A COVID Moonshot: assessment of ligand binding to the SARS-CoV-2 main protease by saturation transfer difference NMR spectroscopy

Anastassia L. Kantsadi[1], Emma Cattermole[1], Minos-Timotheos Matsoukas[2], Georgios A. Spyroulias[2] and Ioannis Vakonakis[1]*

[1]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

[2]Department of Pharmacy, University of Patras, Panepistimioupoli Campus, GR-26504, Greece

*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, Moonshot, M^{pro}, NMR, STD, screening, fragments, molecular dynamics, MD, competition

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. Many efforts aimed at deriving effective M^{pro} inhibitors are currently underway, including an international open-science discovery project, codenamed COVID Moonshot. As part of COVID Moonshot, we used saturation transfer difference nuclear magnetic resonance (STD-NMR) spectroscopy to assess the binding of putative M^{pro} ligands to the viral protease, including molecules identified by crystallographic fragment screening and novel compounds designed as M^{pro} inhibitors. In this manner, we aimed to complement enzymatic activity assays of M^{pro} performed by other groups with information on ligand affinity. We have made the M^{pro} STD-NMR data publicly available. Here, we provide detailed information on the NMR protocols used and challenges faced, thereby placing these data into context. Our goal is to assist the interpretation of M^{pro} STD-NMR data, thereby accelerating ongoing drug design efforts.

Introduction

Infections by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) resulted in approximately 1.8 million deaths in 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 and MERS-CoV, which caused outbreaks in 2002 and 2012, respectively (5-7). 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 3C-like protease and the earlier development of putative inhibitors (11-14). The active sites of these proteases are highly conserved, and peptidomimetic inhibitors active against M^{pro} are also potent against the SARS-CoV 3C-like protease (15, 16). However, to date no M^{pro}-targeting inhibitors have been validated in clinical trials. In order to accelerate M^{pro} inhibitor development, an international, crowd-funded, open-science project was formed under the banner of COVID Moonshot (17), combining high-throughput crystallographic screening (18), computational chemistry, enzymatic activity assays and mass spectroscopy (19) among the many methodologies contributed by collaborating groups.

As part of COVID Moonshot, we utilised saturation transfer difference nuclear magnetic resonance (STD-NMR) spectroscopy (20-22) to investigate the M^{pro} binding of ligands initially identified by crystallographic screening, as well as molecules designed specifically as non-covalent inhibitors of this protease. Our goal was to provide orthogonal information on ligand binding to that which could be gained by enzymatic activity assays conducted in parallel by other groups. STD-NMR is a proven method for characterising the binding of small molecules to biological macromolecules, able to provide both quantitative affinity information and structural data on the proximity of ligand chemical groups to the protein. Here, we provide detailed documentation on the NMR protocols used to record these data and highlight the advantages, limitations and assumptions underpinning our approach. Our aim is to assist the comparison of M^{pro} STD-NMR data with other quantitative measurements, and facilitate the consideration of these data when designing future M^{pro} inhibitors.

65 **Materials and Methods**

66 *Protein production and purification*

67 We created a SARS-CoV-2 M^{pro} genetic construct in pFLOAT vector (23), encoding for the viral
68 protease and an N-terminal His_6-tag separated by a modified human rhinovirus (HRV) 3C protease
69 recognition site, designed to reconstitute a native M^{pro} N-terminus upon HRV 3C cleavage. The M^{pro}
70 construct was transformed into *Escherichia coli* strain Rosetta(DE3) (Novagen) and transformed
71 clones were pre-cultured at 37 °C for 5 h in lysogeny broth supplemented with appropriate
72 antibiotics. Starter cultures were used to inoculate 1 L of Terrific Broth Autoinduction Media
73 (Formedium) supplemented with 10% v/v glycerol and appropriate antibiotics. Cell cultures were
74 grown at 37 °C for 5 h and then cooled to 18 °C for 12 h. Bacterial cells were harvested by
75 centrifugation at 5,000 x g for 15 min.

76 Cell pellets were resuspended in 50 mM trisaminomethane (Tris)-Cl pH 8, 300 mM NaCl, 10 mM
77 imidazole buffer, incubated with 0.05 mg/ml benzonase nuclease (Sigma Aldrich) and lysed by
78 sonication on ice. Lysates were clarified by centrifugation at 50,000 x g at 4 °C for 1 h. Lysate
79 supernatants were loaded onto a HiTrap Talon metal affinity column (GE Healthcare) pre-
80 equilibrated with lysis buffer. Column wash was performed with 50 mM Tris-Cl pH 8, 300 mM NaCl
81 and 25 mM imidazole, followed by protein elution using the same buffer and an imidazole gradient
82 from 25 to 500 mM concentration. The His_6-tag was cleaved using home-made HRV 3C protease. The
83 HRV 3C protease, His_6-tag and further impurities were removed by a reverse HiTrap Talon column.
84 Flow-through fractions were concentrated and applied to a Superdex75 26/600 size exclusion
85 column (GE Healthcare) equilibrated in NMR buffer (150 mM NaCl, 20 mM Na_2HPO_4 pH 7.4).

86

87 *Nuclear magnetic resonance (NMR) spectroscopy*

88 All NMR experiments were performed using a 950 MHz solution-state instrument comprising an
89 Oxford Instruments superconducting magnet, Bruker Avance III console and TCI probehead. A Bruker
90 SampleJet sample changer was used for sample manipulation. Experiments were performed and
91 data processed using TopSpin (Bruker). For direct STD-NMR measurements, samples comprised 10
92 µM M^{pro} and variable concentrations (20 µM – 4 mM) of ligand compounds formulated in NMR
93 buffer supplemented with 10% v/v D_2O and deuterated dimethyl sulfoxide (D_6-DMSO, 99.96% D,
94 Sigma Aldrich) to 5% v/v final D_6-DMSO concentration. In competition experiments, samples
95 comprised 2 µM M^{pro}, 0.8 mM of ligand x0434 and variable concentrations (0 – 20 µM) of competing
96 compound in NMR buffer supplemented with D_2O and D_6-DMSO as above. Sample volume was 140

97 μL and samples were loaded in 3 mm outer diameter SampleJet NMR tubes (Bruker) placed in 96-
98 tube racks. NMR tubes were sealed with POM balls.

99 STD-NMR experiments were performed at 10 °C using a pulse sequence described previously (20)
100 and an excitation sculpting water-suppression scheme (24). Protein signals were suppressed in STD-
101 NMR by the application of a 30 msec spin-lock pulse. We collected time-domain data of 16,384
102 complex points and 41.6 μsec dwell time (12.02 kHz sweepwidth). Data were collected in an
103 interleaved pattern, with on- and off-resonance irradiation data separated into 16 blocks of 16
104 transients each (256 total transients per irradiation frequency). Transient recycle delay was 4 sec and
105 on- or off-resonance irradiation was performed using 0.1 mW of power for 3.5 sec at 0.5 ppm or 26
106 ppm, respectively, for a total experiment time of approximately 50 minutes. Reconstructed time-
107 domain data from the difference of on- and off-resonance irradiation (STD spectra) or only the off-
108 resonance irradiation (reference spectra) were processed by applying a 2 Hz exponential line
109 broadening function and 2-fold zero-filling prior to Fourier transformation. Phasing parameters were
110 derived for each sample from the reference spectra and copied to the STD spectra. ^{1}H peak
111 intensities were integrated in TopSpin using a local-baseline adjustment function. Data fitting to
112 extract K_d values were performed in OriginPro (OriginLab). The folded state of M^{pro} in the presence
113 of each ligand was verified by collecting ^{1}H NMR spectra similar to Fig. 1A from all samples ahead of
114 STD-NMR experiments.

115

116 *Ligand handling*

117 Compounds for the initial STD-NMR assessment of crystallographic fragment binding to M^{pro} were
118 provided by the XChem group at Diamond Light Source in the form of a 384-well plated library (DSI-
119 poised, Enamine), with compounds dissolved in D_6-DMSO at 500 mM nominal concentration. 1 μL of
120 dissolved compounds was aspirated from this library and immediately mixed with 9 μL of D_6-DMSO
121 for a final fragment concentration of 50 mM, from which NMR samples were formulated. For
122 titrations of the same crystallographic fragments compounds were procured directly from Enamine
123 in the form of lyophilized powder, which was dissolved in D_6-DMSO to derive compound stocks at 10
124 mM and 100 mM concentrations for NMR sample formulation.

125 STD-NMR assays of bespoke M^{pro} ligands used compounds commercially synthesised for COVID
126 Moonshot. These ligands were provided to us by the XChem group in 96-well plates, containing 0.7
127 μL of 20 mM D_6-DMSO-disolved compound per well. Plates were created using an Echo liquid
128 handling robot (Labcyte) and immediately sealed and frozen at -20 °C. For use, ligand plates were

129 thoroughly defrosted at room temperature and spun at 3,500 *g* for 5 minutes. In single-
130 concentration STD-NMR experiments, 140 µL of a pre-formulated mixture of M^{pro} and NMR buffer
131 with D_2O and D_6-DMSO were added to each well to create the final NMR sample. For STD-NMR
132 competition experiments, 0.5 µL of ligands were aspirated from the plates and immediately mixed
133 with 19.5 µL of D_6-DMSO for final ligand concentration of 0.5 mM from which NMR samples were
134 formulated.

135

136 *Molecular dynamics (MD) simulations*

137 The monomeric complexes of M^{pro} bound to chemical fragments were obtained from the RCSB
138 Protein Data Bank entries 5R81 (ligand x0195), 5REB (x0387), 5RGI (x0397), 5RGK (x0426), 5R83
139 (x0434) and 5REH (x0540) for MD simulations with GROMACS version 2018 (25) and the
140 AMBER99SB-ILDN force field (26). All complexes were inserted in a pre-equilibrated box containing
141 water implemented using the TIP3P water model (26). Force field parameters for the six ligands
142 were generated using the general Amber force field and HF/6 – 31G*– derived RESP atomic charges
143 (27). The reference system consisted of the protein, the ligand, ~31,400 water molecules, 95 Na and
144 95 Cl ions in a 100 x 100 x 100 Å simulation box, resulting in a total number of ~98,000 atoms. Each
145 system was energy-minimized and subsequently subjected to a 20 ns MD equilibration, with an
146 isothermal-isobaric ensemble using isotropic pressure control (28), and positional restraints on
147 protein and ligand coordinates. The resulting equilibrated systems were replicated 4 times and
148 independent 200 ns MD trajectories were produced with a time step of 2 fs, in constant temperature
149 of 300 K, using separate v-rescale thermostats (28) for the protein, ligand and solvent molecules.
150 Lennard-Jones interactions were computed using a cut-off of 10 Å and electrostatic interactions were
151 treated using particle mesh Ewald (29) with the same real-space cut-off. Analysis on the resulting
152 trajectories was performed using MDAnalysis (30, 31). Structures were visualised using PyMOL (32).

153

154 *Notes*

155 The enzymatic inhibition potential of M^{pro} ligands, measured by RapidFire mass spectroscopy
156 (17), was retrieved from the Collaborative Drug Discovery database (33).

157

158 **Results**

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

160 M^{pro} forms dimers in crystals via an extensive interaction interface involving two domains (15).
161 M^{pro} dimers likely have a sub-μM solution dissociation constant (K_d) by analogy to previously studied
162 3C-like coronavirus proteases (34). At the 10 μM protein concentration of our NMR assays M^{pro} is,
163 thus, expected to be dimeric with an estimated molecular weight of nearly 70 kDa. Despite the
164 relatively large size of M^{pro} for solution NMR, ^{1}H spectra of the protease readily showed the presence
165 of multiple up-field shifted (<0.5 ppm) peaks corresponding to protein methyl groups (Fig. 1A). In
166 addition to demonstrating that M^{pro} is folded under the conditions tested, these spectra allowed us
167 to identify the chemical shifts of M^{pro} methyl groups that may be suitable for on-resonance
168 irradiation in STD-NMR experiments. Trials with on-resonance irradiation applied to different methyl
169 group peaks showed that irradiating at 0.5 ppm (Fig. 1A) produced the strongest STD signal from
170 ligands in the presence of M^{pro}, while simultaneously avoiding ligand excitation that would yield
171 false-positive signals in the absence of M^{pro} (Fig. 1B). Further, we noted that small molecules
172 abundant in the samples but not binding specifically to M^{pro}, such as DMSO, produced pseudo-
173 dispersive residual signal lineshapes in STD spectra, while true M^{pro} ligands produced peaks in STD
174 with absorptive ^{1}H lineshapes. We surmised that STD-NMR is suitable for screening ligand binding to
175 M^{pro}, requiring relatively small amounts (10-50 μgr) of protein and time (under 1 hour) per sample
176 studied.

177 The strength of STD signal is quantified by calculating the ratio of integrated signal intensity of
178 peaks in the STD spectrum over that of the reference spectrum (STD_{ratio}). The STD_{ratio} factor is
179 inversely proportional to ligand K_d, as $STD_{ratio} \propto \frac{1}{K_d+[L]}$ where [L] is ligand concentration.
180 Measuring STD_{ratio} values over a range of ligand concentrations allows fitting of the proportionality
181 constant and calculation of ligand K_d. However, time and sample-amount considerations, including
182 the limited availability of bespoke compounds synthesized for the COVID Moonshot project, made
183 recording full STD-NMR titrations impractical for screening hundreds of ligands. Thus, we evaluated
184 whether measuring the STD_{ratio} value at a single ligand concentration may be an informative
185 alternative to K_d, provided restraints could be placed, for example, on the proportionality constant.

186 Theoretical and practical considerations suggested that three parameters influence our
187 evaluation of single-concentration STD_{ratio} values towards an affinity context. Firstly, the STD_{ratio}
188 factor is affected by the efficiency of NOE magnetisation transfer between protein and ligand, which
189 in turn depends on the proximity of ligand and protein groups, and the chemical nature of these

190 groups (20-22). To minimize the influence of these factors across diverse ligands, we sought to
191 quantify the STD_{ratio} of only aromatic ligand groups, and only consider those showing the strongest
192 STD signal; thus, that are in closest proximity to the protein. Second, STD-NMR assays require ligand
193 exchange between protein-bound and -free states in the timeframe of the experiment; strongly
194 bound compounds that dissociate very slowly from the protein would yield reduced STD_{ratio} values
195 compared to weaker ligands that dissociate more readily. Structures of M^{pro} with many different
196 ligands show that the protein conformation does not change upon complex formation and that the
197 active site is fully solvent-exposed (18), which suggests that ligand association can proceed with high
198 rate ($10^7 – 10^8$ $M^{-1}s^{-1}$). Under this assumption, the ligand dissociation rate is the primary determinant
199 of interaction strength. Given the duration of the STD-NMR experiment in our assays, and the ratios
200 of ligand:protein used, we estimated that significant protein – ligand exchange will take place even
201 for interactions as strong as low-µM K_d. Finally, uncertainties or errors in nominal ligand
202 concentration skew the correlation of STD_{ratio} to compound affinities; as shown in Fig. S1, STD_{ratio}
203 values increase strongly when very small amounts of ligands are assessed. Thus, overly large STD_{ratio}
204 values may be measured if ligand concentrations are significantly lower than anticipated.

205

206 *Quantitating M^{pro} binding of ligands identified by crystallographic screening*

207 Mindful of the limitations inherent to measuring single-concentration STD_{ratio} values, and prior to
208 using STD-NMR to evaluate bespoke M^{pro} ligands, we used this method to assess binding to the
209 protease of small chemical fragments identified in crystallographic screening experiments (18). In
210 crystallographic screening campaigns of other target proteins such fragments were seen to have
211 very weak affinities (> 1 mM K_d, e.g. (35)), thereby satisfying the exchange criterion set out above. 39
212 non-covalent M^{pro} interactors are part of the DSI-poised fragment library to which we were given
213 access, comprising 17 active site binders, two compounds targeting the M^{pro} dimerisation interface
214 and 20 molecules binding elsewhere on the protein surface (18). We initially recorded STD-NMR
215 spectra from these compounds in the absence of M^{pro} to confirm that we obtained no or minimal
216 STD signal when protease is omitted, and to verify ligand identity from reference ^1H spectra. Five
217 ligands gave no solution NMR signal or produced reference ^1H spectra inconsistent with the
218 compound chemical structure; these ligands were not evaluated further. Samples of 10 µM M^{pro} and
219 0.8 mM nominal ligand concentration were then formulated from the remaining 34 compounds
220 (Table S1), and STD-NMR spectra were recorded, from which only aromatic ligand STD signals were
221 considered for further analysis.

222 We observed large variations in STD signal intensity and STD_{ratio} values in the presence of M^{pro}
223 across compounds (Fig. 2A,B; Table S1), with many ligands producing little or no STD signal,
224 suggesting substantial differences in compound affinity for the protease. However, we also noted
225 that ligand reference spectra different substantially in intensity (Fig. 2C), despite compounds being
226 at the same nominal concentration. Integrating ligand peaks in these reference spectra revealed
227 differences in per-^{1}H intensity of up to ~15-fold, indicating significant variation of ligand
228 concentrations in solution (Table S1). Such concentration differences could arise from errors in
229 sample formulation or from concentration inconsistencies in the compound library. To evaluate the
230 former we also integrated the residual ^{1}H signal of D_6-DMSO in our reference spectra, and found it to
231 vary by less than 35% across any pair of samples (11% average deviation). As DMSO was added
232 alongside ligands in our samples, we concluded that sample formulation may have contributed
233 errors in compound concentration of up to ~1/3, but did not account for the ~15-fold differences in
234 concentration observed.

235 Given that differences in compound concentration can skew the relative STD_{ratio} values of ligands
236 (Fig. S1), and that such concentration differences were also observed among newly designed M^{pro}
237 inhibitors (see below), we questioned whether recording STD_{ratio} values under these conditions can
238 provide useful information. To address this question we attempted to quantify the affinity of
239 crystallographic fragments to M^{pro}, selecting ligands that showed clear differences in STD_{ratio} values
240 in the assays above and focusing on compounds binding at the M^{pro} active site; hence, that are of
241 potential interest to inhibitor development. We performed M^{pro} binding titrations monitored by STD-
242 NMR of compounds x0195, x0354, x0426 and x0434 in 50 μM – 4 mM concentrations (Fig. S2), and
243 noted that only compounds x0434 and x0195, which show the highest STD_{ratio} (Fig. 2A), bound
244 strongly enough for an affinity constant to be estimated (K_d of 1.6 ± 0.2 mM and 1.7 ± 0.2 mM,
245 respectively). In contrast, the titrations of x0354 and x0426, which yielded lower STD_{ratio} values,
246 could not be fit to extract a K_d indicating weaker binding to M^{pro}.

247 To further this analysis, we assessed the binding of fragments x0195, x0387, x0397, x0426, x0434
248 and x0540 to the M^{pro} active site using quadruplicate atomistic molecular dynamics (MD) simulations
249 of 200 nsec duration. As shown in Fig. S3A,B, and Movies S1 and S2, fragments with high STD_{radio}
250 values (x0434 and x0195) always located in the M^{pro} active site despite exchanging between
251 different binding conformations (Fig. S4), with average ligand root-mean-square-deviation (RMSD) of
252 3.2 Å and 5.1 Å respectively after the first 100 nsec of simulation. Medium STD_{ratio} value fragments
253 (x0426 and x0540, Fig. S3C,D, and Movies S3 and S4) show average RMSDs of approximately 9 Å in
254 the same simulation timeframe, frequently exchanging to alternative binding poses and with x0540
255 occasionally exiting the M^{pro} active site. In contrast, fragments showing very little STD NMR signal

256 (x0397 and x0387, Fig. S3E,F, and Movies S5 and S6) regularly exit the M^{pro} active site and show
257 average RMSDs in excess of 15 Å with very limited stability. Combining the quantitative K_d and MD
258 information above, we surmised that, despite limitations inherent in this type of analysis and
259 uncertainties in ligand amounts, STD_{ratio} values recorded at single compound concentration can act
260 as proxy measurements of M^{pro} affinity for ligands.

261

262 *Assessment of M^{pro} binding by COVID Moonshot ligands*

263 We proceeded to characterise by STD-NMR the M^{pro} binding of bespoke ligands created as part of
264 the COVID Moonshot project and designed to act as non-covalent inhibitors of the protease (17).
265 Similar to the assays of crystallographic fragments above, we focused our analysis of STD signals to
266 aromatic moieties of ligands binding to the M^{pro} active side and extracted STD_{ratio} values only from
267 the strongest STD peaks. Once again, we noted substantial differences in apparent compound
268 concentrations, judging from reference ^1H spectral intensities (Fig. 3A), which could not be
269 attributed to errors in sample preparation as the standard deviation of residual ^1H intensity in the
270 D_6-DMSO peak did not exceed 5% in any of the ligand batches tested. Crucially, out of 650 different
271 molecules tested, samples of 35 compounds (7.6%) contained no ligand and 86 (13.2%) very little
272 ligand (Fig. 3A). In these cases, NMR assays were repeated using a separate batch of compound;
273 however, 96.2% of repeat experiments yielded the same outcome of no or very little ligand in the
274 NMR samples.

275 We measured STD_{ratio} values from samples were ligands produced sufficiently strong reference ^1H
276 NMR spectra to be readily visible, and deposited these values and associated raw NMR data to the
277 Collaborative Drug Discovery database (33). Some of these ligands were assessed independently for
278 enzymatic inhibition of M^{pro} using a mass spectroscopy method as part of the COVID Moonshot
279 collaboration (17). Where both parameters are available, we compared the STD_{ratio} values and 50%
280 inhibition concentrations (IC_{50}) of these ligands. As shown in Fig. 3B, STD_{ratio} and IC_{50} values show
281 weak correlation (R^2=30%) for most ligands tested; however, a subset of ligands displayed
282 conspicuously low or even no STD signals considering their effect on M^{pro} activity, and presented
283 themselves as outliers in the correlation graph. As these outlier ligands had IC_{50} values below 10 μM,
284 suggesting that their affinities to the protease may be in the μM K_d region, we considered whether
285 our approach gives rise to false-negative STD results, for example through slow ligand dissociation
286 from M^{pro}.

287 To address this question, we derived an assay whereby the bespoke, high-affinity M^{pro} inhibitor
288 would outcompete a lower-affinity ligand known to provide strong STD signal from the protease
289 active site. In these experiments the lower-affinity ligand would act as 'spy' molecule whose STD
290 signal reduces as function of inhibitor concentration. We used fragment x0434, which yields
291 substantial STD signal with M^{pro} (Fig. 1B and 2A), as 'spy', and tested protease inhibitors EDJ-MED-
292 a364e151-1, LON-WEI-ff7b210a-5, CHO-MSK-6e55470f-14 and LOR-NOR-30067bb9-11 as x0434
293 competitors. Of these inhibitors, EDJ-MED-a364e151-1 gave rise to substantial STD signal in earlier
294 assays, whereas the remaining produced little or no STD signal; yet, all four inhibitors were reported
295 to have low-μM or sub-μM IC_{50} values based on M^{pro} enzymatic assays. In these competition
296 experiments, both EDJ-MED-a364e151-1 and LON-WEI-ff7b210a-5 yielded K_d parameters
297 comparable to the reported IC_{50} values (Fig. S5A,B), showing that at least in the case of LON-WEI-
298 ff7b210a-5 the absence of STD signal in the single-concentration NMR assays above represented a
299 false-negative result. In contrast, CHO-MSK-6e55470f-14 and LOR-NOR-30067bb9-11 were unable to
300 compete x0434 from the protease active site (Fig. S5C,D), suggesting that in these two cases the
301 reported IC_{50} values do not reflect inhibitor binding to the protease, and that the weak STD signal of
302 the initial assays was a better proxy of affinity. We surmised that although some low STD_{ratio} values
303 of M^{pro} inhibitors may not accurately reflect compound affinity to the protease, such values cannot
304 be discounted as a whole as they may correspond to non-binding ligands.

305

306

307

308

309

310

311

312

313 **Discussion**

314 Fragment-based screening is a tried and tested method for reducing the number of compounds
315 that need to be assessed for binding against a specific target in order to sample chemical space (36).
316 Combined with X-ray crystallography, which provides information on the target site and binding
317 pose of ligands, initial fragments can quickly be iterated into potent and specifically-interacting
318 compounds. The COVID Moonshot collaboration (17) took advantage of crystallographic fragment-
319 based screening (18) to initiate the design of novel inhibitors targeting the essential main protease
320 of the SARS-CoV-2 coronavirus; however crystallographic structures do not report on ligand affinity
321 and inhibitory potency in enzymatic assays does not always correlate with ligand binding. Thus,
322 supplementing these methods with solution NMR tools highly sensitive to ligand binding can provide
323 a powerful combination of orthogonal information and assurance against false starts.

324 We showed that STD-NMR is a suitable method for characterising ligand binding to M^{pro}, allowing
325 us to assess ligand interactions using relatively small amounts of protein and in under one hour of
326 experiment time per ligand (Fig. 1B). However, screening compounds in a high-throughput manner is
327 not compatible with the time- and ligand-amount requirements of full STD-NMR titrations. Thus, we
328 resorted to using an unconventional metric, the single-concentration STD_{ratio} value, as proxy for
329 ligand affinity. Although this metric has limitations due to its dependency on magnetisation transfer
330 between protein and ligand, and on relatively rapid exchange between the ligand-free and -bound
331 states, we demonstrated that it can nevertheless be informative. Specifically, the relative STD_{ratio}
332 values of chemical fragments bound to the M^{pro} active site provided insight on fragment affinity (Fig.
333 2A), as crosschecked by quantitative titrations (Fig. S2) and MD simulations (Fig. S3). Furthermore,
334 STD_{ratio} values of COVID Moonshot compounds held a weak correlation to enzymatic IC_{50} parameters
335 (Fig. 3B), although false-negative and -positive results from both methods contribute to multiple
336 outliers. Thus, in our view the biggest limitation of using the single-concentration STD_{ratio} value as
337 metric relates to its supra-linear sensitivity to ligand concentration (Fig. S1), which as demonstrated
338 here can vary substantially across ligands in a large project (Fig. 3A).

339 How then should the STD data recorded as part of COVID Moonshot be used? Firstly, we showed
340 that at least for some bespoke M^{pro} ligands the STD_{ratio} value obtained is a better proxy for
341 compound affinity compared to IC_{50} parameters from enzymatic assays (Fig. S5). This, inherently, is
342 the value of employing orthogonal methods thereby minimizing the number of potential false
343 results. Thus, when one is considering existing M^{pro} ligands to base the design of future inhibitors, a
344 high STD_{ratio} value as well as low IC_{50} parameters are both desirable. Second, due to the
345 aforementioned limitations of single-concentration STD_{ratio} value as proxy of affinity, and the

346 influence of uncertainties in ligand concentrations, we believe that comparisons of compounds and
347 derivatives differing by less than ~50% in STD_{ratio} is not meaningful. Rather, we propose that the
348 STD_{ratio} values of M^{pro} ligands measured and available at the CDD database should be treated as a
349 qualitative metrics of compound affinity.

350 In conclusion, we presented here protocols for the assessment of SARS-CoV-2 M^{pro} ligands using
351 STD-NMR spectroscopy, and evaluated the relative qualitative affinities of chemical fragments and
352 compounds designed as part of COVID Moonshot. Although development of novel antivirals to
353 combat COVID-19 is still at an early stage, we hope that this information will prove valuable to
354 groups working towards such treatments.

355

356 **References**

357 1. WHO. Coronavirus disease 2019 [Available from:
358 https://www.who.int/emergencies/diseases/novel-coronavirus-2019.
359 2. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of
360 transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis.
361 2020;20(5):553-8.
362 3. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with
363 human respiratory disease in China. Nature. 2020;579(7798):265-9.
364 4. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with
365 Pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33.
366 5. Bermingham A, Chand MA, Brown CS, Aarons E, Tong C, Langrish C, et al. Severe respiratory
367 illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the
368 Middle East, September 2012. Euro Surveill. 2012;17(40):20290.
369 6. Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, van Amerongen G, van Riel D, et al.
370 Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet.
371 2003;362(9380):263-70.
372 7. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel
373 coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012;367(19):1814-20.
374 8. Thiel V, Ivanov KA, Putics A, Hertzig T, Schelle B, Bayer S, et al. Mechanisms and enzymes
375 involved in SARS coronavirus genome expression. J Gen Virol. 2003;84(Pt 9):2305-15.
376 9. Bredenbeek PJ, Pachuk CJ, Noten AF, Charite J, Luytjes W, Weiss SR, et al. The primary
377 structure and expression of the second open reading frame of the polymerase gene of the
378 coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal
379 frameshifting mechanism. Nucleic Acids Res. 1990;18(7):1825-32.
380 10. Hilgenfeld R. From SARS to MERS: crystallographic studies on coronaviral proteases enable
381 antiviral drug design. FEBS J. 2014;281(18):4085-96.
382 11. Ghosh AK, Xi K, Grum-Tokars V, Xu X, Ratia K, Fu W, et al. Structure-based design, synthesis,
383 and biological evaluation of peptidomimetic SARS-CoV 3CLpro inhibitors. Bioorg Med Chem Lett.
384 2007;17(21):5876-80.
385 12. Verschueren KH, Pumpor K, Anemuller S, Chen S, Mesters JR, Hilgenfeld R. A structural view
386 of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters. Chem Biol.
387 2008;15(6):597-606.
388 13. Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, et al. The crystal structures of severe acute
389 respiratory syndrome virus main protease and its complex with an inhibitor. Proc Natl Acad Sci U S A.
390 2003;100(23):13190-5.
391 14. Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, et al. Design of wide-spectrum inhibitors
392 targeting coronavirus main proteases. PLoS Biol. 2005;3(10):e324.
393 15. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-
394 2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science.
395 2020;368(6489):409-12.
396 16. Rut W, Groborz K, Zhang L, Sun X, Zmudzinski M, Pawlik B, et al. SARS-CoV-2 M(pro)
397 inhibitors and activity-based probes for patient-sample imaging. Nat Chem Biol. 2020.
398 17. , Achdout H, Aimon A, Bar-David E, Barr H, Ben-Shmuel A, et al. COVID Moonshot: Open
399 Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-
400 Throughput Experiments, Computational Simulations, and Machine Learning. bioRxiv. 2020.
401 18. Douangamath A, Fearon D, Gehrtz P, Krojer T, Lukacik P, Owen CD, et al. Crystallographic
402 and electrophilic fragment screening of the SARS-CoV-2 main protease. Nat Commun.
403 2020;11(1):5047.

404 19. El-Baba TJ, Lutomski CA, Kantsadi AL, Malla TR, John T, Mikhailov V, et al. Allosteric Inhibition
405 of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays. Angew Chem Int
406 Edit. 2020.
407 20. Mayer M, Meyer B. Characterization of Ligand Binding by Saturation Transfer Difference
408 NMR Spectroscopy. Angew Chem Int Ed Engl. 1999;38(12):1784-8.
409 21. Becker W, Bhattiprolu KC, Gubensak N, Zangger K. Investigating Protein-Ligand Interactions
410 by Solution Nuclear Magnetic Resonance Spectroscopy. Chemphyschem. 2018;19(8):895-906.
411 22. Walpole S, Monaco S, Nepravishta R, Angulo J. STD NMR as a technique for ligand screening
412 and structural studies. Methods in Enzymology. 615: Elsevier; 2019. p. 423-51.
413 23. Rogala KB, Dynes NJ, Hatzopoulos GN, Yan J, Pong SK, Robinson CV, et al. The Caenorhabditis
414 elegans protein SAS-5 forms large oligomeric assemblies critical for centriole formation. Elife.
415 2015;4:e07410.
416 24. Hwang TL, Shaka AJ. Water Suppression That Works - Excitation Sculpting Using Arbitrary
417 Wave-Forms and Pulsed-Field Gradients. Journal of Magnetic Resonance Series A. 1995;112(2):275-
418 9.
419 25. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High
420 performance molecular simulations through multi-level parallelism from laptops to supercomputers.
421 SoftwareX. 2015;1:19-25.
422 26. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-
423 chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010;78(8):1950-8.
424 27. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based
425 method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem.
426 1993;97(40):10269-80.
427 28. Bussi G, Zykova-Timan T, Parrinello M. Isothermal-isobaric molecular dynamics using
428 stochastic velocity rescaling. J Chem Phys. 2009;130(7):074101.
429 29. Darden T, York D, Pedersen L. Particle mesh Ewald: An N· log (N) method for Ewald sums in
430 large systems. J Chem Phys. 1993;98(12):10089-92.
431 30. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the
432 analysis of molecular dynamics simulations. J Comput Chem. 2011;32(10):2319-27.
433 31. Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, et al., editors. MDAnalysis: A
434 Python package for the rapid analysis of molecular dynamics simulations. 15th Python in Science
435 Conference; 2016; Austin, TX.
436 32. DeLano WL. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.
437 http://www.pymol.org.2002.
438 33. Collaborative Drug Discovery database public access 2020 [Available from:
439 https://www.collaborativedrug.com/public-access/.
440 34. Grum-Tokars V, Ratia K, Begaye A, Baker SC, Mesecar AD. Evaluating the 3C-like protease
441 activity of SARS-Coronavirus: recommendations for standardized assays for drug discovery. Virus
442 Res. 2008;133(1):63-73.
443 35. Davies TG, Wixted WE, Coyle JE, Griffiths-Jones C, Hearn K, McMenamin R, et al. Monoacidic
444 Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2
445 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based
446 Discovery. J Med Chem. 2016;59(8):3991-4006.
447 36. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: the impact of
448 fragments on drug discovery. Nat Rev Drug Discov. 2016;15(9):605-19.

449

450

Acknowledgements

We are grateful to Nick Soffe for maintenance of the Oxford Biochemistry solution NMR facility, to Claire Strain-Damerell, Petra Lukacik and Martin A. Walsh for advice on M^{pro} production, to Anthony Aimon and Frank von Delft for providing the DSI-poised fragment library, to Adrián García, Nil Casajuana and Clàudia Llinàs del Torrent for advice with MD analysis tools, and to Leonardo Pardo for providing access to high-performance computing facilities. This work was supported by philanthropic donations to the University of Oxford COVID-19 Research Response Fund and the Oxford Glycobiology Institute Endowment. The Oxford Biochemistry NMR facility was supported by the Wellcome Trust (094872/Z/10/Z), the Engineering and Physical Sciences Research Council (EP/R029849/1), the Wellcome Institutional Strategic Support Fund, the EPA Cephalosporin Fund and the John Fell OUP Research Fund. This work was also supported by the "Reinforcement of Postdoctoral Researchers - 2nd Cycle" (MIS-5033021), implemented by the Greek State Scholarships Foundation (IKY).

466 *Figure 1: 1D and STD-NMR spectra of SARS-CoV-2 M^{pro}.* A) Methyl regions from ^{1}H NMR spectra of
467 recombinant SARS-CoV-2 M^{pro}. The spectrum on the left was recorded from a 10 μM protein
468 concentration sample in a 5 mm NMR tube at 25 °C using an excitation sculpting water-suppression
469 method (24). 512 acquisitions with recycle delay of 1.25 sec were averaged, for a total experiment
470 time of just over 10 min. The spectrum on the right was recorded from a 10 μM M^{pro} sample in a 3
471 mm NMR tube at 10 °C, using the same pulse sequence and acquisition parameters. For both
472 spectra, data were processed with a quadratic sine function prior to Fourier transformation. Protein
473 resonances are weaker in the 10 °C spectrum due to lower temperature and the reduced amount of
474 sample used for acquisition in the smaller NMR tube. The position where on-resonance irradiation
475 was applied for STD spectra is indicated. B) Vertically offset ^{1}H STD-NMR spectra from ligand x0434
476 binding to M^{pro}. The reference spectrum is in black with the x0434, H_2O and DMSO ^{1}H resonances
477 indicated. The STD spectrum of x0434 in the presence of M^{pro} is shown in red while that in the
478 absence of M^{pro} is in green. STD spectra are scaled up 64x compared to the reference spectrum.
479 Bottom panels correspond to magnified views of the indicated spectral regions, with x0434
480 resonances assigned to chemical groups of that ligand as shown.

481

482 *Figure 2: Assessment of fragment binding to M^{pro}.* A) STD_{ratio} values for chemical fragments identified
483 by crystallographic screening as binding to M^{pro} (18). Ligands binding to the M^{pro} active site are
484 coloured orange, at the M^{pro} dimer interface in red, and elsewhere on the protein surface in blue. B)
485 Overlay of STD-NMR spectra from fragments x0305, x0387 and x434, which bind the M^{pro} active site,
486 showing the ligand aromatic region in the presence of M^{pro}. Spectra are colour coded per ligand as
487 indicated. As seen, the three fragments yield significantly different STD signal intensities captured in
488 the STD_{ratio} values shown in (A). C) Overlay of reference spectra from fragments x305, x376 and x540,
489 showing the ligand aromatic region. Peak intensities vary substantially, suggesting significant
490 differences in ligand concentration.

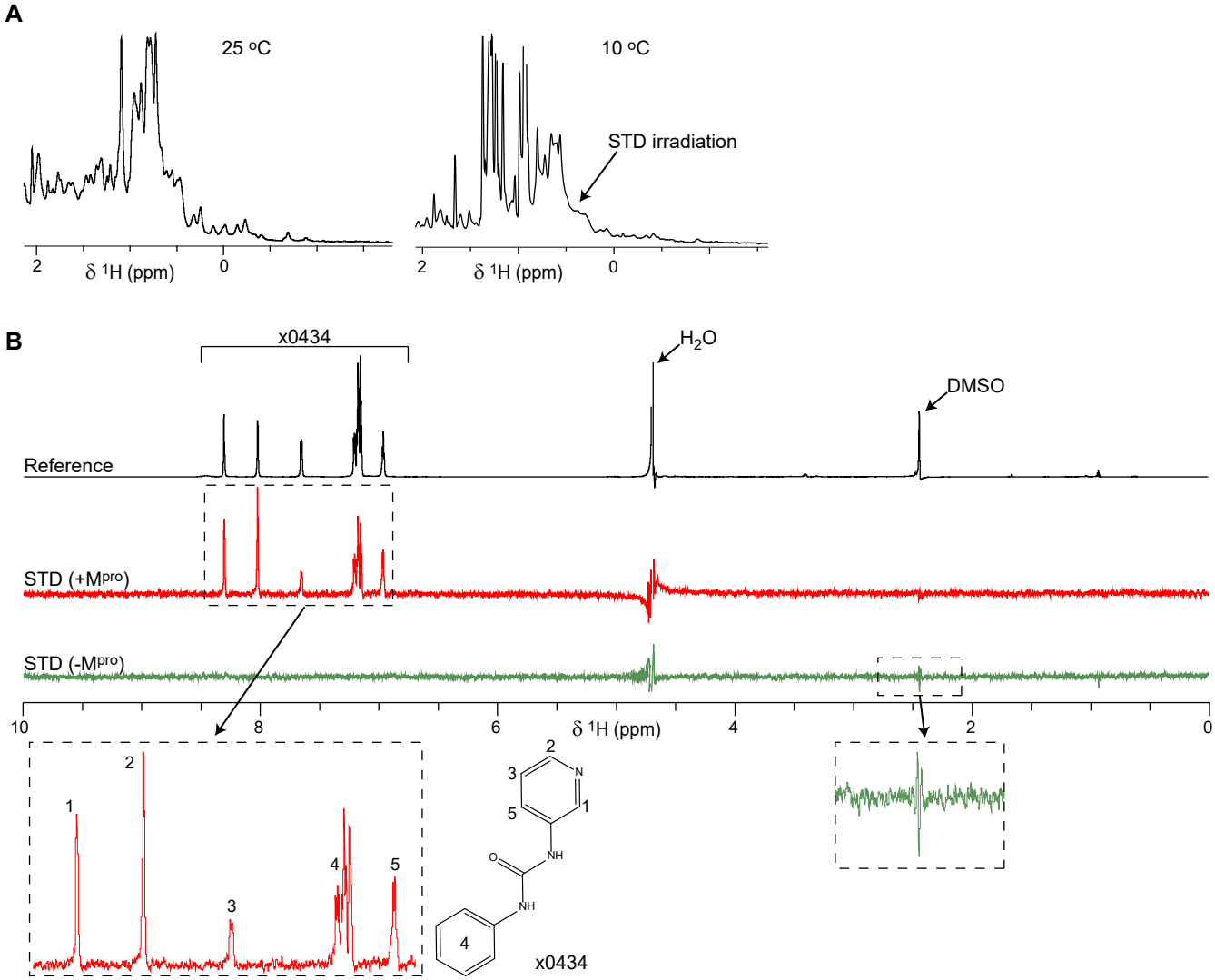
491

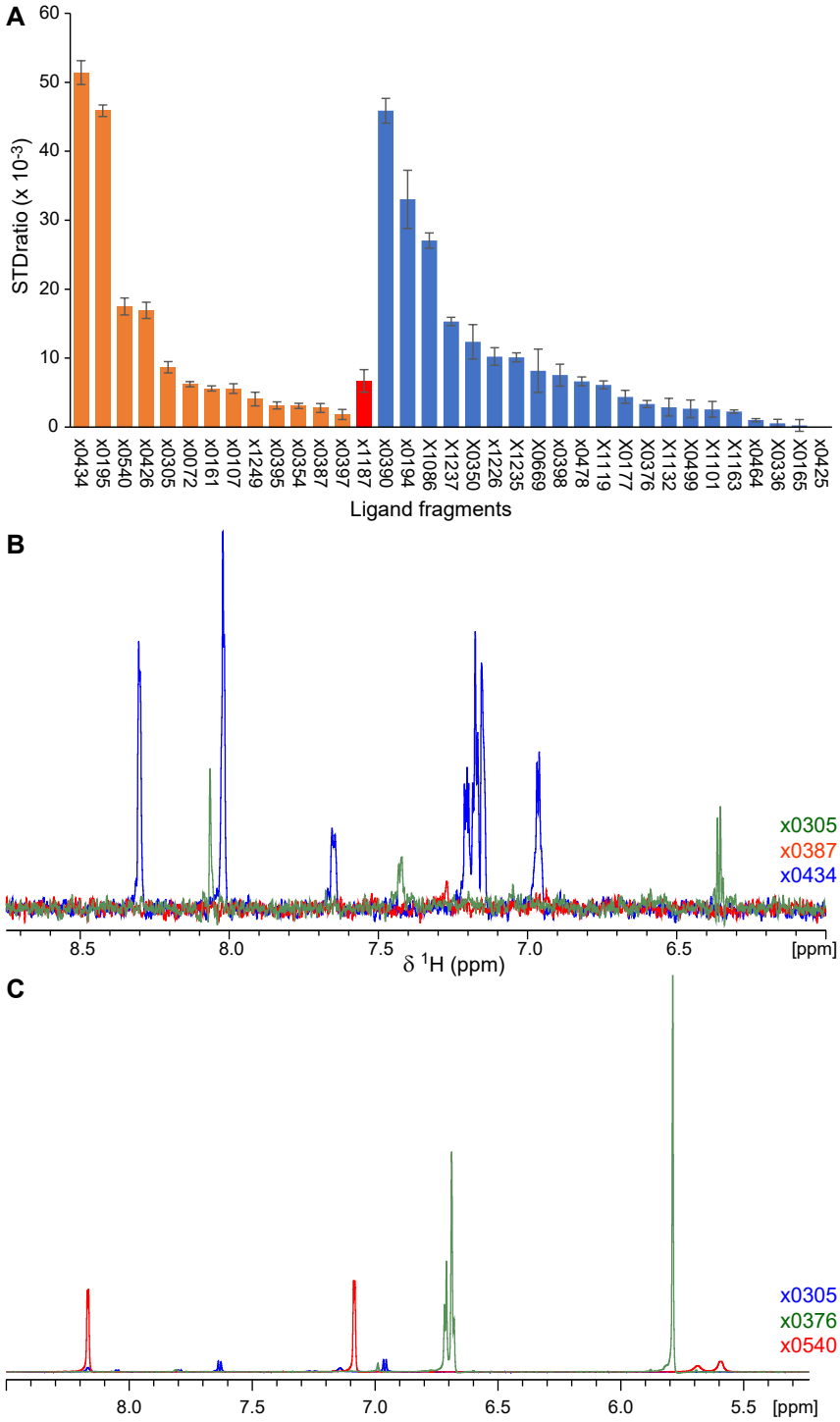
492 *Figure 3. STD-NMR of COVID Moonshot ligands binding to M^{pro}.* A) Overlay of reference spectra from
493 the indicated COVID Moonshot ligands, showing the ligand aromatic region in each case. in the
494 presence of M^{pro}. Spectra are colour coded per ligand as indicated. As seen, peak intensities vary
495 substantially, suggesting significant differences in ligand concentration. Peaks of ligand EDJ-MED-
496 c8e7a002-1 (green) are indicated by arrows; ligand EDJ-MED-e4b030d8-12 (red) produced no peaks
497 in the NMR spectrum. B) Plot of STD_{ratio} values from COVID Moonshot ligands assessed by STD-NMR
498 against their IC_{50} value estimated by RapidFire mass spectroscopy enzymatic assays (17). Ligands in

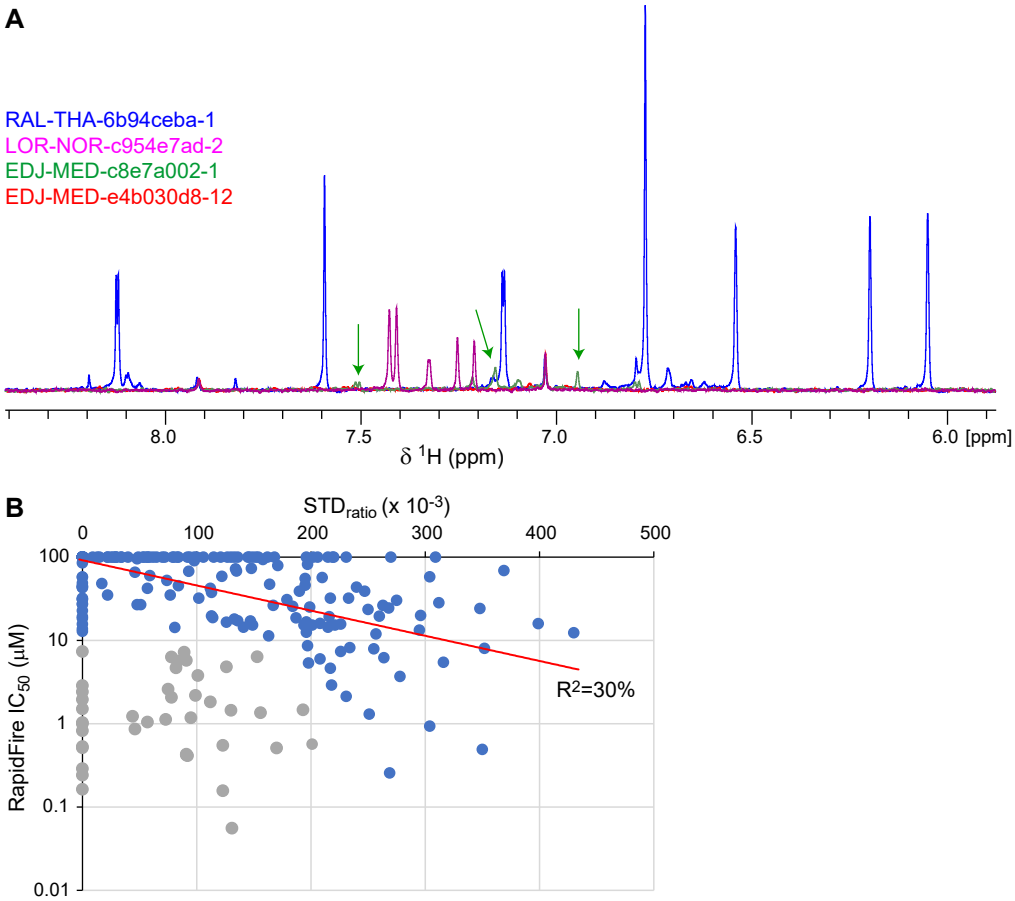
499 blue show weak correlation between the two methods (red line, corresponding to an exponential
500 function along the IC_{50} dimension). Ligands in grey represent outliers of the STD-NMR or enzymatic
501 method as discussed in the text.

502

503


A
25 ºC
10 ºC
STD irradiation
2
0
δ ¹H (ppm)
B
x0434
H2O
DMSO
Reference
STD (+Mpro)
STD (-Mpro)
10
8
6
4
2
0
δ ¹H (ppm)
1
2
3
4
5
N
NH
O
x0434


A
STDratio (x 10-3)
60
50
40
30
20
10
0
x0434
x0195
x0540
x0426
x0305
x0072
x0161
x0107
x1249
x0395
x0354
x0387
x0397
x1187
x0390
x0194
X1086
X1237
X0350
x1226
X1235
X0669
x0398
x0478
X1119
X0177
X0376
X1132
X0499
X1101
X1163
x0464
X0336
X0165
x0425
Ligand fragments
B
x0305
x0387
x0434
8.5
8.0
7.5
7.0
6.5
[ppm]
δ 1H (ppm)
C
x0305
x0376
x0540
8.0
7.5
7.0
6.5
6.0
5.5
[ppm]


A
RAL-THA-6b94ceba-1
LOR-NOR-c954e7ad-2
EDJ-MED-c8e7a002-1
EDJ-MED-e4b030d8-12
8.0
7.5
7.0
6.5
6.0 [ppm]
δ 1H (ppm)
B
STDratio (x 10-3)
0
100
200
300
400
500
100
10
1
0.1
0.01
RapidFire IC50 (μM)
R2=30%