Novel sites for Cathepsin L cleavage in SARS-CoV-2 spike guide treatment strategies

Jin-Kui Yang (✉ jkyang@ccmu.edu.cn)

Beijing Tongren Hospital, Capital Medical University <https://orcid.org/0000-0002-5430-2149>

Miao-Miao Zhao

Beijing Tongren Hospital, Capital Medical University

Yun Zhu

Institute of Biophysics, Chinese Academy of Sciences <https://orcid.org/0000-0001-9382-8592>

Li Zhang

National Institutes for Food and Drug Control

Gongxun Zhong

State key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute

<https://orcid.org/0000-0002-0748-9641>

Linhua Tai

<https://orcid.org/0000-0001-8435-9683>

Shuo Liu

National Institutes for Food and Drug Control

Guoliang Yin

Institute of Biophysics, Chinese Academy of Sciences

Weijin Huang

National Institutes for Food and Drug Control (NIFDC) <https://orcid.org/0000-0002-4246-8889>

Changfa Fan

National Institutes for Food and Drug Control

Lei Shuai

Harbin Veterinary Research Institute, CAAS <https://orcid.org/0000-0003-2122-8365>

Zhiyuan Wen

Harbin Veterinary Research Institute, CAAS <https://orcid.org/0000-0003-3112-0243>

Chong Wang

Harbin Veterinary Research Institute, CAAS <https://orcid.org/0000-0002-3970-9071>

Xijun He

Harbin Veterinary Research Institute, CAAS

Zhigao Bu

Harbin Veterinary Research Institute, CAAS <https://orcid.org/0000-0001-9242-4211>

Youchun Wang

National Institutes for Food and Drug Control <https://orcid.org/0000-0001-9769-5141>

Fei Sun

Institute of Biophysics, Chinese Academy of Sciences <https://orcid.org/0000-0002-0351-5144>

Biological Sciences - Article

Keywords: SARS-CoV-2, COVID-19, vaccines, protease Cathepsin L (CTSL)

Posted Date: July 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-734963/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Cell Discovery on June 6th, 2022. See the published version at <https://doi.org/10.1038/s41421-022-00419-w>.

Abstract

The spike (S) protein of SARS coronavirus 2 (SARS-CoV-2) is an ideal target for the development of specific vaccines or drugs. However, treatments targeting viruses with mutant S proteins that have recently emerged in many countries are limited. Cleavage of the S protein by host proteases is essential for viral infection. Here, we discovered two novel sites (CS-1 and CS-2) in the S protein for cleavage by the protease Cathepsin L (CTSL). Both sites are highly conserved among all SARS-CoV-2 variants of concern. Cryo-electron microscopy structural studies revealed that CTSL cleavage increases the dynamics of the receptor binding domain of S and induces novel conformations. In our pseudovirus (PsV) infection experiment, alteration of the cleavage site significantly reduced the infection efficiency, and CTSL inhibitors markedly inhibited infection with PsVs of both the wild-type and emerged SARS-CoV-2 variants. Furthermore, six highly efficient CTSL inhibitors were found to effectively inhibit live virus infection in human cells *in vitro*, and two of these were further confirmed to prevent live virus infection in human *ACE2* transgenic mice *in vivo*. Our work suggested that the CTSL cleavage sites in SARS-CoV-2 S are emerging new but effective targets for the development of mutation-resistant vaccines and drugs.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic, caused by the novel SARS coronavirus 2 (SARS-CoV-2), has resulted in a global public health crisis. Most treatment strategies for SARS-CoV-2 infection are focused on vaccines or antiviral drugs targeting the viral spike (S) protein^{1,2}. However, the evolution of RNA viruses is driven by their high mutation rates, allowing their evasion of host immune attacks^{3,4}. Despite multiple ongoing efforts to treat SARS-CoV-2 infections, a new wave of infections caused by S mutant strains has emerged in many countries and has exhibited high morbidity³⁻⁶. Therefore, the development of antiviral drugs or vaccines targeting the S protein may encounter the problem of resistant mutations. Indeed, three vaccines (Janssen, Novavax, and AstraZeneca) tested against the B.1.351 variant exhibited reduced effects on preventing symptomatic infection, with effectiveness estimates of 57%, 49% and even a statistically insignificant percentage, respectively⁷. The serum neutralizing activity against the B.1.351 variant among vaccinated persons was reduced by 6.5- and 8.6-fold for the BNT162b2 (Pfizer) and mRNA-1273 (Moderna) mRNA vaccines, respectively⁸. Thus, there is an urgent need to develop broad-spectrum treatment strategies for SARS-CoV-2 infection.

During viral infection, the S protein requires activation by host cell proteases, such as furin, TMPRSS2 and cathepsin L (CTSL)⁹⁻¹¹. CTSL is a member of the lysosomal cysteine protease family, whose major function is proteolysis of antigens produced by pathogens. Recently, we found that in SARS-CoV-2-infected patients, the circulating level of CTSL in blood samples is highly correlated with the severity and course of COVID-19. We also found that CTSL promotes SARS-CoV-2 pseudovirus (PsV) infection by cleaving the S protein and enhancing viral entry into cells⁹. However, the precise cleavage site and mechanism by which CTSL activates the S protein, as well as the effectiveness of CTSL inhibitors in preventing or treating infections with live SARS-CoV-2, remain unknown. In this study, we identified two

novel specific CTSL cleavage sites in SARS-CoV-2 S, which are highly conserved among all SARS-CoV-2 variants of concern. Combining cryo-electron microscopy (cryo-EM) and functional experiments, including treatment of live virus infection, we proved that the novel CTSL cleavage sites in SARS-CoV-2 S provide promising therapeutic targets for the development of drugs to overcome viral mutations.

Results

CTSL cleaves the SARS-CoV-2 S protein at two novel sites

The trimeric S protein incorporated into the SARS-CoV-2 envelope contains S1 and S2 subunits (Fig. 1a and 1b). The S1 subunit binds to the host cellular receptor, while the S2 subunit is involved in the virus-cell membrane fusion process¹², which is followed by the release of viral genetic material into target cells¹³. To study the structural and functional changes in the S protein upon treatment with CTSL, we cloned the mammalian codon-optimized nucleotide sequence encoding the SARS-CoV-2 (Wuhan-Hu-1 strain, GenBank ID: MN908947.3) S protein ectodomain (residues M1–Q1208) with proline substitutions at K986 and V987 and a “GSAS” substitution at the furin cleavage site (S1/S2 site, R682 to R685) and purified the resulting protein. This mutant S protein is widely used for structural and functional analysis due to its enhanced stability¹⁴. After coincubation with CTSL, the purified S protein was cleaved into two major fragments, and the cleavage efficiency exhibited a dose-dependent relationship with the CTSL concentration (Fig. 1c). Using N-terminal amino acid sequencing of these two fragments, we identified two novel CTSL cleavage sites in S: 259T (named CTSL cleavage site 1 or CS-1) and 636Y (named CTSL cleavage site 2 or CS-2) (Fig. 1c and Supplementary Fig. 1a-c). CS-1 is located in the N-terminal domain (NTD) of the S1 subunit, while CS-2 is located in the C-terminal domain (CTD) of the S1 subunit (Fig. 1a). Similar to the furin cleavage site at S1/S2, both CTSL cleavage sites are in the exposed loops in the prefusion S protein, which are accessible to proteases (Fig. 1b). Liquid chromatography–mass spectrometry (LC-MS) analysis of the fragments also proved the accuracy of sequence identification (Supplementary Fig. 1d and e).

Next, to verify these two cleavage sites, we generated a mutant S protein with point mutations in CS-1 and CS-2. Since it has been reported that substrate recognition by CTSL is determined mainly by the cleavage site residue (named the P1 residue) and its adjacent upstream residue (named the P2 residue)¹⁵, we mutated both the P1 and P2 residues in CS-1 and CS-2 to glycine to generate 4-Gly-mutant S (Fig. 1e). After coincubation of CTSL and 4-Gly-mutant S, the two cleaved fragments could not be observed, confirming the specificity of CS-1 and CS-2 (Fig. 1d).

CTSL cleavage sites are highly conserved among SARS-CoV-2 variants

Since the two CTSL cleavage sites in the SARS-CoV-2 S protein have not been reported previously, we investigated their sequence conservation among different variants. To date, many SARS-CoV-2 variants have been found in different countries (Fig. 1f). These variants exhibit potentially enhanced transmission, pathogenicity, immune escape, or a combination of all three and are dominant circulating variants in most countries. Many neutralizing antibody-targeted sites located in the receptor binding domain (RBD)

of the S protein are mutated in these variants. However, the CTSL cleavage sites are highly resistant to viral mutational escape (Fig. 1f), suggesting that these CTSL cleavage sites may be essential for the SARS-CoV-2 life cycle.

It has been reported that SARS-CoV-2 may originate from bat coronaviruses ¹⁶, such as RaTG13; therefore, we performed sequence alignment of residues around these sites in coronaviruses found in bats and pangolins. As shown in Supplementary Fig. 2a, CS-2 is highly conserved among most SARS-CoV- and SARS-CoV-2-related CoVs, while CS-1 is conserved only among SARS-CoV-2-related CoVs. This result suggests that the two CTSL cleavage sites may have different roles during evolution in various animal reservoirs. CS-2 seems to be the most important for the CoV life cycle, while CS-1 may have gradually evolved from SARS-CoV-1 to SARS-CoV-2 to improve viral infection. Moreover, we compared the conservation of the two CTSL cleavage sites among seven known human CoVs and found that these CTSL cleavage sites exist only in SARS-CoV-1 and SARS-CoV-2 (Supplementary Fig. 2b). This finding was consistent with previous studies showing that CTSL is involved in SARS-CoV-1 infection ¹⁷ and suggested that CTSL plays a unique role during the process of both SARS-CoV-1 and SARS-CoV-2 infection.

CTSL cleavage activates SARS-CoV-2 S with increased dynamics

To investigate how CTSL cleavage affects the structure and conformation of SARS-CoV-2 S, we performed cryo-EM studies on both the CTSL-treated and untreated S protein. We found three structural populations of S in the untreated control group (Supplementary Fig. 3) and six populations in the treated group (Supplementary Fig. 4). The resolutions of these structural populations ranged from 4.6 Å to 3.1 Å according to the gold standard Fourier shell correlation coefficient at 0.143 (Supplementary Figs. 3 and 4). The three structural populations in the untreated control group (Fig. 2a) included one inactive state (closed) and two active states (1R-weak and 1R-open), which have been discovered and reported before ^{12,18}. In the closed inactive state, the RBD is buried and not accessible for ACE2 binding, while in the activated states, one RBD becomes upright and exposed (1R-open) or becomes dynamic (1R-weak) and ready for ACE2 binding ¹⁹. In the CTSL-treated group (Fig. 2a), in addition to three populations (closed, 1R-weak and 1R-open) similar to those found in the control group (Supplementary Fig. 5a-c), we found three novel populations: the 1N-attached state, with presumable CTSL binding (Fig. 2b); the active state 2R-weak, with two RBDs missing (Supplementary Fig. 5d); and the ready-for-fusion state 1N-2R-weak, with two RBDs and one NTD missing (Fig. 2c-e). All six populations exhibit shapes generally similar to that of the prefusion S protein. We were aware of that the furin-treated S also kept similar shape with that of the prefusion S ¹⁸. Notably, in cryo-EM, the structure maps of the closed, 1N-attached, 1R-open and 1N-2R-weak states exhibited elongation effects resulting from the preferred orientations of raw particles (Supplementary Fig. 6); however, the functional domains of the S1 subunit, such as the NTD, CTD and RBD, could be clearly distinguished and assigned (Fig. 2 and Supplementary Fig. 6). In addition, we observed no significant differences in the density maps at the cleavage site regions between the untreated and treated groups (Supplementary Fig. 7), similar to the findings for the furin cleavage site, as previously reported ¹⁸.

Next, we sought to analyze the proportions of these populations before and after CTSL treatment. Without CTSL treatment, the proportion of the inactive-state S protein population was approximately 43% in our study, approximately the same as that in previous *in vivo* and *in vitro* results (Supplementary Table 1)^{14,20}. Notably, a previous study reported that in the presence of the ACE2 receptor, the proportion of the inactive state population significantly decreased to 11%¹⁹. In our study, the proportion of the inactive state population (closed) was also significantly reduced from 43–11% after cleavage by CTSL (Fig. 2a and f). In addition, with CTSL treatment, the proportions of the 1N-attached, 1R-open, 1R-weak, 2R-weak and 1N-2R-weak populations were 11%, 17%, 24%, 16% and 21%, respectively (Fig. 2a). The proportion of the 1R-weak population remained the same as that in the untreated group, while the proportion of the 1R-open population was reduced from 34–17% (Fig. 2a). In addition, by categorizing the 1R-open, 1R-weak, 2R-weak and 1N-2R-weak populations as the active state populations, the overall percentage of activation was 78% in the CTSL-treated group but 58% in the untreated group (Fig. 2a and f).

The unique population of 1N-attached emerged in the 3D classification step using the closed state as a reference and exhibited a shape obviously different from a regular threefold symmetry (Supplementary Fig. 4). The extra density surrounds CS-1 (R246-A262 loop region) in one NTD, which can be suitably fitted with the crystal structure of CTSL (PDB entry 3OF8) (Fig. 2b and Supplementary Fig. 5f). Thus, this enzyme-substrate complex population could represent an intermediate state during CTSL cleavage and be ready for activation.

Regarding the novel population 1N-2R-weak, with many domains that exhibited significant changes and offsets compared with the structure in the untreated group (Fig. 2c-e), we postulated that this state is not only an activated state but also a new conformation with the S2 subunit exposed and ready for fusion with the host cell membrane (Fig. 2f). In the structure of this population, a major part of the RBD and the top half of the NTD in one protomer of S (protomer 1) is disordered, while half of the RBD in another protomer (protomer 2) is disordered (Fig. 2c and d). In addition to domain disordering, we also observed significant domain movement in this population compared to the inactive state population (Fig. 2e and Supplementary Fig. 5e). In protomer 1, the remaining part of the RBD moved 8 Å away, while the CTD and the remaining parts of the NTD and CTD moved 7 Å and 15 Å away, respectively. In protomer 2, the remaining part of the RBD moved 14 Å away from its location in the closed configuration. In the third protomer (protomer 3), the intact RBD moved 5 Å away from its original position in the inactive state. These observations imply that this novel population has greatly decreased stability compared with that of the S protein in its prefusion state and is thus ready for fusion with the host membrane (Fig. 2f).

Overall, these results suggested that CTSL cleavage induces significantly increased dynamics of the RBDs and NTDs with an increased proportion of activated S, which is accessible for ACE2 binding and then becomes ready for membrane fusion, resulting in improved potential for infection (Supplementary Movie 1).

CTSL cleavage sites are essential for SARS-CoV-2 infection and cell fusion

To verify the observations from our structural studies, we investigated the functions of CTSL in the viral infection process. We generated several mutants of PsV according to the two cleavage sites, including a CS-1 mutant (CS-1M), CS-2 mutant (CS-2M), and combination mutant (CS-1M + 2M) (Fig. 3a). As the proper control, we generated three PsV mutants proven to escape furin cleavage (Supplementary Fig. 8a): furin mutant (FM)-delta (deletion of residues 682–685), FM-ARAA (₆₈₂ARAA₆₈₅) and FM-GSAS (₆₈₂GSAS₆₈₅). Then, we compared the infectivity of the wild-type (WT) and mutant PsVs in four human and monkey cell lines, which we have previously validated to have the highest susceptibility to SARS-CoV-2 infection ^{3,21}. Mutations in the furin cleavage site did not reduce the infectivity of either the PsV or live SARS-CoV-2 ^{22,23}, suggesting that furin cleavage may not be essential for viral infection. However, once the CTSL cleavage sites were mutated to glycine, these PsVs lost most of their infection ability in all four cell lines. When both CTSL cleavage sites were mutated, the viral infectivity was reduced 100 ~ 700-fold compared with that of WT PsV (Fig. 3b), showing that the two novel CTSL cleavage sites CS-1 and CS-2 are essential for SARS-CoV-2 infection. Moreover, CS-2 seems to play a more important role than CS-1 in SARS-CoV-2 entry. Similar results were observed with different PsV concentrations (Supplementary Fig. 8b).

The fusogenic activity of SARS-CoV-2 S is a characteristic of SARS-CoV-2 infection, and the presence of syncytia in COVID-19 patient lung tissue indicates that this activity plays a role in the pathological process ²⁴. Syncytium formation was quantified with an ESR1-ERE transactivation system (Fig. 3c-e and Supplementary Fig. 9) and observed by bright field microscopy (Fig. 3f). Compared with the control group, the CS-1M group exhibited significantly reduced syncytium formation (Fig. 3e and f), syncytium formation was inhibited even more severely in the CS-2M group, and almost no syncytia were observed in the CS-1M + 2M group (Fig. 3e and f). More importantly, the inhibitory effects of these mutations were not rescued by the addition of CTSL into the system (Fig. 3e and f), confirming that the promotive effect of CTSL on syncytium formation, as previously shown ⁹, was due to S protein cleavage at CS-1 and CS-2. Taken together, our findings proved that CTSL is essential for SARS-CoV-2 infection by cleaving the viral S protein at CS-1 and CS-2.

CTSL cleavage promotes SARS-CoV-2 infection independent of furin

As the CS-2 region is very close to the furin cleavage site, we next compared the function of these two proteases using different FM PsVs. PsV infection assays were performed in the human Huh7 cell line ^{3,21,25} under *CTSL* overexpression or knockdown conditions. Overexpression of the *CTSL* gene markedly increased but knockdown of the *CTSL* gene significantly reduced the infection efficiency of SARS-CoV-2 PsV (Fig. 3g), similar to our previous finding ⁹. This result indicated that CTSL plays a crucial role in the SARS-CoV-2 infection process. Furthermore, overexpression or knockdown of the *CTSL* gene also significantly affected the infection levels of the three FM PsVs in a dose-dependent manner (Fig. 3g), similar to its effects on WT PsV. This result suggested that CTSL-enhanced SARS-CoV-2 viral entry is independent of furin cleavage.

The furin cleavage site has been reported to be required for SARS-CoV-2 S-driven cell-cell fusion¹⁰. Here, we used the same cell fusion system (Fig. 3d) to evaluate the effect of CTSL on FM S protein-mediated syncytium formation. FM-ARAA S protein induced a low level of syncytium formation similar to that in the scramble group (Fig. 3h and 3i). However, after CTSL was added to the mutant S group, many syncytia began to form, and this rescue effect exhibited dose dependency with the concentration of CTSL (Fig. 3h and 3i). This result suggested that CTSL cleavage is another important factor that induces SARS-CoV-2 S-driven cell-cell fusion.

CTSL inhibitors prevent SARS-CoV-2 infection

Since CTSL plays a key role in mediating SARS-CoV-2 infection, we selected six compounds (K777, cathepsin inhibitor 1, E64d, Z-FY-CHO, MDL-28170 and oxocarbazate) that can inhibit CTSL activity by over 90% at a concentration of 50 mM and evaluated their effects on preventing SARS-CoV-2 infection (Supplementary Fig. 10). Z-FY-CHO and oxocarbazate are selective CTSL inhibitors, while the other 4 compounds are pan-cathepsin inhibitors.

First, in *in vitro* cellular experiments using live SARS-CoV-2, all six compounds efficiently blocked SARS-CoV-2 infection at nontoxic doses, with half-maximal effective concentration (EC₅₀) values ranging from 35 nM to 10.9 mM (Fig. 4a). All compounds decreased the viral RNA load in Vero E6 cells by ~ 10³- to 10⁴-fold at the maximal nontoxic dose. This finding further confirmed the critical role of CTSL in SARS-CoV-2 infection.

Then, we further examined the antiviral effects of the six compounds against the recently emerging variants of SARS-CoV-2 using the PsV system (Fig. 4b). As we expected, all six compounds also efficiently inhibited infection with all mutant PsVs. We found that E64d showed a higher inhibitory efficiency against P.1, while MDL-28170 and oxocarbazate were more effective against B.1.429. Thus, these data proved that CTSL inhibitors are resistant to mutational escape of SARS-CoV-2, including B.1.351, against which the effects of most current vaccines are reduced^{7,8}.

Finally, we investigated the *in vivo* antiviral effects of CTSL inhibitors in human *ACE2* (*hACE2*) transgenic mice, which are susceptible to SARS-CoV-2 infection²⁶. We selected two compounds, E64d and Z-FY-CHO, which have been validated to be safe for *in vivo* studies^{9,27}. E64d and Z-FY-CHO were administered in prophylactic and therapeutic regimens, and mice were treated at appropriate doses according to previous studies^{9,27}. Each mouse was infected with 10⁶ PFU SARS-CoV-2 at 0 days post infection (dpi) by intranasal instillation. Tissue samples were collected at 4 dpi (Fig. 5a), when the viral load peaked, and exhibited obvious histopathological changes. Both compounds significantly decreased the number of viral RNA copies in lung tissues by ~ 1 to 4 log₁₀, with undetectable (below the lower limit of detection, LOD) viral loads in lung tissues from 2 mice treated with Z-FY-CHO and 4 mice treated with E64d (Fig. 5b). Both drugs also significantly decreased the viral load in the nasal turbinate (Fig. 5b). Histological analysis of the lungs showed that vehicle-treated mice exhibited moderate pathological changes, as evidenced by the large areas of alveolar septal thickening, inflammatory cell infiltration and bronchiolar epithelial cell degeneration. In contrast, the lungs of E64d- and Z-FY-CHO-treated mice

exhibited improved morphology and less infiltration (Fig. 5c and 5d). Taken together, these results indicated that CTSL inhibitors efficiently block SARS-CoV-2 infection both *in vitro* and *in vivo* and are resistant to viral mutational escape.

Discussion

In this study, we found that CTSL cleaves SARS-CoV-2 S at two novel sites in two loops in the S1 subunit, named CS-1 and CS-2. CS-1 is located in the NTD, while CS-2 is located in the RBD and is near the S1/S2 site. Both sites are highly conserved among all SARS-CoV-2 variants, including recently emerging variants. Our cryo-EM structural studies suggested that CTSL cleavage can greatly enhance S protein dynamics, with an increased proportion of activated state populations as well as novel activation conformations. Our viral infection assay using the PsV system proved that CTSL cleavage, especially cleavage at CS-2, is an essential step in SARS-CoV-2 infection.

Combining these results with those of our structural analysis, we proposed a possible model for the CTSL-mediated SARS-CoV-2 infection process (Fig. 6 and Supplementary Movie 1). When viral S encounters active CTSL, CTSL can bind to CS-1 and CS-2 in the loops on the S protein to form the S-CTSL complex. CTSL cleavage at CS-1 greatly induces the 2R-weak conformation and increases the dynamics of the RBD and NTD to increase the ACE2 binding activity of S. Then, CTSL cleaves S at CS-2, which induces the 1N-2R-weak conformation, and the S1 and S2 subunits are then ready to separate to expose S2 for membrane fusion and complete viral infection. We proved that SARS-CoV-2 utilizes human CTSL to infect host cells independent of furin. In addition, many TMPRSS2-deficient cells, including Vero cells, have been found to be highly susceptible to SARS-CoV-2 infection, suggesting that TMPRSS2 is not indispensable for infection²⁸. Therefore, it seems that instead of furin and TMPRSS2 inhibitors, CTSL-specific inhibitors are a better choice to prevent SARS-CoV-2 infection. Indeed, CTSL inhibitors reduced infection with live SARS-CoV-2 by $\sim 10^3$ - to 10^4 -fold in Vero E6 cells and nearly completely prevented infection with SARS-CoV-2 variant PsVs. Moreover, CTSL inhibitors effectively prevented SARS-CoV-2 infection in *hACE2* transgenic mice.

We identified the precise CTSL cleavage loops in the S protein. This finding provides the structural basis for the development of new drugs and vaccines. In addition to CTSL-specific inhibitors, drugs that can specifically protect the cleavage loops from CTSL-mediated activation or competitively inhibit the binding of CTSL with the cleavage loops are new strategies for COVID-19 treatment. Moreover, vaccines targeting CS-1 and CS-2 to counteract viral mutations are another approach.

In conclusion, the high mutation rate of SARS-CoV-2 often leads to viral escape from neutralizing antibodies or vaccines. An understanding of the precise mechanisms underlying SARS-CoV-2 infection is urgently needed to develop specific antiviral drugs or vaccines with increased resistance to viral mutations. Our study shows that the new CTSL cleavage sites in the SARS-CoV-2 S protein are crucial new targets for the development of drugs and vaccines to treat infections with SARS-CoV-2 variants.

Methods

Protein expression and purification

The S protein was purified as previous reported ¹⁴ by Novoprotein company. Briefly, the prefusion S ectodomain, the mammalian codon-optimized gene coding SARS-CoV-2 (Wuhan-Hu-1 strain, GenBank ID: MN908947.3) S glycoprotein ectodomain (residues M1–Q1208) with proline substitutions at K986 and V987, a “GSAS” substitution at the furin cleavage site (R682 to R685), a C-terminal thrombin tag, a T4 fibrin trimerization motif and a 6xHisTag was synthesized and cloned. To express 4-Gly-mutant S protein, residues W258, T259, V635 and Y636 were substituted into glycine. These expression vectors were used to transiently transfect to Expi293 cells and purified by using cOmpleteHis-Tag Purification Resin to capture the target protein. The protein was then subjected to additional purification by size-exclusion chromatography using Superdex 200 10/300 GL column (GE Healthcare) in PBS.

CTSL treatment

Recombinant CTSL (Novoprotein) was used to cleave SARS-CoV-2 S protein. Purified SARS-CoV-2 S protein was incubated with CTSL (2–8 mg/mL as indicated and 8 mg/ml for cryo-EM sample preparation) in the presence of 100 mM NaAC, 1 mM EDTA and 1 mM DTT, pH 5.5 for 24 h at 25°C. CTSL was preactivated in 30°C for 1 min before use. The resulting protein was then subjected to SDS-PAGE and stained by Coomassie blue or silver staining methods as indicated. For cryo-EM sample preparation, the treated S protein was then subjected to additional purification by size-exclusion chromatography using Superdex 200 10/300 GL column (GE Healthcare) in PBS.

Sequence analysis

Sequence alignments were performed using the CLUSTALW online tool (<https://www.genome.jp/tools-bin/clustalw>). The representative variant sequence of each SARS-CoV-2 PANGO lineage was obtained from Global Initiative on Sharing All Influenza Data (GISAID) database: B.1.617 (GISAID: EPI_ISL_1544002), B.1.1.7 (GISAID: EPI_ISL_1257795), B.1.351 (GISAID: EPI_ISL_935042), P.1 (GISAID: EPI_ISL_1164984), B.1.427 (GISAID: EPI_ISL_1221570), B.1.429 (GISAID: EPI_ISL_1160035), B.1.526 (GISAID: EPI_ISL_1095660), A.23.1 (GISAID: EPI_ISL_954226), A.2 (GISAID: EPI_ISL_462477), A.2.2 (GISAID: EPI_ISL_849668), B.1.258.14 (GISAID: EPI_ISL_1255050), B.1.36 (GISAID: EPI_ISL_751233), B.1.1.214 (GISAID: EPI_ISL_1257799), B.1.1.63 (GISAID: EPI_ISL_1055453), B.1.1.306 (GISAID: EPI_ISL_1201070), B.1.1.157 (GISAID: EPI_ISL_1183464), B.1.1.39 (GISAID: EPI_ISL_1195031). In addition, the following sequences information were obtained from the National Center for Biotechnology Information (NCBI) database: SARS-CoV BJ01 (GenBank: AAP30030.1), MERS-CoV (GenBank: QBM11748.1), HCoV NL63 (GenBank: APF29063.1), HCoV 229E (GenBank: AWH62679.1), HCoV OC43 (GenBank: AIX10763.1), HCoV HKU1 (GenBank: AMN88694.1).

N-terminal Edman sequencing and LC-MS/MS analysis

CTSL treated SARS-CoV-2 S protein was resolved on an SDS-PAGE gel. The protein was transferred to PVDF membrane before Edman sequencing. Each protein band was excised individually. The sequence

of the first 10 amino acids for each band was determined by Edman sequencing by using ABI Procise-cLC machine. LC-MS/MS analysis of excised bands was performed at Laboratory of Proteomics, Institute of Biophysics, Chinese Academy of Sciences.

Cryo-EM sample preparation and data collection

For each grid, 3 μ l of purified protein solution of CTSL treated or untreated SARS-CoV-2 S proteins were applied to newly glow-discharged holey carbon film grids (R1.2/1.3, 200 meshes, Au, Quantifoil, Germany) or holey Ni-Ti film grids²⁹ (R1.2/1.3, 300 meshes, Au, Zhenjiang Lehua Electronic Technology Co., LTD, China). Then the grid was blotted and vitrified by plunge freezing into liquid ethane using Vitrobot Mark IV (Thermo Fisher Scientific, USA) at 4°C and 100% humidity. All movies were collected on a Titan Krios G2 TEM (Thermo Fisher Scientific, USA) operated at 300 KV on EF-TEM mode with nominal magnification set to be 105000x, resulted in a calibrated physical pixel size of 1.36 Å on specimen level. For CTSL treated sample, the total dose was set to be 80 e⁻/Å², with an exposure time of 16 seconds, 0.4 second per frame. For untreated sample, the dose was set to be 60 e⁻/Å², with an exposure time of 15 seconds, 0.3 second per frame. For both datasets, the movies were acquired by Gatan K2 Summit DDD detector (Gatan Company, USA) equipped with a GIF Quantum energy filter, with a slit width of 20 eV, operated in super resolution mode, resulted in a pixel size of 0.68 Å for output super resolution movies. SerialEM³⁰ with in-house scripts were used for data collection³¹, defocus value for either samples were set between - 1.0 and - 2.0 μ m.

Image processing

Image processing steps were performed using RELION^{32,33} and cryoSPARC³⁴. The output super resolution movies were first subjected to motion correction using MotionCor2³⁵, with binning level of 2 in Fourier space, and dose weighting was also performed during the process. Contrast transfer function (CTF) parameters estimation were performed using Gctf³⁶. Gautomatch (<https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/>) was used to pick 364k particles from 1249 micrographs in untreated S dataset, and 1500k particles from 5025 micrographs in CTSL treated S dataset. Then good particles were extracted and sorted by two rounds of 2D classification. After 2D classification, one round of 3D classification was performed to investigate different populations of Ss. Three different populations were found for the untreated dataset and six for the CTSL treated dataset. For the S-1N-attach population, another round of 3D classification was performed to further sort particles. To prevent any model bias, previously reported cryo-EM structure of SARS-CoV-2 S (PDB entry 6VXX) lowpass filtered to 60 Å was used as initial reference and no additional symmetry was imposed throughout the whole image processing pipeline. Then the well-aligned classes with clear secondary structure features were selected separately for subsequent image processing, including auto-refinement, Bayesian polishing, CTF refinement and no-alignment 3D classification. All these steps were performed in RELION. For the final round of refinement, the non-uniform (NU) refinement in cryoSPARC was performed to refine the final maps to a higher resolution (see also **Figures S3 and S4**).

Model building and refinement

To build the atomic model of various CTSL treated or untreated SARS-CoV-2 S structures, the reported cryo-EM structure of SARS-CoV-2 S (PDB entry 6VXX & 6VYB)¹² was used as an initial model. We were able to trace most regions with side chains using COOT³⁷. The built model was further refined in real space using Phenix³⁸. All figures were created by Pymol³⁹, UCSF Chimera⁴⁰ and UCSF ChimeraX⁴¹. The parameters for data collection and structure determination are summarized in Supplementary Table 2–4.

Cell culture

The Huh7 (*Homo sapiens*, liver), 293T-hACE2 (293T cells stably expressed, *Homo sapiens*, embryonic kidney), Vero (*Cercopithecus aethiops*, kidney), VeroE6 (*Cercopithecus aethiops*, kidney) and LLC-MK2 (*Macaca mulatta*, kidney) cells were maintained in high glucose Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% fetal bovine serum (FBS, Gibco, Carlsbad, CA), 100 units/ml Penicillin-Streptomycin (Gibco). All the cells were maintained at 37°C in a humidified atmosphere containing 95% air and 5% CO₂.

Experimental mice

The study used human *ACE2* transgenic mice, a mouse model expressing human *ACE2* (*hACE2*) generated by using CRISPR/Cas9 knock-in technology as previously reported²⁶. The *hACE2* mice used in this manuscript were 17-weeks-old female C57BL/6 mice, with the body weight between 22 g. All animal protocols were approved by the Ethical Review Committee at the Institute of Zoology, Capital Medical University, China.

SARS-2-S plasmids and site-directed mutagenesis

SARS-2-S plasmid was constructed by inserted the codon-optimized S gene of SARS-CoV-2 (GenBank: MN_908947) into pcDNA3.1 as previously described³. For the site-directed mutagenesis, 15 to 20 nucleotides before and after the target mutation site were selected as forward primers, while the reverse complementary sequences were selected as reverse primers. Following site-directed mutagenesis PCR, the template chain was digested using DpnI restriction endonuclease (NEB, USA). The PCR product was transformed into *E. coli* DH5a competent cells; single clones were selected and then sequenced.

Production and quantification of PsVs

PsVs incorporated with S protein from either SARS-CoV-1, SARS-CoV-2, or mutants were constructed using a procedure described by us previously. For this VSV-based PsV system, the backbone was provided by VSV-G pseudotyped virus (G*ΔG-VSV) that packages expression cassettes for firefly luciferase instead of VSV-G in the VSV genome. For quantification of PsV, viral RNA was extracted by using the QIAamp Viral RNA Mini Kit (Cat. No. 52906, QIAGEN), and the reverse transcription was performed with RevertAid™ First Strand cDNA Synthesis Kit (Fermentas K1622) according to the manufacturer. The real-time qPCR was then performed on the LightCycler® 96 Real-Time PCR System (Roche) using SYBR Green I Master Mix reagent (Roche). The P protein gene of VSV virus was quantified and the viral copy number calculated accordingly. The primers were: forward-TCTCGTCTGGATCAGGCGG; reverse-TGCTCTTCCACTCCATCCTCTTGG. All PsVs were normalized to the same amount as previously described³.

Western blot analysis

Seven millilitres of SARS-CoV-2 PsVs with a titre of 1.86×10^5 TCID₅₀/ml were pelleted through a 25% sucrose cushion by ultra-centrifugation at 100,000× g for 3 h. The layer of supernatant was discarded. The pellet was collected and subjected to western-blot analysis as previously described, and detected by anti-S1 mice serum and VSV-M antibody (KeraFast, Cat. No. EB0011).

PsV infection assay

Before infection, the 96-well plates were seeded with cells adjusted to 2×10^5 cells/ml. Then, 100 µL of the normalized PsV with indicated dilution fold was added to wells in 96-well cell culture plate. After 24 h incubation at 37 °C, the activities of firefly luciferase were measured on cell lysates using luciferase substrate (Perkinelmer, BRITELITE PLUS 100 µL KIT, Cat. No. 6066761) following the manufacturer's instructions. Luciferase activity was measured using a luminometer (Promega).

Overexpression and knockdown of CTSL

Overexpression and knockdown of *CTSL* gene in huh7 cells were described by us previously⁹. Briefly, for *CTSL* knockdown, Huh7 cells were plated in 48-well plates, and transfected with indicated concentrations of siRNAs against homo *CTSL* mRNA (si-CTSL) or 50 nM negative control siRNA (scramble) using Lipofectamine 3000 reagent (Thermo Fisher Scientific). For *CTSL* overexpression, Huh7 cells were plated in 48-well plates and transfected with indicated concentrations of human CTSL expression plasmid (GenBank: NM_001912, Vigenebio) or 0.2 mg control plasmid. The overexpression and knockdown efficiencies were validated in both mRNA and protein levels by us previously⁹. Twenty-four hours post transfection, the medium was replaced with fresh medium. Then the cells were infected with SARS-CoV-2 or mutant PsV (1.3×10^4 TCID₅₀/ml) and cultured for another 24 h before firefly luciferase activity analysis. siRNA sequences were provided in Supplementary Table 5.

Syncytium-formation assay

Syncytium-formation assay was performed as previously described by us⁹. Briefly, Huh7 cells were seeded in 24-well plates and transfected with SARS-2-S or mutant protein expression plasmids using Lipofectamine 3000 reagent (Thermo Fisher Scientific). The transfection solutions were changed to standard culture medium 4–6 h post transfection and cells incubated for an additional 6–8 h. Next, cells were treated in the absence (PBS, pH = 5.8) or presence of 2 or 4 mg/ml CTSL (Novoprotein) (in PBS, pH = 5.8) for 20 min at 37°C. Then, the solutions were changed to a standard culture medium, and the cells further incubated for 10–16 h. The pictures were captured under bright-field microscopy (Olympus).

Quantification of syncytium-formation

Huh7 target-cells were seeded in 48-well plate at 10–20 % confluency and transfected with ESR1 expression plasmid. Huh7 effector-cells were seeded in a 6-well dish at 50–70 % confluency and transfected with ERE-Luciferase reporter plasmid as well as expression plasmid for WT or mutant SARS-2-S or empty plasmid (1:2 ratio). At 24 h post-transfection, effector-cells were detached by resuspending

and added into the target-cells at 1:1 ratio. Where indicated, the medium was changed by CTSL, trypsin or PBS 2 h after mixture and reaction was stopped after 20 min by adding 500 μ L culture medium. Luciferase activity was analyzed after 20–30 h.

Analysis of CTSL activity

The CTSL inhibition by drugs was measured using the CTSL substrate Ac-FR-AFC (Abcam, Cat. No. ab157769). CTSL will cleave the synthetic substrate to release free AFC (amino-4-trifluoromethyl coumarin), which can be measured in a spectrophotometer at $E_m = 505$ nm. Purified CTSL (Novoprotein) was incubated with CTSL substrate in the absence (vehicle, control) or presence of 10 mM compounds as indicated in 37 °C for 1–2 h, the CTSL activity was measured by a spectrophotometer (Promega) at $E_m = 505$ nm.

SARS-CoV-2 virus

All experiments with infectious SARS-CoV-2 were performed in the biosafety level 4 and animal biosafety level 4 facilities in the Harbin Veterinary Research Institute (HVRI) of the Chinese Academy of Agricultural Sciences (CAAS), which is approved for such use by the Ministry of Agriculture and Rural Affairs of China.

SARS-CoV-2 strain SARS-CoV-2/HRB25/human/2020/CHN (HRB25, GISAID: EPI_ISL_467430)⁴² was isolated from a patient in Vero E6 cells. Viral stocks were prepared in VeroE6 cells with DMEM containing 5% FBS. Viruses were harvested and the titers were determined by means of plaque assay in Vero E6 cells⁴³.

SARS-CoV-2 infection assay in vitro

Vero E6 cells were seeded in a 96-well plate and grown overnight. Then the cells were pretreated with different concentrations of each compound for 16 h and then infected with SARS-CoV-2 at a MOI of 0.01. After 24 h, the cell supernatants were collected to extract viral RNA and subjected to RT-qPCR analysis. The N gene-specific primers (forward, 5'-GGGGAAGTTCTCCTGCTAGAAT-3'; reverse, 5'-CAGACATTTTGCTCTCAAGCTG-3') and probe (5'-FAM-TTGCTGCTGCTTGACAGATT-TAMRA-3') were utilized according to the information provided by the National Institute for Viral Disease Control and Prevention, China (<http://nmcdc.cn/nCoV>). The EC_{50} values were calculated by using a dose-response model in GraphPad Prism 8.0 software. The cell viability was determined by using the CCK kit (TransGen Biotech) according to the manufacture.

SARS-CoV-2 infection assay in vivo

Prophylactic treatment was given on -2, -1, 0, 1, 2, and 3 dpi, while therapeutic regimen was given on 0, 1, 2, and 3 dpi. The mice were infected at 0 dpi with 10^6 PFU of SARS-CoV-2 by intranasal instillation. Lung and nasal turbinate tissues were collected at 4 dpi. The viral RNAs copies from tissues were measured by RT-qPCR. Tissue pathology of infected mice was examined by H&E staining and immunohistochemical assays as described previously^{43,44}.

Statistical analysis

All statistical analyses were conducted with the software GraphPad Prism version 7.0. Data are presented as means $\pm$ s.e.m. Statistical significance was determined using unpaired Student's *t*-test or the Mann-Whitney *U*-test if $n < 9$, or one-way ANOVA as appropriate.

Data availability

The Electron Microscopy Database (EMD) accession codes of the S-1R-weak, S-1R-open and S-closed states in untreated group are EMD-31506, EMD-31512 and EMD-31511, respectively; and that of the S-1N2R-weak, S-1R-open, S-1R-weak, S-2R-weak, S-closed, S-1N-attach states in CTSL treated group are EMD-31508, EMD-31509, EMD-31513, EMD-31514, EMD-31510 and EMD-31507, respectively. For those maps without preferred orientations, the structural models were deposited to the Protein Data Bank (PDB) with accession code of 7FB0, 7FB1, 7FB3 and 7FB4 for S-closed (UT), S-1R-open (UT), S-1R-weak (CTSL) and S-2R-weak (CTSL), respectively. The other data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Acknowledgment

This work was supported by grants from National Key R&D Program of China (2017YFC0909600) and National Natural Science Foundation of China (81930019), Scientific Project of Beijing Municipal Science & Technology Commission (D171100002817005), Beijing Municipal Administration of Hospitals Clinical Medicine Development of Special Funding Support (ZYLX201823) to J.K.Y. This work was equally supported by grants from Strategic Priority Research Program of Chinese Academy of Sciences (XDB37040102), the National Key Research and Development Program of China (2017YFA0504700) and National Natural Science Foundation of China (31925026). This work was also supported by grants from the Ministry of Science and Technology of China (2018YFA0901102 and 2019YFA0904101).

We thank Jianping Feng (J. K.Y. lab), Ping Shan and Ruigang Su (F.S. lab) for their assistance in lab management. We thank Jifeng Wang and Lili Niu (Laboratory of Proteomics, Institute of Biophysics, Chinese Academy of Sciences) for helping with LC-MS/MS and data analysis. All the cryo-EM work were performed at Center for Biological Imaging (CBI, <http://cbi.ibp.ac.cn>), Core Facilities for Protein Science at Institute of Biophysics (IBP), Chinese Academy of Sciences (CAS). We would also like to thank B. Zhu, X. Huang, X. Li, T. Niu, G. Ji, D. Fan and other staff members from CBI for their help in data collection.

Author contributions

J.-K.Y. conceived the idea for the study, supported the study, designed the experiments, and wrote the manuscript. M.-M.Z., Y.Z. designed, performed the experiments, and wrote the first version of manuscript. L.Z., G.Z. designed and performed the experiments. S.L., L.T., G. Y., W.H., C.F., L.S., Z.W., C.W., X.H. partially

performed the experiments. Z.B., Y.W., F.S. helped with the interpretation of the results, design of experiments, partially supported the study, and proof of the manuscript.

Competing interests

The authors have no competing interests to declare.

Supplementary information

Supplementary data includes 11 figures 5 tables and 1 movie file.

Data availability

The Electron Microscopy Database (EMD) accession codes of the S-1R-weak, S-1R-open and S-closed states in untreated group are EMD-31506, EMD-31512 and EMD-31511, respectively; and that of the S-1N2R-weak, S-1R-open, S-1R-weak, S-2R-weak, S-closed, S-1N-attach states in CTSL treated group are EMD-31508, EMD-31509, EMD-31513, EMD-31514, EMD-31510 and EMD-31507, respectively. For those maps without preferred orientations, the structural models were deposited to the Protein Data Bank (PDB) with accession code of 7FB0, 7FB1, 7FB3 and 7FB4 for S-closed (UT), S-1R-open (UT), S-1R-weak (CTSL) and S-2R-weak (CTSL), respectively. The other data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- 1 Sewell, H. F., Agius, R. M., Kendrick, D. & Stewart, M. Vaccines, convalescent plasma, and monoclonal antibodies for covid-19. *BMJ* **370**, m2722, doi:10.1136/bmj.m2722 (2020).
- 2 Lythgoe, M. P. & Middleton, P. Ongoing Clinical Trials for the Management of the COVID-19 Pandemic. *Trends in pharmacological sciences* **41**, 363-382, doi:10.1016/j.tips.2020.03.006 (2020).
- 3 Li, Q. *et al.* The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **182**, 1284-1294 e1289, doi:10.1016/j.cell.2020.07.012 (2020).
- 4 Dejnirattisai, W. *et al.* Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell*, doi:10.1016/j.cell.2021.03.055 (2021).
- 5 Tegally, H. *et al.* Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. *Nature*, doi:10.1038/s41586-021-03402-9 (2021).
- 6 Ferrante, L. *et al.* Brazil's policies condemn Amazonia to a second wave of COVID-19. *Nature medicine* **26**, 1315, doi:10.1038/s41591-020-1026-x (2020).
- 7 Murray, C. J. L. & Piot, P. The Potential Future of the COVID-19 Pandemic: Will SARS-CoV-2 Become a Recurrent Seasonal Infection? *Jama* **325**, 1249-1250, doi:10.1001/jama.2021.2828 (2021).

- 8 Abdool Karim, S. S. & de Oliveira, T. New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications. *The New England journal of medicine* **384**, 1866-1868, doi:10.1056/NEJMc2100362 (2021).
- 9 Zhao, M. M. *et al.* Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal transduction and targeted therapy* **6**, 134, doi:10.1038/s41392-021-00558-8 (2021).
- 10 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, doi:10.1016/j.cell.2020.02.052 (2020).
- 11 Zang, R. *et al.* TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Science immunology* **5**, doi:10.1126/sciimmunol.abc3582 (2020).
- 12 Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292 e286, doi:10.1016/j.cell.2020.02.058 (2020).
- 13 Shang, J. *et al.* Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* **117**, 11727-11734, doi:10.1073/pnas.2003138117 (2020).
- 14 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263, doi:10.1126/science.abb2507 (2020).
- 15 Biniossek, M. L., Nagler, D. K., Becker-Pauly, C. & Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *Journal of proteome research* **10**, 5363-5373, doi:10.1021/pr200621z (2011).
- 16 Hu, B., Guo, H., Zhou, P. & Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nature reviews. Microbiology* **19**, 141-154, doi:10.1038/s41579-020-00459-7 (2021).
- 17 Simmons, G. *et al.* Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci U S A* **102**, 11876-11881, doi:10.1073/pnas.0505577102 (2005).
- 18 Wrobel, A. G. *et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol* **27**, 763-767, doi:10.1038/s41594-020-0468-7 (2020).
- 19 Benton, D. J. *et al.* Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, doi:10.1038/s41586-020-2772-0 (2020).
- 20 Ke, Z. L. *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498+, doi:10.1038/s41586-020-2665-2 (2020).
- 21 Li, Q. *et al.* SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape. *Cell*, doi:10.1016/j.cell.2021.02.042 (2021).

- 22 Cantuti-Castelvetri, L. *et al.* Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856-860, doi:10.1126/science.abd2985 (2020).
- 23 Papa, G. *et al.* Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. *PLoS Pathog* **17**, e1009246, doi:10.1371/journal.ppat.1009246 (2021).
- 24 Bussani, R. *et al.* Persistence of viral RNA, pneumocyte syncytia and thrombosis are hallmarks of advanced COVID-19 pathology. *EBioMedicine* **61**, 103104, doi:10.1016/j.ebiom.2020.103104 (2020).
- 25 Ou, X. *et al.* Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* **11**, 1620, doi:10.1038/s41467-020-15562-9 (2020).
- 26 Sun, S. H. *et al.* A Mouse Model of SARS-CoV-2 Infection and Pathogenesis. *Cell Host Microbe* **28**, 124-133 e124, doi:10.1016/j.chom.2020.05.020 (2020).
- 27 Woo, J. T. *et al.* Suppressive effect of N-(benzyloxycarbonyl)-L-phenylalanyl-L-tyrosinal on bone resorption in vitro and in vivo. *European journal of pharmacology* **300**, 131-135, doi:10.1016/0014-2999(95)00858-6 (1996).
- 28 Sasaki, M. *et al.* SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells. *PLoS pathogens* **17**, e1009233, doi:10.1371/journal.ppat.1009233 (2021).
- 29 Huang, X. *et al.* Amorphous nickel titanium alloy film: A new choice for cryo electron microscopy sample preparation. *Prog Biophys Mol Biol* **156**, 3-13, doi:10.1016/j.pbiomolbio.2020.07.009 (2020).
- 30 Mastronarde, D. N. SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microscopy Microanalysis* **9**, 1182-1183 (2003).
- 31 Wu, C., Huang, X., Cheng, J., Zhu, D. & Zhang, X. High-quality, high-throughput cryo-electron microscopy data collection via beam tilt and astigmatism-free beam-image shift. *Journal of structural biology*, doi:10.1016/j.jsb.2019.09.013 (2019).
- 32 Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180**, 519-530, doi:10.1016/j.jsb.2012.09.006 (2012).
- 33 Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
- 34 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

- 35 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods* **14**, 331 (2017).
- 36 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).
- 37 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
- 38 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 39 Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* (2015).
- 40 Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
- 41 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*, doi:10.1002/pro.3943 (2020).
- 42 Wu, S. *et al.* A single dose of an adenovirus-vectored vaccine provides protection against SARS-CoV-2 challenge. *Nature communications* **11**, 4081, doi:10.1038/s41467-020-17972-1 (2020).
- 43 Shi, J. *et al.* Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* **368**, 1016-1020, doi:10.1126/science.abb7015 (2020).
- 44 Wang, J. *et al.* Mouse-adapted SARS-CoV-2 replicates efficiently in the upper and lower respiratory tract of BALB/c and C57BL/6J mice. *Protein & cell* **11**, 776-782, doi:10.1007/s13238-020-00767-x (2020).
- 45 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature Methods* **14**, 793+, doi:10.1038/Nmeth.4347 (2017).

Figures

Fig. 1

Figure 1

CTSL cleaves the SARS-CoV-2 S protein at two novel sites. a, Schematic illustration of the SARS-CoV-2 S glycoprotein in which the functional domains and cleavage sites are highlighted (NTD, N-terminal domain; RBD, receptor binding domain; CTD, C-terminal domain; FP, fusion peptide; HR1, heptad repeat 1; HR2, heptad repeat 2; TD, transmembrane domain). CTSL cleaves at CTSL cleavage site 1 (CS-1) and CTSL cleavage site 2 (CS-2). Furin cleaves at the S1/S2 site. b, Overall structure of the SARS-2 S ectodomain (PDB entry: 6VXX). The CTSL cleavage sites, CS-1 and CS-2, are colored red, while the furin cleavage site at S1/S2 is colored orange. The three protomers of the S trimer are colored pink, purple and green. c, SDS-PAGE analysis of CTSL-mediated cleavage of the S protein. The purified SARS-CoV-2 S protein ectodomain was incubated with different concentrations of CTSL (2-8 μg/mL). The N-terminal

sequencing results for band 1 (60 kDa) and band 2 (100 kDa) are also shown. d, 4-Gly-mutant SARS-CoV-2 S cannot be cleaved by CTSL into the 60 kDa (band 1) and 100 kDa (band 2) fragments. WT S protein (1 µg) and mutant S protein (500 ng) were incubated with CTSL (8 µg/mL). All samples were subjected to SDS-PAGE, and bands were detected by silver staining. e, The P1 and P2 residues in CS-1 and CS-2 were mutated to glycine, and the mutant SARS-2-S glycoprotein was named 4-Gly-mutant S. f, Amino acid sequence alignment of residues around CS-1 and CS-2 in SARS-CoV-2 variants. P1 and P2 residues that are the same as the WT residues are highlighted in red. The symbol * indicates amino acid residues that are conserved among all tested sequences.

Fig. 2

Figure 2

Structures of CTSL-untreated and treated SARS-CoV-2 S proteins. a, Side and top views of untreated (untreated) and CTSL-treated S proteins. In the untreated group, the closed form was defined as the inactive state, while the 1R-weak and 1R-open forms were defined as the active state. In the CTSL-treated group, the closed form was defined as the inactive state; the 1R-weak, 1R-open, and 2R-weak forms were defined as the active state; the 1N-attached form was defined as the CTSL-attached state; and the 1N-2R-

weak form was defined as the fusion state. Three newly emerged forms adopted after CTSL treatment (1N-attached, 2R-weak, and 1N-2R-weak) are highlighted in red. The NTD, RBD and CTD functional domains are colored pale green, light pink and pale goldenrod, respectively; the bound CTSL protein is colored medium slate blue; and all other densities are colored light gray. The percentages of each population among the whole group are also indicated in brackets. b-e, Unique conformational states in the CTSL-treated group. (b) Comparison of the superimposed models of the S-1N-attached (CTSL-treated) structure to the S-closed (untreated) structure. The coloring scheme of S-1N-attached is as shown in Fig. 2A. The S-closed (untreated) structure is colored light cyan. The map covering CTSL and the nearest NTD is also shown, and the map density is colored cornflower blue. (c) Comparison of the superimposed models of the S-1N-2R-weak structure (CTSL-treated) to the S-closed (untreated) structure. The coloring scheme of the S-1N-2R-weak (CTSL-treated) S trimer is as shown in Fig. 2A, and S-closed (untreated) is colored light cyan. The two missing RBDs and one missing NTD in the S-1N-2R-weak structure are shown as blue surfaces. (d) The three missing domains in the S-1N-2R-weak (CTSL-treated) structure compared to the S-closed (untreated) structure are magnified and indicated. All missing atoms compared to the S-closed state are shown in blue as a surface map. (e) The five domains in the S-1N-2R-weak state of CTSL-treated S with obvious offsets compared to their locations in S-closed (untreated) S are magnified and indicated. The density map of the S-1N-2R-weak state of CTSL-treated S is colored cornflower blue. The offset distances compared to the corresponding locations in the S-closed (untreated) structure are labeled in angstroms. f, Cartoon illustrations of the conformational changes upon CTSL treatment. First, CTSL bound to the closed form (inactive state) of the S protein and induced a conformational change to the attached state. Second, CTSL was cleaved at CS-1 to increase the dynamics of the RBD and changed the conformation of S into the active state. Finally, CTSL was cleaved at CS-2 to induce a large shift in the S1 subunit and expose the S2 subunit for membrane fusion (fusion state).

Fig. 3

Figure 3

CTSL cleavage sites are essential for SARS-CoV-2 infection and efficient cell-cell fusion. a, Overview of SARS-2 S proteins with mutations in CS-1, CS-2 and the S1/S2 cleavage site. b, Infectivity of PsVs with different point mutations in CS-1 and CS-2 was assessed in 293T/hACE2, Huh7, Vero, and LLC-MK2 cells. PsV infectivity was measured by a luciferase assay and is shown as the raw luciferase activity (n=4). Statistical significance was assessed by one-way ANOVA with Tukey's post hoc test. c-e, Quantitative

analysis of syncytium formation induced by CS mutant SARS-CoV-2 S proteins (see also Supplementary Fig. 11). (c) Luciferase gene expression was driven by the ERE promoter, and ESR1 (activator) bound and activated the ERE promoter to upregulate luciferase expression. (d and e) Effector Huh7 cells were cotransfected with plasmids expressing ERE-luciferase and different S proteins as indicated, and another plate of target Huh7 cells was transfected with plasmid expressing ESR1. After 24 h, the effector cells were detached and added to the target cells for 30-60 min. Then, the supernatant was removed and treated with PBS or CTSL (8 μ g/mL) for 20 min. The reaction was stopped by adding 500 μ L of medium, and culture was continued for another 24 h to allow cell-cell fusion. When a target cell and effector cell fused to form a syncytium, ESR1 bound and activated the ERE promoter to upregulate luciferase expression. Luciferase activity was then measured as a proxy for the fusion rate. The data were normalized to the WT-PBS group (n=3). Statistical significance was assessed between the indicated group and the WT-PBS group by two-way ANOVA with Dunnett's post hoc test. f, Images of syncytium formation induced by CS mutant SARS-CoV-2 S proteins. Huh7 cells were transfected with plasmids to express the WT, CS-1M, CS-2M or CS-1M+2M S protein. Cells were treated in the absence (PBS, pH=5.8) or presence of CTSL (4 μ g/mL, pH=5.8). Images were acquired after an additional 10-16 h of incubation in medium (scale bars, 50 μ m). The black arrowheads indicate syncytia. Representative data from three independent experiments are shown. g, Overexpression or knockdown of the CTSL gene dose-dependently promoted or inhibited, respectively, infection with WT (Wuhan-1) and three mutant SARS-CoV-2 PsVs with different point mutations in the furin cleavage site (FM-delta, FM-ARAA, and FM-GSAS). PsV infectivity in Huh7 cells was measured by a luciferase assay and is shown as the relative luciferase activity (n=3). Statistical significance was assessed by one-way ANOVA with Tukey's post hoc test. h, CTSL promoted syncytium formation induced by the FM-ARAA mutant SARS-CoV-2 S protein. Effector cells were cotransfected with ERE-luciferase plasmids and either FM-ARAA S or scramble vectors (Control). Target cells were transfected with ESR1 expression plasmid. After the effector cells and target cells were mixed, the supernatant was removed and treated with PBS or CTSL (8 μ g/mL and 16 μ g/mL). Luciferase activity was then measured and normalized to that in the control group (n=4). Statistical significance was assessed by one-way ANOVA with Tukey's post hoc test. i, Huh7 cells were transfected with scramble vector or FM-ARAA S protein expression plasmid. Cells were treated in the absence (PBS) or presence of CTSL (2 or 4 μ g/mL). Images were acquired after an additional 10-16 h of incubation in medium (scale bars, 50 μ m). The black arrowheads indicate syncytia. Representative data from four independent experiments are shown. The data are presented as the mean $\pm$ s.e.m. values. * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.

Fig. 4

Figure 4

CTSL inhibitors prevent infection with SARS-CoV-2 and mutant variant PsVs in vitro. a, Vero E6 cells were pretreated with increasing concentrations of each compound for 16 h and were then infected with SARS-CoV-2 at an MOI of 0.01. At 24 h post infection, viral RNA copies in supernatants were quantified by RT-qPCR. The data were normalized to the average value in vehicle-treated cells and are shown as relative infection percentages. The EC₅₀ values for each compound are indicated. Cell viability was evaluated with a CCK kit (TransGen Biotech) (n=3). b, Vero E6 cells were pretreated with increasing concentrations of each compound for 16 h and were then infected with different SARS-CoV-2 variant PsVs as indicated. At 24 h post infection, infectivity was measured by a luciferase assay. The data were normalized to the average value in vehicle-treated cells and are shown as inhibition rates (n=3). Statistical significance was

assessed between the indicated variant and WT PsV by two-way ANOVA with Dunnett's post hoc test. The data are presented as the mean $\pm$ s.e.m. values. * $P < 0.05$, ** $P < 0.01$.

Fig. 5

Figure 5

CTSL inhibitors prevent SARS-CoV-2 infection in vivo. a, E64d and Z-FY-CHO were administered intraperitoneally at -2 ~ 3 dpi as prophylactic treatment, and mice were challenged with 106 PFU at 0 dpi; the two drugs were administered therapeutically at 0~3 dpi. Tissue samples were collected at 4 dpi. b, Viral RNA copies in mouse lung and nasal turbinate tissues (n=5 mice/group). The dotted line indicates the lower limit of detection (LOD). Statistical significance was assessed between the indicated group and

control group by one-way ANOVA with Tukey's post hoc test. c, Representative images from histological analysis of lungs from SARS-CoV-2-infected hACE2 mice at 4 dpi. Magnified views of the boxed regions in each image are shown below the corresponding image. The black arrows indicate inflammatory cell infiltration, the black arrowheads indicate bronchiolar epithelial cell degeneration, and the red arrows indicate alveolar septal thickening. d, Semiquantitative histological scoring of each lung tissue was performed by grading the severity of bronchiolar epithelial cell damage (0-10), alveolar damage (0-10) and inflammatory cell infiltration in blood vessels and bronchioles (0-10) and summing these scores to calculate the total score. Normal = 0, indeterminate = 1-2, mild = 3-4, moderate = 5-7, severe = 8-10. (n=3) Statistical significance was assessed between the indicated group and control group by one-way ANOVA with Tukey's post hoc test.

Fig. 6

Figure 6

Proposed mechanism by which CTSL promotes SARS-CoV-2 infection. SARS-CoV-2 S contains S1 and S2 subunits. CS-1 is located in the NTD of the S1 subunit, and CS-2 is located near the S1/S2 site. CTSL cleaves the SARS-CoV-2 S protein: (1) By binding with the S protein on the surface of SARS-CoV-2, CTSL cleaves S at CS-1 to increase the dynamics of the RBD and make it accessible to ACE2 for binding. (2) CTSL further cleaves S at CS-2 to activate the S2 subunit for membrane fusion. (3) The virus fuses with the cell membrane, and the viral genetic material is released into the host cell.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MovieS1.mp4](#)
- [SupplementaryInformation.pdf](#)