Vulnerabilities of the SARS-CoV-2 virus to proteotoxicity – opportunity for repurposed chemotherapy of COVID-19 infection

Maryam Al-Motawa,[1,2] Hafsa Abbas,[3] Patrick Wijten,[1] Alberto de la Fuente,[1]

Mingzhan Xue,[1,3] Naila Rabbani[3,4]* and Paul J. Thornalley[1,2,3]*

[1]Diabetes Research Center, Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Qatar Foundation, P.O. Box 34110, Doha, Qatar; [2]College of Health and Life Sciences, Hamad Bin Khalifa University, Qatar Foundation, P.O. Box 34110, Doha, Qatar; [3]Clinical Sciences Research Laboratories, Warwick Medical School, University of Warwick, University Hospital, Coventry CV2 2DX, U.K., and [4]Department of Basic Medical Science, College of Medicine, QU Health, Qatar University, P.O. Box 2713, Doha, Qatar.

Correspondence should be addressed to Paul J. Thornalley (pthornalley@hbku.edu.qa)

*Joint corresponding authors.

Word count: abstract 150; main text 3266.

SUMMARY

There is an urgent requirement for improved treatments of COVID-19 disease. A strategy for chemotherapy is to increase levels of endogenous reactive metabolites - such as reactive oxygen species and arginine-directed glycating agent, methylglyoxal - for selective toxicity to SARS-CoV-2. Sequence analysis of functional domains in the SARS-CoV-2 proteome showed 0.8-fold depletion of cysteine residues and 4.9-fold enrichment of arginine residues, suggesting methylglyoxal modification may inactivate the virus. We discovered the peptide motif for MG modification: 3 – 5-fold enrichment of cationic residues preceding the target arginine. There was 5-fold enrichment of methylglyoxal modification sites in the SARS-CoV-2 proteome, compared to the human host - indicating selective toxicity of methylglyoxal to the virus. We found antitumor drugs, doxorubicin and paclitaxel, increase cellular methylglyoxal to virucidal levels. Taken together, these findings reveal a proteomic vulnerability of SARS-CoV-2 to methylglyoxal modification and provide a rationale for repurposing doxorubicin and paclitaxel for COVID-19 treatment.

INTRODUCTION

A global pandemic of COVID-19 disease caused by infection with the SARS-CoV-2 coronavirus has developed in the first three months of 2020. It has produced a global public health emergency with currently (7th April 2020) over 1.3 million infections and 75,000 deaths, with both rapidly increasing. New treatments are urgently required for the pandemic of COVID-19 disease until an effective vaccine is developed for infection with SARS-CoV-2 coronavirus.

We hypothesised that SARS-CoV-2 may have vulnerabilities in the viral proteome to modification by endogenous reactive metabolites. Pharmacological increase of reactive metabolites will then produce a virucidal effect and therapeutic response for COVID-19 disease. Important reactive metabolites producing major quantitative modification of the proteome in physiological systems are: reactive oxygen species (ROS) and methylglyoxal (MG) (Rabbani et al., 2016b; Sies et al., 2017). Key to characterizing the vulnerability of the viral proteome to reactive metabolites, ROS and MG, is location of their susceptible amino acid residue targets in functional domains of viral proteins and activation of these residues towards reaction with reactive metabolites. ROS are formed by mitochondria through trace leakage of electron flux in oxidative phosphorylation, by oxidases and other sources. They are metabolized by antioxidant enzymes, superoxide dismutase, catalase, glutathione peroxidase and peroxiredoxins (Sies et al., 2017). The reactive dicarbonyl metabolite, MG, is formed mainly by trace level degradation of triosephosphate glycolytic intermediates, glyceraldehyde-3-phosphate and dihydroxyacetonephosphate, and is metabolised by glutathione-dependent glyoxalase 1 (Glo1) of the glyoxalase pathway (Rabbani et al., 2016b) (Figure 1A). The most susceptible targets in proteins to modification by ROS are cysteine residues which are oxidised to cystine and cysteine sulfenic and sulfonic acids (Sies et al., 2017). The most susceptible targets in proteins to modification by MG are arginine residues

which are glycated to hydroimidazolone MG-H1 with loss of charge, all related electrostatic interactions and, typically, resistance to proteolytic cleavage close to the site of modification (Figure 1B) (Rabbani et al., 2016b).

Key to exploring if reactive metabolites of the host may be exploited to produce a virucidal response against SARS-CoV-2 is to identify proteomic vulnerabilities of the virus. Currently it is unknown if target amino acid residues of reactive metabolites are enriched in functional domains of the viral proteome, and if these targets are enriched in the viral proteome, compared to the human host. It also unknown if there are investigational new drugs or current clinically approved drugs that increase reactive metabolites to virucidal levels in the cellular environment where SARS-CoV-2 undergo cell fusion and propagation. To address this knowledge gap, we initiated a series of studies using protein informatics tools, available proteomics data and a cell model used in SARS-CoV-2 fusion and propagation. Herein we report the susceptibility of the SARS-CoV-2 virus to increased MG or "dicarbonyl stress" (Rabbani et al., 2016b) based on enrichment of arginine residues in functional domains and enrichment of amino acid motif MG modification sites in the SARS-CoV-2 proteome, and two clinically approved antitumor drugs that increase the cellular concentration of MG to virucidal levels. We propose these drugs be considered for repurposing for treatment of COVID-19 disease.

RESULTS

Enrichment of arginine residues in the functional domains in the SARS-CoV-2 proteome.

We acquired the reference sequences of the 29 proteins of the SARS-CoV-2 proteome (Table S1) and also, for reference, 18,821 reviewed protein sequences of human host proteins from

the UniProt Knowledgebase (UniProtKB; www.uniprot.org). We found a similar prevalence of cysteine and arginine residues in the viral proteome: 3.14% and 3.63%, respectively (Table 1). We applied the sequence-based receptor binding domain (RBD) analysis to identify functional domains of proteins of the viral proteome to thereby deduce the prevalence of cysteine and arginine residue targets therein. In optimised format, RBD analysis is applicable to all proteins and involves a plot of mean hydrophobicity against mean dipole moment of a widow of 5 amino acid residues moved sequentially along the sequence of a protein (with a gyration angle between two consecutive residues in the sequence of 100° assumed). This approach had 80% accuracy when validated against a database of known interacting proteins (Gallet et al., 2000). Functional domains are located in a trapezium-shaped domain on the top-left side of the plot – regions of low mean hydrophobicity and high mean dipole moment of proximate groups of amino acid residues. An example of the RBD analysis of the SARS-CoV-2 Spike protein is given (Figure 1C). This analysis showed that 4.8% of cysteine residues were in functional domains of the SARS-CoV-2 proteome whereas a much greater proportion of arginine residues, 30.7%, were in functional domains. The enrichment of arginine residues in functional domains was 4.9-fold – the highest of any amino acid, whereas there was a slight negative enrichment, 0.8-fold, or depletion of cysteine residues in functional domains. Other amino acid residues susceptible to oxidative damage were also depleted in functional domains. Amino acid enrichment in functional domains: met 0.6, tyr 0.8 and trp 0.3 (Table 1). The SARS-CoV-2 proteome is, therefore, resistant to oxidative inactivation but susceptible to functional inactivation by MG. For individual SARS-CoV-2 proteins, the majority had arginine residue enrichment in the functional domains greater than the mean of the human host proteome of *ca.* 3.5: range 2.3 – 13.5 (Fig. 1d).

Methylglyoxal modification motif in proteins and its enrichment in the SARS-CoV-2 proteome

We next sought to develop a proteomic MG modification motif – the peptide environment of arginine residues at sites reactive towards MG modification, and map this onto the SARS-CoV-2 proteome to assess if there were MG modification sites therein and, moreover, they were enriched with respect to the human host.

Arginine residues of proteins are activated towards reaction with MG by decrease of pK_a of the guanidino side chain which facilitates formation of the MG-guanidino sidechain encounter (Rabbani et al., 2016a; Rabbani and Thornalley, 2012). Arginine residue sidechain pK_a is decreased by interaction with neighbouring amino acid residues with positively charged sidechains. For an α-helix, interactions with lysine or arginine residues at positions -4, -3, +3 and +4 in the sequence with respect to the target arginine residue are expected to decrease the arginine target residue pK_a by side chain interaction along the side of the helix axis. Longer range interactions occur between these and other types of secondary structure domains in the tertiary structures of proteins. For example, in human serum albumin, neighbouring group interactions with R186, R218 and R410 decrease the pK_a values of their sidechain guanidino groups to 12.5, 12.2 and 12.5 from the basal pKa of 13.8 (Fitch et al., 2015). The reactivity with MG theses arginine residues increases by 20 – 40 fold through increase of the trace level conjugate base of the sidechain guanidino group.(Fitch et al., 2015; Rabbani et al., 2016a). In experimental low-level extent of modification of albumin by MG *in vitro* and MG modification of human serum albumin found similarly *in vivo*, MG was detected on these residues preferentially (Ahmed et al., 2005).

To develop a proteomic MG modification motif we collated sequence data from peptides found susceptible to MG modification and all peptides detected in proteomics analysis of cultured human cells. In recent studies of the cytosolic proteome of human

endothelial cells and fibroblasts *in vitro* (Irshad et al., 2019; Rabbani et al., 2016a), we detected 686 unique arginine sites with MG modification – the MG modification motif foreground, and a total of 9,276 unique arginine residue sites – the human arginine-containing peptide background. Making no assumption on protein secondary structure environment of the arginine residue, we compared amino acid frequencies in the proximate sequence around the arginine residues and deduced enrichment of amino acids in the MG-modified peptides to develop the MG modification motif. A remarkable feature is that there was a predominance of activating residues on the N-terminal side of the target arginine residue. This may be linked to the finding that residues with several long range interactions assume a conformation with dihedral angles of φ and ψ near −120° and 140°; that is, the side chains of arginine residues interacting with other positively charged residues tend to point towards amino acid residues earlier in the sequence (Saravanan and Selvaraj, 2017). On the N-terminal side of the target, there was a 3 – 5 fold enrichment of arginine and lysine residues at positions -7 to -2 (Figure 2A and Table S3). We then assessed if the SARS-CoV-2 proteome has evidence of MG modification sites by comparing the proximate sequence to arginine residues in the viral proteome. There are 358 arginine residues in the SARS-CoV-2 proteome – potential MG modification sites (foreground). We compared this SARS-CoV-2 foreground to the human arginine sequence environment proteomic background. We found a 4-fold enrichment of arginine at position -2 in 18 sites and 5-fold enrichments of arginine residues at - 4 and – 5 at 21 and 17 sites, respectively. This shows there is an enrichment of MG modification sites in the SARS-Cov2- proteome compared to the human host (Figure 2B and Table S4). This provides the basis for selective protein modification and selective toxicity by increase of MG to the SARS-CoV-2 virus.

Methylglyoxal modification sites in the spike protein, nucleoprotein and others of SARS-CoV-2 and related coronaviridae

We applied the MG modification motif to identify the location of MG modification sites within the RBD of viral proteins. We identified the following sites: 8 sites in the nucleoprotein - R36, R40, R41, R93, R95, R191, R195 and R262; and one site in each of nsp1 (R124), nsp8 (R75), nsp12 (R555), nsp15 (R138), spike protein (R685), M-protein (R105), ORF8 (R53) and ORF10 (R234); 16 sites in total (Table 2). It is likely that pharmacological increase of endogenous MG concentration will modify the SARS-CoV-2 proteome at multiple susceptible and functional sites, producing protein inactivation and antiviral response.

Uniquely for related coronaviridae, SARS-CoV-2 spike protein has an interacting arginine triad, $R_{682}RAR_{685}$, at the S1/S2 cleavage site (Andersen et al., 2020) with R685 site of MG modification (Figure 2C). Modification of this triad by MG would confer resistance to proteolytic cleavage by transmembrane serine proteases (TMPRSSs) and blocking cell fusion for virion entry into pulmonary alveolar epithelial and other cell target sites, uncoating and replication (Meng et al., 2020). Trapped in the extracellular environment, there is excepted to be an improved host immune response to the virus; *cf.* viral host immunity boosted by similar aldehyde-modifying agents (Herrera-Rodriguez et al., 2019).

SARS-CoV-2 nucleoprotein is highly susceptible to modification and functional inactivation by MG. Nucleoprotein binds the 3' end of the viral single strand RNA genome of SARS-CoV-2 and it is arginine-rich, as is typical of RNA-binding proteins (Tan and Frankel, 1995). By analogy with SARS-Cov, residues 42-187 are involved with RNA binding with R93 playing a critical role (McBride et al., 2014). The crystal structure of a segment of nucleoprotein, residues 50 – 173, is available and this enabled prediction of pK_a values of arginine residues in this region. pK_a shifts from 13.8 of inactivated arginine (Fitch et al.,

2015) for R93 and R95 indicated *ca.* 16 and 40-fold increased reactivity towards MG modification, compared to non-activated target residue. These residues lie in a pentad of reactive arginine residues where MG modification at R93 and R95 is expected to inactive the nucleoprotein (Figure 2D). The SR-rich region of 182 – 196 is important for virus replication (Tylor et al., 2009) and is also a target for MG modification and inactivation. MG modification of the nucleoprotein, and also membrane protein, will block viral replication and virion assembly, respectively.

There are similar MG modification sites in functional domains of related *coronaviridae*. The proteome of severe acute respiratory syndrome coronavirus (SARS-CoV) had 12 similar MG modification sites to those of SARS-CoV-2, and Middle East Respiratory Syndrome (MERS) coronavirus proteome had 7 similar MG modification sites. In all of these *coronaviridae* there were multiple MG modification sites in functional domains of the nucleoprotein protein (Table 2).

Pharmacological increase of cellular methylglyoxal to virucidal levels cell permeable glyoxalase 1 inhibitor and clinical antitumor drugs, doxorubicin and paclitaxel

Antiviral activity of supraphysiological concentrations of MG was reported historically (De Bock et al., 1957); and more recently for some susceptible strains of influenza B with antiviral effect at *ca.* 20 μM MG (Charyasriwong et al., 2016). These studies used exogenous MG in cellular *in vitro* models where MG is rapidly metabolized by Glo1 and onward through the glyoxalase pathway to D-lactate (Rabbani et al., 2016b). The cellular concentration of MG is 1 – 4 μM and the plasma concentration 130 – 250 nM (Irshad et al., 2019; Rabbani and Thornalley, 2014; Xue et al., 2016). The optimum approach to achieve an antiviral effect is to increase cellular MG concentration by inhibition of Glo1. S-p-Bromobenzylglutathione cyclopentyl diester (BBGD) is a potent cell permeable Glo1

inhibitor prodrug. It delivers the Glo1 competitive inhibitor, S-p-bromobenzylglutathione (K_i = 160 nM), into cells and has established antitumor and antimalarial activity (Thornalley et al., 1996; Thornalley et al., 1994) (Figure 3A). We studied the effect of BBDG and clinically approved antitumor drugs on the cellular concentration of MG in human HEK293 cells – a cellular model employed for SARS-CoV-2 propagation (Chien et al., 2016). BBGD increased the endogenous concentration of cellular MG by 4-fold to *ca.* 20 μM – a level similar to that with virucidal activity.

We have a longstanding interest in anticancer activity of Glo1 inhibitors and overexpression of Glo1 in multidrug resistant tumors (Rabbani et al., 2018; Thornalley et al., 1996). Hence, we have been studying the likely involvement of increased MG in the mechanism of action of clinical antitumor drugs. Interestingly, we found clinical antitumor agents, doxorubicin and paclitaxel (Figure 3A), also increased cellular MG by a similar extent (Figure 3b). Increased MG concentration induced by doxorubicin and paclitaxel is linked to increased glucose metabolism and related increased formation of MG as a byproduct of glycolysis. Indeed, flux of formation of D-lactate - a surrogate indicator of flux of formation of MG - was increased by both drugs (Figure 3C).

Increase of cellular MG also likely contributes to the antiproliferative effect of BBGD (Rabbani et al., 2018). The involvement of MG in the antiproliferative activity of doxorubicin and paclitaxel is unknown. We explored the involvement of MG accumulation in the antiproliferative response of doxorubicin and paclitaxel by determining the effect of overexpression of Glo1 on inhibition of HEK293 cell growth. Vector-derived stable transfectant HEK293 cells lines were prepared with Glo1 expression increased 4 – 5 fold and empty vector transfectant control (Figure 3D and 3E). When these transfectant cell lines were treated with growth inhibitory concentrations of drugs, there was an increase of median growth inhibitory concentration GC_{50} value and resistance to inhibition of cell growth in

HEK293 cells with stable overexpression of Glo1 (Figure 3F and 3H). The GC_{50} values were (mean ± SD; empty vector vs Glo1+): doxorubicin, 3.54 ± 0.28 nM vs 55.9 ± 3.4 nM (16-fold resistance); paclitaxel – 6.8 ± 1.0 nM vs 56.4 ± 7.2 nM (8-fold resistance). For treatment with BBGD there was an antiproliferative effect with limited change in GC_{50} value with Glo1 overexpression: 4.78 ± 0.18 μM vs 7.37 ± 0.30 μM (2-fold resistance). The limited effect on antiproliferative effect of BBGD is expected as the delivered Glo1 inhibitor also inhibits the overexpression factor, Glo1.

DISCUSSION

Taken together, our findings describe herein provide evidence of vulnerability of SARS-CoV-2 to modification and inactivation by MG and a scientific rationale for repurposing of doxorubicin and paclitaxel for treatment of COVID-19 disease.

The enrichment of arginine residues in functional domains of the SARS-CoV-2 proteome provides important collateral to support an arginine-modifying agent strategy for inactivation of the virus and virucidal activity. Arginine residues are also enriched in the human host proteome but less so that in SARS-CoV-2; 3.6 versus 4.9. This characteristic of the human proteome was noted previously (Gallet et al., 2000) – now updated by us with UniProtKB updated sequence information.

We report here for the first time a proteomic motif for MG modification. Combination of this with RBD analysis, enabled us to predict number and location of sites in the SARS-CoV-2 proteome susceptible to MG modification and likely to inactivate key functionality of the virus. This provides key evidence to explore expected virucidal activity of exogenous MG and pharmacological agents that may increase cellular MG concentration to virucidal level. Importantly, features of the peptide sequence motif for MG modification were enriched 4 – 5 fold in the SARS-CoV-2 compared to human host – a further feature indicating that

pharmacological agents increasing cellular concentration of MG, inducing dicarbonyl stress, are expected to have selective toxicity to the virus, compared to the human host.

A further important feature for susceptibility of viral proteins to MG modification is protein abundance or protein copy number. Previous studies of the SARS-CoV virion suggested proteins of high copy number (tens to hundreds) in decreasing copy number order are: nucleoprotein, M-protein, Spike protein and nsp3 (Neuman et al., 2008). Assuming similar protein copy number of SARS-CoV-2, we have identified MG modification sites in the 3 most abundant proteins of the SARS-CoV-2 proteome.

Although there were most MG modification sites in functional domains in SARS-CoV-2, 16, other *coronaviridae* - SARS-CoV-2 and MERS - had similar MG modification sites in protein crucial for virion viability – particularly the nucleoprotein protein. This suggests that pharmacological agents increasing cellular concentration of MG, inducing dicarbonyl stress, may have virucidal activity against multiple *coronaviridae*.

Doxorubicin and paclitaxel are clinical antiproliferative antitumor agents with mechanisms of action targeting inhibition of topoisomerase-II in DNA replication and stabilization of the interphase and microtubular network and mitotic spindle in mitosis, respectively (Schiff et al., 1979; Tewey et al., 1984). Herein, we show that increase in MG contributes to their mechanism of antiproliferative activity. Doxorubicin increases glucose metabolism by increasing expression of glucose transporter GLUT1 and hexokinase-2 (Demel et al., 2015). Paclitaxel stabilizes microtubules, decreasing free tubulin concentration; the latter increasing mitochondrial voltage-dependent anion channel (VDAC) activity and thereby *in situ* activity of hexokinase (Maldonado et al., 2010). These mechanisms are available in the lung epithelial cells primarily targeted by SARS-Cov-2 (Lottes et al., 2014; Pezzulo et al., 2011). Increased glucose metabolism produces a corresponding increase in the formation of MG – evidenced herein by increase in flux of formation of D-lactate; there may

be disproportionately large increase in MG if expression of enzymes of onward metabolism of triosephosphates, triosephosphate isomerase and glyceraldehyde-3-phosphate dehydrogenase, are not increased along with hexokinase activity and glycolysis becomes dysregulated or unscheduled (Irshad et al., 2019; Rabbani and Thornalley, 2019). Doxorubicin and paclitaxel have been evaluated for antiviral activity, particularly with respect to inhibition of viral helicase (Ash and Diekema, 1987; Bergamini et al., 1992; Borowski et al., 2002; Briguglio et al., 2011). In the current search for drugs to repurpose for antiviral activity, we suggest they be considered for evaluation against COVID-19 disease. Other investigators have suggested strategies for repurposing drugs based SARS Cov-2 protein interactions with human host proteins and drugs targeted to them, virion endosomal processing and viral protease inhibition (Chen et al., 2020; Gordon et al., 2020; Wang et al., 2020). Our approach rather addresses the intrinsic vulnerability of SARS Cov-2 proteome to endogenous reactive metabolites, with respect to the human host, and identified drugs to exploit this.

We note that Manuka honey contains high levels of MG – up to 760 mg/kg or *ca*. 15 mM (Mavric et al., 2008). It has antiviral activity against alphaherpesvirus and influenzavirus *in vitro* (Shahzad and Cohrs, 2012; Watanabe et al., 2014). It is not known if Manuka honey has virucidal activity against SARS-Cov-2.

Additional materials.

KEY RESOURCES TABLE

LEAD CONTACT AND MATERIALS AVAILABILITY

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

METHOD DETAILS

Sequences of SARS Cov-2, SARS-Cov and MERs and human host proteins.

Receptor binding domain analysis

Peptide motif site of methylglyoxal modification

Culture of HEK293 cells *in vitro*.

QUANTIFICATION AND STATISTICAL ANALYSIS

DATA AVAILABILITY

SUPPLEMENTAL INFORMATION

REFERENCES

Ahmed, N., Dobler, D., Dean, M., and Thornalley, P.J. (2005). Peptide mapping identifies hotspot site of modification in human serum albumin by methylglyoxal involved in ligand binding and esterase activity. JBiolChem *280*, 5724-5732.

Ahmed, U., Dobler, D., Larkin, S.J., Rabbani, N., and Thornalley, P.J. (2008). Reversal of hyperglycemia-induced angiogenesis deficit of human endothelial cells by overexpression of glyoxalase 1 in vitro. AnnNYAcadSci *1126*, 262-264.

Allen, R.E., Lo, T.W.C., and Thornalley, P.J. (1993). A simplified method for the purification of human red blood cell glyoxalase I. Characteristics, immunoblotting and inhibitor studies. JProtChem *12*, 111 - 119.

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nature Medicine.

Arai, M., Nihonmatsu-Kikuchi, N., Itokawa, M., Rabbani, N., and Thornalley, P.J. (2014). Measurement of glyoxalase activities. BiochemSocTrans *42*, 491-494.

Ash, R.J., and Diekema, K.A. (1987). Inhibition of herpes simplex virus replication by anthracycline compounds. Antiviral Research *8*, 71-83.

Bergamini, A., Perno, C.F., Balzarini, J., Capozzi, M., Marinelli, L., Milanese, G., Pesce, C.D., Calió, R., and Rocchi, G. (1992). Selective Inhibition of HIV Replication by Adriamycin in Macrophages But Not in Lymphocytes. AIDS Research and Human Retroviruses *8*, 1239-1247.

Borowski, P., Schalinski, S., and Schmitz, H. (2002). Nucleotide triphosphatase/helicase of hepatitis C virus as a target for antiviral therapy. Antiviral Research *55*, 397-412.

Briguglio, I., Piras, S., Corona, P., and Carta, A. (2011). Inhibition of RNA Helicases of ssRNA(+) Virus Belonging to Flaviviridae, Coronaviridae and Picornaviridae Families. International journal of medicinal chemistry *2011*, 213135.

Chan, J.F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K.K.-W., Yuan, S., and Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerging Microbes & Infections *9*, 221-236.

Charyasriwong, S., Haruyama, T., and Kobayashi, N. (2016). In vitro evaluation of the antiviral activity of methylglyoxal against influenza B virus infection. Drug Discoveries & Therapeutics *10*, 201-210.

Chen, Y., Yiu, C., and Wong, K. (2020). Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates [version 1; peer review: 3 approved]. F1000Research *9*.

Chien, H.-C., Zur, A.A., Maurer, T.S., Yee, S.W., Tolsma, J., Jasper, P., Scott, D.O., and Giacomini, K.M. (2016). Rapid Method To Determine Intracellular Drug Concentrations in Cellular Uptake Assays: Application to Metformin in Organic Cation Transporter 1-Transfected Human Embryonic Kidney 293 Cells. Drug Metab Dispos *44*, 356-364.

Chou, M.F., and Schwartz, D. (2011). Biological Sequence Motif Discovery Using motif-x. Current Protocols in Bioinformatics *35*, 13.15.11-13.15.24.

De Bock, C.A., Brug, J., and Walop, J.N. (1957). Antiviral Activity of Glyoxals. Nature *179*, 706-707.

Demel, H.R., Feuerecker, B., Piontek, G., Seidl, C., Blechert, B., Pickhard, A., and Essler, M. (2015). Effects of topoisomerase inhibitors that induce DNA damage response on glucose metabolism and PI3K/Akt/mTOR signaling in multiple myeloma cells. Am J Cancer Res *5*, 1649-1664.

Fitch, C.A., Platzer, G., Okon, M., Garcia-Moreno, B.E., and McIntosh, L.P. (2015). Arginine: Its pKa value revisited. Protein Sci *24*, 752-761.

Gallet, X., Charloteaux, B., Thomas, A., and Braseur, R. (2000). A fast method to predict protein interaction sites from sequences. JMolBiol *302*, 917-926.

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., O'Meara, M.J., Guo, J.Z., Swaney, D.L., Tummino, T.A., Huettenhain, R., *et al.* (2020). A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. bioRxiv, 2020.2003.2022.002386.

Herrera-Rodriguez, J., Signorazzi, A., Holtrop, M., de Vries-Idema, J., and Huckriede, A. (2019). Inactivated or damaged? Comparing the effect of inactivation methods on influenza virions to optimize vaccine production. Vaccine *37*, 1630-1637.

Irshad, Z., Xue, M., Ashour, A., Larkin, J.R., Thornalley, P.J., and Rabbani, N. (2019). Activation of the unfolded protein response in high glucose treated endothelial cells is mediated by methylglyoxal. Sci Rep *9*, 7889.

Lottes, R.G., Newton, D.A., Spyropoulos, D.D., and Baatz, J.E. (2014). Alveolar type II cells maintain bioenergetic homeostasis in hypoxia through metabolic and molecular adaptation. Am J Physiol Lung Cell Mol Physiol *306*, L947-L955.

Madeira, F., Park, Y.m., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., *et al.* (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research *47*, W636-W641.

Maldonado, E.N., Patnaik, J., Mullins, M.R., and Lemasters, J.J. (2010). Free tubulin modulates mitochondrial membrane potential in cancer cells. Cancer research *70*, 10192-10201.

Mavric, E., Wittmann, S., Barth, G., and Henle, T. (2008). Identification and quantification of methylglyoxal as the dominant antibacterial constituent of Manuka (Leptospermum scoparium) honeys from New Zealand. Molecular Nutrition & Food Research *52*, 483-489.

McBride, R., Van Zyl, M., and Fielding, B.C. (2014). The Coronavirus Nucleocapsid Is a Multifunctional Protein. Viruses *6*, 2991-3018.

Meng, T., Cao, H., Zhang, H., Kang, Z., Xu, D., Gong, H., Wang, J., Li, Z., Cui, X., Xu, H., *et al.* (2020). The insert sequence in SARS-CoV-2 enhances spike protein cleavage by TMPRSS. bioRxiv, 2020.2002.2008.926006.

Neuman, B.W., Joseph, J.S., Saikatendu, K.S., Serrano, P., Chatterjee, A., Johnson, M.A., Liao, L., Klaus, J.P., Yates, J.R., Wüthrich, K., *et al.* (2008). Proteomics Analysis Unravels the Functional Repertoire of Coronavirus Nonstructural Protein 3. Journal of Virology *82*, 5279-5294.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Pezzulo, A.A., Gutiérrez, J., Duschner, K.S., McConnell, K.S., Taft, P.J., Ernst, S.E., Yahr, T.L., Rahmouni, K., Klesney-Tait, J., Stoltz, D.A., *et al.* (2011). Glucose Depletion in the Airway Surface Liquid Is Essential for Sterility of the Airways. PLOS ONE *6*, e16166.

Rabbani, N., Ashour, A., and Thornalley, P.J. (2016a). Mass spectrometric determination of early and advanced glycation in biology. Glycoconjugate Journal *33*, 553-568.

Rabbani, N., and Thornalley, P.J. (2012). Methylglyoxal, glyoxalase 1 and the dicarbonyl proteome. Amino Acids *42*, 1133-1142.

Rabbani, N., and Thornalley, P.J. (2014). Measurement of methylglyoxal by stable isotopic dilution analysis LC-MS/MS with corroborative prediction in physiological samples. Nature Protocols *9*, 1969-1979.

Rabbani, N., and Thornalley, P.J. (2019). Hexokinase-2 Glycolytic Overload in Diabetes and Ischemia–Reperfusion Injury. Trends in Endocrinology & Metabolism *30*, 419-431.

Rabbani, N., Xue, M., and Thornalley, P.J. (2016b). Methylglyoxal-induced dicarbonyl stress in aging and disease: first steps towards glyoxalase 1-based treatments. ClinSci *130*, 1677–1696.

Rabbani, N., Xue, M., Weickert, M.O., and Thornalley, P.J. (2018). Multiple roles of glyoxalase 1-mediated suppression of methylglyoxal glycation in cancer biology—Involvement in tumour suppression, tumour growth, multidrug resistance and target for chemotherapy. Seminars in Cancer Biology *49*, 83-93.

Saethang, T., Hodge, K., Yang, C.-R., Zhao, Y., Kimkong, I., Knepper, M.A., and Pisitkun, T. (2019). PTM-Logo: a program for generation of sequence logos based on position-specific background amino-acid probabilities. Bioinformatics *35*, 5313-5314.

Saravanan, K.M., and Selvaraj, S. (2017). Dihedral angle preferences of amino acid residues forming various non-local interactions in proteins. J Biol Phys *43*, 265-278.

Schiff, P.B., Fant, J., and Horwitz, S.B. (1979). Promotion of microtubule assembly in vitro by taxol. Nature *277*, 665-667.

Shahzad, A., and Cohrs, R.J. (2012). In vitro antiviral activity of honey against varicella zoster virus (VZV): A translational medicine study for potential remedy for shingles. Transl Biomed *3*, 2.

Sies, H., Berndt, C., and Jones, D.P. (2017). Oxidative Stress. Annual Review of Biochemistry *86*, 715-748.

Tan, R., and Frankel, A.D. (1995). Structural variety of arginine-rich RNA-binding peptides. Proceedings of the National Academy of Sciences of the United States of America *92*, 5282-5286.

Tewey, K., Rowe, T., Yang, L., Halligan, B., and Liu, L. (1984). Adriamycin-induced DNA damage mediated by mammalian DNA topoisomerase II. Science *226*, 466-468.

Thornalley, P.J., Edwards, L.G., Kang, Y., Wyatt, C., Davies, N., Ladan, M.J., and Double, J. (1996). Antitumour activity of S-p- bromobenzylglutathione cyclopentyl diester in vitro and in vivo. Inhibition of glyoxalase I and induction of apoptosis. BiochemPharmacol *51*, 1365-1372.

Thornalley, P.J., Strath, M., and Wilson, R.J.M. (1994). Anti-malarial activity in vitro of the glyoxalase I inhibitor diester, S-p-bromobenzylglutathione diethyl ester. BiochemPharmacol *268*, 14189-14825.

Tylor, S., Andonov, A., Cutts, T., Cao, J., Grudesky, E., Van Domselaar, G., Li, X., and He, R. (2009). The SR-rich motif in SARS-CoV nucleocapsid protein is important for virus replication. Canadian Journal of Microbiology *55*, 254-260.

Wang, L., Zhang, M., and Alexov, E. (2016). DelPhiPKa web server: predicting pKa of proteins, RNAs and DNAs. Bioinformatics *32*, 614-615.

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W., and Xiao, G. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Research *30*, 269-271.

Watanabe, K., Rahmasari, R., Matsunaga, A., Haruyama, T., and Kobayashi, N. (2014). Anti-influenza Viral Effects of Honey In Vitro: Potent High Activity of Manuka Honey. Archives of Medical Research *45*, 359-365.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260-1263.

Xue, M., Rabbani, N., and Thornalley, P.J. (2014). Measurement of glyoxalase gene expression. BiochemSocTrans *42*, 495-499.

Xue, M., Weickert, M.O., Qureshi, S., Ngianga-Bakwin, K., Anwar, A., Waldron, M., Shafie, A., Messenger, D., Fowler, M., Jenkins, G., *et al.* (2016). Improved glycemic control and

vascular function in overweight and obese subjects by glyoxalase 1 inducer formulation Diabetes *65*, 2282-2294.

ACKNOWLEDGMENTS

MA-M thanks the Qatar Foundation for a PhD studentship. PJT thanks the Qatar Foundation for a funding his research program. NR thanks Qatar University for funding her research program. We thank Dr Amal Ashour for culture of human periodontal fibroblasts and preparation of tryptic digests for proteomics analysis.

AUTHOR CONTRIBUTIONS

MA-M accessed protein sequence information on the SARS-CoV-2 , applied RBD analysis and molecular graphics. HA cultured HEK293 cells, prepared and propagated plasmids, prepared stable transfectant cell lines and performed metabolite and drug treatment studies. PW checked, collated and curated data on arginine peptides. AdlF performed peptide motif analysis. MX provided technical guidance and support to HA and performed SARS-CoV2, SARS-CoV and MERS sequence alignments. NR analysed the mass spectrometry proteomics data. PJT assisted with MG analysis. NR and PJT acquired the funding, designed and supervised the studies, contributed to the data analysis and wrote the manuscript. All authors read, edited and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Table 1. Receptor binding domain of SARS-CoV-2 viral proteomes.

Amino acid	N		Prevalence		Proportion in RBD (%)	Fold Enrichment
	All	RBD	All	RBD		
Arg	358	110	3.63	17.7	30.7	4.9
Cys	310	15	3.14	2.4	4.8	0.8
Met	203	7	2.06	1.1	3.5	0.6
Tyr	448	23	4.54	3.7	5.1	0.8
Trp	113	2	1.15	0.3	1.8	0.3

RBD analysis was applied to SARS-CoV-2 proteome (see Table S1) using a window of 5 amino acids and gyration angle between two consecutive residues in the sequence of 100° (Gallet et al., 2000). Complete amino acid profile given in Table S2.

Table 2. SARS-CoV-2, SARS-CoV and MERS proteins with putative activated arginine residues in functional domains.

Protein	Activated arginine in RBD of SARS-CoV-2 proteome (16)	Analogous sites in similar *coronaviridae*	
		SARS-CoV (12)	MERS (7)
nsp1	R124 (AYRKVLLR)	R124 (AYRNVLLR)	
nsp8	R75 (TQMYKQAR)		
nsp12	R555 (ISAKNRAR)	R555 (ISAKNRAR)	R556 (ISAKNRAR)
nsp15	R138 (VDLFRNAR)	R138 (VDLFRNAR)	
Spike protein	R685 (TNSPRRAR)		R751 (TLTPRSVR)
M-protein	R105 (ASFRLFAR)	R104 (ASFRLFAR)	R104 (QSIRLFMR)
ORF8	R52 (WYIRVGAR)		
ORF10	R234 (LCCRMNSR)		

Table 2. SARS-CoV-2, SARS-CoV and MERS proteins with putative activated arginine residues in functional domains (cont'd).

Protein	Activated arginine in RBD of SARS-CoV-2 proteome (16)	Analogous sites in similar *coronaviridae* SARS-CoV (12)	MERS (7)
Nucleoprotein (NC)	R36 (NGE<u>R</u>SGA<u>R</u>)	R37 (NGG<u>R</u>NGA<u>R</u>)	R27 (NTNLS<u>R</u>G<u>R</u>)
	R40 (SGA<u>R</u>S<u>K</u>Q<u>R</u>)	R41 (NGA<u>R</u>P<u>K</u>Q<u>R</u>)	
	R41 (GA<u>R</u>S<u>K</u>QR<u>R</u>)	R42 (GA<u>R</u>P<u>K</u>QR<u>R</u>)	
	R93 (YY<u>RR</u>ATR<u>R</u>; $pK_a \approx 12.6$)	R94 (YY<u>RR</u>ATR<u>R</u>)	
	R95 (<u>RR</u>AT<u>RR</u>I<u>R</u>; $pK_a \approx 12.2$)	R96 (<u>RR</u>AT<u>RR</u>V<u>R</u>)	
	R191 (S<u>R</u>SSS<u>R</u>S<u>R</u>)	R192 (S<u>R</u>SSS<u>R</u>S<u>R</u>)	R180 (S<u>R</u>ASSLS<u>R</u>)
	R195 (S<u>R</u>S<u>R</u>NSS<u>R</u>)	R196 (S<u>R</u>S<u>R</u>GNS<u>R</u>)	R184 (SLS<u>R</u>NSS<u>R</u>
	R262 (S<u>K</u>K<u>P</u>R<u>Q</u>K<u>R</u>)	R263 (S<u>K</u>K<u>P</u>R<u>Q</u>K<u>R</u>)	R254 (<u>K</u>N<u>K</u>M<u>R</u>HK<u>R</u>)

Key: <u>R</u>, activated arginine, <u>R</u> and <u>K</u>, neighbouring interacting residues decreasing pKa of the target <u>R</u> residue according to the MG modification. Arginine side chain predicted pK_a values are given, deduced by DelPhiPKa program(Wang et al., 2016) using nucleoprotein

crystal structure (pdb file 6VYO; Chang *et al*., to be published) and AMBER forcefield (predictions were similar with CHARMM and PARSE forcefields). Sequence alignment of SARS-Cov-2, SARS-Cov and MERS proteins was performed using the Clustal Omega software on-line (Madeira et al., 2019).

FIGURE LEGENDS.

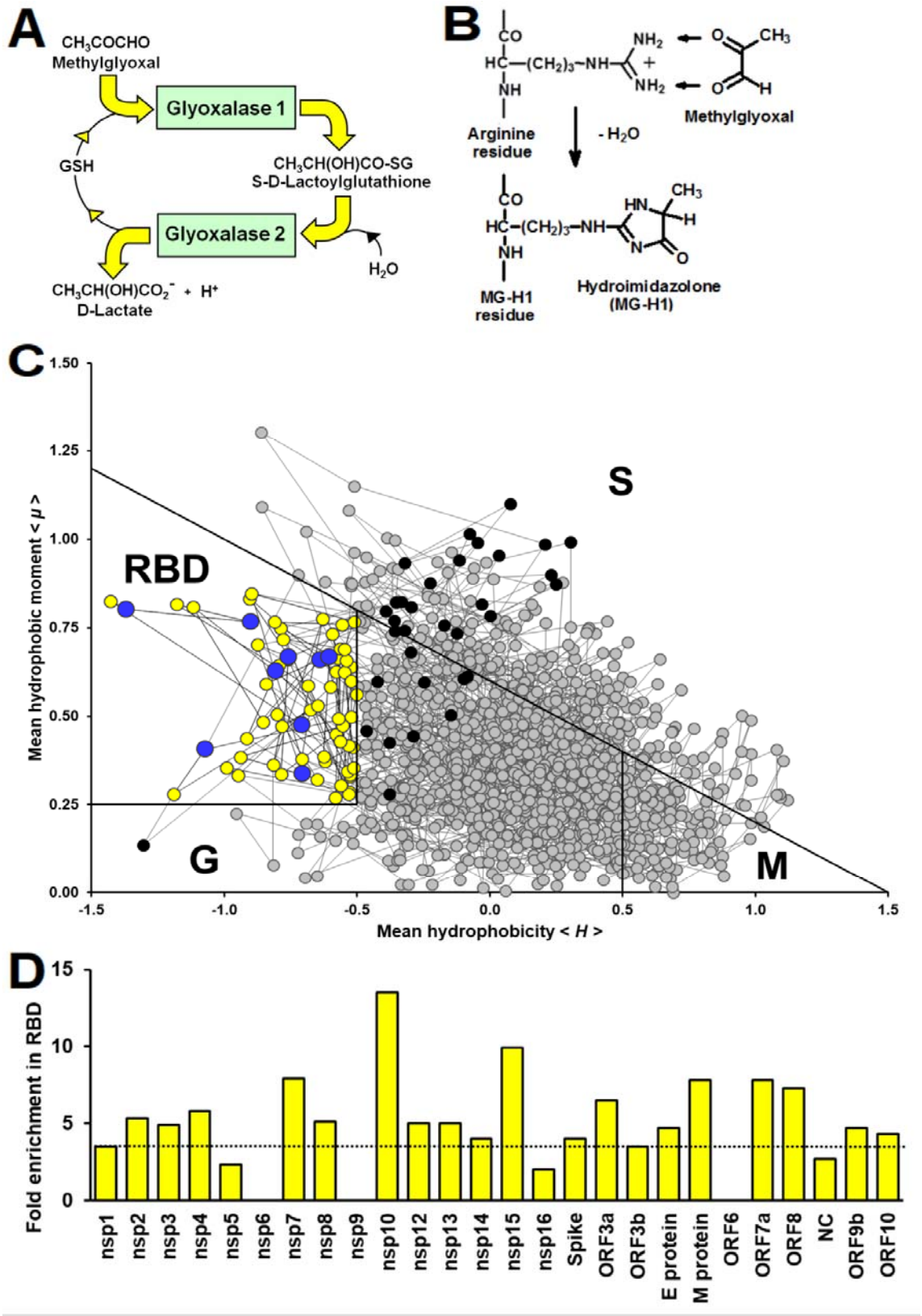
Figure 1. Methylglyoxal – an endogenous arginine-modifying reactive metabolite and receptor binding domain analysis of functional arginines in the SARS-CoV-2 proteome. **A.** Metabolism of methylglyoxal by the glyoxalase pathway (Rabbani et al., 2016b). **B.** Modification of arginine residues by MG to form hydroimidazolone, MG-H1. **C.** Receptor binding domain (RBD) plot for SARS-CoV-2 Spike protein. Line-linked filled circles represent the primary sequence. The RBD is the area bound by the trapezium in the upper left-side region of the chart. Key:  , arginine residue in the RBD;  , arginine residues outside the RBD;  , other amino acid residues in the RBD other amino acid residues in the RBD; and  , other amino acid residues outside the RBD. Other predicted domains: surface (S), globular (G) and membrane (M). **D.** Arginine enrichment in individual proteins of SARS-CoV-2 proteins. Proteins not shown have no arginine residues (nsp11, ORF7b and ORF14).

Figure 2. Development of a methylglyoxal-modification motif and evidence of activation of functional arginine residues towards modification by methylglyoxal in the SARS-CoV-2 proteome. **A**. Methylglyoxal protein modification motif. Over-representation of amino acids in peptide sequences of MG-H1 detected peptides, compared to human background cytosolic proteome. **B**. Over-representation of amino acids in arginine residue environment, compared to human background cytosolic proteome. Threshold count set to 10 and significance, $P<1 \times 10^{-6}$. Structural models of activated arginine residues in the SARS-CoV-2 proteome. **C.** Activated arginine residues triad of Spike protein, $R_{682}R_{683}AR_{685}$. **D**. Activated arginine residue pentad of nucleoprotein, $R_{88}R_{89}ATR_{92}R_{93}IR_{95}$. Molecular graphics produced from Spike protein (pdb file 6VSB) (Wrapp et al., 2020) and nucleoprotein segment crystal

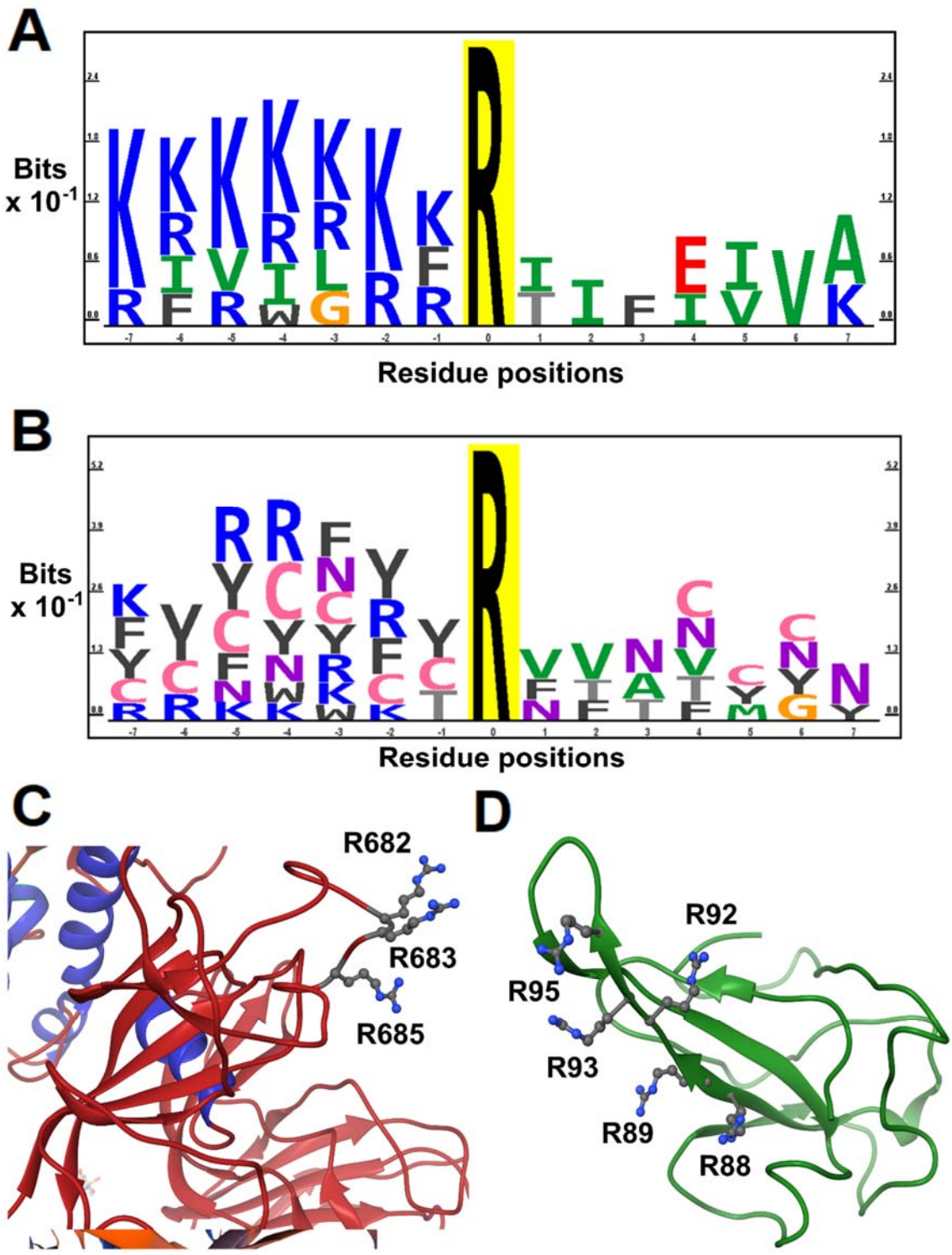
structure (pdb file 6VYO; Chang *et al*., to be published) using Chimera 1.14 (Pettersen et al., 2004).

Figure 3. Glyoxalase 1 inhibitor prodrug, doxorubicin and paclitaxel increase cellular concentration of methylglyoxal to virucidal levels. A. Molecular structures of drugs. Glyoxalase 1 inhibitor prodrug, S-p-bromobenzylglutathione cyclopentyl diester (BBGD). Delivers competitive Glo1 inhibitor, S-p-bromobenzylglutathione, K_i = 160 nM, into cells (Allen et al., 1993; Thornalley et al., 1996). Doxorubicin – topoisomerase inhibitor.(Tewey et al., 1984) Paclitaxel – stabilizer of microtubule assembly (Schiff et al., 1979). **B.** and **C.** Increase in cellular MG in HEK293 cells and flux of formation of D-lactate (surrogate for flux of MG), respectively, incubated *in vitro* with and without investigational agent and drugs indicated. For assay MG, cells were incubated with and without treatment for 3 h and for flux of D-lactate incubated for 24 h. Data are mean ± SEM (n = 3 except n = 4 for MG estimation with Doxorubicin; individual data points are shown). Drug concentrations: BBGD, 7.4 μM; doxorubicin, 6.0 nM, paclitaxel, 21 nM. Significance: b. $P<0.02$ and c. $P<0.01$ (*one-way ANOVA*) and * and ***, $P<0.05$ and $P<0.001$ with respect to control (*t-test*). **D.** and **E.** Activity and protein of Glo1, respectively, in HEK293 cells: wild-type (WT) and cells stably transfected to overexpress Glo1 (Glo1+) and empty vector (EV). Glo1 activity and protein were increased 4 – 5 fold. This was maintained for > 10 passages. **F.** – **H.** Effect of Glo1 expression on anti-proliferative activity. Key: blue – empty vector, red – Glo1 overexpression. HEK293 cells were incubated with and without treatment for 48 h. Data (6 drug concentrations in triplicate) were fitted by non-linear regression to the dose-response equation, $V = 100 \times GC_{50}^n/(GC_{50}^n + [Drug]^n)$, solving for GC_{50} and, n (logistic regression coefficient). **F.** Doxorubicin: empty vector, GC_{50} = 3.54 ± 0.28 nM, n = 0.71 ± 0.05; Glo1+, GC_{50} = 55.9 ± 3.4 nM, n = 1.24 ± 0.10 (16-fold resistance). **G** Paclitaxel: empty vector, GC_{50} = 6.8 ± 1.0 nM, n

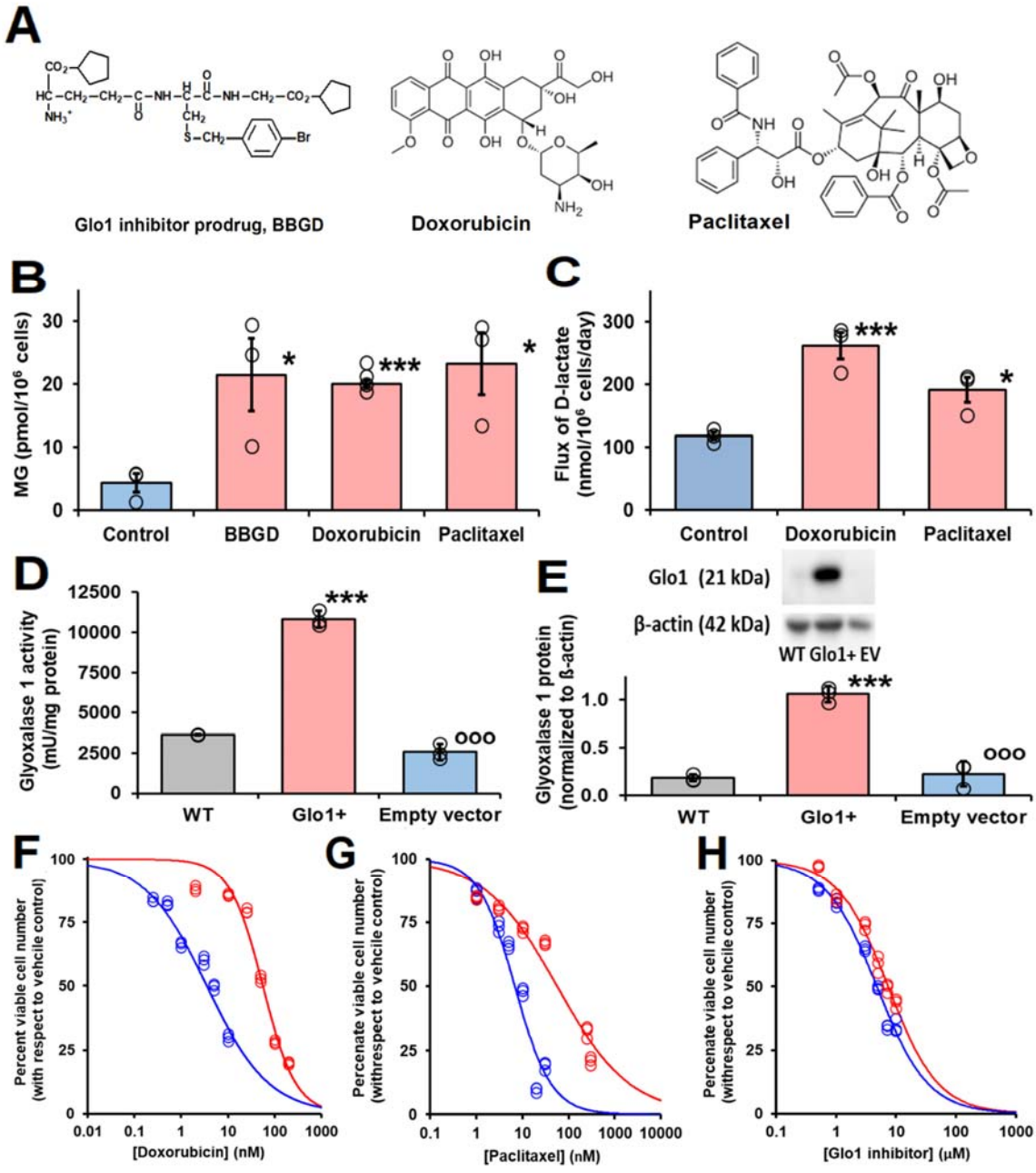
= 1.07 ± 0.17; and Glo1+, GC_{50} = 56.4 ± 7.2 nM, n = 0.55 ± 0.04 (8-fold resistance). **H.** BBGD: GC_{50} = 4.78 ± 0.18 μM, n = 1.02 ± 0.05; and Glo1+, GC_{50} = 7.37 ± 0.30 μM, n = 1.04 ± 0.06 (2-fold resistance).



Vulnerabilities of the SARS-CoV-2 virus to proteotoxicity *etc*., Figure 1



Vulnerabilities of the SARS-CoV-2 virus to proteotoxicity *etc*., Figure 2



Vulnerabilities of the SARS-CoV-2 virus to proteotoxicity *etc*., Figure 3

STAR+METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal anti-Glo1 antibody (rat)	Sigma-Aldrich (Poole, Dorset, UK)	cat. no. SAB4200193
Anti-Rat IgG (whole molecule)–Biotin conjugate	Sigma-Aldrich (Poole, Dorset, UK)	cat. no. B7139
Chemicals		
S-p-Bromobenzylglutathione cyclopentyl diester (BBGD)	PJT laboratory	N/A
Doxorubicin	Sigma-Aldrich (Poole, Dorset, UK)	cat. no. D1515
Paclitaxel	Sigma-Aldrich (Poole, Dorset, UK))	cat. no. T7402
D-Lactic dehydrogenase	Sigma-Aldrich (Poole, Dorset, UK)	cat. no. L9636
Geneticin G-418 (potency rating – 700 μg)	Fisher Scientific (Loughborough, UK)	cat. no. 11811023
Experimental Models: Cell Lines		
HEK293 cell line	ATCC (Virginia, USA)	cat. no. CRL-1573
Human periodontal fibroblasts	ScienCell (Carlsbad, USA)	cat. no. 2630

KEY RESOURCES TABLE (cont'd).

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pIRES2-GLO1-EGFP	PJT laboratory	N/A
pIRES-EGFP	PJT laboratory	N/A
Software and Algorithms		
Rmotifx R-package	General Public License	https://github.com/omarwagih/rmotifx
PTM Logo Java application	Freeware	https://hpcwebapps.cit.nih.gov/PTMLogo/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to the Lead Contact, Paul. J Thornalley (pthornalley@hbku.edu.qa).

Cell lines

Human periodontal fibroblasts (PDLFs, ScienCell, Carlsbad, USA; cat. no. 2630) were cultured in Modified Eagles Medium (MEM) Modified Eagles Medium (MEM) supplemented with 2 mm L-alanyl-L-glutamine, 10% fetal bovine serum, 100 Units/ml penicillin, 100 µg/ml streptomycin at 37ºC under aseptic conditions and 5% carbon dioxide/air with 100% humidity. A cytosolic extract was prepared, processed and analysed as described previously (Irshad et al., 2019).

The HEK293 cell line , seeding density 2 x 10^4 cells cm^{-2}, was cultured in Dulbecco's Modified Eagles Medium (DMEM) containing phenol red, L-glutamine and 4500 mg/L glucose, supplemented with 10% Foetal Bovine Serum (FBS), 100 U penicillin and 0.1 mg/ml streptomycin.

METHODS DETAILS

Sequences of SARS Cov-2, SARS-Cov and MERs and human host proteins.

Reference sequences of the 29 proteins of the SARS-CoV-2 proteome (Table S1) and sequences of analogous proteins of SARS-Cov and MERs *coronaviridae* were obtained from the NBCI reference sequence database (www.ncbi.nlm.nih.gov). Sequences of reviewed proteins of the human proteome, 18,821 – excluding fetal proteins, were obtained from the UniProtKB database (www.uniprot.org).

Receptor binding domain analysis

RBD analysis is a protein primary sequenced based informatics method to deduce amino acid residues in functional domains of proteins – defined as sites of protein-protein, protein-nucleic acid and protein-ligand or substrate interaction. It is applicable to any protein. The optimised protocol uses a window of 5 amino acid residues moved sequentially along the sequence of a protein, with a gyration angle between two consecutive residues in the sequence of 100° assumed, to deduce sequential mean Eisenberg hydrophobicity and mean dipole moment to assign to the central amino acid. Values cannot be deduced for the two amino acids at the N- and C-termini of proteins and they are therefore missing from the amino acid residue prevalence reports (Gallet et al., 2000). We developed an R script to obtain mean hydrophobicity and hydrophobic moment for all UniProtKB proteins and SARS-CoV-2 proteins.

Peptide motif site of methylglyoxal modification

We used proteomic data from our recent study human endothelial cell HMEC-1 cytosolic extract and identified 3862 arginine residue containing peptides in control tryptic and lys-C digests, including 411 MG-H1 containing peptides in cytosolic extract incubated with MG to increase the extent of MG modification to the upper physiological limit of 5 mmol/mol arg (Irshad et al., 2019). A similar cytosolic extract from human periodontal fibroblasts (PDLFs, ScienCell, Carlsbad, USA; cat. no. 2630), cultured in Modified Eagles Medium (MEM) Modified Eagles Medium (MEM) supplemented with 2 mm L-alanyl-L-glutamine, 10% fetal bovine serum, 100 Units/ml penicillin, 100 µg/ml streptomycin at 37^{o}C under aseptic conditions and 5% carbon dioxide/air with 100% humidity, was processed and analysed similarly. These were combined and used with the peptide analysis programme – motif-x (Chou and Schwartz, 2011), with a foreground of 686 MG-H1 peptides and a background of 9,726

unmodified peptides. Experimentally detected arginine peptides rather than whole sequence arginine residues was used as background to control for detection bias; mean sequence coverage in proteomics analysis was *ca.* 23% (Irshad et al., 2019). Motifs are ranked by scores calculated as the sum of the negative log of the binomial probabilities used to generate the motifs. The outcome is presented as a single amino acid letter, size-related fractional presence, compared to background, sequence logo with a threshold P-value of 1×10^{-5} (X^2 test); produced using the PTM Logo, Java application (Saethang et al., 2019). The arginine residue sequence environment in the SARS Cov-2 proteome, 358 sites, was then also compared to the human host background.

Culture of HEK293 cells *in vitro*.

pIRES2-GLO1-EGFP plasmid (Glo1+ vector) and pIRES-EGFP plasmid (empty vector) were prepared as described (Ahmed et al., 2008). HEK293 cells stably transfected with Glo1+ and empty vector Lipofectamine 2000 according to the manufacturer's instructions (plasmid DNA: Lipofectamine 2000, 1:4). After 48 h, cells were sub-cultured, G-418 disulphate was added (2 mg/ml; 405 µg/mg potency) culture continued. Transfected colonies with GFP fluorescence were selected using a cloning disc (3.2 mm) and glass cylinder selector (8 mm, 150 µl) and cultured further with G-418 disulphate (1 mg/ml, 705 µg/mg potency) containing medium. Assessment of Glo1 activity and protein, as described (Arai et al., 2014; Xue et al., 2014), indicated a 4 - 5 fold increase in Glo1 activity and protein. HEK293 cells stably tranfected with empty and GLO1+ vectors were incubated with and without BBGD, doxorubucin and paclitaxel at the concentrations indicated (diluted from 100 mM stock solution in DMSO) for 2 days and effect on cell growth assessed by viable cell number counts, using the Trypan blue exclusion method and median growth inhibitory concentrations GC_{50} deduced. Cellular MG concentration and flux of formation of D-lactate,

a surrogate measure of flux of formation of MG, was assayed as described (Irshad et al., 2019; Rabbani and Thornalley, 2014).

QUANTIFICATION AND STATISTICAL ANALYSIS

Criteria for peptide identification in proteomic data analysis were published previously (Irshad et al., 2019). For comparison of amino acid frequencies in peptide MG modification motif, the sequence logo threshold P-value was 1×10^{-5} (X^2 test). For other significance tests, data normality was assumed and significance of 3 or more groups assessed by one-way *ANOVA* and two groups by *Student's t-test*.

SUPPLEMENTARY TABLES.

Table S1. SARS-CoV-2 viral proteome.

No	Protein (acronym(s))	Reference sequence (no of a.a. residues)	Function
1.	Non-structural protein 1 (nsp1)	YP_009725297.1 (180 residues)	Inhibits host translation by interacting with the 40S ribosomal subunit.
2.	Non-structural protein 2 (nsp2)	YP_009725298.1 (638 residues)	May modulate host cell survival signaling pathway by interacting with host prohibitin-1 and prohibitin-2.
3.	Non-structural protein 3 /Papain-like proteinase (nsp3, PL-PRO)	YP_009725299.1 (1945 residues)	Cleaves N-terminus of the replicase polyprotein. With nsp4, participates in assembly of viral membrane assembly. Antagonizes type I interferon innate immunity by. Prevents host NF-kappa-B signaling.
4.	Non-structural protein 4 (nsp4)	YP_009725300.1 (500 residues)	Participates in the virion double-membrane vesicles assembly.
5.	Non-structural protein 4 (nsp5, 3C-like proteinase, 3CL-PRO, Main protease)	YP_009725301.1 (306 residues)	Cleaves C-terminus of replicase polyprotein R1ab at 11 sites.
6.	Non-structural protein 6 (nsp6)	YP_009725302.1 (290 residues)	Role in initial induction and regulation of autophagosomes from host reticulum endoplasmic.
7.	Non-structural protein 7 (nsp7)	YP_009725303.1 (83 residues)	Forms a hexadecamer with nsp8 that may participate in viral replication by acting as a primase or may synthesize longer oligonucleotides.
8.	Non-structural protein 8 (nsp8)	YP_009725304.1 (198 residues)	Forms a hexadecamer with nsp7 that may participate in viral replication by acting as a primase or may synthesize longer oligonucleotides.
9.	Non-structural protein 9 (nsp9)	YP_009725305.1 (113 residues)	May be an ssRNA-binding protein participating in viral replication.
10.	Non-structural protein 10 (nsp10)	YP_009725306.1 (139 residues)	Pivotal role in viral transcription by stimulating both nsp14 3'-5' exoribonuclease and nsp16 2'-O-methyltransferase activities for viral mRNAs cap methylation.
11.	Non-structural protein 11 (nsp11)	YP_009725312.1 (13 residues)	

Table S1. SARS-CoV-2 viral proteome (cont'd).

No	Protein (acronym(s))	Reference sequence (no of a.a. residues)	Function
12.	RNA-directed RNA polymerase (nsp12, Pol/RdRp)	YP_009725307.1 (932 residues)	Replication and transcription of the viral RNA genome. Part of orf1ab polyprotein.
13.	Helicase (nsp13, Hel)	YP_009725308.1 (601 residues)	Multi-functional protein with RNA and DNA duplex-unwinding activities with 5' to 3' polarity.
14.	Guanine-N7 methyl-transferase (nsp14, ExoN)	YP_009725309.1 (527 residues)	Exoribonuclease activity acting on both ssRNA and dsRNA in a 3' to 5' direction. N^7-guanine methyltransferase activity.
15.	Uridylate-specific endoribonuclease (nsp15, NendoU)	YP_009725310.1 (346 residues)	Mn2+-dependent, uridylate-specific enzyme, which cleaves 2'-3'-cyclic phosphates to 3'-phosphate ester.
16.	2'-O-ribose methyl-transferase (nsp16)	YP_009725311.1 (298 residues)	Methyltransferase mediating mRNA cap 2'-O-ribose methylation to the 5'-cap structure of viral mRNAs - prerequisite for binding of nsp16. Essential function to evade immune system.
17.	Spike glycoprotein (Spike protein S1, S-glycoprotein)	YP_009724390.1 (1273 residues)	Attaches the virion to the cell membrane by interacting with host receptor, initiating the infection. Binding to human ACE2 and CLEC4M/DC-SIGNR receptors.
18.	Protein 3a (ORF3a)	YP_009724391.1 (275 residues)	Forms homotetrameric K^+-sensitive ion channels which may modulate virus release. Up-regulates FGA, FGB and FGG fibrinogen subunit expression in host lung epithelial cells. Downregulates the type 1 interferon receptor.
19.	Protein 3b (ORF3b)*	(57 residues)	Inhibits expression of ß-interferon.
20.	Envelope small membrane protein (E protein)	YP_009724392.1 (75 residues)	Viroporin in host membranes forming pentameric protein-lipid pores allowing ion transport. Also involved in the induction of apoptosis.
21.	Membrane protein (M protein)	YP_009724393.1 (222 residues)	Component of the viral envelope involved in virus morphogenesis and assembly.
22.	Non-structural protein 6 (ns6, ORF6)	YP_009724394.1 (61 residues)	Determinant of virus virulence.

Table S1. SARS-CoV-2 viral proteome (cont'd).

No	Protein (acronym(s))	Reference sequence (no of a.a. residues)	Function
23.	Protein 7a (ORF7a)	YP_009724395.1 (121 residues)	Non-structural protein of unknown function
24.	Protein non-structural 7b (ORF7b)	YP_009725318.1 (43 residues)	Unknown function
25.	Protein non-structural 8 (ORF8)	YP_009724396.1 (121 residues)	May play a role in host virus interaction.
26.	Nucleoprotein (NC, ORF9)	YP_009724397.2 (419 residues)	Packages the positive strand viral genome RNA into a helical ribonucleocapsid. Enhances efficiency of subgenomic viral RNA transcription as well as viral replication.
27.	Protein 9b (ORF9b)	ORF9B_WCPV (97 residues)	Unknown function.
28.	ORF10 protein	YP_009725255.1 (38 residues)	Unknown function.
29.	Uncharacterized protein 14 (ORF14)	Y14_WCPV (73 residues)	Unknown

*Sequence from reference (Chan et al., 2020).

Table S2. Receptor binding domain of SARS-CoV-2 viral proteome.

Amino	N		Prevalence		Proportion	Fold
acid	All	(RBD)	All	RBD	in RBD (%)	Enrichment
Ala	667	18	6.76	2.9	2.7	0.4
Arg	358	110	3.63	17.7	30.7	4.9
Asn	537	42	5.44	6.8	7.8	1.2
Asp	514	45	5.21	7.3	8.8	1.4
Cys	310	15	3.14	2.4	4.8	0.8
Gln	369	29	3.74	4.7	7.9	1.3
Glu	452	41	4.58	6.6	9.1	1.4
Gly	574	22	5.82	3.5	3.8	0.6
His	193	17	1.96	2.7	8.8	1.4
Ile	514	8	5.21	1.3	1.6	0.3
Leu	950	27	9.63	4.4	2.8	0.5
Lys	568	74	5.76	11.9	13.0	2.1
Met	203	7	2.06	1.1	3.5	0.6
Phe	504	13	5.11	2.1	2.6	0.4
Pro	404	23	4.09	3.7	5.7	0.9
Ser	665	53	6.74	8.5	8.0	1.3
Thr	728	33	7.38	5.3	4.5	0.7
Trp	113	2	1.15	0.3	1.8	0.3
Tyr	448	23	4.54	3.7	5.1	0.8
Val	796	18	8.07	2.9	2.3	0.4
Total:	9867	620	100	100		

RBD analysis was applied to SARS-CoV-2 proteome (see Table S1) using a window of 5 amino acids and gyration angle between two consecutive residues in the sequence of 100° (Gallet et al., 2000).

Table S3. Methylglyoxal modification motif. Human MG-H1 residue sequence foreground compared to total arginine residue sequence background. Over-representation of amino acids in peptide sequences of MG-H1 detected peptides, compared to human background cytosolic proteome.

Rank	Motif	Score	Foreground matches	Foreground size	Background matches	Background size	Fold increase
1	KxxxxxxRxxxxxxx	16.0	65	686	278	9726	3.3
2	xxxxxKxRxxxxxxx	13.0	52	621	237	9448	3.3
3	xxxKxxxRxxxxxxx	11.3	37	569	153	9211	3.9
4	xxKxxxxRxxxxxxx	8.5	38	532	214	9058	3.0
5	xKxxxxxRxxxxxxx	6.6	34	494	225	8844	2.7
6	xxxxxRxRxxxxxxx	6.5	26	460	151	8619	3.2
7	xxxxKxxRxxxxxxx	6.7	28	434	175	8468	3.1
8	xxxRxxxRxxxxxxx	5.5	14	372	66	8103	4.6
9	xxxxRxxRxxxxxxx	5.3	18	406	103	8293	3.6
10	xRxxxxxRxxxxxxx	5.2	16	388	87	8190	3.9

Enrichments shown are significant at $P<1 \times 10^{-6}$ and rank ordered by match score matches.

Significance chart (amino acid single letter code):

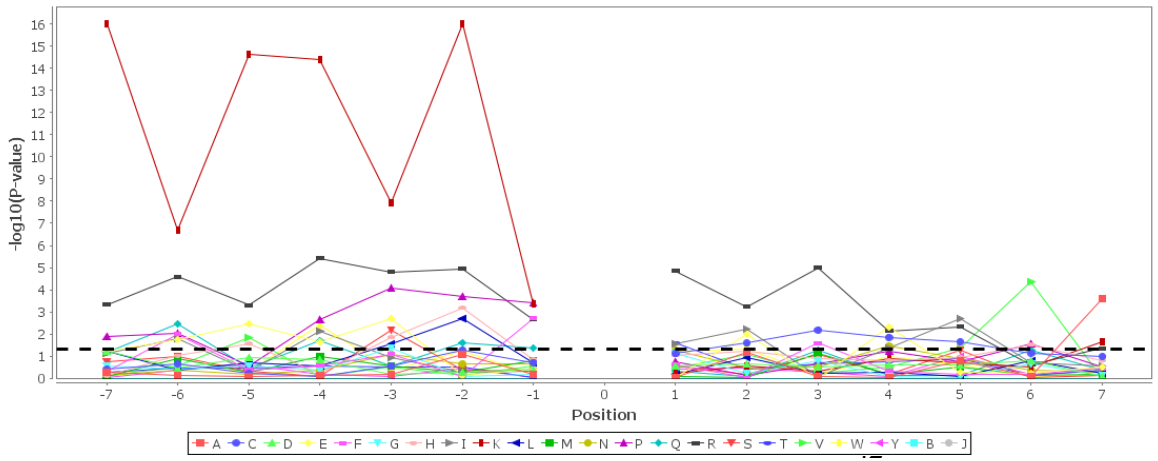


Table S4. MG modification motif. Total SARS-CoV-2 arginine residue sequence foreground compared to total human arginine residue sequence background.

Over-representation of amino acids in arginine residue environment of the SARS-Cov-2, compared to human background cytosolic proteome.

Rank	Motif	Score	Foreground matches	Foreground size	Background matches	Background size	Fold increase
1	xxxCxxxRxxxxxxx	8.3	16	358	65	9726	6.7
2	xxxRxxxRxxxxxxx	8.4	21	342	120	9661	4.9
3	xxRxxxxRxxxxxxx	7.1	17	321	100	9541	5.1
4	xxxxxYxRxxxxxxx	6.1	22	304	199	9441	3.4
5	xxxxxRxRxxxxxxx	6.0	18	282	148	9242	4.0

Enrichments shown are significant at $P<1 \times 10^{-6}$ and rank ordered by match score.

Significance chart (amino acid single letter code):