

Crystal structures and fragment screening of SARS-CoV-2 NSP14 reveal details of exoribonuclease activation and mRNA capping and provide starting points for antiviral drug development

Nergis Imprachim¹, Yuliana Yosaatmadja and Joseph A Newman¹*

Centre for Medicines Discovery, University of Oxford, South Parks Rd, Oxford OX1 3QU, UK

Received March 21, 2022; Revised November 23, 2022; Editorial Decision December 05, 2022; Accepted December 09, 2022

ABSTRACT

NSP14 is a dual function enzyme containing an N-terminal exonuclease domain (ExoN) and C-terminal Guanine-N7-methyltransferase (N7-MTase) domain. Both activities are essential for the viral life cycle and may be targeted for anti-viral therapeutics. NSP14 forms a complex with NSP10, and this interaction enhances the nuclease but not the methyltransferase activity. We have determined the structure of SARS-CoV-2 NSP14 in the absence of NSP10 to 1.7 Å resolution. Comparisons with NSP14/NSP10 complexes reveal significant conformational changes that occur within the NSP14 ExoN domain upon binding of NSP10, including helix to coil transitions that facilitate the formation of the ExoN active site and provide an explanation of the stimulation of nuclease activity by NSP10. We have determined the structure of NSP14 in complex with cap analogue ⁷MeGpppG, and observe conformational changes within a SAM/SAH interacting loop that plays a key role in viral mRNA capping offering new insights into MTase activity. We perform an X-ray fragment screen on NSP14, revealing 72 hits bound to sites of inhibition in the ExoN and MTase domains. These fragments serve as excellent starting point tools for structure guided development of NSP14 inhibitors that may be used to treat COVID-19 and potentially other future viral threats.

INTRODUCTION

The global COVID-19 pandemic is caused by the SARS-CoV-2 virus and has infected over 200 million and caused over 4 million fatalities worldwide at the point of writing. Despite the successes with vaccines, there are currently a lack of effective drugs to treat people infected with SARS-CoV-2, and identification of such agents is a global priority.

SARS-CoV-2 has a relatively large (~30 kb) positive-sense single stranded RNA genome that consists of 12 functional open reading frames (ORF), encoding four structural proteins, six accessory proteins, and 16 non-structural proteins NSP1-16 (1,2). The non-structural proteins (NSPs) consist of the machinery for viral replication including RNA dependent RNA polymerase, primase, proteases, helicase, nucleases and various enzymes involved in RNA capping.

The SARS-CoV-2 NSP14 is a bifunctional protein containing an N-terminal DEDDh-type exoribonuclease (ExoN) domain and a C-terminal SAM-dependent guanine-N7 methyltransferase (N7-MTase) domain (3). The ExoN domain is thought to play a role in RNA proof-reading and maintaining the integrity of the viral genome by removal of mismatched nucleotides in a metal ion dependent manner (4,5). This exonuclease activity is greatly enhanced by the interaction of NSP14 with its activator protein NSP10. NSP10 is a small 139 amino acid with a mixed α/β fold that does not resemble any known prokaryotic or eukaryotic homologues and coordinates two structural zinc ions via C3H1 and C4 zinc binding motifs (6,7). Structures of SARS CoV NSP14/NSP10 complexes have been solved by crystallography (5), and more recently the SARS-CoV-2 NSP14-ExoN domain in complex with NSP10 (8), as well as the NSP14/NSP10 in complex with a mismatch containing RNA by cryo-EM (9). These structures show an extensive predominantly hydrophobic interface (in excess of 2000 Å²) with the N-terminal region of NSP14 packing against two faces of NSP10. The metal active site of NSP14 lies close to the NSP10 interacting regions although a structure of NSP14 in the absence of NSP10 is not known and thus the molecular basis for the nuclease activity stimulation is not fully understood.

The NSP14 N7-MTase domain is located in the C-terminus of NSP14 (residues 288–527) and contributes an important methylation step in the formation of the viral 5' RNA cap, which is essential for hijacking host translational machinery and suppressing innate immune

*To whom correspondence should be addressed. Tel: +44 1865 617752; Email: Joseph.Newman@cmd.ox.ac.uk
Present address: Yuliana Yosaatmadja, School of Biological Sciences, University of Auckland, 3a Symonds Street, Auckland 1010, New Zealand.

responses. The viral 5' RNA cap is formed by the sequential action of multiple enzymes that collectively assemble into a multi protein Replication and Transcription Complex (RTC). The helicase NSP13 is the RNA 5' triphosphatase that cleaves the terminal triphosphate to a diphosphate (10). The SARS-CoV-2 RNA dependent RNA polymerase NSP12 NiRAN domain was recently identified as the guanylyltransferase that enzyme that transfers GMP from GTP to the diphosphate to form a GpppN structure (11). NSP14 then methylates the guanine N7 of the GpppN structure in a S-adenosyl methionine (SAM) dependent manner to form cap-0 structure (⁷MeGpppN) (12). The product of this reaction is further methylated at the ribose O2' position by NSP16 to form the final cap-1 structure (⁷MeGpppN_{2'}OMe) (13), which provides viral RNA stability and host immune evasion. Studies suggest that this 2'-O-MTase activity of NSP16 is stimulated by NSP10 (14). NSP16 forms a complex with NSP10 via distinct yet overlapping interface. The apparent stoichiometric discrepancy is relieved by the fact that NSP10 is present on both open reading frames ORF1a and ORF1ab unlike NSP14 and NSP16. Recent cryo-EM structures have shown how the dual enzymatic activities of NSP14 are coupled to the rest of the RTC machinery via a physical protein interaction between the NSP14 ExoN domain and NSP9 and NSP12 NiRAN domain within the RTC (14). This interaction suggests the existence of a co-transcriptional capping complex (NSP12/NSP9/NSP14/NSP10) and the authors suggest that the ExoN domain may function in an *in trans* backtracking mechanism for proofreading.

In this study, we have determined the 1.7 Å crystal structure of SARS-CoV-2 NSP14 alone in the absence of its partner NSP10. Unexpectedly significant regions of the interaction interface retain some degree of folded structure. The predominantly hydrophobic interface, however, forms a collapsed self-association with significant rearrangement of the loops and residues that form the ExoN active site. The MTase domain also shows conformational differences when compared with previously solved NSP14/NSP10 structures. The MTase SAM binding loop occupies an alternative conformation that may be optimal for binding the ⁷MeGpppN product. NSP14 is one of the most conserved targets in the Coronavirus genome and thus is an excellent target for broad spectrum antivirals. We have used our crystals to perform a crystallographic fragment screen of NSP14 and find multiple fragments bound to sites of inhibition including the MTase and ExoN active sites and sites that may block the potential of NSP14 to interact with its partner NSP10. These fragments are important starting points for structure guided development or optimization of NSP14 inhibitors that may be useful as antiviral therapeutics with the potential for broad spectrum of action.

MATERIALS AND METHODS

Cloning and expression of NSP14

The plasmid for NSP14 with His₆ and Z-Basic tags at the N-terminal was synthesized in a pNIC-ZB vector with codon optimization for expression in *Escherichia coli* (Supplementary Table S1). The plasmid was transformed into

BL21(DE3)-RR-pRARE competent *E. coli*. Cell cultures were grown in Terrific Broth media (Merck) supplemented with 20 μM zinc chloride at 37 °C, shaking at 180 rpm. Once OD₆₀₀ reached 2.5, IPTG was added to the media at a final concentration of 300 μM. Cultures were incubated overnight at 18 °C, shaking at 180 rpm. Cells were harvested by centrifugation in a JLA-8.1000 rotor at 4400 × g for 25 min at 4 °C.

Purification of NSP14 protein

Cells were re-suspended in lysis buffer (50 mM HEPES pH:7.5, 500 mM NaCl, 10 mM imidazole, 5% glycerol, 1 mM TCEP) supplemented with protease inhibitor (Merck Protease inhibitor cocktail III, 1:1000) and 1 × benzonase nuclease. Cells were sonicated for 20 min (5 s pulse-on, 10 s pulse-off). Cell debris was removed by centrifugation with Beckman JA-17 rotor at 39 000 × g at 4 °C for 45 min. The supernatant was incubated with 7 ml pre-equilibrated Ni-IDA resin in tubes at 4 °C for 2 h. After batch binding, the tubes were spun at 700 × g at 4 °C for 5 min and the supernatant was discarded. Resin was washed with 30 ml lysis buffer twice and transferred into a gravity column with 25 ml W20 buffer (50 mM HEPES pH:7.5, 500 mM NaCl, 20 mM imidazole, 5% glycerol, 1 mM TCEP). Resin was washed again with 25 ml W30 buffer (50 mM HEPES pH:7.5, 500 mM NaCl, 30 mM imidazole, 5% glycerol, 1 mM TCEP). Protein was eluted with 25 ml IMAC elution buffer (50 mM HEPES pH:7.5, 500 mM NaCl, 300 mM imidazole, 5% glycerol, 1 mM TCEP) twice. Elution fractions were pooled and applied onto a pre-equilibrated 5 ml-HiTrap-SP High Performance column using a syringe. Column was washed with 50 ml SP wash buffer (50 mM HEPES pH:7.5, 500 mM NaCl, 5% glycerol, 1 mM TCEP). Protein was eluted with 25 ml SP elution buffer (50 mM HEPES pH:7.5, 1 M NaCl, 5% glycerol, 1 mM TCEP). Elution fraction was diluted in 50 mM HEPES pH:7.5, 5% glycerol, 1 mM TCEP to reduce salt concentration to 500 mM before TEV cleavage. Diluted sample was treated with TEV protease (1:5 mass ratio) for 3.5 h at room temperature. To remove ZB tag, the cleavage sample was passed through a 5 ml-HiTrap-SP HP column and the flow-through was collected. The flow-through fraction was concentrated with 30 MWCO Amicon filters and loaded onto a HiLoad 16/600 Superdex 200 pg column pre-equilibrated with SEC buffer (50 mM HEPES pH:7.5, 500 mM NaCl, 5% Glycerol, 1 mM TCEP). Pure NSP14 fractions were collected concentrated and stored at -80 °C for future experiments.

NSP10 expression and purification

The plasmid for NSP10 with GST tag and 3C-protease cleavage site was synthesized in a pNIC-CTH0 vector with codon optimization for expression in *E. coli* (Supplementary Table S1). The plasmid was transformed into BL21(DE3)-RR-pRARE competent *E. coli*. Cell cultures were grown in Terrific Broth media (Merck) supplemented with 20 μM zinc chloride at 37 °C, shaking at 180 rpm. Once OD₆₀₀ reached 2–3, IPTG was added to the media at a final concentration of 300 μM. Cultures were incubated overnight at 18 °C, shaking at 180 rpm. Cells were harvested

by centrifugation at $4400 \times g$ at 4°C for 25 min. Cells were re-suspended in lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM TCEP) supplemented with protease inhibitor (Merck Protease inhibitor cocktail III, 1:1000) and $1 \times$ benzonase nuclease. Cells were sonicated for 20 min (5 s pulse-on, 10 s pulse-off). Cell debris was removed by centrifugation with Beckman JA-17 rotor at $39\,000 \times g$ at 4°C for 45 min. The supernatant was incubated with 10 ml pre-equilibrated GST resin in tubes at room temperature for 45 min. After batch binding, the tubes were spun at $700 \times g$ at 4°C for 10 min and the supernatant was discarded. Resin was washed with 25 ml lysis buffer thrice. Resin was transferred into a gravity column with 25 ml Elution buffer (pH 7.5) (50 mM HEPES pH:7.5, 500 mM NaCl, 5% glycerol, 1 mM TCEP, 20 mM reduced glutathione), incubated for 15 min at RT, and protein was eluted. Elution step was repeated twice. Elution fractions were pooled and treated with 3C-protease in 1:15 mass ratio, and dialysed in 50 mM HEPES pH: 7.5, 300 mM NaCl, 5% Glycerol, 1 mM TCEP overnight at 4°C . To remove GST tag and GST-tagged 3C-protease, the dialysed sample was passed through a gravity column with GST resin and the flow-through was collected. Resin was washed with 25 wash buffer (50 mM HEPES pH: 7.5, 300 mM NaCl, 5% glycerol, 1 mM TCEP) thrice. Fractions with cleaved protein were collected and concentrated with 10 MWCO Amicon filters. Sample was loaded onto a HiLoad 16/600 Superdex 75 pg column pre-equilibrated with SEC buffer (20 mM HEPES pH:7.5, 150 mM NaCl, 5% glycerol, 1 mM TCEP). NSP10 fractions were collected concentrated and stored at -80°C .

NSP10/14 complex expression and purification

NSP10/14 bicistronic construct was expressed and purified as described (Baddock *et al.*, 2022). Vectors and sequences are available at: <https://www.addgene.org/159613/> and (Supplementary Table S1).

Limited proteolysis of NSP10, NSP14 and NSP10/14 complex

Proteolysis strips containing 30 mM HEPES, 75 mM NaCl, 2% glycerol, and with or without $22 \mu\text{g/ml}$ trypsin (Sigma T6567-20UG) were prepared at a volume of 11 μL . NSP10 and NSP14 stock proteins were diluted to 1 mg/ml. NSP10/NSP14 stock was already prepared at 1 mg/ml during purification. 5 μL of NSP10, NSP14 or NSP10/NSP14 (at 1 mg/ml) were separately added to the individual tryptic digest strips in the presence and absence of trypsin. Samples were incubated for 1 h and overnight on ice. After incubation, 6 μL of samples were mixed with 45 μL 0.1% formic acid and subjected to intact mass analysis on electrospray ionization quadrupole time of flight (ESI-QTOF MS), and the rest of the samples were used for SDS-PAGE analysis.

Electrospray ionization quadrupole time of flight (ESI-QTOF) mass spectrometry

Assays were performed on Agilent 6530 RapidFire QTOF mass spectrometer in 384-well polypropylene plates (Greiner, code 781280) with ~ 0.1 mg/ml protein samples

in 0.1% formic acid. Samples were aspirated under vacuum and adsorbed onto a C4 solid-phase extraction (SPE) cartridge (Agilent Technologies), equilibrated and washed with LC-MS grade water containing 0.1% formic acid to remove buffer components. Following the aqueous wash, analytes were eluted from the C4 SPE onto an Agilent 6530 accurate mass Q-TOF in an organic elution step containing 85% acetonitrile and 0.1% formic acid. Masses of fragments were determined using the Agilent Masshunter Qualitative Analysis software (Agilent).

Crystallization

Crystallization was performed by sitting-drop vapor-diffusion method with SGC-HIN3 HT-96 screen (Molecular Dimensions). Crystal plates were set up with 10 mg/ml protein and HIN3 screen, using 150 nl drops (1:1 ratio). Rod shaped crystals were grown in condition containing 1.26 M sodium phosphate monobasic, 0.14 M potassium phosphate dibasic at 4°C . Further crystals were grown in this single condition from Molecular Dimensions. For NSP14- $^{7\text{Me}}\text{GpppG}$ crystals, $\text{m}^7\text{GP}_3\text{G}$ monomethylated cap analogue (Jena Bioscience, NU-852) was dissolved in 10 mM HEPES pH 7.5 at 50 mM to make a $10\times$ stock solution which was diluted with crystallization buffer (final 5 mM concentration). Mature NSP14 crystals (grown in 1.26 M sodium phosphate monobasic, 0.14 M potassium phosphate dibasic) were loop mounted and transferred to a drop containing cap analogue in 96-well plates and incubated overnight at 4°C . Crystals were cryo-protected in a solution consisting of well solution supplemented with 20% glycerol, loop-mounted, and flash-frozen in liquid nitrogen.

Structure determination

All data were collected at Diamond light source beamlines I03 and I04-1 processed using XDS (15). Data were moderately anisotropic and were corrected using an anisotropic cut-off as implemented in the STARANISO server, which gave a significant improvement in the quality and interpretability of electron density maps. The structures were solved by molecular replacement using the program PHASER and the structure of SARS-CoV-1 NSP14 (5C8S) as a search model. Refinement was performed using PHENIX REFINE (16). A summary of the data collection and refinement statistics are shown in Table 1.

X-ray fragment screening

A total of 634 fragments from the DSI poised fragment library (500 mM stock concentration dissolved in DMSO) were transferred directly to NSP14 crystallization drops using an ECHO liquid handler (50 mM nominal final concentration), and soaked for 1–3 h before being loop mounted and flash cooled in liquid nitrogen. A total of 588 datasets were collected at Diamond light source beamline I04-1 and processed using the automated XChem Explorer pipeline (17). Structures were solved by difference Fourier synthesis using the XChem Explorer pipeline. Fragment hits were identified using the PanDDA program (18). Refinement was performed using BUSTER. A summary of data collection

Table 1. Data collection and refinement statistics

	NSP14	NSP14-7MeGpppG
Space group	<i>P</i> 212121	<i>P</i> 212121
Cell dimensions, <i>a, b, c</i> (Å)	69.2, 68.1, 138.5	67.3, 67.3, 138.6
Wavelength (Å)	0.9179	0.9762
Resolution (Å)	69.2–1.66 (1.85–1.66)	69.3–2.53 (2.73–2.53)
<i>R</i> _{merge}	0.125 (1.075)	0.585 (1.990)
<i>I</i> / σ <i>I</i>	9.2 (1.6)	6.0 (1.6)
CC1/2	0.997 (0.716)	0.986 (0.747)
Completeness (spherical) (%)	49.0 (8.7)	55.5 (13.6)
Completeness (ellipsoidal) (%)	89.27 (65.6)	91.7 (65.9)
Multiplicity	5.5 (6.2)	26.5 (27.6)
No. unique reflections	37 654 (1884)	12 041 (602)
Refinement statistics		
Resolution	69.2–1.66	69.3–2.53
<i>R</i> _{work} / <i>R</i> _{free} (%)	19.6/22.1	20.7/24.6
No. atoms	4105	3770
Protein	3661	3660
Solvent	431	45
Ligand/ion	5	65
Average <i>B</i> factors (Å ²)		
All atoms	26	23
Protein	26	24
Solvent	34	19
Ligand/ion	44	60
Wilson <i>B</i>	22	25
R.M.S. deviations		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.508	0.466
Ramachandran plot		
Favoured (%)	96.9	94.9
Allowed (%)	3.1	5.1
PDB ID	7QGI	7QIF

and refinement statistics for all fragment bound datasets is shown in Supplementary Table S3 and an overview of electron density maps for each hit is shown in Supplementary Table S2.

RESULTS

Crystal structure of NSP14 at 1.7 Å resolution

Initial attempts to crystallize the SARS-CoV-2 NSP14/NSP10 complex, either expressed in *E. coli* as a bicistronic construct, or reconstituted from components *in vitro*, failed to produce crystals of diffraction quality. We explored N-terminal truncations of NSP14 as a way to increase stability of NSP14. We found a construct omitting the first 7 amino acids (residues 8–527) gave significantly improved purification yields. Crystallization of this NSP14 construct alone in a condition containing 1.26 M sodium phosphate monobasic and 0.14 M potassium phosphate dibasic produced rod shaped crystals that diffracted to 1.7 Å with moderately anisotropic diffraction. We solved the structure by molecular replacement using NSP14 from SARS-CoV NSP14/NSP10 complex (5) as the starting model (PDB entry 5C8U). Overall, the electron density map is of high quality (Supplementary Figure S1) and the model includes 3 Zinc ions and 2 Phosphate ions. The NSP14 model is complete with the exception of 17 residues in the N-terminus, 3 in the C-terminus and 3 loops (amino acids 95–102, 122–151 and 456–463), which are not visible

in the electron density presumably due to disorder. The model is of high quality with good stereochemistry and has been refined to a final *R* = 0.197 *R*_{free} = 0.220. A summary of the data collection and refinement statistics are shown on Table 1.

Structural comparison of NSP14 alone with NSP14/NSP10 complex

Comparison of our NSP14 structure with previously solved NSP14/NSP10 structures from SARS-CoV-2 or coronavirus species reveals a generally good agreement for the overall structure (overall RMSD values in the range of 1.5 Å) with distinct conformational differences observed for both the ExoN and MTase domains (Figure 1A). The most prominent difference occurs in the fold of the ExoN domain with the regions at the NSP14 N-terminus that interact with NSP10. In the NSP14/NSP10 complex structures, this region forms a small 3 stranded antiparallel β -sheet (strands $\beta 1 \uparrow \beta 5 \downarrow \beta 6 \uparrow$) which makes extensive contacts to the first (C3H1) zinc ion binding site on NSP10 (Figure 1B, C). A long region of largely extended coil structure preceding $\beta 1$ that is interspersed with a small β -hairpin also makes extensive contacts to NSP10 in the region of the NSP10 β -subdomain. This region is also contacted by NSP16 in the NSP16/NSP10 heterodimeric complex (19). The overall interface covers around 2000 Å² of contact area includes 23 hydrogen bonds, 1 salt bridge and several clusters of buried hydrophobic residues. This binding interface has previously been described as similar to a NSP14 ‘hand’ over NSP10 ‘fist’ (20) (Figure 1D).

Somewhat surprisingly, given the intimate association of the N-terminal NSP14 region with NSP10, much this region is still ordered in the absence of NSP10 (103 residues observed out of a possible 153). However, the fold of this region is distinctly different and bears almost no relation to the complex structure (Figure 1B, C). In particular the first 60 residues forms a more compact arrangement with a higher helical content, and the first β -sheet that associates with the first NSP10 zinc binding site is not formed. Several sequences are observed to transition between secondary structure elements, with residues 36–41 transitioning from coil to helix, residues 54–58 transitioning from β -strand to helix, and the helix 146–152 (second α -helix of the NSP14/10 complex) now partially forms a β -strand and associates with the main β -sheet in NSP14 to form an additional 6th strand (mixed parallel/antiparallel) (Supplementary Figure S2). This more compact helical N-terminal region associates closely with the remainder of the ExoN domain in a collapsed arrangement such that much of the interface area in the NSP14/10 complex is not solvent exposed but now buried within the new interface between the N-terminal region of NSP14 and the rest of the ExoN domain (Figure 1B). A long stretch of 27 residues spanning amino acids 123–150 are disordered in our NSP14 structure and given the fact that the neighboring residues lie on different ends of the main central ExoN β -sheet (required to span a gap of around 25 Å) may be assumed to adopt a relatively extended coil like fold rather than forming the remaining part of the small N-terminal β -sheet as for the NSP14/10 complex structure. These dramatic conformational changes

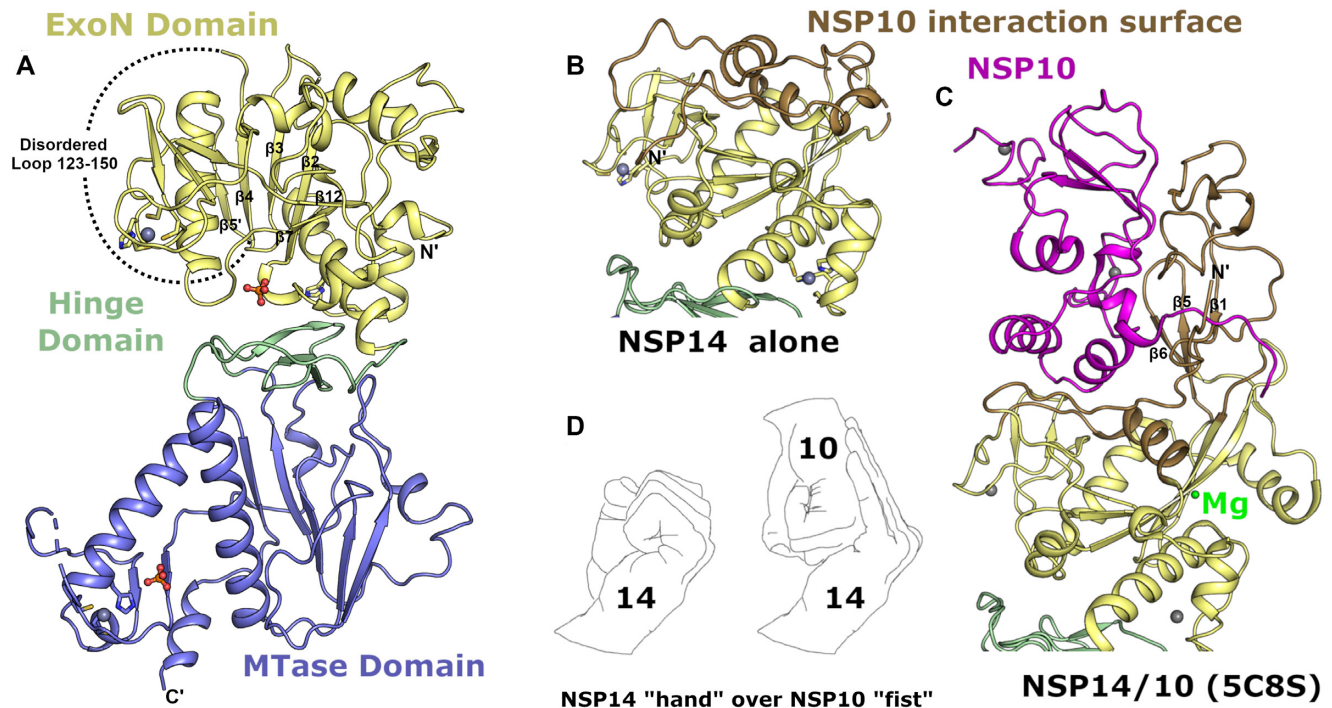


Figure 1. Structure of NSP14 in the absence of NSP10. (A) Overall NSP14 structure with the ExoN domain colored yellow, hinge domain in green and MTase domain slate, zinc ions are shown as blue-grey spheres and phosphate ions are shown in ball and stick format. (B) Close up view of the ExoN domain with regions that were observed to shift conformation highlighted in 'sand' color. (C) Structure of the ExoN domain of SARS-CoV NSP14 in complex with NSP10 (shown in pink) viewed from the same orientation as for panel B. (D) The interaction between NSP14 and NSP10 has been described as 'hand over fist', by the same logic the 'fingers' of NSP14 collapse back toward the core resembling a closed 'fist'.

and collapsing of the N-terminal NSP10 interacting region onto the ExoN core are consistent with predictions made from SAXS analysis of SARS-CoV NSP14 upon binding NSP10 (20). The potential energetic cost of the rearrangements may also explain why the NSP14 - NSP10 interaction appears to have only a moderate affinity (K_d value of around 1 μ M) (21) despite the extensive nature of the interface.

NSP10 complex formation is required to form the ExoN active site

The exonuclease active site is formed in cleft on top of the central ExoN domain β -sheet close to the junction with the N-terminal NSP10 interacting region (Figure 2A & B). Four conserved acidic residues D90, E92, E191 and D273 form the metal binding active site core, whilst the fifth catalytic residue H268 has been proposed in related proteins to function as a general base that deprotonates a catalytic water for nucleophilic attack (22) (Figure 2C). The metal ion binding site has been suggested to bind to two metal ions based on similarities with other enzymes from the family, including exonuclease domains of bacterial polymerases (23). Current structures in the absence of nucleic acid substrates generally show a single metal ion (typically Mg^{2+}) bound in an octahedral environment coordinated with D90, E92, D273 and three water molecules (8,24). Recent cryo-EM structures of NSP14/NSP10 in complex with RNA (9) include a second metal ion binding site adjacent to the first coordinated by E191 and D90 and additional two oxy-

gen atoms on the RNA phosphate substrate, suggesting this metal ion may bind in conjunction with RNA (Figure 2C). This study also established the wider NSP14 RNA binding site which contains contributions from the extreme N-terminus of both NSP14 and NSP10, the $\beta 2$ - $\beta 3$ loop, the $\beta 6$ - $\alpha 2$ loop and the two loops preceding and following $\alpha 5$ (Figure 2A). In the absence of NSP10 binding, no metal ion was observed to bind. Indeed, the loop between $\beta 7$ and $\alpha 4$ is in a different conformation where H188 occupies part of the metal binding pocket and likely prevents the binding of the second metal (Figure 2D). The movement of this loop appears to be influenced by the burying of N-terminal hydrophobic residues F33 and L38 into a hydrophobic environment immediately adjacent to the active site (Figure 2D). Several other regions of the collapsed N-terminal region are in positions where they would form steric clashes with potential substrates (Figure 2B), suggesting that a different mode of RNA binding may be necessary. In the NSP14/NSP10 RNA complex structures, the duplex RNA contains a 5' overhang with a single separated base pair that has been flipped out by the actions of N104 and H95 (9). Both of these residues are no longer part of the extended active site of NSP14 in the absence of NSP10, and the active site is generally more open, raising the possibility of accommodating fully duplex substrates. The requirement for base separation for access to the NSP14/NSP10 active site has been suggested to confer a preference for mismatched nucleotides (4) although there is some debate about the existence of such a preference (25). An intriguing possibility is that NSP14 alone may differ from the

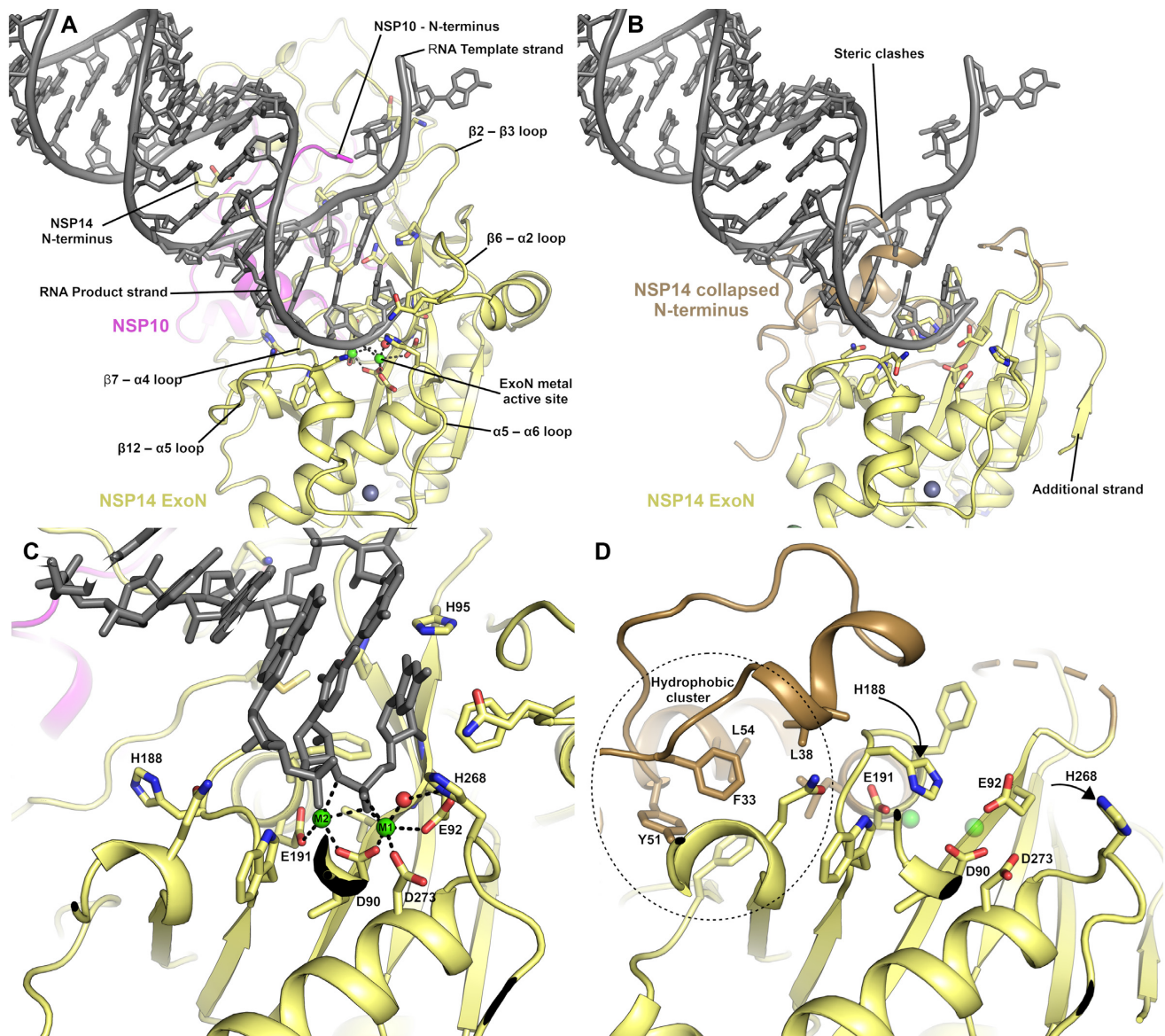


Figure 2. The NSP14 ExoN active site is formed upon binding to NSP10. (A) View of the ExoN domain of the NSP14 NSP10 complex bound to a mismatch containing double stranded RNA molecule as recently solved by cryo EM(9). Key residues in the catalytic center are shown in the stick format and the secondary structure elements that form the wider active site are labelled. (B) View of the NSP14 ExoN domain in the absence of NSP10, the mismatch containing RNA is shown for reference and forms significant steric clashes with residues that constitute the collapsed N-terminus (shown in sand color). (C) Zoomed in view of the active site catalytic center of the NSP14 NSP10 RNA complex with metal ion coordinating residues labelled. (D) NSP14 in the absence of NSP10 viewed from the same orientation. A hydrophobic cluster adjacent to the active site (shown in Sand color) flips H188 into a position where it likely disrupts binding of M2 (shown for reference as semitransparent green spheres).

NSP14/10 complex in this regard, although to date the substrate specificities of NSP14 in the absence of NSP10 remain to be fully characterized. During the late stages of preparation of this manuscript a similar NSP14 structure was released in the PDB (7R2V) from the lab of Krzysztof Pyrc (26) that shows the same dramatic rearrangements of NSP14 upon binding NSP10 that we have seen in our own structures. The two structures are from different crystal forms with different crystal contacts and validates the changes we see as being part of a Nuclease activation mechanism rather than being the result artefacts induced by crystal packing interactions.

Probing the NSP14 ExoN structural rearrangement with limited proteolysis

To provide further experimental validation that our crystal structure of NSP14 in the absence of NSP10 represents a relevant state in solution we have probed the structure using limited Trypsin proteolysis coupled to mass spectrometry. Our crystal structure contains three disordered loops that contain a lysine or arginine within the disorder and would thus be expected to be good substrates for trypsin cleavage. Importantly two of these loops (residues 96–101 and 123–150) are in the N-terminal region, and undergo disorder to order conformational changes upon interaction with

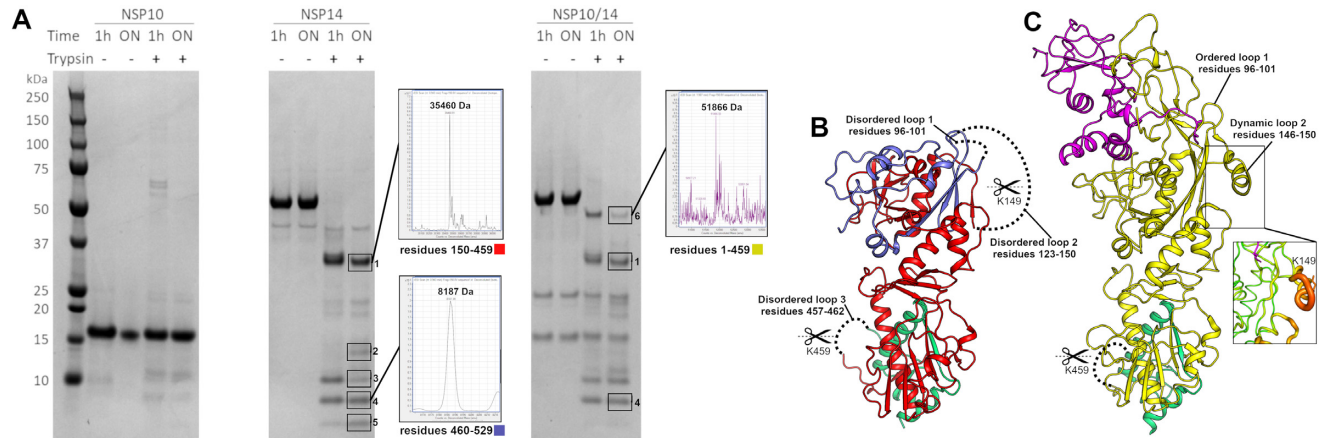


Figure 3. Analysis of NSP14 conformation in solution by limited proteolysis coupled to Mass spectrometry. (A) Coomassie stained SDS PAGE gels of NSP10, NSP14 and the NSP14/10 complex upon limited proteolysis with trypsin. All samples were ran on a single gel which has been divided into sections for clarity. Prominent gel bands are boxed and numbered with deconvoluted spectra shown as insets for fragments that can be matched to cleavage products. (B) Interpretation of the digestion pattern mapped to the NSP14 structure. Red and green regions correspond to matched fragments following cleavage at disordered loops 2 and 3. The remaining residues at the N-terminus (colored in Blue) is a good match by mass to band 2 or may be further processed by cleavage at loop 1 to produce fragments matching bands 3 and 5. (C) Interpretation of the digestion pattern mapped to the NSP14/10 complex structure (5C8S). The yellow fragment is a match for band 6 and contains an intact N-terminal region, consistent with the disorder to order transitions involved in loops 1 and 2. The relative flexibility remaining around K149 may explain why there is still some cleavage at this site.

NSP10. Thus the cleavage of these loops in solution is a sensitive measure of the conformational solution state. A third loop (residues 457–462) is in the MTase domain and does not change upon complex formation and serves as an internal control. As would be expected given its compact nature NSP10 is not cleaved significantly by trypsin during the time course of the experiment (Figure 3A). In contrast NSP14 is cleaved into a prominent band (band 1) at 35 kDa with several smaller species (bands 2–5) appearing between 5–15 kDa. We were able to observe mass spectrometry masses for two of these bands band 1 and 4 which correspond to expected masses for residues 150–459 and 460–529 and are colored in red and green in figure 2B respectively. The blue region in Figure 2B represents the N-terminal fragment following cleavage at residue 149 which would correspond to a mass of 15850 Da and is a good match for band 2 on the gel. Further processing of this fragment within disordered loop 1 would produce fragments with masses of 10750 and 5117 Da which again are good matches for bands 3 and 5 respectively. We have extended this analysis to the NSP14/NSP10 complex, using a bicistronic NSP14/10 construct to produce the complex as has previously been described (25). Under the same digestion conditions the complex displays a different digestion pattern with considerably less evidence for cleavage in the N-terminal region of the ExoN domain. A band at 51886 Da (band 6) corresponds by mass spectrometry residues 1–459 and is still present after overnight digestion (Figure 3A and C). Band 1 is also present (although less prominent) in the complex and may represent the fact that K149 is relatively mobile even in the complex structure (insert Figure 3C) or possible complex dissociation or incomplete saturation of NSP14 by NSP10 as has been described for others (27). Several smaller gel bands between 5 and 15 kDa are also present in the complex sample although it is not clear what these represent (Figure 3A) with the exception of band 4 which was also identified in

this sample by mass spectrometry. The absence of a band corresponding to band 5 in the NSP14 alone suggests that the complex is not cleaved at residue 96 (loop 1), consistent with the observation of an intact N-terminus in band 6. Taken together these results provide experimental support for our crystal structure of NSP14 in the absence of NSP10 and the conformational changes in the ExoN domain observed upon interaction with NSP10.

The SARS-CoV-2 MTase domain and structure with CAP analogue

The MTase domain of NSP14 is located in the C-terminal half and consists of a central five stranded predominantly parallel β -sheet that is capped at its C-terminal end by and additional small three stranded antiparallel β -sheet (inserted between the fourth and fifth strands of the main β -sheet) that has been described as a ‘hinge’ domain (20), as it separates the MTase and ExoN domains (Figure 1A). The MTase fold is distinct from traditional Rossmann type α - β - α folds that feature in other known SAM dependent RNA methyltransferases (28). The MTase active site is formed in a deep pocket between the β -sheets of the MTase and hinge domain with contributions from the long helix (α 7) that precedes the first strand of the MTase domain central β -sheet and the loops between first and second and third and fourth strands (Figure 4A). Structures of the NSP14/NSP10 complex from SARS-CoV in complex with *S*-adenosyl methionine (SAM), and with *S*-adenosyl homocysteine (SAH) and cap analogue GpppA(5) show both substrates binding close together within the same large active site with the SAM/SAH binding and making interactions with the extended loop between second and third strands, whilst the methionine/homocysteine moiety approaches the GpppA in a manner that appears compatible with transfer of the methyl group to the Guanine N7.

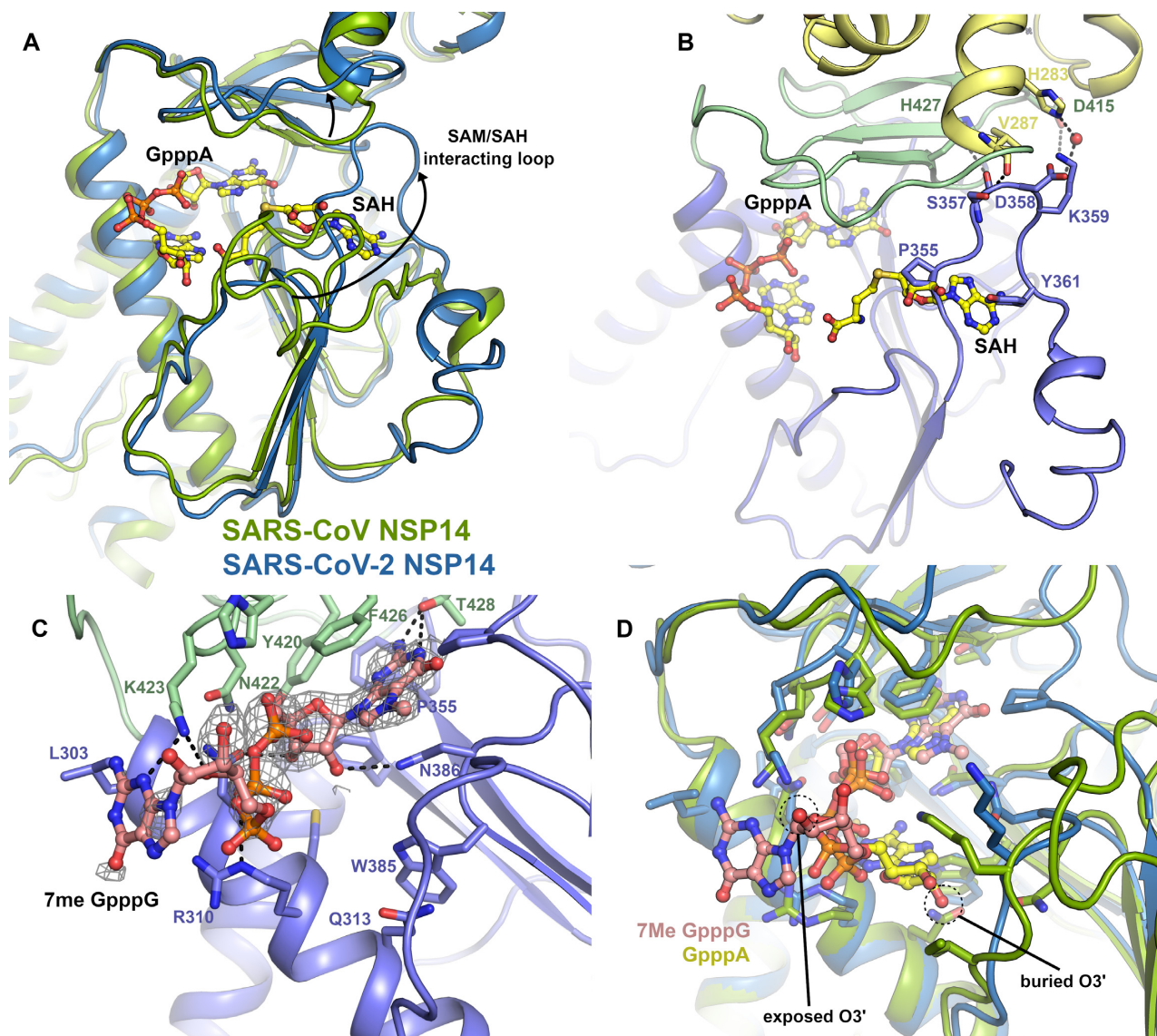


Figure 4. Structural differences in the NSP14 MTase active site. (A) Alignment of the SARS-CoV NSP14 GpppA/SAH ternary complex (shown in green) with the SARS-CoV-2 NSP14 MTase domain (shown in slate), a significant conformational change can be seen for an extended loop that makes interactions with SAH in the SARS-CoV NSP14 structure. Both ligands are shown in stick format for reference. (B) Close up view of the SAM/SAH interacting loop in SARS-CoV-2 NSP14, residues which overlap with SAH or form contacts to residues in the ExoN or hinge domains are shown in the stick format. (C) Structure of SARS-CoV-2 NSP14 in complex with ^7Me GpppG. The $2F_o - 1F_c$ electron density map is shown in grey in the vicinity of the ^7Me GpppG contoured at 1σ . (D) Comparison of the binding poses of the GppA from SARS-CoV NSP14 with the ^7Me GpppG from SARS-CoV-2 NSP14. The proteins are colored as for panel A and the positions of the O2' which is the attachment point for full RNA cap substrates is marked.

Comparing these structures to our NSP14 MTase domain structure reveals generally good agreement (overall RMSD 0.8 \AA) although large differences (displacements of up to 25 \AA) can be seen in the conformation of the extended SAM/SAH interacting loop (residues 356–378). In our NSP14 structure, this loop makes several polar interactions with residues in the ExoN domain (S357 to V287, D358 water mediated to H283) and hinge domain (S577 to H427 and K359 to D415). The loop continues on the opposite side of the central β -sheet, in comparison to the SARS-CoV structures (Figure 4A). As was the case for regions of the N-terminal NSP10 interacting regions, several regions are observed interchange secondary structure ele-

ments with residues 363–366 forming a short section of helix rather than forming an additional sixth strand on the central MTase β -sheet. This loop conformation would not appear to be compatible with binding of SAM or SAH in the same manner as previously reported (5) as the new loop conformation significantly overlaps with the adenine moiety. Consistent with this, attempts to soak either SAH or SAM to our crystals did not give any convincing electron density. On the other hand, the cap binding site appears to be open and we were able to soak in a cap analogue ^7Me GpppG which gave clear electron density despite a slightly lower overall resolution (Figure 4B and Supplementary Figure S3). The ^7Me GpppG has both methylated and

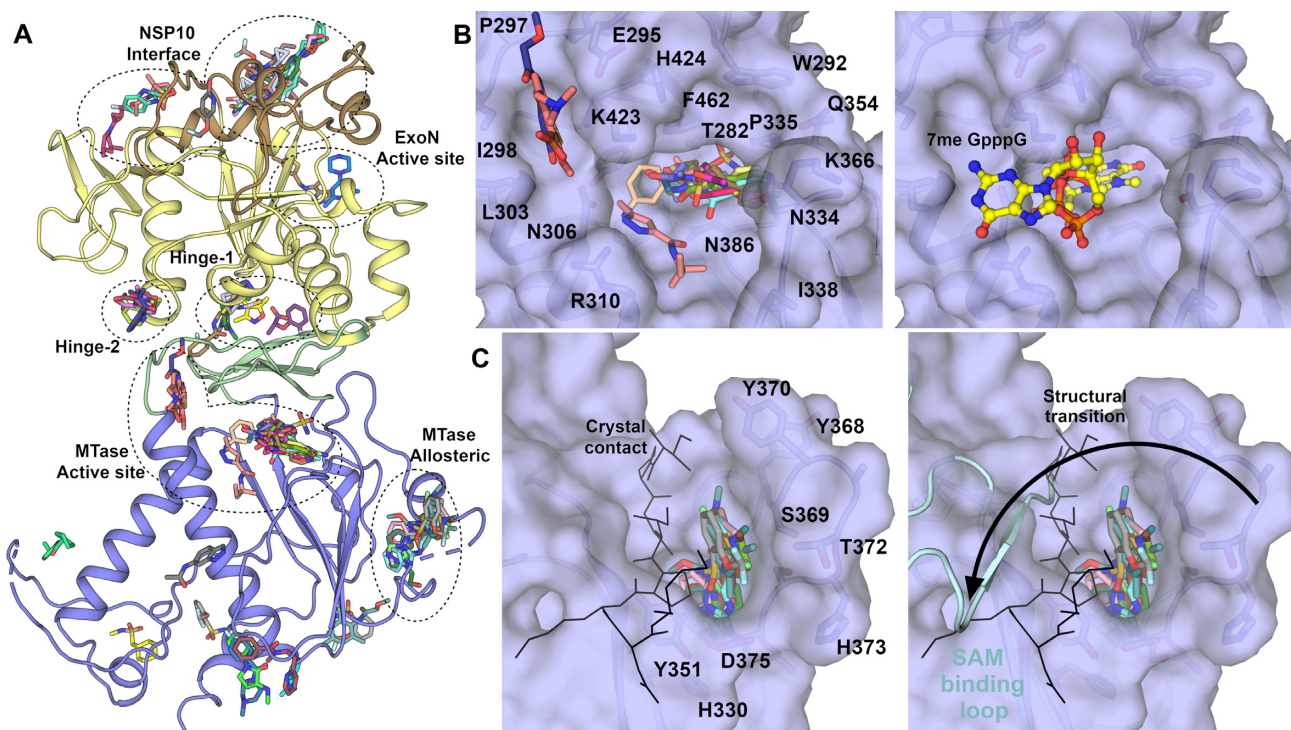


Figure 5. NSP14 fragment screening results overview and details of MTase domain pockets. (A) Overview of all 72 fragments observed to bind to NSP14 from our fragment screen. The major pockets and potential sites of inhibition are labelled. (B) Surface view of fragments bound to the MTase active site, the right hand panel shows the 7^{Me} GpppG viewed from the same perspective for reference. (C) Surface view of pockets bound to a potential allosteric site of inhibition for the MTase activity. The black stick representation shows the contribution of a crystallographic symmetry mate to the upper left side of the pocket. The right hand panel shows the conformation of the equivalent loop (shown in Cyan) in the SARS-CoV NSP14/10 complex. This loop is involved in binding SAM, and the putative inhibitory potential of these fragments against MTase activity may stem from their prevention of this structural transition.

un-methylated guanine moieties and as such is both a substrate and product mimic. The first nucleobase of SARS-CoV-2 transcripts is adenine although NSP14 has been demonstrated to be active on both GpppG and GpppA methyl cap analogues as substrates (29).

The 7^{Me} GpppG binds with the 7^{Me} G in the product state with the 7^{Me} G stacking between F426 and N386. The electron density map shows a clear bulge in the position of the methylated N7 (Supplementary Figure S3) and the methyl group is buried fairly deep within the pocket and makes Van der Waals interactions with P355. The rest of the guanine moiety stacks against F426 and N386 and makes polar interactions with T428 and D388 in a similar manner to that described previously for SARS-CoV NSP14(5). Other elements of the 7^{Me} GpppG binding pose are less similar to previous structures, the second and third phosphates taking a slightly different path and forming polar contacts to K423 and R310. The ribose and nucleobase of the base 1 (adenine in the case of the SARS NSP14 but guanine for our structure) is significantly less well ordered in the electron density map (Supplementary Figure S3) and is modeled in a position where it makes polar contacts to K423 (Figure 4C). The ribose O3 of the first nucleotide is in a solvent exposed position such that it would be possible to be incorporated into the NSP14 active site in the context of a full RNA cap molecule. This is in contrast to the SARS-CoV NSP14-GpppA-SAH ternary complex struc-

ture where the ribose is predominantly buried forming close contacts to K336 and I338. The discrepancy in the binding of the CAP analogues may be explained by the influence of crystal contacts in the SARS-CoV structure which would prevent the binding mode we have observed due to steric clashes (Supplementary Figure S4). The SAH moiety in this structure also makes several steric clashes in the vicinity of the ribose and homocysteine moieties including clashes to the GpppA cap analogue itself (Figure 4D). More widely the conformation of the SAM binding loop observed in the SARS-CoV NSP14/10 complex structures (5C8S, 5C8T and 5C8U) is not compatible with our crystal system as significant clashes would occur between this loop conformation and regions of the N-terminus of a crystallographic neighbor. There are also some crystal contacts with the same regions of our own structure which may play a role in stabilizing this particular conformation. We have also examined this loop in the recent high resolution cryo-EM structures of NSP14 (9) and the density is significantly poorer quality than the rest of the structure indicating that this loop is likely to be flexible and may adopt multiple conformations in solution. Taken together, these observations suggest the potential for a more complex catalytic mechanism involving conformational changes to the flexible SAM/SAH binding loop and the possibility of an ordered sequential mechanism of substrate binding and product release.

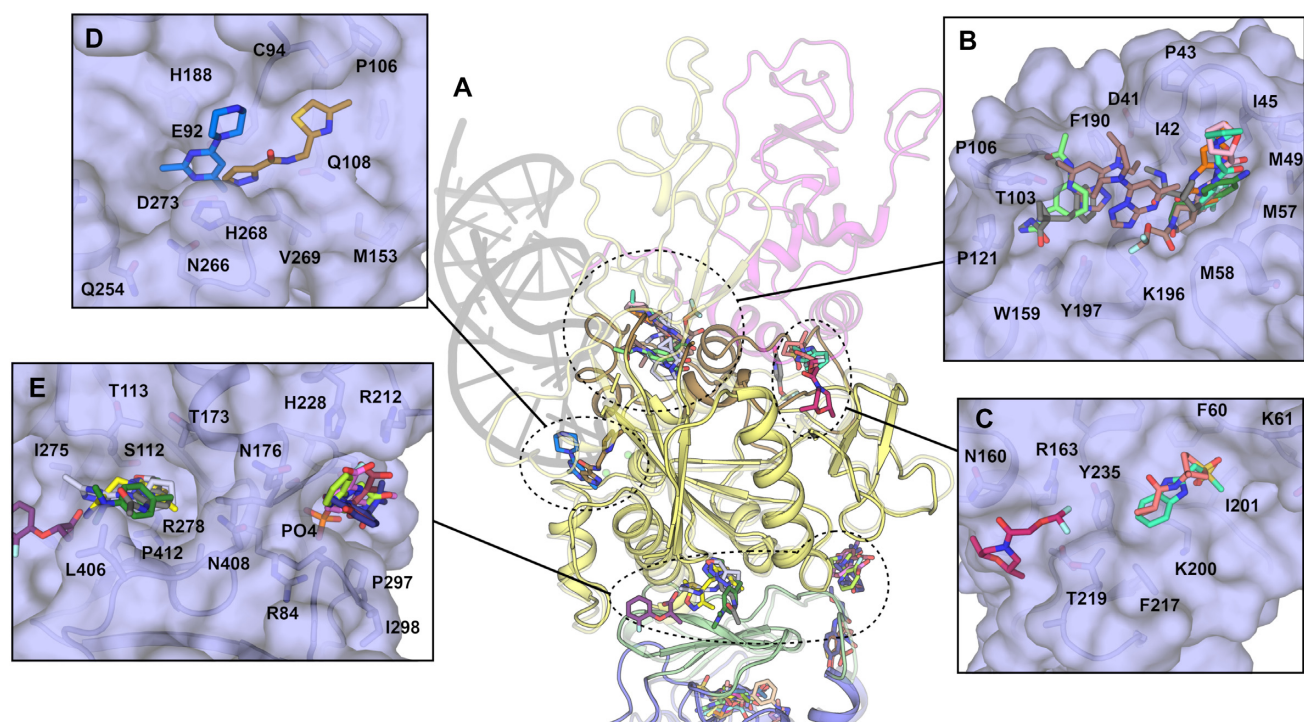


Figure 6. NSP14 fragments bound to pockets in the ExoN and hinge domains. (A) Overview of fragment binding sites with the NSP14/10 RNA complex shown in semitransparent representation for reference. (B) Surface view of fragments bound to NSP14 regions that undergo conformational changes to interact with NSP10. (C) Surface view of fragments bound directly on the NSP10 interface. (D) Surface view of fragments bound to the NSP14 ExoN active site. (E) Surface view of two clusters of fragments bound to pockets in the hinge domain.

X-ray fragment screening of NSP14

The fact that our crystals capture NSP14 in a pre-activated state prior to its interaction with NSP10 represents a unique opportunity to exploit this for drug discovery. Our crystals grow relatively robustly, are tolerant to DMSO and standard soaking protocols and routinely diffract to around 2.0 Å resolution, sufficient for the reliable identification of binders. A total of 634 fragments were soaked using the DSI poised fragment library (30) at a nominal ~50 mM final concentration, yielding 588 diffraction datasets with the majority diffracting better than 2.5 Å. Analysis of the electron density maps using the PANDDA algorithm (18) revealed a total of 72 fragments bound across 59 datasets (Figure 5A and Supplementary Table S2). The most prominent fragment hotspot was found in the MTase active site with 16 fragments (Figure 5B) with the majority of the fragments bound in the vicinity of the methylguanine and SAM moieties making stacking interactions with F426 and N386 and a number of polar contacts to nearby residues including N306, N334, N386, N388 and T428. No fragments were located in the vicinity of the ⁷MeGpppG phosphates although a smaller cluster of 3 fragments is bound close to the second guanine moiety and make polar contacts to Y296, N306 and N442 and hydrophobic interactions with P297 and I298. Another prominent pocket in the MTase domain with 11 fragments bound lies near the extended loop (residues 365–377) that was found to adopt a different conformation than in previous NSP14/10 complex structures. Fragments at this site make polar contacts to E365, L366 and N375 with a subset of the fragments (PDB entries 5SL1,

5SLM, 5SM9 and 5SMA) inducing the ordering of an otherwise disordered loop (residues 370 to 374) within the crystals, allowing the loop to be modelled into good quality electron density. Whilst we note that a significant proportion of the pocket and interactions formed with these fragments are provided by residues in a crystal contact (Figure 5C), the site is a good candidate for allosteric inhibition due to the fragments interfering with the ability of the SAM interacting loop (residues 352–378) to transition to the conformation seen in the SAM bound NSP14/10 complex structures (Figure 5C).

Further potential exists for allosteric inhibition of Exonuclease activity by means of blocking the ability of NSP14 to interact with NSP10 (Figure 6A). Four pockets were found in regions which are on the NSP10 interaction interface or were observed to transition upon complex formation. The most extensive cluster contains 10 fragments which span a relatively broad pocket bounded on one side by the first and second helices at the N-terminus and the disordered loop 95–102 on the other. Fragments make predominantly hydrophobic interactions on one side with I42, P46 and M57 and on the other side more polar contacts to T103, L105, A119, W159 and K196 (Figure 6B). Three fragments in two pockets are bound directly upon the NSP10 interface (Figure 6C), and two fragments bind directly in the ExoN domain active site forming polar contacts to residues N108 and D273, the latter being involved in coordination of the catalytic metal ions (Figure 6D).

Two clusters of fragments are observed to bind in the vicinity of the hinge domain that separates ExoN and

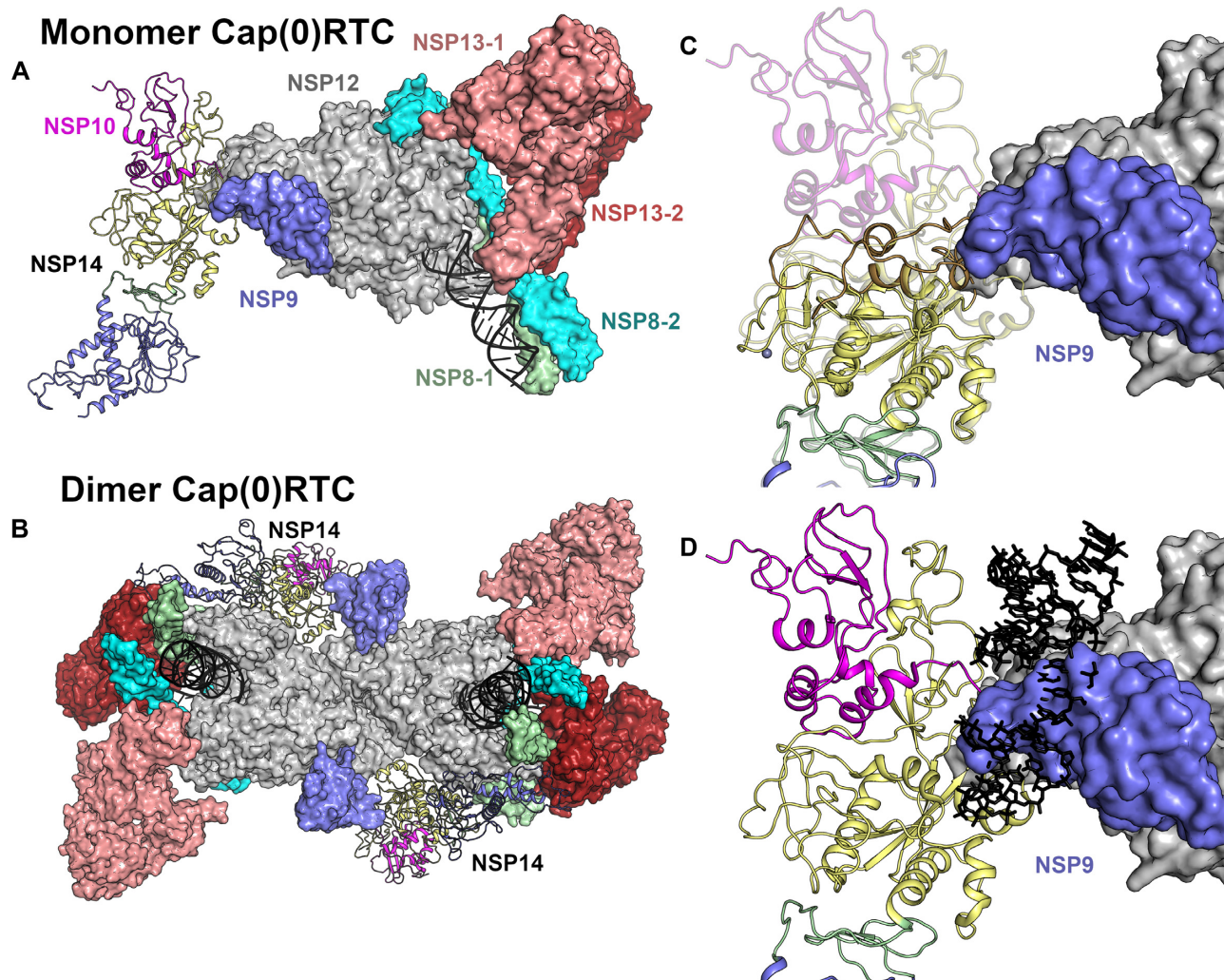


Figure 7. Evaluation of NSP14 as part of the Cap(0)RTC complex. (A) Overview of the positioning of NSP14 within the monomeric Cap(0)RTC complex based on the recently published Cryo EM structure(14). Subunits are colored individually as labeled with NSP14 and NSP10 shown in the cartoon format. (B) Overview of the positioning of NSP14 within the Dimeric Cap(0)RTC complex as viewed along the two-fold symmetry axis. (C) Overlay of NSP14 alone with the NSP14 NSP10 complex within the Cap(0)RTC complex. The N-terminal region of NSP14 forms significant steric clashes with NSP9. (D) Overlay of the NSP14-NSP10-RNA complex (7N0B)(9) onto the Cap(0)RTC complex. Severe steric clashes are evident with the RNA substrate (shown in black in stick format) suggesting this arrangement is not compatible with nuclease activity.

MTase domains. The largest containing 6 fragments in a relatively deep pocket close to the ExoN domain Zn binding site. Fragments make polar contacts to S112 and H264, hydrophobic contacts to I275, L406 and P412 and numerous water mediated contacts (Figure 6E). The second cluster with 4 fragments bound lies directly opposite near a phosphate binding site in the ExoN domain. These fragments make hydrophobic contacts to P297 and I299 and polar contacts to R212, H228 and the Phosphate ion itself (Figure 6E). It is possible that the interactions involving the phosphate ion which are quite extensive at this site may play a role in RNA substrate recognition, furthermore residues Y296 and P297 which line one side of this pocket were identified as being essential for ExoN activity in SARS-CoV NSP14 (20). The reason for this connection is not clear as these residues are distant (over 35 Å) from the ExoN active site and lie on the opposite side of the molecule closer to the MTase active site (Figure 5B & 6E). Whilst this is not

understood the proximity of P297 to a cluster of fragments overlapping with the MTase active site suggests that it may even be possible to develop inhibitors that inhibit both enzymatic activities.

DISCUSSION

In this study, we have determined the crystal structure of SARS-CoV-2 NSP14 in the absence of its partner NSP10 to high resolution. NSP14 is an important target for development of possible antivirals for several reasons. Direct inhibition of the nuclease proofreading activity may increase susceptibility of SARS-CoV-2 to nucleotide inhibitors such as remdesivir (31,32). Our structure offers a unique opportunity for discovery of inhibitors that bind to the ExoN domain in its inactive state with the potential to block the interaction with NSP10. The MTase domain is also a promising site to target for broad spectrum coronavirus inhibition,

and potentially druggable pockets have been identified that are well conserved across coronavirus species (33). To this end, potent SAM analogue inhibitors have been developed against NSP14 that show mid nM potency and promising selectivity profiles (34,35). Our fragment screens identify 72 fragments bound in several promiscuous pockets that represent useful starting points for inhibitor development or may be used for the structure based elaboration of existing hits. It is important to note that these fragments are soaked into crystals at relatively high concentrations and whilst they may have good ligand efficiency, are not expected to be potent inhibitors without further optimization. We suggest the druggable MTase domain active site with its large number of fragment hits, and high degree of sequence conservation may be the best place to start for the development of anti-viral therapeutics that may be able to combat the current pandemic and also future emerging viral threats.

A recent structure has been determined by cryo-EM for NSP14/NSP10 in complex with the rest of the RTC that has been dubbed Cap(0)-RTC(14). This complex is comprised of NSP12, NSP7, NSP9, two copies of NSP8 and NSP13 and a single copy of the NSP14/NSP10 sub-complex. In this model NSP14/NSP10 primarily contacts NSP9 and the NiRAN domain of NSP12 via residues in the NSP14 ExoN domain. A subset (15%) of the particles picked from micrographs belong to a dimeric form which has additional contacts formed between the NSP14 MTase domain and the Zn-binding domain of NSP13 and the ‘fingers’ and ‘thumb’ region of the NSP12 polymerase (14). In both monomeric and dimeric complexes, the conformation of the mobile SAM binding loop is identical to that of the previously reported SARS-CoV NSP14/NSP10 complex structure (5), although we note that the potential movements observed for this loop in our structure can be accommodated in both versions of the complex (Figure 7A, B). Outside of the MTase domain active site, the N-terminal NSP10 interacting region of the ExoN domain forms a significant overlap with residues in NSP9 (suggesting the interface is not compatible with NSP14 in the absence of NSP10) (Figure 7C). However, the recently published cryo-EM structure of NSP14/NSP10 in complex with RNA (9) would lead us to question the possible biological significance of these complexes, due to the fact that the path of the RNA to the ExoN active site is completely blocked by the presence of NSP9 (Figure 7D) suggesting this particular complex is not compatible with exonuclease activity. Moreover the Cap(0)-RTC complexes have been obtained using a modified form of NSP10 in which the proteolytic processing of the link to NSP9 had not occurred, based on the observation that other coronaviruses lacking a cleavage between NSP9 and NSP10 are viable (36). The position of the NSP9 does not change when comparing complexes of the RTC with NSP9 alone (37) or with the NSP9-NSP10/NSP14 fusion. This would suggest that the positioning of the fusion protein is dominated by the contacts made by NSP9. The authors of the Cap(0)-RTC do refer to a reconstruction without a NSP9-10 fusion that had ‘convincing density’ for the NSP14, although this data has not been made available. The lack of other experimental evidence to support the model would lead us to suggest some caution in its interpretation.

DATA AVAILABILITY

Crystallographic coordinates and structure factors for all structures have been deposited in the Protein Data Bank with the following accessing codes: 7QGI, 7QGF, 5KSW, 5SKX, 5SKY, 5SKZ, 5SL0, 5SL1, 5SL2, 5SL3, 5SL4, 5SL5, 5SL6, 5SL7, 5SL8, 5SL9, 5SLA, 5SLB, 5SLC, 5SLD, 5SLE, 5SLF, 5SLG, 5SLH, 5SLI, 5SLJ, 5SLK, 5SLL, 5SLM, 5SLN, 5SLO, 5SLP, 5SLQ, 5SLR, 5SLS, 5SLT, 5SLU, 5SLV, 5SLW, 5SLX, 5SLY, 5SLZ, 5SM0, 5SM1, 5SM2, 5SM3, 5SM4, 5SM5, 5SM6, 5SM7, 5SM8, 5SM9, 5SMA, 5SMB, 5SMC, 5SMD, 5SME, 5SMF, 5SMG, 5SMH, 5SMI.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The crystallographic screen was supported by the XChem facility at Diamond Light Source. We thank all the staff of Diamond Light Source beam-lines I03, I04 and I04-1 for providing support under Proposals LB26998, LB22717 and MX28172.

Author contributions: J.A.N. and Y.Y. initiated the project. Y.Y., J.A.N. and N.I. performed expression, protein purification, and crystallization. N.I. and J.A.N. performed crystal optimization and fragment soaking, Crystal mounting, and XChem data management. J.A.N. performed X-ray data analysis and review of fragment binding. J.A.N. and N.I. wrote the original draft manuscript. All authors read and approved the manuscript.

FUNDING

NCI [P01 CA092584]; Cancer Research UK Programme Award [A24759]. Funding for open access charge: University of Oxford.

Conflict of interest statement. None declared.

REFERENCES

- 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.
- van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T.M., Raj, V.S., Zaki, A.M., Osterhaus, A.D., Haagmans, B.L., Gorbalenya, A.E., Snijder, E.J. *et al.* (2012) Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *Mbio*, **3**, e00473-12.
- Minskaia, E., Hertzog, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B. and Ziebuhr, J. (2006) Discovery of an RNA virus 3'→5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 5108–5113.
- Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B. and Decroly, E. (2012) RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 9372–9377.
- Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R. and Rao, Z. (2015) Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 9436–9441.

6. Rogstam, A., Nyblom, M., Christensen, S., Sele, C., Talibov, V.O., Lindvall, T., Rasmussen, A.A., Andre, I., Fisher, Z., Knecht, W. *et al.* (2020) Crystal structure of non-structural protein 10 from severe acute respiratory syndrome coronavirus-2. *Int. J. Mol. Sci.*, **21**, 7375.
7. Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M. and Rao, Z. (2006) Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. *J. Virol.*, **80**, 7902–7908.
8. Lin, S., Chen, H., Chen, Z., Yang, F., Ye, F., Zheng, Y., Yang, J., Lin, X., Sun, H., Wang, L. *et al.* (2021) Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic Acids Res.*, **49**, 5382–5392.
9. Liu, C., Shi, W., Becker, S.T., Schatz, D.G., Liu, B. and Yang, Y. (2021) Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science*, **373**, 1142–1146.
10. 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.
11. Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M. *et al.* (2020) Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in Cap synthesis. *Cell*, **184**, 184–193.
12. Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T. and Guo, D. (2009) Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 3484–3489.
13. Chen, Y., Su, C., Ke, M., Jin, X., Xu, L., Zhang, Z., Wu, A., Sun, Y., Yang, Z., Tien, P. *et al.* (2011) Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog.*, **7**, e1002294.
14. Yan, L., Yang, Y., Li, M., Zhang, Y., Zheng, L., Ge, J., Huang, Y.C., Liu, Z., Wang, T., Gao, S. *et al.* (2021) Coupling of N7-methyltransferase and 3'-5' exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell*, **184**, 3474–3485.
15. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
16. 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.
17. 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.
18. 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. *et al.* (2017) A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.*, **8**, 15123.
19. Krafcikova, P., Silhan, J., Nencka, R. and Boura, E. (2020) Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat. Commun.*, **11**, 3717.
20. Ferron, F., Subissi, L., Silveira De Moraes, A.T., Le, N.T.T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B. *et al.* (2018) Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E162–E171.
21. Kozielski, F., Sele, C., Talibov, V.O., Lou, J., Dong, D., Wang, Q., Shi, X., Nyblom, M., Rogstam, A., Krojer, T. *et al.* (2021) Identification of fragments binding to SARS-CoV-2 nsp10 reveals ligand-binding sites in conserved interfaces between nsp10 and nsp14/nsp16. *RSC Chem. Biol.*, **3**, 44–55.
22. Hamdan, S., Carr, P.D., Brown, S.E., Ollis, D.L. and Dixon, N.E. (2002) Structural basis for proofreading during replication of the *Escherichia coli* chromosome. *Structure*, **10**, 535–546.
23. Beese, L.S. and Steitz, T.A. (1991) Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.*, **10**, 25–33.
24. Moeller, N.H., Shi, K., Demir, O., Belica, C., Banerjee, S., Yin, L., Durfee, C., Amaro, R.E. and Aihara, H. (2022) Structure and dynamics of SARS-CoV-2 proofreading exoribonuclease ExoN. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2106379119.
25. Baddock, H.T., Brolih, S., Yosaatmadja, Y., Ratnaweera, M., Bielinski, M., Swift, L.P., Cruz-Migoni, A., Fan, H., Keown, J.R., Walker, A.P. *et al.* (2022) Characterization of the SARS-CoV-2 ExoN (nsp14ExoN-nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *Nucleic Acids Res.*, **50**, 1484–1500.
26. Czarna, A., Plewka, J., Kresik, L., Matsuda, A., Karim, A., Robinson, C., O'Byrne, S., Cunningham, F., Georgiou, I., Pachota, M. *et al.* (2022) Refolding of lid subdomain of SARS-CoV-2 nsp14 upon nsp10 interaction releases exonuclease activity. *Structure*, **30**, 1050–1054.
27. Riccio, A.A., Sullivan, E.D. and Copeland, W.C. (2022) Activation of the SARS-CoV-2 NSP14 3'-5' exoribonuclease by NSP10 and response to antiviral inhibitors. *J. Biol. Chem.*, **298**, 101518.
28. Byszewska, M., Smietanski, M., Purta, E. and Bujnicki, J.M. (2014) RNA methyltransferases involved in 5' cap biosynthesis. *RNA Biol.*, **11**, 1597–1607.
29. Jin, X., Chen, Y., Sun, Y., Zeng, C., Wang, Y., Tao, J., Wu, A., Yu, X., Zhang, Z., Tian, J. *et al.* (2013) Characterization of the guanine-N7 methyltransferase activity of coronavirus nsp14 on nucleotide GTP. *Virus Res.*, **176**, 45–52.
30. Cox, O.B., Krojer, T., Collins, P., Monteiro, O., Talon, R., Bradley, A., Fedorov, O., Amin, J., Marsden, B.D., Spencer, J. *et al.* (2016) A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.*, **7**, 2322–2330.
31. Agostini, M.L., Andres, E.L., Sims, A.C., Graham, R.L., Sheahan, T.P., Lu, X., Smith, E.C., Case, J.B., Feng, J.Y., Jordan, R. *et al.* (2018) Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *Mbio*, **9**, <https://doi.org/10.1128/mBio.00221-18>.
32. Robson, F., Khan, K.S., Le, T.K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P. and Ng, W.L. (2020) Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cell*, **80**, 1136–1138.
33. Yazdani, S., De Maio, N., Ding, Y., Shahani, V., Goldman, N. and Schapira, M. (2021) Genetic variability of the SARS-CoV-2 pocketome. *J. Proteome Res.*, **20**, 4212–4215.
34. Devkota, K., Schapira, M., Perveen, S., Khalili Yazdi, A., Li, F., Chau, I., Ghiabi, P., Hajian, T., Loppnau, P., Bolotokova, A. *et al.* (2021) Probing the SAM binding site of SARS-CoV-2 Nsp14 in vitro using SAM competitive inhibitors guides developing selective bisubstrate inhibitors. *SLAS Discov.*, **26**, 1200–1211.
35. Otava, T., Sala, M., Li, F., Fanfrlik, J., Devkota, K., Perveen, S., Chau, I., Pakarian, P., Hobza, P., Vedadi, M. *et al.* (2021) The structure-based design of SARS-CoV-2 nsp14 methyltransferase ligands yields nanomolar inhibitors. *ACS Infect. Dis.*, **7**, 2214–2220.
36. Deming, D.J., Graham, R.L., Denison, M.R. and Baric, R.S. (2007) Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. *J. Virol.*, **81**, 10280–10291.
37. Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M. *et al.* (2021) Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in Cap synthesis. *Cell*, **184**, 184–193.