

# Revealing druggable cryptic pockets in the Nsp1 of SARS-CoV-2 and other $\beta$ -coronaviruses by simulations and crystallography

Alberto Borsatto<sup>1,2</sup>, Obaeda Akkad<sup>1,2</sup>, Ioannis Galdadas<sup>1,2</sup>, Shumeng Ma<sup>3</sup>, Shymaa Damfo<sup>3</sup>, Shozeb Haider<sup>3,4</sup>, Frank Kozielski<sup>3</sup>, Carolina Estarellas<sup>5\*</sup>, Francesco Luigi Gervasio<sup>1,2,6,7\*</sup>

<sup>1</sup>School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland; <sup>2</sup>ISPSO, University of Geneva, Geneva, Switzerland; <sup>3</sup>School of Pharmacy, University College London, London, United Kingdom; <sup>4</sup>UCL Centre for Advanced Research Computing, University College London, London, United Kingdom; <sup>5</sup>Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Sciences, and Institute of Theoretical and Computational Chemistry, University of Barcelona, Barcelona, Spain; <sup>6</sup>Chemistry Department, University College London, London, United Kingdom; <sup>7</sup>Institute of Structural and Molecular Biology, University College London, London, United Kingdom

**Abstract** Non-structural protein 1 (Nsp1) is a main pathogenicity factor of  $\alpha$ - and  $\beta$ -coronaviruses. Nsp1 of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) suppresses the host gene expression by sterically blocking 40S host ribosomal subunits and promoting host mRNA degradation. This mechanism leads to the downregulation of the translation-mediated innate immune response in host cells, ultimately mediating the observed immune evasion capabilities of SARS-CoV-2. Here, by combining extensive molecular dynamics simulations, fragment screening and crystallography, we reveal druggable pockets in Nsp1. Structural and computational solvent mapping analyses indicate the partial crypticity of these newly discovered and druggable binding sites. The results of fragment-based screening via X-ray crystallography confirm the druggability of the major pocket of Nsp1. Finally, we show how the targeting of this pocket could disrupt the Nsp1-mRNA complex and open a novel avenue to design new inhibitors for other Nsp1s present in homologous  $\beta$ -coronaviruses.

## Editor's evaluation

SARS-CoV-2 nonstructural protein (Nsp1) has emerged as an attractive target as it plays an important role in modulating the host and viral gene expression. This study describes multiple druggable sites in Nsp1. A 1.1Å co-crystal structure of Nsp1 with a fragment, together with computational studies, provides a framework for the rational design of potential antiviral candidates. This important study is methodologically convincing and will be of interest to researchers in the fields of structural virology and rational drug design.

## Introduction

Coronaviruses (CoVs) are the largest family of RNA viruses identified to date. CoVs are members of the subfamily Coronavirinae classified into four genera  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -CoV. Common human CoVs belong to the first two genera and include the 229E ( $\alpha$ -), NL63 ( $\alpha$ -), OC43 ( $\beta$ -), and HKU1 ( $\beta$ -).

### \*For correspondence:

cestarellas@ub.edu (CE);  
francesco.gervasio@unige.ch  
(FLG)

**Competing interest:** See page 21

**Funding:** See page 21

**Preprinted:** 20 May 2022

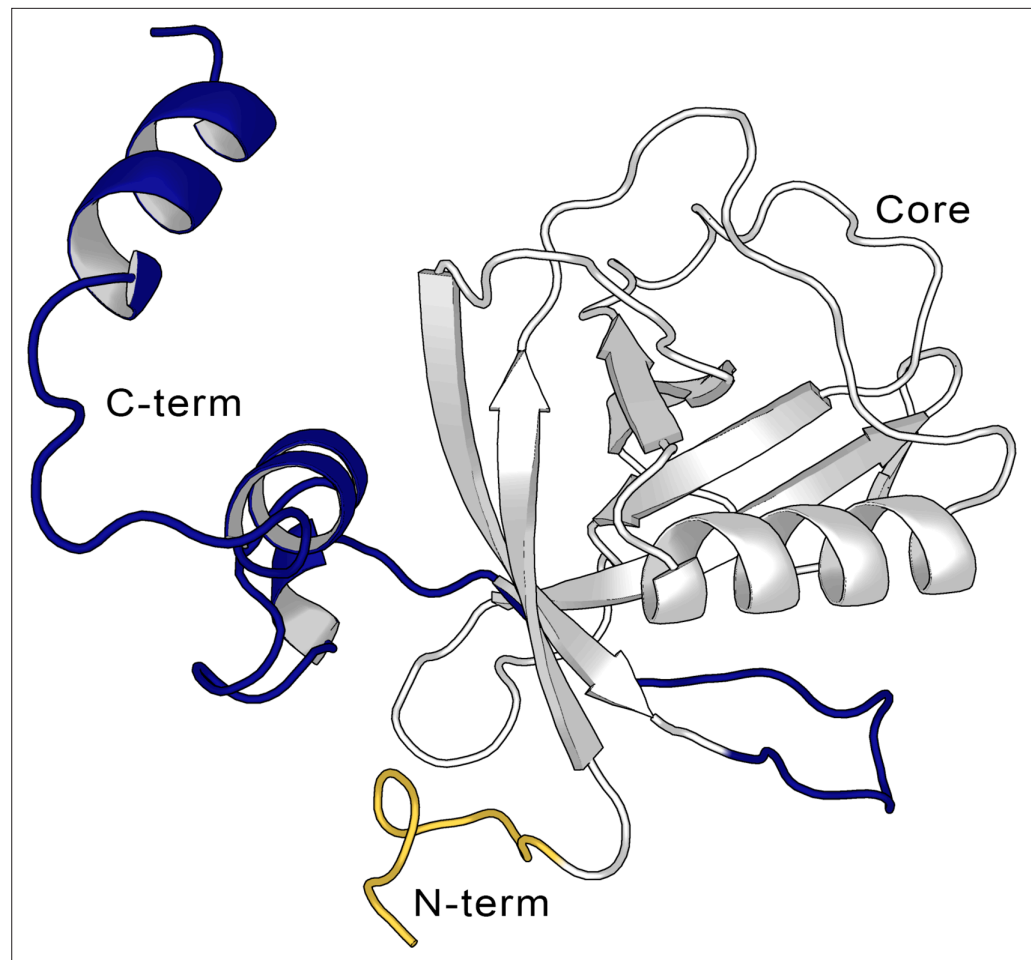
**Received:** 17 June 2022

**Accepted:** 06 November 2022

**Published:** 22 November 2022

**Reviewing Editor:** Yogesh K Gupta, The University of Texas Health Science Center at San Antonio, United States

© Copyright Borsatto et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.



**Figure 1.** Structure of the full-length SARS-CoV-2 Nsp1.

Cartoon representation of the full-length non-structural protein 1 (Nsp1) structure from the AlphaFold (Jumper *et al.*, 2021; Varadi *et al.*, 2022) model, showing the N-terminus (in yellow, aa Met1-Asn9), the Nsp1<sub>N</sub> core (in gray, aa Glu10-Asn126), and the C-terminus (blue, aa Gly127-Gly180).

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which led to the COVID-19 pandemic declared in March 2020 as the severe acute respiratory syndrome-related CoV, SARS-CoV-1, and the Middle East respiratory syndrome-related CoV, MERS-CoV, all belong to the  $\beta$ -CoV genera and are suggested to originate from bats (Andersen *et al.*, 2020; Li *et al.*, 2005; Corman *et al.*, 2014). The genome of different CoVs typically encodes four structural proteins, namely spike, envelope, membrane, and nucleocapsid, and two large polyproteins, pp1a and pp1ab, that are later cleaved into several non-structural proteins (Woo *et al.*, 2010). Due to the SARS-CoV-2 pandemic, significant efforts have been directed to the study and inhibition of these proteins such as the spike protein and the protease M<sup>pro</sup> (Owen *et al.*, 2021). Among the different proteins that are involved in the pathogenicity of SARS-CoV-2 is the non-structural protein 1 (Nsp1), a small 180 residue protein *Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020* whose function has been studied comparatively less than the rest of the pathogenic proteins. The structure of Nsp1 can be divided into three domains, a structured core, and two disordered domains corresponding to the N- and C-termini of the protein. The structured domain is composed of 117 residues (Glu10-Asn126) and displays an  $\alpha/\beta$ -fold. The disordered C-terminal domain (Gly127-Gly180) was shown to partially fold into a helix-loop-helix pattern when bound to the 40S ribosomal subunit (Figure 1; Schubert *et al.*, 2020).

It is worth noting that Nsp1 is only present in the  $\alpha$ - and  $\beta$ -, but not in the  $\gamma$ - and  $\delta$ -CoVs. The structural analysis of Nsp1<sub>N</sub> structures from  $\alpha$ - (PDB entries 3ZBD and 5XBC) and  $\beta$ - (PDB entry 2HSX) CoVs suggests that, despite the low sequence homology, the Nsp1 of the two genera displays a high

structural similarity, with root-mean-square deviation (RMSD) values ranging from 1.8 to 2.4 Å (**Shen et al., 2018; Min et al., 2020**). The structural similarity of the  $\alpha$ -CoVs and  $\beta$ -CoVs Nsp1 is reflected by their biological function as both are involved in the regulation of the host and viral gene expression. Multiple studies have shown that the expression of Nsp1 inhibits the translation in host cells by a combination of different mechanisms. For instance, Nsp1 sterically blocks the mRNA tunnel in the 40S ribosomal subunit (**Schubert et al., 2020**), the 43S preinitiation complex, and the non-translating 80S ribosomes (**Schubert et al., 2020; Thoms et al., 2020; Yuan et al., 2020**). Moreover, it has been shown that Nsp1 can also trigger the cleavage of the host mRNA and hinder its nuclear export to the cytosol (**Zhang et al., 2021**). These mechanisms concur in downregulating the translation-mediated innate immune response of the host cell, hence mediating the observed immune evasion capabilities of SARS-CoV-2 (**Thoms et al., 2020; Narayanan et al., 2008**). Interestingly, only host mRNAs are subjected to Nsp1-mediated endonucleolytic cleavage whereas viral mRNAs escape from this translation suppression mechanism. Experimental results have shown that the interaction of N-terminal Nsp1 with the stem loop 1 (SL1) at the viral mRNA 5' UTR region is necessary to avoid the Nsp1-mediated translation shutdown and cleavage in infected cells (**Tanaka et al., 2012; Vankadari et al., 2020**). The proposed mechanism suggests that Nsp1 plays the role of a ribosome gatekeeper by sterically blocking the mRNA tunnel of the ribosome for the host mRNAs and the blockage is lifted upon the interaction of Nsp1 with the 5'-UTR SL1 sequence of the viral mRNA. In this way, the virus can inhibit selectively the translation of the host mRNA and hijack the ribosome to foster the translation of its own mRNA (**Tidu et al., 2020; Mendez et al., 2021**).

The regulatory role of Nsp1 in viral replication and gene expression has also been demonstrated by mutations in the Nsp1 coding region of the transmissible gastroenteritis virus (TGEV, an  $\alpha$ -CoV infecting pigs) and the murine hepatitis virus (MHV, a  $\beta$ -CoV infecting mice) genomes. According to these mutagenesis studies, blocking the function of Nsp1 in different viruses leads to a drastic reduction or elimination of the infectious virus (**Shen et al., 2019**).

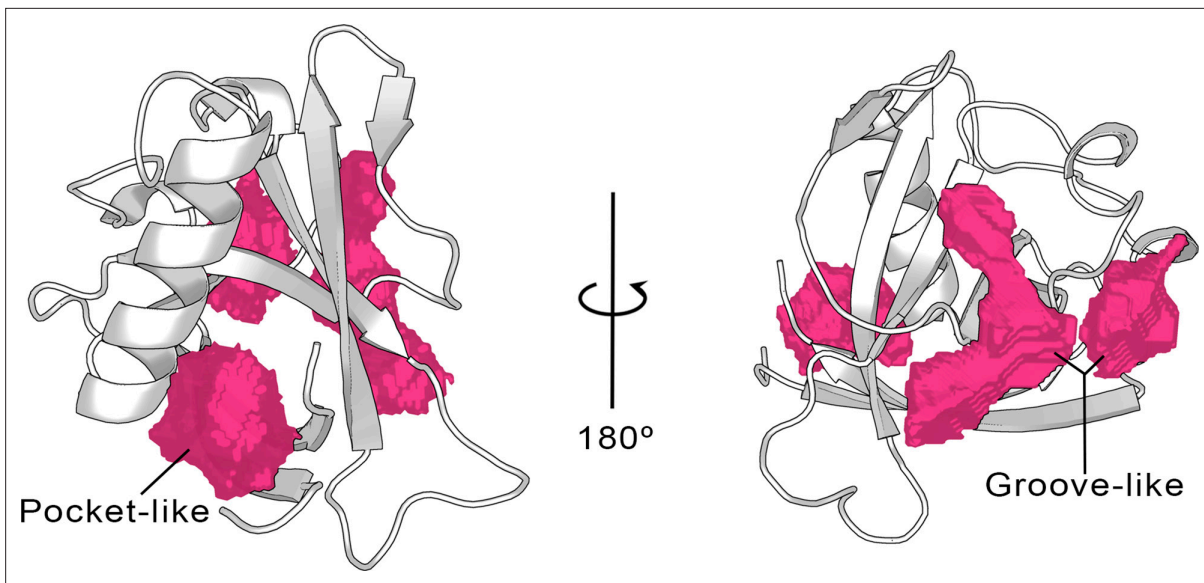
Altogether, the high structural similarity across different  $\alpha$ - and  $\beta$ -CoVs from different organisms, the fact that Nsp1 has no homologues outside of the CoVs, as well as its crucial role in mediating immune evasion, make Nsp1 a valuable target for developing antiviral drugs, not only for the ongoing COVID-19 pandemic but also to prevent future pandemic outbreaks caused by new variants. However, the largest folded domain of Nsp1, namely the N-terminal core region (Nsp1<sub>N</sub>), corresponds to a small compact domain that shows predominantly small superficial cavities, which complicates rational drug design efforts. In spite of Nsp1 being a validated target for therapeutic intervention, very few studies explored Nsp1 for structure-based drug discovery, and only one of them reported the binding of a ligand to the C-terminal domain of SARS-CoV-2 Nsp1 (**Afsar et al., 2022**). To date, no ligand-bound Nsp1 crystal structures are available, making the ones presented here and in another study from our group the first fragment-bound SARS-CoV-2 Nsp1 crystal structures so far (**Ma et al., 2022**).

Here, we used a combination of computational and experimental approaches to explore the drug-gability of Nsp1 in SARS-CoV-2, and the possibility to expand the findings to homologous Nsp1s. In particular, we have used modelling, enhanced sampling simulations, virtual screening, fragment soaking, and X-ray crystallography to identify druggable binding pockets, including hidden (cryptic) ones, and evaluate the potential of ligands to interfere with the formation of Nsp1-RNA complexes. Our enhanced sampling simulations predict four partially cryptic binding pockets of which one is validated by crystallography. Moreover, taking into consideration the 3D similarity between different Nsp1s in  $\alpha$ - and  $\beta$ -CoVs, we have extended our analysis to assess the conservation of the pockets across various Nsp1s of  $\alpha$ - and  $\beta$ -CoVs infecting humans. The results of this research can be used as a stepping stone for the design of Nsp1 inhibitors for SARS-CoV-2 and potentially for other  $\alpha$ - and  $\beta$ -CoVs.

## Results and discussion

### Structural assessment of Nsp1 pockets

To the date of writing this paper, the only available X-ray structure of apo SARS-CoV-2 Nsp1 contains only the N-terminal region (aa 10–126 PDB entry 7K7P) (**Clark et al., 2021**), hereafter referred to as Nsp1<sub>N</sub> system. Starting from this crystal structure, we have characterized the possible binding pockets present in the 3D structure by means of pocket detection algorithms, namely DoGSiteScorer



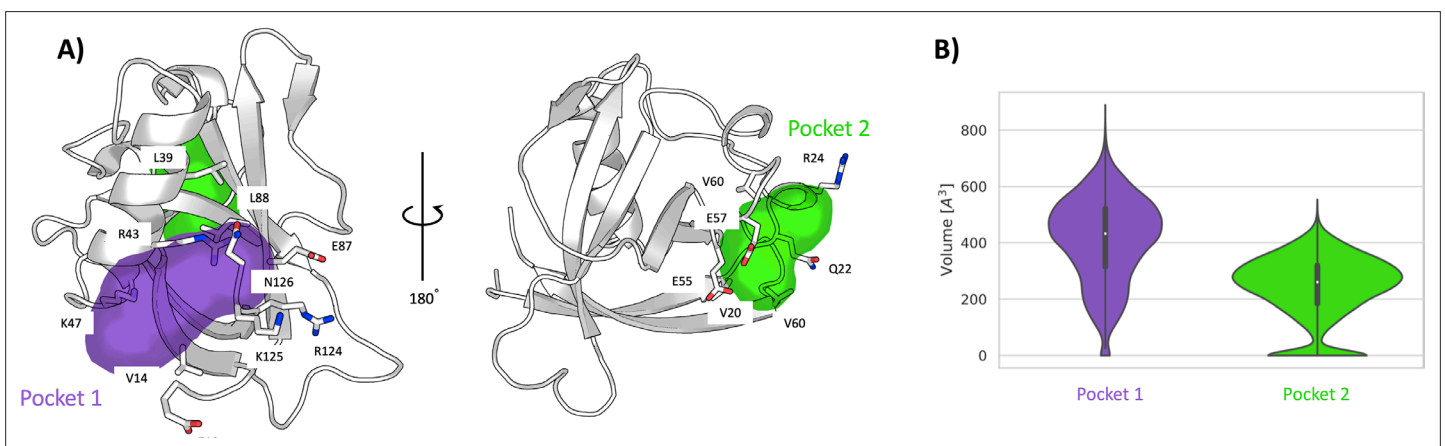
**Figure 2.** Cavities identified on the Nsp1<sub>N</sub> crystal structure (tPDB entry 7K7P) by the ProteinPlus server for the concave pocket-like structure between the  $\beta$ -barrel and the  $\alpha$ -helix and the groove-like topology.

The online version of this article includes the following figure supplement(s) for figure 2:

**Figure supplement 1.** Cavities identified by the Fpocket software on the Nsp1<sub>N</sub> crystal structure (PDB entry 7K7P) (Le Guilloux et al., 2009).

(Volkamer et al., 2012) (available via the ProteinPlus webserver) and Fpocket (Le Guilloux et al., 2009), showing similar results (Figure 2 and Figure 2—figure supplement 1). The analysis indicates the presence of some putative binding sites on the protein surface. Specifically, two different areas were identified to harbor potential binding pockets by both algorithms. The first one is sandwiched between the entrance of the  $\beta$ -barrel and the  $\alpha$ -helix displaying a more evident pocket-like structure with a concave topology (Figure 2, left). The second region, located on the opposite side of the  $\beta$ -barrel, is characterized by a groove-like topology and spans a larger area on the protein surface but is very shallow (Figure 2, right).

To investigate more in-depth the nature of the pocket-like structure and the possible opening of hidden (cryptic) binding sites in the grooved region, we have carried out a 1- $\mu$ s-long unbiased



**Figure 3.** Pockets revealed from unbiased simulations. (A) Cavities identified on Nsp1<sub>N</sub> along the 1  $\mu$ s unbiased molecular dynamics (MD) simulation, namely pocket 1 (purple) and pocket 2 (green), with the main residues displayed in sticks. (B) The volume distribution of each pocket along the unbiased MD simulations.

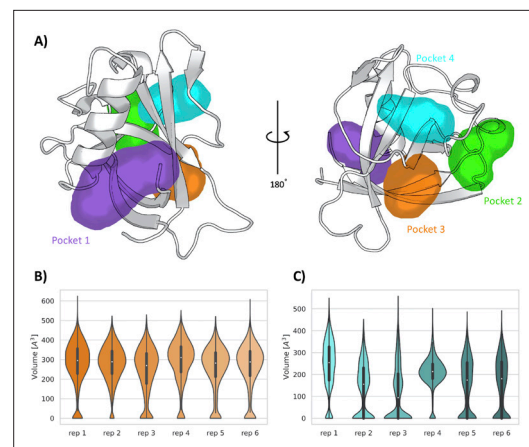
The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Time average of the root-mean-square deviation (RMSD) for Nsp1<sub>N</sub>.

molecular dynamics (MD) simulation of Nsp1N in its apo state. The time series of the RMSD shows that the system remains stable throughout the simulation (**Figure 3—figure supplement 1**). The analysis of the pockets observed along the MD simulation through the MDpocket (Le Guilloux *et al.*, 2009) program confirms the presence of a pocket-like structure, hereafter referred to as pocket 1, at the same place as the one predicted for the X-ray structure by ProteinPlus and Fpocket. A detailed analysis of pocket 1 indicates that the residues forming this pocket are mostly located between the N- and C-termini of the Nsp1<sub>N</sub> protein, namely between the  $\alpha$ -helix and the two  $\beta$ -strands of the  $\beta$ -barrel (**Figure 3A**, left). The analysis of the pocket's volume confirms the stability of pocket 1, with an average value of  $410 \pm 150 \text{ \AA}^3$  (**Figure 3B**). Interestingly, the volume obtained during the simulation is significantly larger than the one predicted based on the crystal structure ( $217 \text{ \AA}^3$ ), since during the simulation the structural features of any pocket are subject to fluctuations due to the rearrangements of the amino acid side chains. However, in this case, the difference in volume between the crystal structure and simulations is the result not only of the residue fluctuations, but also of the different residues identified around the pocket. Specifically, the residues close to the  $\beta$ -barrel form a larger cavity in the simulation than the one observed in the Nsp1<sub>N</sub> crystal structure. This finding suggests that pocket 1 is partially cryptic. It is worth noting that some of these residues are located at the beginning of the C-terminus loop of the protein, suggesting that the binding of a ligand to this region can potentially affect the dynamics of the Nsp1 C-terminus, and possibly release the blockage of host RNA entry into ribosome for translation. Additionally, as pocket 1 is close to the viral RNA binding region (Sakuraba *et al.*, 2021), the binding of a ligand in this pocket could also interfere with the interaction with SL1 of viral RNA, leading to the failure in evasion of translation shutdown and cleavage of viral RNA. The second pocket identified during the MD simulations, namely pocket 2, is located in the grooved region. The structural analysis of pocket 2 over the simulation does not present significant differences with respect to the X-ray structure. In this pocket, the residues are mostly distributed between the loops connecting different  $\beta$ -strands of the  $\beta$ -barrel and the  $\alpha$ -helix (**Figure 3A**, right). Likewise, the volume calculated during the simulation is  $240 \pm 114 \text{ \AA}^3$ , in agreement with the volume predicted by ProteinPlus for the Nsp1<sub>N</sub> crystal structure ( $150 \text{ \AA}^3$ ).

### Crypticity assessment of Nsp1 pockets

Encouraged by the results obtained from the unbiased MD simulations, we used SWISH (Sampling Water Interfaces through Scaled Hamiltonians) with mixed solvents, an enhanced sampling method developed by our group to explore cryptic binding pockets. SWISH is a Hamiltonian replica-exchange method that improves the sampling of hydrophobic cavities by scaling the interactions between water molecules and protein atoms. It has been shown to be very effective in sampling the opening of hidden (cryptic) cavities in several different targets (Comitani and Gervasio, 2018). We have run six replicas of 500 ns each, considering a concentration of 1 M benzene as co-solvent, the presence of which is expected to stabilize any pocket that will open transiently during the simulations (**Figure 4—figure supplement 1**). The resulting trajectories have been analyzed with MDpocket (Le Guilloux *et al.*, 2009). It is worth noting that, in addition to confirming the presence of pocket 1 and pocket



**Figure 4.** Pockets revealed from SWISH simulations. (A) Pockets sampled during the 500 ns of replica-exchange SWISH (sampling water interfaces through scaled Hamiltonians) simulations. Volume distributions of the cryptic binding sites pocket 3 (B) and pocket 4 (C) along the six replicas of the SWISH simulations.

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** Time averages of the root-mean-square deviation (RMSD) for Nsp1<sub>N</sub> across the six SWISH (sampling water interfaces through scaled Hamiltonians) replicas.

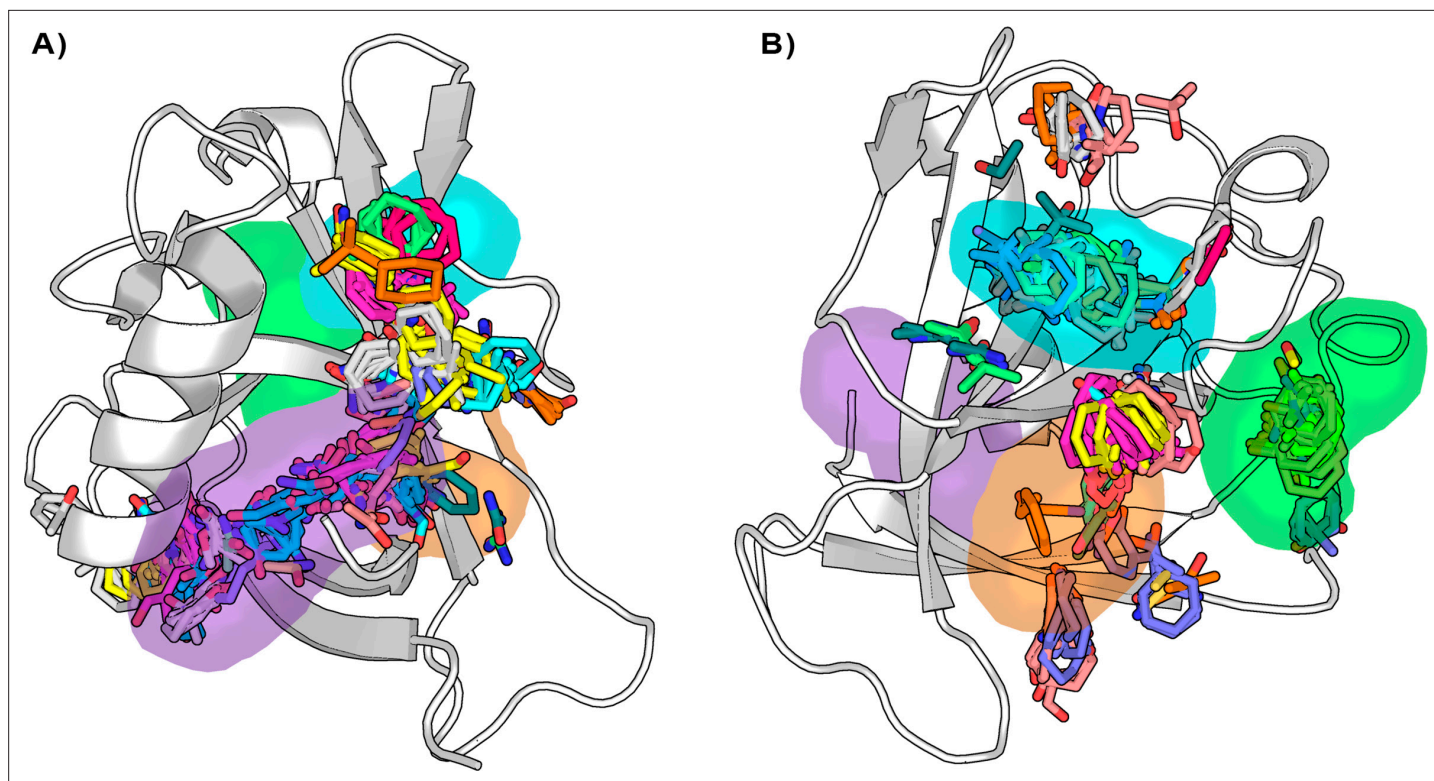
**Figure supplement 2.** Violin plots of the pocket volumes.

2, our analysis shows the presence of two new pockets, namely, pocket 3 and pocket 4. These two pockets are both located on the exterior of the  $\beta$ -barrel, proximal to each other and near pocket 2 (**Figure 4A**). Most of the residues composing pocket 3 are part of the  $\beta$ -sheets of the  $\beta$ -barrel whereas most of the residues in pocket 4 are part of the loops connecting different  $\beta$ -sheets of the barrel. Interestingly, our previous analysis performed over the X-ray structure identified a shallow groove on the surface of Nsp1 connecting pockets 2, 3, and 4. However, neither of the two pocket detection algorithms used on the X-ray structure detected any evident deep cavity in this region. On the contrary, the analysis of the resulting SWISH trajectories clearly shows the presence of two distinct cavities with deep pocket-like structures. The absence of these sites in the crystal structure of Nsp1<sub>N</sub> and the fact that these cavities remained closed during the MD simulations indicate the cryptic nature of pocket 3 and pocket 4.

The average volume of pocket 1 during the SWISH simulations ( $471 \pm 128 \text{ \AA}^3$ ) is comparable to the one obtained in the MD simulations ( $410 \pm 150 \text{ \AA}^3$ ) for the same pocket (**Figure 4—figure supplement 2A**). The topology of pocket 1, that is, the residues identified around the pocket, is the same for the MD and SWISH simulations (**Figure 3**). Likewise, the volume analysis of pocket 2 also indicates that this pocket remains open, and its volume is essentially equal to the one sampled during the unbiased simulation ( $214 \pm 113 \text{ \AA}^3$ , **Figure 4—figure supplement 2B**). More interestingly, the opening of pocket 3 ( $268 \pm 115 \text{ \AA}^3$ ) during the SWISH simulations revealed a pocket of comparable size to pocket 2. The time series of the volume profiles indicates that the pocket starts from a closed-like conformation that opens quickly to a more pocket-like conformation. Once opened, it maintains this open-like conformation in all the replicas (**Figure 4B**). A different behavior was observed for pocket 4. This pocket, the smallest one sampled during the SWISH simulations ( $169 \pm 120 \text{ \AA}^3$ ), displays diverse volume profiles across most SWISH replicas. Nonetheless, we were able to sample an open-like conformation for this pocket in at least two of the six replicas, suggesting that a better combination of molecular probes or higher  $\lambda$  factors could improve the sampling of this cryptic site. Interestingly, as suggested by the preliminary analysis of the X-ray structure, the simulations suggest as well that the shallow groove most likely has a key structural role connecting three cryptic pockets of the  $\beta$ -barrel region of the Nsp1<sub>N</sub>, spanning this area of the protein. Our findings highlight the potential of this region to be exploited in fragment-based drug design to design larger ligands that could bind to different combinations of these pockets, increasing the specificity of the individual pockets.

To further investigate the structural features of the predicted pockets, we analyzed the unbiased MD simulations with FTDyn and FTMap programs (*Kozakov et al., 2015*). These programs have proven to be accurate in locating binding hotspots in proteins. It is based on the fast and accurate distributions of 16 different small organic probes docked and mapped onto the protein surface, retaining the lowest energy probes, which are finally clustered. The clusters obtained for the different probes are clustered together to generate a final consensus cluster, from which the main binding hotspots are then identified. We started by employing the FTDyn webserver, a faster version of the FTMap algorithm without local minimization, to identify the most probable conformation able to bind fragments. We extracted 25 representative conformations from the MD simulation of Nsp1<sub>N</sub> and determined the median number of interactions between the probes and each protein residue across the conformations selected. This step allows us to identify the most likely binding site residues, that is, the residues with the highest number of probe contacts. From the initially selected conformations, we have retained only 11 structures that present higher-than-average probe-residue contacts. Subsequently, the 11 structures retained were post-processed with FTMap to improve the prediction of binding hotspots on Nsp1<sub>N</sub> surface (see details in Materials and methods section and **Figure 5—source data 1**).

Our results are in good agreement with the previous analysis, regarding the location of the binding sites found for 107 consensus clusters obtained for Nsp1<sub>N</sub>. In pocket 1, 50 out of the 107 consensus clusters are mapped into it, most of them corresponding to protein-fragment complexes with the highest binding energy (**Figure 5A** and **Figure 5—source data 1**). Interestingly, the clusters are distributed across the entire MD- and SWISH-sampled volume of pocket 1 and are not only limited to the small region observed in the X-ray structure, suggesting that pocket 1 is partially cryptic. The computational mapping also confirmed the presence of 50 consensus clusters in the vicinity of pockets 2, 3, and 4 (**Figure 5B**). In all selected bound-like Nsp1<sub>N</sub> structures, we identified at least two probes consensus clusters located in the region corresponding to the two cryptic sites previously identified by



**Figure 5.** Distribution of the binding hotspots on the Nsp1<sub>N</sub> surface around (A) pocket 1 and (B) pockets 2, 3, and 4. Multiple consensus clusters are shown in sticks. Each consensus cluster is represented in different colors.

The online version of this article includes the following source data for figure 5:

**Source data 1.** Consensus clusters information as obtained from the FTMap program for the 11 selected Nsp1<sub>N</sub> structures.

our SWISH simulations. In some cases, the second-best binding hotspot is in the presumably cryptic region, suggesting the ability of this region to accommodate molecular fragments (**Figure 5—source data 1; Vajda et al., 2018**). Overall, this analysis strongly suggests that the main binding sites on Nsp1<sub>N</sub> are in the region corresponding to the identified cryptic pockets and highlights their potential use for fragment binding.

### Crystallographic confirmation of Nsp1 cryptic pockets

To validate our computational findings, we proceeded with the soaking of Nsp1<sub>N</sub> crystals with 59 potential fragment hits obtained by computational methods (**Table 1**). Nsp1<sub>N</sub> has been crystallized containing only the structured domain, that is, the domain from Glu10 to Asn126, which displays the characteristic  $\alpha/\beta$ -fold (see Materials and methods for details). This structure was used for crystal soaking experiments in which 59 distinct fragment hits, obtained from the Maybridge Ro3 library, were tested and validated through X-ray diffraction experiments using the Pan-Dataset Density Analysis (PanDDA) method, developed to analyze the data obtained from crystallographic fragment screenings (**Pearce et al., 2015**). Of these 59 fragments, one fragment was found in pocket 1 as previously identified in our simulations. Data collection and refinement statistics are summarized in **Figure 6—figure supplement 1** and **Figure 6—source data 1**. The asymmetric unit contains one molecule of Nsp1. The model covers the sequence from Glu10 to Asn126 (E10 to N126). The chemical structure of the fragment hit is shown in **Figure 6B**. To study the fragment hit using orthogonal biophysical assays, we employed microscale thermophoresis (MST) and thermal shift (TSA) assays (**Figure 6—figure supplement 2** and **Table 2**). 2E10 binds to Nsp1<sub>N</sub> pocket 1 with a KD value of  $15.1 \pm 5.7$  mM, indicating a rather good binding for a fragment. In contrast, we did not observe any stabilization of Nsp1<sub>N</sub> by the fragment. The fragment obeys the rule of 3 (Ro3) with a molecular mass of 172.9 Da, a calculated MolLogP of 2.66 and a total of three hydrogen bond donors and acceptors. Fragment hit 2E10 combines two fused 5- and 6-membered ring systems containing one acetamide substituent

**Table 1.** SMILES and identification number of the 59 tested fragments and corresponding predicted binding.

The predicted pose for each fragment can be found at [https://github.com/Gervasiolab/Gervasio-Protein-Dynamics/tree/master/nsp1/virtual\\_screening](https://github.com/Gervasiolab/Gervasio-Protein-Dynamics/tree/master/nsp1/virtual_screening) (**Gervasiolab, 2022**; copy archived at [swh:1:rev:936e929db11aff39faed53e5f6e6f902f1456a6d](https://www.swh.io/rev/936e929db11aff39faed53e5f6e6f902f1456a6d)). The above reported crystal hit (lig\_1427) is highlighted in bold.

Identification Number	SMILES	Predicted pocket(s)
lig_30	C9H7NO2	3, 4
lig_83	C12H13NO2	1, 3, 4
lig_113	C12H13CIF3NO	1, 3, 4
lig_168	C13H10O3	1, 3, 4
lig_171	C11H10O3	1, 4
lig_194	C8H6F2O3	1, 3, 4
lig_212	C11H12O2	1, 3, 4
lig_243	C14H17NO2S	1
lig_262	C15H16N2O	1, 3, 4
lig_286	C13H13NO2S	1, 3, 4
lig_329	C9H8F3NOS	3, 4
lig_335	C13H11NO2S	1, 3
lig_349	C11H9NO2	1, 4
lig_355	C13H13NO2S	4
lig_369	C11H12O2	1, 3, 4
lig_377	C12H10O2S	1, 4
lig_394	C16H13NO2	1, 3, 4
lig_400	C14H12O2	1, 3, 4
lig_422	C14H9NO2	1, 3, 4
lig_490	C8H5NO2S2	1, 4
lig_502	C12H10O3	1, 3, 4
lig_507	C12H12N2O	1, 4
lig_552	C12H15NO2	1, 3, 4
lig_570	C12H9NO2	1, 4
lig_575	C10H8O3	1, 4
lig_579	C12H9NO2	1, 3, 4
lig_685	C14H12O3	1, 3, 4
lig_706	C12H14N2O	1, 4
lig_752	C9H7CIFN3S2	1, 3, 4
lig_783	C12H11FO2	1, 3, 4
lig_806	C14H13NO2	1, 3
lig_812	C13H11CIN2O	1
lig_864	C15H12O3S	1, 3, 4
lig_892	C7H6N2OS	1, 3
lig_897	C8H6FNO2S	1, 3, 4

Table 1 continued on next page

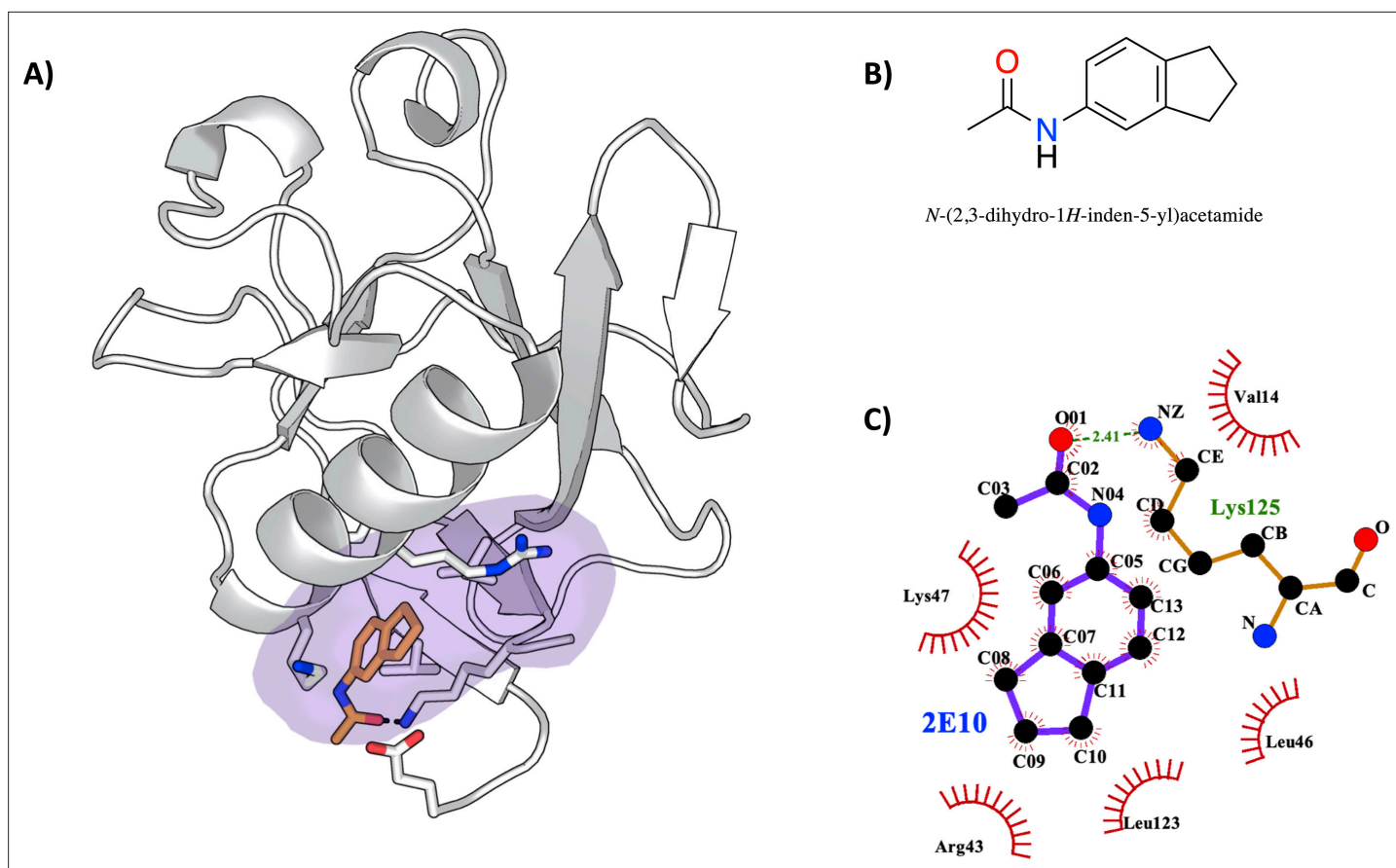


Table 1 continued

Identification Number	SMILES	Predicted pocket(s)
lig_907	C10H11NOS2	1, 4
lig_910	C10H6F3NOS	1, 3, 4
lig_924	C14H12O3S2	1
lig_1009	C11H10CINO2	1, 4
lig_1037	C10H7ClO2S	1, 4
lig_1054	C12H13NO2	1, 3, 4
lig_1057	C12H16N2O	1, 3, 4
lig_1064	C13H12O2	1, 3, 4
lig_1149	C11H10F3NO	1, 3, 4
lig_1157	C13H13NO2	1, 3, 4
lig_1195	C11H10N2O	1, 3, 4
lig_1209	C9H6F3N3S2	1, 3, 4
lig_1216	C12H11F2NO2	1, 3, 4
lig_1220	C11H12O2S	3, 4
lig_1223	C8H9CIN2OS	1, 3
lig_1228	C15H13FN2OS	1, 3, 4
lig_1281	C9H7F3O3	1, 3, 4
lig_1310	C12H13NO2	1, 4
lig_1315	C13H11F3O2	1, 3, 4
lig_1381	C9H6F3NO	3, 4
lig_1382	C12H8F3NO2S	1, 3, 4
lig_1410	C9H11NO	3, 4
lig_1427	C11H13NO	1, 4
lig_1428	C11H11NO2	1, 4

at the phenyl group A range of residues located in binding pocket 1, including Glu10, Val14, Arg43, Leu46, Lys47, and Leu123, establish hydrophobic interactions with the ligand (**Figure 6C**). The acetamide substituent establishes a hydrogen bond interaction with the side chain of Lys125.

Although, at first, this may seem like a very low number of hits, a detailed analysis of the crystal packing provides an explanation. **Figure 7A** shows the central Nsp1<sub>N</sub> structure (white) with the pockets obtained from our SWISH simulations, together with the neighboring Nsp1 structures that interact directly with pocket 1 in the full crystal packing (dark gray). Evidently, the central part of the pocket is completely accessible for fragment binding during the crystal soaking experiments. However, the gray structures are occupying the main hotspots identified by our previous FTMap analysis (**Figure 5A**) impeding its ability to bind more suitable fragments in pocket 1. Likewise, **Figure 7B** shows the central Nsp1 structure in cartoon representation (white) with the pockets obtained from our SWISH simulations, along with the neighboring Nsp1 structures directly interacting with pockets 2, 3, and 4 in the full crystal packing (dark gray). The red regions highlighted in **Figure 7** involve direct contacts made with the pocket environment. It is worth emphasizing that, in this case, these direct contacts involve regions previously determined by the FTMap analysis as potential and key hotspots for the three cryptic pockets identified by SWISH. Taken altogether, soaking Nsp1<sub>N</sub> crystals with fragments presents some limitations to implementing an effective computational fragment screening of the Nsp1. Nevertheless, we would like to stress that these results provide a possible explanation for why we did not obtain more fragments crystallized in the positions of the identified cryptic pockets.



**Figure 6.** Characterization of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) non-structural protein 1 (Nsp1)-2E10 complex. (A) Binding pose of the fragment hit obtained by crystal soaking and structure determination methods. The fragment is located in pocket 1. (B) Chemical structure and name of fragment hit. (C) Magnification of the Nsp1<sub>N</sub>-2E10 binding pocket showing the interactions the fragment establishes with residues of Nsp1. Hydrophobic interactions are shown by red half-moons and the hydrogen bond interaction is displayed with a dotted green line.

The online version of this article includes the following source data and figure supplement(s) for figure 6:

**Source data 1.** Crystallographic data and model refinement statistics for the SARS CoV-2 Nsp1N-2E10 complex.

**Figure supplement 1.** Pan-Dataset Density Analysis (PanDDA) event map and standard single dataset map focused on the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Nsp1<sub>N</sub> binding site of fragment hit 2E10.

**Figure supplement 2.** Microscale thermophoresis (MST) measurements of the binding affinity between fragment hit 2E10 and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) non-structural protein 1 (Nsp1).

Moreover, the crystallization of a single fragment in pocket 1 is an encouraging sign of the reliability and efficacy of our computational techniques for cryptic binding pocket detection.

### Disrupting Nsp1-RNA complex by means of cryptic pockets

The interaction of Nsp1 with viral RNA can release the inhibition of the viral RNA translation in host cells via the combination of different mechanisms, favoring the survival of the virus. Therefore, the disruption of the Nsp1-RNA complex can affect the life cycle of the virus. With this idea in mind, and taking into consideration the pockets found in Nsp1<sub>N</sub>, we asked whether the binding of fragments to pocket 1 could hinder the Nsp1-RNA complex formation. To this end, we have modelled two Nsp1-RNA complexes, testing their stability through MD simulations. It has been shown experimentally that the first stem loop (SL1), comprising nucleotides 7–33 of the 5' UTR region of SARS-CoV-1 and SARS-CoV-2 RNA, is the one that interacts with Nsp1 (Narayanan *et al.*, 2008; Tanaka *et al.*, 2012; Vankadari *et al.*, 2020; Tidu *et al.*, 2020). Therefore, we started by modelling the 3D structure of SL1 RNA. Regarding the Nsp1, we have considered the model of the full-length Nsp1 obtained

**Table 2.** Thermal shift assay (TSA) results for 59 potential fragment hits from computational screening using severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Nsp1<sub>N</sub>.

Fragments showing atypical curves are labelled as atypical curve in the table.

Identification number	Fragments	T <sub>i</sub> ± SD [°C]	ΔT <sub>i</sub> / °C
lig_575	1B9	52.5±0.6	-2.7
lig_490	1C6	51.5±0.4	-3.4
lig_507	1H6	50.5±0.1	-4.1
lig_1427	2E10	53.8±0.3	-1.3
lig_1428	2F10	53.5±0.4	-1.7
lig_1037	2F5	53.7±0.6	-1.6
lig_685	3E8	54.8±0.2	-0.3
lig_924	4B5	54.7±0.5	-0.5
lig_502	4C6	Atypical curve	
lig_369	4E4	53.6±0.4	-1.4
lig_400	5B7	53.9±0.1	-0.9
lig_194	5D3	Atypical curve	
lig_335	5E8	54.8±0.2	-0.4
lig_30	6E3	Atypical curve	
lig_113	6G8	54.3±0.0	-0.6
lig_168	7A2	51.2±0.4	-3.8
lig_212	7A4	Atypical curve	
lig_552	7B6	55.0±0.2	0.1
lig_1220	7D10	54.7±0.3	-0.2
lig_171	7D2	54.6±0.2	-0.3
lig_570	7E6	51.5±0.4	-3.1
lig_579	7F11	51.8±0.2	-3
lig_1054	7F6	54.3±0.3	-0.6
lig_1310	8A4	Atypical curve	
lig_1281	8A5	Atypical curve	
lig_422	8B4	51.2±0.1	-3.7
lig_349	8C4	Atypical curve	
lig_1410	8D3	53.2±0.3	-2.4
lig_1009	8G7	51.5±0.4	-3.3
lig_355	9A11	54.9±0.1	0
lig_243	9C10	54.3±0.1	-0.9
lig_1216	9D8	54.5±0.4	-0.7
lig_812	9E11	53.1±0.1	-1.9
lig_329	9E2	54.4±0.3	-0.5
lig_262	9G9	53.2±0.3	-1.7
lig_783	10A6	54.4±0.2	-1.2
lig_910	10A8	Atypical curve	

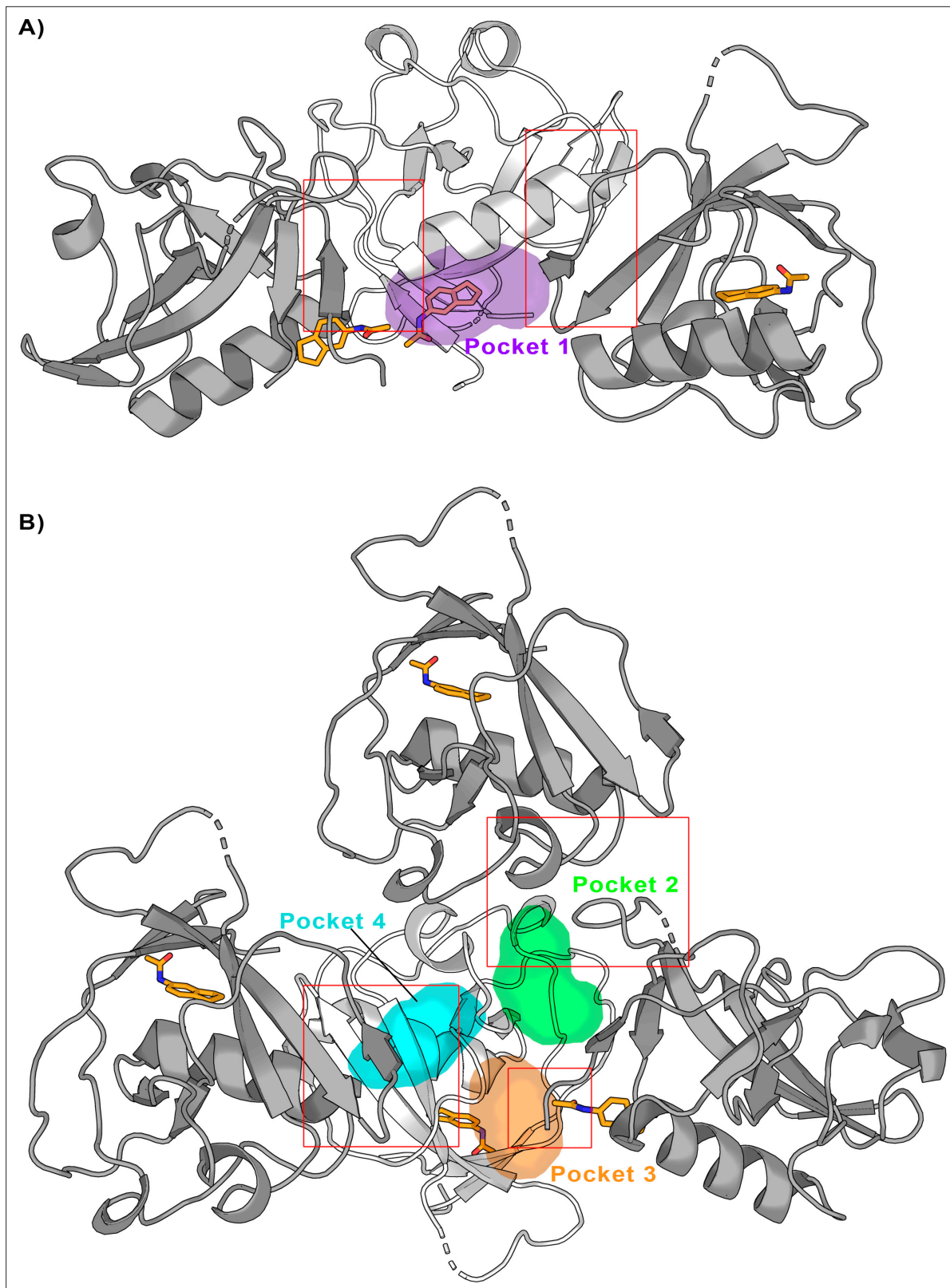
*Table 2 continued on next page*

Table 2 continued

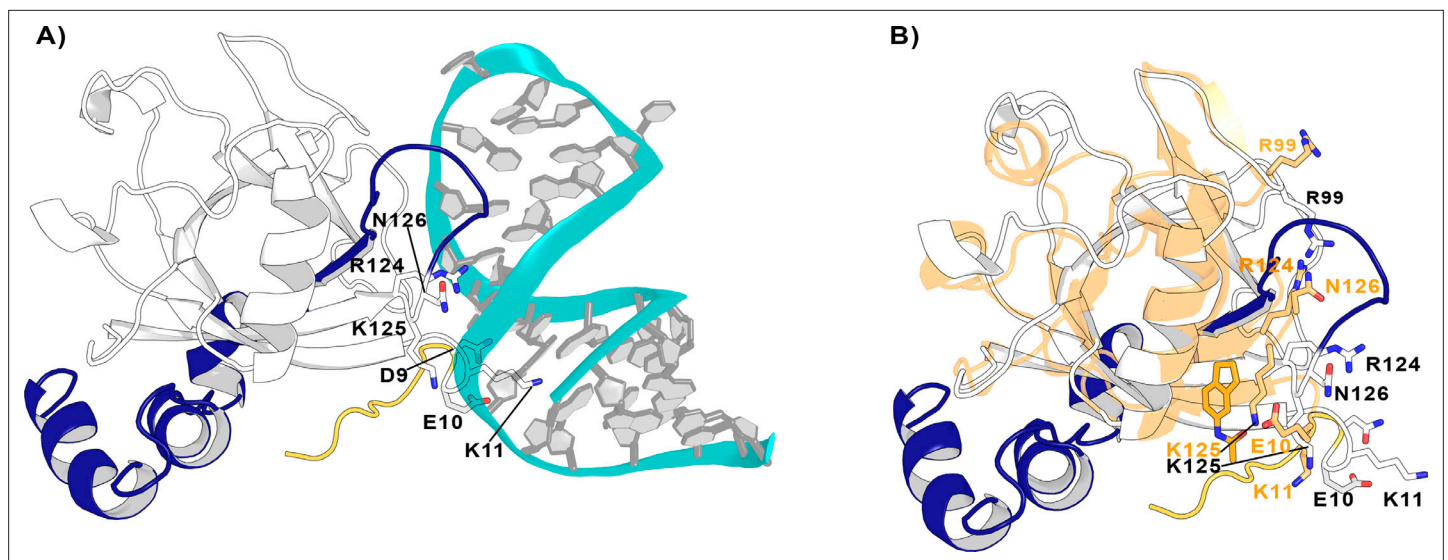
Identification number	Fragments	$T_i \pm SD$ [°C]	$\Delta T_i$ / °C
lig_806	10A9	54.6±0.1	-1
lig_1381	10B7	50.7±0.3	-4.9
lig_286	10G8	54.1±0.2	-0.7
lig_1209	11A5	55.8±0.1	0.2
lig_1057	11B2	Atypical curve	
lig_907	11E4	55.6±0.0	0.7
lig_1064	11E8	54.6±0.1	-0.3
lig_377	11F10	54.8±0.1	-0.8
lig_1195	11F2	54.8±0.2	-0.1
lig_1149	11G3	55.4±0.5	0.5
lig_83	12A11	Atypical curve	
lig_1315	12A3	Atypical curve	
lig_864	12B10	55.1±0.2	-0.1
lig_394	12B5	55.2±0.2	0.1
lig_1382	12C4	55.5±0.4	0.6
lig_752	12D5	54.9±0.1	-0.3
lig_1223	12F7	52.4±0.3	-2.7
lig_1157	12F9	55.0±0.2	-0.2
lig_1228	12G7	52.6±0.5	-2.6
lig_892	12H2	54.8±0.1	-0.4
lig_706	13D6	51.8±0.4	-3.3
lig_897	13G5	54.3±0.1	-0.9

by AlphaFold ([Jumper et al., 2021](#); [Varadi et al., 2022](#)), named Nsp1<sub>FL</sub>. Using the Nsp1<sub>FL</sub> and SL1 models, we have run protein-RNA docking using the HADDOCK program ([Dominguez et al., 2003](#); [van Zundert et al., 2016](#)). The analysis of the docked Nsp1<sub>FL</sub>-RNA complexes suggests that two different protein-RNA complexes, named model A and model B, well capture the experimentally validated contacts between the Nsp1<sub>FL</sub> and SL1. The Nsp1<sub>FL</sub> in models A and B binds to diametrically opposite positions on the SL1, with Nsp1 in model A interacting predominately with the groove of SL1 ([Figure 8A](#)) and in model B with its backbone ([Figure 8—figure supplement 1](#)). Since there was no structural or experimental reason to exclude one over the other, we assessed their stability over the course of a 500-ns-long MD simulation. The RMSD of the two Nsp1<sub>FL</sub>-RNA complexes fluctuates between 5.5 and 6.5 Å for models A and B, respectively ([Figure 8—figure supplement 2](#)). However, a more detailed RMSD analysis of each component of the two protein-RNA systems reveals that Nsp1<sub>FL</sub> is more stable in model A than in model B, with an average RMSD of 5.0 and 7.0 Å, respectively. This result suggests that the Nsp1<sub>FL</sub> in model A displays a more stable interaction with RNA. However, the observed RMSD difference alone between the two models is not sufficient to let us propose model A over B as the most probable mode of Nsp1-RNA interaction ([Figure 8—figure supplement 2](#)).

Since both models not only capture the important interactions between Nsp1<sub>FL</sub> and viral RNA described experimentally but are also stable over the course of the MD simulations, we used them to verify if the previously identified pockets could be exploited to impede the Nsp1<sub>FL</sub>-RNA complex. Interestingly, out of the four pockets in Nsp1<sub>N</sub>, pocket 1 is positioned close to the Nsp1<sub>FL</sub>-RNA interface. More specifically, some residues of the C-terminal loop of Nsp1<sub>FL</sub> around pocket 1 interact with the viral RNA, namely Arg124, Lys125, and Asn126. Moreover, Glu10 is oriented toward pocket 1 and both Asp9 and Lys11 establish contacts with the SL1 moiety ([Figure 8A](#)). Therefore, these results



**Figure 7.** Crystal packing of the Nsp1<sub>N</sub>-fragment complex (PDB entry 8A4Y) was obtained from X-ray soaking experiments. The direct crystal contacts around (A) pocket 1 and (B) pockets 2, 3, and 4 are highlighted with squares.



**Figure 8.** Nsp<sub>1FL</sub>-RNA complex obtained from the HADDOCK program. **(A)** Residues in the proximity of pocket 1 involved in crucial Nsp<sub>1FL</sub>-SL1 contacts are displayed in sticks. **(B)** Crystal structure from soaking experiments (orange transparent cartoon) is superimposed over the Nsp<sub>1FL</sub> of the Nsp<sub>1FL</sub>-RNA model. The most important residues for the Nsp<sub>1FL</sub>-RNA interaction are highlighted in white (Nsp<sub>1FL</sub> model) and orange (PDB entry 8A4Y from our soaking experiments) sticks. Figures obtained from model A.

The online version of this article includes the following figure supplement(s) for figure 8:

**Figure supplement 1.** Different Nsp<sub>1FL</sub>-RNA models proposed in this work.

**Figure supplement 2.** RMSD for model A and B of the Nsp1-RNA complex.

**Figure supplement 3.** Ligand accessibility to pocket 1 in Nsp1FL.

suggest that pocket 1 can be a crucial target for rational drug design since the residues that define this pocket are directly interacting with the viral RNA and the binding of a ligand to this pocket can impair the Nsp<sub>1FL</sub>-RNA interaction.

To further investigate the effect of ligand binding to pocket 1 and how this would affect the Nsp1-SL1 interface, we superimposed the crystal structure from the soaking experiments on the Nsp<sub>1FL</sub>-RNA model A (**Figure 8B**). Interestingly, in the Nsp<sub>1FL</sub>-RNA complex, some residues of the N-terminus (Asp9, Glu10, and Lys11) and the core (Arg99, Arg124, Lys125, and Asn126) of Nsp<sub>1FL</sub> are directly interacting with the SL1 of RNA. When compared with the structure obtained from soaking experiments (Glu10 to Asn126), some of these residues have changed orientation. Specifically, Glu10 and Lys11 are pointing toward the fragment, Arg99 is pointing to the surface, and Arg124 and Asn126 are displaced due to the ligand binding. In the orientation that these residues have adopted in the structure obtained from soaking experiments, these residues could not interact with the RNA. As these residues are essential for the Nsp1-SL1 interaction in SARS-CoVs, the binding of a ligand in pocket 1 could diminish the Nsp1-mRNA interaction (*Tanaka et al., 2012; Terada et al., 2017*). Additionally, a recent study where site-directed mutagenesis was performed on the Nsp1 N-terminal and core region (Nsp1<sub>N</sub>) has demonstrated the functional role of Arg99, Arg124, and Lys125 residues in host expression shutdown, since the mutation of these residues to Ala compromised the binding of Nsp1 to the host 40S ribosomal subunit and increased the dissociation constants with purified ribosomes (*Mendez et al., 2021*). Therefore, these findings support the hypothesis that a disruption of the interactions between these residues and SL1 by the presence of a fragment would lead to a decrease in virus-induced host shutoff.

To assess whether the C-terminal end would affect the accessibility of a fragment to pocket 1, we quantified the volume of the pocket throughout three independent replicas. The C-terminal domain indeed affects the accessibility of the identified pocket. The distribution of the pocket volume (**Figure 8—figure supplement 3**) shows how the volume of pocket 1 reduces to approximately 200 Å<sup>3</sup> in the case of Nsp<sub>1FL</sub>, reaching similar values to the ones found in the crystal structure of the apo Nsp1 (PDB entry 7K7P). The loop that connects the N- to the C-terminal (aa Arg124-Pro153) is flexible

and a portion of it (aa Arg124-Gly137) occasionally occludes pocket 1. Nonetheless, we clustered the last 200 ns of the trajectories and extracted representative structures per replica corresponding to the most populated clusters. For the structure representing the second-most populated cluster in replica 2, we were able to dock the fragment within pocket 1, obtaining a similar pose to the one seen in the presented crystal structure (**Figure 8—figure supplement 3**). This indicates that, even in the setting of full-length Nsp1, pocket 1 is still accessible in the context of the full-length Nsp1. Moreover, we obtained a docked pose also for replica 3, even if the site is partially closed due to the inward orientation of Arg43 (**Figure 8—figure supplement 3**). A conformation where the pocket was fully closed was instead obtained for replica 1, where no fragment could be docked to the pocket. Taken together, these results indicate that pocket 1 is still accessible for fragment binding in our full-length Nsp1 model in solution. Moreover, the flexible nature of the loop surrounding the pocket supports the hypothesis of the cryptic nature of the site, which is partially detectable in the crystal structures where the C-terminal domain is absent. Ultimately, the accessibility predictions from our simulations are confirmed by the MST measurements conducted on Nsp1<sub>FL</sub> (**Figure 6—figure supplement 2**). The measured KD has a larger error bar, but within the error is comparable to the one obtained for Nsp1<sub>N</sub>, indicating that the C-terminal domain present in Nsp1<sub>FL</sub> does not prevent the binding of the fragment.

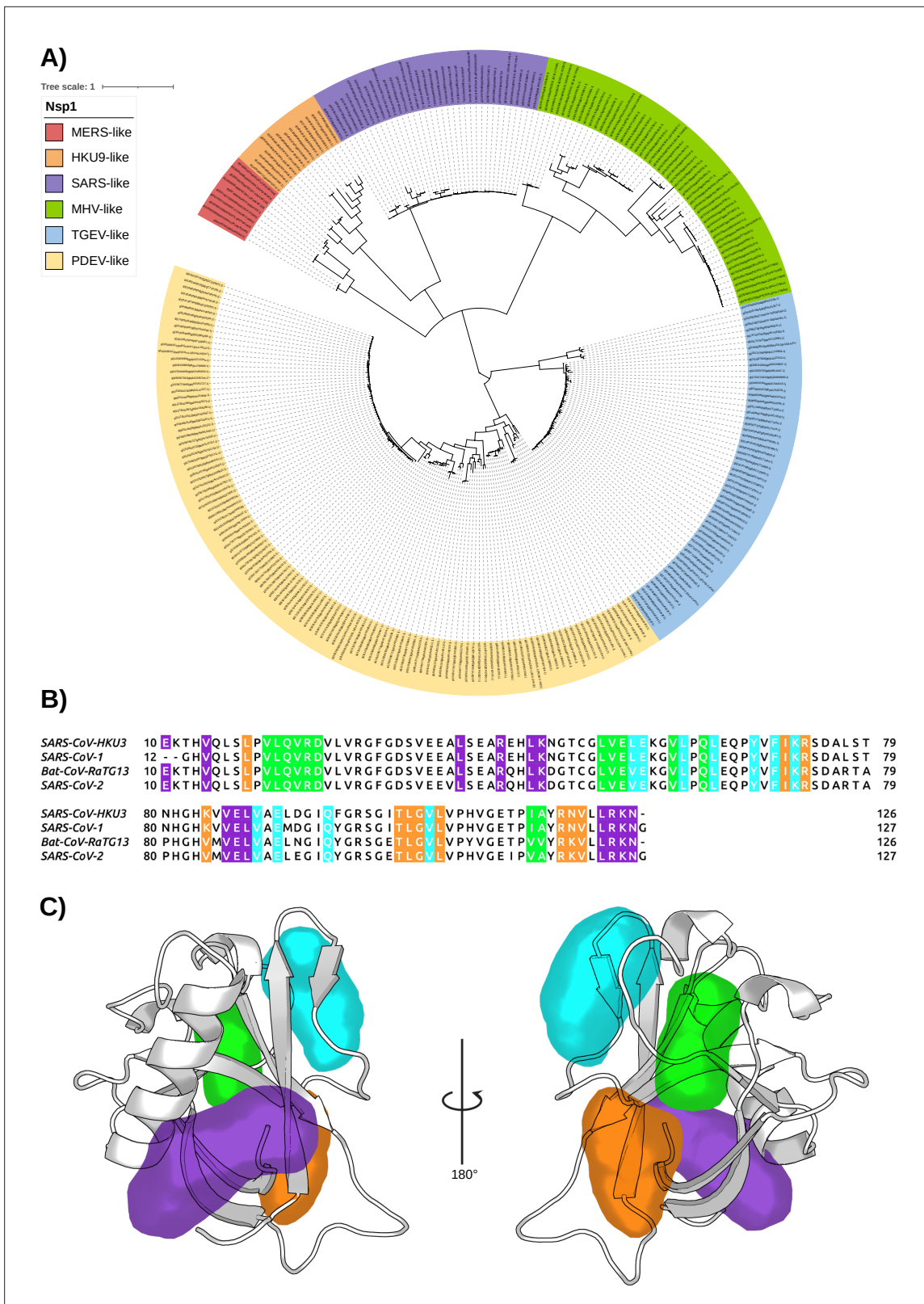
### The conservation of Nsp1 in different CoV genera

Given the identified cryptic pockets on the Nsp1 of SARS-CoV-2 and their implication in the interaction with the RNA, we sought to analyze thoroughly the structural similarity of the Nsp1 among different CoVs. We asked whether a drug designed for Nsp1 of SARS-CoV-2 could also be useful to inhibit Nsp1 of other CoVs. As mentioned in the Introduction, of the four CoV genera, only  $\alpha$ - and  $\beta$ - are common human CoVs, and Nsp1 is expressed only in these two genera. To demonstrate the homology relationship of the Nsp1<sub>N</sub> domain within different viruses belonging to the  $\alpha$ - and  $\beta$ -genera, we have performed a phylogenetic analysis considering this domain, for both  $\alpha$ - and  $\beta$ -CoV genera from the Conserved Domain Database (CDD) hosted at the National Center for Biotechnology Information (NCBI) (*Lu et al., 2020*). The analysis of the 283 Nsp1<sub>N</sub> sequences available until the end of February 2022 shows the distribution of Nsp1 homologues for different  $\alpha$ - (TGEV- and PDEV-like) and  $\beta$ - (MERS-, HKU9-, SARS-, and MHV-like) CoVs (**Figure 9A**). In particular, we have performed a pair sequence alignment considering one representative Nsp1 protein from each of the  $\alpha$ - and  $\beta$ -CoVs subfamilies presented in our phylogenetic tree (*Needleman and Wunsch, 1970*). The alignment demonstrates that the sequence identity varies widely depending on the homologue proteins under consideration (**Table 3**). Subsequently, to decrease the high heterogeneity, we have considered the Nsp1<sub>N</sub> sequence from SARS-CoV-2 and three high identity homologues corresponding to Nsp1 proteins from the human SARS-CoV-1, and two Bat CoVs, namely BatCoV HKU3 and RaTG13 (NCBI accession numbers: MT782115.1 and MN996532.2, respectively; *Min et al., 2020*). The selected homologues share a high sequence identity with SARS-CoV-2 Nsp1 ranging from 86% to 93%. **Figure 9B** shows this high sequence identity between CoVs from different organisms, especially of the residues surrounding the cryptic pockets found in Nsp1 from SARS-CoV-2. Our analysis shows that for the four sequences, all the important residues of the pockets are conserved in the selected  $\beta$ -CoVs.

Finally, taking into account the previous data and the fact that Nsp1 proteins share a high 3D structural core identity across CoV species (*Shen et al., 2018; Min et al., 2020*), we have run 1  $\mu$ s of unbiased MD simulations for each of the three Nsp1 homologues considered, which show a stable conformation (**Figure 9—figure supplement 1**). Additionally, for all of them, we have evaluated the conservation of the cryptic binding pockets via extensive SWISH simulations (six replicas of 500 ns per homologue). The resulting trajectories indicate that the four pockets we reported for Nsp1<sub>N</sub> of SARS-CoV-2 are conserved in the considered homologues (**Figure 9C** and **Figure 9—figure supplement 2**). Taken together, these results suggest that ligands binding to any of the pockets identified in SARS-CoV-2 Nsp1<sub>N</sub> could also target the corresponding pocket in the evaluated homologues, ultimately paving the way to the development of a drug targeting the Nsp1<sub>N</sub> of different  $\beta$ -CoVs.

### Conclusions

Nsp1 is a promising drug target for CoVs both due to its crucial role in suppressing host immune response and its sequence conservation and structural similarity across the  $\alpha$ - and  $\beta$ -CoV families. In this paper, by using a multidisciplinary approach combining modelling, simulations, X-ray crystallography,



**Figure 9.** The conservation of Nsp1 sequence. **(A)** Phylogenetic tree based on 283 sequences from distinct  $\alpha$ - and  $\beta$ -coronaviruses (CoVs) of different subgenera. The sequences were obtained from the Conserved Domain Database (CDD) with accession number cl41742. The scale bar indicates the number of substitutions per site in the amino acid sequence. Six different Nsp1 domain models can be identified, two for the  $\alpha$ -CoVs (transmissible gastroenteritis virus [TGEV]-like and PDEV-like), and four for the  $\beta$ -CoVs genus (MERS-like, HKU9-like, SARS-like, and murine hepatitis virus [MHV]-like).

Figure 9 continued on next page



Figure 9 continued

(B) Multi-sequence alignment of the four homologues selected. The alignment was performed with the MUSCLE algorithm (Edgar, 2004). The residues of the different pockets are highlighted in the corresponding color, namely pocket 1 in purple, pocket 2 in green, pocket 3 in orange, and pocket 4 in cyan. (C) Representation of the pockets found in the severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1) Nsp1<sub>N</sub> variant.

The online version of this article includes the following figure supplement(s) for figure 9:

**Figure supplement 1.** Time averages of the backbone root-mean-square deviation (RMSD) along the 1  $\mu$ s unbiased molecular dynamics (MD) simulation for Nsp1<sub>N</sub> of (A) severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), (B) SARS-HKU3, and (C) Bat-CoV-RaTG13.

**Figure supplement 2.** Representation of the pockets found in the Bat-CoV-HKU3 (A) and Bat-Cov-RaTG13 (B) Nsp1<sub>N</sub> variants.

and fragment screening, we revealed druggable and partially cryptic pockets in the folded main domain of Nsp1. Our enhanced sampling simulations revealed four candidate pockets and predicted that fragments can bind to them. Not all of these are predicted to be accessible in a crystal structure, due to crystallographic contacts. The fragment screening and subsequent crystallographic structures confirm the presence of the deepest and most accessible pocket predicted by the simulations. Interestingly, the pockets are conserved across multiple CoV species. Moreover, we show how the fragment bound to one of these pockets can disrupt the Nsp1-mRNA complex. In particular, the binding of the fragment to the identified pocket interferes with crucial interactions between the 5'-UTR SL1 and Nsp1, namely R124 and K125. This could prevent the viral mRNA to be efficiently translated, ultimately impairing the viral translation strategy. At the same time, the fragment is expected also to lower the affinity of Nsp1 for the ribosome by hindering the interaction mediated by R99, R124, and R125. Altogether, the crucial information arising from our multidisciplinary approach can provide a solid foundation for the rational drug discovery of new inhibitors not only for SARS-CoV-2, but also for other  $\alpha$ - and  $\beta$ -CoVs with pandemic potential.

## Materials and methods

### Setup of the systems

#### Preparation of Nsp1<sub>N</sub>

The structure of the N-terminal domain of SARS-CoV-2 Nsp1 (Nsp1<sub>N</sub>) was obtained from the PDB entry 7K7P (resolution 1.77 Å), comprising residues Glu10-Asn126 (Martínez-Rosell et al., 2017).

**Table 3.** Pairwise identity percentages of selected non-structural protein 1 (Nsp1) sequences for representative  $\alpha$ - and  $\beta$ -coronaviruses (CoVs) subfamilies.

National Center for Biotechnology Information (NCBI) accession numbers the sequences used for the analysis are as follows: transmissible gastroenteritis virus (TGEV) *6IVC\_A*, PDEV *5XBC\_A*, severe acute respiratory syndrome coronavirus 1 (SARS CoV-1) *NP\_828860.2*, severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) *YP\_009725297.1*, HKU9 *P0C6T6*, MERS *YP\_009047229*, and murine hepatitis virus (MHV) *YP\_209244*.

CoV-like	<i>SARS-1</i>	<i>SARS-2</i>	<i>HKU9</i>	<i>MERS</i>	<i>MHV</i>	<i>TGEV</i>	<i>PDEV</i>
<i>SARS-1</i>	100						
<i>SARS-2</i>	86.1	100					
<i>HKU9</i>	20.9	19.5	100				
<i>MERS</i>	20.9	17.1	24.5	100			
<i>MHV</i>	16.9	16.4	4.1	18.0	100		
<i>TGEV</i>	17.4	16.9	13.3	7.9	9.0	100	
<i>PDEV</i>	17.9	16.1	15.2	8.7	11.5	18.2	100

## Preparation of Nsp1<sub>FL</sub>

The Nsp1<sub>FL</sub> structure model was obtained from AlphaFold 2.0 neural network-based structural prediction (*Jumper et al., 2021; Varadi et al., 2022*), which is based on the NCBI Reference Sequence: YP\_009725297.1.

## Preparation of Nsp1<sub>FL</sub>-RNA complex

The SL1 structure corresponding to nucleotides 7–33 at the 5' UTR of the viral mRNA was modelled with RNAComposer (*Popenda et al., 2012; Antczak et al., 2016*). The NCBI accession number for the reference viral RNA is NC\_045512 (<https://www.ncbi.nlm.nih.gov/genome/viruses/>). Then, the SL1 structure obtained was docked to the Nsp1<sub>FL</sub> model from AlphaFold using the HADDOCK software (*Dominguez et al., 2003; van Zundert et al., 2016*). Two docking poses of Nsp1<sub>FL</sub>-RNA complexes were selected based on their score and the non-covalent interactions on the interface of Nsp1<sub>FL</sub> with RNA (*Figure 8—figure supplement 1*).

## Preparation of Nsp1 in different CoV genera

The NMR structure of the SARS-CoV-1 Nsp1<sub>N</sub> was obtained from PDB entry 2HSX. Since there are no experimentally determined structures for the Nsp1 of Bat CoVs HKU3 nor RatG13, we modelled their structure using Nsp1<sub>N</sub> of SARS-CoV-2 (PDB entry 7K7P) as a template. The NCBI accession number for HKU3 and RatG13 is QND76032.1 and QHR63299.2, respectively, using the *automodel* function of MODELLER (*Sali and Blundell, 1993*).

Each of these systems has been processed in the same way. First, the standard protonation state at physiological pH 7.4 was assigned to ionizable residues with the ProteinPrepare tool of the Play-Molecule server (*Martínez-Rosell et al., 2017*). Then, the systems were placed in a pre-equilibrated octahedral box using the four-point water model from the a99SB-disp force field (*Robustelli et al., 2018*), which is a modified version of TIP4P-D (*Piana et al., 2015*). The final systems considering the full-size enzyme contain the model protein, around 6500 water molecules, and 0.15 M of NaCl, forcing the system to be neutral, leading to simulation systems comprising around 29,000–30,400 atoms for Nsp1<sub>N</sub> systems, and approximately 96,400 atoms for the Nsp1<sub>FL</sub>-RNA complex.

All the simulations were performed using the a99SB-disp force field, which is a modified form of the a99SB force field that improves the modelling of intrinsically disordered peptides while retaining the accurate description of folded proteins. To parameterize the SL1 from RNA, the RNA-Shaw force field was used (*Tan et al., 2018*).

## MD simulations

### Unbiased MD simulations of Nsp1<sub>N</sub>

All the atomistic MD simulations were performed using the GROMACS 2021.3 package (*Abraham et al., 2015*) employing the a99SB-disp force field (*Robustelli et al., 2018*). Energy minimization was conducted using 50,000 steps of the steepest descent algorithm and setting the tolerance to 100 kJ mol<sup>-1</sup> nm<sup>-1</sup>. The equilibration process was performed in two steps, applying harmonic restraints to all heavy atoms in the system (harmonic constant: 1000 kJ mol<sup>-1</sup> nm<sup>-1</sup>). First, a 5 ns heating in the NVT ensemble was performed, using the V-rescale ( $\tau = 0.1$  ps) as thermostat (*Bussi et al., 2007*). Two different groups were used for temperature coupling: one for the protein and one comprising water molecules and ions. The reference temperature was set to 310 K. Second, the system was equilibrated during 15 ns in the NPT ensemble using V-rescale (*Bussi et al., 2007*) ( $\tau = 0.5$  ps) and Berendsen (*Berendsen et al., 1984*) ( $\tau = 0.5$  ps) as thermostat and barostat, respectively. The same temperature coupling groups were kept during the NPT equilibration step. The final structure from the equilibration process was used as a starting point for the MD simulations. All systems were simulated in the NPT ensemble with periodic boundary conditions using the same parameters as in the equilibration step and removing the harmonic restraints. The particle mesh Ewald method was used for treating long-range electrostatics using a cutoff of 12 Å (*Darden et al., 1993*). A time step of 2 fs was used for

all simulations after imposing constraints on the hydrogen stretching modes. We ran one replica of 1  $\mu$ s of Nsp1<sub>N</sub> from SARS-CoV-2, and one replica of 1  $\mu$ s of the three Nsp1 homologues (human SARS-CoV-1 and two Bat CoVs). For the Nsp1<sub>FL</sub>-RNA complex, we ran 1 replica of 500 ns for each model, A and B. Considering all the systems simulated leads to a total simulation time of 5 ms.

### Unbiased MD simulations of Nsp1<sub>FL</sub>

The AlphaFold 2.0 Nsp1<sub>FL</sub> structure model that was used to construct the Nsp1<sub>FL</sub>-RNA complex was isolated from the complex, parameterized with the a99SB-disp force field, and solvated with TIP4P-D water molecules in an octahedral box. We used the same energy minimization and equilibration protocol as the one we applied in the Nsp1<sub>N</sub> simulations. Post equilibration, we ran three independent replicas of Nsp1<sub>FL</sub> for 1  $\mu$ s each starting from the same conformation but with different initial velocities each.

### SWISH simulations

SWISH (Woo *et al.*, 2010; Min *et al.*, 2020) is a Hamiltonian replica-exchange enhanced sampling technique that increases the conformational sampling of proteins by scaling the interaction of the apolar atoms of the protein with the water molecules. In this way, water molecules acquire more hydrophobic physicochemical properties that allow them to induce the opening of hydrophobic cavities. Including organic fragments in the solvent during the SWISH simulations has been shown to stabilize the cavities that open up during the simulation (Comitani and Gervasio, 2018; Oleinikovas *et al.*, 2016). All SWISH simulations presented in this work, that is, of the Nsp1<sub>N</sub> of SARS-CoV-2 and the three CoV homologues, were run employing the same protocol: six different parallel 500 ns replicas, each one at a specific scaling factor ( $\lambda$ , ranging evenly from 1.00 to 1.35) value, in the presence of benzene (1 M concentration) as co-solvent. Since we ran six replicas per SWISH simulation, the total accumulated time is 12  $\mu$ s. Besides the scaling factors, all other parameters for both equilibration and production are the same as in the ones used for the unbiased MD simulations. Before each production run, six independent equilibration steps were carried out, one for each  $\lambda$  value. A contact-map-based bias was introduced for each replica to prevent the possible unfolding of the protein in high scaling factor replicas. The optimal upper wall value for the contact map was tuned based on the unbiased simulations. The benzene molecules for the mixed-solvent simulations were parametrized using Gaussian 16 (Frisch *et al.*, 2016) with Amber GAFF-2 force field (Mukherjee *et al.*, 2011) and RESP charges.

### Pocket detection

#### Crystal structures

In order to evaluate whether binding pockets exist in the Nsp1 and quantify their physicochemical properties, the crystal structure of Nsp1<sub>N</sub> (PDB entry 7K7P) was analyzed with Fpocket12 and DoGSiteScorer that is available as part of the modelling server ProteinPlus (Volkamer *et al.*, 2012). Fpocket is a geometry-based cavity detection algorithm that employs Voronoi tessellation and  $\alpha$  spheres to identify pockets in the protein structure. In this context, an  $\alpha$  sphere is defined as a sphere that contacts four atoms on its boundary and contains no internal atom. Similarly, DoGSiteScorer is an algorithm for pocket and druggability prediction that employs 3D difference of Gaussian filters to detect cavities and a support vector machine to score the identified binding sites.

#### MD simulations

The trajectories were analyzed with MDpocket (Dominguez *et al.*, 2003), an open source to detect binding pockets along MD simulations. MDpocket was run over down-sampled and reference-aligned trajectories with a time step of 100 ps between each frame. The corresponding minimized Nsp1<sub>N</sub> structure was chosen as a reference structure for each different system. The outputs let us identify and visualize the pockets observed throughout the whole simulation time with the PyMol software (PyMOL, 2022). This analysis has been performed along all the trajectories resulting from MD and SWISH simulations. Additionally, the software lets us compute the volume of the selected pockets.

## Druggability assessment

*FTDyn* and *FTMap* algorithms proved to be accurate in locating binding hotspots in proteins, that is, regions of the surface that majorly contribute to the free energy of binding. This approach is based on the fast and accurate distributions of 16 different small organic probes on the protein surface. Each probe is docked billions of times and map onto the target surface and scored according to an energy-based function (Dennis et al., 2002). Hence, the lowest energy probes are retained, locally minimized, and clustered. Ultimately, clusters of different probes are clustered together into a consensus cluster. The main binding hotspot is then identified as the consensus cluster containing the highest number of different probe fragments. We first employed *FTDyn*, a faster version of the *FTMap* algorithm without local minimization, to identify the most bound-like conformation in our unbiased MD simulations. We extracted 25 representative conformations from the MD simulation of SARS-CoV-2 Nsp1<sub>N</sub>, one from each of the first 25 most populated clusters (<https://github.com/Gervasiolab/Gervasio-Protein-Dynamics/tree/master/nsp1/ftmap>);). The 10,000-frame trajectory was clustered employing the *gromos* algorithm with a cutoff of 0.1 nm. Hence, we processed these structures with *FTDyn* server and determined the median number of interactions between probe molecules and each protein residue across the structural ensemble. We then assigned a contact score to each structure in the ensemble. The contact score is calculated as the sum of the number of residues in a given structure with a number of contacts higher than the median. We then calculated an ensemble score averaging over the 25 contact scores of the structures in the ensemble. Finally, we retained the structures in our 25-structure ensemble with a contact score higher than the average, as they were considered to be the best approximations of bound-like conformations. We further processed these 11 structures with *FTMap* to have a better prediction of binding hotspots on the Nsp1<sub>N</sub> surface.

## Crystallographic data

The N-terminal domain of SARS-CoV-2 Nsp1 was purified and crystallized as previously described (Ma et al., 2022). Fifty-nine potential fragment hits obtained from computational fragment screening of the Maybridge Ro3 library were purchased and fragments soaked into Nsp1<sub>N</sub> crystals and validated through X-ray diffraction experiments in quasi-automated mode at ESRF beamline MASSIF-1. Data analysis of the 59 datasets was conducted in the multi-crystal system PanDDA (Pearce et al., 2017). One hit was subsequently verified by manual inspection in COOT followed by refinement in Phenix.

## Thermal shift assay

The potential 59 fragments hits were tested using TSA at 2 mM containing 1% (v/v) DMSO using SARS-CoV-2 NSP1<sub>N</sub> at 1.25 mg/mL. The change in inflection temperature (Ti) would indicate a change in protein stability as the result of the protein-fragment interaction. NSP1<sub>N</sub> in the presence of 1% (v/v) DMSO was used as a control. Each fragment was tested in triplicate. The averaged Ti values for each fragment were calculated from each triplicate group. The change of inflection temperature ( $\Delta$ Ti) for fragments was calculated by subtracting the averaged Ti of the control group from the values of each triplicate group. Some fragments display atypical unfolding curves at 2 mM concentration. The Ti values of these were remeasured at lower concentrations of 1, 0.5, 0.25, or 0.125 mM. Correspondingly, 0.5%, 0.25%, 0.125%, or 0.0625% (v/v) DMSO in 1.25 mg/mL protein were used as control.

$$\Delta\text{Ti} = \text{averaged Ti (protein\_fragments)} - \text{averaged Ti (protein\_DMSO)}$$

## Microscale thermophoresis

To label SARS-CoV-2 Nsp1<sub>N</sub> with a fluorescent dye, 100 nM of RED-TRIS-NTA second-generation dye (MO-L018, NanoTemper, München, Germany) was mixed with 800 nM SARS-CoV-2 Nsp1<sub>N</sub> and incubated for 30 min on ice. The mixture was centrifuged in a Thermo Scientific Pico 17 MicroCentrifuge, 24-PI Rotor at 15,000× g for 10 min at 4°C to remove aggregates. The labelled protein was diluted to 100 nM in assay buffer (buffer E supplemented with 0.05% Pluronic(R) F-127) and the fragment hits 2E10 were diluted from 200 to 40 mM with assay buffer. Fifty  $\mu$ L of DMSO was mixed with 200  $\mu$ L of assay buffer as ligand buffer. The fragment solutions were serially diluted with ligand buffer with a dilution factor of 1.5, obtaining 16 fragment solutions. Then, an equal volume of protein solution was mixed with each diluted fragment solution and incubated for 30 min at 4°C to reach the binding equilibrium. The final fragment concentrations were from 20mM to 45.7  $\mu$ M, each containing 10%

DMSO. The final protein concentration in each sample was 50 nM. The samples were centrifuged at 6000 rpm for 10 min before the supernatant was being loaded into capillaries and detected in the Monolith NT.115 Pico instrument (NanoTemper, München, Germany) under the Pico-RED channel with 20% excitation power and 40% MST power under the Expert mode in the MO.Control software. The temperature was set at 25°C. The fragment hit 2E10 was also tested on SARS-CoV-2 Nsp1<sub>FL</sub> using the same method. All measurements were performed in triplicate. The data were analyzed, and the figures were generated in the MO.Affinity Analysis software.

## Phylogenetic analysis

The 283 sequences of different  $\alpha$ - and  $\beta$ -CoVs of different subgenera were obtained from the CDD, family accession number cl41742. The sequences were aligned with MEGA, version 11.0.11 (Tamura *et al.*, 2021). The resulting multi-sequence alignment was used to construct the maximum-likelihood tree with MEGA. The resulting tree was rendered with the iTol webserver (Letunic and Bork, 2021).

## Acknowledgements

We acknowledge PRACE and the Swiss National Supercomputing Centre (CSCS) for large supercomputer time allocations on Piz Daint, project IDs: pr126, s1107. FLG acknowledges the Swiss National Science Foundation and Bridge for financial support (projects number: 200021\_204795 and 40B2-0\_203628). We thank Dr Matthew Bowler at ESRF beamline Massif-1 for his excellent support.

---

## Additional information

### Competing interests

Shozeb Haider: Reviewing editor, eLife. The other authors declare that no competing interests exist.

### Funding

Funder	Grant reference number	Author
Swiss National Science Foundation	200021_204795	Francesco Luigi Gervasio
Swiss National Science Foundation	40B2-0_203628	Francesco Luigi Gervasio
Swiss National Supercomputing Centre	pr126	Francesco Luigi Gervasio
Swiss National Supercomputing Centre	s1107	Francesco Luigi Gervasio

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Alberto Borsatto, Data curation, Software, Investigation; Obaeda Akkad, Shymaa Damfo, Investigation; Ioannis Galdadas, Formal analysis, Writing – original draft; Shumeng Ma, Formal analysis, Investigation; Shozeb Haider, Conceptualization, Writing – review and editing; Frank Kozielski, Carolina Estarellas, Supervision, Investigation, Writing – original draft; Francesco Luigi Gervasio, Conceptualization, Resources, Supervision, Funding acquisition, Methodology, Writing – original draft, Project administration, Writing – review and editing

### Author ORCIDs

Alberto Borsatto  <http://orcid.org/0000-0002-8889-6491>

Ioannis Galdadas  <http://orcid.org/0000-0003-2136-9723>

Shozeb Haider  <http://orcid.org/0000-0003-2650-2925>

Carolina Estarellas  <http://orcid.org/0000-0002-0944-9053>

Francesco Luigi Gervasio  <http://orcid.org/0000-0003-4831-5039>

**Decision letter and Author response**Decision letter <https://doi.org/10.7554/eLife.81167.sa1>Author response <https://doi.org/10.7554/eLife.81167.sa2>**Additional files****Supplementary files**

- MDAR checklist

**Data availability**

The PDB file of the crystallographic structure (8A4Y) is available on the protein data bank: <https://www.rcsb.org/structure/8A4Y>.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Gervasio FL	2022	Revealing druggable cryptic pockets in the Nsp1 of SARS-CoV-2 and other $\beta$ -coronaviruses by simulations and crystallography	<a href="https://www.rcsb.org/structure/8A4Y">https://www.rcsb.org/structure/8A4Y</a>	RCSB Protein Data Bank', 8A4Y

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz Marc, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer Aron	2020	Conserved Domain Database (CDD)	<a href="https://www.ncbi.nlm.nih.gov/cdd">https://www.ncbi.nlm.nih.gov/cdd</a>	NCBI, CDD

**References**

- Abraham MJ**, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**:19–25. DOI: <https://doi.org/10.1016/j.softx.2015.06.001>
- Afsar M**, Narayan R, Akhtar MN, Das D, Rahil H, Nagaraj SK, Eswarappa SM, Tripathi S, Hussain T. 2022. Drug targeting nsp1-ribosomal complex shows antiviral activity against SARS-cov-2. *eLife* **11**:e74877. DOI: <https://doi.org/10.7554/eLife.74877>, PMID: 35323109
- Andersen KG**, Rambaut A, Lipkin WI, Holmes EC, Garry RF. 2020. The proximal origin of SARS-cov-2. *Nature Medicine* **26**:450–452. DOI: <https://doi.org/10.1038/s41591-020-0820-9>, PMID: 32284615
- Antczak M**, Popena M, Zok T, Sarzynska J, Ratajczak T, Tomczyk K, Adamiak RW, Szachniuk M. 2016. New functionality of rnacomposer: an application to shape the axis of mir160 precursor structure. *Acta Biochimica Polonica* **63**:737–744. DOI: [https://doi.org/10.18388/abp.2016\\_1329](https://doi.org/10.18388/abp.2016_1329), PMID: 27741327
- Berendsen HJC**, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**:3684–3690. DOI: <https://doi.org/10.1063/1.448118>
- Bussi G**, Donadio D, Parrinello M. 2007. Canonical sampling through velocity rescaling. *J Chem Phys* **126**:014101. DOI: <https://doi.org/10.1063/1.2408420>, PMID: 17212484
- Clark LK**, Green TJ, Petit CM. 2021. Structure of nonstructural protein 1 from SARS-cov-2. *Journal of Virology* **95**:e02019-20. DOI: <https://doi.org/10.1128/JVI.02019-20>, PMID: 33234675
- Comitani F**, Gervasio FL. 2018. Exploring cryptic pockets formation in targets of pharmaceutical interest with SWISH. *Journal of Chemical Theory and Computation* **14**:3321–3331. DOI: <https://doi.org/10.1021/acs.jctc.8b00263>, PMID: 29768914
- Corman VM**, Ithete NL, Richards LR, Schoeman MC, Preiser W, Drosten C, Drexler JF. 2014. Rooting the phylogenetic tree of middle east respiratory syndrome coronavirus by characterization of a conspecific virus

- from an african bat. *Journal of Virology* **88**:11297–11303. DOI: <https://doi.org/10.1128/JVI.01498-14>, PMID: 25031349
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses**. 2020. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it SARS-cov-2. *Nature Microbiology* **5**:536–544. DOI: <https://doi.org/10.1038/s41564-020-0695-z>, PMID: 32123347
- Darden T**, York D, Pedersen L. 1993. Particle mesh ewald: an N-log(N) method for ewald sums in large systems. *J Chem Phys* **98**:10089–10092. DOI: <https://doi.org/10.1063/1.464397>
- Dennis S**, Kortvelyesi T, Vajda S. 2002. Computational mapping identifies the binding sites of organic solvents on proteins. *PNAS* **99**:4290–4295. DOI: <https://doi.org/10.1073/pnas.062398499>, PMID: 11904374
- Dominguez C**, Boelens R, Bonvin A. 2003. Haddock: a protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society* **125**:1731–1737. DOI: <https://doi.org/10.1021/ja026939x>, PMID: 12580598
- Edgar RC**. 2004. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**:1792–1797. DOI: <https://doi.org/10.1093/nar/gkh340>, PMID: 15034147
- Frisch MJ**, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X. 2016. Gaussian 16, Revision C.01. Wallingford CT: Gaussian, Inc.
- Gervasiolab**. 2022. Gervasio-protein-dynamics. 4f1bb65. GitHub. <https://github.com/Gervasiolab/Gervasio-Protein-Dynamics>
- Jumper J**, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021. Highly accurate protein structure prediction with alphafold. *Nature* **596**:583–589. DOI: <https://doi.org/10.1038/s41586-021-03819-2>, PMID: 34265844
- Kozakov D**, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S. 2015. The ftmap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nature Protocols* **10**:733–755. DOI: <https://doi.org/10.1038/nprot.2015.043>, PMID: 25855957
- Le Guilloux V**, Schmidtke P, Tuffery P. 2009. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**:168. DOI: <https://doi.org/10.1186/1471-2105-10-168>, PMID: 19486540
- Letunic I**, Bork P. 2021. Interactive tree of life (ITOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**:W293–W296. DOI: <https://doi.org/10.1093/nar/gkab301>, PMID: 33885785
- Li W**, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Cramer G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF. 2005. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**:676–679. DOI: <https://doi.org/10.1126/science.1118391>, PMID: 16195424
- Lu S**, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* **48**:D265–D268. DOI: <https://doi.org/10.1093/nar/gkz991>, PMID: 31777944
- Ma S**, Damfo S, Lou J, Pinotsis N, Bowler MW, Haider S, Kozielski F. 2022. Two Ligand-Binding Sites on SARS-CoV-2 Non-Structural Protein 1 Revealed by Fragment-Based x-Ray Screening. *bioRxiv*. DOI: <https://doi.org/10.1101/2022.06.12.495816>
- Martínez-Rosell G**, Giorgino T, De Fabritiis G. 2017. PlayMolecule proteinprepare: a web application for protein preparation for molecular dynamics simulations. *Journal of Chemical Information and Modeling* **57**:1511–1516. DOI: <https://doi.org/10.1021/acs.jcim.7b00190>, PMID: 28594549
- Mendez AS**, Ly M, González-Sánchez AM, Hartenian E, Ingolia NT, Cate JH, Glaunsinger BA. 2021. The N-terminal domain of SARS-cov-2 nsp1 plays key roles in suppression of cellular gene expression and preservation of viral gene expression. *Cell Reports* **37**:109841. DOI: <https://doi.org/10.1016/j.celrep.2021.109841>, PMID: 34624207
- Min YQ**, Mo Q, Wang J, Deng F, Wang H, Ning YJ. 2020. SARS-cov-2 nsp1: bioinformatics, potential structural and functional features, and implications for drug/vaccine designs. *Frontiers in Microbiology* **11**:587317. DOI: <https://doi.org/10.3389/fmicb.2020.587317>, PMID: 33133055
- Mukherjee G**, Patra N, Barua P, Jayaram B. 2011. A fast empirical GAFF compatible partial atomic charge assignment scheme for modeling interactions of small molecules with biomolecular targets. *Journal of Computational Chemistry* **32**:893–907. DOI: <https://doi.org/10.1002/jcc.21671>, PMID: 21341292
- Narayanan K**, Huang C, Lokugamage K, Kamitani W, Ikegami T, Tseng CTK, Makino S. 2008. Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. *Journal of Virology* **82**:4471–4479. DOI: <https://doi.org/10.1128/JVI.02472-07>, PMID: 18305050
- Needleman SB**, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**:443–453. DOI: [https://doi.org/10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4), PMID: 5420325
- Oleinikovas V**, Saladino G, Cossins BP, Gervasio FL. 2016. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *Journal of the American Chemical Society* **138**:14257–14263. DOI: <https://doi.org/10.1021/jacs.6b05425>, PMID: 27726386
- Owen DR**, Allerton CMN, Anderson AS, Aschenbrenner L, Avery M, Berritt S, Boras B, Cardin RD, Carlo A, Coffman KJ, Dantonio A, Di L, Eng H, Ferre R, Gajiwala KS, Gibson SA, Greasley SE, Hurst BL, Kadar EP, Kalgutkar AS, et al. 2021. An oral SARS-cov-2 mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **374**:1586–1593. DOI: <https://doi.org/10.1126/science.abl4784>, PMID: 34726479

- Pearce NM**, Kelm S, Shi J, Deane CM, von Delft F. 2015. PANDDAs: multi-dataset methods for finding hits from fragment screening by X-ray crystallography. *Acta Crystallographica Section A Foundations and Advances* **71**:s258. DOI: <https://doi.org/10.1107/S2053273315096072>
- Pearce NM**, Krojer T, Bradley AR, Collins P, Nowak RP, Talon R, Marsden BD, Kelm S, Shi J, Deane CM, von Delft F. 2017. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nature Communications* **8**:15123. DOI: <https://doi.org/10.1038/ncomms15123>, PMID: 28436492
- Piana S**, Donchev AG, Robustelli P, Shaw DE. 2015. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The Journal of Physical Chemistry. B* **119**:5113–5123. DOI: <https://doi.org/10.1021/jp508971m>, PMID: 25764013
- Popenda M**, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW. 2012. Automated 3D structure composition for large rnas. *Nucleic Acids Research* **40**:e112. DOI: <https://doi.org/10.1093/nar/gks339>, PMID: 22539264
- PyMOL**. 2022. The pymol molecular graphics system. 2.0. Schrödinger, LLC. <https://pymol.org/2/>
- Robustelli P**, Piana S, Shaw DE. 2018. Developing a molecular dynamics force field for both folded and disordered protein states. *PNAS* **115**:E4758–E4766. DOI: <https://doi.org/10.1073/pnas.1800690115>, PMID: 29735687
- Sakuraba S**, Qilin X, Kasahara K, Iwakiri J, Kono H. 2021. Modeling the SARS-CoV-2 Nsp1–5′-UTR Complex via Extended Ensemble Simulations. *bioRxiv*. DOI: <https://doi.org/10.1101/2021.02.24.432807>
- Sali A**, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234**:779–815. DOI: <https://doi.org/10.1006/jmbi.1993.1626>, PMID: 8254673
- Schubert K**, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler LA, Leibundgut M, Thiel V, Mühlemann O, Ban N. 2020. SARS-cov-2 nsp1 binds the ribosomal mrna channel to inhibit translation. *Nature Structural & Molecular Biology* **27**:959–966. DOI: <https://doi.org/10.1038/s41594-020-0511-8>, PMID: 32908316
- Shen Z**, Ye G, Deng F, Wang G, Cui M, Fang L, Xiao S, Fu ZF, Peng G. 2018. Structural basis for the inhibition of host gene expression by porcine epidemic diarrhoea virus nsp1. *Journal of Virology* **92**:e01896–17. DOI: <https://doi.org/10.1128/JVI.01896-17>, PMID: 29237834
- Shen Z**, Wang G, Yang Y, Shi J, Fang L, Li F, Xiao S, Fu ZF, Peng G. 2019. A conserved region of nonstructural protein 1 from alphacoronaviruses inhibits host gene expression and is critical for viral virulence. *The Journal of Biological Chemistry* **294**:13606–13618. DOI: <https://doi.org/10.1074/jbc.RA119.009713>, PMID: 31350335
- Tamura K**, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution* **38**:3022–3027. DOI: <https://doi.org/10.1093/molbev/msab120>, PMID: 33892491
- Tan D**, Piana S, Dirks RM, Shaw DE. 2018. Rna force field with accuracy comparable to state-of-the-art protein force fields. *PNAS* **115**:E1346–E1355. DOI: <https://doi.org/10.1073/pnas.1713027115>, PMID: 29378935
- Tanaka T**, Kamitani W, DeDiego ML, Enjuanes L, Matsuura Y. 2012. Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mrna. *Journal of Virology* **86**:11128–11137. DOI: <https://doi.org/10.1128/JVI.01700-12>, PMID: 22855488
- Terada Y**, Kawachi K, Matsuura Y, Kamitani W. 2017. Mers coronavirus nsp1 participates in an efficient propagation through a specific interaction with viral RNA. *Virology* **511**:95–105. DOI: <https://doi.org/10.1016/j.virol.2017.08.026>, PMID: 28843094
- Thoms M**, Buschauer R, Ameisemeier M, Koepke L, Denk T, Hirschenberger M, Kratzat H, Hayn M, Mackens-Kiani T, Cheng J, Straub JH, Stürzel CM, Fröhlich T, Berninghausen O, Becker T, Kirchhoff F, Sparrer KMJ, Beckmann R. 2020. Structural basis for translational shutdown and immune evasion by the nsp1 protein of SARS-cov-2. *Science* **369**:1249–1255. DOI: <https://doi.org/10.1126/science.abc8665>, PMID: 32680882
- Tidu A**, Janvier A, Schaeffer L, Sosnowski P, Kuhn L, Hammann P, Westhof E, Eriani G, Martin F. 2020. The viral protein NSP1 acts as a ribosome gatekeeper for shutting down host translation and fostering SARS-cov-2 translation. *RNA* **27**:253–264. DOI: <https://doi.org/10.1261/rna.078121.120>, PMID: 33268501
- Vajda S**, Beglov D, Wakefield AE, Egbert M, Whitty A. 2018. Cryptic binding sites on proteins: definition, detection, and druggability. *Current Opinion in Chemical Biology* **44**:1–8. DOI: <https://doi.org/10.1016/j.cbpa.2018.05.003>, PMID: 29800865
- van Zundert GCP**, Rodrigues J, Trellet M, Schmitz C, Kastiris PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin A. 2016. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology* **428**:720–725. DOI: <https://doi.org/10.1016/j.jmb.2015.09.014>, PMID: 26410586
- Vankadari N**, Jeyasankar NN, Lopes WJ. 2020. Structure of the SARS-cov-2 nsp1/5′-untranslated region complex and implications for potential therapeutic targets, a vaccine, and virulence. *The Journal of Physical Chemistry Letters* **11**:9659–9668. DOI: <https://doi.org/10.1021/acs.jpcllett.0c02818>, PMID: 33135884
- Varadi M**, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Židek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, et al. 2022. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**:D439–D444. DOI: <https://doi.org/10.1093/nar/gkab1061>, PMID: 34791371
- Volkamer A**, Kuhn D, Rippmann F, Rarey M. 2012. DoGSiteScorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **28**:2074–2075. DOI: <https://doi.org/10.1093/bioinformatics/bts310>, PMID: 22628523



- Woo PCY**, Huang Y, Lau SKP, Yuen KY. 2010. Coronavirus genomics and bioinformatics analysis. *Viruses* **2**:1804–1820. DOI: <https://doi.org/10.3390/v2081803>, PMID: 21994708
- Yuan S**, Peng L, Park JJ, Hu Y, Devarkar SC, Dong MB, Shen Q, Wu S, Chen S, Lomakin IB, Xiong Y. 2020. Nonstructural protein 1 of SARS-cov-2 is a potent pathogenicity factor redirecting host protein synthesis machinery toward viral RNA. *Molecular Cell* **80**:1055–1066. DOI: <https://doi.org/10.1016/j.molcel.2020.10.034>, PMID: 33188728
- Zhang K**, Miorin L, Makio T, Dehghan I, Gao S, Xie Y, Zhong H, Esparza M, Kehrer T, Kumar A, Hobman TC, Ptak C, Gao B, Minna JD, Chen Z, García-Sastre A, Ren Y, Wozniak RW, Fontoura BMA. 2021. Nsp1 protein of SARS-cov-2 disrupts the mrna export machinery to inhibit host gene expression. *Science Advances* **7**:6. DOI: <https://doi.org/10.1126/sciadv.abe7386>, PMID: 33547084