bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Structure, Mechanism and Crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase

Joseph A Newman[1], Alice Douangamath[2], Setayesh Yazdani[3], Yuliana Yosaatmadja[1], Anthony Aimon[2], José Brandão-Neto[2], Louise Dunnett[2], Tyler Gorrie-stone[2], Rachel Skyner[2], Daren Fearon[2], Matthieu Schapira[3,4], Frank von Delft[1,2,5], Opher Gileadi[1*]

[1]Centre for Medicines Discovery, University of Oxford, ORCRB, Roosevelt Drive, Oxford, OX3 7DQ, UK
[2]Diamond Light Source Ltd., Harwell Science and Innovation Campus, Didcot OX11 0QX, UK
[3]Structural Genomics Consortium, University of Toronto, MaRS South Tower, Suite 700, 101 College Street, Toronto, ON, M5G 1L7, Canada
[4]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada
[5]Faculty of Science, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa
* Corresponding author. Tel. +44 1865 617572, email: opher.gileadi@cmd.ox.ac.uk

Abstract

The global COVID-19 pandemic is caused by the SARS-CoV-2 virus and has infected over 100 million and caused over 2 million fatalities worldwide at the point of writing. There is currently a lack of effective drugs to treat people infected with SARS-CoV-2. The SARS-CoV-2 Non-structural protein 13 (NSP13) is a superfamily1B helicase that has been identified as a possible target for anti-viral drugs due to its high sequence conservation and essential role in viral replication. In this study we present crystal structures of SARS-CoV-2 NSP13 solved in the APO form and in the presence of both phosphate and the non-hydrolysable ATP analogue (AMP-PNP). Comparisons of these structures reveal details of global and local conformational changes that are induced by nucleotide binding and hydrolysis and provide insights into the helicase mechanism and possible modes of inhibition. Structural analysis reveals two pockets on NSP13 that are classified as “druggable” and include one of the most conserved sites in the entire SARS-CoV-2 proteome. To identify possible starting points for anti-viral drug development we have performed a crystallographic fragment screen against SARS-CoV-2 NSP13 helicase. The fragment screen reveals 65 fragment hits across 52 datasets, with hot spots in pockets predicted to be of functional importance, including the druggable nucleotide and nucleic acid binding sites, opening the way to structure guided development of novel antiviral agents.

Introduction

SARS-CoV-2 is the causative agent of the current global coronavirus (COVID-19) pandemic, a severe respiratory disease that emerged in the Chinese city of Wuhan in late 2019(1,2). Genome sequencing indicates a zoonotic origin for SARS-CoV-2(3), as was the case for previous coronavirus outbreaks of SARS-CoV-1 in 2002(4) and MERS in 2012(5). SARS-CoV-2 belongs to the genus *Betacoronavirus* which have a positive sense single stranded RNA genome of approximately 30 KB in length and amongst the largest know of any viral RNA genomes. The SARS-CoV-2 genome encodes two open reading frames ORF1a and 1b, that when translated produce polyproteins that are processed by viral proteases into 16 non-structural proteins (NSP1-16)(6) that collectively form the machinery for viral replication and transcription.

NSP13 is a 67 kDa protein that belongs to the helicase superfamily 1B, it utilizes the energy of nucleotide triphosphate hydrolysis to catalyze the unwinding of double stranded DNA or RNA in a 5’ to 3’

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

direction(7). Although NSP13 is believed to act on RNA *in vivo* enzymatic characterization shows a significantly more robust activity on DNA in *in vitro* assays with relatively weak non processive helicase activity when compared to other superfamily 1B enzymes(8,9). NSP13 has been shown to interact with the viral RNA-dependent RNA polymerase NSP12 (10,11), and acts in concert with the replication-transcription complex (NSP7/NSP8/NSP12)(12). This interaction has been found to significantly stimulate the helicase activity of NSP13 possibly by means of mechanoregulation (11,13). In addition to its helicase activity NSP13 also possess RNA 5' triphosphatase activity within the same active site(14), suggesting a further essential role for NSP13 in the formation of the viral 5' mRNA cap.

NSP13 contains 5 domains, a N-terminal Zinc binding domain (ZBD) that coordinates 3 structural Zinc ions, a helical "stalk" domain, a beta-barrel 1B domain and two "RecA like" helicase subdomains 1A and 2A that contain the residues responsible for nucleotide binding and hydrolysis. Previous X-ray structures of NSP13 have been solved for MERS-CoV and the highly related SARS-CoV to 3.0 Å and 2.8 Å respectively (10,15). More recently, cryo-electron microscopy studies have revealed the architecture of the NSP13 containing replication and transcription complex, which contains two copies of NSP13 that interact with NSP8 via the N-terminal ZBD(12,16,17) (7CXM). One of the NSP13 protomers makes additional interactions with NSP12 and is located with its RNA binding site in the path of downstream RNA(12). However the polarity of the helicase translocation 5' to 3' and polymerase are in opposition, leading to the suggestion that NSP13 may play a role in backtracking, template switching or disruption of downstream secondary structures (12). It is also not well understood why two copies of the helicase are present, although mutagenesis of specific residues involved in discrete domain contacts indicate a role for both domains for the enhanced helicase activity of the complex (7XCM).

Whilst the precise role for NSP13 in the viral life cycle has yet to be determined, it was found to be a critical component for viral replication in the highly similar SARS-CoV-1(10), and in other viral species more distantly related to SARS-CoV-2(18-20). For this reason NSP13 has been suggested as a good target for the development of new antiviral drugs(21,22). To this end, several efforts have identified compounds which inhibit SARS-CoV NSP13 with IC50's in the low µM range and display anti-viral activity in cellular assays(23-25). The lack of structural information on the binding mode and unknown mode of action of these compounds is likely a barrier to further development although these studies demonstrate that NSP13 may be a tractable target. NSP13 is also among the most conserved of the non-structural proteins in the SARS-CoV-2 genome, differing from SARS-CoV-1 in only a single amino acid (V570I) . Thus, compounds targeting SARS-CoV-2 NSP13 would likely be effective against SARS-CoV-1 and potentially other future emerging coronaviruses, making it an ideal target for the development of new antiviral therapeutics with a broad spectrum of action.

To accelerate the development of anti-viral therapeutics targeting NSP13 we have determined the X-ray crystal structure to high resolution in a variety of nucleotide bound and conformational states. These structures provide insights into the catalytic mechanism and provide a robust platform for virtual screening. We have analyzed potential druggable pockets on NSP13 and find two such pockets, one of which is highly conserved across a variety of viral species. Finally, we have performed a crystallographic fragment screen on NSP13 which demonstrates the ligandability of these pockets and provides structural information and chemical starting points for the rational design or structure guided optimization of new anti-viral therapeutics.

Results and Discussion

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Crystal structures of NSP13 in APO, phosphate and nucleotide bound form

We have determined the crystal structures of full length SARS-CoV-2 NSP13 (residues 1-901) in three different forms that represent different conformations of the ATPase catalytic cycle. The APO form was determined at 2.4 Å resolution with two NSP13 molecules in the asymmetric unit. A phosphate-bound form was determined with the same crystal form but diffracted to significantly higher resolution (1.9 Å). A nucleotide bound form of NSP13 with the ATP analogue AMP-PNP was determined to 3.0 Å resolution in a monoclinic crystal system with 4 molecules of NSP13 in the asymmetric unit. In general, the electron density is of high quality throughout (Figure S1), with the exception of the final 7 residues at the C-terminus and two loops within the 1B domain (residues 185-194 and 203-207), which display variable degrees of disorder across the various chains. The models have been restrained to standard bond lengths and angles and a summary of the data collection and refinement statistics can be found in Table 1.

All chains feature the same basic architecture which has been described previously for NSP13 structures of related organisms(10,15) (Figure 1A). The ZBD coordinates two structural zinc ions via a RING like module in the N-terminus and an additional single zinc ion via a treble-clef Zinc finger. The ZBD packs against the three-helix bundle of the "stalk" domain which also makes contacts with 1A domain on its opposite side. Following closely from the end of the third helix of the stalk domain is the 1B domain which forms a 6-stranded RIFT type anti-parallel β-barrel(26). A long 30 amino acid linker that does not adopt any secondary structure links the 1B to the 1A and 2A RecA-like core helicase domains. The nucleotide binding site is situated in a cleft between the 1A and 2A domains with specific contacts to the nucleotide provided by conserved helicase motifs I, II and III in the 1A domain and IV, V and VI in the 2A domain.

Comparisons of NSP13 nucleotide bound structures in different conformational states

In the nucleotide bound form, AMP-PNP is bound to all 4 molecules in the asymmetric unit but in two distinctly different modes (Figure 1B, 1C). In mode A (exemplified by chain A of the nucleotide form) both AMP-PNP and a hydrated Mg^{2+} ion are bound, with the adenine base making stacking interactions to H290 on one side and R442 on the other. The ribose O3 makes a single hydrogen bond to E540 (part of motif V), and the α-phosphate interacts with the main chain of H290 and the side chain of K320 via a salt bridge. The β-phosphate is situated directly within the phosphate binding motif I and makes interactions with the main chain amides of residues 287-289. The γ-phosphate makes extensive interactions with the sidechains of Q404, R443 and R567 (from motif III, IV and VI respectively). The Mg^{2+} ion is in an octahedral coordination and makes direct contacts to both the β and γ phosphates as well as the side chain of S289 and additional three waters. The conserved D374 and E375 from motif II (walker B) make further interactions with the waters and are poised to perform their presumed catalytic role of activating a hydroxyl for nucleophilic attack (Figure 1B).

In nucleotide binding mode B (chains B, C and D of the nucleotide form) the interaction environment of the adenine base and γ-phosphate are largely unchanged, with the ribose O2 contacting the side chain of K320. The α-phosphate is situated in the equivalent position of the β-phosphate from mode A, with the β-phosphate interacting with R443 and the invariant K288 from motif I contacting oxygens on both α and γ phosphates (Figure 1C). No density could be observed for Mg^{2+} ions in these three chains with the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sidechain of S289 contacting instead the α-phosphate and the residues from motif II more distant. In the phosphate bound structure the two phosphates bind in essentially equivalent positions to the α and γ phosphates from binding mode B, although very slight shifts are observed for the equivalent of the γ-phosphate which probably represents the expected binding poise for the hydrolyzed phosphate product (Figure S2).

The structures of NSP13 in complex with nucleotide also offer insights into the ability of NSP13 to perform the RNA 5' triphosphatase reaction. In both binding modes (but particularly mode A) large portions of the adenine and ribose moiety are solvent exposed with the ribose O3 group pointing towards the outside of the active site such that it would be possible to accommodate additional phosphodiester linkages such as those found on a 5' triphosphate containing RNA without steric clashes (Figure S3). The lack of specific contacts in this area would also explain the relaxed nucleotide specificity that NSP13 displays in *in vitro* assays(14).

Flexibility and domain movements in the structures of NSP13

In addition to the specific changes in the nucleotide binding site detailed above there are also significant domain movements that become evident when comparing the various chains of NSP13 structures. A systematic pairwise comparison of Cα RMSD values for all NSP13 chains from this study and relevant PDB entries is shown in supplementary table S1. Whilst the individual domains of NSP13 are generally well conserved (in the region of 0.8 Å RMSD or lower), the differences when comparing entire chains are in some cases significantly greater (in the region of 3Å). These differences can be attributed to rigid domain movements and structural superpositions identify multiple sources of domain movements (Figure 2A-D). Firstly, the Zinc binding domain appears to be fairly flexible with rotations of 10-15° around a single hinge point located near residues 104-105 (Figure 2C). There is no obvious connection between the conformation of this domain and crystal form or nucleotide bound status as variability between chains within a single crystal forms is equally large as when comparing between forms. More directly related to the nucleotide status is the relative conformations adopted by the 1A and 2A domains which shifts by around 20° and adopts a closed conformation the Mg^{2+} AMP-PNP bound crystals structure, and a more open conformation in the APO, Phosphate bound and other AMP-PNP bound chains (Figure 2D). The 1B domain also displays significant conformational variation with rotations of around 25 ° (Figure 2B). In this case the movements do correlate with the relative positioning of the 1A and 2A domains, such that the respective interfaces are maintained whist the 1B domain pivots around its relatively narrow point of attachment at the terminus of the stalk domain (Figure 2B).

Model for the translocation mechanism of NSP13

Previous structural studies on UPF-1(27) and recent Cryo-EM structures of NSP13(16) within the viral replication-transcription complex reveal the expected DNA/RNA binding site which is formed from a channel bounded on one side by the 1A and 2A domains and the other by the 1B and Stalk domains. Although the precise details are still to be revealed, the RNA at its 5' end appears to interact with the 2A domain via the phosphodiester backbone with the nucleobases pointing towards the 1A and 1B domains, whilst at the 3' end the RNA backbone contacts the 1A domain with the bases pointing towards the stalk and 1B domain (Figure 3A). Positioning RNA into its expected binding site in both the more "open" APO/product state and the "closed" ATP-Mg^{2+} bound conformations, the RNA contacting motifs on the 1A and 2A domains are shifted such that conserved contacts may be formed to

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

phosphates that differ by a single RNA base. This observation has allowed us to construct an inchworm type mechanistic model for the translocation of NSP13 along single stranded RNA in a 5' to 3' direction. In our model the "closed" conformation is assumed to represent an activated pre-hydrolysis state which interacts with the RNA most strongly via the 2A domain and more weakly via the 1A domain at the 3' end (Figure 3A). ATP hydrolysis triggers the conformational change to the open product state with the 1A domain sliding a single nucleotide step toward the 3' end of the RNA (Figure 3A). Based on the requirement for directional movement, we assume that following this structural transition, the 1A domain would switch to bind the RNA more tightly either as a result of the remodeled RNA interface or from ADP and phosphate release. The nature of the nucleotide binding site would suggest that ADP would be released before phosphate due to steric effects. The catalytic cycle is completed by binding a new ATP molecule which triggers the closed conformation, although this time the 2A domain moves relative to the RNA (Figure 3A & Supplementary Movie S1). A requirement of this model is a mechanism for ATP binding and subsequent hydrolysis to induce conformational changes and for the RNA binding affinity of one or both of the 1A or 2A domains to be modulated according to the nucleotide or conformational status. Comparing the mode of nucleotide binding in the open and closed states, the nucleotide is bound in two different modes, with the γ-phosphate in mode B (product state) being considerably more distant from the phosphate binding residues in motif I. We presume that charge repulsion following ATP hydrolysis would induce this conformation and movement of the 2A domain is induced due to the extensive interactions between it and the γ-phosphate provided by residues R443, Q537 and R567 from motifs IV, V and VI respectively (Figure 3B). The means by which the RNA binding affinities may be modulated are more uncertain partly due to the lack of a detailed understanding of the RNA-protein interface in both conformational states. One possibility is that this remodeling is a result of the coordinated movements of the 1B domain which forms a significant part of the RNA protein interface and appears to contact the RNA more extensively close to the 1A domain in the open conformation and close to the 2A domain in the closed conformation (Figure 3C). The calculated interface areas between RNA and protein is thus more extensive with the 3' end in the closed conformation (558 $Å^2$ versus 528 $Å^2$) and with the 5' end in the open conformation (595 $Å^2$ versus 574 $Å^2$), matching the preference required for 5' to 3' directional translocation. Another possibility is that the RNA interface is remodeled via contacts from regions proximal to key helicase motifs such as the loop following motif III in the 1A domain, and helices preceding motif V and following motif VI in the 2A domain which form part of the RNA interface whilst also contacting the γ-phosphate or hydrolyzed product in a conformation dependent manner (Figure 3B).

This mechanistic model is similar to the mechanisms suggested previously for other SF1B helicases such as RecD2(28), and provides a framework for the understanding of the NSP13 translocation mechanism and possible sites of inhibition including possible allosteric sites that may differ between the two states and block structural transitions that occur as part of the catalytic cycle. We do not describe an active base separating mechanism for NSP13 consistent with biochemical and single molecule analysis of NSP13 which was found to be a predominantly passive helicase (advancing upon the spontaneous opening of base pairs), with a strong force dependent stimulation of activity that suggests mechanoragulation by the RNA polymerase NSP12(8).

X-ray Crystallographic fragment screening of NSP13

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To identify possible starting points for the development of NSP13 inhibitors we have performed an X-ray crystallographic fragment screen. The phosphate-bound NSP13 crystals show robust crystallization behavior and routinely diffract to around 2.0 Å resolution, sufficient for the reliable identification of binders. 648 crystals were soaked individually with a library of chemical fragments at a ~50 mM. X-ray analysis of the majority of crystals showed no bound compounds; however, 65 bound fragments were found within 52 datasets using the PANDDA algorithm (29)., Some of the hits were found in pockets predicted to be of functional importance, including the nucleotide and nucleic acid binding sites (Figure 4A, & Supplementary table I; the bound fragment structures are be referred to by the corresponding PBD depositions). In the nucleotide binding site, 15 fragments were bound in positions overlapping with the ATP ribose and adenine moiety (Figure 4B). three of these fragments (PDB:5RLI, 5RLJ and 5RLW) contain sulfonamide functional groups which make polar contacts to key residues within motif I and the phosphate ion occupying the α-phosphate position. The rest of the fragments explore the wider vicinity of the pocket occupied by the adenine moiety, making interactions with nearby residues such as H290, K320, Y342, R442 and N464.

Several fragment clusters were found in or close to the RNA/DNA binding channel (Figure 4C). Close to the 5' end of the channel, three fragments (5RLH, 5RLZ and 5RMM) occupy a pocket formed between the 2A and 1B domains and make polar contacts to residues S486, N516, Y515 and T552 on the 2A domain (Figure 4C & S4). These contacts are direct mimics contacts formed by two successive RNA phosphates in the structure of the related UPF1 helicase in complex with RNA(27), making these fragments particularly attractive starting points for the design of RNA competitive inhibitors. In the central cavity of the channel a single fragment (5RML) occupies a hydrophobic pocket created by the side chains of F145, K146, Y149, Y180 and T410, and makes further polar interactions to E142 and P408 on the periphery (Figure 4C). A single sulfonamide containing fragment (PDB entry 5RLK) occupies a similar conserved phosphate binding region on the 1A domain and makes contacts to T359 and H311 as well as inducing the ordering of loop 337-340 (via contacts to R337) which is disordered in the ground state model (Figure 4C & S5). This fragment would also be expected to be RNA competitive based on the UPF1 RNA complex(27). Also bound near the RNA interface where the 3' end would be expected to exit are two fragments (5RL8 and 5RMC) which form contacts to residues in the 1A domain and would be expected to block RNA from entering the cleft.

Our analysis of the NSP13 mechanism and flexibility also suggests the importance of fragments that may be potential starting points for allosteric inhibitors, due to binding in sites that are specific to one conformational state and may block structural transitions that occur as part of the catalytic cycle. In addition to the nucleotide and RNA binding pockets identified above, three fragment sites were identified that span domains that exhibit conformational variability. The most prominent of such sites is found in a cleft between the Zinc domain and the stalk domain, which bound to 11 fragments, several of which share a common mode of interaction and are candidates for fragment merging (Figure 5A). Two fragments (5RMF and 5RMB) were bound to a shallow, predominantly hydrophobic pocket between the 1A and 2A domains that is formed on the opposite face to the nucleotide binding site, and appears to be present only in the open product state conformation (Figure 5B). A further 3 fragments were bound to at the junction of the 1B and stalk domain (5RL6, 5RL7 and 5RLU) which is also close to the 3' end of the RNA binding interface (Figure 5C), although part of this site is formed by contributions from a crystallographic neighbor.

Druggablilty and conservation analysis of NSP13 binding pockets

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In addition to the experimental fragment screening, we used a computational approach (ICM; Molsoft, San Diego) to find potentially druggable binding sites in our NSP13 structures. We identified two pockets of interest that are expected to be functionally relevant (Figure 6A): the binding site occupied by AMP-PNP, at the interface of domains 1A and 2A, and a pocket lined by domains 1A, 1B and 2A which is occupied by the 5'-end of the substrate RNA in the SARS-Cov-2 transcription complex [PDB:7CXM](16). This 5'-RNA pocket is also occupied by some of our soaked fragments [PDB:5RMM, 5RLH, 5RLZ]. Both sites are accessible in the context of the transcription complex structure (Figure 6A), suggesting that they could be targeted pharmacologically. We used SiteMap to evaluate the druggability of these two sites(30). The 5'-RNA pocket is clearly druggable (druggability score 1.03). The nucleotide site is highly charged and developing drugs targeting this site is predicted to be more challenging (druggability score 0.91).

Ideally, a drug against COVID-19 would also be effective against past and future coronaviruses. To behave as a broad-spectrum inhibitor, a pharmacological agent targeting NSP13 would need to exploit a binding site that is highly conserved across coronaviruses. To evaluate the relevance of the nucleotide and 5'-RNA binding pockets for the development of pan-coronavirus drugs, we analyzed the conservation of amino-acids lining these two pockets across twenty-seven α- and β-coronaviruses with reviewed sequences in Uniprot. We find that 79% and 87% of residues lining the nucleotide and 5'-RNA sites respectively are conserved in all analyzed coronaviruses (Figure 6B). Additionally, 100% of side-chains in direct contact with our co-crystallized fragments (PDB:5RMM, 5RLZ) are conserved. In a systematic analysis of the sequence conservation of nineteen binding pockets from fifteen SARS-CoV-2 proteins across the same twenty-seven coronaviruses, we find that the active site of RdRp/NSP12 (94% sidechains conserved) and the ADP-bound pocket of the NSP12 NiRAN domain (87% sidechains conserved) are the only two cavities with a degree of conservation as high as the 5'RNA site of NSP13 (Yazdani et al. bioRxiv 2021). Together, our results indicate that the binding pocket occupied by the 5'-end of the RNA substrate is druggable and highly conserved. As such it is a good candidate for the development of broad-spectrum inhibitors.

Summary

We have determined the crystal structure of SARS-CoV-2 NSP13 in APO, phosphate-bound and nucleotide bound states. These structures are of good quality and the crystals significantly higher resolution than previous structures of related NSP13 proteins and are thus a good starting point for virtual ligand screening. Our analysis of the structures of NSP13 revealed a high degree of conformational heterogeneity, with two distinct "open" and "closed" forms identified in the AMP-PNP complex crystals which represent different states in the catalytic cycle. We have used this information to suggest a translocation mechanism for NSP13 based on the transition between these states, with concomitant modulation of RNA binding interfaces producing directional movement along single stranded RNA. We have utilized the robust crystallization and high-resolution diffraction of the phosphate bound crystals to perform crystallographic fragment screening on NSP13. The screen identified over 50 binders with fragment hotspots in both nucleotide and RNA binding channels that can be used as starting points for design of novel anti-virals. Our analysis of the mechanism has also allowed us to identify fragments that may serve as possible starting points for allosteric inhibition based on their ability to interfere with or block structural transitions that form part of the catalytic cycle. Finally, we have assessed the druggability and sequence conservation of pockets on NSP13, this analysis shows that a fragment containing pocket on the 5' end of the RNA binding site is highly druggable and amongst the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

most well conserved pockets in the entire SARS-CoV-2 proteome, making it a good target for the development of anti-viral therapeutics that may be able to combat the current pandemic and also future emerging viral threats.

Materials and Methods

Cloning and Expression of NSP13

The plasmid for N-terminally His-ZB tagged NSP13 was synthesized in a pNIC-ZB vector (Twist biosciences) with codon optimization for expression in *E.coli.* The plasmid and its full sequence have been deposited in Addgene (https://www.addgene.org/159614/). For overexpression the plasmid was transformed into *E. coli* BL21 Rosetta2 cells. Cell cultures were grown in Terrific Broth media at 37°C, with shaking at 180 rpm. When the OD_{600} reached 2-3, IPTG (300uM) was added to the media and cultures were incubated overnight at 18°C, shaking 180 rpm.

Protein Purification

Cell pellets were re-suspended in lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 10 mM Imidazole, 0.5 mM TCEP) with protease inhibitors (Merck Protease inhibitor cocktail III, 1:500). Cells were disrupted by sonication for 15 mins 10sec on 5sec off, and cell debris was removed by centrifugation in a JA25.5 rotor at 24500 RPM for 30 mins. The supernatant was incubated for 40 mins with 5ml of Ni resin (IMAC sepharose) for batch binding. The tubes containing the lysate were centrifuged at 700 x g at 4°C for 5 minutes and the supernatant discarded. Beads were loaded on a gravity flow column and washed with 40 ml lysis buffer, 25 ml wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 45 mM Imidazole, 0.5 mM TCEP). A further wash with 10 ml Hi-salt buffer (50 mM HEPES pH 7.5, 1 M NaCl, 5% Glycerol, 0.5 mM TCEP) and again with another 10 ml of wash buffer. Proteins were eluted with addition of 15ml of elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% Glycerol, 300 mM Imidazole, 0.5 mM TCEP). The elution fraction was immediately applied to a 5ml HItrap SP column using a syringe, collecting the flow through. The SP column was washed with 10ml elution buffer and proteins were eluted with 15ml Hi-salt buffer. The NSP13 protein was found to be present in flow-through and elution fractions and both fractions were pooled and treated separately from this point onward. For further purification protein samples were incubated overnight with TEV protease (1:40 mass ratio) and loaded onto gel filtration using a superdex 200 16/60 column equilibrated in 50 mM HEPES, 500 mM NaCl, 0.5 mM TCEP. Both samples were found to crystallize with the majority of the crystals coming from the SP flow through which had greater yield although slightly less pure.

Crystallization

All crystals were grown from and optimized using pre-prepared mixes from the Morpheus screen from molecular dimensions. Phosphate bound crystals (used for fragment screening) were grown at 20 degrees from conditions containing 20 % Ethylene Glycol, 10 % PEG 8K, 0.05 M HEPES pH , 0.05 M MOPS, 0.03 M Sodium Nitrate, 0.03 M Sodium Phosphate, 0.03 M Ammonium Sulphate using 10 mg/ml protein. For crystal optimization seeding was performed: 5-10 crystals were crushed with glass probe and transferred to 25 ul of well solution. A seed bead was added, and the mixture was sonicated for around 30-60 seconds with pulsing. Final seeding was performed with a 1 in 400 dilution of seed stock. Final plates were set up with protein at 5 mg/ml (diluted 4 fold in water from 20 mg/ml stock) with a

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

slightly reduced precipitant concentration (16 % Ethylene Glycol, 8 % PEG 8K, 0.05 M HEPES, 0.05 M MOPS, 0.03 M Sodium Nitrate, 0,03 M Sodium Phosphate, 0.03 M Ammonium Sulphate), using 300 nl drops (1:1 ratio) with 20 nl seeds (added last). Crystals without phosphate were grown at 10 mg/ml from conditions containing 20 % Ethylene Glycol, 10 % PEG 8K, 0.05 M MES pH 6.5, 0.05 M Imidazole pH 6.5, 10% v/v Ethylene glycols mix (contains 0.3M Diethylene glycol, 0.3M Triethylene glycol, 0.3M Tetraethylene glycol, 0.3M Pentaethylene glycol). For crystals containing AMP-PNP, 10 mM AMP-PNP and 10 mM $MgCl_2$ were added to the protein and incubated on ice for 10 minutes. Crystallization was performed at 5 mg/ml and crystals appeared at 20 degrees in conditions containing 20 % Ethylene Glycol, 10 % PEG 8K, 0.05 M MES pH 6.5, 0.05 M Imidazole pH 6.5, 10% v/v Alcohols mix (contains 0.2 M 1,6-Hexanediol, 0.2 M 1-Butanol, 0.2 M 1,2-Propanediol, 0.2 M 2-Propanol, 0.2 M 1,4-Butanediol, 0.2 M 1,3-Propanediol). All crystals were loop mounted and flash cooled in liquid nitrogen without the addition of further cryoprotectant.

Structure Determination

All data were collected at Diamond light source beamline I04-1 and processed using XDS(31) and DIALS(32). The structures were solved by molecular replacement using the program PHASER and the structure of SARS-CoV-1 NSP13 (6JYT) as a search model. Refinement was performed using PHENIX REFINE(33). A summary of the data collection and refinement statistics are shown in table 1.

X-ray Fragment Screening

A total of 648 fragments from the DSI poised and York3D libraries (500 mM stock concentration dissolved in DMSO) were transferred directly to NSP13 crystallization drops using an ECHO liquid handler (50 mM nominal final concentration), and soaked for 1-3hr before being loop mounted and flash cooled in liquid nitrogen. A total of 616 datasets were collected at a resolution of 2.8 Å or higher with the majority being in the range of 1.8 Å to 2.4 Å. Data were collected at Diamond light source beamline I04-1 and processed using the automated XChem Explorer pipeline. Structures were solved by difference Fourier synthesis using the XChem Explorer pipeline(34). Fragment hits were identified using the PanDDA(29) program. Refinement was performed using REFMAC(35) or BUSTER.

Druggabity and conservation analysis

The PocketFinder function implemented in ICM (version 3.9-2b) (Molsoft, San Diego) was used to map binding pockets on our apo and AMP-PNP complex structures of NSP13(36). Druggability was calculated with SiteMap (Release 2019-4 of Maestro – Schrodinger, New-York)(30). The AMP-PNP complex structure was used to calculate the druggability of the nucleotide site after removing the bound nucleotide and its coordinating magnesium. The druggability of the 5'-RNA site was calculated on the fragment-bound complex structure (PDB 5RMM) after removing the fragment. The multiple alignment of twenty-seven coronavirus helicase sequences was carried-out with ICM(37).

Acknowledgements

The crystallographic screen was supported by the XChem facility at Diamond Light Source (proposal ID LB26998). We thank all the staff of Diamond Light Source for providing support and encouragement which allowed us to carry out this work during the COVID-19 lockdown.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, Merck KGaA Darmstadt Germany, MSD, Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome [106169/ZZ14/Z].

Data and Materials Availability

The plasmid for expression of nsp13 is available from Addgene (https://www.addgene.org/159614/).

Crystallographic coordinates and structure factors for all structures have been deposited in the Protein Data Bank with the following accessing codes: 6ZSL, 7NIO, 7NN0, 5RL6, 5RL7, 5RL8, 5RL9, 5RLB, 5RLC, 5RLD, 5RLE, 5RLF, 5RLG, 5RLH, 5RLI, 5RLJ, 5RLK, 5RLL, 5RLM, 5RLN, 5RLO, 5RLP, 5RLQ, 5RLR, 5RLS, 5RLT, 5RLU, 5RLV, 5RLW, 5RLY, 5RLZ, 5RM0, 5RM1, 5RM2, 5RM3, 5RM4, 5RM5, 5RM6, 5RM7, 5RM8, 5RM9, 5RMA, 5RMB, 5RMC, 5RMD, 5RME, 5RMF, 5RMG, 5RMH, 5RMI, 5RMJ, 5RMK, 5RML, 5RMM, 5ROB, 7NNG

References

1. Chan, J. F., Yuan, S., Kok, K. H., To, K. K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C., Poon, R. W., Tsoi, H. W., Lo, S. K., Chan, K. H., Poon, V. K., Chan, W. M., Ip, J. D., Cai, J. P., Cheng, V. C., Chen, H., Hui, C. K., and Yuen, K. Y. (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514-523
2. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., and Zhang, Y. Z. (2020) A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269
3. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., Zhao, K., Chen, Q. J., Deng, F., Liu, L. L., Yan, B., Zhan, F. X., Wang, Y. Y., Xiao, G. F., and Shi, Z. L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273
4. Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguiere, A. M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D., Schmitz, H., and Doerr, H. W. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **348**, 1967-1976
5. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* **367**, 1814-1820
6. Chan, J. F., Kok, K. H., Zhu, Z., Chu, H., To, K. K., Yuan, S., and Yuen, K. Y. (2020) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* **9**, 221-236

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

7. Tanner, J. A., Watt, R. M., Chai, Y. B., Lu, L. Y., Lin, M. C., Peiris, J. S., Poon, L. L., Kung, H. F., and Huang, J. D. (2003) The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5' to 3' viral helicases. *J Biol Chem* **278**, 39578-39582
8. Mickolajczyk, K. J., Shelton, P. M. M., Grasso, M., Cao, X., Warrington, S. E., Aher, A., Liu, S., and Kapoor, T. M. (2020) Force-Dependent Stimulation of RNA Unwinding by SARS-CoV-2 nsp13 Helicase. *Biophys J*
9. Jang, K. J., Jeong, S., Kang, D. Y., Sp, N., Yang, Y. M., and Kim, D. E. (2020) A high ATP concentration enhances the cooperative translocation of the SARS coronavirus helicase nsP13 in the unwinding of duplex RNA. *Sci Rep* **10**, 4481
10. Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., Zheng, L., Ming, Z., Zhang, L., Lou, Z., and Rao, Z. (2019) Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res* **47**, 6538-6550
11. Adedeji, A. O., Marchand, B., Te Velthuis, A. J., Snijder, E. J., Weiss, S., Eoff, R. L., Singh, K., and Sarafianos, S. G. (2012) Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS One* **7**, e36521
12. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P. M. M., Olinares, P. D. B., Maruthi, K., Eng, E. T., Vatandaslar, H., Chait, B. T., Kapoor, T. M., Darst, S. A., and Campbell, E. A. (2020) Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **182**, 1560-1573 e1513
13. Mickolajczyk, K. J., Shelton, P. M. M., Grasso, M., Cao, X., Warrington, S. R., Aher, A., Liu, S., and Kapoor, T. M. (2020) Force-dependent stimulation of RNA unwinding by SARS-CoV-2 nsp13 helicase. *bioRxiv*
14. Ivanov, K. A., Thiel, V., Dobbe, J. C., van der Meer, Y., Snijder, E. J., and Ziebuhr, J. (2004) Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol* **78**, 5619-5632
15. Hao, W., Wojdyla, J. A., Zhao, R., Han, R., Das, R., Zlatev, I., Manoharan, M., Wang, M., and Cui, S. (2017) Crystal structure of Middle East respiratory syndrome coronavirus helicase. *PLoS Pathog* **13**, e1006474
16. Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., Jia, Z., Wang, H., Huang, Y., Li, M., Wang, Q., Rao, Z., and Lou, Z. (2020) Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat Commun* **11**, 5874
17. Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M., Liu, Z., Wang, H., Li, Y., Chen, Y., Guddat, L. W., Wang, Q., Rao, Z., and Lou, Z. (2020) Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell*
18. Borowski, P., Schalinski, S., and Schmitz, H. (2002) Nucleotide triphosphatase/helicase of hepatitis C virus as a target for antiviral therapy. *Antiviral Res* **55**, 397-412
19. Kleymann, G., Fischer, R., Betz, U. A., Hendrix, M., Bender, W., Schneider, U., Handke, G., Eckenberg, P., Hewlett, G., Pevzner, V., Baumeister, J., Weber, O., Henninger, K., Keldenich, J., Jensen, A., Kolb, J., Bach, U., Popp, A., Maben, J., Frappa, I., Haebich, D., Lockhoff, O., and Rubsamen-Waigmann, H. (2002) New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. *Nat Med* **8**, 392-398
20. Seybert, A., Posthuma, C. C., van Dinten, L. C., Snijder, E. J., Gorbalenya, A. E., and Ziebuhr, J. (2005) A complex zinc finger controls the enzymatic activities of nidovirus helicases. *J Virol* **79**, 696-704
21. Habtemariam, S., Nabavi, S. F., Banach, M., Berindan-Neagoe, I., Sarkar, K., Sil, P. C., and Nabavi, S. M. (2020) Should We Try SARS-CoV-2 Helicase Inhibitors for COVID-19 Therapy? *Arch Med Res* **51**, 733-735

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

22. Cavasotto, C. N., Lamas, M. S., and Maggini, J. (2020) Functional and druggability analysis of the SARS-CoV-2 proteome. *Eur J Pharmacol*, 173705
23. Tanner, J. A., Zheng, B. J., Zhou, J., Watt, R. M., Jiang, J. Q., Wong, K. L., Lin, Y. P., Lu, L. Y., He, M. L., Kung, H. F., Kesel, A. J., and Huang, J. D. (2005) The adamantane-derived bananins are potent inhibitors of the helicase activities and replication of SARS coronavirus. *Chem Biol* **12**, 303-311
24. Yu, M. S., Lee, J., Lee, J. M., Kim, Y., Chin, Y. W., Jee, J. G., Keum, Y. S., and Jeong, Y. J. (2012) Identification of myricetin and scutellarein as novel chemical inhibitors of the SARS coronavirus helicase, nsP13. *Bioorg Med Chem Lett* **22**, 4049-4054
25. Adedeji, A. O., Singh, K., Calcaterra, N. E., DeDiego, M. L., Enjuanes, L., Weiss, S., and Sarafianos, S. G. (2012) Severe acute respiratory syndrome coronavirus replication inhibitor that interferes with the nucleic acid unwinding of the viral helicase. *Antimicrob Agents Chemother* **56**, 4718-4728
26. Coles, M., Hulko, M., Djuranovic, S., Truffault, V., Koretke, K., Martin, J., and Lupas, A. N. (2006) Common evolutionary origin of swapped-hairpin and double-psi beta barrels. *Structure* **14**, 1489-1498
27. Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell* **41**, 693-703
28. Saikrishnan, K., Powell, B., Cook, N. J., Webb, M. R., and Wigley, D. B. (2009) Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell* **137**, 849-859
29. Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., Marsden, B. D., Kelm, S., Shi, J., Deane, C. M., and von Delft, F. (2017) A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat Commun* **8**, 15123
30. Halgren, T. A. (2009) Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* **49**, 377-389
31. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132
32. Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K., and Evans, G. (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85-97
33. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367
34. Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., Dias, A., Marsden, B., and von Delft, F. (2017) The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallogr D Struct Biol* **73**, 267-278
35. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255
36. An, J., Totrov, M., and Abagyan, R. (2005) Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics* **4**, 752-761
37. Abagyan, R. A., and Batalov, S. (1997) Do aligned sequences share the same fold? *J Mol Biol* **273**, 355-368

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1 Data collection and refinement statistics

	Phosphate bound	**APO**	**AMP-PNP**
Space group	P1	P1	C 2
Cell dimensions, *a,b,c* (Å)	59.1, 70.1, 84.6	56.7, 70.1, 84.0	324.9, 59.5, 132.4
Angles α,β,γ (°)	102.5, 95.6, 112.8	104.4, 93.3, 112.2	90, 93.9,90
Wavelength (Å)	0.91	0.91	0.91
Resolution (Å)	62.1 – 1.94 (1.94 - 1.92)	80.2 – 2.20 (2.27 – 2.20)	81.0 – 3.04 (3.09 – 3.04)
R_{merge}	0.053 (0.904)	0.077 (0.459)	0.258 (1.390)
I/σI	11.1 (1.1)	6.1 (0.7)	6.1 (1.2)
CC1/2	0.998 (0.669)	0.946 (0.605)	0.978 (0.359)
Completeness (%)	97.3 (97.2)	91.8 (90.7)	97.7 (95.2)
Multiplicity	3.5 (3.2)	2.5 (2.5)	3.4 (3.2)
No. Unique reflections	86341 (4418)	53358 (4635)	48372 (2343)
Refinement statistics			
Resolution	57.6-1.94	80.2 – 2.20	81.0 – 3.04
R_{work}/R_{free} (%)	20.9/25.3	22.9/28.6	24.4/28.4
No. atoms			
Protein	8917	9042	18063
Solvent	456	221	3
Ligand/ion	26	6	137
Average B factors (Å^2)			
All atoms	53	46	62
Protein	54	46	62
Solvent	51	42	35
Ligand/ion	50	41	77
Wilson B	38	32	50
R.M.S. deviations			
Bond lengths (Å)	0.005	0.003	0.004
Bond angles (°)	0.75	0.57	0.85
Ramachandran plot			
Favoured (%)	96	94	95
Allowed (%)	4	6	5
PDB ID	6ZSL	7NIO	7NN0

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

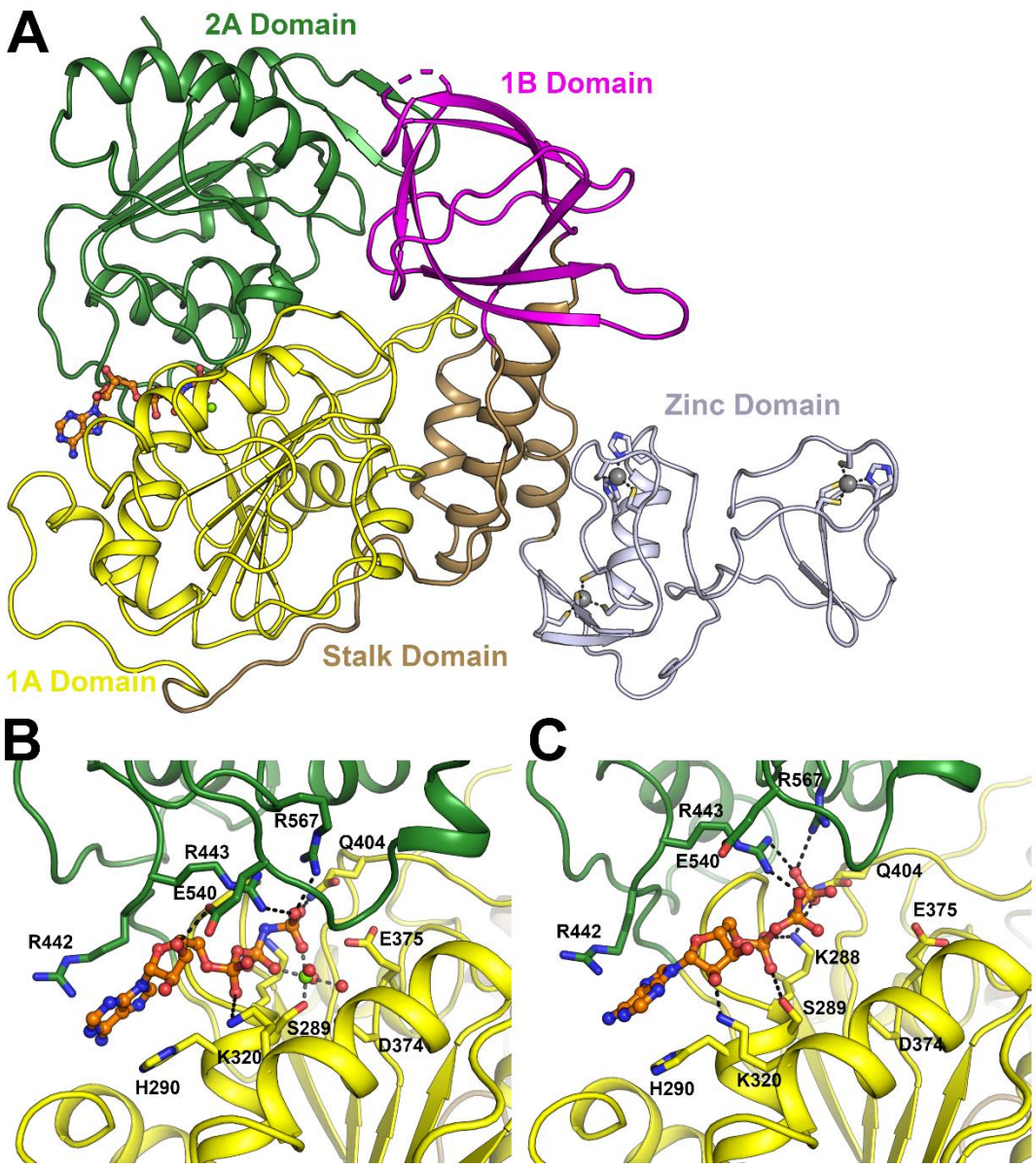


Figure 1 – Overall structure of SARS-CoV-2 NSP13. (A) Structure overview with domains labelled and colored individually (the same color scheme is used throughout). The AMP-PNP nucleotide is shown in stick format in the nucleotide binding site between the 1A and 2A domains. (B) Close up view of the nucleotide binding mode for the AMP-PNP Mg^{2+} complex (mode A), with interacting residues labeled and shown in the stick format. (C) Close up view of the nucleotide binding form the AMP-PNP complex (mode B), viewed from the same orientation as panel B.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

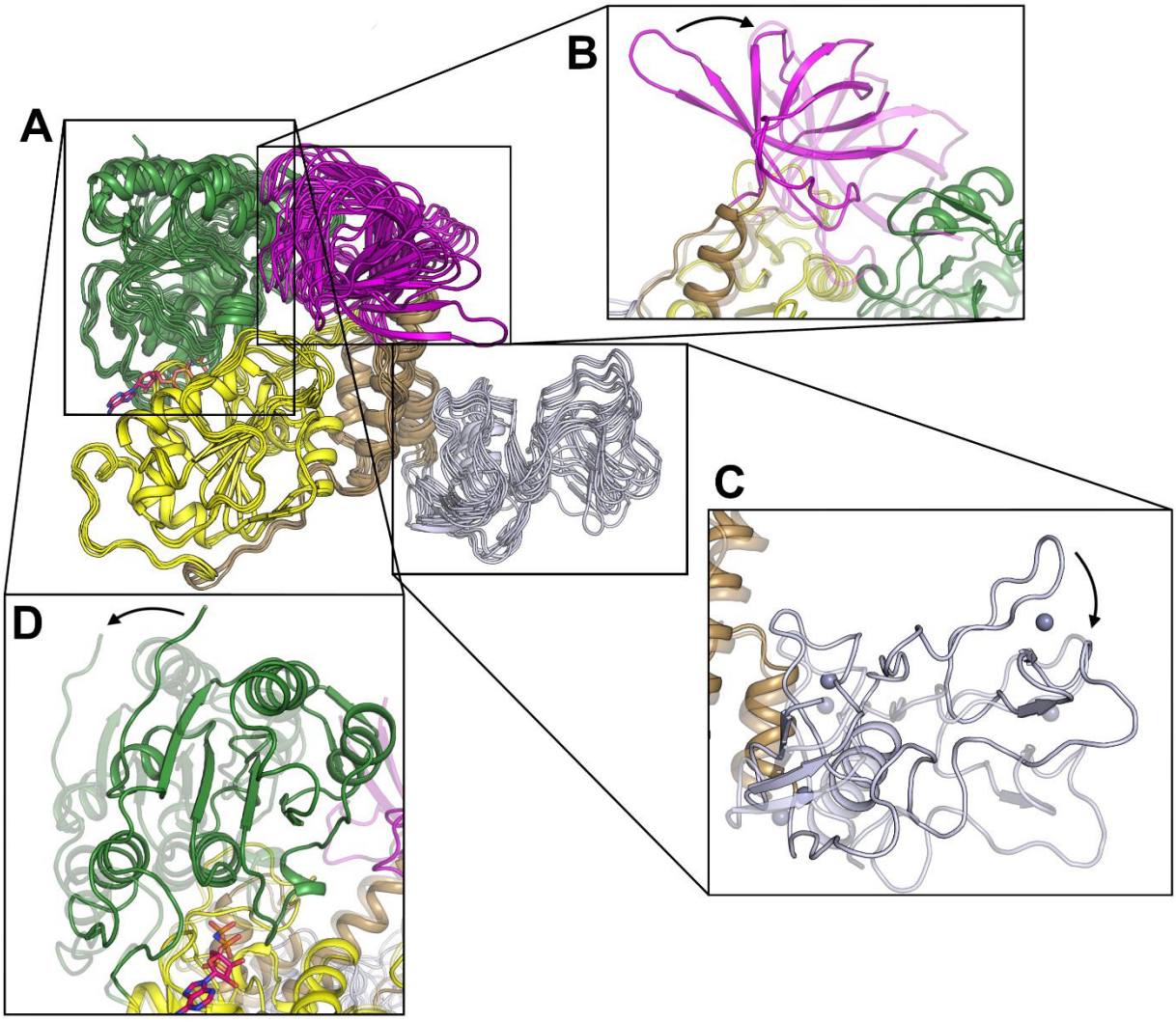
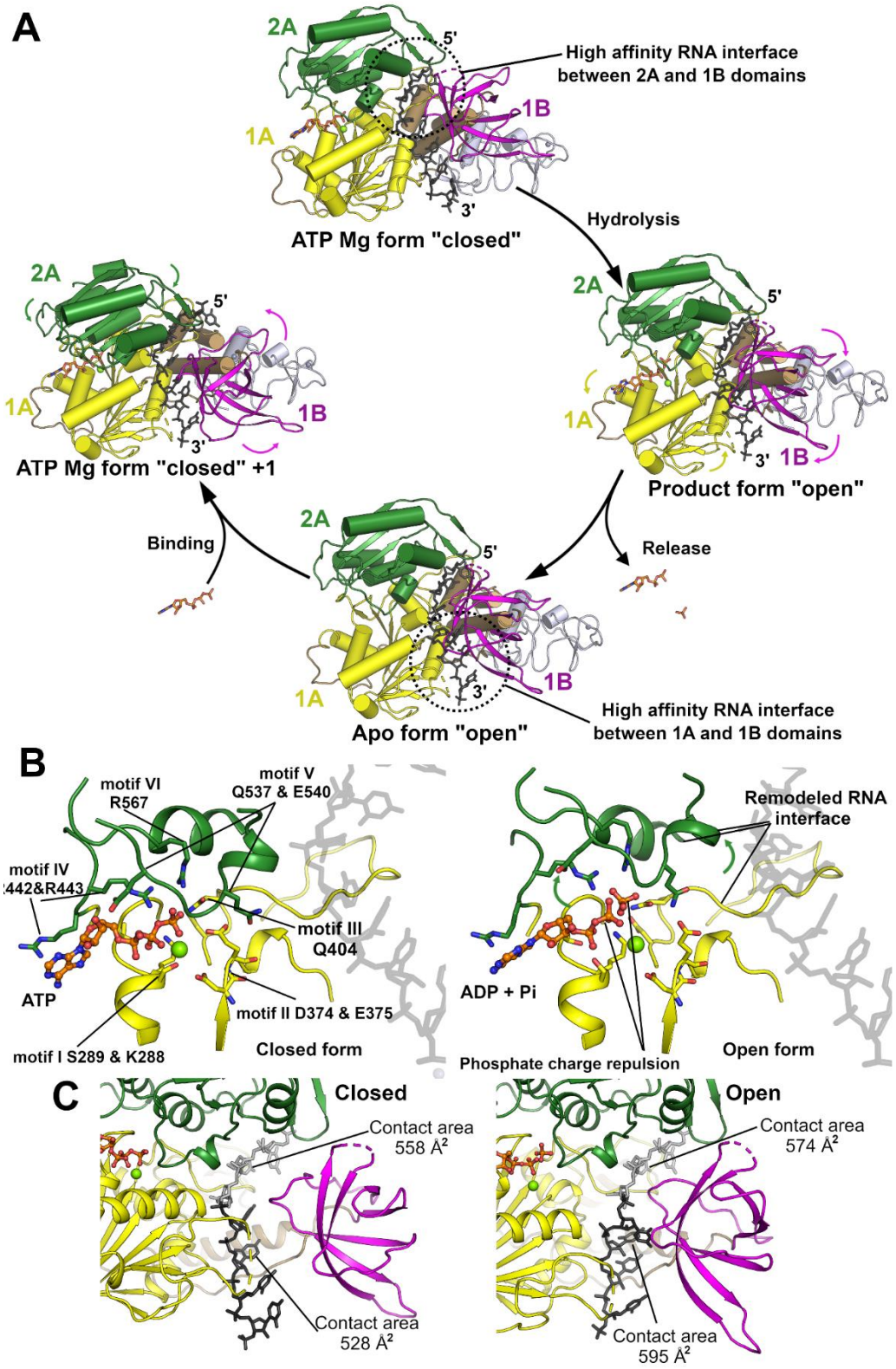


Figure 2 – Conformational flexibility exhibited in the NSP13 structures. (A) Overall view showing a structural superposition of the phosphate bound (2 chains), APO (2 chains) and AMP-PNP bound (4 chains) NSP13 structures. The 1A domain was used as the reference for the structural superposition. (B) Close up view of the variable conformation of the 1B domain, the two most different conformations are shown. (C) Close up view of the variable conformation of the Zinc domain. (D) Close up view of the variable conformation of the 2A domain).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3 – Model for the NSP13 5’ to 3’ translocation mechanism. (A) Proposed translocation mechanism for NSP13 based on the transition from the closed (pre hydrolysis) to open (Product and APO) forms. The transitions are initiated by the binding, hydrolysis and release of ATP which triggers the conformational changes and remodels the RNA interface. (B) Close up view of the active site with ATP in the closed conformation (left) and ADP and Pi in the open conformation (right). Hydrolysis as subsequent charge repulsion could trigger the opening of the cleft between the two domains with conserved motifs on the 2A domain primarily contacting the product phosphate whilst the ADP product interacts with the 1A domain. Several of the phosphate interacting motifs are proximal to regions of the RNA binding interface indicating the possibility of modulation based on hydrolysis status. (C) Remodeling of the RNA interface based on the position adopted by the 1B domain. The closed conformation is show on the left and the open on the right. The contact areas for the 5’ and 3’ RNA regions (depicted in grey and black) is shown.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

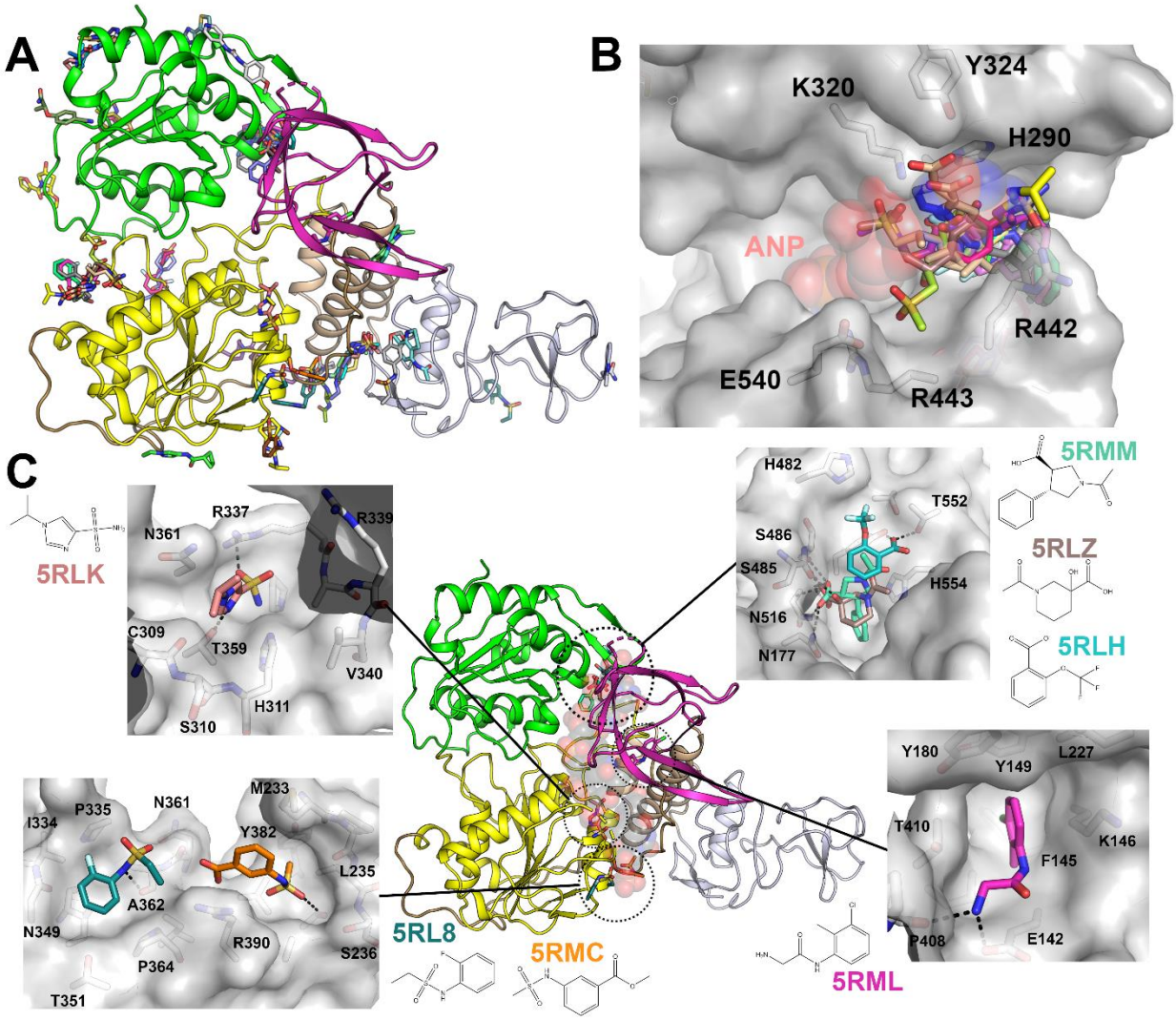


Figure 4 – Fragments bound to NSP13 in the nucleotide and RNA binding interfaces. (A) Overall view of all fragments bound to NSP13. (B) Close up view of the 15 fragments identified in the nucleotide binding site. The AMP-PNP moiety (ANP) is shown in a semi-transparent sphere representation for reference. (C) Overview of the fragments bound to the NSP13 RNA interface. The main panel shows the positioning of the clusters with the RNA shown in a semi-transparent sphere representation for reference. Each binding interface is shown in the inset in a detailed surface view with polar contacts shown in black.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

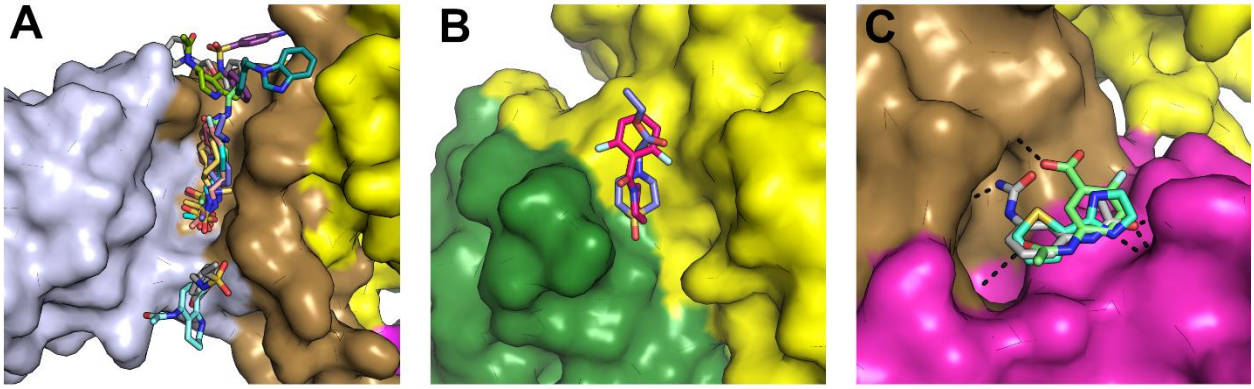


Figure 5 – Fragments bound on the interfaces between domains that appear to move as part of the catalytic cycle and thus may be starting points for allosteric inhibitor design. (A) Surface view of a prominent fragment binding site which bound to 11 fragments in a cleft between the Zinc and Stalk domains with the domains colored individually as for the color scheme in Fig1. (B) Two fragments were observed to bind in a cleft between the 1A and 2A domains approximately opposite the hinge. (C) Three fragments were bound in a shallow pocket between the Stalk and 1B domain (also close to the RNA interface) and make polar contacts to both regions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

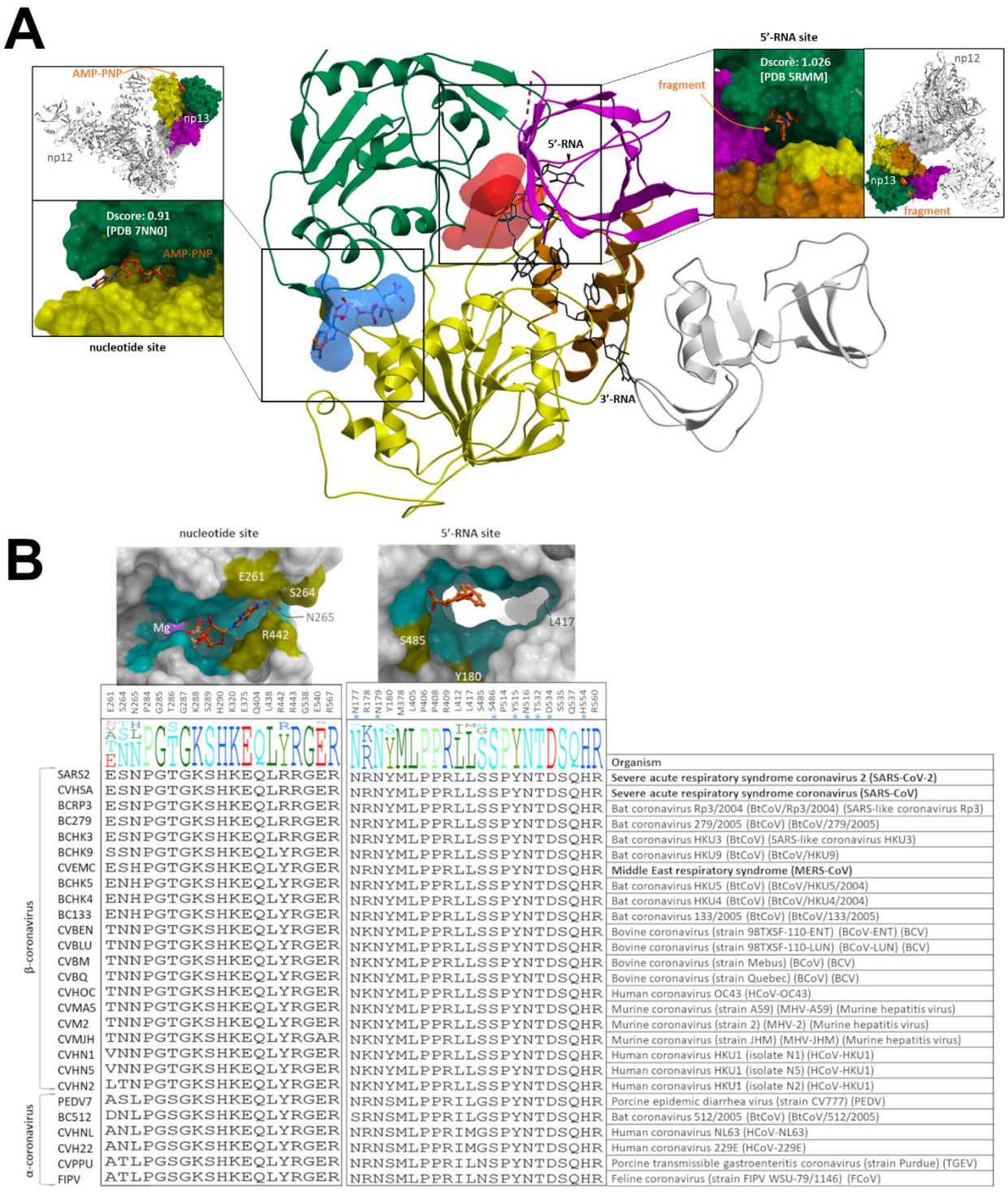


Figure 6. Druggability and genetic variability of the nucleotide and 5'-RNA sites. (A) The nucleotide binding pocket (blue) occupied by AMP-PNP and a cavity (red) that is partially occupied by the 5'-end of the substrate RNA in the transcription complex structure (PDB 7CXM) were identified as druggable sites in our crystal structures. Druggability scores (Dscores) calculated with SiteMap (30) are indicated (insets). The location of each binding site in the context of the transcription complex structure is also

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.15.435326; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

shown (insets). (B) Conservation of sidechains lining the AMP-PNP-bound site (left) and the fragment-bound 5'-RNA site (right) is shown across 27 α- and β-coronavirus sequences reviewed in the Uniprot database. Conserved (cyan) and non-conserved (mustard) sidechains are mapped on 3D structures. Asterisks indicate sidechains that are in direct contact with the fragment in structure 5RMM. Residue numbering in SARS-CoV-2 is shown at the top. Viruses associated with human epidemics are highlighted in bold.