bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Characterisation of protease activity during SARS-CoV-2 infection identifies novel viral cleavage sites and cellular targets for drug repurposing.

Bjoern Meyer [1], Jeanne Chiaravalli [2], Philip Brownridge [3], Dominic P. Bryne [4], Leonard A. Daly [3], Fabrice Agou [2], Claire E. Eyers [3], Patrick A. Eyers [4], Marco Vignuzzi [1], and Edward Emmott [3,✉]

[1]Viral Populations and Pathogenesis Unit, CNRS, UMR 3569, Institut Pasteur, Paris, CEDEX 15, France
[2]Chemogenomic and Biological Screening Core Facility, C2RT, Departments of Cell Biology & Infection and of Structural Biology Chemistry, Institut Pasteur, Paris, CEDEX 15, France
[3]Centre for Proteome Research, Department of Biochemistry Systems Biology, Institute of Systems, Molecular & Integrative Biology, Biosciences Building, Crown Street, University of Liverpool, Liverpool, L69 7ZB, UK
[4]Department of Biochemistry & Systems Biology, Institute of Systems, Molecular & Integrative Biology, Biosciences Building, Crown Street, University of Liverpool, Liverpool, L69 7ZB, UK

Abstract

SARS-CoV-2 is the causative agent behind the COVID-19 pandemic, and responsible for tens of millions of infections, and hundreds of thousands of deaths worldwide. Efforts to test, treat and vaccinate against this pathogen all benefit from an improved understanding of the basic biology of SARS-CoV-2. Both viral and cellular proteases play a crucial role in SARS-CoV-2 replication, and inhibitors targeting proteases have already shown success at inhibiting SARS-CoV-2 in cell culture models. Here, we study proteolytic cleavage of viral and cellular proteins in two cell line models of SARS-CoV-2 replication using mass spectrometry to identify protein neo-N-termini generated through protease activity. We identify multiple previously unknown cleavage sites in multiple viral proteins, including major antigenic proteins S and N, which are the main targets for vaccine and antibody testing efforts. We discovered significant increases in cellular cleavage events consistent with cleavage by SARS-CoV-2 main protease, and identify 14 potential high-confidence substrates of the main and papain-like proteases. We showed that siRNA depletion of these cellular proteins inhibits SARS-CoV-2 replication, and that drugs targeting two of these proteins: the tyrosine kinase SRC and Ser/Thr kinase MYLK, showed a dose-dependent reduction in SARS-CoV-2 titres. Overall, our study provides a powerful resource to understand proteolysis in the context of viral infection, and to inform the development of targeted strategies to inhibit SARS-CoV-2 and treat COVID-19 disease.

SARS-CoV-2 | Coronavirus | Protease | N-terminomics | Proteomics
Correspondence: ***e.emmott @liverpool.ac.uk***

Introduction

SARS-CoV-2 emerged into the human population in late 2019, as the latest human coronavirus to cause severe disease following the emergence of SARS-CoV and MERS-CoV over the preceding decades (1, 2). Efforts to develop vaccines and therapeutic agents to treat COVID-19 disease are well underway, however it is widely expected that this first generation of treatments might provide imperfect protection from disease. As such, in-depth characterisation of the virus and its interactions with the host cell can inform current and next-generation efforts to test, treat and vaccinate against SARS-CoV-2. Past efforts in this area have included the proteome, phosphoproteome, ubiquitome and interactome of SARS-CoV-2 viral proteins and infected cells (3–9). Proteolytic cleavage plays a crucial role in the life cycle of SARS-CoV-2, and indeed most positive-sense RNA viruses. Inhibitors targeting both viral and cellular proteases have previously shown the ability to inhibit SARS-CoV-2 replication in cell culture models (10–13). Here we present a first unbiased study of proteolysis during SARS-CoV-2 infection, and its implications for viral antigens, as well as cellular proteins that may represent options for antiviral intervention.

Proteolytic cleavage of the two coronavirus polyproteins generates the various viral proteins needed to form a replication complex, required for transcription and replication of the viral genome and subgenomic mRNAs. The key viral enzymes responsible are the papain-like (PLP, nsp3) and main proteases (MPro, nsp5). Aside from cleaving viral substrates, these enzymes can also act on cellular proteins, modifying or neutralising substrate activity to benefit the virus. A recent study highlighted the ability of the viral proteases to cleave proteins involved in innate immune signaling including IRF3, NLRP12 and TAB1 (14). However, there has yet to be an unbiased study to identify novel substrates of the coronavirus proteases in the context of viral infection. The identification of such substrates can identify cellular enzymes or pathways required for efficient viral replication that may represent suitable targets for pharmaceutical repurposing and antiviral intervention for the treatment of COVID-19 disease.

Viral proteins can also be the targets of cellular proteases, with the most prominent example for coronaviruses being the cleavage of the Spike glycoprotein by the cellular proteases furin, TMPRSS2 and Cathepsins (10, 11, 15, 16), but

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the exact cleavage sites within Spike for most of these individual cellular proteases are not yet characterised. Proteolytic processing can also be observed for other coronavirus proteins, for example, signal peptide cleavage of SARS-CoV ORF7A (17) and caspase cleavage of the nucleocapsid protein (18, 19). Many of these viral proteins, and especially the Spike glycoprotein form part of vaccine candidates currently undergoing clinical trials. For a functional immune response, it is vital that the antigens presented to the immune system, as part of these vaccines, closely mimic those seen in natural infection. An understanding of any modifications to these antigens observed during natural infection, such as glycosylation, phosphorylation and proteolytic cleavage, is critical to enable the rational design and validation of vaccine antigens and the selection of appropriate systems for their production. Mass spectrometry-based proteomic approaches have already led to rapid advances in our understanding of SARS-CoV-2, with notable examples including the rapid release of the cellular interactome (6) and proximity interactome (7) for a majority of SARS-CoV-2 proteins, as well as proteomic (3, 5), phosphoproteomic (4, 8) and ubiquitomic analyses (9) Larger scale-initiatives have been launched focusing on community efforts to profile the immune response to infection, and provide in-depth characterization of viral antigens (20). Mass spectrometry has particular advantages for investigation of proteolytic cleavage as analysis can be conducted in an unbiased manner, and identify not only the substrate, but the precise site of proteolytic cleavage (21).

In this work we have applied mass spectrometry-based methods for N-terminomics to study proteolysis and the resulting proteolytic proteoforms generated in the context of SARS-CoV-2 infection, enabling the identification of novel cleavage and processing sites within viral proteins. We also identify cleavage sites within cellular proteins that match the coronavirus protease consensus sequences for Mpro and PLP, show temporal regulation during infection, and demonstrate these proteins are required for efficient SARS-CoV-2 replication. These potential SARS-CoV-2 protease substrates include proteins that can be targeted with drugs in current clinical use to treat other conditions (22). Indeed, we demonstrate potent inhibition of SARS-CoV-2 replication with two compounds that are well-established chemical inhibitors of the SARS-CoV-2 protease substrates SRC and myosin light chain kinase (MYLK).

Results

Proteomic analysis of SARS-CoV-2-infected cell lines identifies alterations to the N-terminome. To investigate proteolysis during SARS-CoV-2 infection, N-terminomic analysis at various timepoints during the course of SARS-CoV-2 infected Vero E6 and A549-Ace2 cells (Fig. 1A) was performed. Vero E6 cells are an African Green Monkey kidney cell line commonly used for the study of a range of viruses, including SARS-CoV-2 which replicates in this cell line to high titres. A549-Ace2 cells are a human lung cell line which has been transduced to overexpress the ACE2 receptor to allow for SARS-CoV-2 entry. Cells were infected in biological triplicates at a multiplicity of infection (MOI) of 1, and harvested at 4 timepoints (0, 6, 12 and 24h) post-infection. Mock-infected samples were collected at 0 and 24h post-infection. These timepoints were chosen to cover SARS-CoV-2 infection from virus entry, over replication to virus egress: RNA levels increased from 9h post-infection (Fig. 1B), protein levels showed steady increases throughout infection (Fig. 1C), and viral titres increased at the 24h timepoint (Fig. 1D). These features were shared in both cell lines, with the Vero E6 cells showing greater RNA and protein levels, as well as viral titres compared with the A549-Ace2 cells.

Analysis of the N-termini-enriched samples was performed by LC-MS/MS following basic reverse phase fractionation. For the purposes of this analysis, neo-N-termini were taken to be those beginning at amino acid 2 in a given protein or later. By this definition these neo-N-termini will include those with post-translational removal of methionine, signal peptide cleavage, as well as those cleaved by viral or cellular proteases. The modified N-terminomic enrichment strategy used (21) employed isobaric labelling (TMTpro) for quantification as this permitted all samples to be combined prior to enrichment, minimising sample variability. This strategy meant that only those peptides with a TMTpro-labelled N-terminus or lysine residue were quantified. As only unblocked N-termini are labelled with undecanal, this approach results in the selective retention of undecanal-tagged tryptic peptides on C18 in acidified 40% ethanol, with N-terminal and neo-N-terminal peptides enriched in the unbound fraction (21).

Quality filtering of the dataset was performed (Fig. S1), infected and mock samples separated by PCA and 0h Mock, 0h infected and 6h infected clustered together, and away from the 12h and 24h infected samples (Fig. S1A-D). With the exception of the enriched Vero E6 dataset the 24h mock sample clustered with the 0h mocks. The Vero E6 24h Mock clustered away from the 0h and infected samples which may reflect regulation due to cell confluence as this was not observed with the paired unenriched sample. Sample preparation successfully enriched for blocked N-termini consisting of acetylated, pyroglutamine-N-termini and TMTpro-labelled N-termini (Fig. S1), and blocked N-termini were more abundant in the enriched samples. In both datasets, TMTpro-labelled N-termini represent approximately 50% of the blocked N-termini, with the rest split evenly between pyroglutamine and N-terminal acetylation (Fig. S1). After filtering, over 2700 TMTpro-labelled N-termini representing neo-N-termini were identified from each cell line.

When the 24h infected and mock-infected timepoints were compared, both cellular and viral neo-N-termini in A549-Ace2 cells (Fig. 1) and Vero E6 cells (Fig. 1F) were identified as showing significant alterations in their abundance. In line with expectation, N-termini from viral proteins were solely identified as showing increased abundance during infection in both cell lines. N-termini from cellular proteins showed both increased and decreased abundance during infection. We reasoned that those neo-N-termini showing increased abundance would include viral neo-N-termini, as well as those cellular

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

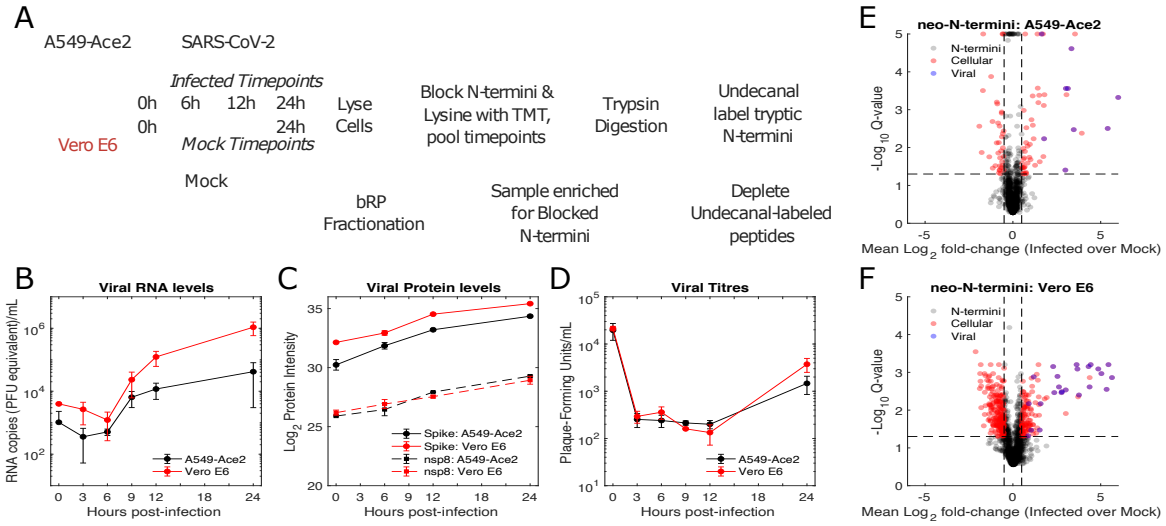

Fig. 1. N-terminomic analysis of SARS-CoV-2 infection of A549-Ace2 and Vero E6 cells. A) Experimental design. B) Viral RNA levels were determined by qRT-PCR. C) Protein levels were determined based on the TMTpro fractional intensity of the total protein intensity for the unenriched proteomic samples. D) Infectious virus production (PFU). E) A549-Ace2 and F) Vero E6 neo-N-terminomic analysis reveals significant increases in peptides corresponding to viral and cellular neo-N-termini, where neo-N-termini must begin from amino acid 2 or later. Error bars represent standard deviation, P-values were obtained by t-test, correction for multiple hypothesis testing to obtain Q-values was performed as described Storey (2002)(23)

proteins cleaved by the SARS-CoV-2 PLP and Mpro proteases. For this study we therefore focused specifically on viral N-termini and those cellular neo-N-termini identified as showing significantly increased abundance (t-test, multiple hypothesis testing corrected Q value ≤ 0.05) during infection.

Novel proteolytic processing of SARS-CoV-2 proteins is observed during infection. The 30kb SARS-CoV-2 genome encodes a large number of proteins including two long polyproteins formed through ribosomal frameshifting, the structural proteins S, E, M and N and a range of accessory proteins (Fig. 2). Coronavirus proteins, in line with those of other positive-sense RNA viruses are known to undergo post-translational modifications, including proteolytic cleavage in some cases. Across all datasets we identified the S, M and N structural proteins, with the exception of E which has also not been observed in other proteomics datasets due to both short length and sequence composition (3, 5). We identified the ORF3a, ORF6, ORF8 and ORF9b accessory proteins, and all domains of the polyprotein aside from nsp6, 7 and 11.

We first sought to characterise neo-N-termini from viral proteins to understand potential patterns of cleavage that might generate functional proteolytic protoforms of the viral proteins. Neo- and N-termini were identified from 8 viral proteins including the polyprotein (Fig. 2B-D; Fig. S2). Of these the nucleocapsid (N), ORF3a accessory protein and Spike were most prominent. More cleavage sites were observed from infected Vero E6 cells than A549-Ace2 cells, which is in line with expectation given the higher levels of viral protein expression, and superior infectivity of this cell line compared to the A549-Ace2 cell line.

The coronavirus N protein is highly expressed during infection, and also represents a major antigen detected by the host immune response. Prior studies have identified cleavage of the SARS-CoV N protein by cellular proteases (18, 19), and our data identified multiple neo-N-termini consistent with proteolytic cleavage from both infected A549-Ace2 and Vero E6 cells (Fig. 2B). neo-N-termini common to both datasets include amino acids 17, 19, 69, 71, 76, 78, 154, and 263. Many of these cleavage sites were spaced closely together (e.g. 17/19, 69/71), consistent with a degree of further endo-proteolytic processing.

The ORF3a putative viroporin also shows N-terminal processing, possibly reflecting signal peptide cleavage (Fig. 2C). In a recent study, cryoEM of ORF3a in lipid nanodiscs did not resolve the first 39 N-terminal suggesting this region is unstructured (24). We observed N-terminal processing sites in the first 22 residues of the protein, with neo-N-termini beginning at amino acids, 10, 13 and 16 identified in both datasets, giving a possible explanation for the lack of N-terminal amino acids in cryoEM experiments.

Proteolytic cleavage of the Spike glycoprotein is of major interest as it can play an important role in cell entry, with different distributions of cellular proteases between cell types resulting in the usage of different entry pathways, as well as potentially changing availability of surface epitopes for antibody recognition. Key proteases include furin, TMPRSS2 and cathepsins, though in the latter two cases the actual cleavage sites targeted by these enzymes to process Spike into S1 and S2 remain unclear. Consistent with previous observations (3, 5), we do not detect a neo-N-terminus deriving from the furin cleavage site as the trypsin digestion we employed would not be expected to yield peptides of suitable length for analysis. However, while beneficial for replication, furin cleavage is not essential and other cleavage events within Spike can compensate (15, 16). We detect neo-N-terminal peptides from S637 in both datasets (Fig. 2D). In line with the pattern of viral gene expression observed in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

unenriched datasets this neo-N-terminus showed consistent increases in abundance throughout the experimental time-course (Fig. 2E). S637 is located on a flexible loop near the furin cleavage site (Fig. 2F), suggesting it is accessible for protease cleavage (25). A mass spectrum for the S637 neo-N-terminus from the A549-Ace2 dataset is shown in Fig. 2G, the same peptide was observed with both 2+ and 3+ charge states in the Vero E6 dataset, and with a higher Andromeda score (124.37 vs. 104.82). Intriguing, S637 was identified as a phosphorylation site in Davidson et al. (3). As phosphorylation can inhibit proteolytic cleavage when close the the cleavage site, this suggests potential post-translational regulation of this cleavage event.

Further neo-N-termini from Spike were identified in the Vero E6 dataset alone, including a neo-N-terminus beginning at Q14. This is slightly C-terminal of the predicted signal peptide which covers the first 12 amino acids. This peptide featured N-terminal pyroglutamic acid formed by cyclization of the N-terminal glutamine residue. The peptide does not follow an R or K residue in the Spike amino acid sequence and thus represents non-tryptic cleavage. The absence of TMT-pro labelling at the N-terminus suggests that this N-terminus was blocked prior to tryptic digestion, with this modified N-terminus preventing TMTpro modification. Artifactual cyclization of N-terminal glutamine or glutamic acid residues typically results from extended trypsin digestion and acidic conditions (26). However, the order of labelling and digestion steps in our protocol, and non-tryptic nature of this peptide suggests that this N-terminal pyroglutamic acid residue is an accurate reflection of the state of this neo-N-terminus in the original biological sample. Three further N-terminal pyroglutamic acid residues were identified in SARS-CoV-2 proteins within the Vero E6 dataset and can be found in table S2.

We detected viral neo-N-termini and N-termini in M, ORF7a, ORF9b and pp1ab. Due to conservation with SARS-CoV ORF7a, the first 15 residues of SARS-CoV-2 ORF7a are expected to function as a signal peptide which is post-translationally cleaved (17, 27). neo-N-termini were identified in both datasets consistent with this hypothesis. Due to inclusion of the ORF7a iORF1 proposed N-terminal truncation of ORF7a which lacks the first two amino acids in ORF7a in the SARS-CoV-2 sequences used for data analysis, the start position of this neo-N-terminal peptide is given as 14 (28). However, this would be position 16 in ORF7a, consistent with removal of the signal peptide (MKIILFLALITLATC , in Uniprot P0DTC7), and conserved with that in SARS-CoV ORF7a.

The native N-terminus of ORF9b was also identified, and several sites mapping to the replicase polyprotein, including a conserved neo-N-terminus consistent with predicted nsp10-nsp12 cleavage by Mpro. A neo-N-terminus consistent with nsp15-nsp16 cleavage by Mpro was identified in A549-Ace2 cells, and several internal neo-N-termini deriving from nsp1, -2 and -3 were also observed, though not common to both datasets. All the viral neo-N-termini and N-termini identified in this study can be found in tables S1 (A549-Ace2) and S2 (Vero E6) respectively. Table S3 includes all viral peptides identified in this study in both enriched and unenriched datasets.

SARS-CoV-2 infection induces proteolytic cleavage of multiple host proteins. The consensus sequences for coronavirus proteases are conserved between coronaviruses, with PLP recognising a P4 to P1 LxGG motif, and Mpro recognising a (A|P|S|T|V)xLQ motif (29). No strong preference has been identified for either protease at the P3 residue (Fig. 3A). Analysis of both datasets showed strong enrichment for neo-N-termini consistent with cleavage at Mpro motifs (two-tailed Kolmogorov-Smirnov test, $p<0.001$, Fig. 3B,C). However, no comparable enrichment could be seen for neo-N-termini consistent with cleavage at PLP motifs (Fig. 3D,E). This may reflect fewer cellular protein substrates of PLP compared to Mpro, or higher background levels of neo-N-termini generated by cellular proteases with similar P4 to P1 cleavage specificities as PLP.

Neo-N-termini matching, or close to the consensus sequences, for either Mpro or PLP and showing significant upregulation (t-test, $q \leq 0.05$ after correction for multiple hypothesis testing) at 24h post-infection compared to the 24h mock sample were selected for further analysis. Perfect matches to the consensus sequences from A549-Ace2 cells included NUP107, PAICS, PNN, SRC and XRCC1. GOLGA3 and MYLK (MCLK) were identified from Vero E6 cells. Hits from both cell lines that resembled, but did not completely match the consensus sequence were ATAD2, ATP5F1B, BST1, KAT7, KLHDC10, NUCKS1 and WNK1 (Fig. 3F, G). Adding confidence to these observations, approximately half of these hits were also identified in a recent SARS-CoV-2 proximity labelling study (ATP5F1B, GOLGA3, NUP107, PNN, SRC and WNK)(7), and GOLGA3 was additionally identified in an interactome study as an nsp13 interaction partner (6).

SRC, MYLK and WNK are all protein kinases, one of the protein families best studied as drug targets (30). MYLK is especially interesting as dysregulation of MYLK has been linked to acute respiratory distress syndrome - one of the symptoms of severe COVID-19 disease (31). NUP107 is a member of the nuclear pore complex, with nucleo-cytoplasmic transport a frequent target for viral disregulation (32). GOLGA3 is thought to play a role in localisation of the Golgi and Golgi-nuclear interactions, and was identified in two recent studies of SARS-CoV-2 interactions (6, 7). PNN is a transcriptional activator, forming part of the exon junction complex, with roles in splicing and nonsense-mediated decay. The coronavirus mouse hepatitis virus has previously been shown to target nonsense mediated decay, with pro-viral effects of inhibition (33). PAICS and BST1 both encode enzymes with roles in ADP ribose and purine metabolism respectively, with PAICS previously identified as binding the influenza virus nucleoprotein (34).

The majority of these neo-N-termini showed enrichment at 24h, with levels remaining largely unchanged at earlier time-points, especially for Mpro substrates (Fig. 3F,G). This matches the timing for peak viral RNA, protein expression

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

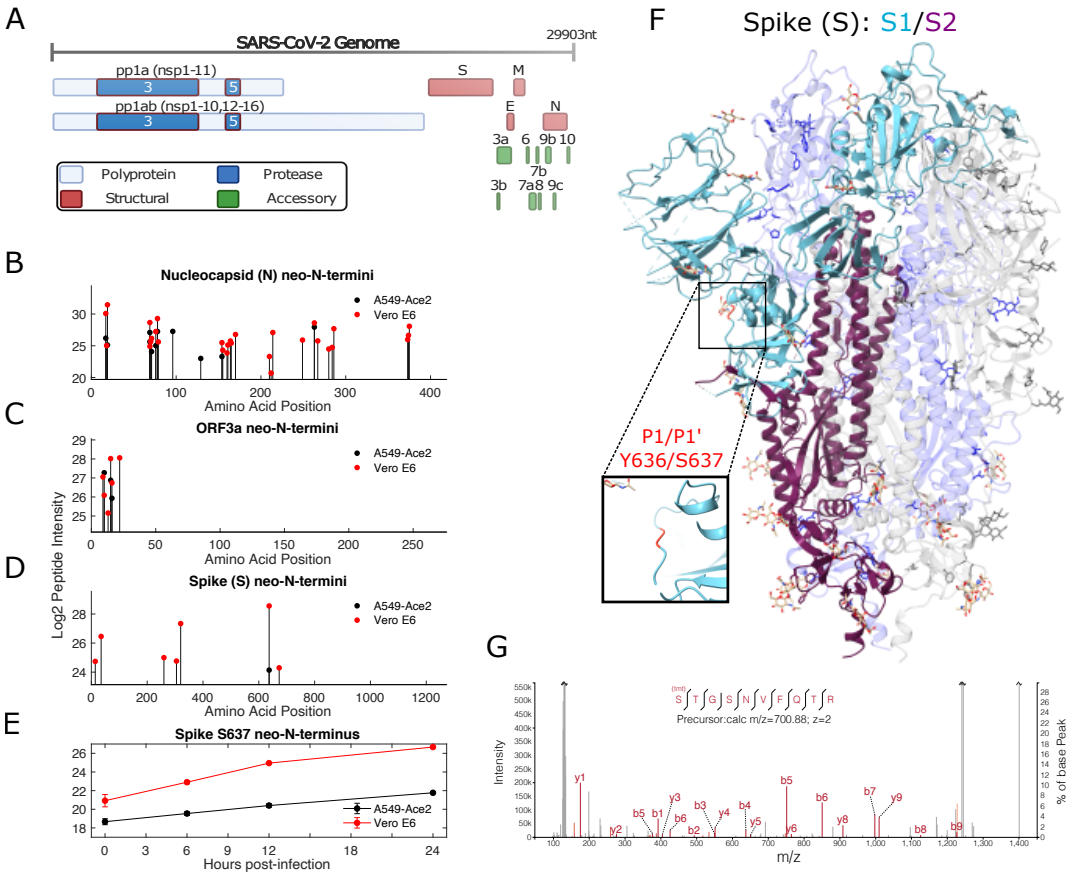

Fig. 2. Proteolysis of viral proteins during SARS-CoV-2 infection. A) Schematic of the SARS-CoV-2 genome and proteome, with the nsp3 (PLP) and nsp3 (MPro) highlighted. Proteolytic processing of SARS-CoV-2 proteins during infection of A549-Ace2 and Vero E6 cells includes B) extensive cleavage of the nucleocapsid protein C) N-terminal processing of the ORF3a putative viroporin, and D) a novel cleavage site between Y636 and S637 in Spike, N-terminal of the furin cleavage site. E) The abundance of the S637 spike neo-N-terminus increases over the infection timecourse, F) This cleavage site is present on an flexible region, C-terminal of the RBD (PDB: 6X6P G) A mass spectrum for the S637 peptide from A549-Ace2 cells is shown.

and titres over the timepoints examined (Fig. 1B-D). Exceptions to this trend include the potential PLP substrates, 2/3 of which begin to show increased abundance at 12h post-infection, with BST1 appearing to peak at 12h rather than 24h, indicating a potential temporal regulation of the two viral proteases. Data for all quantified and filtered N- and neo-N-termini from A549-Ace2 and Vero E6 cells is available in tables S4 and S5 respectively.

Prospective MPro and PLP substrates are necessary for efficient viral replication, and represent targets for pharmacological intervention. To investigate if the putative cellular substrates of MPro and PLP identified in the N-terminomic analyses are necessary for efficient viral replication, an siRNA screen was conducted Fig. 4. Where proteolytic cleavage inactivates cellular proteins or pathways inhibitory for SARS-CoV-2 replication, siRNA depletion would be anticipated to result in inreased viral titres and/or RNA levels. If proteolysis results in altered function that is beneficial for the virus, we would expect siRNA depletion to result in a reduction in viral titres/RNA levels. Proteins with neo-N-termini showing statistically significant increased abundance during SARS-CoV-2 infection and either matching, or similar to the viral protease consensus sequences were selected for siRNA depletion.

Infection of A549-Ace2 cells was performed 24h post-transfection with the indicated siRNA and allowed to proceed for 72h (Fig. 4A). Cell viability for all targets was comparable to untreated controls (Fig. S3). siRNA knockdown efficiency at the time of infection was confirmed by qRT-PCR (Fig. S4), with a low of 77% efficiency for NUCKS1, and averaging over 95% efficiency for most targets.10/14 coronavirus protease substrates showed significant reductions (one-way ANOVA, $p \leq 0.01$) in viral RNA levels, averaging a 100-1000-fold median decrease in viral RNA equating to pfu equivalents per ml at 72h post-infection compared to treatment with a control siRNA (Fig. 4B). PAICS, GOLGA3, NUCKS1, and XRCC1 did not show a significant drop in RNA copy number following siRNA treatment. Plaque assays were then conducted on these samples to determine whether this observed reduction in viral RNA levels reflected a reduction in infectious virus titres (Fig. 4C). All 14 potential substrates showed a statistically significant (one-way ANOVA, $p \leq 0.01$) reduction in viral titres following siRNA depletion. For PAICS and GOLGA3, which did not show reduced RNA levels, these reductions were approx-

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

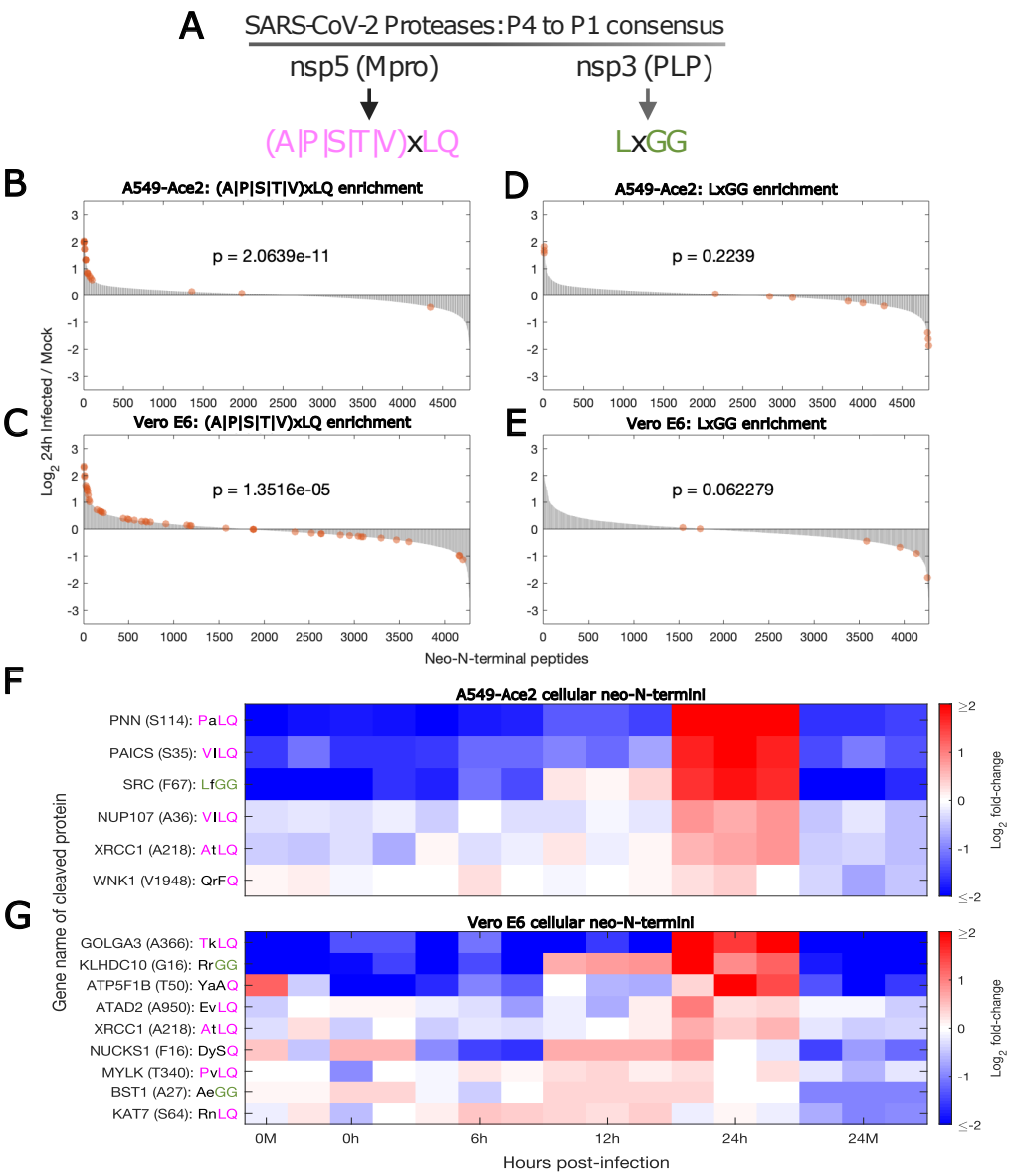

Fig. 3. Increased abundance of novel cellular neo-N-termini consistent with SARS-CoV-2 protease consensus sequences suggests viral protease activity on cellular substrates. a) Consensus motifs for Mpro and PLP. b) and c) Distribution of neo-N-termini consistent with the Mpro consensus motif in A549-Ace2 and Vero E6 cells respectively. d) and e) Distribution of neo-N-termini consistent with the PLP consensus motif in A549-Ace2 and Vero E6 cells respectively. Distributions cover all three biological replicates. Enrichment was determined by two-tailed Kolmogorov-Smirnov test. f) and g) show the relative abundance of cellular neo-N-termini identified as significantly upregulated (t-test, multiple-hypothesis corrected q <= 0.05) and matching or resembling the Mpro or PLP consensus motifs from A549-Ace2 or Vero E6 cells respectively. Sequence match to the consensus is indicated by the pink or green coloring of the P4 to P1 positions of the relevant cleavage sites indicating match to the Mpro or PLP P4, P2 or P1 positions respectively.

imately 10-fold. Most other siRNA targets showed reduced titres in the 100-1000-fold range. These differences in outcome between viral RNA levels and plaque assays may result from a subset of proteins required for efficient viral replication. While efficient mRNA knockdown was shown for all targets Fig. S4, it is also possible that this discrepancy between viral RNA levels and titres may result from differences in protein half-life of the knockdown targets. This could result in proteins with longer half lives only giving a phenotype at later stages of infection when infectious virus is produced.

A subset of the prospective viral protease substrates have commercially-available inhibitors, notably SRC and MYLK. In the case of SRC these include tyrosine kinase inhibitors in current clinical use. In light of the siRNA screening results we concluded that pharmacological inhibition of SARS-CoV-2 protease substrates could represent a viable means to inhibit SARS-CoV-2 infection. Dose-response experiments were conducted with 7 inhibitors to determine whether pharmacological inhibition of SARS-CoV-2 protease substrates could be employed as a potential therapeutic strategy (Fig. 5; Fig. S5). Of these, two tyrosine kinase inhibitors: Bafetinib and Sorafenib showed inhibition at concentrations which did not result in cytotoxicity in the human cell line A549-Ace2 (Fig. 5). In the case of Bafetinib, a Lyn/Bcr-Abl inhibitor which has off-target activity against SRC the IC_{50} was in the nanomolar range (IC_{50}: 0.79 µM, 95% confidence in-

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

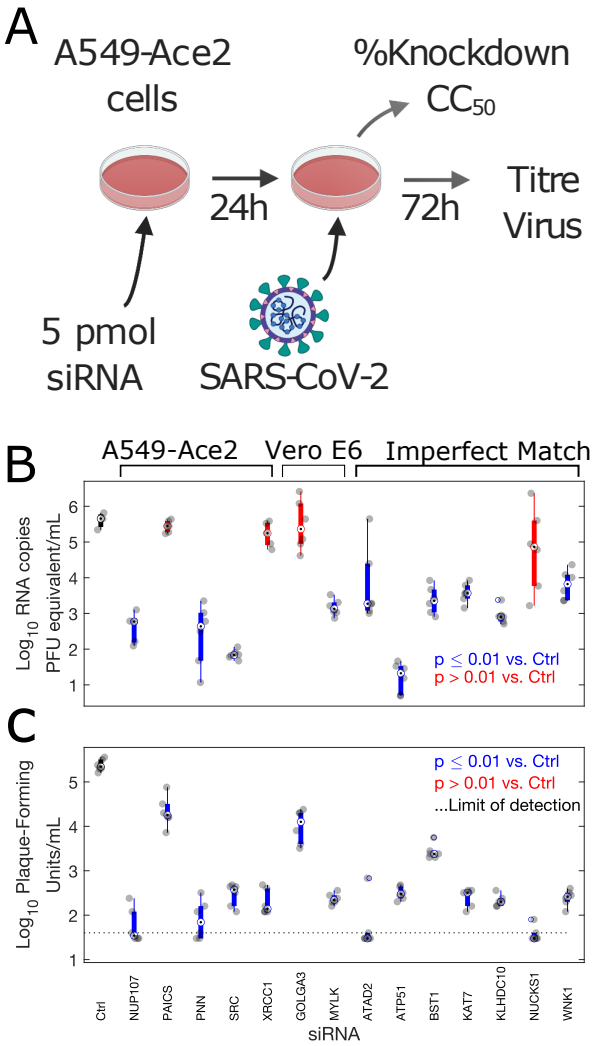

Fig. 4. siRNA depletion of potential MPro and PLP substrates results in significant reductions to viral RNA copies and titres in A549-Ace2 cells. A) Experimental design, B) Viral RNA copies and C) titres in supernatant at 72h post-infection, following infection 24h post-transfection with the indicated siRNA. Colored bars indicate median (hollow circle) and 25th and 75th percentiles. Individual datapoints are shown in grey. Significance was calculated by one-way ANOVA. Blue bars indicate samples with significantly reduced viral RNA copies or titres (p<=0.01). Those not meeting this threshold are shown in red. The control siRNA-treated sample is indicated with a black bar. The limit of detection in the plaque assay was calculated to be 40 PFU/mL (dotted line).

terval 0.23-1.35 µM). Bafetinib has recently been independently identified as an inhibitor of the coronaviruses OC43 and SARS-COV-2 in a large-scale drug-repurposing screen (35). Inhibition with Sorafenib which was included as a positive control and does not directly target any of the protease substrates was in the low micromolar range (Fig. 5), in line with a previously published report (4). Two inhibitors were trialed against MYLK. These were MLCK inhibitor peptide 18, and ML-7. Only ML-7 showed inhibition of SARS-CoV-2, with inhibition in the low micromolar range (IC_{50}: 1.7 µM, 95% confidence interval 1.51-1.80 µM), at concentrations which did not induce cytotoxicity (Fig. 5; Fig. S6). ML-7 and MLCK inhibitor peptide 18 have different mechanisms of action, with MLCK inhibitor peptide 18 outcompeting kinase substrate peptides, and ML-7 inhibiting ATPase activity. All four had CC_{50} values over the 10µM maximum

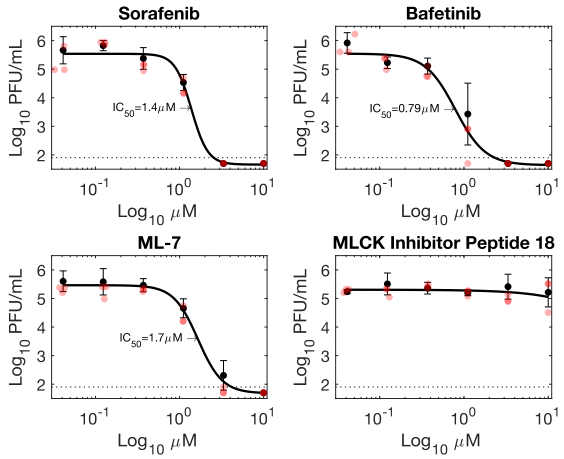

Fig. 5. Inhibitors targeting viral protease substrates reduce SARS-CoV-2 titres in A549-Ace2 cells. Sorafenib is a tyrosine kinase inhibitor previously shown to inhibit SARS-CoV-2 replication (4). Bafetinib is a tyrosine kinase inhibitor with off-target activity against SRC kinase. ML-7 and MLCK target myosin light chain kinase (MYLK/MLCK). Error bars represent standard deviation from 3 biological replicates. Red circles indicate individual datapoints..

concentration tested, except ML-7 which had a CC_{50} of 5 µM. Bafetinib did show reduced viability at the two highest concentrations tested (10 µM, 3.3 µM), though not reaching 50% reduction (Fig. S6).

The other 3 tyrosine kinase inhibitors tested (Bosutinib, Saracatinib, Dasatinib) all showed inhibition Fig. S5, however cytotoxicity results obtained with the assay used were also high preventing the unambiguous determination of whether inhibition was specific or due to cytotoxicity Fig. S6. However, it should be noted that these agents have been reported to be cytostatic in A549 cells, and the CellTiter Gloassay used to assess viability measures cellular metabolism so will not distinguish between cytostatic and cytotoxic effects.

Discussion

We employed a mass spectrometry approach to study proteolytic cleavage events during SARS-CoV-2 infection. Substrates of viral proteases are frequently inferred through studies of related proteases (36). However, such approaches are unable to identify novel substrates, and even closely-related proteases can differ in their substrate specificity (37). Mass spectrometry-based approaches to identify protease substrates by identifying the neo-N-terminal peptides generated by protease activity have existed for a number of years (21, 38–40), however, they have seen only limited application to the study of viral substrates (41), and have not been previously applied to the study of proteolysis during coronavirus infection.

While our approach identified multiple novel viral and cellular cleavage sites, it also failed to identify multiple known cleavage sites, including the furin cleavage site in Spike, and multiple cleavage sites within the viral polyprotein. This can be understood from the dependence of the approach on the specific protease used for mass spectrometry analysis. Isobaric labelling prior to trypsin digestion blocks tryptic cleav-

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

age at lysine residues and causes trypsin to cleave solely after arginine residues. This results in the generation of long peptides and if the specific cleavage site does not produce a peptide of suitable length for analysis (typically 8-30 amino acids) then it will be missed. This can be alleviated through the application of multiple mass spectrometry-compatible proteases in parallel, yielding multiple peptides of different length for each cleavage site (42, 43). This would both increase the number of sites identified and cross-validate previously identified cleavage sites. These methods will likely prove a fruitful avenue for future investigations of proteolysis during infection with SARS-CoV-2 and other viruses that employ protease-driven mechanisms of viral replication.

Our approach identified multiple cleavage sites within viral proteins. In some cases, such as the nucleocapsid protein, cleavage by cellular proteases has been observed for SARS-CoV (18, 19), though the number of cleavage products observed was much higher in our study (Fig. 2). Compared to the gel-based approaches used in the past, our approach is much more sensitive for detecting when protease activity results in N-termini with ragged ends, due to further endoproteolytic activity. Examples of this in our data are particularly evident in Fig. 2 for the nucleocapsid and ORF3a where neo-N-termini appear in clusters. Cleavage sites within the nucleocapsid and Spike protein are of particular interest as these are the two viral antigens to which research is closely focuses for both testing and vaccination purposes. In this context, neo-N-termini are of interest as N-termini can be recognised by the immune response, as they are typically surface-exposed. Antibodies recognising neo-N-termini such sites will not be detected in tests using complete or recombinant fragments that do not account for such cleavage sites. Understanding cleavage events can also inform interpretation of protein structural analysis, for example in the ORF3a viroporin (24). Knowledge of cleavage sites can permit further analysis of Spike entry mechanisms, and vaccine design, especially when considering N-terminal modifications such as pyroglutamine which will impact antibody binding in this region.

Proteolytic cleavage can alter protein function in several ways, including inactivation, re-localisation, or altered function including the removal of inhibitory domains. Our siRNA screen showed knockdown of the majority of potential protease targets we identified was inhibitory to SARS-CoV-2 replication (Fig. 4). Indeed, no siRNA treatment resulted in higher viral titres or RNA levels, suggesting that inactivation is not the prime purpose of these cleavage events. This suggests that in many cases, proteolytic cleavage by viral proteases may be extremely targeted, serving to fine-tune protein activity, rather than merely serving as a blunt instrument to shut down unfavorable host responses. An improved understanding of the exact ways in which proteolytic cleavage modulates protein activity and serves to benefit viral replication will be crucial for targeting cellular substrates of viral proteases as a therapeutic strategy.

Limitations

In this study, we used two cell line models to characterise the effects of SARS-CoV-2 infection on protease activity and the generation of viral and cellular cleavage products. Notably, we tested the efficiency of several inhibitors against SARS-CoV-2 infection only in the context of the A549-Ace2 cell line model. These results present preliminary data that must be further validated in other models, in vivo, and through clinical trials before use in patients for the treatment of COVID-19 disease.

Materials & Methods

Cell culture Virus. Vero E6 (Vero 76, clone E6, Vero E6, ATCC® CRL-1586TM) authenticated by ATCC and tested negative for mycoplasma contamination prior to commencement were maintained in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific) containing 10% (v/v) fetal bovine serum (FBS, ThermoFisher Scientific) and penicillin/streptavidin (ThermoFisher Scientific). A549-Ace2 cells, a human lung epithelial cell line that over-expresses ACE2, were kindly provided by Oliver Schwartz (Institut Pasteur) (44). A549-Ace2 cells were cultured in DMEM supplemented with 10% FBS, penicillin/streptavidin and 10 μg/ml blasticidin (Sigma) and maintained at 37°C with 5% CO_2. The SARS-CoV-2 isolate BetaCoV/France/IDF0372/2020 was supplied through the European Virus Archive goes Global (EVAg) platform. Viral stocks were prepared by propagation in Vero E6 cells in DMEM supplemented with 2% FBS. All experiments involving live SARS-CoV-2 were performed in compliance with Institut Pasteur Paris's guidelines for Biosafety Level 3 (BSL-3) containment procedures in approved laboratories. All experiments were performed in at least three biologically independent samples.

SARS-CoV-2 titration by plaque assay. Vero E6 cells were seeded in 24-well plates at a concentration of 7.5×10^4 cells/well. The following day, serial dilutions were performed in serum-free MEM media. After 1 hour absorption at 37°C, 2x overlay media was added to the inoculum to give a final concentration of 2% (v/v) FBS / MEM media and 0.4% (w/v) SeaPrep Agarose (Lonza) to achieve a semi-solid overlay. Plaque assays were incubated at 37° C for 3 days. Samples were fixed using 4% Formalin (Sigma Aldrich) and plaques were visualized using crystal Violet solution (Sigma Aldrich).

Infections for N-terminomic/proteomic analysis. N-terminomic sample preparation is based around Weng et al. 2019 Mol. Cell. Proteomics, adapted for TMTpro-based quantitation (21, 45). Vero E6 or A549-Ace2 cells were seeded using 2×10^6 cells in T25 flasks. The following day cells were either mock infected or infected with SARS-CoV-2 at a MOI of 1 in serum-free DMEM at 37°C for 1 hour. After absorption, the 0 hour samples were lysed immediately, while the media for other samples was replaced with 2% FBS / DMEM (ThermoFisher Scientific) and incubated at 37°C

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

for times indicated before lysis. Cells were washed 3x with PBS (ThermoFisher Scientific) before lysing them in 100 mM HEPES pH 7.4 (ThermoFisher Scientific), 1% Igepal (Sigma Aldrich), 1% sodium dodecyl sulfate (SDS; ThermoFisher Scientific), and protease inhibitor (mini-cOmplete, Roche). Samples were then heated to 95°C for 5 minutes, before immediately freezing at -80°C. Samples were then thawed and incubated with benzonase for 30 min at 37°C. Sample concentrations were normalized by BCA assay, and 25µg of material from each sample was used for downstream processing.

DTT was added to 10mM and incubated at 37°C for 30 min, before alkylation with 50mM 2-chloroacetamide at room temperature in the dark for 30 min. DTT at 50mM final concentration was added to quench the 2-chloroacetamide for 20 min at room temperature. Samples were washed by SP3-based precipitation (REF). Each sample was resuspended in 22.5µL 6M GuCl, 30µL of 0.5M HEPES pH8, and 4.5µL TCEP (10mM final) and incubated for 30 minutes at room temperature.

0.5mg of individual TMTpro aliquots (Lot VB294905) were resuspended in 62uL of anhydrous DMSO. 57µL of the TMTpro was then added to each sample, mixed and incubated for 1.5h. Label allocation was randomized using the Matlab Randperm function. Excess TMTpro was quenched with the addition of 13µL of 1M ethanolamide and incubated for 45 min. All samples were combined for downstream processing. SP3 cleanup was performed on the combined samples. These were resuspended in 400µL of 200mM HEPES pH8, containing Trypsin gold at a concentration of 25ng/µL and incubated overnight at 37°C.

Samples were placed on a magnetic rack for 5 min. 10% of the samples was retained for the unenriched analysis. The remaining material was supplemented with 100% ethanol to a final concentration of 40%, undecanal added at an undecanal:peptide ratio of 20:1 and sodium cyanoborohydride to 30mM. pH was confirmed to be between pH7-8 and the samples were incubated at 37°C for 1h. Samples were then sonicated for 15 seconds, and bound to a magnetic stand for 1 min. The supernatant was retained and then acidified with 5% TFA in 40% ethanol. Macrospin columns (Nest group) were equilibrated in 0.1% TFA in 40% ethanol. The acidified sample was applied to the column, and the flow through retained as the N-terminal-enriched sample.

Both unenriched and enriched samples were desalted on macrospin columns (Nest group), before drying down again. Off-line basic reverse phase fractionation for both unenriched and enriched samples was performed on a Waters nanoAcquity with an Acquity UPLC M-Class CSH C18 130A 1.7µm, 300µm x 150µm column. The sample was run on a 70 minute gradient at 6µL/min flow rate. Gradient parameters were 10 min 3% B, 10-40 min 3-34% B, 40-45 min 34-45% B, 45-50 min 45-99%B, 50-60 min 99% B, 60.1-70 min 3% B. Buffers A and B were 10mM ammonium formate pH10, and 10mM ammonium formate pH10 in 90% acetonitrile respectively. Both samples were resuspended in buffer A, and 1 minute fractions were collected for 1-65 min of the run. These were concatenated into 12 (1:13:24...) or 5 fractions (1:6:11...) for unenriched and enriched samples respectively using a SunChrom Micro Fraction Collector. Samples were dried down and resuspended in 1% formic acid for LC-MS/MS analysis.

Mass spectrometry. LC-MS/MS analysis was conducted on a Dionex 3000 coupled in-line to a Q-Exactive-HF mass spectrometer. Digests were loaded onto a trap column (Acclaim PepMap 100, 2 cm x 75 microM inner diameter, C18, 3 microM, 100 Å) at 5 µL per min in 0.1%(v/v) TFA and 2%(v/v) acetonitrile. After 3 min, the trap column was set in-line with an analytical column (Easy-Spray PepMap® RSLC 15 cm x 50cm inner diameter, C18, 2 microlM, 100 Å) (Dionex). Peptides were loaded in 0.1%(v/v) formic acid and eluted with a linear gradient of 3.8–50% buffer B (HPLC grade acetonitrile 80%(v/v) with 0.1%(v/v) formic acid) over 95 min at 300 nl per min, followed by a washing step (5 min at 99% solvent B) and an equilibration step (25 min at 3.8% solvent). All peptide separations were carried out using an Ultimate 3000 nano system (Dionex/Thermo Fisher Scientific).. The Q-Exactive-HF was operated in data-dependent mode with survey scans aquired at a resolution of 60,000 at 200m/z over a scan range of 350-2000m/z. The top 16 most abundant ions with charge states +2 to +5 from the survey scan were selected for MS2 analysis at 60,000 m/z resolution with an isolation window of 0.7m/z, with a (N)CE of 30. The maximum injection times were 100ms and 90ms for MS1 and MS2 respectively, and AGC targets were 3e6 and 1e5 respectively. Dynamic exclusion (20 seconds) was enabled.

Data analysis. All data were analysed using Maxquant version 1.6.7.0 (46). Custom modifications were generated to permit analysis of TMTpro 16plex-labelled samples. FASTA files corresponding to the reviewed Human proteome (20,350 entries, downloaded 8th May 2020), and African Green monkey proteome (Chlorocebus sabeus, 19,223 entries, downloaded 16th May 2020). A custom fasta file for SARS-CoV-2 was generated from the Uniprot-reviewed SARS-CoV-2 protein sequences (2697049). This file was modified to additionally include the processed products of pp1a and pp1b, novel coding products identified by ribo-seq (28), as well as incorporate two coding changes identified during sequencing (Spike: V367F, ORF3a: G251V). All FASTA files, TMT randomisation strategy, and the modifications.xml file containing TMTpro modifications have been included with the mass spectrometry data depositions. Annotated spectra covering peptide N-termini of interest were prepared using xiSPEC v2(47).

Several different sets of search parameters were used for analysis of different experiments.

For analysis of unenriched material from fractionated lysates. Default MaxQuant settings were used with the following alterations. As the experimental design meant unenriched samples contained a majority of peptides lacking N-terminal TMT labelling, quantification was performed at MS2-level with the correction factors from Lot VB294905

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

on lysine labelling only with the N-Terminal label left unused. Digestion was trypsin/p with a maximum of 3 missed cleavages. Carbamidomethylation of cysteines was selected as a fixed modification. Oxidation (M), Acetylation (Protein N-terminus), and N-terminal TMTpro labelling were selected as variable modifications. PSM and Protein FDR were set at 0.01.

For analysis of fractionated, N-terminally-enriched material. Default MaxQuant settings were used with the following alterations. Quantification was performed at MS2-level with the correction factors from Lot VB294905. Digestion was semi-specific ArgC, as TMTpro labelling of lysines blocks trypsin-cleavage. Carbamidomethylation of cysteines was selected as a fixed modification. Oxidation (M), Acetylation (Protein N-terminus), Gln/Glu to pyroglutamine were selected as variable modifications. PSM and Protein FDR were set at 0.01.

For analysis of viral protein neo-N-termini from fractionated, N-terminally-enriched material. Default MaxQuant settings were used with the following alterations. MS1-based quantitation was selected. Digestion was ArgC, sei-specific N-terminus. Carbamidomethylation of cysteines was selected as a fixed modification. Oxidation (M), Acetylation (Protein N-terminus), Gln/Glu to pyroglutamine, and TMTpro modification of N-termini and lysine residues were selected as variable modifications. PSM and Protein FDR were set at 0.01.

All downstream analysis was conducted in Matlab. Reverse hits and contaminants were removed, peptides were filtered to meet PEP ≤ 0.02. For quantitative analysis, peptides were further filtered at PIF ≥ 0.7. TMTpro data was normalised for differences in protein loading by dividing by the label median, rows were filtered to remove rows with more than 2/3 missing data. Missing data was KNN imputed, and individual peptides were normalised by dividing by their mean abundance accross all TMTpro channels. As the objective was to identify protein cleavage events, peptides were further filtered to remove those beginning at the first or second amino acid in a protein sequence that represent the native N-terminus. +/- methionine. neo-N-termini were annotated if they matched known signal peptides. For non-quantitative analysis (e.g. mapping of viral neo-N-termini), peptides were filtered to retain only blocked (acetylated, TMTpro labelled, and pyroglutamine) N-termini. Pyroglutamine-blocked N-termini were discarded if they were preceeded by arginine or lysine as these could represent artifactual cyclization of tryptic N-termini. Fractional protein or peptide intensity was calculated as the total intensity for the protein or peptide, multiplied by the fraction of the summed normalised TMTpro intensity represented by a particular TMTpro label of interest.

Visualisation of the Y636/S637 cleavage site within the SARS-CoV-2 Spike glycoprotein structure was performed using PDB: 6X6P (25), in UCSF ChimeraX v1.0 (48).

Virus infections in siRNA-based cellular protein knockdowns . Host proteins were knocked-down in A549-Ace2 cells using specific dsiRNAs from IDT. Briefly, A549-Ace2 cells seeded at $1x10^4$ cells/well in 96-well plates. After 24 hours, each well was transfected with 5 pmol of individual dsiRNAs using Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions. 24 hours post transfection, the cell culture supernatant was removed and replaced with virus inoculum (MOI of 0.1 PFU/cell). Following a 1 hour adsorption at 37°C, the virus inoculum was removed and replaced with fresh 2% FBS/DMEM media. Cells were incubated at 37°C for 3 days before supernatants were harvested. Samples were either heat-inactivated at 80°C for 20 min and viral RNA was quantified by RT-qPCR, using previously published SARS-CoV-2 specific primers targeting the N gene (49). RT-qPCR was performed using the Luna Universal One-Step RT-qPCR Kit (NEB) in an Applied Biosystems QuantStudio 7 thermocycler, using the following cycling conditions: 55 °C for 10 min, 95 °C for 1 min, and 40 cycles of 95 °C for 10 sec, followed by 60 °C for 1 min. The quantity of viral genomes is expressed as PFU equivalents, and was calculated by performing a standard curve with RNA derived from a viral stock with a known viral titer. Alternatively, infectious virus titers were quantified using plaque assays as described above. To quantify siRNA-based cellular protein knockdowns, A549-Ace2 cells were seeded and transfected with individual dsiRNAs as described above. After 24 hours incubation at 37 °C cells were lysed and RNA was extracted using Trizol (ThermoFisher Scientific) followed by purification using the Direct-zol-96 RNA extraction kit (Zymo) following the manufacturer's instructions. RNA levels of target proteins were subsequently quantified by using RT-with the Luna Universal One-Step RT-qPCR Kit (NEB) in an Applied Biosystems QuantStudio 7 thermocycler using gene-specific primers. Expression levels were compared to scrambled dsiRNA-transfected cells und normalized to expression of human beta-actin. Knockdown efficiencies were calculated using $\Delta\Delta$Ct in Matlab.

To assess cell viability after siRNA knockdowns, cells were seeded and transfected as described above. 24 hours after transfection cell viability was measured using alamarBlue reagent (ThermoFisher Scientific), media was removed and replaced with alamarBlue and incubated for 1h at 37 °C and fluorescence measured in a Tecan Infinite M200 Pro plate reader. Percentage viability was calculated relative to untreated cells (100% viability) and cells lysed with 20% ethanol (0% viability), included in each plate.

Drug Screens and Cytotoxicity analysis. Black with clear bottom 384 well plates were seeded with $2x10^3$ A549-Ace2 cells per well. The following day, individual compounds were added using the Echo 550 acoustic dispenser at concentrations indicated 2 hours prior to infection. DMSO-only (0.5%) and remdesivir (10μM; SelleckChem) controls were added in each plate. After the pre-incubation period, the drug-containing media was removed, and replaced with virus inoculum (MOI of 0.1 PFU/cell). Following a one-

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

hour adsorption at 37°C, the virus inoculum was removed and replaced with 2% FBS/DMEM media containing the individual drugs at the indicated concentrations. Cells were incubated at 37°C for 3 days. Supernatants were harvested and heat-inactivated at 80°C for 20 min. Detection of viral genomes from heat-inactivated was performed by RT-qPCR as described above. Cytotoxicity was determined using the CellTiter-Glo luminescent cell viability assay (Promega). White with clear bottom 384 well plates were seeded with $2x10^3$ A549-Ace2 cells per well. The following day, individual compounds were added using the Echo 550 acoustic dispenser at concentrations indicated. DMSO-only (0.5%) and camptothecin (10 μM; Sigma Aldrich) controls were added in each plate. After 72 h incubation, 20μl/well of Celltiter-Glo reagent was added, incubated for 20 min and the luminescence was recorded using a luminometer (Berthold Technologies) with 0.5 sec integration time. Curve fits and IC_{50}/CC_{50} values were obtained in Matlab.

Data availability. All mass spectrometry data, database FASTA files, and the matlab scripts used to generate the data in this manuscript can be found on the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE repository (50), and on GitHub respectively. Specifically the proteomics datasets have been deposited as described in table S6, where reviewer usernames and passwords are provided.

The Matlab scripts used to process the mass spectrometry data and produce the figures in this manuscript have been tested in Matlab versions R2019b with the Statistics Machine Learning Toolbox, on Mac OS Catalina. These can be accessed through the Emmott Lab Github page at: https://github.com/emmottlab/sars2nterm/.

Other reagent and oligo sequence details are described in table S7.

ACKNOWLEDGEMENTS

We thank members of the Centre for Proteome Research, especially Rob Beynon and Jos Sarsby, as well as Nikolai Slavov & Aleksandra Petelski (Northeastern University) for constructive comments. We also thank Agnès Zettor and Soizick Lucas-Staat from the Chemogenomic and Biological Screening Platform for their technical assistance. The A549-Ace2 cells were a gift from Olivier Schwartz (Institut Pasteur). This work was supported by the Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant ANR-10-LABX-62-IBEID) to MV. EE is supported by startup funding from the University of Liverpool, as well as a Wellcome Trust ISSF Interdisciplinary Industry Award. EE is grateful for the support of GoFundMe donors for sponsoring SARS-CoV-2 research in his laboratory.

References

1. Chen Wang, Peter W Horby, Frederick G Hayden, and George F Gao. A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223):470–473, 2020. doi: 10.1016/s0140-6736(20)30185-9.
2. Na Zhu, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, Baoying Huang, Weifeng Shi, Roujian Lu, Peihua Niu, Faxian Zhan, Xuejun Ma, Dayan Wang, Wenbo Xu, Guizhen Wu, George F. Gao, and Wenjie Tan. A novel coronavirus from patients with pneumonia in china, 2019. *New England Journal of Medicine*, 382(8):727–733, 2020. doi: 10.1056/NEJMoa2001017. PMID: 31978945.
3. Andrew D. Davidson, Maia Kavanagh Williamson, Sebastian Lewis, Deborah Shoemark, Miles W. Carroll, Kate J. Heesom, Maria Zambon, Joanna Ellis, Philip A. Lewis, Julian A. Hiscox, and David A. Matthews. Characterisation of the transcriptome and proteome of sars-cov-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Medicine*, 12(1):68, Jul 2020. ISSN 1756-994X. doi: 10.1186/s13073-020-00763-0.
4. Kevin Klann, Denisa Bojkova, Georg Tascher, Sandra Ciesek, Christian Münch, and Jindrich Cinatl. Growth factor receptor signaling inhibition prevents sars-cov-2 replication. *Molecular Cell*, Aug 2020. ISSN 1097-2765. doi: 10.1016/j.molcel.2020.08.006. PMC7418786[pmcid].
5. Denisa Bojkova, Kevin Klann, Benjamin Koch, Marek Widera, David Krause, Sandra Ciesek, Jindrich Cinatl, and Christian Münch. Proteomics of sars-cov-2-infected host cells reveals therapy targets. *Nature*, 583(7816):469–472, 2020. doi: 10.1038/s41586-020-2332-7.
6. David E. Gordon, Gwendolyn M. Jang, Mehdi Bouhaddou, Jiewei Xu, Kirsten Obernier, Kris M. White, Matthew J. O'Meara, Veronica V. Rezelj, Jeffrey Z. Guo, Danielle L. Swaney, Tia A. Tummino, Ruth Hüttenhain, Robyn M. Kaake, Alicia L. Richards, Beril Tutuncuoglu, Helene Foussard, Jyoti Batra, Kelsey Haas, Maya Modak, Minkyu Kim, Paige Haas, Benjamin J. Polacco, Hannes Braberg, Jacqueline M. Fabius, Manon Eckhardt, Margaret Soucheray, Melanie J. Bennett, Merve Cakir, Michael J. McGregor, Qiongyu Li, Bjoern Meyer, Ferdinand Roesch, Thomas Vallet, Alice Mac Kain, Lisa Miorin, Elena Moreno, Zun Zar Chi Naing, Yuan Zhou, Shiming Peng, Ying Shi, Ziyang Zhang, Wenqi Shen, Ilsa T. Kirby, James E. Melnyk, John S. Chorba, Kevin Lou, Shizhong A. Dai, Inigo Barrio-Hernandez, Danish Memon, Claudia Hernandez-Armenta, Jiankun Lyu, Christopher J. P. Mathy, Tina Perica, Kala Bharath Pilla, Sai J. Ganesan, Daniel J. Saltzberg, Ramachandran Rakesh, Xi Liu, Sara B. Rosenthal, Lorenzo Calviello, Srivats Venkataramanan, Jose Liboy-Lugo, Yizhu Lin, Xi-Ping Huang, YongFeng Liu, Stephanie A. Wankowicz, Markus Bohn, Maliheh Safari, Fatima S. Ugur, Cassandra Koh, Nastaran Sadat Savar, Quang Dinh Tran, Djoshkun Shengjuler, Sabrina J. Fletcher, Michael C. O'Neal, Yiming Cai, Jason C. J. Chang, David J. Broadhurst, Saker Klippsten, Phillip P. Sharp, Nicole A. Wenzell, Duygu Kuzuoglu-Ozturk, Hao-Yuan Wang, Raphael Trenker, Janet M. Young, Devin A. Cavero, Joseph Hiatt, Theodore L. Roth, Ujjwal Rathore, Advait Subramanian, Julia Noack, Mathieu Hubert, Robert M. Stroud, Alan D. Frankel, Oren S. Rosenberg, Kliment A. Verba, David A. Agard, Melanie Ott, Michael Emerman, Natalia Jura, Mark von Zastrow, Eric Verdin, Alan Ashworth, Olivier Schwartz, Christophe d'Enfert, Shaeri Mukherjee, Matt Jacobson, Harmit S. Malik, Danica G. Fujimori, Trey Ideker, Charles S. Craik, Stephen N. Floor, James S. Fraser, John D. Gross, Andrej Sali, Bryan L. Roth, Davide Ruggero, Jack Taunton, Tanja Kortemme, Pedro Beltrao, Marco Vignuzzi, Adolfo García-Sastre, Kevan M. Shokat, Brian K. Shoichet, and Nevan J. Krogan. A sars-cov-2 protein interaction map reveals targets for drug repurposing. *Nature*, 583(7816):459–468, Jul 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2286-9.
7. Estelle M.N. Laurent, Yorgos Sofianatos, Anastassia Komarova, Jean-Pascal Gimeno, Payman Samavarchi Tehrani, Dae-Kyum Kim, Hala Abdouni, Marie Duhamel, Patricia Cassonnet, Jennifer J. Knapp, Da Kuang, Aditya Chawla, Dayag Sheykhkarimli, Ashyad Rayhan, Roujia Li, Oxana Pogoutse, David E. Hill, Michael A. Calderwood, Pascal Falter-Braun, Patrick Aloy, Ulrich Stelzl, Marc Vidal, Anne-Claude Gingras, Georgios A. Pavlopoulos, Sylvie Van Der Werf, Isabelle Fournier, Frederick P. Roth, Michel Salzet, Caroline Demeret, Yves Jacob, and Etienne Coyaud. Global bioid-based sars-cov-2 proteins proximal interactome unveils novel ties between viral polypeptides and host factors involved in multiple covid19-associated mechanisms. *bioRxiv*, 2020. doi: 10.1101/2020.08.28.272955.
8. Mehdi Bouhaddou, Danish Memon, Bjoern Meyer, Kris M. White, Veronica V. Rezelj, Miguel Correa Marrero, Benjamin J. Polacco, James E. Melnyk, Svenja Ulferts, Robyn M. Kaake, Jyoti Batra, Alicia L. Richards, Erica Stevenson, David E. Gordon, Ajda Rojc, Kirsten Obernier, Jacqueline M. Fabius, Margaret Soucheray, Lisa Miorin, Elena Moreno, Cassandra Koh, Quang Dinh Tran, Alexandra Hardy, Rémy Robinot, Thomas Vallet, Benjamin E. Nilsson-Payant, Claudia Hernandez-Armenta, Alistair Dunham, Sebastian Weigang, Julian Knerr, Maya Modak, Diego Quintero, Yuan Zhou, Aurelien Dugourd, Alberto Valdeolivas, Trupti Patil, Qiongyu Li, Ruth Hüttenhain, Merve Cakir, Monita Muralidharan, Minkyu Kim, Gwendolyn Jang, Beril Tutuncuoglu, Joseph Hiatt, Jeffrey Z. Guo, Jiewei Xu, Sophia Bouhaddou, Christopher J.P. Mathy, Anna Gaulton, Emma J. Manners, Eloy Félix, Ying Shi, Marisa Goff, Jean K. Lim, Timothy McBride, Michael C. O'Neal, Yiming Cai, Jason C.J. Chang, David J. Broadhurst, Saker Klippsten, Emmie De wit, Andrew R. Leach, Tanja Kortemme, Brian Shoichet, Melanie Ott, Julio Saez-Rodriguez, Benjamin R. tenOever, R. Dyche Mullins, Elizabeth R. Fischer, Georg Kochs, Robert Grosse, Adolfo García-Sastre, Marco Vignuzzi, Jeffery R. Johnson, Kevan M. Shokat, Danielle L. Swaney, Pedro Beltrao, and Nevan J. Krogan. The global phosphorylation landscape of sars-cov-2 infection. *Cell*, 182(3):685–712.e19, Aug 2020. ISSN 0092-8674. doi: 10.1016/j.cell.2020.06.034.
9. Alexey Stukalov, Virginie Girault, Vincent Grass, Valter Bergant, Ozge Karayel, Christian Urban, Darya A. Haas, Yiqi Huang, Lila Oubraham, Anqi Wang, Sabri M. Hamad, Antonio Piras, Maria Tanzer, Fynn M. Hansen, Thomas Enghleitner, Maria Reinecke, Teresa M. Lavacca, Rosina Ehmann, Roman Wölfel, Jörg Jores, Bernhard Kuster, Ulrike Protzer, Roland Rad, John Ziebuhr, Volker Thiel, Pietro Scaturro, Matthias Mann, and Andreas Pichlmair. Multi-level proteomics reveals host-perturbation strategies of sars-cov-2 and sars-cov. *bioRxiv*, 2020. doi: 10.1101/2020.06.17.156455.
10. Xiuyuan Ou, Yan Liu, Xiaobo Lei, Pei Li, Dan Mi, Lili Ren, Li Guo, Ruixuan Guo, Ting Chen, Jiaxin Hu, Zichun Xiang, Zhixia Mu, Xing Chen, Jieyong Chen, Keping Hu, Qi Jin, Jianwei Wang, and Zhaohui Qian. Characterization of spike glycoprotein of sars-cov-2 on virus entry and its immune cross-reactivity with sars-cov. *Nature Communications*, 11(1):1620, Mar 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15562-9.
11. Markus Hoffmann, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S. Schiergens, Georg Herrler, Nai-Huei Wu, Andreas Nitsche, Marcel A. Müller, Christian Drosten, and Stefan Pöhlmann. Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2): 271 – 280.e8, 2020. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2020.02.052.
12. Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, Xiaofeng Li, Leike Zhang, Chao Peng, Yinkai Duan, Jing Yu, Lin Wang, Kailin Yang, Fengjiang Liu, Rendi Jiang, Xinglou Yang, Tian You, Xiaoce Liu, Xiuna Yang, Fang Bai, Hong Liu, Xiang Liu, Luke W. Guddat, Wenqing Xu, Gengfu Xiao, Chengfeng Qin, Zhengli Shi, Hualiang Jiang, Zihe Rao, and Haitao Yang. Structure of mpro from sars-cov-2 and discovery of its inhibitors. *Nature*, 582(7811):289–293, Jun 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2223-y.
13. Wioletta Rut, Zongyang Lv, Mikolaj Zmudzinski, Stephanie Patchett, Digant Nayak, Scott J. Snipas, Farid El Oualid, Tony T. Huang, Miklos Bekes, Marcin Drag, and Shaun K. Olsen. Activity profiling and structures of inhibitor-bound sars-cov-2-plpro protease provides a framework for anti-covid-19 drug design. *bioRxiv*, 2020. doi: 10.1101/2020.04.29.068890.
14. Mehdi Moustaqil, Emma Ollivier, Hsin-Ping Chiu, Sarah Van Tol, Paulina Rudolffi-Soto, Christian Stevens, Akshay Bhumkar, Dominic J.B. Hunter, Alex Freiberg, David Jacques,

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Benhur Lee, Emma Sierecki, and Yann Gambin. Sars-cov-2 proteases cleave irf3 and critical modulators of inflammatory pathways (nlrp12 and tab1): implications for disease presentation across species and the search for reservoir hosts. *bioRxiv*, 2020. doi: 10.1101/2020.06.05.135699.

15. Guido Papa, Donna L. Mallery, Anna Albecka, Lawrence Welch, Jérôme Cattin-Ortolá, Jakub Luptak, David Paul, Harvey T. McMahon, Ian G. Goodfellow, Andrew Carter, Sean Munro, and Leo C. James. Furin cleavage of sars-cov-2 spike promotes but is not essential for infection and cell-cell fusion. *bioRxiv*, 2020. doi: 10.1101/2020.08.13.243303.
16. Jian Shang, Yushun Wan, Chuming Luo, Gang Ye, Qibin Geng, Ashley Auerbach, and Fang Li. Cell entry mechanisms of sars-cov-2. *Proceedings of the National Academy of Sciences*, 117(21):11727–11734, 2020. ISSN 0027-8424. doi: 10.1073/pnas.2003138117.
17. Christopher A. Nelson, Andrew Pekosz, Chung A. Lee, Michael S. Diamond, and Daved H. Fremont. Structure and intracellular targeting of the sars-coronavirus orf7a accessory protein. *Structure*, 13(1):75 – 85, 2005. ISSN 0969-2126. doi: https://doi.org/10.1016/j.str.2004.10.010.
18. Claudia Diemer, Martha Schneider, Judith Seebach, Janine Quaas, Gert Frösner, Hermann M. Schätzl, and Sabine Gilch. Cell type-specific cleavage of nucleocapsid protein by effector caspases during sars coronavirus infection. *Journal of Molecular Biology*, 376 (1):23 – 34, 2008. ISSN 0022-2836. doi: https://doi.org/10.1016/j.jmb.2007.11.081.
19. John Mark, Xuguang Li, Terry Cyr, Sylvie Fournier, Bozena Jaentschke, and Mary Alice Hefford. Sars coronavirus: Unusual lability of the nucleocapsid protein. *Biochemical and Biophysical Research Communications*, 377(2):429 – 433, 2008. ISSN 0006-291X. doi: https://doi.org/10.1016/j.bbrc.2008.09.153.
20. Weston Struwe, Edward Emmott, Melanie Bailey, Michal Sharon, Andrea Sinz, Fernando J Corrales, Kostas Thalassinos, Julian Braybrook, Clare Mills, Perdita Barran, and et al. The covid-19 ms coalition—accelerating diagnostics, prognostics, and treatment. *The Lancet*, 395(10239):1761–1762, 2020. doi: 10.1016/s0140-6736(20)31211-3.
21. Samuel S. H. Weng, Fatih Demir, Enes K. Ergin, Sabrina Dirnberger, Anuli Uzozie, Domenic Tuscher, Lorenz Nierves, Janice Tsui, Pitter F. Huesgen, Philipp F. Lange, and et al. Sensitive determination of proteolytic proteoforms in limited microscale proteome samples. *Molecular Cellular Proteomics*, 18(11):2335–2347, 2019. doi: 10.1074/mcp.tir119.001560.
22. Sudeep Pushpakom, Francesco Iorio, Patrick A. Eyers, K. Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, Tim Guilliams, Joanna Latimer, Christine McNamee, Alan Norris, Philippe Sanseau, David Cavalla, and Munir Pirmohamed. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery*, 18(1):41–58, Jan 2019. ISSN 1474-1784. doi: 10.1038/nrd.2018.168.
23. John D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002. doi: 10.1111/1467-9868.00346.
24. David M. Kern, Ben Sorum, Christopher M. Hoel, Savitha Sridharan, Jonathan P. Remis, Daniel B. Toso, and Stephen G. Brohawn. Cryo-em structure of the sars-cov-2 3a ion channel in lipid nanodiscs. *bioRxiv*, 2020. doi: 10.1101/2020.06.17.156554.
25. Natalia G. Herrera, Nicholas C. Morano, Alev Celikgil, George I. Georgiev, Ryan J. Malonis, James H. Lee, Karen Tong, Olivia Vergnolle, Aldo B. Massimi, Laura Y. Yen, Alex J. Noble, Mykhailo Kopylov, Jeffrey B. Bonanno, Sarah C. Garrett-Thomson, David B. Hayes, Robert H. Bortz, Ariel S. Wirchnianski, Catalina Florez, Ethan Laudermilch, Denise Haslwanter, J. Maximilian Fels, M. Eugenia Dieterle, Rohit K. Jangra, Jason Barnhill, Amanda Mengotto, Duncan Kimmel, Johanna P. Daily, Liise-anne Pirofski, Kartik Chandran, Michael Brenowitz, Scott J. Garforth, Edward T. Eng, Jonathan R. Lai, and Steven C. Almo. Characterization of the sars-cov-2 s protein: Biophysical, biochemical, structural, and antigenic analysis. *bioRxiv*, 2020. doi: 10.1101/2020.06.14.150607.
26. Dirk Chelius, Kay Jing, Alexis Lueras, Douglas S. Rehder, Thomas M. Dillon, Alona Vizel, Rahul S. Rajan, Tiansheng Li, Michael J. Treuheit, and Pavel V. Bondarenko. Formation of pyroglutamic acid from n-terminal glutamic acid in immunoglobulin gamma antibodies. *Analytical Chemistry*, 78(7):2370–2376, Apr 2006. ISSN 0003-2700. doi: 10.1021/ac051827k.
27. Burtram C Fielding, Yee-Joo Tan, Shen Shuo, Timothy H P Tan, Eng-Eong Ooi, Seng Gee Lim, Wanjin Hong, and Phuay-Yee Goh. Characterization of a unique group-specific protein (u122) of the severe acute respiratory syndrome coronavirus. *Journal of virology*, 78(14): 7311—7318, July 2004. ISSN 0022-538X. doi: 10.1128/jvi.78.14.7311-7318.2004.
28. Yaara Finkel, Orel Mizrahi, Aharon Nachshon, Shira Weingarten-Gabbay, David Morgenstern, Yfat Yahalom-Ronen, Hadas Tamir, Hagit Achdout, Dana Stein, Ofir Israeli, Adi Beth-Din, Sharon Melamed, Shay Weiss, Tomer Israely, Nir Paran, Michal Schwartz, and Noam Stern-Ginossar. The coding capacity of sars-cov-2. *Nature*, Sep 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2739-1.
29. Wioletta Rut, Katarzyna Groborz, Linlin Zhang, Xinyuanyuan Sun, Mikolaj Zmudzinski, Bartlomiej Pawlik, Wojciech Młynarski, Rolf Hilgenfeld, and Marcin Drag. Substrate specificity profiling of sars-cov-2 main protease enables design of activity-based probes for patient-sample imaging. *bioRxiv*, 2020. doi: 10.1101/2020.03.07.981928.
30. Stefan Knapp. New opportunities for kinase drug repurposing and target discovery. *British Journal of Cancer*, 118(7):936–937, 2018. doi: 10.1038/s41416-018-0045-6.
31. Keely L. Szilágyi, Cong Liu, Xu Zhang, Ting Wang, Jeffrey D. Fortman, Wei Zhang, and Joe G. N. Garcia. Epigenetic contribution of the myosin light chain kinase gene to the risk for acute respiratory distress syndrome. *Translational research : the journal of laboratory and clinical medicine*, 180:12–21, Feb 2017. ISSN 1878-1810. doi: 10.1016/j.trsl.2016.07.020. 27543902[pmid].
32. Valeria Nofrini, Danika Di Giacomo, and Cristina Mecucci. Nucleoporin genes in human diseases. *European journal of human genetics : EJHG*, 24(10):1388–1395, Oct 2016. ISSN 1476-5438. doi: 10.1038/ejhg.2016.25. 27071718[pmid].
33. Masami Wada, Kumari G. Lokugamage, Keisuke Nakagawa, Krishna Narayanan, and Shinji Makino. Interplay between coronavirus, a cytoplasmic rna virus, and nonsense-mediated mrna decay pathway. *Proceedings of the National Academy of Sciences*, 115(43):E10157–E10166, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1811675115.
34. Alex Generous, Molly Thorson, Jeff Barcus, Joseph Jacher, Marc Busch, and Heidi Sleister. Identification of putative interactions between swine and human influenza a virus nucleoprotein and human host proteins. *Virology journal*, 11:228–228, Dec 2014. ISSN 1743-422X. doi: 10.1186/s12985-014-0228-6. 25547032[pmid].
35. Nir Drayman, Krysten A. Jones, Saara-Anne Azizi, Heather M. Froggatt, Kemin Tan, Natalia Ivanovna Maltseva, Siquan Chen, Vlad Nicolaescu, Steve Dvorkin, Kevin Furlong, Rahul S. Kathayat, Mason R. Firpo, Vincent Mastrodomenico, Emily A. Bruce, Madaline M. Schmidt, Robert Jedrzejczak, Miguel Á. Muñoz-Alía, Brooke Schuster, Vishnu Nair, Jason W. Botten, Christopher B. Brooke, Susan C. Baker, Bryan C. Mounce, Nicholas S. Heaton, Bryan C. Dickinson, Andrzej Jaochimiak, Glenn Randall, and Savaş Tay. Drug repurposing screen identifies masitinib as a 3clpro inhibitor that blocks replication of sars-cov-2 in vitro. *bioRxiv*, 2020. doi: 10.1101/2020.08.31.274639.
36. Edward Emmott, Frederic Sorgeloos, Sarah L. Caddy, Surender Vashist, Stanislav Sosnovtsev, Richard Lloyd, Kate Heesom, Nicolas Locker, and Ian Goodfellow. Norovirus-mediated modification of the translational landscape via virus and host-induced cleavage of translation initiation factors. *Molecular & Cellular Proteomics*, 16(4 suppl 1):S215–S229, 2017. ISSN 1535-9476. doi: 10.1074/mcp.M116.062448.
37. Edward Emmott, Trevor R. Sweeney, and Ian Goodfellow. A cell-based fluorescence resonance energy transfer (fret) sensor reveals inter- and intragenogroup variations in norovirus protease activity and polyprotein cleavage. *Journal of Biological Chemistry*, 290(46): 27841–27853, 2015. doi: 10.1074/jbc.m115.688234.
38. Kris Gevaert, Marc Goethals, Lennart Martens, Jozef Van Damme, An Staes, Grégoire R. Thomas, and Joël Vandekerckhove. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted n-terminal peptides. *Nature Biotechnology*, 21(5):566–569, May 2003. ISSN 1546-1696. doi: 10.1038/nbt810.
39. Lucy McDonald and Robert J. Beynon. Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nature Protocols*, 1 (4):1790–1798, Nov 2006. ISSN 1750-2799. doi: 10.1038/nprot.2006.317.
40. Oded Kleifeld, Alain Doucet, Ulrich auf dem Keller, Anna Prudova, Oliver Schilling, Rajesh K. Kainthan, Amanda E. Starr, Leonard J. Foster, Jayachandran N. Kizhakkedathu, and Christopher M. Overall. Isotopic labeling of terminal amines in complex samples identifies protein n-termini and protease cleavage products. *Nature Biotechnology*, 28(3):281–288, Mar 2010. ISSN 1546-1696. doi: 10.1038/nbt.1611.
41. Julienne M. Jagdeo, Antoine Dufour, Theo Klein, Nestor Solis, Oded Kleifeld, Jayachandran Kizhakkedathu, Honglin Luo, Christopher M. Overall, and Eric Jan. N-terminomics tails identifies host cell substrates of poliovirus and coxsackievirus b3 3c proteinases that modulate virus infection. *Journal of Virology*, 92(8), 2018. ISSN 0022-538X. doi: 10.1128/JVI.02211-17.
42. Danielle L. Swaney, Craig D. Wenger, and Joshua J. Coon. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *Journal of Proteome Research*, 9(3): 1323–1329, 2010. doi: 10.1021/pr900863u.
43. Piero Giansanti, Liana Tsiatsiani, Teck Yew Low, and Albert J. R. Heck. Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nature Protocols*, 11 (5):993–1006, May 2016. ISSN 1750-2799. doi: 10.1038/nprot.2016.057.
44. Julian Buchrieser, Jeremy Dufloo, Mathieu Hubert, Blandine Monel, Delphine Planas, Maaran Michael Rajah, Cyril Planchais, Françoise Porrot, Florence Guivel-Benhassine, Sylvie Van der Werf, Nicoletta Casartelli, Hugo Mouquet, Timothée Bruel, and Olivier Schwartz. Syncytia formation by sars-cov-2 infected cells. *bioRxiv*, 2020. doi: 10.1101/2020.07.14.202028.
45. Jiaming Li, Jonathan G. Van Vranken, Laura Pontano Vaites, Devin K. Schweppe, Edward L. Huttlin, Chris Etienne, Premchendar Nandhikonda, Rosa Viner, Aaron M. Robitaille, Andrew H. Thompson, Karsten Kuhn, Ian Pike, Ryan D. Bomgarden, John C. Rogers, Steven P. Gygi, and Joao A. Paulo. Tmtpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nature Methods*, 17(4): 399–404, Apr 2020. ISSN 1548-7105. doi: 10.1038/s41592-020-0781-4.
46. Stefka Tyanova, Tikira Temu, and Juergen Cox. The maxquant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, 11(12):2301–2319, 2016. doi: 10.1038/nprot.2016.136.
47. Lars Kolbowski, Colin Combe, and Juri Rappsilber. xiSPEC: web-based visualization, analysis and sharing of proteomics data. *Nucleic Acids Research*, 46(W1):W473–W478, 05 2018. ISSN 0305-1048. doi: 10.1093/nar/gky353.
48. Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Eric F. Pettersen, Gregory S. Couch, John H. Morris, and Thomas E. Ferrin. Ucsf chimerax: Meeting modern challenges in visualization and analysis. *Protein Science*, 27(1):14–25, 2018. doi: 10.1002/pro.3235.
49. Daniel K W Chu, Yang Pan, Samuel M S Cheng, Kenrie P Y Hui, Pavithra Krishnan, Yingzhi Liu, Daisy Y M Ng, Carrie K C Wan, Peng Yang, Quanyi Wang, Malik Peiris, and Leo L M Poon. Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. *Clinical Chemistry*, 66(4):549–555, 01 2020. ISSN 0009-9147. doi: 10.1093/clinchem/hvaa029.
50. Yasset Perez-Riverol, Attila Csordas, Jingwen Bai, Manuel Bernal-Llinares, Suresh Hewapathirana, Deepti J Kundu, Avinash Inuganti, Johannes Griss, Gerhard Mayer, Martin Eisenacher, Enrique Pérez, Julian Uszkoreit, Julianus Pfeuffer, Timo Sachsenberg, Şule Yılmaz, Shivani Tiwary, Jürgen Cox, Enrique Audain, Mathias Walzer, Andrew F Jarnuczak, Tobias Ternent, Alvis Brazma, and Juan Antonio Vizcaíno. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research*, 47(D1):D442–D450, 11 2018. ISSN 0305-1048. doi: 10.1093/nar/gky1106.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Material

Table S1. Viral neo-N-termini identified from SARS-CoV-2-infected A549-Ace2 cells - .csv

Table S2. Viral neo-N-termini identified from SARS-CoV-2-infected Vero E6 cells - .csv

Table S3. All Viral peptides identified accross enriched and unenriched A549-Ace2 and Vero E6 datasets - .csv

Table S4. Quantification data for all N- and neo-N-termini quantified from SARS-CoV-2 infected A549-Ace2 cells - .csv

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table S5. Quantification data for all N- and neo-N-termini quantified from SARS-CoV-2 infected Vero E6 cells - .csv

Dataset:	PRIDE accession:	Reviewer Username:	Reviewer Password:
A549-Ace2 Unenriched	PXD021145	reviewer34884@ebi.ac.uk	ZKTFOUZA
A549-Ace2 Enriched	PXD021152	reviewer62064@ebi.ac.uk	xBgebTgO
A549-Ace2 Enriched (Variable Search)	PXD021402	reviewer_pxd021402@ebi.ac.uk	iSbBtAdt
Vero E6 Unenriched	PXD021154	reviewer98284@ebi.ac.uk	x3fU4szp
Vero E6 Enriched	PXD021153	reviewer53878@ebi.ac.uk	BiDiQ0VG
Vero E6 Enriched (Variable Search)	PXD021403	reviewer_pxd021403@ebi.ac.uk	x7ZdiAt6

Table S6. Proteomic datasets and access details

Table S7. Oligo sequences and reagent details - .csv

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

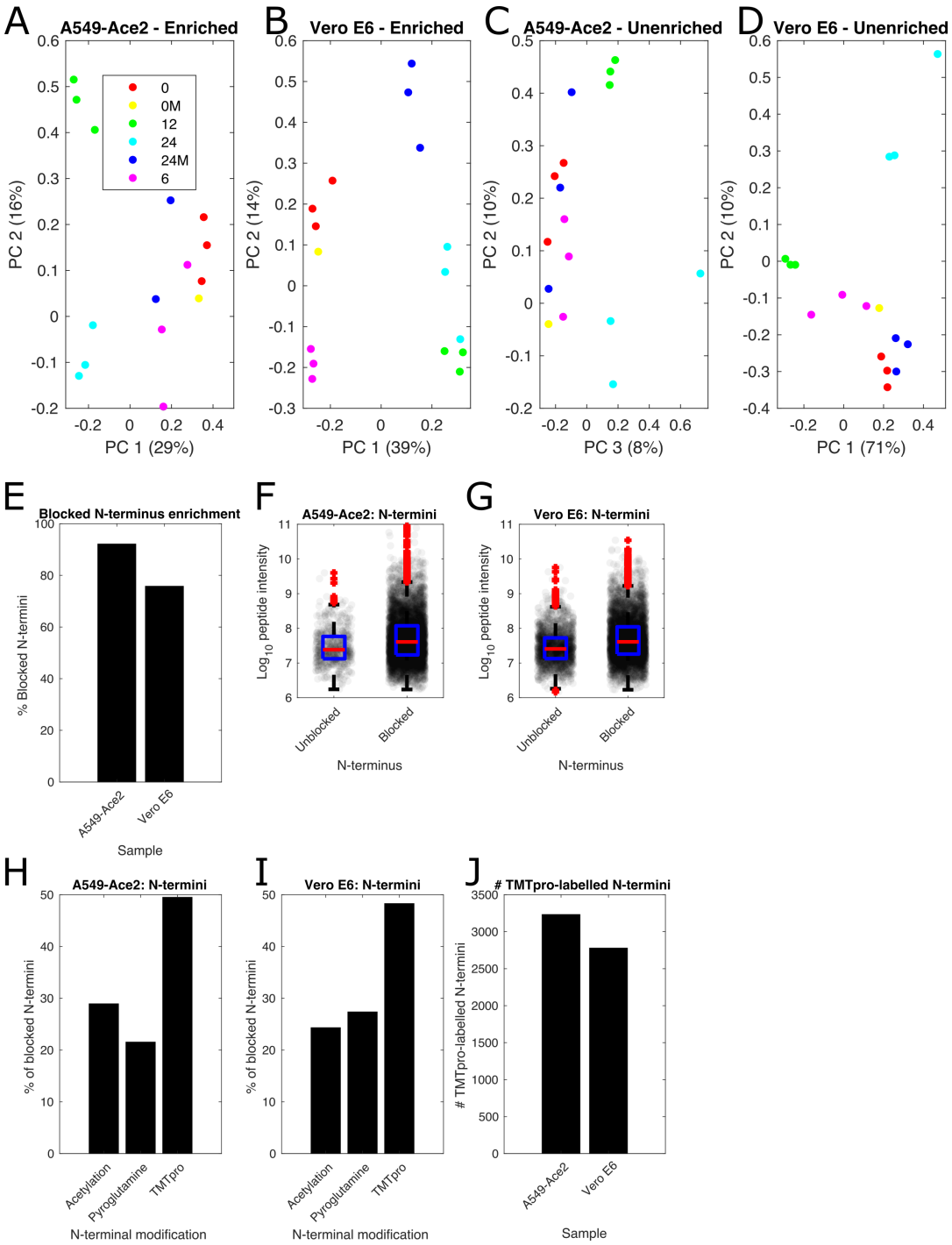

Fig. S1. Quality control of the proteomic datasets, both pre- and post-enrichment for N-termini. A-D) Principal component analysis separates infected from mock cells and shows reproducible clustering of biological replicates. E) Enrichment results in a majority of peptide identifications belonging to blocked N-termini, with blocked N-termini most abundant in both F) A549-Ace2 and G) Vero E6 cells. In both cell lines, TMTpro-labelled N-termini are the most abundant enriched N-termini H), I). J) over 2700 TMTpro-labelled N-termini were identified from each dataset.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

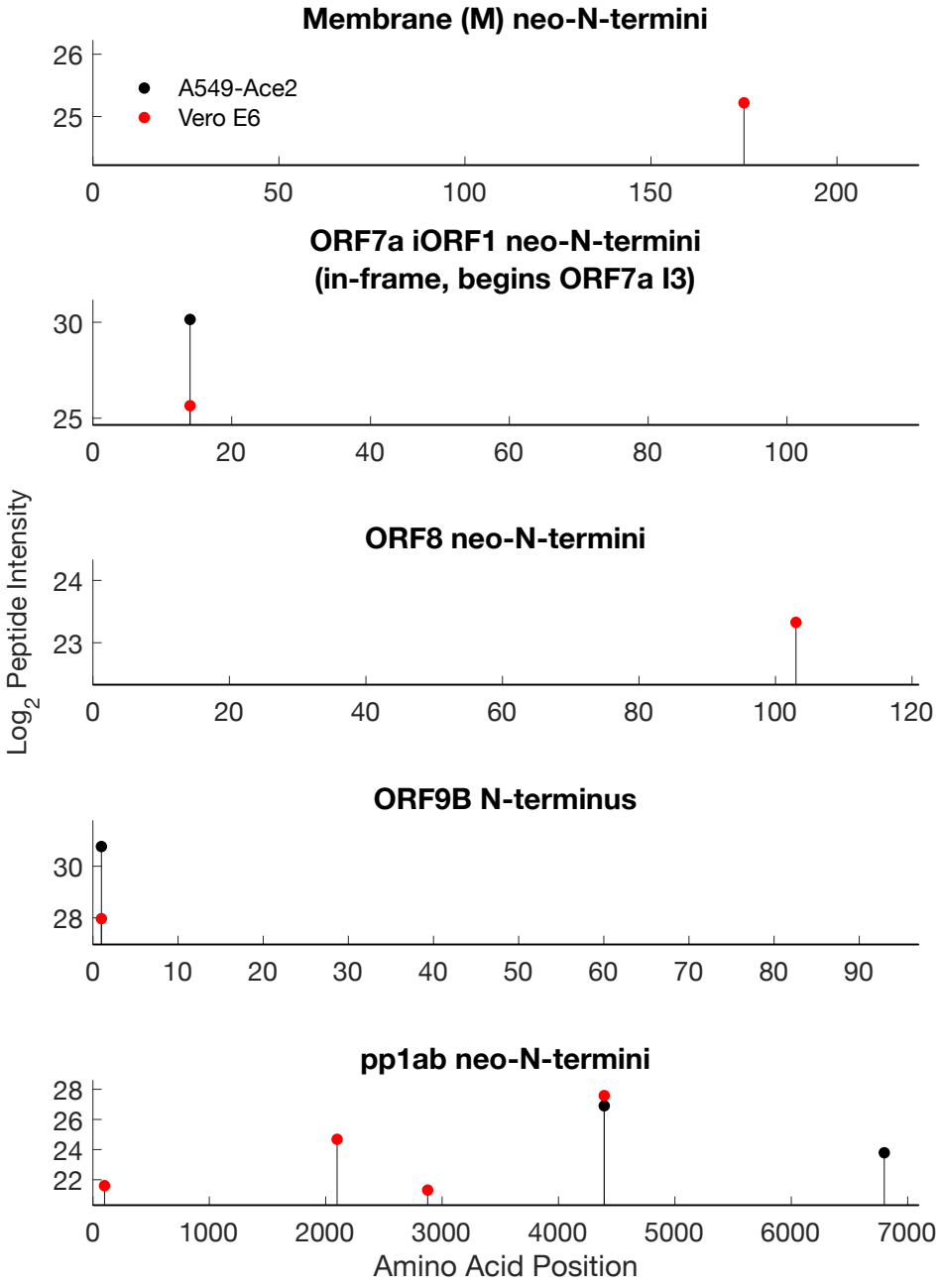

Fig. S2. Viral N-termini and neo-N-termini identified from the viral M, ORF7A/ORF7A iORF1, ORF8, ORF9B and pp1ab replicase. Please note that ORF7A iORF1 is an N-terminally truncated form of ORF7A that initiates at isoleucine 3 in the ORF7A sequence. The indicated neo-N-terminus beginning at amino acid 14, would therefore be amino acid 16 in ORF7A.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

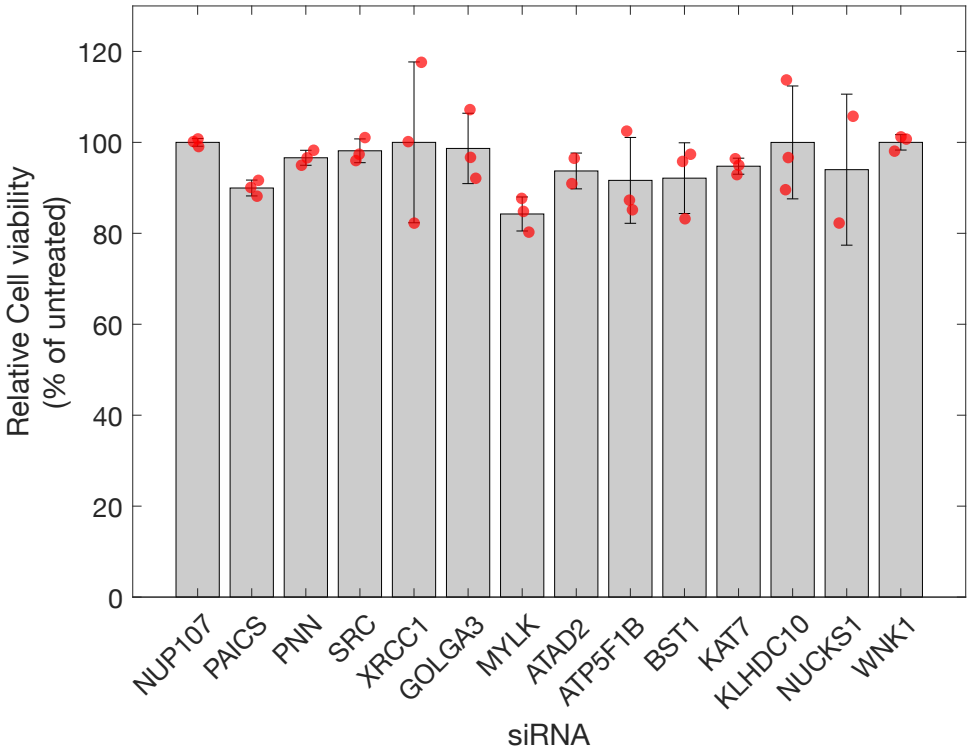

Fig. S3. Cell viability of siRNA-treated A549-Ace2 cells. Cell viability was assessed by alamarBlueAlamar blue staining and compared to untreated control cells, and a 20% ethanol-lysed control. Error bars represent standard deviation from 3 biological replicates. Red markers indicate individual datapoints.

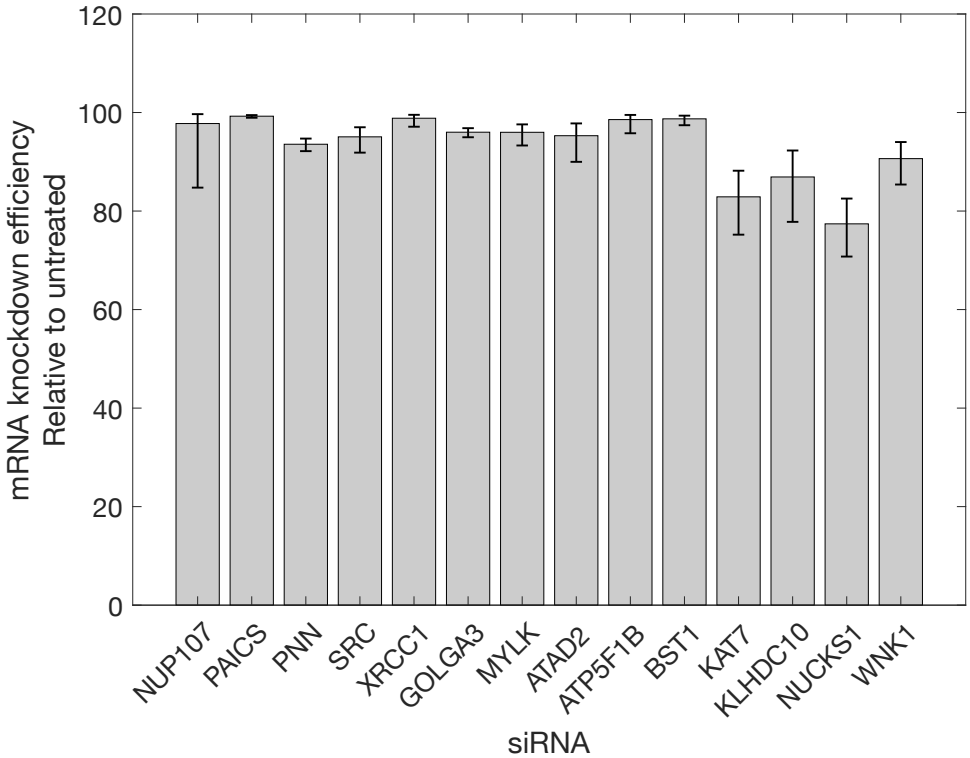

Fig. S4. mRNA knockdown efficiency for SARS-CoV-2 protease substrates in siRNA-treated A549-Ace2 cells. Knockdown efficiency was calculated by qRT-PCR compared to a untreated control by the $2^{-\Delta\Delta Ct}$ method. Error bars represent standard deviation from a minimum of 3 biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

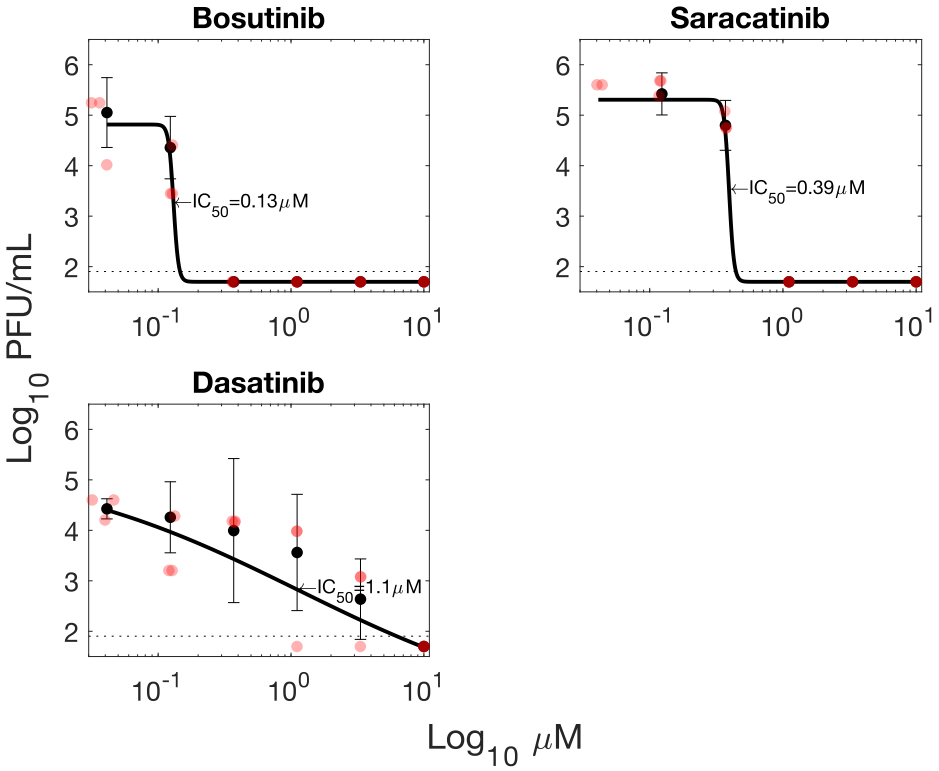

Fig. S5. Additional Inhibitors targeting SRC kinase reduce SARS-CoV-2 titres in A549-Ace2 cells. Error bars represent standard deviation from 3 biological replicates. Red circles indicate individual datapoints.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.297945; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

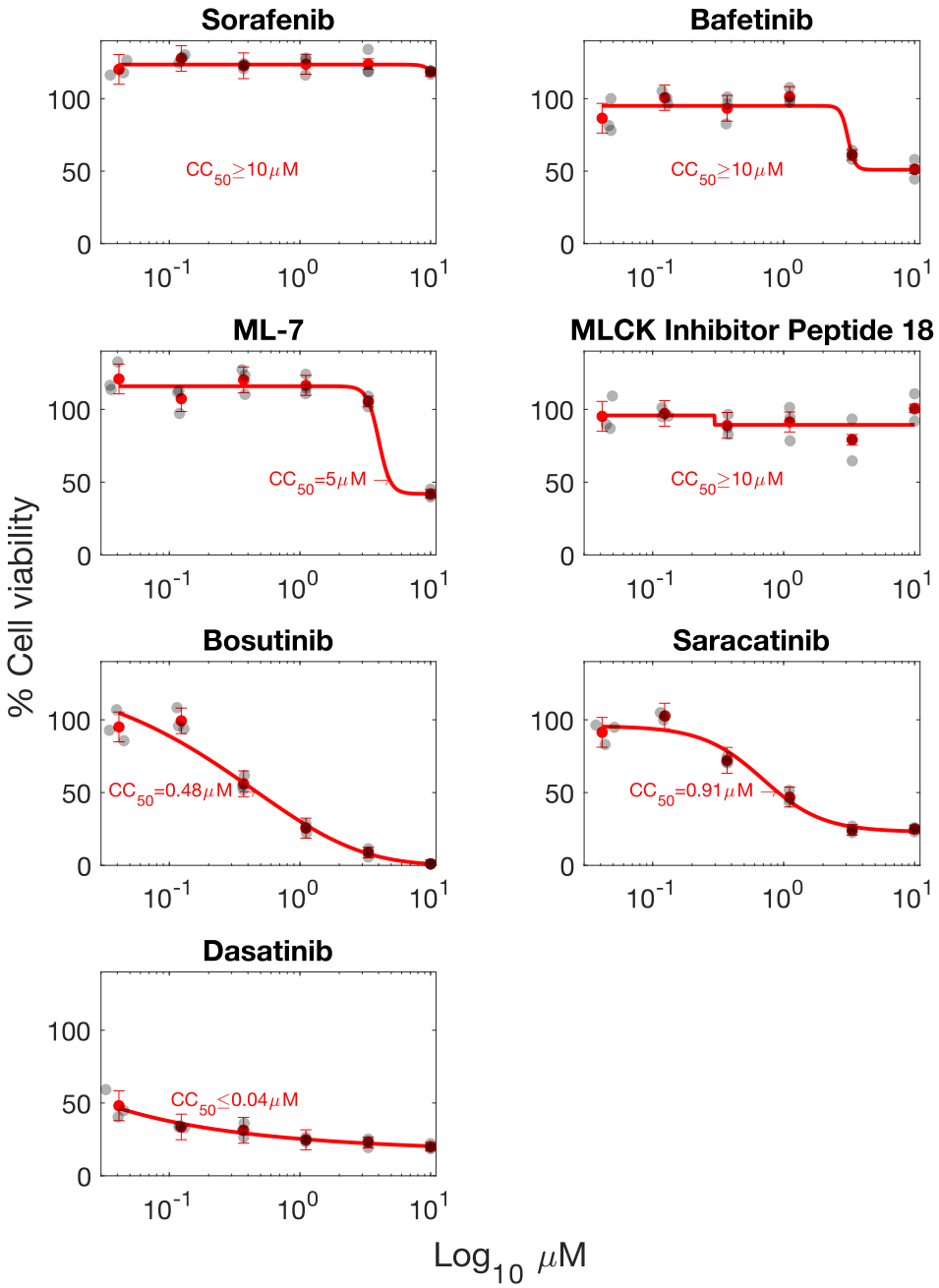

Fig. S6. Cell viability and CC_{50} calculations for inhibitor-treated A549-Ace2 cells. Cell viability was assessed by Celltiter Glo staining and compared to untreated control cells, and a 20% ethanol-lysed control. Line represents best fit. Error bars represent standard deviation from 3 biological replicates. Black markers indicate individual data points.