NMR spectroscopy of the main protease of SARS-CoV-2 and fragment-based screening identify three protein hotspots and an antiviral fragment

François-Xavier Cantrelle, Emmanuelle Boll, Lucile Brier, Danai Moschidi, Sandrine Belouzard, Valérie Landry, Florence B Leroux, Frédérique Dewitte, Isabelle Landrieu, Jean Dubuisson, et al.

A Journal of the Gesellschaft Deutscher Chemiker

Angewandte International Edition GDCh Chemie

www.angewandte.org

Accepted Article

Title: NMR spectroscopy of the main protease of SARS-CoV-2 and fragment-based screening identify three protein hotspots and an antiviral fragment

Authors: François-Xavier Cantrelle, Emmanuelle Boll, Lucile Brier, Danai Moschidi, Sandrine Belouzard, Valérie Landry, Florence Leroux, Frédérique Dewitte, Isabelle Landrieu, Jean Dubuisson, Benoit Deprez, Julie Charton, and Xavier Hanoulle

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.202109965

Link to VoR: https://doi.org/10.1002/anie.202109965

WILEY-VCH

10.1002/anie.202109965

WILEY-VCH

RESEARCH ARTICLE

NMR spectroscopy of the main protease of SARS-CoV-2 and fragment-based screening identify three protein hotspots and an antiviral fragment

François-Xavier Cantrelle[a][b]#, Emmanuelle Boll[a][b]#, Lucile Brier[c]#, Danai Moschidi[a][b], Sandrine Belouzard[d], Valérie Landry[c], Florence Leroux[c], Frédérique Dewitte[a][b], Isabelle Landrieu[a][b], Jean Dubuisson[d], Benoit Deprez*[c], Julie Charton[c], and Xavier Hanoulle*[a][b]

[a] Dr.F-X. Cantrelle, E. Boll, D. Moschidi, F. Dewitte, Dr. I. Landrieu, Dr. X. Hanoulle
CNRS ERL9002 - BSI - Integrative Structural Biology
50 avenue Halley, F-59658 Villeneuve d'Ascq, Lille, France
E-mail: xavier.hanoulle@univ-lille.fr
[b] Dr.F-X. Cantrelle, E. Boll, D. Moschidi, F. Dewitte, Dr. I. Landrieu, Dr. X. Hanoulle
Univ. Lille, INSERM, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk Factors and Molecular Determinants of Aging-Related Diseases
1 rue du Professeur Calmette, F-59019, Lille, France
E-mail: xavier.hanoulle@univ-lille.fr
[c] L. Brier, V. Landry, Dr. F. Leroux, Pr. B. Deprez, Dr. J. Charton
Univ. Lille, INSERM, Institut Pasteur de Lille, U1177 - Drugs and Molecules for Living Systems, F-59000, Lille, France; European Genomic Institute for Diabetes, EGID, University of Lille
3 rue du Professeur Laguesse, F-59006, Lille, France
E-mail: benoit.deprez@univ-lille.fr
[d] Dr. S. Belouzard, Dr. J. Dubuisson
Univ. Lille, CNRS, INSERM, CHU Lille, Institut Pasteur de Lille, U1019-UMR 9017 - CIIL - Center for Infection and Immunity of Lille
1 rue du Professeur Calmette, F-59019, Lille, France
# The authors contributed equally

Supporting information for this article is given via a link at the end of the document.

Abstract: The main protease (3CLp) of the SARS-CoV-2, the causative agent for the COVID-19 pandemic, is one of the main targets for drug development. To be active, 3CLp relies on a complex interplay between dimerization, active site flexibility, and allosteric regulation. The deciphering of these mechanisms is a crucial step to enable the search for inhibitors. In this context, using NMR spectroscopy, we studied the conformation of dimeric 3CLp from the SARS-CoV-2 and monitored ligand binding, based on NMR signal assignments. We performed a fragment-based screening that led to the identification of 38 fragment hits. Their binding sites showed three hotspots on 3CLp, two in the substrate binding pocket and one at the dimer interface. **F01** is a non-covalent inhibitor of the 3CLp and has antiviral activity in SARS-CoV-2 infected cells. This study sheds light on the complex structure-function relationships of 3CLp, and constitutes a strong basis to assist in developing potent 3CLp inhibitors.

Introduction

Since the end of 2019, the world faces the global COVID-19 pandemic that represents a major health burden worldwide with strong societal and economic impacts. The etiological agent is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with a case fatality rate of ~2%[1]. This virus represents the seventh coronavirus that infects humans and causes the third β-coronavirus outbreak that emerged in the 21st century. Even though, both vaccines[2–5] and neutralizing antibodies[6–8] are now available to fight against SARS-CoV-2, specific and efficient antivirals against β-coronaviruses are urgently needed to overcome the limited vaccine coverage, variant escapes from antibodies and the future outbreaks.

The RNA genome of SARS-CoV-2 encodes for up to 27 different proteins[9,10]: the structural proteins, the nonstructural proteins (Nsp) and finally several accessory proteins. The Nsp, corresponding to the replicase-transcriptase, are first translated in two polyproteins, pp1a and pp1ab, which are then cleaved by two viral proteases, the main protease (Mpro or 3CLp) and papain-like protease to release 16 functional proteins. 3CLp cleaves at 11 sites (Nsp4-Nsp16), including its own release. Native 3CLp (306 aa) is composed of three domains[11]. Domains I and II are chymotrypsin-like domains with a β-barrel fold and domain III is a 5 α-helices globular domain that is involved in the regulation of 3CLp dimerization. A long linker (L3)[12] connects the domains II and III whereas the N-ter and C-ter (N-terminal and C-terminal) ends are located at the interface between the protomers (Fig. S1). The functional and active SARS-CoV-2 3CLp corresponds to a homodimeric[13] cysteine protease with an unusual catalytic dyad (Cys145, His41). These are buried in a cleft between the domains I and II that is highly conserved among coronaviruses. The recognition sequence, (L,F)Q↓(S,A,G)[14], for the proteolytic cleavage (↓) requires a Gln at position P1 that is a hallmark feature shared by 3CLp of others coronaviruses[15,16], and which in contrast is not present in human proteases[17]. The substrate binding site is made by 4 pockets named S1', S1, S2

10.1002/anie.202109965

and S4[11,18] formed residues from domains I and II and also by residues from the linker (L3). The active conformation of the active site is further stabilized by Ser1 from the second protomer, which stresses the functional importance of 3CLp dimerization.

The function, conservation, substrate specificity, and the absence of human homologue all contribute to make 3CLp an attractive drug target. Structural biology here plays a tremendous role as it helps to select the ligands, to get the molecular details of their interactions, and to find the proper ways to improve their potency. So far, the 3CLp inhibitors correspond either to compounds that covalently react with the catalytic Cys145, or to non-covalent molecules that bind either in the active site or at several allosteric sites. Among the first category, there are boceprevir[19,20], GC376[21], inhibitors 11a[22] and 13b[11]; and more recently PF-00835231[23,24]. In the second, molecules that bind either in its active site (MUT056399, 23R) or elsewhere on its molecular surface, including two allosteric sites (Pelitinib, AT7519), have been identified[25,26]. Moreover, fragment screening has also been performed to identify fragments[27] that can be grown, linked or merged to develop potent 3CLp inhibitors[28]. Numerous structural biology methods, including crystallography (X-ray[11,18,22,25] or neutrons[29]), mass spectrometry[13,19,27], computational analyses[30,31,31,32], have been used to get a better understanding of the complex structure-function relationships in 3CLp, including the conformational flexibility of its active site[11,12,18,19,25,30,33–36], and then to find or conceive inhibitors.

In this work, we used NMR spectroscopy to study the dimeric SARS-CoV-2 3CLp. We obtained its NMR chemical shift backbone assignment and used these data in a fragment-based screening that led to the identification of 38 fragment hits. The deciphering of their binding sites and the conformational consequences they induced in 3CLp led to the identification of 3 protein hotspots, two located in the active site of the protease, with two different NMR signatures, and one at the dimerization interface. We further show that the fragment lead **F01** binds in the active site and is, without optimization, a reversible 3CLp inhibitor with antiviral activity in SARS-CoV-2 infected Vero-81 cells. The crystal structure of F01-bound 3CLp that we have solved will help its optimization. These NMR data should help to get a better understanding of the complex interplays between the active site plasticity, the dimerization and the enzymatic activity of 3CLp. This also constitutes a new tool to assist the development of potent 3CLp inhibitors for the present or future outbreaks.

Results and Discussion

NMR spectroscopy of SARS-CoV-2 3CLp dimer. We produced SARS-CoV-2 3CLp samples with different isotopic labeling schemes to study by liquid-state NMR spectroscopy. The purified protease (306 aa, 67.6 kDa) has both native N- and C-terminal ends (SI and Fig. S1), which is crucial for both its enzymatic activity and its proper dimerization. We obtained good quality $^{1}H,^{15}N$-TROSY HSQC spectrum, with ~280 resonances (Figure 1) and then recorded 3D $^{1}H,^{15}N,^{13}C$ TROSY- HNCACB, -HN(CO)CACB, -HNCO, -HN(CA)CO, -HN(CO)CA spectra. Due to unfavorable magnetic relaxation properties some ^{13}C signals were not observed and thus we had to record additional data on other samples, including 3CLp bound to boceprevir, and a monomeric 3CLp R298A mutant, in order to reduce the protein dynamics and the molecular weight, respectively. To perform the NMR backbone assignments of SARS-CoV-2 3CLp, we used a combined and integrated strategy that includes classical sequential assignment, analyses of chemical shift perturbations (CSPs) upon boceprevir binding, CS predictions and previous NMR assignments for the isolated N-ter and C-ter domains of SARS-CoV 3CLp[37] (see SI). We assigned 183 proton amide correlations (183/293, 63%) and further obtained 239, 207 and 234 chemical shifts for Cα, Cβ and C', respectively (Figure 1, SI, BMRB entry 50780).

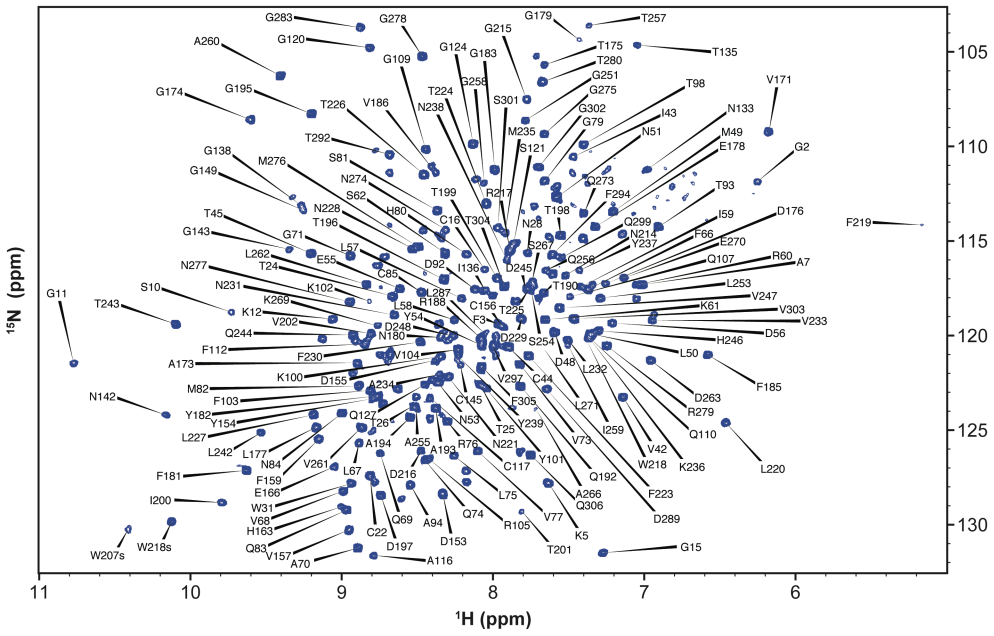


Figure 1. 2D $^{1}H,^{15}N$-TROSY-HSQC NMR spectrum of SARS-CoV-2 3CLp dimer. The assignments are annotated with black labels.

10.1002/anie.202109965

Most of the unassigned proton amides lie in the first two β-barrel domains or at the dimerization interface (Fig. S2). Whereas previous attempts to record multidimensional NMR data on SARS-CoV[38] and SARS-CoV-2[39] 3CLp have failed, these new NMR data open the field to a large range of future studies of the dimeric 3CLp in solution and at temperature close to physiological, an important parameter when considering dynamics. To assess the potential of our experimental system, we analyzed the 3CLp spectral perturbations upon binding of either boceprevir or GC376 (Figs. S3-S4). In both cases, the perturbations induced are highest in the active site but also propagate further in the two catalytic domains, and even toward its C-terminal end with GC376. NMR perturbations may arise from ligand binding but also from the subsequent conformational changes. GC376 indeed induces perturbations both at the active site and at the dimerization interface, the two regions of the protease that are targeted to develop inhibitors[13,25,27,40]. Moreover, in the presence of GC376, a few 3CLp NMR resonances split into two new ones (Fig. S5), probably highlighting the two conformations of the P3 moiety of the bound inhibitor[20]. The split resonances notably match with Val42, Asn142, Gln192 and Gly2. The later one showing that we can detect inter-protomer conformational consequences. Interestingly, when using a R298A 3CLp monomeric mutant, we observed ~115 additional resonances in the 2D ^{1}H,^{15}N NMR spectrum that is ~100 more than expected. This could be due to the two orientations of the domain III that have been described for SARS-CoV 3CLp R298A[41]. These data highlight the potential for in-solution studies of the 3CLp. Based on the NMR assignments we are able to not only detect ligand binding and map the binding site(s), but also to analyze the conformational rearrangement(s) throughout the dimer, providing essential molecular detail for medicinal chemistry.

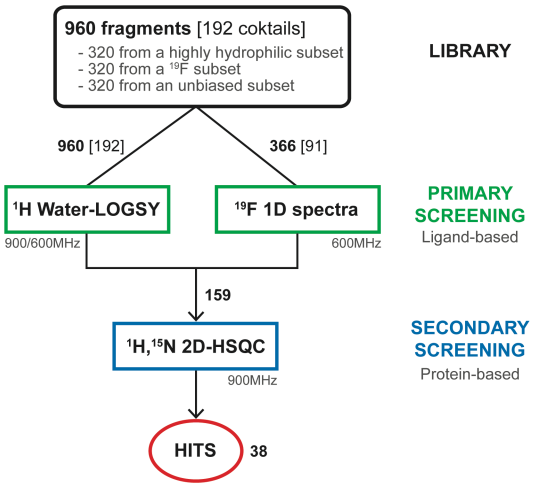


Figure 2. NMR fragment screening.

NMR Fragment-based screening set up. Fragment screening is widely used in drug discovery as it allows to efficiently probe the chemical space while keeping reasonable the numbers of molecule that have to be assessed[28]. The fragment hits identified (low MW) that bind to the target are then optimized to give lead compounds. We used a library of 960 commercially available fragments with physio-chemical properties that mostly fulfill the 'rule of three' criteria[42] (Fig. S7a-d). We designed a strategy with a primary and a secondary screening using ligand- and protein-observed NMR methods, respectively (Figure 2). The screening steps were performed in the presence of DTT, a nucleophile and reducing agent, to minimize the selection of highly electrophilic and nonspecific compounds that would covalently bind to the protease.

^{1}H and ^{19}F NMR ligand-based primary screening. The 960 fragments were split into 192 cocktails of 5 fragments, as this strategy already proved efficient[43]. All the cocktails have been analyzed with ^{1}H Water-LOGSY[44] and additionally with ^{19}F spectroscopy for 91 of them (Figure 2), as our library contains 427 fluorine fragments in total. With Water-LOGSY, the detection of the hits is straightforward since their signals have opposite phase (Figure 3a). When using ^{19}F spectroscopy, the spectra only contain one NMR signal for each ^{19}F-fragment present in the cocktail and we monitored both CSPs and signal broadening (Figure 3b). The primary screening led to the identification of 159 binders (Scheme S1), corresponding to a 16.6 % hit rate (Figure 2).

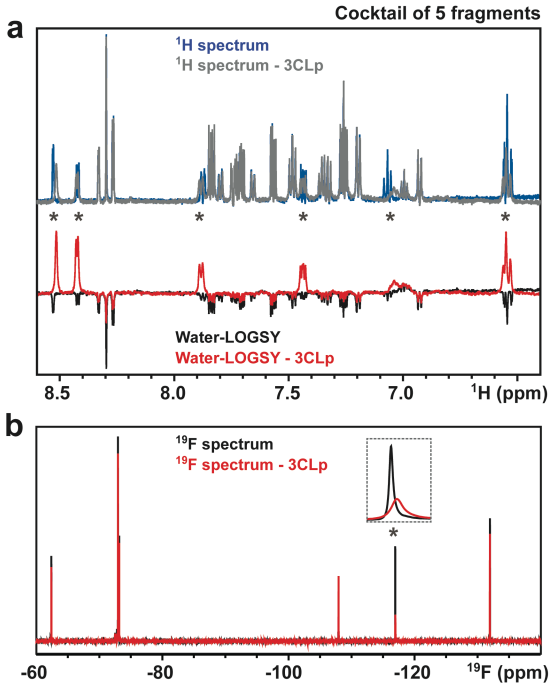


Figure 3. Ligand-based NMR primary screening. Analyses of a 5-fragment cocktail in the absence and in the presence of unlabeled 3CLp. (a) 1D ^{1}H and ^{1}H Water-LOGSY spectra. (b) 1D ^{19}F-NMR of the same cocktail. Signals, annotated with an asterisk, correspond to the F04 fragment that is a direct binder. See Scheme S1 for other cocktails.

Secondary screening using NMR spectra of 3CLp. We performed the secondary screening using 2D ^{1}H,^{15}N TROSY-HSQC spectra that have been acquired on SARS-CoV-2 3CLp in the presence of each of the 159 binders identified in the primary screening. Using both CSPs and signal broadening (Figure 4), we confirmed 38 fragments as direct binders of 3CLp, corresponding to an overall ~4% hit rate (Figures 2 and 4 and Scheme S2, Tables S1-S2). This value can be compared with the ~6% obtained in a combined MS and X-ray approach[27]. The ratio of

10.1002/anie.202109965

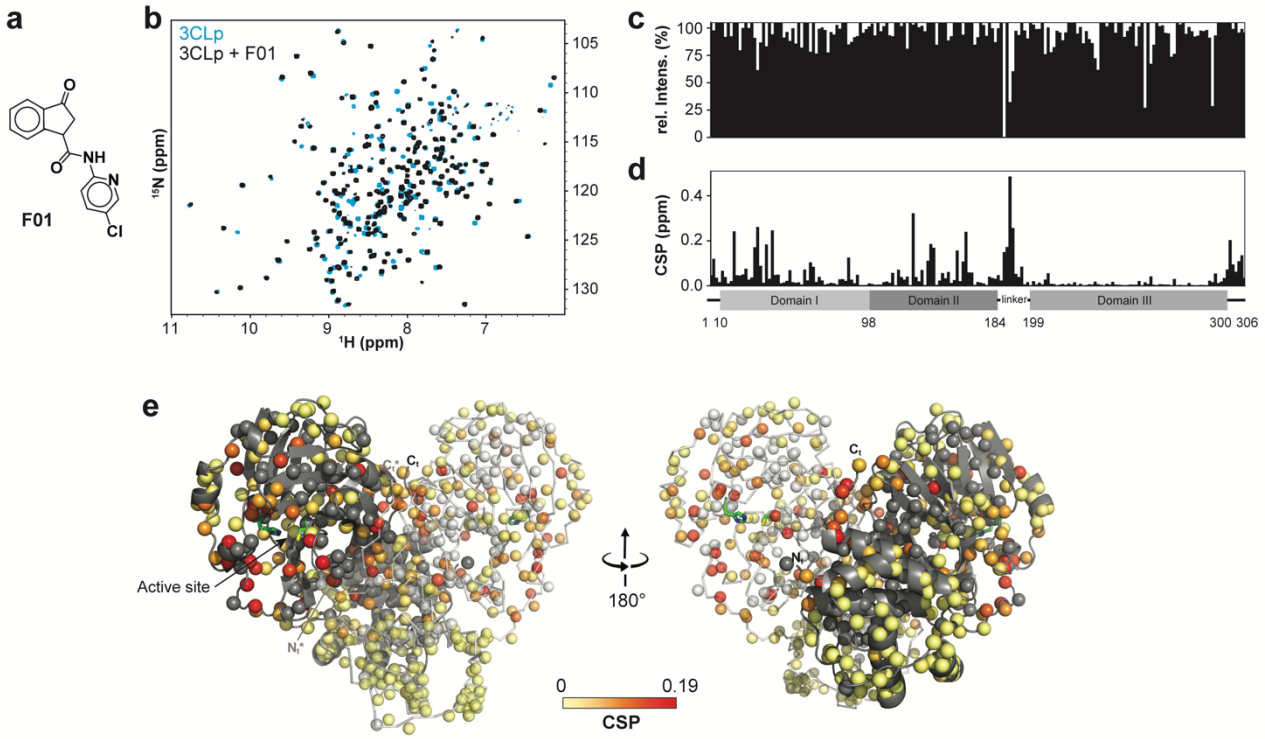


Figure 4. Protein-based NMR secondary screening. (a) Fragment **F01** structure. (b) Overlay of two 2D ^{1}H,^{15}N-TROSY-HSQC spectra acquired on 3CLp in the absence (in light blue) and in the presence (in black) of fragment **F01**. The broadening of the resonances and ^{1}H and ^{15}N-combined CSPs induced upon fragment binding are shown along the 3CLp sequence in (c) and (d), respectively. (e) Structure of the 3CLp dimer (PDB: 7k3t), with protomers A and B shown in grey and white, respectively. Each small ball represents a proton amide and thus should correspond to a resonance in the ^{1}H,^{15}N 2D spectrum. The CSPs, shown in (d), have been color coded (from light yellow to red) and are displayed on these balls. Unassigned residues were kept in the original color of the protomer. Catalytic His41 and Cys145, are shown in green. See Scheme S2 and Table S2 for other hits. See the SI for a color-blind-friendly version of this figure.

^{19}F-containing fragments in the hits (~40%) is close to the ratio in the library used. In contrast, both the average MW and lipophilicity of the fragment hits are higher than those in the entire library (Fig. S7a-d)

Identification of three different classes of binders corresponding to three protein hotspots. Using the backbone assignments, the analysis of the CSPs induced by the 38 hits shows that they can be grouped into three classes corresponding to three 3CLp hotspots (Figure 5; Fig. S8). In Class I (24 hits), CSPs are observed for resonances assigned to residues distributed in the active site cleft, in the loop L3, and in the C-ter end, whereas residues from the N-ter end are only moderately affected. Class II is made by 8 hits that induce CSPs for only a restricted set of residues, in the substrate binding site, that belong exclusively to either the domain I or the tip of the loop L3 and that corresponds to the S2 and S3 binding sites. Class III (5 hits) is defined by CSPs for residues located at the dimerization interface of 3CLp (N-ter and C-ter ends). As to the fragment F27, it induces a strong reduction in the signal intensity all along the 3CLp sequence (Fig. S8 and Table S2), and may correspond to a false positive.

Analysis of the 3CLp binding hotspots. The CSPs pattern in Class I, illustrated by fragment **F01**, is similar to the ones observed in 3CLp upon binding of either boceprevir or GC376 (Fig. S8), two potent inhibitors. The NMR CSPs induced upon binding of **F01** (see Figure 4) correspond to residues distributed all along the 3CLp active site cleft (S1-S4 pockets) and indeed match with the residues involved in the binding of GC376 (Fig. S9a). Moreover, the CSPs propagate toward the 3CLp dimerization interface, as with GC376 (Fig. S4).

These NMR data are fully supported by the crystal structure of fragment **F01**-bound 3CLp that we solved (Figure 6 and Fig. S10 and Table S3; PDB: 7p51). The 3-oxo-2,3-dihydro-indene ring and 5-chloro-2-pyridyl group of **F01** occupy the S1 and S2 pockets of 3CLp, respectively. Three hydrogen bonds are formed between **F01** and 3CLp. One of them involves the ketone in the indene ring of **F01** that is electrophilic and could covalently react with the catalytic Cys145. This group, located in a key position of the active site, rather behaves as a H-bond acceptor and interacts with His163 (see SI). The binding of **F01** induces conformational changes in all the active site of 3CLp (see SI, Figure 6 and Fig. S10b). It induces the displacement of: the α-helix (Ser46-Leu50) around the S2 pocket, the loop L3 and of Asn142 and Glu166 residues around the S1 pocket. This last movement propagates to the 3CLp dimeric interface with Ser1 of protomer B being slightly displaced. It has been shown that in the 3CLp dimer, Ser1 from protomer B interacts with Glu166 of protomer A and stabilizes the active conformation of the S1 pocket[11,18]. Thus, the CSPs observed in 3CLp spectrum upon **F01** binding both match with its binding site and the induced conformational changes through allosteric pathways (Fig. S10c).

Our data show that conformational plasticity[29,36] and allosteric regulations [13,25,35] within 3CLp can be studied using NMR

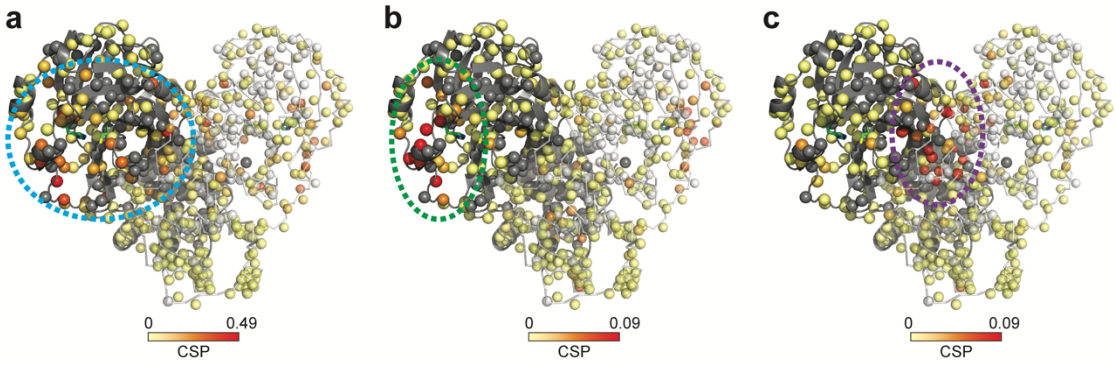


Figure 5. The 38 hits identified in the NMR screening can be grouped into three classes according to the CSPs they induced on the 2D NMR spectrum of $^{2}H,^{15}N$-3CLp upon binding. The representation is similar to that in Figure 4e. (a) Class I (**F01**) - The CSPs are distributed in all the active site cleft, including the S1-S4 substrate pockets, and extend toward the dimerization interface of the protease. (b) Class II (**F30**) - The CSPs induced correspond to a binding of the fragments in the S2 and S3 pockets, with the highest perturbations observed for residues located in a short α-helix (Ser46-Leu50). (c) Class III (**F15**) - Upon binding these fragments induce CSPs at the dimerization interface of 3CLp. See Figs. S8-S9. See the SI for a color-blind-friendly version of this figure.

spectroscopy, especially the tight interplay between substrate binding, active site conformation and dimerization.

The hits from Class II, such as **F30**, induced CSPs that would correspond to their binding into the S2 and S3 pockets located in the domain I-side of the 3CLp substrate binding site, as SEN1269[25] (Fig. S9b). This molecule binds to S2 and induced the displacement of the short α-helix (Ser46-Leu50), for which we observed the highest CSPs upon binding of Class II hits (Fig. S8).

The NMR CSPs induced upon binding of the Class III hits, which includes **F15**, are localized at the 3CLp dimeric interface and could be predicted to resemble the binding of x1086 and x1187[13,27] in the hydrophobic pocket made by residues both in the N-ter (Met6, Phe8) and C-ter (Arg298, Gln299, Val303) ends (Fig. S9c). With **F15**, we also observed a high CSP for the resonance corresponding to Gln127, which is at the dimeric interface, and that has been shown to make a hydrogen bond with x1086.

Interestingly, no NMR perturbations observed in our screening match with fragment binding into the allosteric sites 1 and 2 that have been identified by Günther *et al*[25]. It could be that the binding in these two sites requires bigger and more complex molecule structures, or simply that the fragment library used did not allow to probe all the possible binding sites.

Looking at the chemical properties of the fragment hits on the basis of their Class I, II or III belonging, we found that in average Class II hits are smaller than Class I hits (avg. 233.3 Da *vs* 245.7 Da), and that Class III hits are even smaller (avg. 206.85 Da) and are also in more hydrophobic (80% with 2<AlogP<3) (Fig. S7).

Among the 38 hits identified in this work, **F01** induced the highest CSPs in the NMR spectrum of SARS-CoV-2 3CLp (Table S2 and Scheme S2).

F01 is a reversible inhibitor of 3CLp and has antiviral activity against SARS-CoV-2. We further characterized **F01**, the main hit of our screening. First, using NMR titration experiments, we determined a dissociation constant K_D = 73±14 μM for the interaction between **F01** and 3CLp (Figure 7a and Fig. S11).

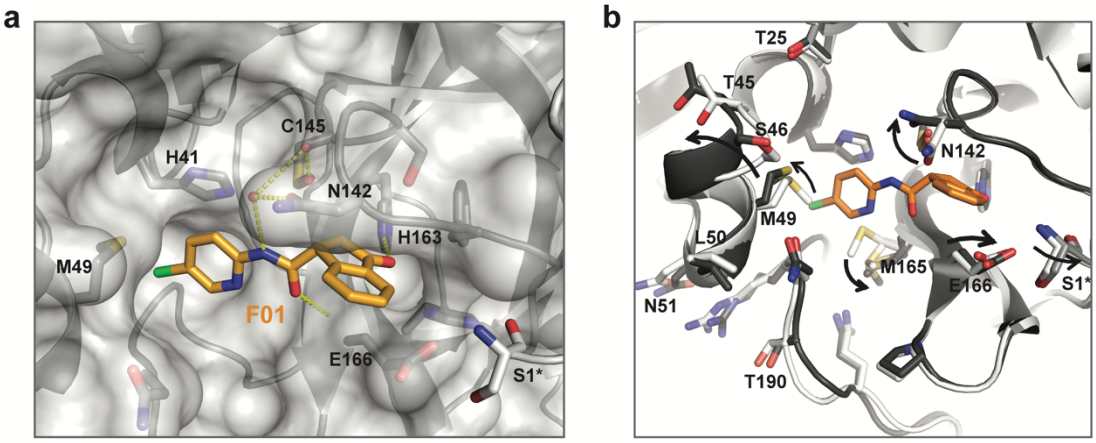


Figure 6. Crystal structure of the fragment **F01**-bound 3CLp. (a) Close-up view of the **F01** binding in the active site. Protomer A is shown in grey and with surface representation, whereas protomer B is displayed in white and in cartoon representation. Three hydrogen bonds between **F01** and 3CLp are displayed as yellow dashes. Residue from protomer A are labeled in black and residue S1 from protomer B is marked with an asterisk. (b) Conformational changes in the **F01**-bound 3CLp structure (PDB: 7p51) compared to the apo 3CLp structure (PDB: 7nts). See Fig. S10.

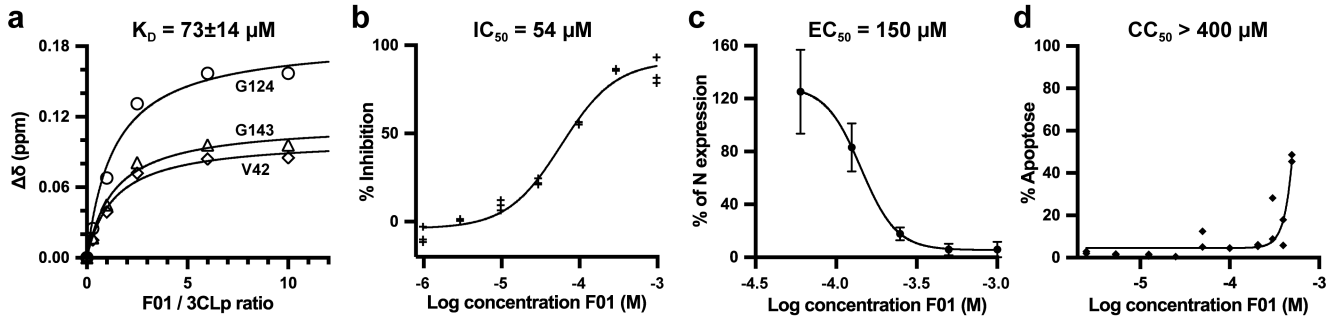


Figure 7. F01 is an inhibitor of 3CLp and is active against SARS-CoV-2 in Vero-81 cells. (a) Affinity of the Interaction between **F01** and 3CLp. NMR titration curves where the $^{1}H,^{15}N$-combined CSPs ($\Delta\delta$, ppm) were plotted as a function of the F01/3CLp ratios. The K_D value (µM) corresponds to the mean (±SD) calculated over 18 3CLp resonances (Fig. S11). (b) **F01** inhibits the *in vitro* enzymatic activity of 3CLp. The half-maximal inhibitory concentration (IC_{50}) has been calculated using the initial velocities of the reactions. (c) The antiviral activity of **F01** against SARS-CoV-2 has been tested on Vero-81 infected cells. After infection in the presence of increasing **F01** concentrations, the cells were lysed (t=16h) and the viral N-protein content was quantified and was used to determine the half-maximal effective concentration (EC_{50}). Viral titers were also measured in the cell supernatants (Fig. S13). (d) The 50% cytotoxic concentration (CC_{50}) of **F01** has been assayed on Vero-81 cells (t=20h).

This affinity is higher than expected, as initial hits from fragment-based screening usually bind to their target with a low affinity, in the 1-10 mM range[45]. Second, using an *in-vitro* enzymatic assay, we showed that **F01** is an inhibitor of 3CLp with a moderate potency (IC_{50} = 54 µM) (Figure 7b). Third, using jump dilution assay, we showed that **F01** is a reversible inhibitor of the protease (Fig. S12), which agrees with the crystal structure (see Figure 6). Finally, Vero-81 cells were infected with SARS-CoV-2 in the presence of increasing concentrations of **F01** and then both the viral N protein cellular content was assayed and the number of infectious viral particles was determined in the cell supernatants. The results showed that **F01** has antiviral activity (EC_{50} = 150 µM) against SARS-CoV-2 (Figure 7c and Fig. S13) while displaying a low cytotoxicity (CC_{50} > 400 µM) (Figure 7d).

Usually, the initial fragment hits have neither *in vitro* nor biological activity, as they often are too small and bind to their target with very low affinity. In this work, we identified the fragment **F01** that even without optimization has antiviral activity against SARS-CoV-2. Very recently, Bajusz *et al*. have reported a fragment, SX013, that blocks the SARS-CoV-2 replication in Vero E6 cells with an EC_{50} of 304 µM[46], which is double of that for **F01** in Vero-81 cells. The ligand efficiency of **F01** is 0.29-0.30 kcal.mol^{-1}.heavy atom^{-1} showing that **F01** is a good fragment lead and deserved to be optimized in order to increase its potency and other drug related properties[28,47].

Conclusion

Whereas structural biology plays a central role in drug discovery and drug development, up to date, NMR spectroscopy has not successfully been pushed forward to study the 3CLp from coronaviruses[37–39]. In this work, we used solution-state NMR spectroscopy to study the dimeric 3CLp protease of the SARS-CoV-2, which is one of the main targets to develop efficient antivirals to fight against the COVID-19 pandemic. Considering the high sequence conservation between the 3CLps[20,40], our data will also be valuable for others β-coronaviruses, such as MERS-CoV and SARS-CoV (67% and 98% sequence similarity, respectively), and possibly for future emerging β-coronaviruses. Even being incomplete, the 3CLp backbone chemical shift assignment, obtained at pH and temperature close to physiological ones, has proved to be highly valuable in a medicinal chemistry project as these new NMR data allowed the study of both the structure and conformation of the dimeric protease. As a complement to the molecular dynamics[12,30,35,48], these data also provide, for future studies, an experimental mean to assess the 3CLp dynamics in solution, an important point to consider in drug development.

Since mid-2020, the world faces the apparition of SARS-CoV-2 variants that may, at least partially, escape to currents vaccines. This stresses that there is a need for direct acting antiviral(s) and also that there is a high risk for emergence of resistance mutations in 3CLp if targeted. To help resolve this common issue in drug development, a promising strategy consists in the combination of both orthosteric and allosteric drugs[49,50]. In this way, our NMR data could be valuable to identify both the allosteric sites of SARS-CoV-2 3CLp and the molecules that bind into, and to identify the allosteric pathways along which resistance mutations may also occur.

Using a two-step fragment screening, we identified 38 hits, including the promising fragment **F01**, and three binding sites, or hotspots, located in the active site and at the dimerization interface of 3CLp. It has been shown that 3CLp can indeed be efficiently targeted at its active site, at its dimerization interface and even at different allosteric sites[11,13,14,18,22,24,25,27,29,51]. We showed that **F01** binds to 3CLp active site with a rather good affinity (K_D = 73 µM), is a non-covalent reversible inhibitor of the protease (IC_{50} = 54 µM) and demonstrates antiviral activity against SARS-CoV-2 (EC_{50} = 150 µM), despite no optimization. Our results indicates that **F01** is a promising fragment lead that deserved to be optimized to give more potent compounds[28,52]. Structure-activity relationship studies, guided by the the crystal structure, will help this process and two approaches could be considered: first, **F01** (Class I) could be linked or merged to Class II hits, and second, **F01** could be studied in combination with fragments from Class III that bind at the dimerization interface.

This work and our NMR results will benefit to the better understanding of the complex structure-function relationships in the dimer of 3CLp and assist the rational design of potent 3CLp inhibitors, that may both block its active site and interfere with its dimerization, in order to tackle current, or even future, coronaviruses pandemics.

Accession codes

Backbone NMR assignments of SARS-CoV-2 3CLp have been deposited in the Biological Magnetic Resonance Data Bank (Entry 50780).

Crystal structures of apo 3CLp and 3CLp in complex with fragment F01 have been deposited in the Protein Data Bank as entries 7NTS and 7P51, respectively.

Acknowledgements

The NMR facilities were funded by the Nord Region Council, CNRS, Institut Pasteur de Lille, European Union (FEDER), French Research Ministry and University of Lille. Financial support from the IR-RMN-THC (FR 3050 CNRS) for the infrastructure is gratefully acknowledged.

This study was supported by the I-site ULNE (project 3CLPRO-SCREEN-NMR), The CPER CTRL (Transdisciplinary Research Center on Longevity) program, and the Institut Pasteur de Lille.

Prof. B. Luy (Karlsruhe Institute of Technology) and Dr. D. Sinnaeve are thanked for advice about the ^{19}F BURBOP pulses.

We would like to thank T. Isabet, S. Sirigu and W. Shepard for their valuable support during data collection at beamlines PX1 and PX2A at the SOLEIL synchrotron facility (Paris, France). We thank Dr V. Villeret and Dr E. Dupre for their advice on crystallogenesis and data processing.

Keywords: NMR spectroscopy • Viruses • Protein structure • Fragment screening • Drug discovery

[1] World Health Organization, "WHO Coronavirus (COVID-19) Dashboard," can be found under https://covid19.who.int, **2021**.
[2] A. Mullard, *Lancet* **2020**, *395*, 1751–1752.
[3] F. P. Polack, S. J. Thomas, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart, J. L. Perez, G. Pérez Marc, E. D. Moreira, C. Zerbini, R. Bailey, K. A. Swanson, S. Roychoudhury, K. Koury, P. Li, W. V. Kalina, D. Cooper, R. W. Frenck, L. L. Hammitt, Ö. Türeci, H. Nell, A. Schaefer, S. Ünal, D. B. Tresnan, S. Mather, P. R. Dormitzer, U. Şahin, K. U. Jansen, W. C. Gruber, C4591001 Clinical Trial Group, *N Engl J Med* **2020**, *383*, 2603–2615.
[4] L. R. Baden, H. M. El Sahly, B. Essink, K. Kotloff, S. Frey, R. Novak, D. Diemert, S. A. Spector, N. Rouphael, C. B. Creech, J. McGettigan, S. Khetan, N. Segall, J. Solis, A. Brosz, C. Fierro, H. Schwartz, K. Neuzil, L. Corey, P. Gilbert, H. Janes, D. Follmann, M. Marovich, J. Mascola, L. Polakowski, J. Ledgerwood, B. S. Graham, H. Bennett, R. Pajon, C. Knightly, B. Leav, W. Deng, H. Zhou, S. Han, M. Ivarsson, J. Miller, T. Zaks, COVE Study Group, *N Engl J Med* **2021**, *384*, 403–416.
[5] M. Voysey, S. A. C. Clemens, S. A. Madhi, L. Y. Weckx, P. M. Folegatti, P. K. Aley, B. Angus, V. L. Baillie, S. L. Barnabas, Q. E. Bhorat, S. Bibi, C. Briner, P. Cicconi, A. M. Collins, R. Colin-Jones, C. L. Cutland, T. C. Darton, K. Dheda, C. J. A. Duncan, K. R. W. Emary, K. J. Ewer, L. Fairlie, S. N. Faust, S. Feng, D. M. Ferreira, A. Finn, A. L. Goodman, C. M. Green, C. A. Green, P. T. Heath, C. Hill, H. Hill, I. Hirsch, S. H. C. Hodgson, A. Izu, S. Jackson, D. Jenkin, C. C. D. Joe, S. Kerridge, A. Koen, G. Kwatra, R. Lazarus, A. M. Lawrie, A. Lelliott, V. Libri, P. J. Lillie, R. Mallory, A. V. A. Mendes, E. P. Milan, A. M. Minassian, A. McGregor, H. Morrison, Y. F. Mujadidi, A. Nana, P. J. O'Reilly, S. D. Padayachee, A. Pittella, E. Plested, K. M. Pollock, M. N. Ramasamy, S. Rhead, A. V. Schwarzbold, N. Singh, A. Smith, R. Song, M. D. Snape, E. Sprinz, R. K. Sutherland, R. Tarrant, E. C. Thomson, M. E. Török, M. Toshner, D. P. J. Turner, J. Vekemans, T. L. Villafana, M. E. E. Watson, C. J. Williams, A. D. Douglas, A. V. S. Hill, T. Lambe, S. C. Gilbert, A. J. Pollard, Oxford COVID Vaccine Trial Group, *Lancet* **2021**, *397*, 99–111.
[6] C. Wang, W. Li, D. Drabek, N. M. A. Okba, R. van Haperen, A. D. M. E. Osterhaus, F. J. M. van Kuppeveld, B. L. Haagmans, F. Grosveld, B.-J. Bosch, *Nat Commun* **2020**, *11*, 2251.
[7] X. Chen, R. Li, Z. Pan, C. Qian, Y. Yang, R. You, J. Zhao, P. Liu, L. Gao, Z. Li, Q. Huang, L. Xu, J. Tang, Q. Tian, W. Yao, L. Hu, X. Yan, X. Zhou, Y. Wu, K. Deng, Z. Zhang, Z. Qian, Y. Chen, L. Ye, *Cell Mol Immunol* **2020**, *17*, 647–649.
[8] Y. Cao, B. Su, X. Guo, W. Sun, Y. Deng, L. Bao, Q. Zhu, X. Zhang, Y. Zheng, C. Geng, X. Chai, R. He, X. Li, Q. Lv, H. Zhu, W. Deng, Y. Xu, Y. Wang, L. Qiao, Y. Tan, L. Song, G. Wang, X. Du, N. Gao, J. Liu, J. Xiao, X.-D. Su, Z. Du, Y. Feng, C. Qin, C. Qin, R. Jin, X. S. Xie, *Cell* **2020**, *182*, 73-84.e16.
[9] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, Y.-Z. Zhang, *Nature* **2020**, *579*, 265–269.
[10] Y. Chen, Q. Liu, D. Guo, *J Med Virol* **2020**, *92*, 418–423.
[11] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, *Science* **2020**, *368*, 409–412.
[12] N. Verma, J. A. Henderson, J. Shen, *J. Am. Chem. Soc.* **2020**, *142*, 21883–21890.
[13] T. J. El-Baba, C. A. Lutomski, A. L. Kantsadi, T. R. Malla, T. John, V. Mikhailov, J. R. Bolla, C. J. Schofield, N. Zitzmann, I. Vakonakis, C. V. Robinson, *Angewandte Chemie International Edition* **2020**, *59*, 23544–23548.
[14] W. Rut, K. Groborz, L. Zhang, X. Sun, M. Zmudzinski, B. Pawlik, X. Wang, D. Jochmans, J. Neyts, W. Młynarski, R. Hilgenfeld, M. Drag, *Nature Chemical Biology* **2020**, DOI 10.1038/s41589-020-00689-z.
[15] K. Fan, P. Wei, Q. Feng, S. Chen, C. Huang, L. Ma, B. Lai, J. Pei, Y. Liu, J. Chen, L. Lai, *Journal of Biological Chemistry* **2004**, *279*, 1637–1642.
[16] C.-P. Chuck, H.-F. Chow, D. C.-C. Wan, K.-B. Wong, *PLoS One* **2011**, *6*, e27228.
[17] L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E. J. Snijder, H. Liu, R. Hilgenfeld, *J. Med. Chem.* **2020**, *63*, 4562–4578.
[18] 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.
[19] C. Ma, M. D. Sacco, B. Hurst, J. A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M. T. Marty, Y. Chen, J. Wang, *Cell Res* **2020**, *30*, 678–692.
[20] L. Fu, F. Ye, Y. Feng, F. Yu, Q. Wang, Y. Wu, C. Zhao, H. Sun, B. Huang, P. Niu, H. Song, Y. Shi, X. Li, W. Tan, J. Qi, G. F. Gao, *Nat Commun* **2020**, *11*, 4417.
[21] W. Vuong, M. B. Khan, C. Fischer, E. Arutyunova, T. Lamer, J. Shields, H. A. Saffran, R. T. McKay, M. J. van Belkum, M. A. Joyce, H. S. Young, D. L. Tyrrell, J. C. Vederas, M. J. Lemieux, *Nature Communications* **2020**, *11*, DOI 10.1038/s41467-020-18096-2.
[22] W. Dai, B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang, H. Liu, *Science* **2020**, *368*, 1331–1335.
[23] R. L. Hoffman, R. S. Kania, M. A. Brothers, J. F. Davies, R. A. Ferre, K. S. Gajiwala, M. He, R. J. Hogan, K. Kozminski, L. Y. Li, J. W. Lockner, J. Lou, M. T. Marra, L. J. Mitchell, B. W. Murray, J. A. Nieman, S. Noell, S. P. Planken, T. Rowe, K. Ryan, G. J. Smith, J. E. Solowiej, C. M. Steppan, B. Taggart, *J Med Chem* **2020**, DOI 10.1021/acs.jmedchem.0c01063.
[24] B. Boras, R. M. Jones, B. J. Anson, D. Arenson, M. A. Bakowski, N. Beutler, J. Binder, E. Chen, H. Eng, J. Hammond, R. Hoffman, E. P. Kadar, R. Kania, M. G. Kirkpatrick, L. Lanyon, E. K. Lendy, J. R. Lillis, S. A. Luthra, C. Ma, S. Noell, R. S. Obach, M. N. O. Brien, R. O'Connor, K. Ogilvie, D. Owen, M. Pettersson, M. R. Reese, T. F. Rogers, M. I. Rossulek, J. G. Sathish, C. Steppan, L. W. Updyke, Y. Zhu, J. Wang, A. K. Chatterjee, A. S. Anderson, C. Allerton, **n.d.**, 32.
[25] S. Günther, P. Y. A. Reinke, Y. Fernández-García, J. Lieske, T. J. Lane, H. M. Ginn, F. H. M. Koua, C. Ehrt, W. Ewert, D. Oberthuer, O. Yefanov, S. Meier, K. Lorenzen, B. Krichel, J.-D. Kopicki, L. Gelisio, W. Brehm, I. Dunkel, B. Seychell, H. Gieseler, B. Norton-Baker, B. Escudero-Pérez, M. Domaracky, S. Saouane, A. Tolstikova, T. A. White, A. Hänle, M. Groessler, H. Fleckenstein, F. Trost, M. Galchenkova, Y. Gevorkov, C. Li, S. Awel, A. Peck, M. Barthelmess, F. Schluenzen, P. L. Xavier, N. Werner, H. Andaleeb, N. Ullah, S. Falke, V. Srinivasan, B. A. França, M. Schwinzer, H. Brognaro, C. Rogers, D. Melo, J. J. Zaitseva-Doyle, J. Knoska, G. E. Peña-Murillo, A. R. Mashhour, V. Hennicke, P. Fischer, J. Hakanpää, J. Meyer, P. Gribbon, B. Ellinger, M. Kuzikov, M. Wolf, A. R. Beccari, G. Bourenkov, D. von Stetten, G. Pompidor, I. Bento, S. Panneerselvam, I. Karpics, T. R. Schneider, M. M. Garcia-Alai, S. Niebling, C. Günther, C. Schmidt, R. Schubert, H. Han, J. Boger, D. C. F. Monteiro, L. Zhang, X. Sun, J. Pletzer-Zelgert, J. Wollenhaupt, C. G. Feiler, M. S. Weiss, E.-C. Schulz, P. Mehrabi, K. Karničar, A. Usenik, J. Loboda, H. Tidow, A. Chari, R. Hilgenfeld, C. Uetrecht, R. Cox, A. Zaliani, T. Beck, M. Rarey, S.

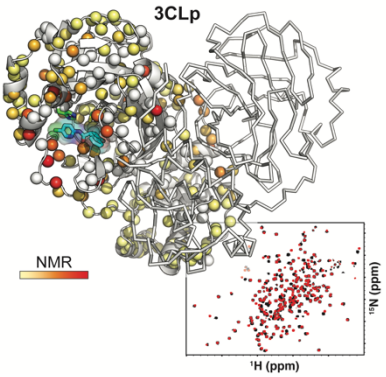
Günther, D. Turk, W. Hinrichs, H. N. Chapman, A. R. Pearson, C. Betzel, A. Meents, *Science* **2021**, DOI 10.1126/science.abf7945.
[26] N. Kitamura, M. D. Sacco, C. Ma, Y. Hu, J. A. Townsend, X. Meng, F. Zhang, X. Zhang, M. Ba, T. Szeto, A. Kukuljac, M. T. Marty, D. Schultz, S. Cherry, Y. Xiang, Y. Chen, J. Wang, *J. Med. Chem.* **2021**, DOI 10.1021/acs.jmedchem.1c00509.
[27] A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keeley, G. M. Keserü, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O'Brien, A. Powell, R. N. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft, M. A. Walsh, *Nature Communications* **2020**, *11*, DOI 10.1038/s41467-020-18709-w.
[28] D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, H. Jhoti, *Nat Rev Drug Discov* **2016**, *15*, 605–619.
[29] D. W. Kneller, G. Phillips, K. L. Weiss, S. Pant, Q. Zhang, H. M. O'Neill, L. Coates, A. Kovalevsky, *Journal of Biological Chemistry* **2020**, *295*, 17365–17373.
[30] V. Mody, J. Ho, S. Wills, A. Mawri, L. Lawson, M. C. C. J. C. Ebert, G. M. Fortin, S. Rayalam, S. Taval, *Commun Biol* **2021**, *4*, 93.
[31] M. Sencanski, V. Perovic, S. B. Pajovic, M. Adzic, S. Paessler, S. Glisic, *Molecules* **2020**, *25*, DOI 10.3390/molecules25173830.
[32] J. Breidenbach, C. Lemke, T. Pillaiyar, L. Schäkel, G. A. Hamwi, M. Diett, R. Gedschold, N. Geiger, V. Lopez, S. Mirza, V. Namasivayam, A. C. Schiedel, K. Sylvester, D. Thimm, C. Vielmuth, L. P. Vu, M. Zyulina, J. Bodem, M. Gütschow, C. E. Müller, *Angewandte Chemie International Edition* **2021**, *60*, 10423–10429.
[33] M. A. Walsh, J. M. Grimes, D. I. Stuart, *Biochemical and Biophysical Research Communications* **2021**, *538*, 40–46.
[34] J. Qiao, Y.-S. Li, R. Zeng, F.-L. Liu, R.-H. Luo, C. Huang, Y.-F. Wang, J. Zhang, B. Quan, C. Shen, X. Mao, X. Liu, W. Sun, W. Yang, X. Ni, K. Wang, L. Xu, Z.-L. Duan, Q.-C. Zou, H.-L. Zhang, W. Qu, Y.-H.-P. Long, M.-H. Li, R.-C. Yang, X. Liu, J. You, Y. Zhou, R. Yao, W.-P. Li, J.-M. Liu, P. Chen, Y. Liu, G.-F. Lin, X. Yang, J. Zou, L. Li, Y. Hu, G.-W. Lu, W.-M. Li, Y.-Q. Wei, Y.-T. Zheng, J. Lei, S. Yang, *Science* **2021**, *371*, 1374–1378.
[35] D. Suárez, N. Díaz, *J. Chem. Inf. Model.* **2020**, *60*, 5815–5831.
[36] D. W. Kneller, G. Phillips, H. M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates, A. Kovalevsky, *Nature Communications* **2020**, *11*, DOI 10.1038/s41467-020-16954-7.
[37] S. Zhang, N. Zhong, X. Ren, C. Jin, B. Xia, *Biomolecular NMR Assignments* **2011**, *5*, 143–145.
[38] J. Shi, J. Song, *FEBS Journal* **2006**, *273*, 1035–1045.
[39] A. L. Kantsadi, E. Cattermole, M.-T. Matsoukas, G. A. Spyroulias, I. Vakonakis, *J Biomol NMR* **2021**, *75*, 167–178.
[40] B. Goyal, D. Goyal, *ACS Combinatorial Science* **2020**, *22*, 297–305.
[41] J. Shi, J. Sivaraman, J. Song, *Journal of Virology* **2008**, *82*, 4620–4629.
[42] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discovery Today* **2003**, *8*, 876–877.
[43] D. Valenti, J. F. Neves, F.-X. Cantrelle, S. Hristeva, D. L. Santo, T. Obšil, X. Hanoulle, L. M. Levy, D. Tzalis, I. Landrieu, C. Ottmann, *Med. Chem. Commun.* **2019**, DOI 10.1039/C9MD00215D.
[44] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *Journal of biomolecular NMR* **2001**, *21*, 349–359.
[45] **N.d.**
[46] D. Bajusz, W. S. Wade, G. Satała, A. J. Bojarski, J. Ilaš, J. Ebner, F. Grebien, H. Papp, F. Jakab, A. Douangamath, D. Fearon, F. von Delft, M. Schuller, I. Ahel, A. Wakefield, S. Vajda, J. Gerencsér, P. Pallai, G. M. Keserű, *Nat Commun* **2021**, *12*, 3201.
[47] C. W. Murray, D. C. Rees, *Nature Chem* **2009**, *1*, 187–192.
[48] M. Macchiagodena, M. Pagliai, P. Procacci, *Chemical Physics Letters* **2020**, *750*, 137489.
[49] D. Ni, Y. Li, Y. Qiu, J. Pu, S. Lu, J. Zhang, *Trends in Pharmacological Sciences* **2020**, *41*, 336–348.
[50] S. Lu, Y. Qiu, D. Ni, X. He, J. Pu, J. Zhang, *Drug Discovery Today* **2020**, *25*, 177–184.
[51] C. Liu, S. Boland, M. D. Scholle, D. Bardiot, A. Marchand, P. Chaltin, L. M. Blatt, L. Beigelman, J. A. Symons, P. Raboisson, Z. A. Gurard-Levin, K. Vandyck, J. Deval, *Antiviral Res* **2021**, *187*, 105020.
[52] A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees, C. H. Reynolds, *Nat Rev Drug Discov* **2014**, *13*, 105–121.

10.1002/anie.202109965

WILEY-VCH

RESEARCH ARTICLE

Entry for the Table of Contents



We herein report the liquid-sate NMR spectroscopy analysis of the dimeric SARS-CoV-2 main protease (3CLp), including its backbone assignments, to study its complex conformational regulation. Using fragment-based NMR screening, we highlighted three hotspots on the protein, two in the substrate binding pocket and one at the dimer interface, and we identified a non-covalent reversible inhibitor of 3CLp that has antiviral activity in infected cells.