Two ligand-binding sites on SARS-CoV-2 non-structural protein 1 revealed by fragment-based x-ray screening

Shumeng Ma[1], Shymaa Damfo[1], Jiaqi Lou[1], Nikos Pinotsis[2], Matthew W. Bowler[3], Shozeb Haider[1,4] and Frank Kozielski[1]

[1]School of Pharmacy, University College London, 29-39 Brunswick Square, London WC1N 1AX, United Kingdom
[2]Institute of Structural and Molecular Biology, Birkbeck College, WC1E 7HX London, United Kingdom
[3]European Molecular Biology Laboratory, Grenoble, France
[4]UCL Centre for Advanced Research Computing, University College London, WC1H 9RN

Corresponding author: Frank Kozielski, Department of Pharmaceutical and Biological Chemistry, School of Pharmacy, University College London, 29-39 Brunswick Square, London WC1N 1AX, UK

Abbreviations: CoV, Coronavirus; EDTA, Ethylenediaminetetraacetic Acid;; IPTG, Isopropyl-d-thiogalactopyranoside; MERS, Middle East Respiratory Syndrome; NSP, Nonstructural protein; ORF, Open Reading Frame; PMSF, Phenylmethylsulfonylfluorid; SDS-PAGE, Sodium Dodecyl Sulphate–polyacrylamide Gel Electrophoresis; SARS, Severe Acute Respiratory Syndrome; CoV, Coronaviruses; SARS-CoV, Severe Acute Respiratory Syndrome Coronavirus; MERS-CoV, Middle East Respiratory Syndrome Coronavirus;; SAR, Structure activity relationship; IFN, Interferon; MR, Molecular replacement; GS, Ground state; DMSO, Dimethyl Sulfoxide; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid; LB, Terrific Broth; TB, Terrific Broth; SOC, Super Optimal Broth; MST, Microscale Thermophoresis; TSA, Thermal Shift Assay; SUMO, Small Ubiquitin-like Modifier; His-tag, Polyhistidine-tag; β-ME, β-mercaptoethanol; ULP1, Ubiquitin-like-specific Protease 1; IPTG, Isopropylthio-β-galactoside; PDB, Protein Data Bank; COOT, Crystallographic Object-oriented Toolkit; PanDDA, Pan-Dataset Density Analysis; NanoDSF, Nanoscale Differential Scanning Fluorimetry; PEG, Polyethylene Glycol; PBS-T, 1X Phosphate-buffered Saline with 0.1% Tween; *E. coli*, *Escherichia coli*; SD, Standard Deviation; SL, Stem Loop; Pol α, DNA Polymerase α.

Keywords: SARS-CoV-2, Covid-19, non-structural protein 1, fragment-based x-ray screening, ligands, structure-based drug design, cross-inhibitory effects

Abstract

The regular reappearance of coronavirus (CoV) outbreaks over the past 20 years has caused significant health consequences and financial burdens worldwide. The most recent and still ongoing novel CoV pandemic, caused by Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) has brought a range of devastating consequences. Due to the exceptionally fast development of vaccines, the mortality rate of the virus has been curbed to a significant extent. However, the limitations of vaccination efficiency and applicability, coupled with the still high infection rate, emphasise the urgent need for discovering safe and effective antivirals against SARS-CoV-2 through suppressing its replication and or attenuating its virulence. Non-structural protein 1 (nsp1), a unique viral and conserved leader protein, is a crucial virulence factor for causing host mRNA degradation, suppressing interferon (IFN) expression and host antiviral signalling pathways. In view of the essential role of nsp1 in the CoV life cycle, it is regarded as an exploitable target for antiviral drug

discovery. Here, we report a variety of fragment hits against SARS-CoV-2 nsp1 identified by fragment-based screening via X-ray crystallography. We also determined the structure of nsp1 at atomic resolution (0.95 Å). Binding affinities of hits against nsp1 were determined by orthogonal biophysical assays such as microscale thermophoresis and thermal sift assays. We identified two ligand-binding sites on nsp1, one deep and one shallow pocket, which are not conserved between the three medially relevant SARS, SARS-CoV-2 and MERS coronaviruses. Our study provides an excellent starting point for the development of more potent nsp1-targeting inhibitors and functional studies on SARS-CoV-2 nsp1.

Introduction

In the past two decades, coronaviruses (CoVs) have posed a constant threat to humans in various forms, from Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), the etiological agent for the outbreak of severe acute respiratory syndrome (SARS) in 2002 [1], to the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in 2012 [2], and the most recent novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), leading to the ongoing coronavirus pandemic first emerging at the end of 2019 in Wuhan, China [3]. The high infectivity and fatality of CoVs have caused great public concern [4] and its worldwide impact is causing a huge financial burden [5-7]. These events stress the importance of timely precaution and effectively curative strategies such as vaccination and the development of antiviral agents, which significantly curbed the effects of the disease [8, 9]. However, the vaccination efficacy partially depends on the physical [10] and psychological [11] conditions of inoculators and is challenged by rapidly appearing variants of the virus [12, 13]. Additionally, there is a large group of people, e.g. pregnant women, young children, people having certain comorbidities, who are not recommended to become vaccinated due to their medical condition [14]. With respect to the repurposing of drugs and the development of new agents, their efficiency is far from satisfactory [15]. Therefore, the discovery of novel anti-viral drugs specific to SARS-CoV-2 is urgently required.

The CoV genome is composed of a single, 5′-capped and 3′-polyadenylated RNA [16]. The first two-thirds of the genome encode two overlapping open reading frames (ORFs), ORF1a and ORF1b that produce polyproteins 1a and 1ab in a ribosomal frameshift. Downstream of this region, the remaining one-third of the genome encodes structural proteins (spike, envelope, membrane, and nucleocapsid) and several ORFs that produce the accessory proteins [17]. The polyproteins 1a and 1ab are then processed post-translationally by a virus-encoded protease, known as non-structural protein 5 (nsp5 or main protease) to produce other non-structural proteins (nsps) [18].

Among these proteins non-structural protein 1 (nsp1) encoded by the gene closest to the 5′ end of the viral genome received extensive attention. Nsp1 proteins of β-CoVs display similar biological functions in suppressing host gene expression [19, 20] and inhibiting the innate immune response to virus infection [18, 19]. SARS-CoV and SARS-CoV-2 nsp1 induce a near-complete shutdown of host protein translation by a three-pronged strategy: First, it binds to the small ribosomal subunit and stalls canonical mRNA translation at various stages [21, 22]. Second, nsp1 binding to the host ribosome leads to endonucleolytic cleavage and subsequent degradation of host mRNAs [21]. The interaction between nsp1 and a conserved region in the 5′ untranslated region (UTR) of viral mRNA prevents the shutdown of viral protein expression as a self-protection mechanism [23, 24]. Third, nsp1 interacts with heterodimeric host messenger RNA (mRNA) export receptor NXF1-NXT1, which is responsible for nuclear export of cellular mRNAs, causing a significant number of cellular mRNAs to be retained in the nucleus during infection [25]. Likewise, MERS-CoV nsp1 also exhibits a conserved function to negatively regulate host gene expression, despite employing

a different strategy that selectively inhibits the host translationally competent mRNAs [26]. After infection, nsp1 also contributes to the evasion of the innate anti-viral immune response, mainly by antagonising interferon (IFN) induction and downstream signalling. In view of the crucial roles of nsp1 in the replication and virulence of CoVs, it represents an attractive target for both vaccine development and drug discovery.

The crystal structure of the N-terminal domain of SARS-CoV-2 nsp1 has been determined to 1.65 Å resolution (PDB entry 7K3N) [27] and the small C-terminal domain in complex with the human 40S ribosome (PDB entry 7K5I) [24] has been determined by cryo-electron microscopy providing useful structural information although the entire nsp1 structure could not be determined, probably due to its flexible linker region connecting these two distinct domains.

In this project we optimised crystallisation conditions and determined the structure of the N-terminal domain of nsp1 to atomic resolution of 0.95 Å, which served as a basis for fragment-based screening via x-ray crystallography. We identified two ligand binding pockets on nsp1, which bind a range of fragments. The fragment hits were further characterised by orthogonal biophysical assays. These fragments will serve as a basis for structure-based drug design.

Results and discussion

Structure determination of nsp1 to atomic resolution

SARS-CoV-2 is a protein of 180 residues consisting of an N-terminal domain (residues 1 to 128), followed by a linker region (residues 129 to 148) and a smaller C-terminal domain (residues 149 to 180). We generated an expression construct covering almost the entire globular domain (residues 10 to 126) (Figure 1A). A DNA insert coding for residues 10 to 126 of SARS-CoV-2 nsp1, named $nsp1_{10\text{-}126}$ throughout the manuscript, was codon optimised for expression in *E. coli* and the protein was expressed and purified. $Nsp1_{10\text{-}126}$ was subjected to crystallisation trials using published crystallisation conditions[27], but this did not yield any crystals. We therefore tested a variety of commercial crystallisation screens, which yielded three types of crystals with distinct morphology (supplementary material section Table S1). Crystals were cryo-cooled in liquid nitrogen, tested for diffraction quality and data were processed, all yielding the same space group $P4_32_12$ with very similar cell parameters (a=b=36.81 Å, c=140.96 Å, α=β=γ=90°) and identical to the published space group[27].

The structure of $Nsp1_{10\text{-}126}$ was determined by molecular replacement (MR) and refined to atomic resolution at 0.95 Å. Data collection and refinement statistics are summarised in Table 1. The asymmetric unit contains one molecule of nsp1. The model covers the sequence from Glu10 to Asn126. Currently, it is the highest resolution structure of a CoV Nsp1.

The structure of SARS-CoV-2 $nsp1_{10\text{-}126}$ features a unique topological arrangement resulting in the formation of a seven-stranded (n = 7) β-barrel that is primarily antiparallel, except for strands β1 (His13 – Val20) and β2 (Cys51 – Val54) (Figures 1B and 1C). Additional major structural features include helix α1 (Val35 – Asp48), which is positioned as a cap along one opening of the β-barrel, three 3_{10} helices that run parallel to each other, and the strand β5 (ILe95 – Tyr97), which is not part of the β-barrel but forms a β-sheet interaction with the β4 strand (Val84 – Leu92). Compared to the recently published structures determined at 1.6 Å and 1.77 Å resolution (PDB entries 7K3N and 7K7P) [27] [28], our atomic structure identified a third 3_{10} helix in the flexible loop region between β3 and β4 (residues 80 to 82). As the density indicates that this (loop) region may have various conformations, we tentatively fitted one conformation into the density. In addition, our structure also reveals a new strand, β6, and shows a longer strand β5 than the previously

published structures. This more complete model refined to higher resolution provides a more accurate structural picture of nsp1 and will serve as an excellent starting point for fragment-based screening.

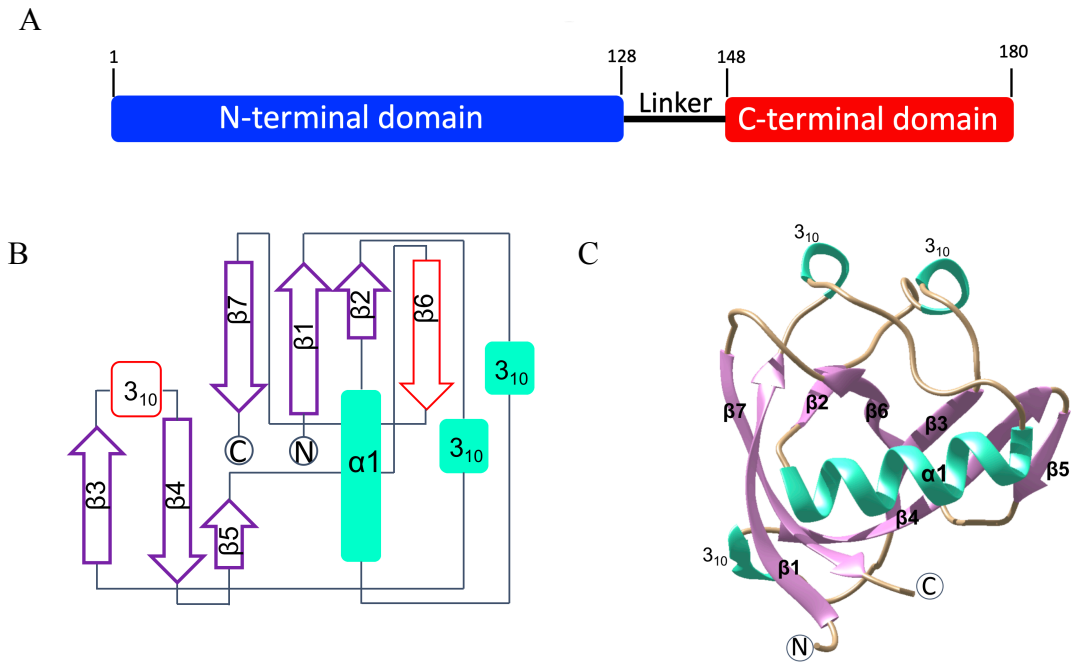


Figure 1. The structure of SARS-CoV-2 nsp1$_{10\text{-}126}$. A) Bardiagram of the nsp1 domain arrangement, including the N-terminal domain (blue), the flexible linker region (black) and the C-terminal domain (red). B) Topological arrangement of SARS-CoV-2 nsp1$_{10\text{-}126}$ at high resolution, where newly identified structural features are coloured in red. C) Cartoon representation of the structure. The secondary structure elements are depicted in different colours in the right panel with α-helixes coloured in green, β-strands shaded in purple, and loops are shown in tan.

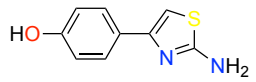
Table 1. Data collection, structure determination and refinement statistics for SARS-CoV-2 nsp1$_{10\text{-}126}$. Data in parenthesis correspond to the highest resolution shell.

	SARS-CoV-2 nsp1$_{10\text{-}126}$
Wavelength [Å]	0.82656
Resolution range [Å]	32.63 - 0.95 (0.984 - 0.95)
Space group	$P4_32_12$
Unit cell parameters [Å, °]	36.81, 36.81, 140.96, 90, 90, 90
Molecules per asymmetric unit	1
Total reflections	469243 (66169)
Unique reflections	62200 (9745)
Multiplicity	7.5 (6.8)
Completeness [%]	99.4 (98.2)
Mean I/sigma(I)	11.31 (0.51)
Wilson B-factor (Å^2)	12.1
R_{meas} [%]	6.4 (309.0)

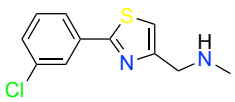
$CC_{1/2}$	0.998 (0.172)
Reflections used in refinement	61563 (5572)
R_{cryst} / R_{free} [%]	16.9 (36.1) / 18.8 (38.1)
Total no. of non-hydrogen atoms (protein)	1088
No. of protein/ligand/solvent residues	117 / 0 / 140
RMSD bonds length, bond angles [Å, °]	0.012 / 1.44
Ramachandran favoured/allowed/outliers/rotamer outliers [%]	97.4 / 2.6 / 0/0 / 0/0
Clashscore	8.3
Average B-factor/protein/ligands/solvent	23.4 / 21.2 / 38.5

Fragment screening via X-ray crystallography

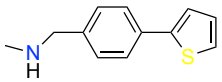
Fragment screening was performed by soaking 584 fragments from the Maybridge Ro3 library into $nsp1_{10-126}$ crystals followed by validation through X-ray diffraction experiments in fully automated mode at beamline MASSIF-1 [29, 30]. In addition, 40 nsp1 crystals without fragments were prepared to construct the Ground State (GS) model for data analysis in the multi-crystal software system PanDDA [31]. The overall resolution of data sets for the GS was 1.3 Å. 398 fragments were accepted for PanDDA analysis. Datasets that were rejected for the high R_{free} values (14) were reprocessed and manually inspected. Potential fragment hits were discovered by PanDDA (34), and nine of them were verified in the single crystal system by manual inspection in COOT [32] followed by refinement in Phenix [33]. These fragment hits can be defined in two groups binding to two distinct ligand binding sites named binding pocket 1 and binding pocket 2. Here we characterise five selected fragment hits. Data collection and refinement statistics for SARS-CoV-2 $nsp1_{10-126}$-fragment complexes are summarized in Table S2 of the supplementary material section and the chemical structures of the fragment hits are shown in Figure 2. Fragments **10B6**, **11C6** and **5E11** represent hits in which a phenyl group is connected to a 5-membered heterocyclic ring system. The second group of fragments represented by **7H2** and **8E6** contain a single phenyl ring system containing single or double substitutions.



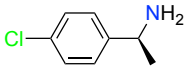
4-(2-aminothiazol-4-yl)phenol
10B6
1



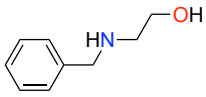
1-(2-(3-chlorophenyl)thiazol-4-yl)
-N-methylmethanamine
11C6
2



N-methyl-1-(4-(thiophen-2-yl)phenyl)methanamine
5E11
3



(*S*)-1-(4-chlorophenyl)ethan-1-amine
7H2
4



2-(benzylamino)ethan-1-ol
8E6
5

Figure 2. Chemical structures, systematic names, plate positions and numbering of fragment hits binding to two distinct binding sites in SARS-CoV-2 nsp1$_{10\text{-}126}$.

Identification of ligand binding sites in SARS-CoV-2 nsp1

By overlaying all nsp1-fragment complexes, we could identify two distinct ligand-binding pockets. The majority of fragment hits, 3, bind to ligand binding site I, whereas 2 of the 5 hits bind to the shallower binding site II (Figure 3a). Ligand binding site I involves the N- and C-termini, strands β1 and β7 as well as residues of helix α1. Ligand-binding site II is formed by nsp1 and one of its symmetry mates involving the loop region between the first 3_{10} helix and helix α1, two residues of α1 and residues in the loop connecting strands β6 and β7. Residues from the symmetry mate involved in binding are located in and close to the first 3_{10} helix and in the second 3_{10} helix. The shortest distance between the two binding pockets is about 18 Å and seems to be challenging to bridge for fragment linking strategies. The characteristics of these two pockets are summarised in Table 2.

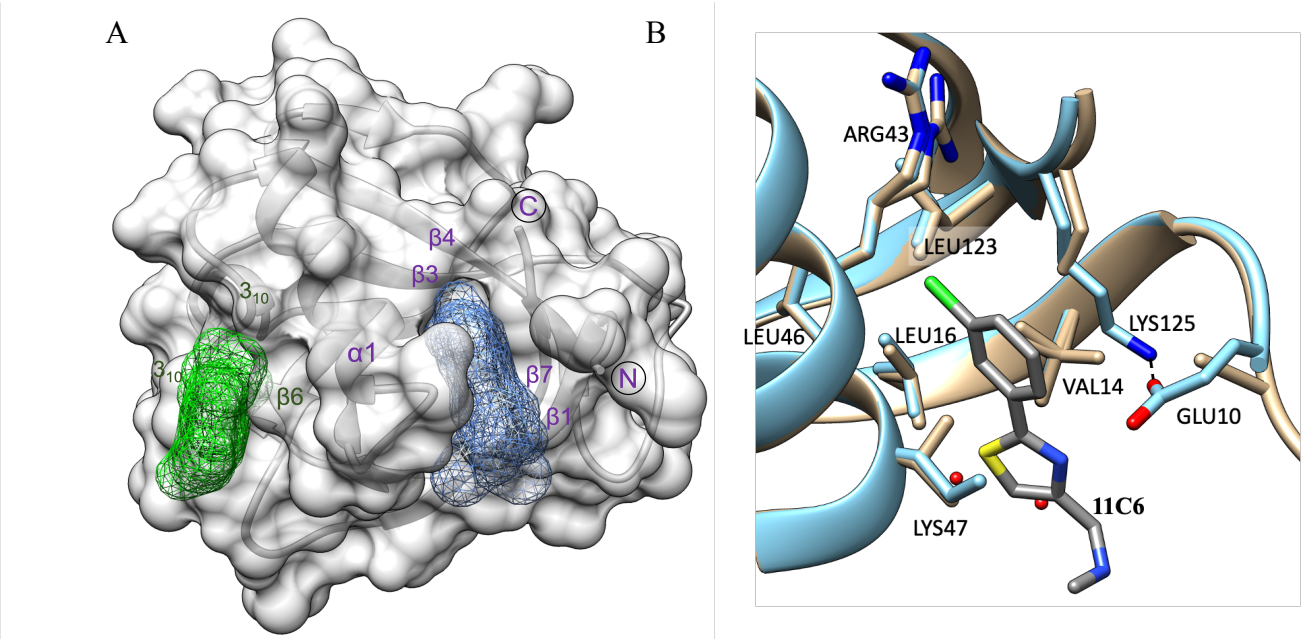


Figure 3. Binding sites I and II in SARS-CoV-2 nsp1$_{10\text{-}126}$. A) Surface representation of SARS-CoV-2 nsp1$_{10\text{-}126}$ shown in light grey indicating binding sites I (blue) and II (green). The secondary structures forming the two binding pockets are labelled in the corresponding colours with α1 contributing to the formation of both pockets. B) Overlay between native nsp1 (beige) and the SARS-CoV-2 nsp1$_{10\text{-}126}$-11C6 complex (blue) showing the structural changes induced by fragment binding.

Table 2. List of SARS-CoV-2 nsp1$_{10\text{-}126}$ amino acids defining binding sites I and II.

Binding site	Fragments	Residues
I	10B6, 11C6, 5E11	Glu10, Val14, Leu16, Arg43, Leu46, Lys47, Leu123 and Lys125
II	7H2, 8E6	Val28, Phe31, Glu41, His45, Pro109, Val111; Symmetry mate: Arg24, Arg29, Pro62, Gln63

Fragment hits binding to pocket I induce small structural rearrangements

Having high-resolution native and fragment-bound structures allows us to discuss ligand-induced structural changes in detail. Fragment **11C6** was chosen as an example. With respect to binding site I, Glu10, Lys47 and Lys125 show the most obvious changes. Whereas in the native atomic structure the side chain of Lys125 is only partially visible, it is fully visible in the ligand bound structure, as it runs parallel to the fragment and forms hydrophobic interactions with one of the phenol rings and displays a hydrogen bond interaction with Glu10 (Figure 3b). In the fragment bound structure the Glu10 side chain density is therefore mostly visible, while it is absent in the native structure. The density for the Lys47 side chain is also only partially visible in both native and fragment bound structures but some positive density in the native structure suggests one conformation extending toward the space, which the fragment would occupy. In the native structure, two water molecules are located in, or close to, the space otherwise occupied by the 3-chlorophenyl group of the ligand, but these are absent in the fragment bound structure. In conclusion, binding of fragments to this site leads to a stabilisation of surrounding residues indicating a small but noticeable induced fit mechanism caused by fragment binding. In contrast, binding of fragment **7H2** to binding site II does not induce any significant structural changes in its surrounding environment (data not shown).

Detailed description of SARS-CoV-2 nsp1$_{10\text{-}126}$-fragment interactions

Fragments hits **10B6**, **11C6** and **5E11** are chemically related as they contain a substituted phenyl ring system connected to a five-membered heterocyclic thiophen or thiazol ring system. All three bind to ligand binding site I, establishing various interactions.

10B6 establishes various hydrogen bond interactions with a range of residues of binding pocket I (Figure 4a). The 2-amino group establishes a hydrogen bond interaction with a water molecule. Fragment **11C6** maintains the overall binding conformation in the same pocket (Figure 4b). However, this fragment contains a chloride substituent in the *meta*-position instead of the *para*-substituted hydroxyl group. Binding occurs exclusively through hydrophobic interactions, but compared to **10B6, 11C6** displays significantly less interaction with nsp1. Interestingly, the orientation for fragment **5E11** is reversed whereby the 6- and 5-membered rings switch positions (Figure 4c). Overall, many hydrophobic interactions of the fragment hit with residues in binding pocket I are maintained, but the nitrogen of the Lys47 side chain now establishes a hydrogen bond interaction with the sulphur of the thiophen ring system, while a N-methyl methanamine substituent on the 6-membered ring system does not establish a hydrogen bond interaction. This example shows the importance of high-quality and high-resolution structural information as mostly flat fragments can be easily fit into the electron density in various orientations and a wrong orientation of the ligand in subsequent SAR studies could easily lead to incorrect conclusions.

Fragments **7H2** and **8E6** bind to a distinct site on SARS-CoV-2 nsp1$_{10\text{-}126}$ named binding pocket II (Figure 4 d-g). The common feature of these two hits is that they both possess a single substituted phenyl ring system.

7H2 displays various hydrophobic interactions with residues of SARS-CoV-2 nsp1$_{10\text{-}126}$ (Figure 4d). The ethan-1-amine group establishes a hydrogen-bond interaction through a water molecule with the main chain oxygen of Phe31. The opposite chloride substituent does not seem to contribute to binding to SARS-CoV-2 nsp1$_{10\text{-}126}$. It is noticeable that compared to fragments binding to pocket I, **7H2** has a reduced number of interactions with residues of the protein, probably because this ligand binding site II is shallower than pocket I. Compared to **7H2**, **8E6** is a mono-substituted fragment lacking the chloride substituent but carrying an amin-1-ethan-1-ol substituent instead of the ethan1-1-amine (Figure 4e). As the data for this

complex were particularly good, with a resolution of 1.18 Å, we placed the fragment in all possible orientations into the electron density and conducted refinement for six distinct orientations. The best refinement was obtained in which the substituent points towards the solvent, establishing a hydrogen-bond interaction with the side chains of Glu41 and His45 through a water molecule. Despite essentially observing electron density for only part of the fragment, **8E6** also establishes hydrophobic interactions with Pro109 and Val111, similar to **7H2**. In contrast to **10B6**, **11C6** and **5E1, 7H2** (Figure 4f) and **8E6** (Figure 4g) also establish interactions with a symmetry mate generating a set of novel interactions, which requires orthogonal assays to establish if there is a "real" binding site, or if these only bind in the presence of the arrangement dictated by the crystal.

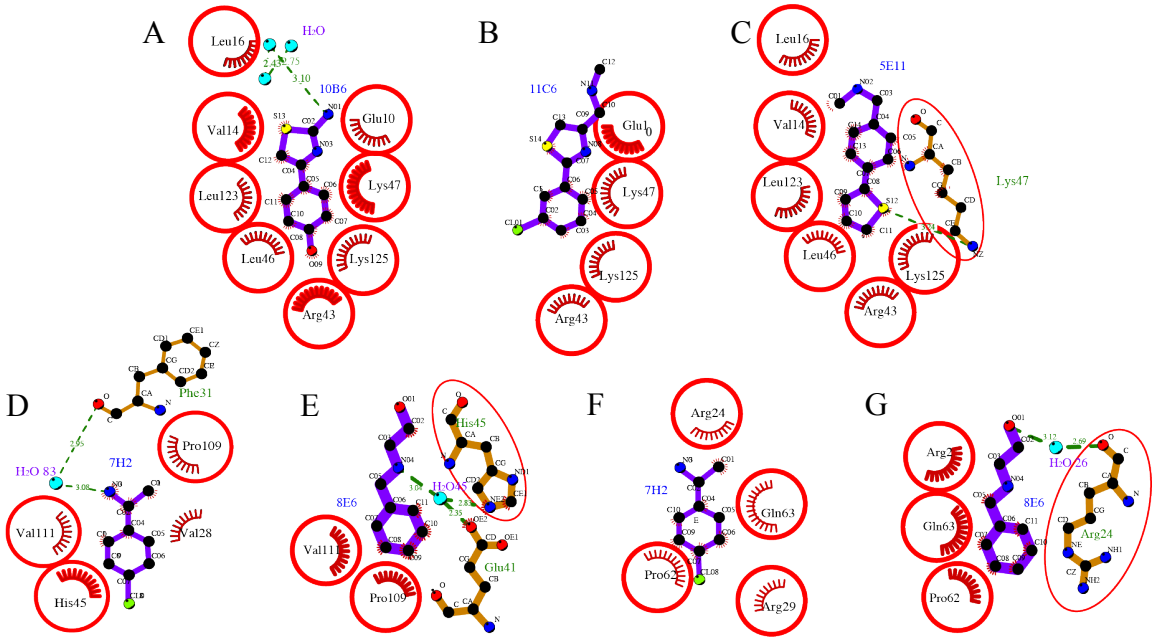


Figure 4. Molecular interactions of fragment hits binding to SARS-CoV-2 nsp1$_{10\text{-}126}$. Hydrophobic interactions are displayed by red half moons and identical residues are shown with a red circle. Hydrogen-bond interactions are shown with a dotted green line. Three fragment hits (A) **10B6**, (B) **11C6** and (C) **5E11** interacting with the residues of SARS-CoV-2 nsp1$_{10\text{-}126}$ in binding pocket I. The binding conformation is switched in **5E11**, in which the 5- and 6-membered ring system exchange their positions compared to **10B6** and **11C6**. Fragment hits **7H2** and **8E6** interacting with the residues located in binding site II. **7H2** binding to D) SARS-CoV-2 nsp1$_{10\text{-}126}$ and F) a symmetry mate. **8E6** interacting with E) SARS-CoV-2 nsp1$_{10\text{-}126}$ and G) a symmetry-related molecule. The substituent of **8E6** is not visible in the electron density.

Orthogonal biophysical assays to further validate fragment hits

To further validate SARS-CoV-2 nsp1$_{10\text{-}126}$-targeting fragment hits, we employed two orthogonal biophysical assays, thermal shift assays (TSA) and microscale thermophoresis (MST). For MST experiments, the binding affinities between SARS-CoV-2 nsp1$_{10\text{-}126}$ and fragment hits were evaluated using the Monolith NT. 115 instrument with the pico-red fluorescent channel. The MST results of the fragments are summarised in Table 3. The K_d values vary between 475 µM and larger than 20 mM. Interestingly, fragment hits binding in the "deeper" binding site systematically show better K_d values than the two hits binding in the

shallower site II which display weaker K_d values. Crystallographic analysis of the binding site for **7H2** and **8E6** revealed that a symmetry mate contributed to the binding of the two fragments to nsp1, raising the possibility that both proteins in the particular crystallographic arrangement are required for binding and that these two fragments may not bind to nsp1 in solution. However, the calculated K_d values clearly show that binding to a SARS-CoV-2 $nsp1_{10-126}$ monomer in solution is sufficient.

For TSA, nsp1 in the presence of an equal concentration of DMSO was used as a control to offset the influence of DMSO on the protein. The averaged change of inflection temperature (ΔT_i) of the protein in the presence of fragments higher than three-fold of standard deviation ($|\Delta Ti|-3*SD>0$) was regarded as statistically significant. As can be seen from Table 3, except for **5E11**, which showed atypical unfolding curves, all fragment hits influenced the unfolding property of SARS-CoV-2 $nsp1_{10-126}$. A stabilisation effect was observed for **7H2**, while destabilisation effects were measured for **8E6**, **10B6** and **11C6**. In this regard, TSA may not be suitable yet for the validation of hits against SARS-CoV-2 $nsp1_{10-126}$, as we have recently also observed for another SARS-CoV-2 protein, nsp10 [34]. With respect to physicochemical properties, calculated values for fragments are well within the expected range, indicating good starting points for structure-based drug design.