

## Inhibitors

## Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays\*\*

Tarick J. El-Baba, Corinne A. Lutomski, Anastassia L. Kantsadi, Tika R. Malla, Tobias John, Victor Mikhailov, Jani R. Bolla, Christopher J. Schofield, Nicole Zitzmann, Ioannis Vakonakis, and Carol V. Robinson\*

**Abstract:** The SARS-CoV-2 main protease ( $M^{\text{Pro}}$ ) cleaves along the two viral polypeptides to release non-structural proteins required for viral replication.  $M^{\text{Pro}}$  is an attractive target for antiviral therapies to combat the coronavirus-2019 disease. Here, we used native mass spectrometry to characterize the functional unit of  $M^{\text{Pro}}$ . Analysis of the monomer/dimer equilibria reveals a dissociation constant of  $K_d = 0.14 \pm 0.03 \mu\text{M}$ , indicating  $M^{\text{Pro}}$  has a strong preference to dimerize in solution. We characterized substrate turnover rates by following temporal changes in the enzyme-substrate complexes, and screened small molecules, that bind distant from the active site, for their ability to modulate activity. These compounds, including one proposed to disrupt the dimer, slow the rate of substrate processing by  $\approx 35\%$ . This information, together with analysis of the x-ray crystal structures, provides a starting point for the development of more potent molecules that allosterically regulate  $M^{\text{Pro}}$  activity.

The coronavirus, SARS-CoV-2, is the etiological agent of the 2020 pandemic that has claimed  $> 550\,000$  lives and affected more than 12 million people as of July 2020.<sup>[1]</sup> Coronaviruses have long existed in nature and have made zoonotic transmission to humans. Despite the tragic and widespread effects of these sudden occurrences, we do not yet have validated anti-viral treatments targeting coronavirus infections. SARS-CoV-2 packages a large RNA genome of  $\approx 30$  k bases, two-thirds of which encodes for two polyproteins (pp1a and pp1b). These polyproteins are processed into 16 non-

How to cite: *Angew. Chem. Int. Ed.* **2020**, 59, 23544–23548  
International Edition: doi.org/10.1002/anie.202010316  
German Edition: doi.org/10.1002/ange.202010316

structural proteins (nsps) that are liberated from the long polypeptide chains by two viral proteases, the papain-like protease (nsp 3) and the 3C-like protease (nsp 5). The latter species, named the main protease  $M^{\text{Pro}}$  is a cysteine protease that cleaves the viral polyproteins at eleven sites to generate twelve non-structural proteins (nsp5–nsp16). Included in these nsps are those involved in the replication machinery (e.g., the RNA-dependent RNA polymerase nsp12).<sup>[2]</sup> Inhibition of  $M^{\text{Pro}}$  impairs the ability of the virus to replicate.

Analogous to the 2004 SARS-CoV main protease, the functional unit of the SARS-CoV-2  $M^{\text{Pro}}$  is a homodimer (Figure 1).<sup>[3]</sup> Several encouraging strategies to inhibit  $M^{\text{Pro}}$  have been explored via its covalent inhibition.<sup>[3a,b]</sup> Non-covalent modulators, including compounds that disrupt the dimer interface, have not been investigated to the same extent. Drugs that bind non-covalently can often be fine-tuned to diffuse through membranes and bind to target proteins with high affinity.<sup>[4]</sup> Moreover, these species lack reactive warheads often leading to higher chemical stability than their covalent counterparts and a reduction in undesirable toxic effects due to irreversible binding to host proteins and nucleic acids. Therefore, non-covalent compounds serve as a promising means to inhibit viral proliferation.

Previous studies of the SARS-CoV  $M^{\text{Pro}}$  dimer identified nano to micromolar dissociation constants.<sup>[5]</sup> We report the dissociation constant for SARS-CoV-2  $M^{\text{Pro}}$  determined using native mass spectrometry (MS), which can directly identify and quantify the relative amounts of the oligomeric state of a protein in solution.<sup>[6]</sup> To probe the monomer/dimer equilibrium we recorded native mass spectra over a range

[\*] Dr. T. J. El-Baba, Dr. C. A. Lutomski, Dr. J. R. Bolla, Prof. C. V. Robinson  
Physical and Theoretical Chemistry Laboratory, University of Oxford  
South Parks Rd. OX1 3QZ Oxford (UK)  
E-mail: carol.robinson@chem.ox.ac.uk

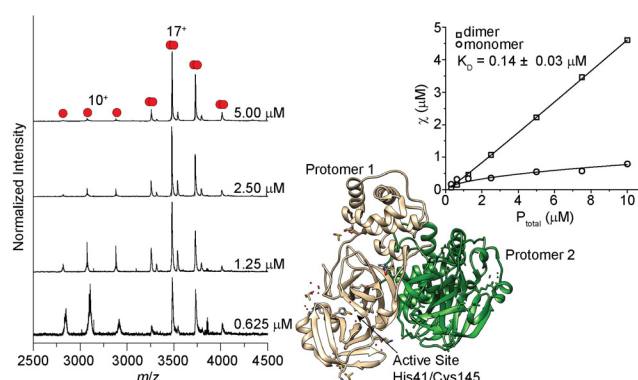
Dr. A. L. Kantsadi, Prof. N. Zitzmann, Prof. I. Vakonakis  
Department of Biochemistry, University of Oxford  
South Parks Rd. OX1 3QU, Oxford (UK)

T. R. Malla, T. John, Dr. V. Mikhailov, Prof. C. J. Schofield  
Chemistry Research Laboratory, University of Oxford  
12 Mansfield Rd, OX1 3TA, Oxford (UK)

[\*\*] A previous version of this manuscript has been deposited on a preprint server (<https://doi.org/10.1101/2020.07.29.226761>).

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: <https://doi.org/10.1002/anie.202010316>.

© 2020 The Authors. Published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

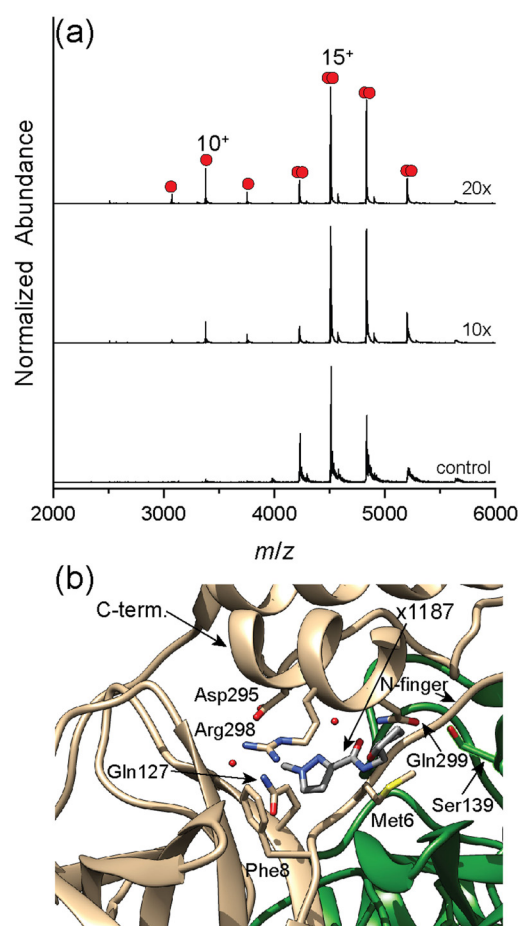


**Figure 1.** Analysis of  $M^{\text{Pro}}$  by native MS. Left: Native mass spectra for  $M^{\text{Pro}}$  at different concentrations. Right: Representative plot of mole fraction versus concentration to quantify the dissociation constant and a view from an X-ray structure of unligated  $M^{\text{Pro}}$  (PDB ID: 6YB7).

of  $M^{\text{Pro}}$  concentrations from 0.313 to 10.0  $\mu\text{M}$  in an aqueous buffer containing 200 mM ammonium acetate (pH 7.4). At a protein concentration of 5  $\mu\text{M}$ , two well-resolved charge state distributions are readily identified. The high-abundant signals centered at the 17<sup>+</sup> charge state correspond to a deconvoluted mass of  $67591 \pm 0.5$  Da, consistent with the expected sequence mass of dimeric  $M^{\text{Pro}}$  (67 592 Da). A minor peak series between  $m/z \approx 2750$  and 3500, is also observed. These signals, centered at a 10<sup>+</sup> charge state, correspond to a deconvoluted mass of  $33795 \pm 3$  Da, in excellent agreement with the mass of monomeric  $M^{\text{Pro}}$  with wildtype N- and C-termini (33 796 Da). As the concentration of protein is decreased in a series of stepwise dilutions from 5.0  $\mu\text{M}$  to 0.625  $\mu\text{M}$ , the signals corresponding to the  $M^{\text{Pro}}$  monomer increase concurrent with a decrease in the peak intensities assigned to the dimer.

To extract a monomer/dimer equilibrium constant we performed measurements in triplicate at seven different protein concentrations (from 0.325 to 10  $\mu\text{M}$ ) and plotted the mole fraction of each species as a function of total protein concentration (Figure 1). Excellent agreement was found between the measured values and a monomer/dimer equilibrium binding model (see Experimental section in the Supporting Information) providing high confidence in the derived dissociation constant,  $K_d = 0.14 \pm 0.03$   $\mu\text{M}$ . This compares with an estimated  $K_d \approx 2.5$   $\mu\text{M}$  reported recently determined using analytical ultracentrifugation.<sup>[3b]</sup> The difference between the estimated value and our measurement may originate from the different buffer conditions and temperatures (20 mM Tris, 150 mM NaCl at 20 °C versus 200 mM ammonium acetate at  $\approx 26$  °C used here) both of which could perturb the monomer/dimer equilibrium. Nevertheless, we conclude that, analogous to SARS-CoV  $M^{\text{Pro}}$ , the SARS-CoV-2 variant has a high propensity to form dimers.

Following the deposition of high-resolution SARS-CoV-2  $M^{\text{Pro}}$  structures to the protein databank,<sup>[3a,b]</sup> a fragment screen was released providing structural insight towards 96 candidates that bind  $M^{\text{Pro}}$ .<sup>[3c]</sup> Three of these candidates were identified as binding to the dimer interface; 23 non-covalent and 48 covalent hits were found to bind the active site. We focused our study to those small molecules that bind non-covalently to  $M^{\text{Pro}}$  and considered their potential to both destabilize the dimer and modulate substrate cleavage. We identified four candidates, from the subset that were immediately available to us, due to their binding properties: three bind to the solvent exposed surface (x0390, x0425, and x0464) and one binds within the dimer interface (x1187). We incubated each species with 5  $\mu\text{M}$   $M^{\text{Pro}}$  at a ligand concentration range of 1 to 100  $\mu\text{M}$  (0.125–20-fold molar excess). Following incubation for 30 mins we recorded mass spectra to investigate their effect on the monomer/dimer equilibrium. x0390, x0425, and x0464 showed no appreciable perturbation of the monomer/dimer ratios (Figure S1). However, the effect of x1187 on dimerization was clearly apparent (Figure 2). In the absence of x1187, under the solvent conditions used to dissolve the small molecules, the charge state distribution is centered at 15<sup>+</sup> reinforcing the notion that  $M^{\text{Pro}}$  is predominantly a dimer under a range of solution conditions. Following the addition of a 10-fold molar excess of x1187



**Figure 2.** a) Native mass spectra of 5  $\mu\text{M}$   $M^{\text{Pro}}$  with the addition of different molar equivalents of x1187. To maintain similar solution conditions to the samples containing x1187, the control contains 10% DMSO. b) Detailed view of dimer interface where x1187 binds (PDB ID: 5FRA).

the peaks assigned to monomers, centered at the 10<sup>+</sup> charge state, are observed to increase. After addition of a 20-fold molar excess of x1187 the fractional abundance of the monomer peaks increases further to  $\approx 15\%$ . Saturation transfer difference nuclear magnetic resonance spectroscopy studies indicate the small molecules bind tightly to  $M^{\text{Pro}}$  in solution.<sup>[7]</sup> The absence of peaks in the mass spectrum for the ligand-bound forms of  $M^{\text{Pro}}$  likely results from the labile nature of small-molecule binding upon transit to the gas-phase. Moreover, comparing ligand bound species via mass spectrometry would require a different set of parameters to those optimized to monitor the monomer/dimer equilibrium. Our experiments allow us to conclude however that x1187, while bound in solution, disrupts the dimeric form of  $M^{\text{Pro}}$ .

Inspection of the crystal structure identifies the likely origin of this destabilization; x1187 rests across the dimer interface, packing into a hydrophobic pocket partially comprised of key residues Met6 and Phe8 on the N-finger (Figure 2). The long axis of x1187 rests along the C-terminal helix of protomer 1 and is proximal to Ser139 on protomer 2, a critical residue for proper function and assembly.<sup>[8]</sup> The position of x1187 is important as it rests in a pocket crucial to

formation of a dimer with a productive topology,<sup>[9]</sup> prompting us to consider whether or not this, or the other small molecules that bind non-covalently, mediate catalysis.

To identify the ability for these molecules to mediate the proteolytic activity of M<sup>Pro</sup> we first characterized the substrate turnover in the absence of small molecules using an MS-based kinetic assay. Rather than using a simple peptide cleavage assay, which here we found to be unreliable due to the variability in the ionization of the peptides generated, we elected to quantify changes in observable enzyme-substrate complexes as a function of time. These complexes are notoriously difficult to isolate and quantify as they are turned over rapidly; but as they transition from solution to the gas phase, they are effectively “flash-frozen”. Quenching the reaction enables interrogation of these transient species captured in the MS instrument.<sup>[10]</sup> A mass spectrum collected 30 s after initiating the cleavage reaction is shown (Figure 3). Several satellite peaks are observed alongside the main charge state series, one of which corresponds to the enzyme substrate complex with a mass of  $68\,781 \pm 2$  Da, in agreement with dimeric M<sup>Pro</sup> bound to a single 11-mer substrate (1192 Da). We also identified a series of peaks with a deconvoluted mass of  $68\,188 \pm 0.5$  Da, consistent with the acyl-enzyme complex that is formed by reaction of the nucleophilic Cys145 with the scissile Q-S peptide bond of the substrate to give a covalent TSAVLQ-enzyme complex of mass 68192 Da (Figure 3). A summary of the M<sup>Pro</sup>-substrate complexes we identified are listed in Table 1.

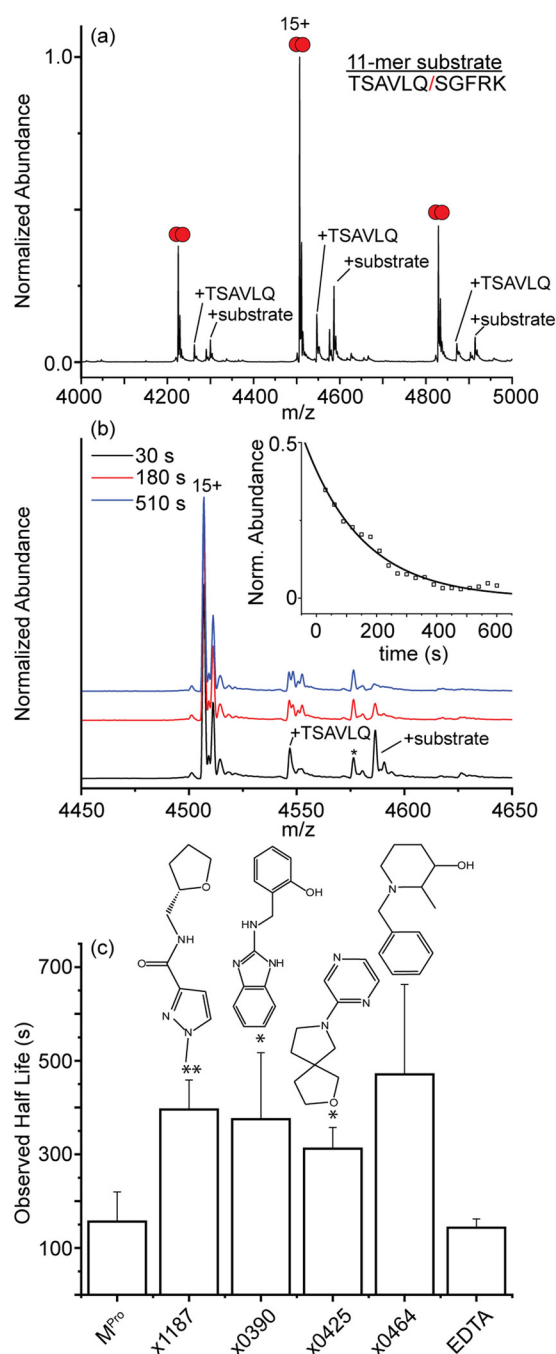
Interestingly, we did not observe any peaks corresponding to monomeric M<sup>Pro</sup> bound to the 11-mer substrate or acyl-enzyme complex (Figure S2). We considered several reasons for the absence of signals for substrate-bound monomers. One possibility is that the monomers have enhanced activity compared to the dimer. The timescale we measure for substrate cleavage by the dimer is on the order of minutes. Thus, if monomers were to bind and process substrates the monomer activity would need to be enhanced by  $\approx 100$ -fold (complete turnover by  $\approx 6$  s), otherwise we would readily capture monomer-substrate complexes. This is unlikely, as other assays have indicated that monomeric M<sup>Pro</sup> is inactive.<sup>[5c]</sup> Our results indicate that monomers are not only inactive but also they do not bind the 11-mer substrate with high affinity.

An expansion of the 15<sup>+</sup> charge state at three representative time points reveals that as time evolves, the signal for the enzyme-substrate complex at 4586 *m/z* is depleted (Figure 3). At our longest timepoint (10 min), the signals for the enzyme-substrate complex are no longer present in detectable quantities indicating that by this time the substrate 11-mer

**Table 1:** M<sup>Pro</sup> and M<sup>Pro</sup>-substrate complexes identified by native MS

Species	Measured Mass [Da] <sup>[a]</sup>	Expected Mass [Da]	$\Delta$ mass [Da]
M <sup>Pro</sup> (Monomer)	$33\,795 \pm 3$	33 796	-1
M <sup>Pro</sup> (Dimer)	$67\,591 \pm 0.5$	67 592	-1
Enzyme-substrate complex	$68\,781 \pm 2$	68 784	-2
Enzyme-TSAVLQ complex	$68\,188 \pm 0.5$	68 192	-4

[a] Uncertainty in deconvoluted mass determined using at least three charge states.



**Figure 3.** a) Native mass spectrum for 5  $\mu$ M M<sup>Pro</sup> with 50  $\mu$ M of the 11-mer substrate at  $t = 30$  s. Peaks labelled TSAVLQ and + substrate indicate acyl-enzyme complex and the non-covalent enzyme-substrate complex, respectively. b) mass spectra for the 15<sup>+</sup> charge state at three representative times along the substrate cleavage reaction. Satellite peaks adjacent to the 15<sup>+</sup> charge state are consistent with oxidation of between 4 and 8 of the ten methionine residues (+16 Da). Peaks marked with asterisk corresponds to an impurity (+1042 Da). Inset: plot of the relative abundance of the enzyme-substrate complex as a function of time. Solid line indicates the fit to a unimolecular kinetics model. c) Bar plot summarizing half-lives of the enzyme-substrate complex in the presence of different small molecules. Error bars represent standard deviation ( $n = 3$  independent replicates). \* $p < 0.05$ , \*\* $p < 0.001$  (to M<sup>Pro</sup> values). Representative mass spectra and kinetic plots for each dataset are shown in Figure S4.



has been proteolyzed and the products released (Figure S3). A plot of the relative abundance of the decay of the enzyme-substrate complex versus reaction time reveals that the decay is exponential (Figure 3). The experimental data can be readily fit to a single-step unimolecular kinetics model, providing a relatively straightforward means to compare half-life changes in the presence of inhibitors. A plot comparing the half-lives obtained by incorporating the 20-fold molar excess of the different small molecules into the reaction mixtures is shown (Figure 3 and S4). We note that the presence of these small molecules increases the lifetime of the enzyme-substrate complex, in some cases by as much as  $\approx 40\%$ .

As control experiments, we first characterized the propensity for substrate turnover using a potent inhibitor IPA3 that covalently modified Cys145 and found no appreciable changes in substrate turnover, formation of the enzyme-substrate complex, or acyl-enzyme complex (Figure S5). We also characterized the kinetic behavior of  $M^{\text{Pro}}$  in the presence of 50-fold molar excess of EDTA. We found no appreciable change in the depletion of the enzyme-substrate complex, suggesting that trace amounts of divalent metal cations, (e.g.,  $\text{Zn}^{2+}$ ) when added in isolation, are not modulators of  $M^{\text{Pro}}$  activity.<sup>[11]</sup>

All of the small molecule candidates tested inhibit  $M^{\text{Pro}}$  proteolytic activity and bind to locations distant from the active site (Figure 4). Specifically, x1187 binds at the dimer interface and x0390, x0425, and x0464 bind to solvent-exposed pockets remote from the active site. While x1187 disrupts the dimer, its proximity to key residues in the vicinity of the N-finger may also allow it to intervene with the N-finger in regulating activity<sup>[9]</sup> and allows us to conclude that all of these fragments inhibit substrate cleavage via allosteric regulation.

More generally, the therapeutic potential of these, or analogous small molecules, can be readily assessed through the dual MS approaches described here: probing the monomer/dimer equilibrium and measuring efficacy in reducing substrate turnover rates. These relatively straightforward

measurements delineate the mechanism of action of these allosteric regulators—either via disruption of the dimer interface or reversible non-competitive binding, distal from the active site. Overall, in addition to highlighting new ways of assaying these small molecules, the results provide a starting point for lead optimization chemistry with the ultimate goal of deactivating SARS-CoV-2  $M^{\text{Pro}}$  and reducing the burden of the ongoing COVID-19 pandemic.

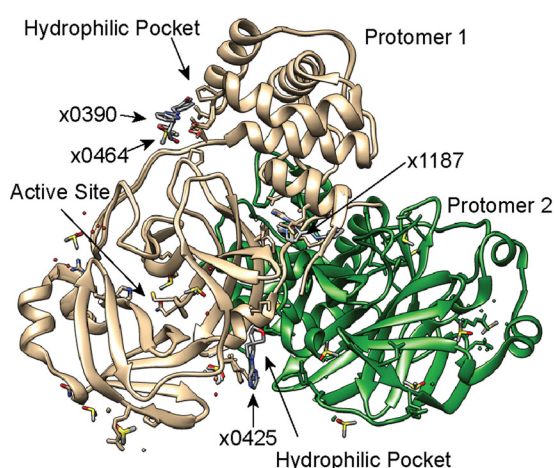
## Acknowledgements

We are grateful to Claire Strain-Damerell, Petra Lukacik, David Owen and Martin A. Walsh for providing the  $M^{\text{Pro}}$  expression construct and advice on protein production, and to the Diamond Light Source XChem facility, led by Frank von Delft, for providing small molecule ligands. Work in the C.V.R. laboratory is supported by a Medical Research Council (MRC) program grant (MR/N020413/1), a European Research Council Advanced Grant ENABLE (695511), and a Wellcome Trust Investigator Award (104633/Z/14/Z). T.J.E. is supported by the Royal Society as a Royal Society Newton International Fellow. C.A.L. is supported by the European Commission as a Marie Skłodowska Curie Postdoctoral Fellow. A.L.K. was supported by the Oxford Glycobiology Institute Endowment. T.R.M. thanks Biotechnology and Biological Sciences Research Council (BB/M011224/1). T.J. was supported by the Oxford-GSK-Crick Doctoral Program in Chemical Biology, EPSRC (EP/R512060/1) and GlaxoSmithKline. C.J.S. thanks the Wellcome Trust and the Medical Research Council for support. We are grateful for generous support provided by the University of Oxford COVID-19 Research Response fund and its donors (BRD00230). Carol Robinson is also a part of the COVID-19 mass spectrometry consortium.<sup>[12]</sup>

## Conflict of interest

The authors declare no conflict of interest.

**Keywords:** allosteric inhibitors · drug development · proteases · native mass spectrometry · SARS-CoV-2



**Figure 4.** Superimposition of views of x-ray structures for unligated  $M^{\text{Pro}}$  (PDB 6YB7) and  $M^{\text{Pro}}$  bound to x1187 (PDB 5RFA), x0390 (PDB 5REC), x0464 (PDB 5REE), and x0425 (5RGJ).

[1] Values taken from: <https://coronavirus.jhu.edu>.

[2] Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L. W. Guddat, Q. Wang, Z. Lou, Z. Rao, *Science* **2020**, *368*, 779–782.

[3] a) Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, *Nature* **2020**, *582*, 289–293; b) L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, *Science* **2020**, *368*, 409–412; c) A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandaõ-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L.

- Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keely, G. M. Keserü, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O'Brien, A. Powell, R. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft, M. A. Walsh, *BioRxiv* **2020** <https://doi.org/10.1101/2020.05.27.118117>.
- [4] S. Zhang, X. Xue, L. Zhang, L. Zhang, Z. Liu, *Chem. Biol. Drug Des.* **2015**, *86*, 1411–1424.
- [5] a) W.-C. Hsu, H.-C. Chang, C.-Y. Chou, P.-J. Tsai, P.-I. Lin, G.-G. Chang, *J. Biol. Chem.* **2005**, *280*, 22741–22748; b) V. Graziano, W. J. McGrath, L. Yang, W. F. Mangel, *Biochemistry* **2006**, *45*, 14632–14641; c) B. Xia, X. Kang, *Protein Cell* **2011**, *2*, 282–290.
- [6] K. Gupta, J. A. C. Donlan, J. T. S. Hopper, P. Uzdavinys, M. Landreh, W. B. Struwe, D. Drew, A. J. Baldwin, P. J. Stansfeld, C. V. Robinson, *Nature* **2017**, *541*, 421–424.
- [7] A. L. Kantsadi, I. Vakonakis, *BioRxiv* **2020** <https://doi.org/10.1101/2020.06.17.156679>.
- [8] S. Chen, J. Zhang, T. Hu, K. Chen, H. Jiang, X. Shen, *J. Biochem.* **2007**, *143*, 525–536.
- [9] a) N. Zhong, S. Zhang, P. Zou, J. Chen, X. Kang, Z. Li, C. Liang, C. Jin, B. Xia, *J. Virol.* **2008**, *82*, 4227–4234; b) X. Xue, H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, Z. Rao, *J. Mol. Biol.* **2007**, *366*, 965–975.
- [10] D. R. Fuller, C. R. Conant, T. J. El-Baba, C. J. Brown, D. W. Woodall, D. H. Russell, D. E. Clemmer, *J. Am. Chem. Soc.* **2018**, *140*, 9357–9360.
- [11] F. Wang, C. Chen, X. Liu, K. Yang, X. Xu, H. Yang, T. S. Dermody, *J. Virol.* **2016**, *90*, 1910–1917.
- [12] W. Struwe, E. Emmott, M. Bailey, M. Sharon, A. Sinz, F. J. Corrales, K. Thalassinou, J. Braybrook, C. Mills, P. Barran, *Lancet* **2020**, *395*, 1761–1762.

Manuscript received: July 28, 2020

Revised manuscript received: August 20, 2020

Accepted manuscript online: August 25, 2020

Version of record online: October 15, 2020