

# Multiple sites on SARS-CoV-2 spike protein are susceptible to proteolysis by cathepsins B, K, L, S, and V

Keval Bollavaram<sup>1</sup> | Tiffanie H. Leeman<sup>1</sup> | Maggie W. Lee<sup>1</sup> | Akhil Kulkarni<sup>1</sup> |  
Sophia G. Upshaw<sup>1</sup> | Jiabei Yang<sup>1,2</sup> | Hannah Song<sup>1</sup> | Manu O. Platt<sup>1</sup> 

<sup>1</sup>Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology & Emory University, Atlanta, Georgia

<sup>2</sup>Biomedical Engineering, Peking University, Beijing, China

## Correspondence

Manu O. Platt, Department of Biomedical Engineering, Georgia Institute of Technology & Emory University, 950 Atlantic Drive, Suite 3015, Atlanta, GA 30332, USA.  
Email: manu.platt@bme.gatech.edu

## Funding information

Science and Technology Center Emergent Behaviors of Integrated Cellular Systems, Grant/Award Number: CBET-0939511; National Science Foundation

## Abstract

SARS-CoV-2 is the coronavirus responsible for the COVID-19 pandemic. Proteases are central to the infection process of SARS-CoV-2. Cleavage of the spike protein on the virus's capsid causes the conformational change that leads to membrane fusion and viral entry into the target cell. Since inhibition of one protease, even the dominant protease like TMPRSS2, may not be sufficient to block SARS-CoV-2 entry into cells, other proteases that may play an activating role and hydrolyze the spike protein must be identified. We identified amino acid sequences in all regions of spike protein, including the S1/S2 region critical for activation and viral entry, that are susceptible to cleavage by furin and cathepsins B, K, L, S, and V using PACMANS, a computational platform that identifies and ranks preferred sites of proteolytic cleavage on substrates, and verified with molecular docking analysis and immunoblotting to determine if binding of these proteases can occur on the spike protein that were identified as possible cleavage sites. Together, this study highlights cathepsins B, K, L, S, and V for consideration in SARS-CoV-2 infection and presents methodologies by which other proteases can be screened to determine a role in viral entry. This highlights additional proteases to be considered in COVID-19 studies, particularly regarding exacerbated damage in inflammatory preconditions where these proteases are generally upregulated.

## KEYWORDS

cathepsin, computational modeling, COVID-19, extracellular matrix remodeling, inflammation, molecular docking, proteolysis, viral entry

## 1 | INTRODUCTION

Proteases are central to the infection process of SARS-CoV-2,<sup>1</sup> the coronavirus responsible for the COVID-19 pandemic.<sup>2</sup> Cleavage of the spike protein on the virus's capsid into S1/S2 subunits causes the conformational change that leads to membrane fusion and viral entry into the target cell.<sup>3</sup> This can happen at the cell surface through the action of the membrane bound protease,

TMPRSS2<sup>4</sup> and soluble, extracellular furin,<sup>5</sup> and also after endocytosis of the virion facilitating endolysosomal escape through proteolytic activity of cysteine cathepsins B and L<sup>6</sup> in acidic microenvironments.<sup>1</sup> These proteases have been the primary focus and implicated in SARS-CoV or SARS-CoV-2 infection,<sup>7</sup> but a number of other proteases may be contributing to spike protein activation, but have not yet been implicated.<sup>8</sup> Since inhibition of one protease may not be sufficient to block SARS-CoV-2 entry

into cells,<sup>4</sup> other proteases that may play an activating role and hydrolyze the spike protein must be identified.

Cathepsins B and L have already been implicated in cleavage of SARS-CoV-2 spike protein,<sup>4</sup> and there have been strong suggestions to prioritize cathepsin L inhibitors for COVID-19 therapies that have already been developed—some are even FDA approved.<sup>9</sup> Cathepsins K, S, and V share 60% sequence homology with cathepsin L with cathepsin V actually sharing 80% sequence homology with cathepsin L.<sup>10</sup> These cathepsins are also promiscuous with their substrates, suggesting cross-reactivity.<sup>11</sup> Here, we test the hypothesis that the spike protein might have sites susceptible to cleavage by cathepsins K, S, or V and suggest additional mechanisms by which proteases are involved in SARS-CoV-2 infection.

SARS-CoV-2 infection and severity of COVID-19 were shown to be elevated in patients with pre-existing conditions, resulting in a greater risk of mortality in those patients.<sup>12</sup> These pre-existing conditions include diseases that are associated with upregulation of cathepsins K, L, S, and V such as diabetes,<sup>13–15</sup> hypertension,<sup>16–18</sup> sickle cell disease,<sup>19,20</sup> cardiovascular disease,<sup>14,21</sup> and emphysema.<sup>22,23</sup> Together, this led us to investigate spike protein amino acid sequences preferred in the active site of these cathepsins for peptide bond hydrolysis.

Here, we used an unbiased approach, applying computational methods to determine the potential of other closely related cysteine cathepsins to bind to and cleave the spike protein. Computational analyses of the SARS-CoV-2 genome-encoded protease necessary for viral replication<sup>24</sup> suggested inhibitor molecules and strategies, but the focus here was to investigate human proteases that could also be targeted for inhibition. We used a program we previously developed called PACMANS, Protease-Ase Cleavages from MEROPS ANalyzed Specificities<sup>25</sup> to determine putative cleavage site locations on SARS-CoV-2 spike protein. This algorithm works with users' input of a substrate peptide's amino acid sequence for cleavage by an identified protease that is included in the MEROPS database.<sup>26</sup> PACMANS uses a sliding-window approach, where individual sub-sequences that fill the active site pocket are scored, based on amino acids preferred at the active site of the particular protease for maximal hydrolytic ability against a region on the input substrate to generate scores for preferred cleavage sites. This pocket slides along the length of the substrate amino acid sequence, scores all possible 8-amino acid sub-sequences, then ranks them by the normalized version of this score. PACMANS also calculates a normalized score to enable comparisons across multiple proteases on the same substrate. Coupling sequence analysis with molecular docking tools, together, are an

unbiased approach presented here to rationally identify other protease targets for new therapeutics against SARS-CoV-2, its infectivity, disease severity, and potentially longer-term complications.<sup>27</sup>

## 2 | RESULTS

### 2.1 | Bioinformatic analyses using the protein sequence of spike protein indicates susceptibility to multiple cysteine cathepsins beyond those previously identified

Using PACMANS, possible putative cleavage sites on SARS-CoV-2 spike protein, the substrate, were identified with furin and cathepsins B, K, L, S, and V as the proteases. Furin, and cathepsins B and L have already been identified to cleave spike protein in previous studies.<sup>4,5</sup> From PACMANS analysis, furin had the highest scoring cleavage site and identified its cleavage at the RRAR/SVAS site known to be the activating cleavage at S1/S2 site that causes conformational change and membrane fusion<sup>5</sup> (Table 1). This confirms that PACMANS does identify valid cleavage sites along the full length of the spike protein based solely on the amino acid sequence and its susceptibility to hydrolysis by these proteases. Cathepsin B, like furin, also prefers basic amino acid residues at its active site, and scored high at multiple sites on spike protein. Putative cleavage sites for cathepsins B, K, L, S, and V are included in the rank ordered list sorted by their normalized scores. Cleavage sites were identified in multiple domains of the spike protein (Table 1).

High ranking cleavage sites in the S1/S2 region was also identified for cathepsin K at amino acid 700 at the S2' site which was the 7 highest rank according to PACMANS. A number of sequences were identified as highly ranked for their susceptibility to multiple proteases, indicating a strong likelihood of cleavage by one or more proteases. Ranks indicate the preferred site on spike protein for each specific protease, so each protease has a rank#1 for highest preference in the table for cleaving spike protein. The normalized score allows for comparison across all the proteases for the same substrate. There were potential cleavage sites in all regions of the spike protein including the receptor binding domain, the S1/S2 cleavage domain for spike protein activation, N-terminus, C-terminus, and heptad repeat domains by cathepsins B, K, S, V, and L. Cathepsin B's highest score region was in the S1 domain. Cathepsins K and V had the same top ranked site in the heptad repeat region, cathepsin S was highest scored in the S2 region, and cathepsin L was in the receptor binding domain.

**TABLE 1** Results of PACMANS and docking analysis to identify preferred cleavage sites by multiple proteases on SARS-CoV-2 spike protein. Sequences for SARS-CoV-2 spike protein cleavage sites by six proteases were identified by PACMANS and sorted by their PACMANS normalized score. Ranks are for each individual protease on spike protein, based on their normalized score and likelihood of cleavage (lower the rank number, higher likelihood of cleavage). The “/” indicates the hydrolytic site with amino acids on either side. Based on amino acid number, the region of the spike protein is also identified. Receptor binding domain (RBD), N-terminal domain (N-Term), fusion peptide (FP), connector domain (CD), and heptad repeat (HR)

	Rank	Normal score	Starting A.A.	Sequence	Region of spike protein
FURIN	1	2.574	682	RRAR/SVAS	S1/S2 SITE
FURIN	2	1.992	454	RLFR/KSNL	RBD
FURIN	3	1.971	812	PSKR/SFIE	S2'
FURIN	4	1.699	75	GTKR/FDNP	N-TERM
FURIN	6	1.63	354	NRKR/ISNC	RBD
FURIN	8	1.61	355	RKRI/SNCV	RBD
FURIN	11	1.491	243	ALHR/SYLT	N-TERM
FURIN	13	1.478	343	NATRFASV	RBD
FURIN	14	1.452	762	QLNR/ALTG	S2
CATB	1	1.121	542	NFNG/LTGT	S1
CATB	6	1.046	520	APAT/VCGP	RBD
CATB	4	1.042	69	HVSG/TNGT	N-TERM
CATB	8	1.033	440	NLDS/KVGG	RBD
CATB	7	1.029	410	IAPG/QTGK	RBD
CATB	9	1.015	262	AAAYVGY	N-TERM
CATB	11	0.999	763	LNRA/LTGI	S2
CATB	12	0.998	498	QPTN/GVGY	RBD
CATB	14	0.982	826	VTLA/DAGF	FP
CATB	15	0.972	479	PCNG/VEGF	RBD
CATB	20	0.958	694	AYTM/SLGA	S1/S2
CATB	26	0.933	1,093	GVFV/SNGT	CD
CATK	1	0.928	943	SALG/KLQD	HR1
CATB	25	0.924	751	NLLL/QYGS	HR1
CATB	24	0.922	490	FPLQ/SYGF	RBD
CATV	1	0.914	943	SALG/KLQD	HR1
CATS	1	0.91	856	NGLT/VLPP	S2
CATB	37	0.908	792	PPIK/DFGG	S1/S2
CATS	2	0.899	516	ELLH/APAT	RBD
CATL	1	0.89	516	ELLH/APAT	RBD
CATK	2	0.884	1,139	DPLQ/PELD	CD
CATK	3	0.883	221	SALE/PLVD	N-TERM
CATB	43	0.882	793	PIKD/FGGF	S1/S2
CATB	32	0.877	766	ALTG/IAVE	S1/S2
CATS	5	0.877	936	DSLS/STAS	HR1
CATK	4	0.876	544	NGLT/GTGV	S1
CATL	2	0.871	876	ALLA/GTIT	S2
CATS	10	0.868	943	SALG/KLQD	HR1
CATB	49	0.864	425	LPDD/FTGC	RBD
CATV	2	0.863	516	ELLH/APAT	RBD

(Continues)

TABLE 1 (Continued)

	Rank	Normal score	Starting A.A.	Sequence	Region of spike protein
CATL	3	0.861	856	NGLT/VLPP	S2
CATK	5	0.861	946	GKLQDVVN	HR1
CATS	11	0.86	697	MSLGAENS	S1/S2
CATS	12	0.857	826	VTLA/DAGF	FP
CATV	3	0.851	856	NGLT/VLPP	S2
CATK	7	0.85	697	MSLG/AENS	S1/S2
CATK	8	0.845	187	KNLR/EFVF	N-TERM
CATB	57	0.843	441	LDSK/VGGN	RBD
CATK	10	0.843	765	RALT/GIAV	S2
CATL	5	0.84	943	SALG/KLQD	HR1
CATL	4	0.839	697	MSLG/AENS	S1/S2
CATB	65	0.832	470	TEIY/QAGS	RBD
CATK	12	0.83	46	SVLH/STQD	N-TERM
CATS	17	0.822	46	SVLH/STQD	N-TERM
CATB	70	0.82	496	GFQP/TNGV	RBD
CATS	25	0.812	221	SALE/PLVD	N-TERM
CATV	7	0.808	697	MSLGAENS	S1/S2
CATK	16	0.807	826	VTLA/DAGF	FP
CATL	7	0.803	936	DSLS/STAS	HR1
CATV	11	0.799	221	SALE/PLVD	N-TERM
CATL	8	0.796	765	RALT/GIAV	S2
CATV	13	0.79	46	SVLH/STQD	N-TERM
CATV	14	0.79	210	INLV/RDLP	N-TERM
CATL	10	0.777	46	SVLH/STQD	N-TERM
CATL	12	0.77	218	QGFS/ALEP	N-TERM
CATS	41	0.769	405	DEVV/QIAP	RBD
CATL	19	0.755	826	VTLA/DAGF	FP
CATB	106	0.751	398	DSFV/IRGD	RBD
CATL	23	0.745	210	INLV/RDLP	N-TERM
CATB	102	0.73	413	GQTG/KIAD	RBD
CATK	65	0.717	366	SVLY/NSAS	RBD
CATL	46	0.705	211	NLVR/DLPQ	N-TERM
CATB	121	0.696	336	CPFG/EVFN	RBD
CATV	82	0.681	116	SLLI/VNNA	N-TERM
CATL	81	0.673	116	SLLI/VNNA	N-TERM
CATL	87	0.669	1,073	KNFT/TAPA	CD

## 2.2 | Molecular docking analyses corroborate protease active site binding at identified cleavage sites

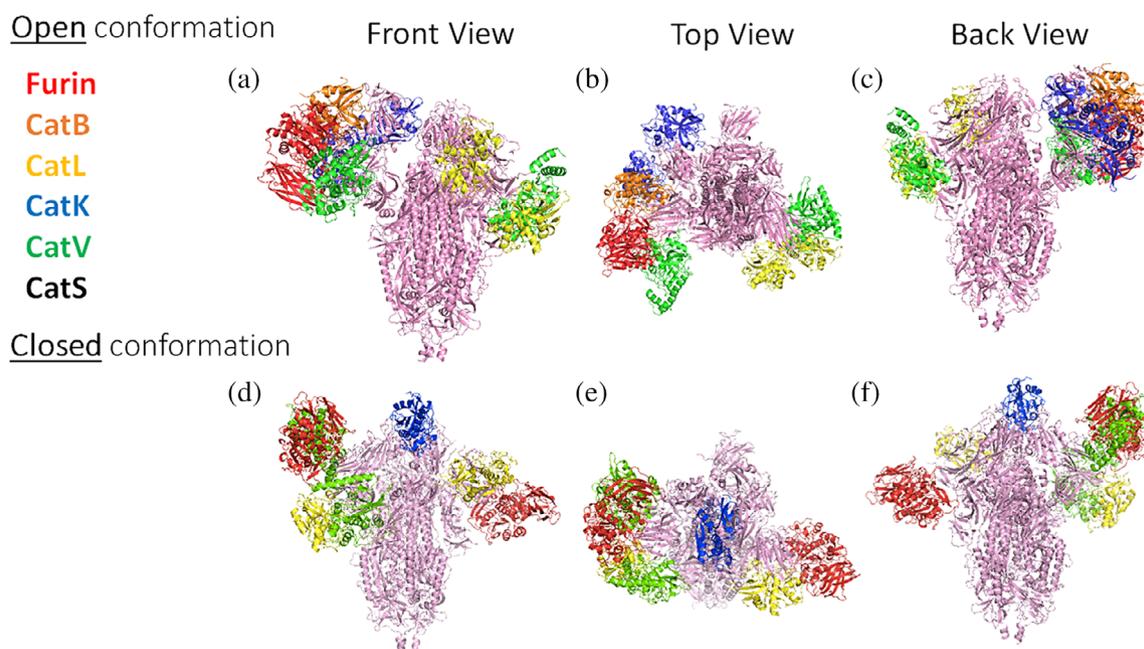
Since PACMANS only analyzes the sequence of amino acids, consideration for the three-dimensional structure

was helpful to determine access of the protease active sites to dock within close proximity of the spike protein sequence for cleavage is necessary. To assess this separately in an unbiased manner, furin, and cathepsins B, K, L, S, and V were used in molecular docking studies in ClusPro 2.0 to identify spike protein locations where the

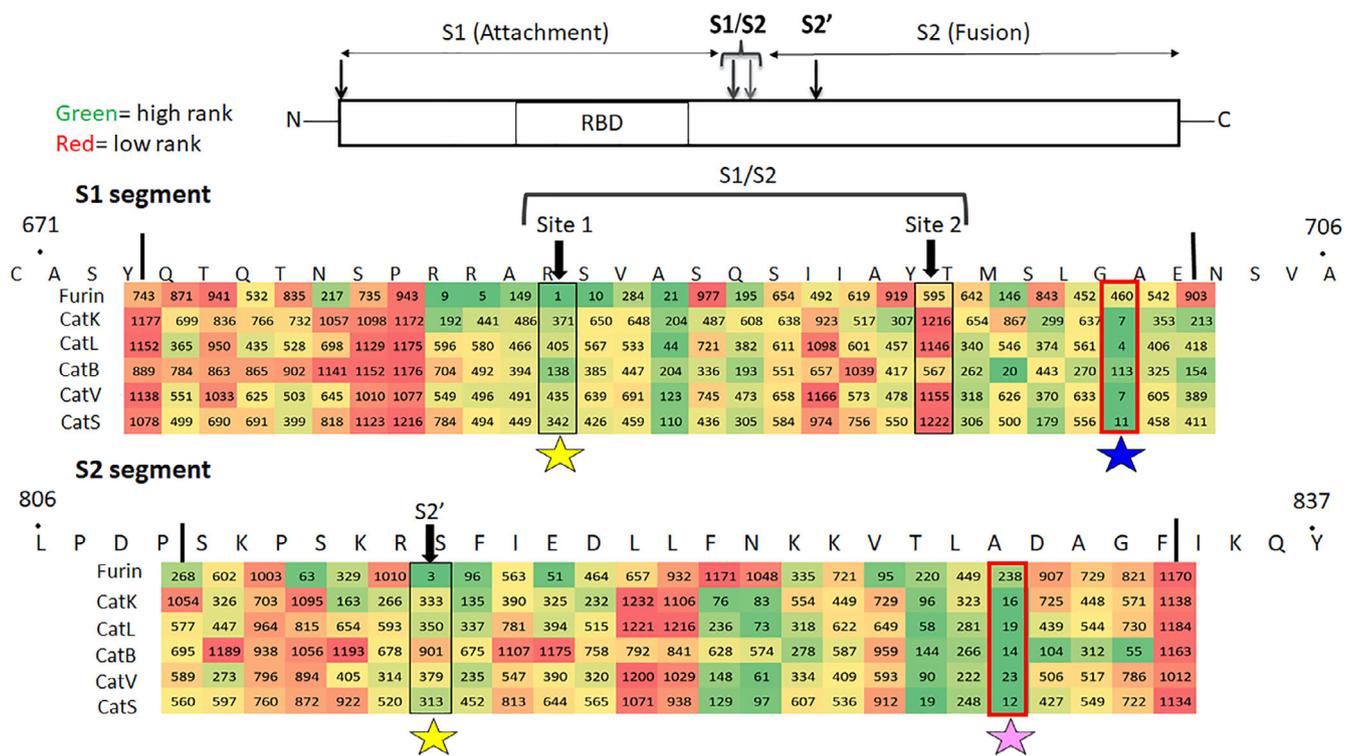
proteases' active site would be sufficiently close to the SARS-CoV-2 spike protein to bind. Priority was given to those models with the protease catalytic triad facing the SARS-CoV-2 spike protein with distances near 10 Å or less. Both open and closed conformations of the spike protein were used in the analysis. From the open conformation, molecular docking sites were compiled onto one spike protein trimer and shown from multiple viewpoints in Figure 1(a-c). Cathepsins L and K had the closest binding distances with cathepsin L being 3.6 Å from V213 (N-terminal domain) and 5.8 Å from A123 (N-terminal domain), and cathepsin K being 3.8 Å from A372 (receptor binding domain). Cathepsin B was 10 Å from E340 (receptor binding domain). Cathepsin V's closest distance was 12.2 Å from R214 (N-terminal domain) and 14.9 Å from A123 (N-terminal domain). From the closed conformation (Figure 1(d-f)), again, cathepsins K and L were the closest distances with cathepsin K only 6.2 Å from V367 (receptor binding domain), and cathepsin L being 5.1 Å from R214 (N-terminal domain) and 5.3 Å from A123 (N-terminal domain). Cathepsin V again was less close, but still produced calculated distances of 11 Å from R214 (N-terminal domain) and 10 Å from N122 (N-terminal domain).

### 2.3 | Putative cleavage of SARS-CoV-2-S protein at S1/S2 and S2'

The spike protein has two key domains, S1 subunit contains angiotensin I converting enzyme 2 (ACE2) receptor binding domain and S2 domain mediates membrane fusion.<sup>28</sup> S1/S2 site is the critical site where cleavage of spike protein by proteases causes a conformational change promoting virus entry to the host cells. The locations of putative maturation sites S1/S2 site 1 and site 2 as well as S2' on the SARS-CoV-2 spike protein sequence were detailed by Coutard et al.<sup>5</sup> The regions surrounding these cleavage sites were highlighted areas of analysis. From PACMANS, the rankings are represented by a heatmap with green indicating higher probability of cleavage and red indicating lower Furin ranked highest at the previously identified furin cleavage sites S1/S2 Site 1 and S2' (yellow stars), as expected (Figure 2). Cathepsins K, L, S, and V also shared high PACMANS rankings just outside the canonical S1/S2 Site 2, at MSLG/AENS (blue star), suggesting these proteases may also be involved in cleavage that might induce conformational change and membrane fusion. Additional sites such as VTLA/DAGF (pink star) also demonstrated high probabilities of cleavage across multiple cathepsins, with rankings in the top 20 for cathepsins B, K, L, S, and V (Figure 2).



**FIGURE 1** Closest molecular docking sites of each protease on SARS-CoV-2 spike protein. From the open conformation, molecular docking sites with bound proteases are compiled onto one spike protein trimer and shown from a front (a), top (b), and back (c) view. Furin: 12.2 Å from N81; Cat L: 3.6 Å from V213 and 5.8 Å from A123; Cat B: 10 Å from E340; Cat V: 12.2 Å from R214 and 14.9 Å from A123; Cat K: 3.8 Å from A372 and 11.1 Å from L226. From the closed conformation, docking is shown from a front (d), top (e), and back (f) view. Furin: 14.1 Å from S689 and 14.6 Å from N81; Cat K: 6.2 Å from V367; Cat L: 5.1 Å from R214 and 5.3 Å from A123; Cat V: 11 Å from R214 and 10 Å from N122



**FIGURE 2** Identification of putative cleavage sites in the S1/S2 region of spike protein based on PACMANS scoring. For each protease, spike protein cleavage sites in the S1 and S2 regions were ranked using PACMANS, and scores are shown with a heat map. Heat scores of ranks are colored from green to red, with green indicating a higher rank (closer to 1) and red indicating lower rank (greater numbers) that are less likely to be susceptible to cleavage by the specific protease. Boxed sequences indicate published cleavage sites of spike protein that promote conformational change for entry to cell. Stars indicate locations where multiple proteases have higher scores of cleavage probability: yellow for furin, blue for the cathepsins K, L, S, and V near S1/S2 site, and pink for cathepsins K, L, S, and V cleaving outside the S2' site

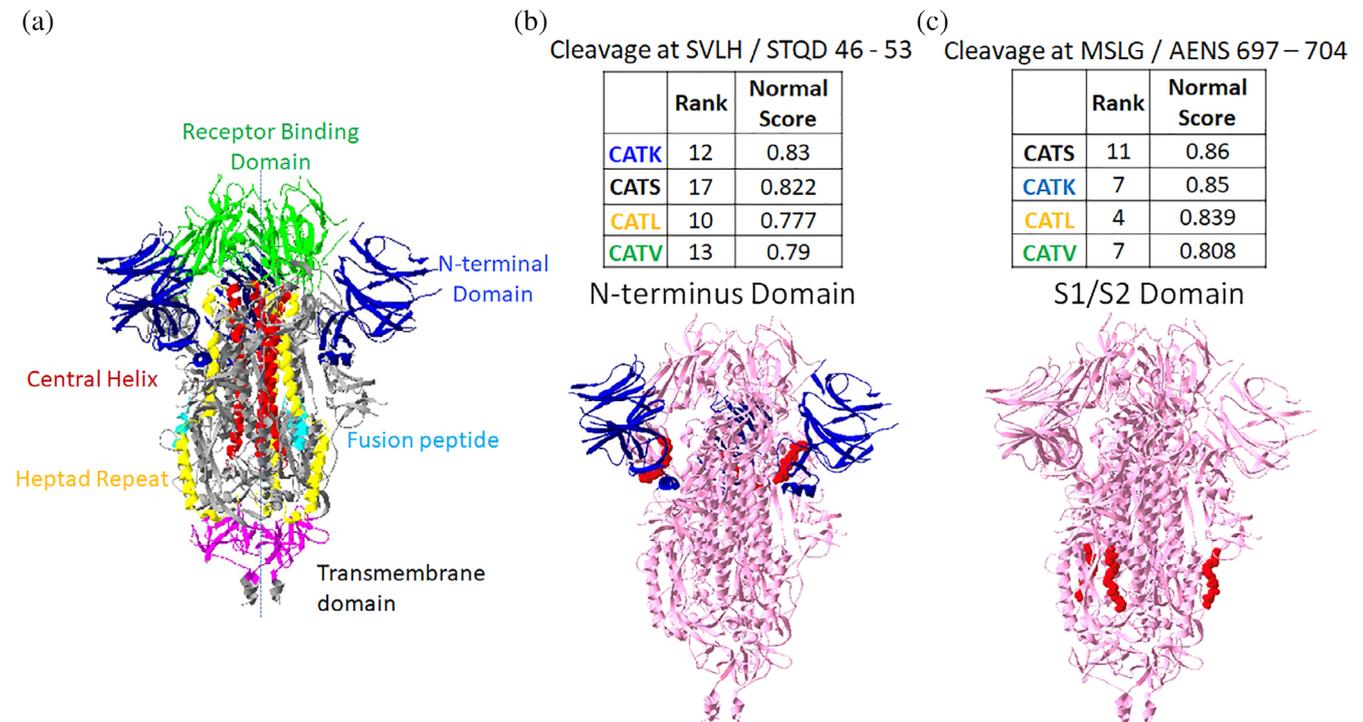
## 2.4 | Spike protein sequences susceptible to cleavage by multiple proteases

All four cathepsins, cathepsins K, S, L and V showed high ranks for two of the identified sequences on spike protein, so these were investigated more closely. SVLH/STQD (46–53) in N-terminus domain and MSLG/AENS (697–704) just outside the S2' site were the two sequences (Figure 3). Three of the cathepsins showed high ranked scores for other sequences. SALE/PLVD (221–228) was scored to be susceptible to cleavage by cathepsins S, V and K in the N-terminus domain (Figure 4(a)). Cathepsins S, L, and V also showed high ranks for ELLH/APAT (516–523) in the receptor binding domain, and NGLT/VLPP (856–863) in the S2 domain (Figure 4(b,c)).

## 2.5 | Susceptibility for inactivating cleavage of spike protein by multiple cathepsins

With multiple cathepsins having preferred cleavage sites on SARS-CoV-2 spike protein, we proposed the

hypothesis that cleavage by cathepsins at multiple sites could be inactivating cleavage events. For example, if two proteases hydrolyzed spike protein at adjacent regions, freeing a part of the protein, then the peptide fragment released could be destabilizing to protein conformation. Using high ranked sequences and molecular docking, cleavage events that would produce fragments were examined. We hypothesize that some of these cleavage events in the N-terminal domain and receptor binding domain might be inactivating, preventing virus binding or cell entry. Using these analyses, fragments to be released from the N-terminal domain were determined and their molecular weights predicted: S50 – V213 (19 kDa fragment), S50 – E224 (21 kDa fragment), S50 – I119 (8 kDa fragment) are in the N-terminal domain and included one of the sites susceptible to cleavage by cathepsins K, L, S, and V (Figure 5(a-c)). V120 – E224 (12 kDa fragment) is a sequence susceptible to cleavage by cathepsins S, V, and K. Cleavage of these sites and release of these fragments could affect the total protein structure in a destabilizing or unfolding manner that could reduce spike protein's ability to bind to the ACE2 receptor.



**FIGURE 3** Putative sites susceptible to cleavage by four cathepsins. From PACMANS analysis, there were two amino acid sequences susceptible to cleavage by four cathepsins as indicated by high normalized scores and relatively high rank orders. From the three-dimensional model of spike protein, the domains are color coded, (a) and susceptible sequences are highlighted in red. (b) Cleavage after H49 is in the N-terminus domain (blue) of spike protein, and (c) cleavage after G700 was even more highly ranked and cleaves in the S1/S2 domain

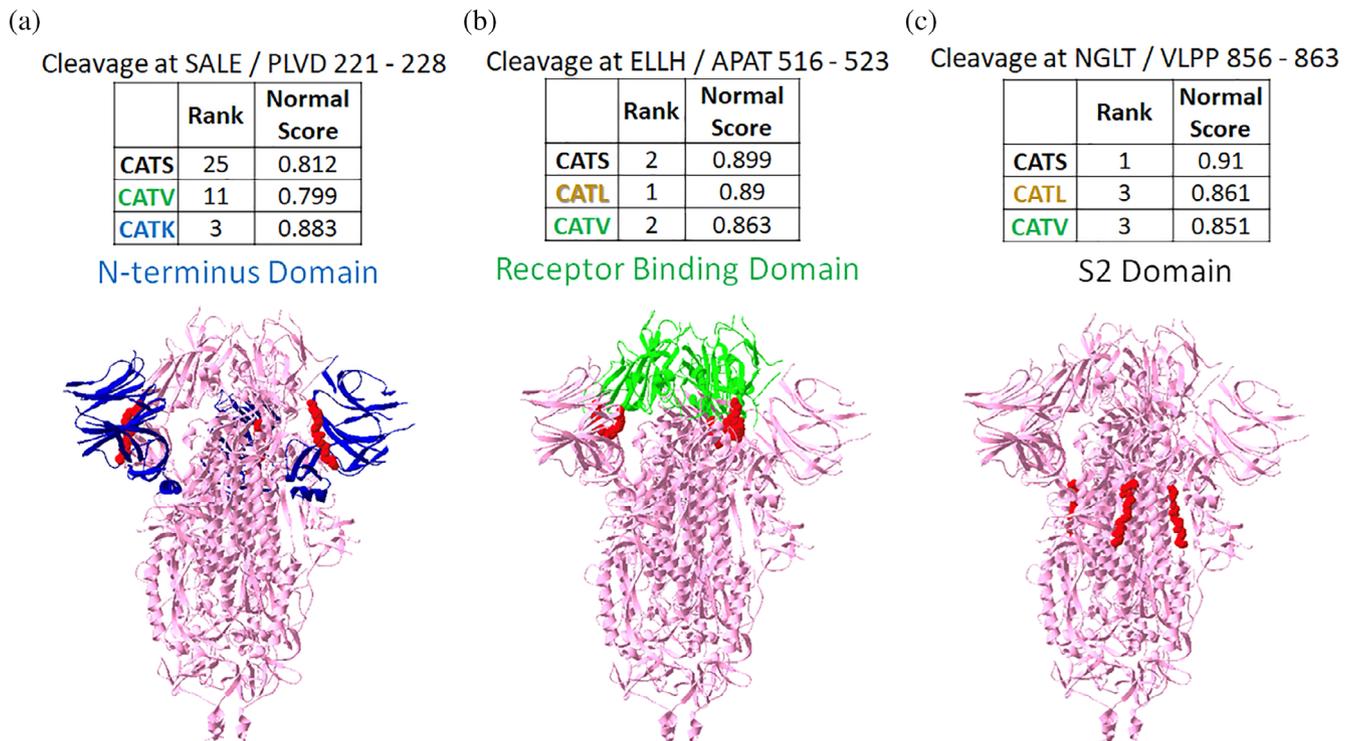
For cleavage in the S1/S2 region, again the site susceptible to cleavage by multiple cathepsins were analyzed. G700 - G769 (7 kDa fragment), G700 - V860 (17 kDa fragment), and T768 - V860 (10 kDa fragment) were modeled (Figure 6). Fragments are shown assuming the proteases are cleaving from the protomer of the spike protein, and from the assembled trimer of spike protein. This S1/S2 region is where proteolytic cleavage can be activating, meaning it promotes the conformational change that causes viral membrane fusion with cell membrane and subsequent viral entry. However, if there were multiple cleavage events freeing a fragment from the total spike protein, then the fusion event may not occur when these multiple cathepsins are present.

## 2.6 | SARS-CoV-2 spike protein cleavage by individual cathepsins generate unique cleavage fragments

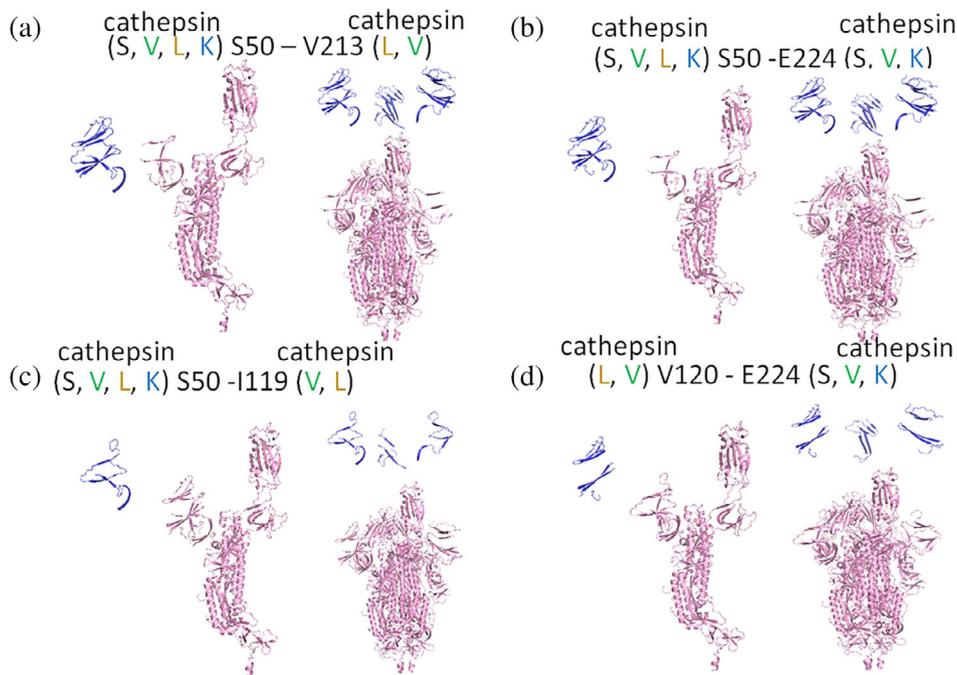
To assess the results from our computational analyses, we conducted experiments testing cleavage of spike protein by cathepsins B, K, L, S, and V. To do this, recombinant spike protein was incubated separately with each cathepsin for increasing periods of time in separate

aliquots. Then these samples were probed for immunoblots against SARS-CoV-2 spike protein to visualize cleavage fragments caused by cathepsin hydrolysis. Cathepsins B, K, L, S, and V generated different cleavage fragment patterns from full length spike protein (Figure 7). The cleavage products obtained by the individual cathepsins are unique which corroborate the implications from PACMANS that there are multiple cleavage sites on spike protein where multiple cysteine cathepsins can act, beyond just cathepsins B and L. From the blot, cathepsins B and L actually generated the fewest number of unique immunodetectable cleavage products whereas cathepsins K and V generated a number of uniquely visible cleavage products.

The fragment sizes generated reflect some of the predicted cleavage sizes suggested by PACMANS analysis. For example, cathepsin K was suggested to cleave spike protein at G700 and G769 which would generate three fragments: 7, 56, 78 kDa (Figure 6); a 78 kDa band appears in the blot. Cathepsin S was suggested to cleave spike protein at both G700 and V860 to release a 17 kDa fragment, but also 46 and 78 kDa fragments; while the 78 kDa fragment is not detected, there is a band at 46 kDa and the lower 17 kDa. CatV also has this 46 kDa band as it was suggested to cleave at both of these sites as



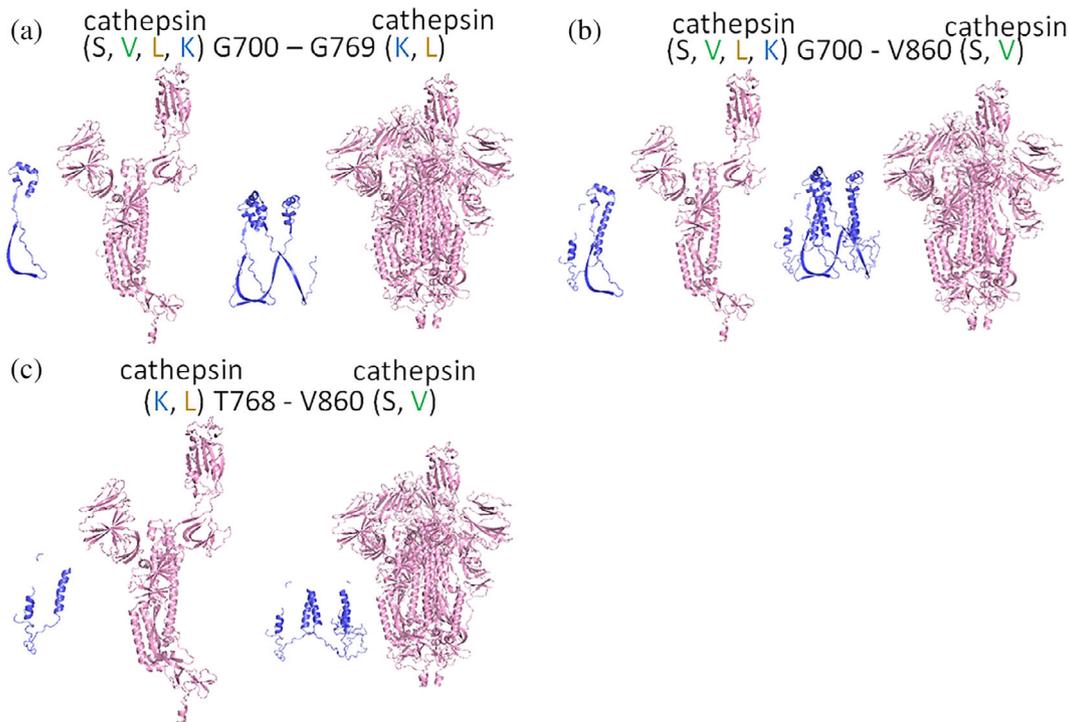
**FIGURE 4** Putative sites susceptible to cleavage by three cathepsins. From the PACMANS analysis, several eight amino acid sequences were susceptible to cleavage by three cathepsins as indicated by high normalized scores and relatively high rank orders. (a) Cleavage after E224 was in the N-terminus domain of spike protein (blue). (b) Cleavage after H519 was even more highly ranked and cleaves in the receptor binding domain (green). (c) Cleavage after T859 was in the S2 domain. Susceptible sequences are highlighted in red



**FIGURE 5** Fragments of spike protein generated due to cleavage at cathepsin putative sites in the N-terminal domain. From PyMol, fragments created from cleavage by multiple cathepsins at two sites on spike protein are shown for (a) S50 through V213, a 19 kDa fragment, (b) S50 through E224, a 21 kDa fragment, (c) S510 - I119, an 8 kDa fragment, and (d) V120 through E224, a 12 kDa fragment. Protomers are shown on the left and spike protein trimers are shown on the right for each cleavage site pair

well. Fragments below 20 kDa in size are present for all of the proteases tested except for cathepsin B. A number of other calculations of band sizes from cleavage fragments can be done, but the validation of specific

sequences will need to be done by mass spectrometry to specifically identify peptide fragments that, when re-assembled, recapitulate full length spike protein. However, these computational tools help motivate those



**FIGURE 6** Fragments of spike protein generated due to cleavage in S1/S2 domains. From the docking software, fragments created from cleavage by two or more cathepsins at two sites in the S1/S2 region of spike protein are shown for (a), G700 – G769, 7 kDa fragment (b), T768-V860, a 10 kDa fragment and (c) G700 – V860, a 17 kDa fragment. Fragments are shown from the protomer and from the trimer for each cleavage site pair

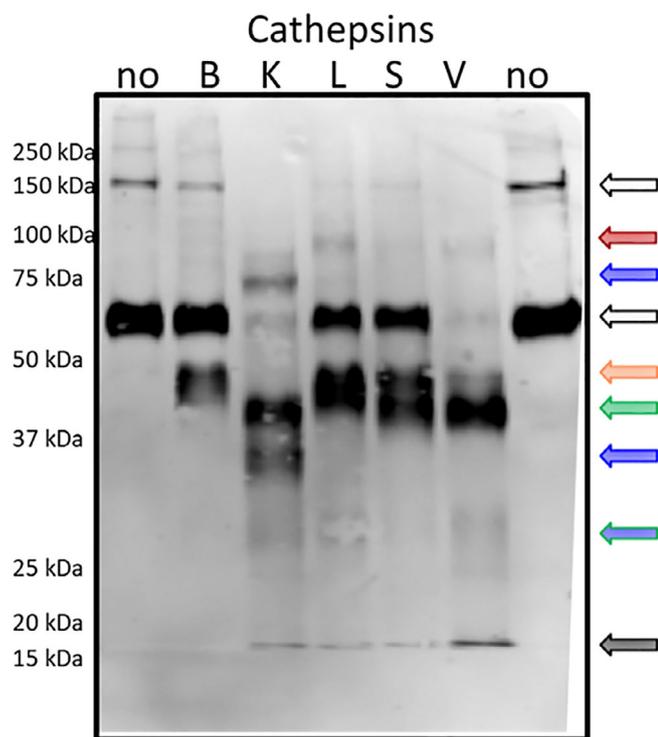
further analyses with suggestive hypotheses of multiple cathepsin cleavage events.

### 3 | DISCUSSION

From our computational-based PACMANS analysis and computational simulation of protein docking analysis, spike protein is susceptible to cleavage by cathepsins B, K, L, S, V, and furin at multiple sites, which can impact the infectivity and cell entry by SARS-CoV-2. This is the first implication of cathepsins K, S, and V for a role in COVID-19, but we have shown in this computational study that they can potentially cause a maturation cleavage event near the S2' site that facilitates membrane fusion. A table containing a rank ordered list of sites where cathepsins might cleave spike protein was generated, and we identified two locations that cathepsins K, L, S and V all showed high sequence cleavage preference. As given in Table 1 and Figures 3 and 4, there are various regions in the spike protein that are accessible and susceptible to cleavage by multiple cathepsins other than S1/S2 domains. N-terminus region and receptor binding domain cleavage sites were identified that could also inactivatingly hydrolyze the spike protein by removing fragments necessary for binding or for destabilizing the

spike protein structure. These computational predictions motivated experimental testing of spike protein cleavage by cathepsins B, K, L, S, and V, then immunoblotting confirmed a number of unique fragments of spike protein generated by cathepsin hydrolysis of the full length protein. Taken together, these results have generated a number of hypotheses that can be confirmed experimentally, but also can guide mutational analyses to investigate other potential loss or gain of function mutations in SARS-CoV-2 spike protein using the informed tools described in this study.

Cleavage of the spike protein can occur at the cell surface or in the endolysosomes which puts this cleavage event in two very different microenvironments. At the cell membrane, membrane bound protease, TMPRSS2, or other soluble proteases such as furin have been implicated, and in the acidic microenvironment, cathepsins B and L were shown to be key. While cathepsins play key roles in protein turnover in endosomes and lysosomes, they are also overexpressed and secreted in a number of diseases,<sup>11,29</sup> including those that are pre-existing conditions escalating risk of death due to COVID-19. These conditions could elevate the levels of other cathepsins working in the immediate extracellular spaces near the cell surface where they might facilitate spike protein activation or mitigate its actions by inactivatingly cleaving



**FIGURE 7** Individual cathepsins cleave SARS-CoV-2 spike protein generating unique cleavage fragments. 5 ng/ $\mu$ l of SARS-CoV-2 spike protein was incubated with 5 p mol of cathepsins B, K, L, S, V or no enzyme for increasing periods of time. Samples were collected and prepared for reducing immunoblot with anti-SARS-CoV-2 spike. Representative blot after 10 min of incubation is shown

spike protein. Indeed, cathepsin L has been shown to play a role in protecting from influenza infection through an intracellular mechanism, so there is precedent for proteases acting on both sides.<sup>30</sup> To determine this possibility, we undertook this study of an unbiased search for cathepsin cleavage sites along the spike protein, and then supported those findings with molecular docking analyses and recombinant protein cleavage analyses.

Identification of cathepsin K, L, S, and V cleavage sites in the S1/S2 fusion peptide region suggests their assistance in activating cleavage of spike protein to promote membrane fusion. In addition to the sites discussed, there are two more sequences that are highly susceptible to cleavage by cathepsins K, L, V, and S, in which cleavage sequences are located internally on the spike protein in either fusion peptide region (VLTA/DAGF) or heptad repeat region (SALG/KLQD). Cleavage in the fusion protein region of SARS-CoV by cathepsin L, not at the canonical site, still promoted membrane fusion and viral entry<sup>8</sup> so these locations could later be shown to affect spike protein function as well. As more is learned about the mechanisms of spike protein folding and membrane fusion, then the significance of these protein sequences,

and the impact of their proteolysis on either SARS-CoV-2 infectivity or function may become evident.

Cardiac injury has been an important factor associated with mortality of COVID-19.<sup>31</sup> ACE2, the cell receptor that binds spike protein, has previously been shown to be upregulated in cardiomyocytes in patients with cardiovascular disease,<sup>32</sup> and SARS-CoV-2 infection of induced pluripotent stem cell-derived cardiomyocytes has been confirmed.<sup>33</sup> In a separate study, RNA sequencing showed that ACE2 as well as cathepsins B and L were highly expressed in human induced pluripotent stem cell-derived cardiomyocytes, but interestingly, the protease TMPRSS2 was only expressed at low levels.<sup>34</sup> This shows that COVID-19 infection in cardiomyocytes may proceed through cathepsin-mediated pathways more so than TMPRSS2-mediated mechanisms; therefore, confirmation studies need to be completed. It is well reported that protease inhibitors reduce viral infection. FDA approved protease inhibitors used to suppress COVID-19 infection found that ritonavir was most effective at inhibiting SARS-CoV-2 main protease and human TMPRSS2. Cathepsins B and L were also capable of being blocked by indinavir and atazanavir.<sup>35</sup> Again, the other proteases implicated by the results of this study have not been tested but may also be targeted by these inhibitors as well, as cathepsin inhibitors have been notoriously cross-reactive.<sup>36–39</sup>

Based on the computational analysis of this study, sites were identified where proteolysis might be an inactivating cleavage event. These sites were shown in Figures 5 and 6 where adjacent cleavage sites, susceptible to proteolysis by multiple cathepsins, could generate fragments of the spike protein. Then the varying fragments generated by incubation of spike protein with the individual cathepsins, confirmed multiple cleavage events were possible, at least in vitro. Once these fragments were removed from spike and the viral capsid surface, then impaired binding to ACE2 receptor or impaired membrane fusion could be sufficient to prevent infection by that virion. Mature, active cathepsins are much smaller than spike protein (~25 vs. 140 kDa), so while it may be possible that multiple cathepsins could bind to one spike protein, it is more likely that with the distribution of spike proteins along the surface of the virus capsid, that cathepsins would be binding to and hydrolyzing individual spike proteins iteratively, but subsequent hydrolytic events by multiple proteases could lead to the results predicted in Figures 5 and 6 and then shown in Figure 7. This would also be dependent on concentrations of proteases present in the extracellular environment, in a tissue-specific manner, where virions are trying to attach to cell surfaces. Patients with cardiovascular disease showed dysregulated inflammation,<sup>40</sup> increased permeability,<sup>41</sup>

upregulated protease,<sup>42</sup> and epithelial dysfunction.<sup>43</sup> Severe inflammation, such as the cytokine storm triggered by COVID-19,<sup>44</sup> can also cause an influx of cysteine cathepsins produced by inflammatory macrophages, alveolar macrophages, endothelial cells, and smooth muscle cells, all of which produce cathepsins in response to cytokines such as TNF $\alpha$ , IFN $\gamma$ , IL-1 $\beta$ .<sup>19,45–47</sup>

Though we find this study to be useful, there are limitations to be considered. The results presented identify hypothetical protease cleavage sites on SARS-CoV-2-S, and spike protein hydrolysis by cathepsins K, S, and V was verified with recombinant *in vitro* studies. However, confirmation of the identity of these sequences to confirm they match with PACMANS predictions must be completed with mass spectrometry or site-directed mutagenesis for confirmation. The benefits of the computational approaches applied here are that a process has been presented of PACMANS followed by molecular docking to generate hypotheses for investigating other proteases, even beyond those included in this study. Upregulated plasmin, thrombin, and other serine proteases may be present at the site as well, implicated in the hypercoagulable state induced by SARS-CoV-2 infection.<sup>48</sup> Another example involves the D614G SARS-CoV-2 spike mutant that has been shown to be a more infectious virus with an enhanced susceptibility of spike protein to furin cleavage, while also altering the dynamics of the spike protein conformation.<sup>49</sup> This site was not identified in the unbiased approach used in the studies presented here, but if other mutations are identified, they can be input into PACMANS to determine their putative result on activating cleavage. Another study identified a cleavage site on spike protein from SARS-CoV, the virus that caused the SARS outbreak in 2009, at position T678,<sup>8</sup> but that site also was not highly ranked in the unbiased approaches used here. Despite these sites not appearing in the analysis, we are confident that the PACMANS analysis identifies cleavage sites as the furin site was the top ranked site, and additional cleavage sites that had been validated by other experimental approaches were identified from this analysis.

The extension from these analyses would be the further discovery of the role multiple cathepsins interacting, binding, and cleaving spike protein in activating or inactivating ways that can be taken in the aggregate to predict the responses in cells and tissues when exposed to SARS-CoV-2. We also have identified interactions of cathepsins with each other that remove them from the system, as they work in a proteolytic network and affect substrate degradation.<sup>50</sup> Together, these systems considerations of multiple cathepsins working with or against other proteases in the presence of spike protein during

SARS-CoV-2 infection may yield insight into optimizing pharmaceutical targeting to mitigate the effects.

## 4 | MATERIALS AND METHODS

### 4.1 | PACMANS analysis to identify putative cleavage sites

Protease-Ase Cleavage from MEROPS Analyzed Specificities (PACMANS) algorithm developed by Ferrall-Fairbanks et al. was utilized to rank potential cleavage sequences.<sup>25</sup> Protease specificity matrixes were obtained from the MEROPS peptidase database. PACMANS ranks each eight amino acid sequence in the substrate by its likelihood of hydrolysis by the selected protease.

### 4.2 | 3D molecular docking simulation

The possible binding of SARS-CoV-2-S protein (PDB ID: 6VYB) and furin (4Z2A), cathepsins L (5F02), cathepsin S (4P6E), cathepsin K (5TUN), cathepsin B (2IPP), and cathepsin V (3H6S) were simulated using Clus Pro 2.0, in which 3D protein crystal structures from protein data bank were automatically stripped of their water molecules and additional ligands before beginning docking simulations.<sup>51,52</sup> Each simulation yielded 15–30 possible docking models using a balanced binding interaction (electrostatic, hydrophobic, and Van der Waals) model. Top 10 simulation models were viewed in PyMOL followed by highlighting the active sites of catalytic triad of each protease. Among those, only the models that catalytic triad facing the SARS-CoV-2 S protein were further analyzed for possible cleavage. The distance was measured in Angstroms between the closest residue on the S protein and the catalytic triad. Distances of less than 10 Angstroms were considered of most interest.

### 4.3 | SARS-CoV-2-S protein domain color coding

The N terminal domain, receptor binding domain, fusion peptide, heptad repeat 1, central helix, and connector domains of each subunit of the SARS-CoV-2-S protein were identified in SwissPDB by comparison of the SARS-CoV-2 PDB sequence and the domain sequences outlined by Wrapp et al.<sup>53</sup> Each domain was colored in the SwissPDB model to visualize potential cleavage site locations in relation to these domains.

## 4.4 | Cathepsin hydrolysis of SARS-CoV-2 spike protein and immunoblotting

A 5 p mol of recombinant human cathepsins B, K, L, S, and V (Enzo) was incubated with 5 ng/ $\mu$ l of SARS-CoV-2 (2019-nCoV) Spike S1 + S2 ECD-His Recombinant Protein (Sino Biological #40589-V08B1) in phosphate buffer (pH 6.0), 2 mM DTT, 1 mM EDTA for at 37°C for defined time periods. Reducing SDS-PAGE loading buffer was added and then heated at 95°C to terminate experiments. After SDS-PAGE, protein was transferred to a nitrocellulose membrane (Bio-Rad) and probed with SARS-CoV/SARS-CoV-2 Spike Protein S2 Monoclonal Antibody (1A9) (ThermoFisher # MA5-35946) and donkey anti-mouse secondary antibodies tagged with an infrared fluorophore (Rockland Immunochemicals). Membranes were imaged with a LI-COR Odyssey scanner.

### ACKNOWLEDGMENTS

This work was supported by the National Science Foundation through Science and Technology Center Emergent Behaviors of Integrated Cellular Systems (EBICS) Grant CBET-0939511 (Manu O. Platt).

### AUTHOR CONTRIBUTIONS

**Keval Bollavaram:** Formal analysis; investigation; methodology; writing-review & editing. **Tiffany H. Leeman:** Formal analysis; investigation; methodology; writing-original draft; writing-review & editing. **Maggie W. Lee:** Formal analysis; investigation; methodology; writing-review & editing. **Akhil Kulkarni:** Formal analysis; investigation; writing-review & editing. **Sophia G. Upshaw:** Formal analysis; investigation. **Jiabei Yang:** Investigation; writing-original draft; writing-review & editing. **Hannah Song:** Conceptualization; formal analysis; investigation; methodology; project administration; supervision; writing-original draft; writing-review & editing. **Manu O. Platt:** Conceptualization; formal analysis; funding acquisition; methodology; project administration; writing-review & editing.

### ORCID

Manu O. Platt  <https://orcid.org/0000-0001-6390-4320>

### REFERENCES

- Ou X, Liu Y, Lei X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun.* 2020;11:1620.
- Zhu N, Zhang D, Wang W, et al. China novel coronavirus I, research T. a novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* 2020;382:727–733.
- Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol.* 2016;3:237–261.
- Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* 2020;181:271–280.
- Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* 2020;176:104742.
- Huang IC, Bosch BJ, Li F, et al. SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. *J Biol Chem.* 2006;281:3198–3203.
- Simmons G, Zmora P, Gierer S, Heurich A, Pohlmann S. Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. *Antiviral Res.* 2013;100:605–614.
- Bosch BJ, Bartelink W, Rottier PJ. Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. *J Virol.* 2008;82:8887–8890.
- Liu T, Luo S, Libby P, Shi GP. Cathepsin L-selective inhibitors: A potentially promising treatment for COVID-19 patients. *Pharmacol Ther.* 2020;213:107587.
- Wilder CL, Park KY, Keegan PM, Platt MO. Manipulating substrate and pH in zymography protocols selectively distinguishes cathepsins K, L, S, and V activity in cells and tissues. *Arch Biochem Biophys.* 2011;516:52–57.
- Turk V, Stoka V, Vasiljeva O, et al. Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim Biophys Acta.* 2012;1824:68–88.
- Guan WJ, Liang WH, Zhao Y, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: A nationwide analysis. *Eur Respir J.* 2020;55:17650.
- Hie M, Shimono M, Fujii K, Tsukamoto I. Increased cathepsin K and tartrate-resistant acid phosphatase expression in bone of streptozotocin-induced diabetic rats. *Bone.* 2007;41:1045–1050.
- Liu J, Ma L, Yang J, et al. Increased serum cathepsin S in patients with atherosclerosis and diabetes. *Atherosclerosis.* 2006;186:411–419.
- Garsen M, Rops AL, Dijkman H, et al. Cathepsin L is crucial for the development of early experimental diabetic nephropathy. *Kidney Int.* 2016;90:1012–1022.
- Hua Y, Xu X, Shi GP, Chicco AJ, Ren J, Nair S. Cathepsin K knockout alleviates pressure overload-induced cardiac hypertrophy. *Hypertension.* 2013;61:1184–1192.
- Cheng XW, Murohara T, Kuzuya M, et al. Superoxide-dependent cathepsin activation is associated with hypertensive myocardial remodeling and represents a target for angiotensin II type 1 receptor blocker treatment. *Am J Pathol.* 2008;173:358–369.
- Lu Y, Sun X, Peng L, et al. Angiotensin II-induced vascular remodeling and hypertension involves cathepsin L/V- MEK/-ERK mediated mechanism. *Int J Cardiol.* 2020;298:98–106.
- Keegan PM, Surapaneni S, Platt MO. Sickle cell disease activates peripheral blood mononuclear cells to induce cathepsins k and v activity in endothelial cells. *Anemia.* 2012;2012:201781.
- Song H, Keegan PM, Anbazhakan S, et al. Sickle cell anemia mediates carotid artery expansive remodeling that can be prevented by inhibition of JNK (c-Jun N-terminal kinase). *Arterioscler Thromb Vasc Biol.* 2020;40:1220–1230.

21. Liu CL, Guo J, Zhang X, Sukhova GK, Libby P, Shi GP. Cysteine protease cathepsins in cardiovascular disease: From basic research to clinical trials. *Nat Rev Cardiol*. 2018;15:351–370.
22. Zheng T, Zhu Z, Wang Z, et al. Inducible targeting of IL-13 to the adult lung causes matrix metalloproteinase- and cathepsin-dependent emphysema. *J Clin Invest*. 2000;106:1081–1093.
23. Dai R, Wu Z, Chu HY, et al. Cathepsin K: The action in and beyond bone. *Front Cell Dev Biol*. 2020;8:433.
24. Chen YW, Yiu CB, Wong KY. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL [pro]) structure: Virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Res*. 2020;9:129.
25. Ferrall-Fairbanks MC, Barry ZT, Affer M, Shuler MA, Moomaw EW, Platt MO. PACMANS: A bioinformatically informed algorithm to predict, design, and disrupt protease-on-protease hydrolysis. *Protein Sci*. 2017;26:880–890.
26. Rawlings ND, Barrett AJ, Bateman A. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*. 2012;40:D343–D350.
27. Becker RC. Anticipating the long-term cardiovascular effects of COVID-19. *J Thromb Thrombolysis*. 2020;50:512–524.
28. Xiu S, Dick A, Ju H, et al. Inhibitors of SARS-CoV-2 entry: Current and future opportunities. *J Med Chem*. 2020;63:12256–12274.
29. Vasiljeva O, Reinheckel T, Peters C, Turk D, Turk V, Turk B. Emerging roles of cysteine cathepsins in disease and their potential as drug targets. *Curr Pharm Des*. 2007;13:387–403.
30. Xu X, Greenland JR, Gotts JE, Matthay MA, Caughey GH. Cathepsin L helps to defend mice from infection with influenza A. *PLoS One*. 2016;11:e0164501.
31. Shi S, Qin M, Shen B, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. *JAMA Cardiol*. 2020;5:802–810.
32. Nicin L, Abplanalp WT, Mellentin H, et al. Cell type-specific expression of the putative SARS-CoV-2 receptor ACE2 in human hearts. *Eur Heart J*. 2020;41:1804–1806.
33. Perez-Bermejo JA, Kang S, Rockwood SJ, et al. SARS-CoV-2 infection of human iPSC-derived cardiac cells reflects cytopathic features in hearts of patients with COVID-19. *Science Translational Medicine*, 2021:eabf7872. <http://doi.org/10.1126/scitranslmed.abf7872>.
34. Bojkova D, Wagner JUG, Shumliakivska M, et al. SARS-CoV-2 infects and induces cytotoxic effects in human cardiomyocytes. *Cardiovasc Res*. 2020;116:2207–2215.
35. Abhinand CS, Nair AS, Krishnamurthy A, Oommen OV, Sudhakaran PR. Potential protease inhibitors and their combinations to block SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, 2020;1–15. <http://doi.org/10.1080/07391102.2020.1819881>.
36. Lee-Dutra A, Wiener DK, Sun S. Cathepsin S inhibitors: 2004–2010. *Expert Opin Ther Patents*. 2011;21:311–337.
37. Li YY, Fang J, Ao GZ. Cathepsin B and L inhibitors: A patent review (2010 - present). *Expert Opin Ther Patents*. 2017;27:643–656.
38. Mullard A. Merck & Co. drops osteoporosis drug odanacatib. *Nat Rev Drug Discov*. 2016;15:669.
39. Platt MO, Shockey WA. Endothelial cells and cathepsins: Biochemical and biomechanical regulation. *Biochimie*. 2016;122:314–323.
40. Williams JW, Huang LH, Randolph GJ. Cytokine circuits in cardiovascular disease. *Immunity*. 2019;50:941–954.
41. Sedding DG, Boyle EC, Demandt JAF, et al. Vasa vasorum angiogenesis: Key player in the initiation and progression of atherosclerosis and potential target for the treatment of cardiovascular disease. *Front Immunol*. 2018;9:706.
42. Wu H, Du Q, Dai Q, Ge J, Cheng X. Cysteine protease cathepsins in atherosclerotic cardiovascular diseases. *J Atheroscler Thromb*. 2018;25:111–123.
43. Barthelmes J, Nagele MP, Ludovici V, Ruschitzka F, Sudano I, Flammer AJ. Endothelial dysfunction in cardiovascular disease and Flammer syndrome-similarities and differences. *EPMA J*. 2017;8:99–109.
44. De la Rica R, Borges M, Gonzalez-Freire M. COVID-19: In the eye of the cytokine storm. *Front Immunol*. 2020;11:558898.
45. Sukhova GKSG, Simon DI, Chapman HA, Libby P. Expression of the elastolytic cathepsins S and K in human atheroma and regulation of their production in smooth muscle cells. *J Clin Invest*. 1998;102:576–583.
46. Keegan PM, Wilder CL, Platt MO. Tumor necrosis factor alpha stimulates cathepsin K and V activity via juxtacrine monocyte-endothelial cell signaling and JNK activation. *Mol Cell Biochem*. 2012;367:65–72.
47. Shi GP, Sukhova GK, Grubb A, et al. Cystatin C deficiency in human atherosclerosis and aortic aneurysms. *J Clin Invest*. 1999;104:1191–1197.
48. Ji HL, Zhao R, Matalon S, Matthay MA. Elevated plasmin(ogen) as a common risk factor for COVID-19 susceptibility. *Physiol Rev*. 2020;100:1065–1075.
49. Plante JA, Liu Y, Liu J, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*. 2020;592:116–121.
50. Ferrall-Fairbanks MC, Kieslich CA, Platt MO. Reassessing enzyme kinetics: Considering protease-as-substrate interactions in proteolytic networks. *Proc Natl Acad Sci U S A*. 2020;117:3307–3318.
51. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein-protein docking. *Nat Protoc*. 2017;12:255–278.
52. Ferrall-Fairbanks MC, West DM, Douglas SA, Averett RD, Platt MO. Computational predictions of cysteine cathepsin-mediated fibrinogen proteolysis. *Protein Sci*. 2018;27:714–724.
53. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020;367:1260–1263.

**How to cite this article:** Bollavaram K, Leeman TH, Lee MW, et al. Multiple sites on SARS-CoV-2 spike protein are susceptible to proteolysis by cathepsins B, K, L, S, and V. *Protein Science*. 2021;30:1131–1143. <https://doi.org/10.1002/pro.4073>