Rapid assessment of ligand binding to the SARS-CoV-2 main protease by saturation transfer difference NMR spectroscopy

Anastassia L. Kantsadi and Ioannis Vakonakis[1]

From the Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

[1]To whom correspondence should be addressed, e-mail: ioannis.vakonakis@bioch.ox.ac.uk, Tel.: +44 1865 275725, Fax: +44 1865 613201

Short title: Assessment of ligand binding to SARS-CoV-2 M^{pro} by STD-NMR

Keywords: SARS-CoV-2, COVID-19, Mpro, NMR, STD, screening, fragments

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological cause of the coronavirus disease 2019, for which no effective therapeutics are available. The SARS-CoV-2 main protease (M^{pro}) is essential for viral replication and constitutes a promising therapeutic target. Recently, crystallographic fragment screening uncovered a large number of molecules binding at the M^{pro} active site as well as at sites of potential allostery. However, no information on fragment affinity is available. Here, we describe an efficient protocol for screening chemical compounds for binding to M^{pro} using saturation transfer difference nuclear magnetic resonance (STD-NMR) spectroscopy. We characterised the relative affinity of M^{pro}-binding fragments using STD-NMR, revealing a ~200-fold difference in binding potential. Combined with crystallographic structures of these fragments bound to M^{pro}, this information can assist ongoing drug design efforts.

Introduction

Infections by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have resulted in over 400,000 deaths as of early June 2020 (1) and led to the coronavirus 2019 (COVID-19) pandemic (2-4). SARS-CoV-2 is a zoonotic betacoronavirus highly similar to SARS-CoV-1 and MERS-CoV, which caused outbreaks in 2002 and 2012, respectively (5-7). Much of SARS-CoV-2 biology has been described thus far by analogy to previously studied coronaviruses, including SARC-CoV-1 and MERS-CoV. SARS-CoV-2 encodes its proteome in a single, positive-sense, linear RNA molecule of ~30 kb length, the majority of which (~21.5 kb) is translated into two polypeptides, pp1a and pp1ab, via ribosomal frame-shifting (8, 9). Key viral enzymes and factors, including most proteins of the reverse-transcriptase machinery, inhibitors of host translation and molecules signalling for host cell survival, are released from pp1a and pp1ab via post-translational cleavage by two viral cysteine proteases (10). These proteases, a papain-like enzyme cleaving pp1ab at three sites, and a 3C-like protease cleaving the polypeptide at 11 sites, are primary targets for the development of antiviral drugs.

The 3C-like protease of SARS-CoV-2, also known as the viral main protease (M^{pro}), has been the target of intense study owing to its centrality in viral replication. M^{pro} studies have benefited from previous structural analyses of the SARC-CoV-1 3C-like protease and the development of putative inhibitors (11-14). The active sites of these proteases are highly conserved, and at least one family of SARS-CoV-1 3C-like protease inhibitors, comprising peptidomimetic α-ketoamides, has been shown to be effective against M^{pro}, thereby offering a promising lead for antiviral development (15). However, to date no M^{pro}-targeting inhibitors have been validated in clinical trials. Recently, structures of M^{pro} in complex with 74 fragments were resolved as part of a high-throughput screening using X-ray crystallography (16). Hits included covalently and non-covalently bound molecules targeting the protease active site, as well as the M^{pro} dimerisation interface and surface epitopes. These structures offer opportunities for the development of potent compounds that would directly inhibit M^{pro} activity or allosterically regulate it, for example by disrupting M^{pro} dimerisation, which is known to reduce activity of coronavirus 3C-like proteases (17). However, crystallographic structures of protein – ligand complexes are silent on compound affinities, which is key information to select the most promising leads. Further, the combined high concentration of protein (solid crystals) and of compounds (50 mM typical) in crystallographic fragment screening assays raises the possibility that some very weak binders may have been selected, which could lead inhibitor development towards blind alleys.

Here, we show that saturation transfer difference nuclear magnetic resonance (STD-NMR) spectroscopy can rapidly assess compound binding to M^{pro}. We use this method to derive the relative

58 affinities of non-covalent interactors identified by crystallographic fragment screening. This
59 information can aid selection of the most promising fragment 'hits' for further development.

60

61

62 **Materials and Methods**

63 *Protein production and purification*

64 A SARS-CoV-2 M^{pro} genetic construct including the viral protease, a glutathione S-transferase N-
65 terminal tag, and a C-terminal His_6-tag (16) was transformed into *E. coli* strain Rosetta(DE3).
66 Transformed clones were pre-cultured at 37 °C in lysogeny broth supplemented with appropriate
67 antibiotics and grown to exponential phase for approximately 5 h. Starter cultures were used to
68 inoculate 1 L of Terrific Broth Autoinduction Media (Formedium) supplemented with 10% v/v glycerol
69 and appropriate antibiotics. Cell cultures were grown at 37 °C for 5 h and then cooled to 18 °C for 12
70 h. Bacterial cells were harvested by centrifugation at 5,000 x g for 15 min.

71 Cell pellets were resuspended in 50 mM trisaminomethane (Tris)-Cl pH 8, 300 mM NaCl, 10 mM
72 imidazole buffer, incubated with 0.05 mg/ml benzonase nuclease (Sigma Aldrich) and lysed by
73 sonication on ice. Lysates were clarified by centrifugation at 50,000 x g at 4 °C for 1 h. Lysate
74 supernatants were loaded onto a HiTrap Talon metal affinity column (GE Healthcare) pre-equilibrated
75 with lysis buffer. Column wash was performed with 50 mM Tris-Cl pH 8, 300 mM NaCl and 25 mM
76 imidazole, followed by protein elution using the same buffer and an imidazole gradient from 25 to 500
77 mM concentration. The His_6-tag was cleaved using home-made recombinant human rhinovirus (HRV)
78 3C protease. The HRV 3C protease, His_6-tag and further impurities were removed by a reverse HiTrap
79 Talon column. Flow-through fractions were concentrated and applied to a Superdex75 26/600 size
80 exclusion column (GE Healthcare) equilibrated in NMR buffer (see below).

81 *Nuclear magnetic resonance (NMR) spectroscopy*

82 All NMR experiments were performed using the Oxford Biochemistry 950 MHz (22.3 T) solution-
83 state instrument comprising an Oxford Instruments superconducting NMR magnet, Bruker Avance III
84 console and TCI probehead. A Bruker SampleJet sample changer was used for sample manipulation.
85 Experiments were performed and data processed using TopSpin (Bruker). Samples comprised 10 μM
86 M^{pro} alone, 0.8 mM of compounds alone, or 10 μM M^{pro} mixed with 0.8 mM compounds, formulated
87 in NMR buffer (150 mM NaCl, 20 mM Na_2HPO_4 pH 7.4) supplemented with 10% v/v D_2O and
88 deuterated dimethyl sulfoxide (D_6-DMSO, 99.96% D, Sigma Aldrich) to 5% v/v final D_6-DMSO
89 concentration. Sample volume was 140 μL and samples were loaded in 3 mm outer diameter
90 SampleJet NMR tubes (Bruker) placed in 96-tube racks. NMR tubes were sealed with POM balls.

91 STD-NMR experiments were performed at 10 °C using a pulse sequence described previously (18)
92 and an excitation sculpting water-suppression scheme (19). Protein signal were suppressed in STD-
93 NMR by the application of a 30 msec spin-lock pulse. We collected time-domain data of 16,384

complex points and 41.6 μsec dwell time (12.02 kHz sweepwidth). Data were collected in an interleaved pattern, with on- and off-resonance irradiation data separated into 16 blocks of 16 transients each (256 transients per irradiation frequency). Transient recycle delay was 4 sec and on- or off-resonance irradiation was performed using 0.1 mW of power for 3.5 sec at 0.5 ppm or 26 ppm, respectively, for a total experiment time of approximately 45 minutes. Reconstructed time-domain data from the difference of on- and off-resonance irradiation (STD spectra) or only the off-resonance irradiation (reference spectra) were processed by applying a 2 Hz exponential line broadening function prior to Fourier transformation. Phasing parameters were derived for each sample from the reference spectra and copied to the STD spectra. ^{1}H peak intensities were integrated in TopSpin using a local-baseline adjustment function. For the calculation of STD parameters and for estimating ligand concentrations we integrated ^{1}H peaks originating from ligand aromatic groups, as peak dispersion in this area of the spectra allowed accurate consideration of the number of underlying ^{1}H atoms. Ligand concentrations are reported relative to the most abundant compound.

Ligand binding was quantified by calculating the ratio of integrated signal for a peak or peaks in the ^{1}H STD spectrum over that of the reference spectrum (STD_{ratio}). From this the STD amplification factor (STD_{amp}) is derived as $STD_{amp} = STD_{ratio}\ x\ ligand\ excess$. STD_{amp} is inversely proportional to the ligand dissociation constant, K_d, as $STD_{amp} \propto \frac{[L]}{K_d+[L]}$, where [L] is the ligand concentration.

Notes

Compounds were procured from Enamine (Riga, Latvia) in the form of a plated library (DSI-poised), dissolved in D_6-DMSO at 500 mM nominal concentration. 1 μL of dissolved compounds was aspirated from the library and immediately mixed with 9 μL of D_6-DMSO for final fragment concentration of 50 mM, from which NMR samples were formulated. Crystallographic structures of M^{pro} with bound fragments were visualised using PyMOL (20) and correspond to Protein Data Bank entries 5R81 (ligand x0195), 5REC (x0390), 5R83 (x0434) and 5RGQ (x1086) (16).

119 **Results**

120 *STD-NMR assays of M^{pro} ligand binding*

121 M^{pro} forms dimers in crystals (15), likely with a sub-μM solution dissociation constant (K_d) based on
122 analogy to previously studied 3C-like coronavirus proteases (17). At the 10 μM protein concentration
123 of our NMR assays M^{pro} is thus expected to be dimeric with an estimated molecular weight of
124 approximately 68 kDa. Despite the relatively large size of M^{pro} for NMR, we were able to record one-
125 dimensional ^{1}H spectra of M^{pro} alone showing the presence of multiple up-field shifted (<0.5 ppm)
126 resonances, likely corresponding to methyl groups in the protein hydrophobic core, as well as down-
127 field shifted resonances (>8.5 ppm) originating from hydrogen-bonded amides (Fig. 1A). In addition to
128 demonstrating that M^{pro} is folded under the conditions tested, these spectra allowed us to identify
129 multiple resonance frequencies where irradiation for STD-NMR experiments might be applied in a
130 protein-specific manner.

131 As the vast majority of compounds assessed feature aromatic groups (see below), we chose to
132 irradiate M^{pro} at up-field shifted frequencies in STD-NMR experiments. Trials revealed that irradiation
133 at 0.5 ppm frequency produced unambiguous ^{1}H signals in STD spectra from ligands bound to M^{pro},
134 while simultaneously avoiding ligand excitation that would yield false-positive ^{1}H signals in the
135 absence of M^{pro} (Fig. 1B). Further, contributions of small molecules not binding to M^{pro}, such as DMSO,
136 to the STD spectra were pseudo-dispersive in nature, due to imperfect signal subtraction, while true
137 ligands produced peaks in STD with absorptive ^{1}H lineshapes. We conclude that STD-NMR provides an
138 efficient method for detecting ligand binding to M^{pro}.

139

140 *Relative affinities of M^{pro}-binding fragments*

141 To better evaluate the information provided by STD-NMR, we assessed using this method the
142 binding of compounds previously seen complexed to M^{pro} in crystallographic fragment screening
143 experiments (16). As STD-NMR signal relies on ligand exchange between protein-bound and -free
144 states, we focused on non-covalent M^{pro} interactors. 39 such compounds are part of the DSI-poised
145 fragment library to which we were given access, comprising 17 active site binders, two compounds
146 targeting the M^{pro} dimerisation interface and 20 molecules binding elsewhere on the protein surface.
147 We initially recorded STD-NMR experiments in the absence of M^{pro} to confirm that under these
148 conditions we obtained no or minimal STD signal from the ligands, and to verify ligand identity from
149 the reference ^{1}H spectra. Five ligands were either absent from our DSI-poised library or produced
150 reference ^{1}H spectra inconsistent with the compound chemical structure; these ligands were not

151 evaluated further. Samples of 10 μM M^{pro} and 0.8 mM nominal ligand concentration were formulated
152 from the remaining 34 compounds, and STD-NMR spectra were recorded. The folded state of M^{pro} in
153 the presence of each ligand was verified by collecting ^{1}H NMR spectra similar to that of Fig. 1A from
154 all samples ahead of STD-NMR experiments.

155 We observed large variation in ^{1}H STD signal intensity in the presence of M^{pro} across ligands, with
156 many ligands producing little or no STD signal, suggesting substantial differences in compound affinity
157 for the protease (Table 1). However, we also recorded up to ~15-fold differences in the per-^{1}H
158 integrated intensity of reference ligand spectra, indicating significant variation in ligand
159 concentrations in solution. Such concentration differences may arise from errors in sample
160 formulation or from concentration inconsistencies in the ligand library. To evaluate the former we
161 integrated the residual ^{1}H signal of deuterated DMSO in our reference spectra, which should have
162 been constant as DMSO was added alongside ligands in the samples, and found it to vary by less than
163 35% across any pair of samples (11% average deviation). Thus, although sample formulation did
164 introduce ligand concentration errors by up to ~1/3, the extend of these errors did not account for
165 the up to ~15-fold differences in relative ligand concentrations.

166 To factor in these concentration differences we calculated the STD amplification ratio (STD_{amp})
167 upon M^{pro} binding for all compounds, which is related to ligand affinity. Strikingly, M^{pro} binding varied
168 by ~200-fold across ligands (Fig. 2A), with the strongest interactions recorded for compounds x0434,
169 x0390, x1086 and x0195 (Table 1). Large differences in ligand affinity were also readily apparent by
170 incubating M^{pro} with mixtures of two different compounds, where STD signal for only one compound
171 was visible (Fig. 2B). We surmise that the relative affinities of small molecules towards M^{pro} can be
172 readily assessed by STD-NMR, and that M^{pro}-binding fragments identified by crystallography vary
173 substantially in their interaction affinities for this protease.

174

175

176

177

178 **Discussion**

179 Fragment-based screening is a tried and tested method for reducing the number of compounds
180 that need to be screened against a specific target in order to sample chemical space (21). Combined
181 with X-ray crystallography, which provides information on the target site and binding pose of ligands,
182 identified fragments are thought of as 'bricks' that, when combined, can yield potent and specifically-
183 interacting compounds. However, crystallographic structures do not report on ligand affinity, and
184 obtaining fragment affinities is often difficult by biophysical methods, such as calorimetry, due to the
185 weak nature of these interactions. However, NMR methods, such as STD-NMR, provide highly
186 powerful and sensitive tools for identifying ligand binding to macromolecules, even when interaction
187 affinities are in the mM K_d range (22). Here, we showed that STD-NMR is a suitable method for
188 characterising the binding of fragments to M^{pro}, an essential protease of SARC-CoV-2.

189 Despite its considerable molecular weight (~68 kDa dimer) and the low protein concentration (10
190 μM) of samples in our study, M^{pro} yielded ^{1}H NMR spectra of high quality and with strong signal/noise
191 in relatively little acquisition time (Fig. 1A). Importantly, we observed no protein precipitation during
192 sample preparation even in M^{pro} concentrations of hundreds of μM, and samples were stable over
193 time yielding highly reproducible protein ^{1}H spectra during the course of several days. These
194 observations suggest that this viral protease is amenable to further NMR studies, such as experiments
195 to characterise the M^{pro} dynamic behaviour during catalysis.

196 STD-NMR allowed us to screen for ligand binding to M^{pro} using relatively small amounts of protein
197 and in under one hour of experiment time per ligand. Ligands interacting strongly with M^{pro} yielded
198 high signal/noise STD data (Fig. 1B), which suggests that further improvements in our protocol are
199 possible. For example, we anticipate that halving the amount of M^{pro} used in STD-NMR experiments,
200 or reducing the experiment time to ¼ of its current duration, will still provide us with sufficient signal
201 to detect interacting compounds. Although we refrained from doing so in this study, as our goal was
202 to characterise the interaction of even weekly-binding M^{pro} ligands, such modifications should be
203 considered if STD-NMR was used in a high-throughput manner to identify compounds binding to this
204 protease.

205 Our work showed that fragments identified by previous crystallography-based assays as binding to
206 M^{pro} in a non-covalent manner (16) do so with a wide-range of relative affinities. Four ligands, x0195,
207 x0390, x0434 and x1086, were shown to be most promising on the basis of their M^{pro} interaction
208 strength in solution. Of these, both x0195 and x0434 bind at the M^{pro} active site (Fig. 2C). However,
209 the x0196 and x0434 binding poses overlap only in part, with x0434 being in contact with the catalytic
210 cysteine residue and x0195 in the active site periphery. Thus, a likely direction towards the synthesis

211 of a potent M^{pro} inhibitor would be combining these two fragments, thereby exploiting a larger
212 fraction of the active site interface.

213 Interestingly, both x0390 and x1086 bind at the M^{pro} surface at the junction between the protease
214 N-terminal catalytic subdomain and its C-terminal subdomain (Fig. 2D). Previous work showed that
215 coronavirus 3C-like proteases are extensively regulated by the C-terminal subdomain, likely as this
216 subdomain enhances protease dimerisation that, in turn, increases catalytic activity by several fold
217 (23). The x1086 binding site locates proximal to the interface between the two M^{pro} protomers; thus,
218 further development of this ligand holds potential for directly disrupting the M^{pro} dimer and reducing
219 protease activity. However, x0390-derivatives may also yield means towards the allosteric control of
220 M^{pro}, by altering the relative conformation of the catalytic and regulatory subdomains, thereby
221 impacting on protease function.

222 In conclusion, we presented here a protocol for the rapid assessment of SARS-CoV-2 M^{pro} ligands
223 using STD-NMR spectroscopy and evaluated the relative affinities of fragments seen to bind this
224 protease. Although development of novel antivirals to combat COVID-19 is still at an early stage, we
225 hope that this information will prove valuable to groups working towards such treatments.

226

227 **References**

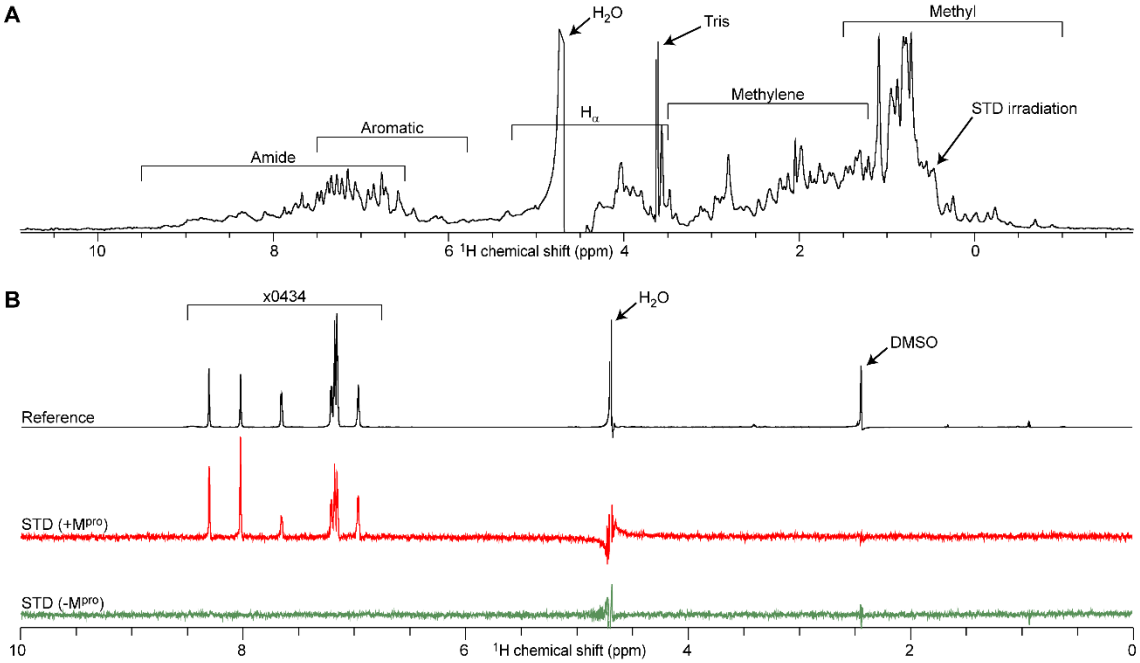
228 1. WHO. Coronavirus disease 2019 [Available from:
229 https://www.who.int/emergencies/diseases/novel-coronavirus-2019.
230 2. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of
231 transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis.
232 2020;20(5):553-8.
233 3. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with
234 human respiratory disease in China. Nature. 2020;579(7798):265-9.
235 4. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with
236 Pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33.
237 5. Bermingham A, Chand MA, Brown CS, Aarons E, Tong C, Langrish C, et al. Severe respiratory
238 illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle
239 East, September 2012. Euro Surveill. 2012;17(40):20290.
240 6. Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, van Amerongen G, van Riel D, et al.
241 Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet.
242 2003;362(9380):263-70.
243 7. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel
244 coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012;367(19):1814-20.
245 8. Thiel V, Ivanov KA, Putics A, Hertzig T, Schelle B, Bayer S, et al. Mechanisms and enzymes
246 involved in SARS coronavirus genome expression. J Gen Virol. 2003;84(Pt 9):2305-15.
247 9. Bredenbeek PJ, Pachuk CJ, Noten AF, Charite J, Luytjes W, Weiss SR, et al. The primary
248 structure and expression of the second open reading frame of the polymerase gene of the coronavirus
249 MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting
250 mechanism. Nucleic Acids Res. 1990;18(7):1825-32.
251 10. Hilgenfeld R. From SARS to MERS: crystallographic studies on coronaviral proteases enable
252 antiviral drug design. FEBS J. 2014;281(18):4085-96.
253 11. Ghosh AK, Xi K, Grum-Tokars V, Xu X, Ratia K, Fu W, et al. Structure-based design, synthesis,
254 and biological evaluation of peptidomimetic SARS-CoV 3CLpro inhibitors. Bioorg Med Chem Lett.
255 2007;17(21):5876-80.
256 12. Verschueren KH, Pumpor K, Anemuller S, Chen S, Mesters JR, Hilgenfeld R. A structural view
257 of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters. Chem Biol.
258 2008;15(6):597-606.
259 13. Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, et al. The crystal structures of severe acute
260 respiratory syndrome virus main protease and its complex with an inhibitor. Proc Natl Acad Sci U S A.
261 2003;100(23):13190-5.
262 14. Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, et al. Design of wide-spectrum inhibitors targeting
263 coronavirus main proteases. PLoS Biol. 2005;3(10):e324.
264 15. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-
265 2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science.
266 2020;368(6489):409-12.
267 16. Douangamath A, Fearon D, Gehrtz P, Krojer T, Lukacik P, Owen CD, et al. Crystallographic and
268 electrophilic fragment screening of the SARS-CoV-2 main protease. bioRxiv. 2020:2020.05.27.118117.
269 17. Grum-Tokars V, Ratia K, Begaye A, Baker SC, Mesecar AD. Evaluating the 3C-like protease
270 activity of SARS-Coronavirus: recommendations for standardized assays for drug discovery. Virus Res.
271 2008;133(1):63-73.
272 18. Mayer M, Meyer B. Characterization of Ligand Binding by Saturation Transfer Difference NMR
273 Spectroscopy. Angew Chem Int Ed Engl. 1999;38(12):1784-8.
274 19. Hwang TL, Shaka AJ. Water Suppression That Works - Excitation Sculpting Using Arbitrary
275 Wave-Forms and Pulsed-Field Gradients. J Magn Reson Ser A. 1995;112(2):275-9.

20. DeLano WL. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.2002.
21. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: the impact of fragments on drug discovery. Nat Rev Drug Discov. 2016;15(9):605-19.
22. Becker W, Bhattiprolu KC, Gubensak N, Zangger K. Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. Chemphyschem. 2018;19(8):895-906.
23. Shi J, Song J. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. FEBS J. 2006;273(5):1035-45.

286 **Acknowledgements**

287 We are grateful to Nick Soffe for maintenance of the Oxford Biochemistry solution NMR facility, to
288 Claire Strain- Damerell, Petra Lukacik and Martin A. Walsh for providing the M^{pro} expression construct
289 and advise on protein production, and to Anthony Aimon and Frank von Delft for providing the DSI-
290 poised fragment library. This work was supported by philanthropic donations to the University of
291 Oxford COVID-19 Research Response Fund and the Oxford Glycobiology Institute Endowment. The
292 Oxford Biochemistry NMR facility was supported by the Wellcome Trust (094872/Z/10/Z), the
293 Engineering and Physical Sciences Research Council (EP/R029849/1), the Wellcome Institutional
294 Strategic Support Fund, the EPA Cephalosporin Fund and the John Fell OUP Research Fund.
295

296 **Figures and captions**



297

298 *Figure 1: 1D and STD-NMR spectra of SARS-CoV-2 M^{pro}.* A) 1D ^{1}H NMR spectrum of recombinant
299 SARS-CoV-2 M^{pro}, recorded from a 10 μM protein concentration sample in a 5 mm NMR tube at 25 °C
300 using an excitation sculpting water-suppression method (19). 512 acquisitions with recycle delay of
301 1.25 sec were averaged, for a total experiment time of just over 10 min. Data were processed with a
302 quadratic Sine function prior to Fourier transformation. The broad areas where ^{1}H from protein
303 chemical groups resonate are indicated by brackets. The residual signals of H_2O and Tris, from the
304 protein purification, as well as the position where on-resonance irradiation was applied for STD
305 experiments, are indicated. B) Vertically offset ^{1}H STD-NMR spectra from ligand x0434 binding to M^{pro}.
306 The reference spectrum is shown on top (black) with the x0434, H_2O and DMSO ^{1}H resonances
307 indicated. The STD spectrum of x0434 in the presence of M^{pro} is shown in the middle (red), while that
308 in the absence of M^{pro} is at the bottom (green). STD spectra are scaled up 64x compared to the
309 reference spectrum.

310

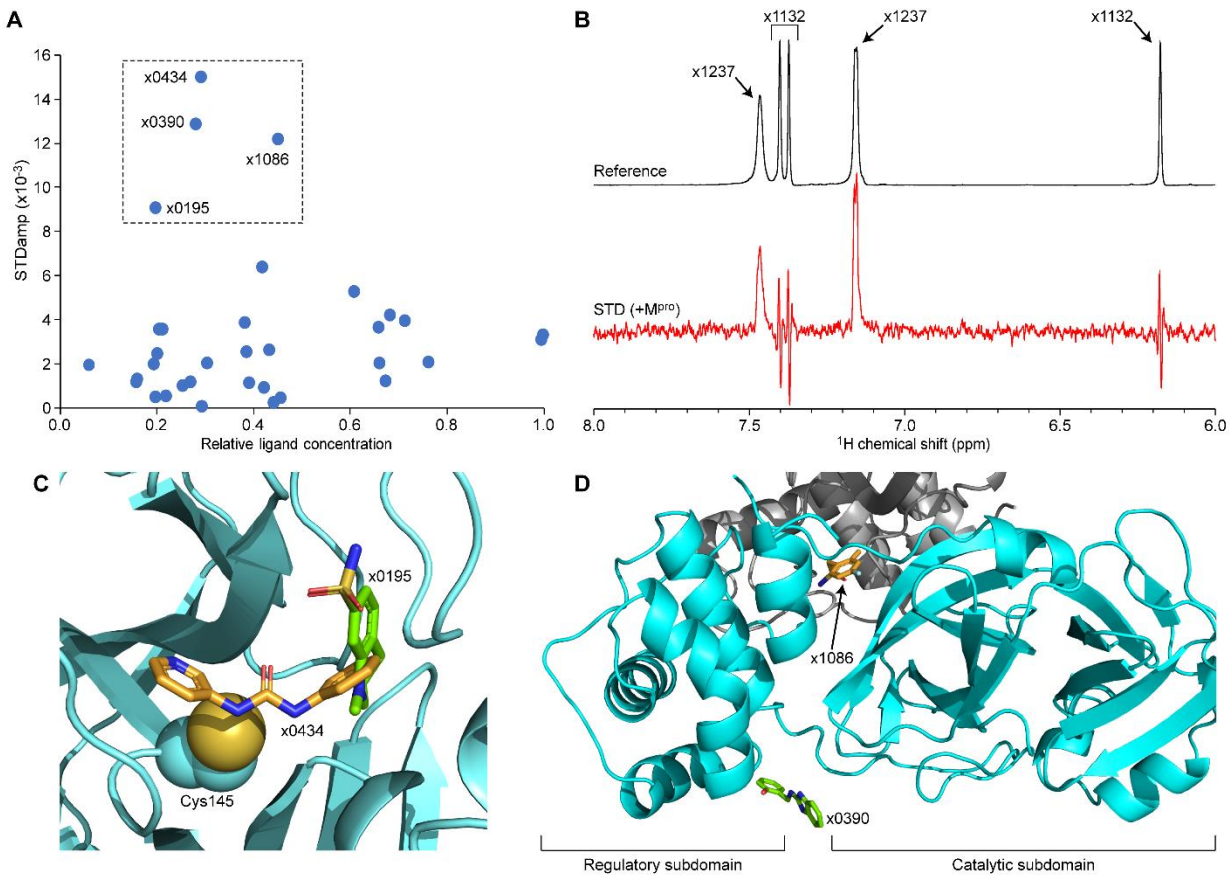
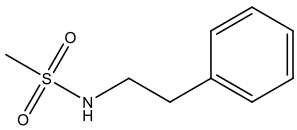
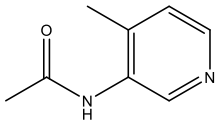
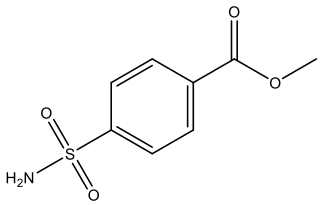
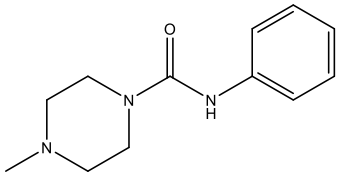
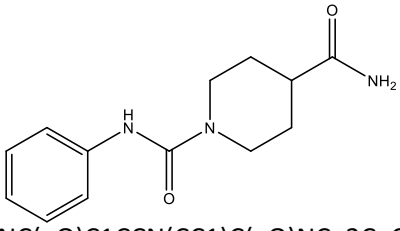
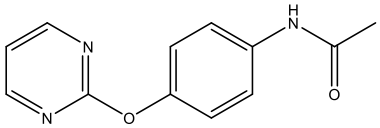
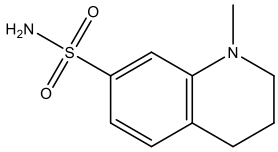
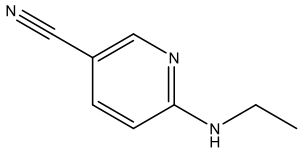
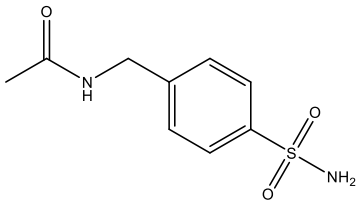
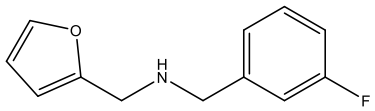
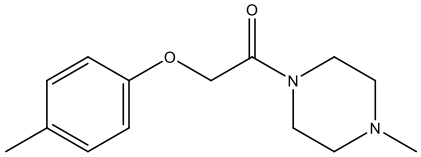
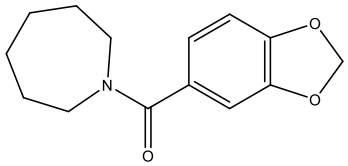
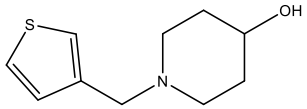
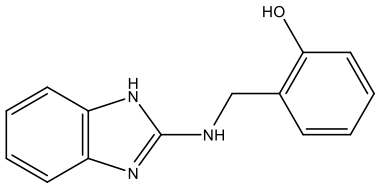
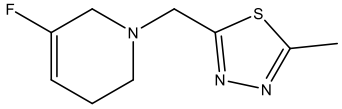


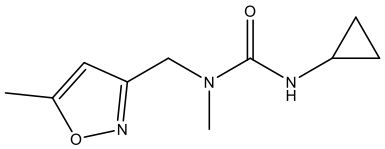
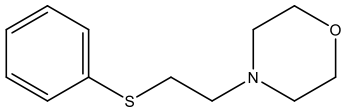
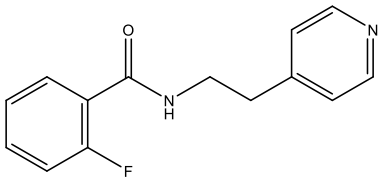
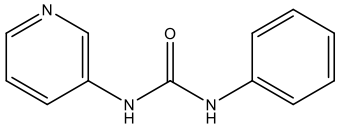
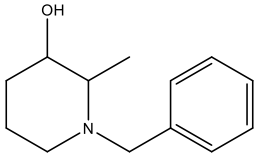
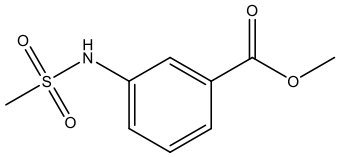
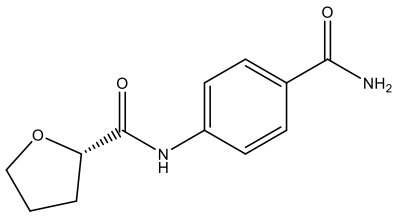
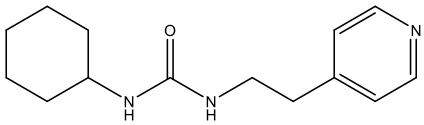
Figure 2: Assessment of non-covalent ligand binding to M^{pro}. A) Plot of the STD amplification factor (STD_{amp}) versus relative ligand concentration for non-covalent fragments seen as binding to M^{pro} in crystallographic screening (16). STD_{amp} is inversely proportional to K_d for a given ligand concentration. The strongest interacting ligands are boxed. B) Magnification of the aromatic ^1H area of STD-NMR spectra from a mixture of ligands (x1237 and x1132, each at 0.8 mM concentration) in the presence of M^{pro}. Both ligands interacted with the same M^{pro} site in crystallographic structures (16). Reference spectrum is shown on top (black) and STD spectrum at the bottom (red, scaled up by 64x). The resonances originating from each ligand are indicated. Whereas x1237 binds to M^{pro} under our solution conditions and gives rise to STD peaks with absorptive lineshapes, x1132 does not and only produces dispersive peak artefacts. C) Magnification of the M^{pro} active site and position of ligands x0434 and x0195 within it from crystallographic structures (16). The sidechain of the catalytic cysteine residue is shown as spheres; ligands are shown as sticks. D) View of a M^{pro} protomer (cyan) with the catalytic (N-terminal) and regulatory (C-terminal) subdomains indicated. The second M^{pro} protomer in the dimer is shown in grey. Ligands x0390 and x1086 (sticks) bind at the junction of the two protease subdomains, with x1086 facing the dimerisation interface.

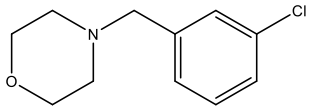
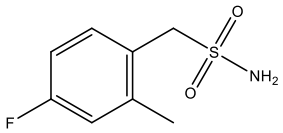
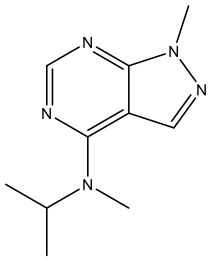
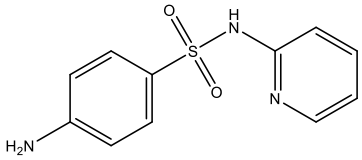
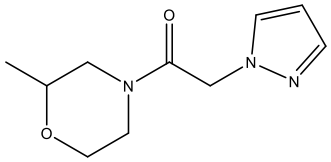
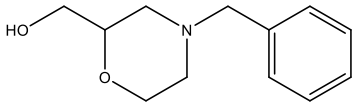
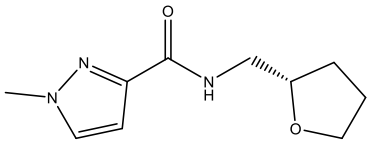
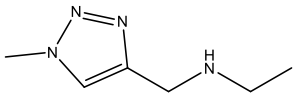
328 **Tables**

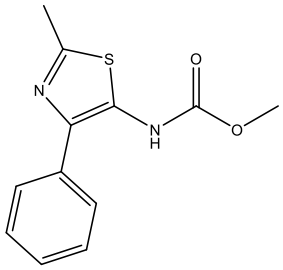
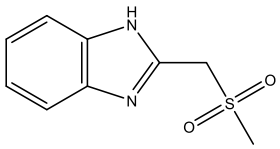
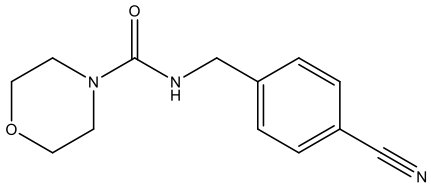
329 *Table 1: M^{pro}-binding fragments assessed by STD-NMR*

Ligand	M^{pro} site	SMILES and structure	Relative ligand concentration	STDamp ($x10^{-3}$)
x0072	Active	 CS(=O)(=O)NCCC=1C=CC=CC1	0.68	4.24
x0107	Active	 CC(=O)NC=1C=NC=CC1C	0.71	3.98
x0161	Active	 COC(=O)C=1C=CC(=CC1)S(=O)(=O)N	0.66	3.69
X0165	Surface	 CN1CCN(CC1)C(=O)NC=2C=CC=CC2	0.29	0.07
X0177	Surface	 NC(=O)C1CCN(CC1)C(=O)NC=2C=CC=CC2	0.27	1.19
x0194	Surface	 CC(=O)NC=1C=CC(OC=2N=CC=CN2)=CC1	0.06	1.97
x0195	Active	 CN1CCCC=2C=CC(=CC12)S(=O)(=O)N	0.20	9.09

x0305	Active	 CCNC=1C=CC(C#N)=CN1	0.61	5.29
X0336	Surface	 CC(=O)NCC=1C=CC(=CC1)S(=O)(=O)N	0.44	0.25
X0350	Surface	 FC=1C=CC=C(CNCC2=CC=CO2)C1	0.20	2.49
x0354	Active	 CN1CCN(CC1)C(=O)COC=2C=CC(C)=CC2	0.66	2.03
X0376	Surface	 O=C(N1CCCCCC1)C=2C=CC=3OCOC3C2	1.00	3.35
x0387	Active	 OC1CCN(CC=2C=CSC2)CC1	0.76	2.11
x0390	Surface	 OC=1C=CC=CC1CNC2=NC=3C=CC=CC3N2	0.28	12.86
x0395	Active	 CC1=NN=C(CN2CCC=C(F)C2)S1	0.99	3.13

x0397	Active	 CN(CC=1C=C(C)ON1)C(=O)NC2CC2	0.67	1.24
x0398	Surface	 C(CN1CCOCC1)SC=2C=CC=CC2	0.16	1.19
x0426	Active	 FC=1C=CC=CC1C(=O)NCCC=2C=CN=CC2	0.21	3.60
x0434	Active	 O=C(NC=1C=CC=CC1)NC=2C=CC=NC2	0.29	15.03
x0464	Surface	 CC1C(O)CCCN1CC=2C=CC=CC2	0.46	0.47
x0478	Surface	 COC(=O)C=1C=CC=C(NS(=O)(=O)C)C1	0.39	2.56
X0499	Surface	 NC(=O)C=1C=CC(NC(=O)[C@@H]2CCCO2)=C C1	0.20	0.52
x0540	Active	 O=C(NCCC=1C=CN=CC1)NC2CCCCC2	0.21	3.59

x0669	Surface	 ClC=1C=CC=C(CN2CCOCC2)C1	0.16	1.31
x1086	Dimer	 CC=1C=C(F)C=CC1CS(=O)(=O)N	0.45	12.18
x1101	Surface	 CC(C)N(C)C=1N=CN=C2N(C)N=CC12	0.22	0.57
x1119	Surface	 NC=1C=CC(=CC1)S(=O)(=O)NC=2C=CC=CN2	0.43	2.64
x1132	Surface	 CC1CN(CCO1)C(=O)CN2C=CC=N2	0.39	1.13
x1163	Surface	 OCC1CN(CC=2C=CC=CC2)CCO1	0.42	0.94
x1187	Dimer	 CN1C=CC(=N1)C(=O)NC[C@@H]2CCCO2	0.30	2.04
x1226	Surface	 CCNCC1=CN(C)N=N1	0.19	1.99

x1235	Surface	 COC(=O)NC=1SC(C)=NC1C=2C=CC=CC2	0.38	3.88
x1237	Surface	 CS(=O)(=O)CC1=NC=2C=CC=CC2N1	0.42	6.41
x1249	Active	 O=C(NCC=1C=CC(C#N)=CC1)N2CCOCC2	0.25	1.04