LDHSAWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	PQCTVFMNLRLRGG|VGAVQG
Oryctolagus cuniculus	LDHSTWFLIALRQG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	PRCTVY**L**NLRLRGG|SAGTQP
Canis lupus familiaris	LDRSAWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	PQCTVFMNLRLRGG|GGNWAG
Felis catus	LDHSCWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	PQCTLIMNLRLRGG|------
Sus scrofa	SDHSAWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	PMCTVYMNLRLRGG|GTGPGE
Rhinolophus ferrumequinum	LDHSAWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	PHCTVLMNLRLRGG|GTGPEG
Phyllostomus discolor	VDQSAWFLIAVRDG|KLEIQV	HRLDEAEQIALKGG|KKQLQK	PQCTVQMNLYLRGG|RVGPGE
Pteropus Alecto	LDHSAWFLIALRDG|KIEIQF	HRLDEAEQIALKGG|KKQLQK	SLCTVHMNLRLRGG|AAGLRG
Myotis lucifugus	LDHSAWLLIALRDG|KLEIQF	HRLDEAEQIALKGG|KKQLQK	----(no coverage)----
Myotis davidii	LDGSAWLLIALRGG|KLEIQF	HRLDEAEQIALKGG|KKQLQK	PHCTVQMNLRLRGG|EEGLGG
Salmo salar	SSQDSWFMLGLRGG|RIEVQF	HRLDEAEQIAMKGG|KKQIQK	NQSTIHLALRLRGG|------

SSHHPS Alignments COVID-19 nsP1/2 and nsP2/3

Protease	nsp1/2 FOXP3	nsP1/2 erbB4(HER4)	nsP2/3 von Willebrand factor
COVID-19	SGVTRELMRELNGG↓AYTRYV	SGVTRELMRELNGG↓AYTRYV	NMMVTNNTFTLKGG↓APTKVT
MERS	GKYAQNLLKKLIGG|DVTPVD	GKYAQNLLKKLIGG|DVTPVD	NGHAVPTLFRLKGG|APVKKV
SARS	SGALRELTRELNGG|AVTRYV	SGALRELTRELNGG|AVTRYV	GLLATNNVFRLKGG|APIKGV
Ubiquitin	KESTLHLVLRLRGG|	KESTLHLVLRLRGG|	KESTLHLVLRLRGG|
Organism	↓	↓	↓
Homo sapiens	GPGGTFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLPTACTIQLRGG|QIMTLK
Pan troglodytes	GPGGTFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLPTACTIQLRGG|QIMTLK
Macaca mulatta	GPGGIFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLPTVCTIRLRGG|QIMTLK
Bos taurus	SOGTTFQGRDLRSG|AHTSSS	ELGLKNLTEILNGG|VYVDQN	RCLPTACTIQLRGG|QIMMLK
Camelus dromedarius	GPGTTFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLSTACTIQLQGG|QILTLK
Camelus bactrianus	GPGTTFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLSTACTIQLQGG|QILTLK
Cricetulus griseus	GPGGTFQGRDLRGG|AHTSSS	ELGLKNLTEILNGG|VYVDQN	TLGYHIQLGPLQGG|SLRRVE
Mesocricetus auratus	GPGGTFQGRDLRGG|AHTSSL	ELGLKNLTEILNGG|VYVDQN	RCLPTACTIQLRGG|HITTLK
Mus musculus	GSGGPFQGRDLRSG|AHTSSS	ELGLKNLTEILNGG|VYVDQN	RCLPTACTIQLRGG|QIMTLK
Rattus norvegicus	GPGGPFQGRDLRSG|AHTSSS	ELGLKNLTEILNGG|VYVDQN	RCLPTACTIQLRGG|RVMTLK
Mustela putorius furo	GPGATFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	KCLPTACTIQLRGG|QIMTLK
Mustela ermina	GPGATFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	KCLPTACTIQLRGG|QIMTLK
Oryctolagus cuniculus	GPGGTFQGRDLRGG|AHATTT	ELGLKNLTEILNGG|VYVDQN	RCLPMGCTIQLRGG|QIMTLK
Canis lupus familiaris	GPGVTFQGRDLRGG|THASSS	ELGLKNLTEILNGG|VYVDQN	RCLPIACTIQLRGG|QIMTLK
Felis catus	GPGATFQGRDLRGG|THASSS	ELGLKNLTEILNGG|VYVDQN	RCLPVACTIQLRGG|QTVMLK
Sus scrofa	GPGAAFQGRELRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLPTACTIQLRGG|QIMTLK
Rhinolophus ferrumequinum	GPGATFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	KCRLMACTIQLRGG|QIMTLK
Phyllostomus discolor	GPGATFQGRDPRGG|THASSS	ELGLKNLTEILNGG|VYVDQN	RCLLLACTIQLRGG|QTMTLK
Pteropus alecto	GPGATFQGRDLRGV|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLPMACTIQLRGG|QIMTLK
Myotis lucifugus	GPGATFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLPVACTIQLRGG|QIMTLQ
Myotis davidii	GPGATFQGRDLRGG|AHASSS	ELGLKNLTEILNGG|VYVDQN	RCLPVACTIQLRGG|QIMTLQ
Salmo salar	NSGRQQQQRDRRGE|HGEEPD	QLGLKNLTEILQGG|VYVDQN	-----no homologue----

Figure 8. Multispecies sequence alignments of predicted and *in vitro*-validated cleavage site sequences. Residues that match a predicted cleavage site sequence in consecutive order are shown in red, and residues that are tolerated at the subsite but are present in other cleavage sequences are shown in green. The human PROS1, MYH6/MYH7, FOXP3 and ErbB4(HER4), sequences were inserted between CFP and YFP and tested *in vitro*. Deubiquitinase and deISGylase activity was shown by Freitas *et al.*[22]

known to remain active in human cells for months. The role of double membrane vesicles, which are derived from the cell's secretory system, also is unclear; the CoV PLpro is anchored to the ER membrane. In proteases such as those of CoV, the DUB activity may also contribute to their cellular lifetimes.

For reservoir species, an obvious rule or trend was not apparent. Bats are the proposed reservoir species of CoV. Although some bat genomes are complete, others are only partial, and some homologues are absent in bats. RNA sequences matching those found in the virus at the junctional regions spanning the protease cleavage sites could be readily

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

found in several bat species (Figure S2); however, these were relatively short sequences. Many of the bat protein sequences in the targeted host proteins had key substitutions in the cleavage site motifs, while others were identical to those in humans (Figure 8). *Myotis davidii*, a bat found near lakes in Siberia as well as in parts of Europe, China, and Korea,[81] had several cleavage site sequences (protein) that were almost identical to those of humans (*e.g.*, PROS1). Sequence alignments of the viral polyproteins of SARS, SARS-CoV-2, and MERS show substitutions, insertions, and deletions in and around these protease cleavage sequences, suggesting that they may be RNA recombination hot spots (Figure S2).

Finally, while the general logic (*i.e.*, post-translational silencing) described here may apply only to Group IV and possibly some Group VI viruses, there are a total of seven groups of viruses, and similar logic has not been reported for each of these. Viral bioinformatics is still very much in its infancy.[82,83] Viral genomes contain algorithms, and their logic may become the basis of computer programs that can predict symptoms and virulence directly from viral sequences.

MATERIALS

General chemicals and BugBuster were purchased from Fisher Scientific (Waltham, MA) or Millipore-Sigma (Burlington, MA). MPER was obtained from Thermo Scientific. Chelating Sepharose, SP-Sepharose, and Q-Sepharose were purchased from Cytiva (Marlborough, MA) (formerly GE Healthcare Life Sciences). Recombinant FOXP3 was purchased from Abcam (Cambridge, MA). Novagen-competent *Escherichia coli* was obtained from Millipore-Sigma.

METHODS

Bioinformatics. PHI-BLAST was used for host protein target predictions. The default settings for a short sequence search select a PAM30 scoring matrix. A 20 amino acid sequence from each junctional region (nsp1/2, SGVTRELMRELNGG↓AYTRYV; nsP2/3, NMMVTNNTFTLKGG↓APTKVT; nsP3/4, VVNVVTTKIALKGG↓KIVNNW) was input with a pattern sequence (L[RKN]GG), and a search for hits in *Homo sapiens* (taxid: 9606) was performed. For the Zika viral protease a 14-residue sequence centered around the scissile bond (FAAGKRG↓AALGVME) was used. The PHI-BLAST HitTable.csv file was downloaded and used for analysis.

Expression and Purification of the COVID-19 SARS-CoV-2 PLpro. A pet15b plasmid was constructed with an N-terminal thrombin-cleavable His_6 tag. The plasmid encoded amino acids 1564−1878 of the nonstructural polyprotein papain-like protease (PLpro) from NCBI reference sequence YP_009724389.1.

BL21(DE3) pLysS *E. coli* was transformed with the plasmid, and 3 L of Luria−Bertani (LB) medium was inoculated with ∼50 mL of overnight culture. Ampicillin (100 μg/mL final concentration) and chloramphenicol (4 μg/mL) were added to the medium. The culture was grown to an optical density at 600 nm (OD_{600}) of approximately ∼1.0 at 37 °C. Protein expression was induced with IPTG (0.5 mM final concentration), and the temperature was lowered to 17 °C. The protein was allowed to express during overnight incubation (∼17 h). The next day, cells were collected by centrifugation (5000*g*, 10 min, 4 °C) and lysed in lysis buffer containing 50 mM Tris (pH 7.2), 500 mM NaCl, 2 mM β-mercaptoethanol (BME), 5% glycerol, ∼30 mg lysozyme and DNase, and 35% BugBuster. The lysate was sonicated for ∼2 min on level 5 at 20 s intervals while the tubes were submerged in ice water. The lysate was clarified by centrifugation (20500*g* for 30 min at 4 °C) and then loaded onto a nickel-charged chelating Sepharose column equilibrated with 3 column volumes of 50 mM Tris (pH 7.2) containing 500 mM NaCl, 5% glycerol, and 2 mM BME. The column was washed with ∼2 column volumes of the same buffer after loading and then with ∼2 column volumes of buffer containing 60 mM imidazole. The protein was then eluted with the same buffer containing 300 mM imidazole. The fractions containing the protease were dialyzed overnight at 4 °C in 50 mM Tris (pH 7.2) containing 250 mM NaCl, 5 mM DTT (dithiothreitol), 1 mM EDTA (ethylenediaminetetraacetic acid), and 5% glycerol. Protein was removed from dialysis, diluted, loaded onto an SP-Sepharose column equilibrated with 50 mM Tris (pH 7.2) containing 100 mM NaCl, 5 mM DTT, and 5% glycerol, and collected from the flow through. Purified protein was aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

Expression and Purification of the MERS PLpro. BL21(DE3) pLysS *E. coli* was transformed with the plasmid, and 3 L of LB containing ampicillin and chloramphenicol was grown at 37 °C and then induced with 0.5 mM IPTG at 17 °C overnight (∼17 h). Cell pellets were lysed in 50 mM Tris (pH 8.5) containing 500 mM NaCl, 2 mM BME, 5% glycerol, 35% BugBuster, ∼30 mg of lysozyme, and DNase. The lysate was sonicated and clarified by centrifugation (20500*g* for 30 min at 4 °C). The supernatant was loaded onto a nickel column equilibrated with 50 mM Tris (pH 8.5) containing 500 mM NaCl, 5% glycerol, and 2 mM BME. The column was washed with the same buffer containing 60 mM imidazole, and the protein was eluted with the same buffer containing 300 mM imidazole. The protein was then dialyzed against 50 mM Tris (pH 8.5) containing 250 mM NaCl, 5 mM DTT, 1 mM EDTA, and 5% glycerol at 4 °C overnight and then for 2 h in buffer containing 125 mM NaCl. The protein was loaded onto a Q-Sepharose column equilibrated with 50 mM Tris (pH 8.5) containing 125 mM NaCl, 5 mM DTT, and 5% glycerol, and the protein in the flow through was collected. The protein was then aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C. The C112A variant was purified using the same methods.

Immunoblotting. Bovine heart lysates were prepared using ∼600−700 mg of tissue cut from various regions of the heart and mixed with 600−700 μL of 25 mM HEPES (pH 7.4) containing 0.1% Tween-20. The lysates were sonicated for 10−15 s and further diluted for protease assays. Lysates were diluted 1:100 with MPER for immunoblots. Digest reactions with 2.6 μM enzyme were run for 16 h to 5 days at room temperature, stopped with 2× Laemelli buffer containing BME (975:25 v/v), and then separated by SDS-PAGE and transferred to nitrocellulose. Blots were blocked for 1 h at room temperature with 1× PBST (phosphate-buffered saline, 0.1% Tween-20) containing 5% w/v dry milk, and primary antibodies were co-incubated with fresh 1× PBST and 5% w/v dry milk overnight at 4 °C with rocking. An anti-MYH7 mouse mAb primary antibody from Abcam (ab11083) was used at a concentration of 1:1000 with a horseradish peroxidase (HRP)-conjugated anti-mouse secondary antibody (1:1000). Actin was detected with the C4 mAb (1:500, MAB1501R, Millipore). The HRP-conjugated polyclonal goat anti-mouse antibody (Upstate no. 12-349, Sigma-Millipore or NXA931,

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

Sigma) was used as the secondary antibody at 1:1000 dilution. MYH6 was detected using a mouse monoclonal Ab (Abcam, ab207926) at 1:1000 dilution.

Commercially available pooled human serum (Sigma, H4522) was diluted 1:100 in pH 7.6 MPER (Thermo Scientific, 78503) and incubated with each PLpro ($[E]_f$ = 4 μM) at room temperature for 20 h. The C112A MERS PLpro variant was used as a control. Products were separated on SDS-PAGE gels (12%), blotted onto nitrocellulose, and cut below the 75 kDa MW marker. The upper part of the blot was incubated overnight in 1× PBST and 5% w/v dry milk at 4 °C with the anti-PROS1 polyclonal rabbit (Rb) antibody (Abcam, ab97387) at 1:1000 dilution and anti-Rb-HRP antibodies (1:1000, NA934, Cytiva and sc-2030, Santa Cruz Biotech (Dallas, TX). Samples were loaded in triplicate. The lower part of the blot was incubated with an anti-β-actin C4 antibody (1:500, MAB1501R, Millipore) overnight at 4 °C. The secondary antibody was an HRP-conjugated polyclonal goat anti-mouse antibody (NXA931, Sigma-Millipore) used at 1:1000. Blots were developed with SuperSignal West Pico Plus chemiluminescent substrate according to the manufacturer's directions. Band intensities were measured using Image Lab 3.0 (BioRad Inc., Hercules, CA).

Recombinantly produced N-terminally His-tagged FOXP3 fragment containing residues 1−260 (acc. no. Q9BZS1, ab226445, 0.05 mg/mL final concentration) was mixed with each of the PLpro enzymes (10 μM final concentration) for 20 h at room temperature in MPER. Reactions were stopped by addition of an equal volume of 2× Laemelli buffer containing BME (975:25 v/v). The uncut protein has a reported MW of 32 kDa. Cleavage at the anticipated site would produce ∼9 and 23 kDa N- and C-fragments, respectively (*n.b.*, the N-terminal fragment was not observed). Gel bands were excised and processed for mass spectrometry.

Mass Spectrometry (MS). MS/MS spectra were using a high-resolution Orbitrap MS analyzer. Gel bands were cut and digested in gel by AspN, LysC, or arginyl endopeptidase (ArgC) overnight. Peptides were extracted from gel pieces with 0.1% formic acid in 50% acetonitrile and 100% acetonitrile and then dried on the speed vac. Prior to analysis by LC−MS/MS, peptides were reconstituted in 0.1% formic acid in water (20 μL) and placed in a vial. Subsequently 3 μL of the peptide mixture was injected into a U3000 HPLC coupled to a Thermo Scientific Orbitrap Fusion Lumos mass spectrometer equipped with a Nanospray Flex ion source and analyzed by shotgun proteomics.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsinfecdis.0c00866.

Additional immunoblots, alignments, and substrate sequences (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Patricia M. Legler − *Center for Bio/molecular Science and Engineering (CBMSE), U.S. Naval Research Laboratory, Washington, DC 20375, United States;* orcid.org/0000-0003-1196-1061; Phone: 202-404-6037; Email: patricia.legler@nrl.navy.mil

Authors

Nathanael D. Reynolds − *Center for Bio/molecular Science and Engineering (CBMSE), U.S. Naval Research Laboratory, Washington, DC 20375, United States*

Nathalie M. Aceves − *California State University, Fresno, California 93740, United States*

Jinny L. Liu − *Center for Bio/molecular Science and Engineering (CBMSE), U.S. Naval Research Laboratory, Washington, DC 20375, United States*

Jaimee R. Compton − *Center for Bio/molecular Science and Engineering (CBMSE), U.S. Naval Research Laboratory, Washington, DC 20375, United States*

Dagmar H. Leary − *Center for Bio/molecular Science and Engineering (CBMSE), U.S. Naval Research Laboratory, Washington, DC 20375, United States;* orcid.org/0000-0003-2325-7143

Brendan T. Freitas − *Center for Drug Discovery, College of Pharmacy, University of Georgia, Athens, Georgia 30602, United States*

Scott D. Pegan − *Center for Drug Discovery, College of Pharmacy, University of Georgia, Athens, Georgia 30602, United States;* orcid.org/0000-0002-2958-5319

Katarina Z. Doctor − *Navy Center for Applied Research in AI (NCARAI) Information Technology Division, U.S. Naval Research Laboratory, Washington, DC 20375, United States*

Fred Y. Wu − *Indiana University Health Systems, Indiana University School of Medicine, Bloomington, Indiana 47401, United States*

Xin Hu − *National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsinfecdis.0c00866

Author Contributions

N.D.R., N.M.A., J.R.C., and P.M.L. collected data. K.Z.D., F.Y.W., and X.H. contributed to the writing of the manuscript. P.M.L. conceived the ideas and wrote the manuscript. All of the authors approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Capt. David Bauer (ret.) and Cheryl Walton of the Assistant Secretary of the Navy (Research, Development, and Acquisition) and Christopher Marchefsky of the Office of Naval Research for their support of this work. We thank Barry Pineles for proofreading. We also thank the clinicians who reported the vast number of symptoms associated with COVID-19 and other viral infections. This work was supported by the Office of the Secretary of the Navy, Naval Innovated Science and Engineering (NISE) Funding (1X42), and ONR Core 6.1 Base Funding (IL73). N.M.A. was supported by the NRL HBCU/MI Summer Internship Program sponsored by the Office of Naval Research (ONR) (N0017315RMV04). The interpretations and opinions expressed belong to the authors and do not necessarily represent the official views of the U.S. Navy, U.S. Army, U.S. Department of Defense, or U.S. Government.

REFERENCES

(1) Morazzani, E. M., Compton, J. R., Leary, D. H., Berry, A. V., Hu, X., Marugan, J., Glass, P. J., and Legler, P. M. (2019) Proteolytic cleavage of host proteins by the Group IV viral proteases of Venezuelan equine encephalitis virus and Zika virus. *Antiviral Res. 164*, 106−122.

(2) Falk, M. M., Grigera, P. R., Bergmann, I. E., Zibert, A., Multhaup, G., and Beck, E. (1990) Foot-and-mouth disease virus protease 3C induces specific proteolytic cleavage of host cell histone H3. *J. Virol. 64* (2), 748−56.

(3) Li, W., Ross-Smith, N., Proud, C. G., and Belsham, G. J. (2001) Cleavage of translation initiation factor 4AI (eIF4AI) but not eIF4AII by foot-and-mouth disease virus 3C protease: identification of the eIF4AI cleavage site. *FEBS Lett. 507* (1), 1−5.

(4) Griger, P. R., and Tisminetzky, S. G. (1984) Histone H3 modification in BHK cells infected with foot-and-mouth disease virus. *Virology 136* (1), 10−19.

(5) Ventoso, I., Blanco, R., Perales, C., and Carrasco, L. (2001) HIV-1 protease cleaves eukaryotic initiation factor 4G and inhibits cap-dependent translation. *Proc. Natl. Acad. Sci. U. S. A. 98* (23), 12966−71.

(6) Badorff, C., Berkely, N., Mehrotra, S., Talhouk, J. W., Rhoads, R. E., and Knowlton, K. U. (2000) Enteroviral protease 2A directly cleaves dystrophin and is inhibited by a dystrophin-based substrate analogue. *J. Biol. Chem. 275* (15), 11191−7.

(7) Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2017) The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *mBio 8* (5), 1−18.

(8) Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S., and Yang, S. H. (2015) Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol. Ther.–Nucleic Acids 4* (11), No. e264.

(9) Hu, X., Compton, J. R., and Legler, P. M. (2019) Analysis of Group IV Viral SSHHPS Using In Vitro and In Silico Methods. *J. Vis. Exp.* No. 154, e60421.

(10) Zhou, Z., Jia, X., Xue, Q., Dou, Z., Ma, Y., Zhao, Z., Jiang, Z., He, B., Jin, Q., and Wang, J. (2014) TRIM14 is a mitochondrial adaptor that facilitates retinoic acid-inducible gene-I-like receptor-mediated innate immune response. *Proc. Natl. Acad. Sci. U. S. A. 111* (2), E245−54.

(11) Montgomery, S. A., and Johnston, R. E. (2007) Nuclear import and export of Venezuelan equine encephalitis virus nonstructural protein 2. *J. Virol. 81* (19), 10268−79.

(12) Báez-Santos, Y. M., St. John, S. E., and Mesecar, A. D. (2015) The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res. 115*, 21−38.

(13) Gorbalenya, A. E. (1992) Host-related sequences in RNA viral genomes. *Semin. Virol. 3*, 359−371.

(14) Meyers, G., Tautz, N., Dubovi, E. J., and Thiel, H. J. (1991) Viral cytopathogenicity correlated with integration of ubiquitin-coding sequences. *Virology 180* (2), 602−616.

(15) Lai, M. M. C. (1992) Genetic Recombination in RNA Viruses. In *Genetic Diversity of RNA Viruses* (Holland, J. J., Ed.) pp 21−32, Springer, Berlin. DOI: 10.1007/978-3-642-77011-1_2.

(16) Banner, L. R., Keck, J. G., and Lai, M. M. (1990) A clustering of RNA recombination sites adjacent to a hypervariable region of the peplomer gene of murine coronavirus. *Virology 175* (2), 548−55.

(17) Kirkegaard, K., and Baltimore, D. (1986) The mechanism of RNA recombination in poliovirus. *Cell 47* (3), 433−443.

(18) Becher, P., and Tautz, N. (2011) RNA recombination in pestiviruses: cellular RNA sequences in viral genomes highlight the role of host factors for viral persistence and lethal disease. *RNA Biol. 8* (2), 216−24.

(19) Gallei, A., Pankraz, A., Thiel, H.-J. C., and Becher, P. (2004) RNA recombination in vivo in the absence of viral replication. *J. Virol. 78* (12), 6271−6281.

(20) Chen, Y., Liu, Q., and Guo, D. (2020) Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol. 92* (4), 418−423.

(21) Mielech, A. M., Deng, X., Chen, Y., Kindler, E., Wheeler, D. L., Mesecar, A. D., Thiel, V., Perlman, S., and Baker, S. C. (2015) Murine coronavirus ubiquitin-like domain is important for papain-like protease stability and viral pathogenesis. *J. Virol. 89* (9), 4907−17.

(22) Freitas, B. T., Durie, I. A., Murray, J., Longo, J. E., Miller, H. C., Crich, D., Hogan, R. J., Tripp, R. A., and Pegan, S. D. (2020) Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infect. Dis. 6*, 2099−2109.

(23) Viswanathan, K., Fruh, K., and DeFilippis, V. (2010) Viral hijacking of the host ubiquitin system to evade interferon responses. *Curr. Opin. Microbiol. 13* (4), 517−23.

(24) Davis, M. E., and Gack, M. U. (2015) Ubiquitination in the antiviral immune response. *Virology 479−480*, 52−65.

(25) Lindner, H. A., Fotouhi-Ardakani, N., Lytvyn, V., Lachance, P., Sulea, T., and Menard, R. (2005) The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J. Virol. 79* (24), 15199−208.

(26) Kiemer, L., Lund, O., Brunak, S., and Blom, N. (2004) Coronavirus 3CLpro proteinase cleavage sites: possible relevance to SARS virus pathology. *BMC Bioinf. 5*, 72.

(27) Miczi, M., Golda, M., Kunkli, B., Nagy, T., Tőzsér, J., and Mótyán, J. A. (2020) Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease. *Int. J. Mol. Sci. 21* (24), 9523.

(28) Blom, N., Hansen, J., Blaas, D., and Brunak, S. (1996) Cleavage site analysis in picornaviral polyproteins: Discovering cellular targets by neural networks. *Protein Sci. 5* (11), 2203−2216.

(29) Badorff, C., Lee, G.-H., Lamphear, B. J., Martone, M. E., Campbell, K. P., Rhoads, R. E., and Knowlton, K. U. (1999) Enteroviral protease 2A cleaves dystrophin: Evidence of cytoskeletal disruption in an acquired cardiomyopathy. *Nat. Med. 5* (3), 320−326.

(30) Lim, B.-K., Peter, A. K., Xiong, D., Narezkina, A., Yung, A., Dalton, N. D., Hwang, K.-K., Yajima, T., Chen, J., and Knowlton, K. U. (2013) Inhibition of Coxsackievirus-associated dystrophin cleavage prevents cardiomyopathy. *J. Clin. Invest. 123* (12), 5146−5151.

(31) Stabell, A. C., Meyerson, N. R., Gullberg, R. C., Gilchrist, A. R., Webb, K. J., Old, W. M., Perera, R., and Sawyer, S. L. Dengue viruses cleave STING in humans but not in nonhuman primates, their presumed natural reservoir. *eLife* 2018, 7, DOI: 10.7554/eLife.31919.

(32) Legler, P. M., Morazzani, E., Glass, P. J., and Compton, J. R. (2018) Proteome Editing System and a Biomarker of VEEV Infection, US 20180149651.

(33) Storer, A. C., and Ménard, R. (1996) Recent insights into cysteine protease specificity: Lessons for drug design. *Perspect. Drug Discovery Des. 6* (1), 33−46.

(34) Ten Dam, E., Flint, M., and Ryan, M. D. (1999) Virus-encoded proteinases of the Togaviridae. *J. Gen. Virol. 80*, 1879−1888.

(35) Karlin, S., and Altschul, S. F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. U. S. A. 87* (6), 2264−8.

(36) Florian, C., Bahi-Buisson, N., and Bienvenu, T. (2011) FOXG1-Related Disorders: From Clinical Description to Molecular Genetics. *Mol. Syndromol. 2* (3−5), 153−163.

(37) Hong, S. T., and Mah, W. (2015) A Critical Role of GIT1 in Vertebrate and Invertebrate Brain Development. *Experimental neurobiology 24* (1), 8−16.

(38) Li, H., Saucedo-Cuevas, L., Yuan, L., Ross, D., Johansen, A., Sands, D., Stanley, V., Guemez-Gamboa, A., Gregor, A., Evans, T., Chen, S., Tan, L., Molina, H., Sheets, N., Shiryaev, S. A., Terskikh, A. V., Gladfelter, A. S., Shresta, S., Xu, Z., and Gleeson, J. G. (2019) Zika Virus Protease Cleavage of Host Protein Septin-2 Mediates Mitotic Defects in Neural Progenitors. *Neuron 101* (6), 1089−1098.

(39) Bailey-Elkin, B. A., Knaap, R. C., Johnson, G. G., Dalebout, T. J., Ninaber, D. K., van Kasteren, P. B., Bredenbeek, P. J., Snijder, E. J., Kikkert, M., and Mark, B. L. (2014) Crystal structure of the Middle East respiratory syndrome coronavirus (MERS-CoV) papain-like protease bound to ubiquitin facilitates targeted disruption of

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

deubiquitinating activity to demonstrate its role in innate immune suppression. *J. Biol. Chem. 289* (50), 34667−82.
(40) England, J., and Loughna, S. (2013) Heavy and light roles: myosin in the morphogenesis of the heart. *Cell. Mol. Life Sci. 70* (7), 1221−39.
(41) CDC. *Symptoms of Coronavirus*. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html#print (accessed 2021-04-30).
(42) Wichmann, D., Sperhake, J.-P., Lütgehetmann, M., Steurer, S., Edler, C., Heinemann, A., Heinrich, F., Mushumba, H., Kniep, I., Schröder, A. S., Burdelski, C., de Heer, G., Nierhaus, A., Frings, D., Pfefferle, S., Becker, H., Bredereke-Wiedling, H., de Weerth, A., Paschen, H.-R., Sheikhzadeh-Eggers, S., Stang, A., Schmiedel, S., Bokemeyer, C., Addo, M. M., Aepfelbacher, M., Püschel, K., and Kluge, S. (2020) Autopsy Findings and Venous Thromboembolism in Patients with COVID-19: A Prospective Cohort Study. *Ann. Intern. Med. 173* (4), 268−277.
(43) Freaney, P. M., Shah, S. J., and Khan, S. S. (2020) COVID-19 and Heart Failure With Preserved Ejection Fraction. *JAMA 324* (15), 1499−1500.
(44) Perez-Bermejo, J. A., Kang, S., Rockwood, S. J., Simoneau, C. R., Joy, D. A., Ramadoss, G. N., Silva, A. C., Flanigan, W. R., Li, H., Nakamura, K., Whitman, J. D., Ott, M., Conklin, B. R., and McDevitt, T. C. (2020) SARS-CoV-2 infection of human iPSC-derived cardiac cells predicts novel cytopathic features in hearts of COVID-19 patients. *bioRxiv*, DOI: 10.1101/2020.08.25.265561.
(45) Farouk, S. S., Fiaccadori, E., Cravedi, P., and Campbell, K. N. (2020) COVID-19 and the kidney: what we think we know so far and what we don't. *J. Nephrol. 33* (6), 1213−1218.
(46) Taylor, K. C., Buvoli, M., Korkmaz, E. N., Buvoli, A., Zheng, Y., Heinze, N. T., Cui, Q., Leinwand, L. A., and Rayment, I. (2015) Skip residues modulate the structural properties of the myosin rod and guide thick filament assembly. *Proc. Natl. Acad. Sci. U. S. A. 112* (29), E3806.
(47) Sharma, A., Garcia, G., Arumugaswami, V., and Svendsen, C. N. (2020) Human iPSC-Derived Cardiomyocytes are Susceptible to SARS-CoV-2 Infection. *bioRxiv*, DOI: 10.1101/2020.04.21.051912.
(48) Dolhnikoff, M., Ferreira Ferranti, J., de Almeida Monteiro, R. A., Duarte-Neto, A. N., Soares Gomes-Gouvea, M., Viu Degaspare, N., Figueiredo Delgado, A., Montanari Fiorita, C., Nunes Leal, G., Rodrigues, R. M., Taverna Chaim, K., Rebello Pinho, J. R., Carneiro-Sampaio, M., Mauad, T., Ferraz da Silva, L. F., Brunow de Carvalho, W., Saldiva, P. H. N., and Garcia Caldini, E. (2020) SARS-CoV-2 in cardiac tissue of a child with COVID-19-related multisystem inflammatory syndrome. *Lancet Child Adolesc Health 4* (10), 790−794.
(49) Wang, Y., Liu, S., Liu, H., Li, W., Lin, F., Jiang, L., Li, X., Xu, P., Zhang, L., Zhao, L., Cao, Y., Kang, J., Yang, J., Li, L., Liu, X., Li, Y., Nie, R., Mu, J., Lu, F., Zhao, S., Lu, J., and Zhao, J. (2020) SARS-CoV-2 infection of the liver directly contributes to hepatic impairment in patients with COVID-19. *J. Hepatol. 73* (4), 807−816.
(50) Dofferhoff, A. S. M., Piscaer, I., Schurgers, L. J., Visser, M. P. J., van den Ouweland, J. M. W., de Jong, P. A., Gosens, R., Hackeng, T. M., van Daal, H., Lux, P., Maassen, C., Karssemeijer, E. G. A., Vermeer, C., Wouters, E. F. M., Kistemaker, L. E. M., Walk, J., and Janssen, R. (2020) Reduced vitamin K status as a potentially modifiable risk factor of severe COVID-19. *Clin. Infect. Dis.*, DOI: 10.1093/cid/ciaa1258.
(51) Connors, J. M., and Levy, J. H. (2020) COVID-19 and its implications for thrombosis and anticoagulation. *Blood 135* (23), 2033−2040.
(52) Edler, C., Schröder, A. S., Aepfelbacher, M., Fitzek, A., Heinemann, A., Heinrich, F., Klein, A., Langenwalder, F., Lütgehetmann, M., Meißner, K., Püschel, K., Schädler, J., Steurer, S., Mushumba, H., and Sperhake, J.-P. (2020) Dying with SARS-CoV-2 infection—an autopsy study of the first consecutive 80 cases in Hamburg, Germany. *Int. J. Legal Med. 134* (4), 1275−1284.
(53) Giannis, D., Ziogas, I. A., and Gianni, P. (2020) Coagulation disorders in coronavirus infected patients: COVID-19, SARS-CoV-1, MERS-CoV and lessons from the past. *J. Clin. Virol. 127*, 104362−104362.
(54) Lemke, G., and Silverman, G. J. (2020) Blood clots and TAM receptor signalling in COVID-19 pathogenesis. *Nat. Rev. Immunol. 20* (7), 395−396.
(55) Stoichitoiu, L. E., Pinte, L., Balea, M. I., Nedelcu, V., Badea, C., and Baicus, C. (2020) Anticoagulant protein S in COVID-19: low activity, and associated with outcome. *Rom. J. Intern. Med. 58* (4), 251−258.
(56) Van Cott, E. M., Ledford-Kraemer, M., Meijer, P., Nichols, W. L., Johnson, S. M., and Peerschke, E. I. (2005) Protein S assays: an analysis of North American Specialized Coagulation Laboratory Association proficiency testing. *Am. J. Clin. Pathol. 123* (5), 778−85.
(57) Dahlback, B. (2008) Advances in understanding pathogenic mechanisms of thrombophilic disorders. *Blood 112* (1), 19−27.
(58) Martin-Rojas, R. M., Perez-Rus, G., Delgado-Pinos, V. E., Domingo-Gonzalez, A., Regalado-Artamendi, I., Alba-Urdiales, N., Demelo-Rodriguez, P., Monsalvo, S., Rodriguez-Macias, G., Ballesteros, M., Osorio-Prendes, S., Diez-Martin, J. L., and Pascual Izquierdo, C. (2020) COVID-19 coagulopathy: An in-depth analysis of the coagulation system. *Eur. J. Haematol. 105* (6), 741−750.
(59) Zhang, Y., Cao, W., Jiang, W., Xiao, M., Li, Y., Tang, N., Liu, Z., Yan, X., Zhao, Y., Li, T., and Zhu, T. (2020) Profile of natural anticoagulant, coagulant factor and anti-phospholipid antibody in critically ill COVID-19 patients. *J. Thromb. Thrombolysis 50* (3), 580−586.
(60) Nie, J., Li, Y. Y., Zheng, S. G., Tsun, A., and Li, B. (2015) FOXP3(+) Treg Cells and Gender Bias in Autoimmune Diseases. *Front. Immunol. 6*, 493.
(61) Kim, D., Nguyen, Q. T., Lee, J., Lee, S. H., Janocha, A., Kim, S., Le, H. T., Dvorina, N., Weiss, K., Cameron, M. J., Asosingh, K., Erzurum, S. C., Baldwin, W. M., 3rd, Lee, J. S., and Min, B. (2020) Anti-inflammatory Roles of Glucocorticoids Are Mediated by Foxp3(+) Regulatory T Cells via a miR-342-Dependent Mechanism. *Immunity 53* (3), 581−596 e5.
(62) Li, S., Gowans, E. J., Chougnet, C., Plebanski, M., and Dittmer, U. (2008) Natural Regulatory T Cells and Persistent Viral Infection. *J. Virol. 82* (1), 21−30.
(63) Huang, I., and Pranata, R. (2020) Lymphopenia in severe coronavirus disease-2019 (COVID-19): systematic review and meta-analysis. *Journal of Intensive Care 8* (1), 36.
(64) Vabret, N., Britton, G. J., Gruber, C., Hegde, S., Kim, J., Kuksin, M., Levantovsky, R., Malle, L., Moreira, A., Park, M. D., Pia, L., Risson, E., Saffern, M., Salome, B., Esai Selvan, M., Spindler, M. P., Tan, J., van der Heide, V., Gregory, J. K., Alexandropoulos, K., Bhardwaj, N., Brown, B. D., Greenbaum, B., Gumus, Z. H., Homann, D., Horowitz, A., Kamphorst, A. O., Curotto de Lafaille, M. A., Mehandru, S., Merad, M., and Samstein, R. M. (2020) mmunology of COVID-19: Current State of the Science. *Immunity 52* (6), 910−941.
(65) Gu, J., Gong, E., Zhang, B., Zheng, J., Gao, Z., Zhong, Y., Zou, W., Zhan, J., Wang, S., Xie, Z., Zhuang, H., Wu, B., Zhong, H., Shao, H., Fang, W., Gao, D., Pei, F., Li, X., He, Z., Xu, D., Shi, X., Anderson, V. M., and Leong, A. S. Y. (2005) Multiple organ infection and the pathogenesis of SARS. *J. Exp. Med. 202* (3), 415−424.
(66) Zhao, G. J., Zheng, J. Y., Bian, J. L., Chen, L. W., Dong, N., Yu, Y., Hong, G. L., Chandoo, A., Yao, Y. M., and Lu, Z. Q. (2017) Growth Arrest-Specific 6 Enhances the Suppressive Function of CD4(+)CD25(+) Regulatory T Cells Mainly through Axl Receptor. *Mediators Inflammation 2017*, 6848430.
(67) Magg, T., Mannert, J., Ellwart, J. W., Schmid, I., and Albert, M. H. (2012) Subcellular localization of FOXP3 in human regulatory and nonregulatory T cells. *Eur. J. Immunol. 42* (6), 1627−38.
(68) Yang, Y., Xiong, Z., Zhang, S., Yan, Y., Nguyen, J., Ng, B., Lu, H., Brendese, J., Yang, F., Wang, H., and Yang, X.-F. (2005) Bcl-xL inhibits T-cell apoptosis induced by expression of SARS coronavirus E protein in the absence of growth factors. *Biochem. J. 392*, 135−143.
(69) Jagger, A., Shimojima, Y., Goronzy, J. J., and Weyand, C. M. (2014) Regulatory T cells and the immune aging process: a mini-review. *Gerontology 60* (2), 130−7.

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

(70) Ghosh, S., and Klein, R. S. (2017) Sex Drives Dimorphic Immune Responses to Viral Infections. *J. Immunol. 198* (5), 1782−1790.
(71) Meckiff, B. J., Ramírez-Suástegui, C., Fajardo, V., Chee, S. J., Kusnadi, A., Simon, H., Eschweiler, S., Grifoni, A., Pelosi, E., Weiskopf, D., Sette, A., Ay, F., Seumois, G. g., Ottensmeier, C. H., and Vijayanand, P. (2020) Imbalance of regulatory and cytotoxic SARS-CoV-2-reactive CD4+ T cells in COVID-19. *Cell 183* (5), 1340−1353.
(72) Tidcombe, H., Jackson-Fisher, A., Mathers, K., Stern, D. F., Gassmann, M., and Golding, J. P. (2003) Neural and mammary gland defects in ErbB4 knockout mice genetically rescued from embryonic lethality. *Proc. Natl. Acad. Sci. U. S. A. 100* (14), 8281.
(73) Liu, W., Purevdorj, E., Zscheppang, K., von Mayersbach, D., Behrens, J., Brinkhaus, M. J., Nielsen, H. C., Schmiedl, A., and Dammann, C. E. (2010) ErbB4 regulates the timely progression of late fetal lung development. *Biochim. Biophys. Acta, Mol. Cell Res. 1803* (7), 832−9.
(74) Logan, C. V., Lucke, B., Pottinger, C., Abdelhamed, Z. A., Parry, D. A., Szymanska, K., Diggle, C. P., van Riesen, A., Morgan, J. E., Markham, G., Ellis, I., Manzur, A. Y., Markham, A. F., Shires, M., Helliwell, T., Scoto, M., Hubner, C., Bonthron, D. T., Taylor, G. R., Sheridan, E., Muntoni, F., Carr, I. M., Schuelke, M., and Johnson, C. A. (2011) Mutations in MEGF10, a regulator of satellite cell myogenesis, cause early onset myopathy, areflexia, respiratory distress and dysphagia (EMARDD). *Nat. Genet. 43* (12), 1189−92.
(75) Ye, R., Gordillo, R., Shao, M., Onodera, T., Chen, Z., Chen, S., Lin, X., SoRelle, J. A., Li, X., Tang, M., Keller, M. P., Kuliawat, R., Attie, A. D., Gupta, R. K., Holland, W. L., Beutler, B., Herz, J., and Scherer, P. E. (2018) Intracellular lipid metabolism impairs N2 cell compensation during diet-induced obesity. *J. Clin. Invest. 128* (3), 1178−1189.
(76) Zhao, Y., Sun, X., Nie, X., Sun, L., Tang, T. S., Chen, D., and Sun, Q. (2012) COX5B regulates MAVS-mediated antiviral signaling through interaction with ATG5 and repressing ROS production. *PLoS Pathog. 8* (12), No. e1003086.
(77) Kulasegaran-Shylini, R., Thiviyanathan, V., Gorenstein, D. G., and Frolov, I. (2009) The 5′UTR-specific mutation in VEEV TC-83 genome has a strong effect on RNA replication and subgenomic RNA synthesis, but not on translation of the encoded proteins. *Virology 387* (1), 211−21.
(78) Kinney, R. M., Chang, G. J., Tsuchiya, K. R., Sneider, J. M., Roehrig, J. T., Woodward, T. M., and Trent, D. W. (1993) Attenuation of Venezuelan equine encephalitis virus strain TC-83 is encoded by the 5′-noncoding region and the E2 envelope glycoprotein. *J. Virol. 67* (3), 1269−77.
(79) Hardy, W. R., and Strauss, J. H. (1989) Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. *J. Virol. 63* (11), 4653−64.
(80) Strauss, E. G., De Groot, R. J., Levinson, R., and Strauss, J. H. (1992) Identification of the active site residues in the nsP2 proteinase of Sindbis virus. *Virology 191* (2), 932−40.
(81) Zhigalin, A. (2019) New data on David's myotis, Myotis davidii (Peters, 1869) (Mammalia, Chiroptera, Vespertilionidae), in Siberia and the Urals. *Biodiversity Data Journal 7*, No. e34211.
(82) Marz, M., Beerenwinkel, N., Drosten, C., Fricke, M., Frishman, D., Hofacker, I. L., Hoffmann, D., Middendorf, M., Rattei, T., Stadler, P. F., and Töpfer, A. (2014) Challenges in RNA virus bioinformatics. *Bioinformatics 30* (13), 1793−1799.
(83) Ibrahim, B., McMahon, D. P., Hufsky, F., Beer, M., Deng, L., Mercier, P. L., Palmarini, M., Thiel, V., and Marz, M. (2018) A new era of virus bioinformatics. *Virus Res. 251*, 86−90.
(84) Barral, P. M., Sarkar, D., Fisher, P. B., and Racaniello, V. R. (2009) RIG-I is cleaved during picornavirus infection. *Virology 391* (2), 171−6.
(85) Das, S., and Dasgupta, A. (1993) Identification of the cleavage site and determinants required for poliovirus 3CPro-catalyzed cleavage of human TATA-binding transcription factor TBP. *J. Virol. 67* (6), 3326−31.
(86) Kundu, P., Raychaudhuri, S., Tsai, W., and Dasgupta, A. (2005) Shutoff of RNA polymerase II transcription by poliovirus involves 3C protease-mediated cleavage of the TATA-binding protein at an alternative site: incomplete shutoff of transcription interferes with efficient viral replication. *J. Virol. 79* (15), 9702−13.
(87) Kuyumcu-Martinez, N. M., Van Eden, M. E., Younan, P., and Lloyd, R. E. (2004) Cleavage of poly(A)-binding protein by poliovirus 3C protease inhibits host cell translation: a novel mechanism for host translation shutoff. *Mol. Cell. Biol. 24* (4), 1779−90.
(88) de Breyne, S., Bonderoff, J. M., Chumakov, K. M., Lloyd, R. E., and Hellen, C. U. (2008) Cleavage of eukaryotic initiation factor eIF5B by enterovirus 3C proteases. *Virology 378* (1), 118−22.
(89) Yalamanchili, P., Banerjee, R., and Dasgupta, A. (1997) Poliovirus-encoded protease 2APro cleaves the TATA-binding protein but does not inhibit host cell RNA polymerase II transcription in vitro. *J. Virol 71* (9), 6881−6.
(90) Neznanov, N., Chumakov, K. M., Neznanova, L., Almasan, A., Banerjee, A. K., and Gudkov, A. V. (2005) Proteolytic cleavage of the p65-RelA subunit of NF-kappaB during poliovirus infection. *J. Biol. Chem. 280* (25), 24153−8.
(91) Wang, D., Fang, L., Li, K., Zhong, H., Fan, J., Ouyang, C., Zhang, H., Duan, E., Luo, R., Zhang, Z., Liu, X., Chen, H., and Xiao, S. (2012) Foot-and-mouth disease virus 3C protease cleaves NEMO to impair innate immune signaling. *Journal of Virology 86* (17), 9311−22.
(92) Foeger, N., Glaser, W., and Skern, T. (2002) Recognition of eukaryotic initiation factor 4G isoforms by picornaviral proteinases. *J. Biol. Chem. 277* (46), 44300−9.
(93) Yang, Y., Liang, Y., Qu, L., Chen, Z., Yi, M., Li, K., and Lemon, S. M. (2007) Disruption of innate immunity due to mitochondrial targeting of a picornaviral protease precursor. *Proc. Natl. Acad. Sci. U. S. A. 104* (17), 7253−8.
(94) Qu, L., Feng, Z., Yamane, D., Liang, Y., Lanford, R. E., Li, K., and Lemon, S. M. (2011) Disruption of TLR3 signaling due to cleavage of TRIF by the hepatitis A virus protease-polymerase processing intermediate, 3CD. *PLoS Pathog. 7* (9), No. e1002169.
(95) Xiang, Z., Li, L., Lei, X., Zhou, H., Zhou, Z., He, B., and Wang, J. (2014) Enterovirus 68 3C protease cleaves TRIF to attenuate antiviral responses mediated by Toll-like receptor 3. *J. Virol. 88* (12), 6650−9.
(96) Lei, X., Sun, Z., Liu, X., Jin, Q., He, B., and Wang, J. (2011) Cleavage of the adaptor protein TRIF by enterovirus 71 3C inhibits antiviral responses mediated by Toll-like receptor 3. *J. Virol 85* (17), 8811−8.
(97) Lei, X., Xiao, X., Xue, Q., Jin, Q., He, B., and Wang, J. (2013) Cleavage of Interferon Regulatory Factor 7 by Enterovirus 71 3C Suppresses Cellular Responses. *J. Virol. 87* (3), 1690−1698.
(98) Kuyumcu-Martinez, M., Belliot, G., Sosnovtsev, S. V., Chang, K. O., Green, K. Y., and Lloyd, R. E. (2004) Calicivirus 3C-like proteinase inhibits cellular translation by cleavage of poly(A)-binding protein. *J. Virol. 78* (15), 8172−82.
(99) Yu, C. Y., Chang, T. H., Liang, J. J., Chiang, R. L., Lee, Y. L., Liao, C. L., and Lin, Y. L. (2012) Dengue virus targets the adaptor protein MITA to subvert host innate immunity. *PLoS Pathog. 8* (6), No. e1002780.
(100) Ding, Q., Gaska, J. M., Douam, F., Wei, L., Kim, D., Balev, M., Heller, B., and Ploss, A. (2018) Species-specific disruption of STING-dependent antiviral cellular defenses by the Zika virus NS2B3 protease. *Proc. Natl. Acad. Sci. U. S. A. 115* (27), E6310−E6318.
(101) Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R., and Tschopp, J. (2005) Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature 437* (7062), 1167−72.
(102) Bellecave, P., Sarasin-Filipowicz, M., Donze, O., Kennel, A., Gouttenoire, J., Meylan, E., Terracciano, L., Tschopp, J., Sarrazin, C., Berg, T., Moradpour, D., and Heim, M. H. (2010) Cleavage of mitochondrial antiviral signaling protein in the liver of patients with

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX

chronic hepatitis C correlates with a reduced activation of the endogenous interferon system. *Hepatology 51* (4), 1127−36.
(103) Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr, and Lemon, S. M. (2005) Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc. Natl. Acad. Sci. U. S. A. 102* (8), 2992−7.
(104) Mohamud, Y., Xue, Y. C., Liu, H., Ng, C. S., Bahreyni, A., Jan, E., and Luo, H. (2021) The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochem. Biophys. Res. Commun. 540*, 75−82.
(105) Moustaqil, M., Ollivier, E., Chiu, H. P., Van Tol, S., Rudolffi-Soto, P., Stevens, C., Bhumkar, A., Hunter, D. J. B., Freiberg, A. N., Jacques, D., Lee, B., Sierecki, E., and Gambin, Y. (2021) SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerging Microbes Infect. 10* (1), 178−195.
(106) Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., Schulz, L., Widera, M., Mehdipour, A. R., Tascher, G., Geurink, P. P., Wilhelm, A., van der Heden van Noort, G. J., Ovaa, H., Müller, S., Knobeloch, K.-P., Rajalingam, K., Schulman, B. A., Cinatl, J., Hummer, G., Ciesek, S., and Dikic, I. (2020) Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature 587* (7835), 657−662.
(107) *Introduction to CRISPR and Cas9*. http://parts.igem.org/Collections/CRISPR (accessed 2017).
(108) Pfaffl, M. W. http://www.gene-quantification.de/siRNA-mechanism.png (accessed 2021).
(109) Bult, C. J, Blake, J. A, Smith, C. L, Kadin, J. A, Richardson, J. E, Anagnostopoulos, A, Asabor, R, Baldarelli, R M, Beal, J S, Bello, S M, Blodgett, O, Butler, N E, Christie, K R, Corbani, L E, Creelman, J, Dolan, M E, Drabkin, H J, Giannatto, S L, Hale, P, Hill, D P, Law, M, Mendoza, A, McAndrews, M, Miers, D, Motenko, H, Ni, L, Onda, H, Perry, M, Recla, J M, Richards-Smith, B, Sitnikov, D, Tomczuk, M, Tonorio, G, Wilming, L, and Zhu, Y (2019) Mouse Genome Database (MGD) 2019. *Nucleic Acids Res. 47* (D1), D801−D806.
(110) Bekes, M., van der Heden van Noort, G. J., Ekkebus, R., Ovaa, H., Huang, T. T., and Lima, C. D. (2016) Recognition of Lys48-Linked Di-ubiquitin and Deubiquitinating Activities of the SARS Coronavirus Papain-like Protease. *Mol. Cell 62* (4), 572−85.

https://doi.org/10.1021/acsinfecdis.0c00866
ACS Infect. Dis. XXXX, XXX, XXX−XXX