The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine

Monica Rosas-Lemus[1,2], George Minasov[1,2], Ludmilla Shuvalova[1,2], Nicole L. Inniss[1,2], Olga Kiryukhina[1,2], Grant Wiersum[1,2], Youngchang Kim[2,3], Robert Jedrzejczak[2,3], Michael Enders[2,3], Lukasz Jaroszewski[2,3], Adam Godzik[2,3], Andrzej Joachimiak[2,3], and Karla J. F. Satchell[1,2*]

[1]Department of Microbiology-Immunology, [2]Center for Structural Genomics of Infectious Diseases, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

[3] Consortium for Advanced Science and Engineering, University of Chicago, Chicago, IL 60667, USA

[4]Department of Biomedical Sciences, University of California, Riverside School of Medicine, Riverside, CA, USA

*Correspondence: k-satchell@northwestern.edu

Abstract

SARS-CoV-2 is a member of the coronaviridae family and is the etiological agent of the respiratory Coronavirus Disease 2019. The virus has spread rapidly around the world resulting in over two million cases and nearly 150,000 deaths as of April 17, 2020. Since no treatments or vaccines are available to treat COVID-19 and SARS-CoV-2, respiratory complications derived from the infections have overwhelmed healthcare systems around the world. This virus is related to SARS-CoV-1, the virus that caused the 2002-2004 outbreak of Severe Acute Respiratory Syndrome. In January 2020, the Center for Structural Genomics of Infectious Diseases implemented a structural genomics pipeline to solve the structures of proteins essential for coronavirus replication-transcription. Here we show the first structure of the SARS-CoV-2 nsp10-nsp16 2'-O-methyltransferase complex with S-adenosylmethionine at a resolution of 1.8 Å. This heterodimer complex is essential for capping viral mRNA transcripts for efficient translation and to evade immune surveillance.

Introduction

On December 31st, 2019, the World Health Organization (WHO) was alerted of a pneumonia outbreak with an unknown etiology, originating in the Chinese province of Wuhan, Hubei (1,2). The etiological agent was identified as a coronavirus, closely related to the virus responsible for Severe Acute Respiratory Syndrome (SARS). The new SARS coronavirus-2 (SARS-CoV-2) causes the severe respiratory infection, Coronavirus Disease 2019 (COVID-19) (3). Within four months, SARS-CoV-2 rapidly spread, sparking a global pandemic. The COVID-19 pandemic has also forced government-enacted "stay-at-home" orders around the world. According to the World Health Organization, 2,074,529 SARS-CoV-2 infections have been confirmed, of which 139,378 were fatal as of April 17th. These data are similar to Johns Hopkins University tracking system, that reported 2,182,734 of infections and 147,384 deaths (1).

The coronaviridae family of viruses causes disease in birds and mammals, including bats, camels, pigs, and humans. In lower vertebrates, pathogenic coronaviruses cause acute and severe gastrointestinal infections, fevers and organ failure. Of the seven human-tropic coronaviruses, hCoV-229E, hCoV-NL63, hCoVB-OC43 cause only asymptomatic or mild infections, including the common cold (4-6). Four of the viruses are linked to severe infections; including, hCoV-HKU1, a common cause of pneumonia, SARS-CoV-1 with a 10% mortality rate, Middle East Respiratory Syndrome Virus (MERS-CoV) with a 37% mortality rate, and SARS-CoV-2 currently with a 6% mortality rate (3). As SARS-CoV-2 continues to spread, the need for effective vaccines and therapeutics increases. In addition, there is currently little information about the immunological response to the virus or the potential for re-infection (7-10). Therefore, it is urgent to study SARS-CoV-2 mechanisms of infection and replication in order to find effective targets for drug and vaccine development.

Coronaviruses have a large (~ 30 kb) single-stranded, positive RNA genome that is 5'-capped, contains a 3'-poly-A tail, and are direct templates for the transcription of sub-genomic mRNAs for the translation of viral proteins. The first open reading frame produces the large non-structural polyprotein 1a (pp1a) and read-through across a frameshift results in translation of the larger non-structural polyprotein 1ab (pp1a/b). These polyproteins are subsequently processed into sixteen non-structural proteins (nsps) that assemble to form the Replication-Transcription Complex (RTC) or function as accessory proteins necessary for viral replication (11-14).

The components of the RTC include enzymes that regulate mRNA and genomic RNA synthesis, proofreading, and mRNA maturation. Two of these enzymes are critical for capping viral mRNAs, a tactic employed by multiple RNA viruses to avoid immune detection by toll-like receptor 7 (TLR7) and toll-like receptor 8 (TLR8) (15). In eukaryotic cells, mRNA capping is initiated by an RNA triphosphatase (TPase), which removes the γ-phosphate from the 5'-end of the nascent mRNA transcript, generating a diphosphate 5'-ppN end. An RNA guanylyltransferase (GTase) subsequently catalyzes the hydrolysis of pyrophosphate (PPi) from a guanidine

triphosphate (GTP) molecule forming GMP, followed by the transfer of the α-phosphate of guanidine monophosphate (GMP) to the diphosphate 5'-ppN transcript end, forming the cap core structure, referred to as GpppN. The GpppN formation is followed by N^7-methylation of the capping guanylate by a guanine-N^7-methyltransferase (N^7-MTase) to generate the Cap-0. Further methylation at the ribose 2'-O position of first nucleotide of the RNA is catalyzed by a ribose 2'-O-methyltransferases (2'-O-MTase) to generate Cap-1 and sometimes at the second nucleotide to generate Cap-2 (4). Both the N^7-MTase and 2'-O-MTase use S-adenosyl-L-methionine (SAM) as the methyl group donor (4,16).

For coronavirus mRNA maturation, the host cell TPases and GTase are used to guanylate the 5'-end of the nascent mRNA and the viral non-structural protein 14 (nsp14) N^7-MTase activity generates the Cap-0 (4). Nsp14 is a bifunctional enzyme with an exonuclease domain in addition to its N^7-MTase domain (17). Its activity is modulated by the binding of the small viral protein, nsp10, which specifically stimulates its exonuclease activity with no effect on its N^7-MTase activity (16). The coronavirus mRNAs are further modified to have a Cap-1 by the viral non-structural protein 16 (nsp16). Nsp16 is a m^7GpppA-specific, SAM-dependent, 2'-O-MTase (18,19) and is activated by binding to nsp10 (20). Nsp10 is a stable protein that can self-dimerize (5) or form dodecamers with nsp11 (6), in addition to modulating nsp14 and nsp16 activity (21). Although no specific enzymatic activity has been identified for nsp10, it is known as a zinc binding protein that binds nonspecifically to RNA and stabilizes the SAM binding pocket in nsp16 to form a stable complex (20). The nsp10-nsp16-mediated 2'-O-methylation of coronavirus RNA is essential for preventing host recognition and decreasing immune response, while the viral RNA is translated (16,18,22).

Multiple crystal structures of the nsp16-nsp10 heterodimer have been solved, from both SARS-CoV-1 and MERS-CoV (4,19,23) and provide some insight into the structural basis of substrate binding and the proposed sn2-mechanism of methyl transfer. To facilitate the development of vaccine and drug targets for SARS-CoV-2, the Center for Structural Genomics

for Infectious Diseases (CSGID) has implemented a high-throughput structural determination pipeline in order to solve the structure of SARS-CoV-2 proteins that are amenable to drug discovery and vaccine development. Here we present the 1.80 Å X-ray crystal structure of SARS-CoV-2 nsp10-nsp16 heterodimer with SAM bound.

Results

Overall structure of SARS-CoV-2 nsp10-nsp16 heterodimer. Nsp10 and nsp16 are encoded in the polycistronic *orf1ab* of the (+)ssRNA from SARS-CoV-2 (Fig. 1A) and are released by processing of the polyproteins by the nsp5 protease (14). Nsp10 is released from pp1a and pp1ab, while nsp16 is part of only the pp1ab polyprotein created after a -1 ribosome shift (24). Previous studies from SARS-CoV-1 demonstrated that nsp16 is an unstable protein in several conditions and thus the protein was purified in complex following co-expression with nsp10 (25). Here, we successfully expressed and purified nsp10 and nsp16 separately using nickel affinity and size exclusion chromatography (SEC). After removal of the 6xHis-tags, a heterodimer of nsp16 with nsp10 was formed and concentrated by centrifugation across a 30 kDa cutoff membrane and dialyzed into crystallization buffer. The protein complex was further incubated with 2 mM SAM and diffraction quality crystals grew in A7 of the ComPAS screen (0.2 M Calcium acetate, 0.1 M HEPES, pH 7.5, 18% w/v PEG 8000).

The structure of nsp16 in complex with nsp10 and SAM was solved by molecular replacement at 1.8 Å (PDB code 6w4h) using the nsp10-nsp16 heterodimer structure from SARS-CoV-1 (PDB code 3r24) as the model. The diffraction data are presented in Table 1 and a representative image of the structures is shown in Figure 1. The crystal belonged to the space group $P3_121$ with two polypeptide chains, one chain of nsp10 and one chain of nsp16 in the asymmetric unit, which form a heterodimer. The heterodimer has a total solvent-exposed surface area of 19,710 $Å^2$ estimated by PISA (Fig. 1B). This heterodimer is mainly stabilized by hydrophobic interactions and hydrogen bounds in the interface of nsp10 and nsp16. In parallel

with this work, a nearly identical structure with the same space group was solved at 2.0 Å (PDB code 6w61). The crystal obtained for this structure was prepared by a distinct procedure (see materials and methods). A third structure was also solved at 1.95 Å (PDB code 6w75) by the same methods used for the first structure. The third crystal form belongs to the $P3_221$ space group with twice the unit cell than first two. In this structure, there are four polypeptide chains, which form a tetramer, and this structure will be described later. For this report, only highest resolution structure (PDB code 6w4h) is discussed in detail.

In the representative crystal structure (PDB code 6w4h), there are several ligands bound to the heterodimer: two molecules of acetate (ACT), two molecules of β-D-fructopyranose (BDF), one molecule of sulfite (SO_3), two Zinc ions and one molecule of SAM. All of these ligands, except the MTase co-factor SAM and two Zinc ions, are derived from the crystallization and freezing solutions. It is likely that their interaction with the protein complex results from its electrostatic potential, as predicted by the Adaptive Poisson-Boltzmann Solver program (26) and modeled as vacuum surface electrostatics in the software package PyMOL. For example, one of the ACT molecules interacts in the positively charged nucleotide binding groove (Fig. 1C).

Structure of nsp10. The structure of SARS-CoV-2 nsp10 consist of residues 19-133 (residues 4272-4392 of pp1a). It contains a positively charged and hydrophobic surface that interacts with a hydrophobic pocket and a negatively charged surface from nsp16, which contributes to the stabilization of the SAM binding site (Fig. 1C). The structure of nsp10 of SARS-CoV-2 is composed of a central anti-parallel pair of β-strands surrounded on one side by a crossover large loop. The other side is a helical domain with loops that form two zinc fingers (Figure 2 A). These structures are involved in the non-specific binding of RNA in other coronaviruses (4,27).The Zn-binding site 1 is coordinated by the residues C74, C77, H83 and C90. The Zn-binding site 2 is coordinated by C117, C120, C128, and C130.

Structure of nsp16. The nsp16 structure has all 298 residues (orf1ab residues 6799-7096) with the Serine-Asparagine-Alanine residues at the N-terminus derived from the tobacco etch virus (TEV) protease cleavage site. The structure of nsp16 contains the 2′-O-MTase catalytic core comprised of a Rossmann-like β-sheet fold decorated by eleven α-helices, seven β-strands, and loops (Fig. 2C and D). The SARS-CoV-2 nsp16 fold is formed by a β-sheet with the canonical 3-2-1-4-5-7-6 arrangement, in which β7 is the only antiparallel strand (Figure 2C). This β-sheet is sandwiched by loops and α-helices. The catalytic core binds one molecule of SAM near the β1 and β2 strands of the Rossmann-like fold, and the negatively charged binding SAM binding pocket is formed by the loops 71-79 and 100-108, and 130-148 (Figure 2D). Residues Lysine-46, Aspartic acid-130, Lysine-170, and Glutamic acid-203, comprising the canonical SAM MTase motif, are in close proximity to the SAM methyl group catalytically transferred to the 2'-O sugar (Fig. 2 E) (4,16,19,28).

nsp10 and nsp16 are highly conserved among lineage B of the beta coronaviruses. In order to explore whether SARS-CoV-2 has amino acid variation in the 2-O'MTase that impact function, we compared the primary amino acid sequences of nsp10 and nsp16 proteins to that of other beta-coronaviruses (Fig. 3 A and B). Nsp10 was found to be 100% identical to Bat-CoV-RaTG13, 99% identical to SARS-CoV-1, and 98% identical to Bat-SARS-like coronavirus (Bat-SL-CoV) isolate Rs4247, but only 59% identical to MERS-CoV (Figure 3A). In the aligned sequences, the zinc coordinating residues C74, C77, H83, C90, C117, C120, C128, and C130 are 100% conserved, emphasizing the importance of zinc binding for nsp10 activity during the viral replication-transcription processes. Thus, nsp10 is highly conserved in the lineage B beta-coronaviruses, and less conserved with lineage C beta coronaviruses.

Similarly, nsp16 from SARS-CoV-2 is 99% identical to Bat-CoV-RaTG13, 94% identical to Bat-SL-CoV Rs4247, and 93% identical to SARS-CoV-1 (Figure 3B), but only 66% identical to

MERS-CoV. SARS-CoV-2 nsp16 contains the highly conserved K46, D130, K170, E203 (KDKE) motif for methyl-transfer found in many 2′-O-MTases (22), including the more phylogenetically distant MERS-CoV. Furthermore, we confirmed that residues N43, Y47, G71, A72, S74, G81, D99, N101, L100, D114, and M131, which coordinate SAM substrate binding in SARS-CoV-1 through hydrogen bonds and water-mediated interactions (4), are also 100% conserved in SARS-CoV-2 (Fig. 2F). The high conservation of nsp10 and nsp16 sequences, and the identification of 100% conserved catalytic and substrate-binding residues suggest that SARS-CoV-2 nsp10-nsp16 function and structure are also highly conserved.

Structural comparison of 2′-O-MTase from SARS-CoV-2 with SARS-CoV-1 and MERS-CoV. In the primary sequence alignments of nsp10, there are only two amino acid differences (A34P and K113) between SARS-CoV-2 and SARS-CoV-1. These changes do not affect the structure of nsp10 or its interactions with Zn^{2+} or nsp16. On the other hand, there are 16 amino acid differences between the SARS-CoV-2 and SARS-CoV-1 nsp16 sequences (Figure 3 A and B). All of these amino acid variations in SARS-Cov-2 occurred on the solvent-exposed surface opposite of the nsp10-nsp16 interface (Figure 3C), suggesting that they do not alter dimerization. Furthermore, it appears that the amino acid changes did not alter the catalytic site, substrate binding pocket, or heterodimer interface.

Structural alignment using the DALI server shows the highest coverage, percent identity, and structural conservation with the SARS-CoV-1 complex bound to SAM (PDB code 3r24 (4) , z= 48.3, r.m.s.d=0.9) or to S-adenosylhomocysteine (PDB code 2xyq (4) z=47.5, r.m.s.d.=1). The structure is also highly similar to the unbound MERS-CoV complex (PDB code 5yn5 (29) , z=47.2, r.m.s.d.=0.8) and the structure bound to SAM (PDB code 5YN6 (29), z=46.7, r.m.s.d.= 0.7). The differences found between the structures is mostly influenced by the absence of some residues from the N and C-terminal sequences and the flexibility of the loops 131-148 and 28-38 (Figure

3C), which are disordered in the MERS-CoV nsp16 structures (PDB codes 5yn5 and 5yn6 (29)) or have a different conformation in the structure of SARS-CoV-1 bound to SAM (PDB code 3r24).

Overall, the structure of SARS-CoV-2 nsp10-nsp16 heterodimer is very conserved among coronaviruses and further supports the idea that there is functional and mechanistic conservation of the nsp16 2'-O-MTases for mRNA maturation in coronaviruses.

Comparison with other eukaryotic MTases. Although the primary sequences vary, MTases of viral MTases have high similarity with eukaryotic MTases (28,30). To identify structurally-related, non-viral MTases, the nsp16-nsp10 structure was submitted to the DALI server (30), which identified eukaryotic MTases that are structurally homologous. These include the human cap-specific 2′-O-MTase (PDB code 4n48, z=13.7, r.m.s.d= 3.4) and the yeast 27S pre-rRNA guanosine (2922)-2'-O)-MTase (PDB code 6em5-W, z=12.4, r.m.s.d.=3.1). The main differences are located in the loops that surround the beta sheet that is fully conserved, which corroborates the high degree of structural conservation in this pathway. The close alignment reveals that the mechanism of coronavirus 2'-O-methylation is conserved with eukaryotic cell mechanisms.

Conclusions

The SARS-CoV-2 pandemic has yielded a world-wide effort to understand the molecular mechanisms involved in virus transmission, virulence, and replication. The ultimate goal is to identify viral proteins that are amenable for drug targeting and epitopes suitable for vaccine development. Although, previous studies conducted in SARS-CoV-1 and MERS-CoV enlighten the pathway for drug discovery and vaccine development, no approved treatments were fully developed. In addition, due to the high replication error rate of ssRNA viruses, small amino acid variations could cause resistance to previously described compounds or increase the pathogenicity. Thus, in order to ensure an accurate approach for drug discovery, the CSGID implemented a structural genomics pipeline to solve the structure of proteins essential for the

replication of SARS-CoV-2. Here we presented the structure of the 2′-O-MTase, nsp10-nsp16 heterodimer in complex with SAM at 1.8 Å resolution, which shares 98% identity with the 2′-O-MTase of SARS-CoV-1. All the amino acid variations in the new SARS-CoV-2 were located at the solvent exposed surface of the heterodimer, and not among the residues of the catalytic site, substrate binding site, or interface between nsp10-nsp16. Furthermore, the only identified discrepancies in our model versus previous models were in the flexible loop regions 28-38 and 131-148 and the N- and C-terminal residues. Notedly, these residues were derived from SARS-CoV-1 nsp10 fusion to nsp11. As the 2′-O-MTase activity of nsp10-nsp16 is required for viral replication-transcription in host cells, these structural data can be used for developing new specific treatments against SARS-CoV-2. Importantly, the structure reported here is the highest resolution structure thus far of a coronavirus 2′-O-MTase and is amenable for computational modeling of potential inhibitors.

Materials and Methods

In order to increase the probability to obtain high quality crystals suitable for structure determination, the CSGID implemented two different approaches, which are listed here and identified by their PDB accession code.

Method 1, PDB code 6wh4

Cloning and protein production

The genes from nsp10 and nsp16 were codon optimized for expression in *E. coli*, synthetized (Twist Bioscience) and cloned into pMCSG53 (31) vector, which contains a TEV cleavable N-terminal 6xHis-tag, encodes ampicillin resistance, and genes for rare codons. The plasmids were transformed into *E. coli* BL21(DE3)(Magic) cells (32). The proteins were expressed separately in 10 ml LB supplemented with 130 µg/ml and 50 µg/ml KAN at 37°C and 220 rpm, overnight. The

next day, 3 liters of TB medium supplemented with 200 µg/ml AMP were inoculated at 1:100 dilution with the overnight culture and incubated at 37°C and 220 rpm. Protein expression was induced at OD_{600}=1.8-2 by addition of 0.5 mM isopropyl β- d-1-thiogalactopyranoside and the cultures were further incubated at 25°C, at 200 rpm for 14 h (33). The cells were harvested by centrifugation and resuspended (1g of cells: 5 ml of lysis buffer) in lysis buffer (50 mM Tris, 0.5 M NaCl, 10% glycerol, 0.1% IGEPAL CA-630) and frozen at -30°C until purification.

Protein purification

Frozen pellets were thawed and sonicated at 50% amplitude, in 5 s x 10 s cycle for 20 min at 4°C. The lysate was cleared by centrifugation at 18,000 x *g* for 40 min at 4°C, the supernatant was collected, and the protein was purified as previously described with some modifications (34). The supernatant was loaded into a His-Trap FF (Ni-NTA) column using a GE Healthcare ÅKTA Purifier system in loading buffer (10 mM Tris-HCl pH 8.3, 500 mM NaCl, 1 mM Tris(2-carboxyethyl) phosphine (TCEP), 2 mM $MgCl_2$ and 5% glycerol). The column was washed with 10 column volumes (cv) of loading buffer and 10 (cv) of washing buffer (10 mM Tris-HCl pH 8.3, 500 mM NaCl, 25 mM imidazole), and was eluted with elution buffer (10 mM Tris pH 8.3, 500 mM NaCl, 1 M imidazole). The protein was loaded onto a Superdex 200 26/600 column and ran with loading buffer. The protein was collected and mixed with TEV protease (1:20 protein:protease) and incubated overnight at room temperature for removal of the 6xHis-tag. The cleaved tag was separated from protein by Ni-NTA-affinity chromatography using loading buffer. nsp10 was collected in the flow through, whereas nsp16 eluted at 50 mM imidazole. To form the nsp10-nsp16 complex, the pure cleaved proteins were mixed at a 1:1 molar ratio at approximately 2 mg/ml in loading buffer and incubated for 1 h, then dialyzed in crystallization buffer (10 mM Tris-HCl pH 7.5, 150 mM) NaCl for 2 h (23). SAM was added to a final concentration of 2 mM. The complex was concentrated to 5.5 mg/ml and set up immediately for crystallization.

Crystallization and data collection

The pure nsp10-nsp16 complex (1:1 ratio for nsp10 and nsp16) + SAM was set up as 2 µl crystallization drops in (1µl protein:1µl reservoir solution) in 96-well plates and equilibrated using commercially available Classics II, PEG's II, Anions and ComPas Suites. Diffraction quality crystals appeared after 5 days in conditions A7 of the ComPAS screen (0.2 M Calcium acetate, 0.1 M HEPES, pH 7.5, 18% w/v PEG 8000).

The data set was collected on the LS-CAT 21-ID-F beamline at the Advanced Photon Source (APS) at the Argonne National Laboratory. A total of 200 images, which corresponded to 120 degrees of the spindle axis rotation were collected. Images were indexed, integrated and scaled using HKL-3000 (35).

Structure solution and refinement

The structure of native nsp10-nsp16 complex crystal from SARS-CoV-2 was solved by Molecular Replacement with Phaser (36) from the CCP4 Suite (37) using the crystal structure of the nsp10/nsp16 heterodimer from SARS (PDB ID 3r24) as a search model. Initial solutions went through several rounds of refinement in REFMAC v. 5.8.0258 (38) and manual model corrections using Coot (39). The water molecules were generated using ARP/wARP (40), S-Adenosylmethionine (SAM), Zinc ions and ligands were added to the model manually during visual inspection in Coot. Translation–Libration–Screw (TLS) groups were created by the TLSMD server (41) (http://skuldbmsc.washington.edu/~tlsmd/) and TLS corrections were applied during the final stages of refinement. MolProbity (42) (http://molprobity.biochem.duke.edu/) was used for monitoring the quality of the model during refinement and for the final validation of the structure. Final model and diffraction data were deposited to the Protein Data Bank (https://www.rcsb.org/) with the assigned PDB code 6wh4.

Method 2 PDB code 6w61

Cloning and protein production

The cloning strategy for nsp10 and nsp16 was the same than the previous protocol except that they were transformed in BL21(DE3)-Gold strain cells. For protein expression, 4 L culture of LB Lennox medium was grown at 37°C (190 rpm) in presence of 150 μg/ml ampicillin to OD_{600} ~1.0, then incubated at 4°C until the cultures reached 18°C.Protein production was induced by addition of 0.2 mM IPTG, 0.1% glucose, 40 mM K_2HPO_4 to the medium, and cultures were incubated at 18 °C for 20 hours. Bacterial cells were harvested by centrifugation at 7,000 x *g* and cell pellets were resuspended in lysis buffer (500 mM NaCl, 5% (v/v) glycerol, 50 mM HEPES pH 8.0, 20 mM imidazole and 10 mM β-mercaptoethanol).

Protein purification

Purification of nsp10 and nsp16 (PDB ID 6W61) from the fresh suspension of cells in lysis buffer was sonicated at 120W for 5 minutes (4s ON, 20 sec OFF cycle). The cellular debris was removed by centrifugation at 30,000 *g* for one hour at 4°C. The supernatant was mixed with 4 ml of Ni^{2+} Sepharose (GE Healthcare Life Sciences) equilibrated with lysis buffer supplemented to 50 mM imidazole pH 8.0 and suspension was applied on Flex-Column (420400-2510) connected to Vac-Man vacuum manifold (Promega). Unbound proteins were washed out via controlled suction with 160 ml of lysis buffer (50 mM imidazole). Bound proteins were eluted with 20 ml of lysis buffer supplemented to 500 mM imidazole pH 8.0. 2 mM DTT was added followed by TEV protease treatment at 1:20 protease:protein ratio at 4°C overnight. Next, proteins were concentrated and run on size exclusion column Superdex 200 in lysis buffer supplemented with 1 mM TCEP. The fractions containing protein of interest were pooled and run through Ni^{2+} Sepharose and the proteins were collected in the flow through. Finally, buffer was exchanged to crystallization buffer (150 mM NaCl, 20 mM HEPES pH 7.5, 1 mM TCEP) (23) via 10X concentration/dilution repeated

3 times. Final concentrations of Nsp10 and Nsp16 were 8.5 mg/ml and 5 mg/ml, respectively. Proteins were either flash cooled or used immediately for crystallization trials.

Crystallization and data collection

To obtain crystal form 2, the complex was screened by the sitting-drop vapor-diffusion method in 96-well CrystalQuick plates (Greiner Bio-One) which had been set up using Mosquito liquid dispenser (TTP LabTech), (400nl protein: 400 nl reservoir solution) equilibrated against 135 nl reservoir solution against MCSG1, MCSG4 (Anatrace), SaltRX (Hampton) and INDEX (Hampton) screens were used for protein crystallization at 16°C. Diffraction quality crystals appeared after 3 days in MCSG4 G4 (0.1 M Sodium citrate pH 5.6, 10 %(w/v) PEG4000).

Data were collected at the SBC ID beamline sector 19 at the APS. A total of 954 diffraction images from five different parts of the rod-shaped crystal were collected to 2.00 A on the Dectris Pilatus3 X 6M detector. The data were processed and scaled by HKL3000 (43).

Structure solution and refinement

The structure was determined by molecular replacement using MolRep (44) implemented in HKL3000 and 6w4h structure as a search model. The structure was refined iteratively using Phenix.refine (45) and manually corrected in Coot. Water molecules were automatically generated in Phenix.refine while SAM, Zinc ions and ligands were added manually in Coot. The stereochemistry of the structure was validated using MolProbity and the structure was deposited to PDB with the assigned PDB code 6w61.

Structural and sequence alignment

The protein sequence of nsp10 and nsp16 from SARS-CoV-2 (YP_009725295.1), Bat-CoV-RaTG13 (QHR63299.1) and Bat-SL-CoV Rs4247 (ATO98179.1) SARS-CoV-1 (ACZ72252.1)

and MERS-CoV (YP_009047238.1) were obtained from the NCBI database. The multiple sequence alignment was performed using ClustalO (https://www.ebi.ac.uk/Tools/msa/clustalo/) and merged with the PDB file of 6w4h using ESPript 3.x (46). The PDB coordinates of the SARS-CoV-1 nsp10-nsp16 complex were analyzed on the DALI (47), POSA and (48) FATCAT (49) servers to perform structural and sequence alignment. Structural alignments and structure figures were conducted using Matchmaker in Pymol (50) open source V 2.1.

Acknowledgements

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases under Contract Nos. HHSN272201200026C and HHSN272201700060C and Grant Nos. Grant Nos. U01AI124316 and R01 GM05789 and from the National Institute of General Medical Sciences grant GM118187, both from National Institutes of Health, Department of Health and Human Services.

References

1. Dong, E., Du, H., and Gardner, L. (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*
2. 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
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. Chen, Y., Su, C., Ke, M., Jin, X., Xu, L., Zhang, Z., Wu, A., Sun, Y., Yang, Z., Tien, P., Ahola, T., Liang, Y., Liu, X., and Guo, D. (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
5. Algahtani, H., Subahi, A., and Shirah, B. (2016) Neurological Complications of Middle East Respiratory Syndrome Coronavirus: A Report of Two Cases and Review of the Literature. *Case Rep Neurol Med* **2016**, 3502683
6. Chau, T. N., Lee, K. C., Yao, H., Tsang, T. Y., Chow, T. C., Yeung, Y. C., Choi, K. W., Tso, Y. K., Lau, T., Lai, S. T., and Lai, C. L. (2004) SARS-associated viral hepatitis caused by a novel coronavirus: report of three cases. *Hepatology* **39**, 302-310

7. Moon, C. (2020) Fighting COVID-19 exhausts T cells. *Nat Rev Immunol*
8. Vaninov, N. (2020) In the eye of the COVID-19 cytokine storm. *Nat Rev Immunol*
9. Weiss, S. R., and Navas-Martin, S. (2005) Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev* **69**, 635-664
10. Zhou, L., Liu, K., and Liu, H. G. (2020) [Cause analysis and treatment strategies of "recurrence" with novel coronavirus pneumonia (covid-19) patients after discharge from hospital]. *Zhonghua Jie He He Hu Xi Za Zhi* **43**, E028
11. Furuichi, Y., and Shatkin, A. J. (2000) Viral and cellular mRNA capping: past and prospects. *Adv Virus Res* **55**, 135-184
12. Hussain, S., Pan, J., Chen, Y., Yang, Y., Xu, J., Peng, Y., Wu, Y., Li, Z., Zhu, Y., Tien, P., and Guo, D. (2005) Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. *J Virol* **79**, 5288-5295
13. Sola, I., Almazan, F., Zuniga, S., and Enjuanes, L. (2015) Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* **2**, 265-288
14. Ziebuhr, J. (2005) The coronavirus replicase. *Curr Top Microbiol Immunol* **287**, 57-94
15. Hyde, J. L., and Diamond, M. S. (2015) Innate immune restriction and antagonism of viral RNA lacking 2-O methylation. *Virology* **479-480**, 66-74
16. Bouvet, M., Ferron, F., Imbert, I., Gluais, L., Selisko, B., Coutard, B., Canard, B., and Decroly, E. (2012) [Capping strategies in RNA viruses]. *Med Sci (Paris)* **28**, 423-429
17. Minskaia, E., Hertzig, 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
18. von Grotthuss, M., Wyrwicz, L. S., and Rychlewski, L. (2003) mRNA cap-1 methyltransferase in the SARS genome. *Cell* **113**, 701-702
19. Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., Gluais, L., Papageorgiou, N., Sharff, A., Bricogne, G., Ortiz-Lombardia, M., Lescar, J., and Canard, B. (2011) Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog* **7**, e1002059
20. Chen, Y., and Guo, D. (2016) Molecular mechanisms of coronavirus RNA capping and methylation. *Virol Sin* **31**, 3-11
21. Pan, J., Peng, X., Gao, Y., Li, Z., Lu, X., Chen, Y., Ishaq, M., Liu, D., Dediego, M. L., Enjuanes, L., and Guo, D. (2008) Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV replication. *PLoS One* **3**, e3299
22. Bollati, M., Milani, M., Mastrangelo, E., Ricagno, S., Tedeschi, G., Nonnis, S., Decroly, E., Selisko, B., de Lamballerie, X., Coutard, B., Canard, B., and Bolognesi, M. (2009) Recognition of RNA cap in the Wesselsbron virus NS5 methyltransferase domain: implications for RNA-capping mechanisms in Flavivirus. *J Mol Biol* **385**, 140-152
23. Debarnot, C., Imbert, I., Ferron, F., Gluais, L., Varlet, I., Papageorgiou, N., Bouvet, M., Lescar, J., Decroly, E., and Canard, B. (2011) Crystallization and diffraction analysis of the SARS coronavirus nsp10-nsp16 complex. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**, 404-408
24. Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G., and Jiang, T. (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* **27**, 325-328
25. (!!! INVALID CITATION !!! (24)).
26. Konecny, R., Baker, N. A., and McCammon, J. A. (2012) iAPBS: a programming interface to Adaptive Poisson-Boltzmann Solver (APBS). *Comput Sci Discov* **5**

27. Matthes, N., Mesters, J. R., Coutard, B., Canard, B., Snijder, E. J., Moll, R., and Hilgenfeld, R. (2006) The non-structural protein Nsp10 of mouse hepatitis virus binds zinc ions and nucleic acids. *FEBS Lett* **580**, 4143-4149
28. Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* **28**, 329-335
29. Wei, S. M., Yang, L., Ke, Z.H., Liu, Q.Y., Che, Y., Yang, Z.Z., Guo, D.Y., Fan, C.P. (2017) Crystal structure of MERS-CoV nsp10/nsp16 complex bound to SAM.
30. Chouhan, B. P. S., Maimaiti, S., Gade, M., and Laurino, P. (2019) Rossmann-Fold Methyltransferases: Taking a "beta-Turn" around Their Cofactor, S-Adenosylmethionine. *Biochemistry* **58**, 166-170
31. Eschenfeldt, W. H., Makowska-Grzyska, M., Stols, L., Donnelly, M. I., Jedrzejczak, R., and Joachimiak, A. (2013) New LIC vectors for production of proteins from genes containing rare codons. *J Struct Funct Genomics* **14**, 135-144
32. Kwon, K., and Peterson, S. N. (2014) High-Throughput Cloning for Biophysical Applications. in *Structural Genomics and Drug Discovery: Methods and Protocols* (Anderson, W. F. ed.), Springer New York, New York, NY. pp 61-74
33. Millard, C. S., Stols, L., Quartey, P., Kim, Y., Dementieva, I., and Donnelly, M. I. (2003) A less laborious approach to the high-throughput production of recombinant proteins in Escherichia coli using 2-liter plastic bottles. *Protein Expr Purif* **29**, 311-320
34. Shuvalova, L. (2014) Parallel protein purification. in *Methods Mol Biol*, 2014/03/05 Ed. pp 137-143
35. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. *Acta Crystallographica Section D* **62**, 859-866
36. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
37. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
38. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367
39. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
40. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* **374**, 229-244
41. Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-450
42. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21
43. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **62**, 859-866
44. Lebedev, A. A., Vagin, A. A., and Murshudov, G. N. (2008) Model preparation in MOLREP and examples of model improvement using X-ray data. *Acta Crystallogr D Biol Crystallogr* **64**, 33-39

45. Adams PD, A. P., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. (2010) *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* **66**, 213-221
46. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324
47. Holm, L., and Laakso, L. M. (2016) Dali server update. *Nucleic Acids Res* **44**, 351-355
48. Li, Z., Natarajan, P., Ye, Y., Hrabe, T., and Godzik, A. (2014) POSA: a user-driven, interactive multiple protein structure alignment server. *Nucleic Acids Res* **42**, W240-245
49. Ye, Y., and Godzik, A. (2004) FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Res* **1**, 582-585
50. DeLano, W. L. (2002) Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallog*, 82-92

PDB Accession Code	6w4h	6w61
Crystallization Conditions		
Screen conditions	0.2M Calcium acetate, 0.1M HEPES pH 7.5, 18% (w/v) PEG8000	0.1 M sodium citrate, pH 5.6, 10% (w/v) PEG4000, 10%(w/v) isopropanol
Protein concentration (mg/ml)	5.3	16.7
Data Collection		
Space group	*$P3_121$*	*$P3_121$*
Unit cell parameters (Å; º)	*a* = 167.7, *b* = 167.7, *c* = 51.9; $\alpha = \beta$=90.0, γ = 120.0	*a = 168.6, b = 168.6, c = 52.0;* $\alpha = \beta$ *=90.0,* γ *= 120.0*
Resolution range (Å)	30.00 – 1.80 (1.83 - 1.80)	50.00 - 2.00 (2.04 - 2.00)
No. of reflections	77,886 (3,873)	57,010 (2,781)
R_{merge} (%)	6.0 (76.5)	13.9 (233.0)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
$\langle I/\sigma(I)\rangle$	29.3 (2.6)	37.1 (1.0)
Multiplicity	7.3 (6.8)	6.2 (6.3)
Wilson *B* factor	25.4	49.0
Structure Determination		
MR initial model (PDB ID)	3r24	6w4h
Refinement		
Resolution range (Å)	29.76 - 1.80 (1.85 - 1.80)	42.33 - 2.00 (2.04 - 2.00)
Completeness (%)	100.0 (100.0)	99.7 (96.3)
No. of reflections	73,752 (5,432)	56,859 (2,522)
R_{work}/R_{free}, (%)	14.9/16.3 (22.9/23.9)	17.2/19.3 (30.4/32.3)
Protein chains/atoms	2/3,201	2/3,390
Ligand/Solvent atoms	65/442	6/141
Mean temperature factor (Å^2)	31.5	41.0
Coordinate Deviations		
R.m.s.d. bonds (Å)	0.004	0.002
R.m.s.d. angles (º)	1.183	0.473
Ramachandran plot		
Favored (%)	97.0	96.8
Allowed (%)	3.0	3.2
Outside allowed (%)	0.0	0.0

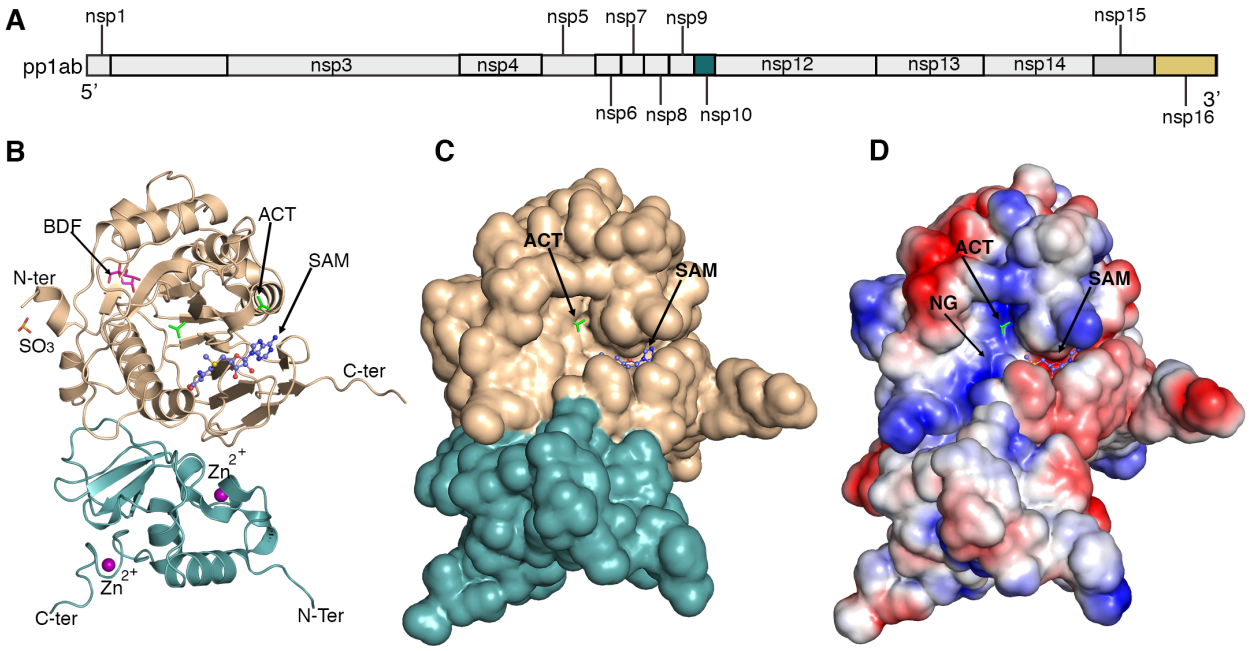

Figure 1. Overall structure of the nsp10-nsp16. A, Linear schematic of the orf1ab translation produce prior to nsp3 and nsp5 processing. B, Cartoon representation of the heterodimer of nsp16 (tan) in complex with nsp10 (teal). Ligands are sulfite (SO_3), β-D-fructopyranose, (BDF), acetate, (ACT), S-adenosyl-methionine (SAM), and Zn^{2+}. C, Solvent exposed surface, showing the SAM pocket and ACT. D, Surface charge showing the positively charged surface at the groove for nucleotide (NG) binding site and the negatively charged SAM binding pocket (C).

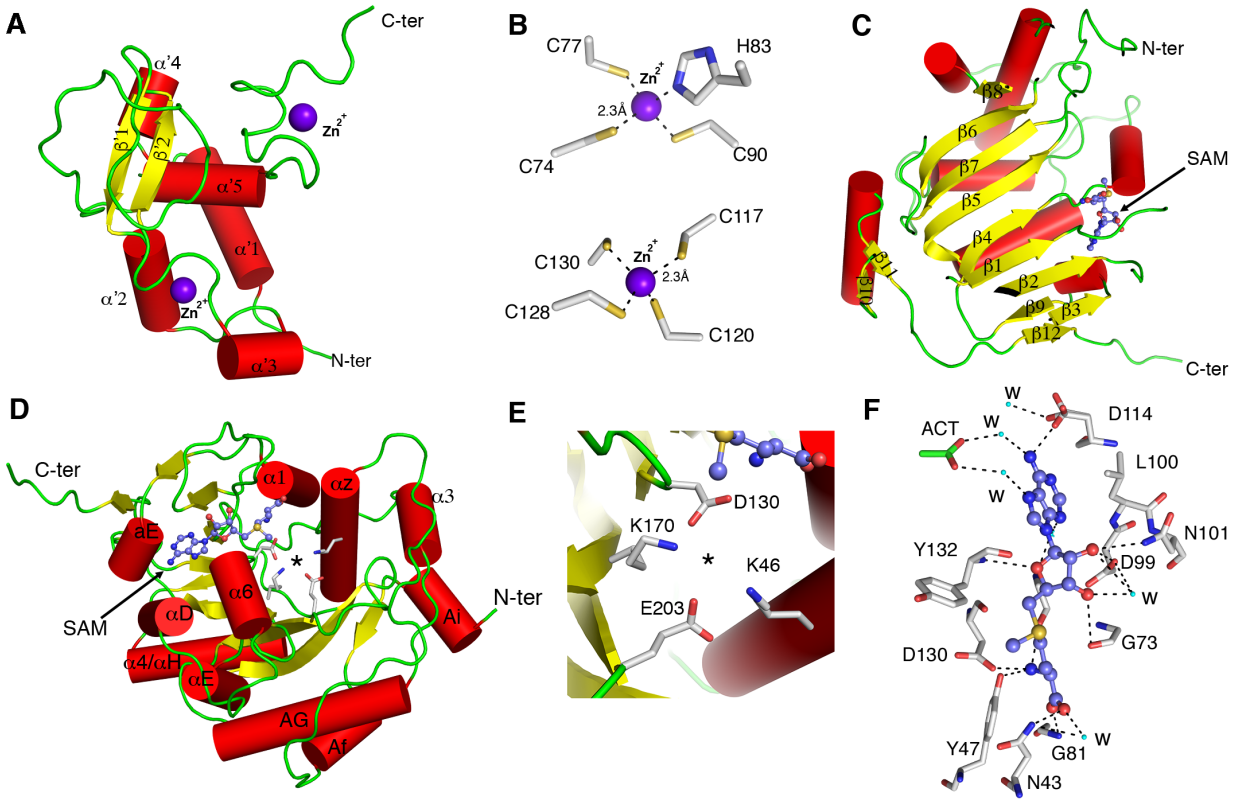

Figure 2. Detailed representation of nsp10 and nsp16 structures and ligands. A, representation of the structure of nsp10 with 5 helices (red) and 2 beta-strands (yellow) and two zinc ions (purple). B, Close-up view of the two Zn^{2+} binding sites in nsp10. C, canonical beta sheet from nsp16 bound with SAM (ball and stick). D, nsp16 helices, SAM (ball and stick) and catalytic residues (gray sticks). Asterisk indicates the catalytic methyl transferase residues. E, Close-up view of the catalytic residues in nsp16. F, Detailed view of coordination of SAM binding. Small cyan dots indicate water molecules (w) and dashed lines the hydrogen bonds. Acetate (ACT) is in green sticks.

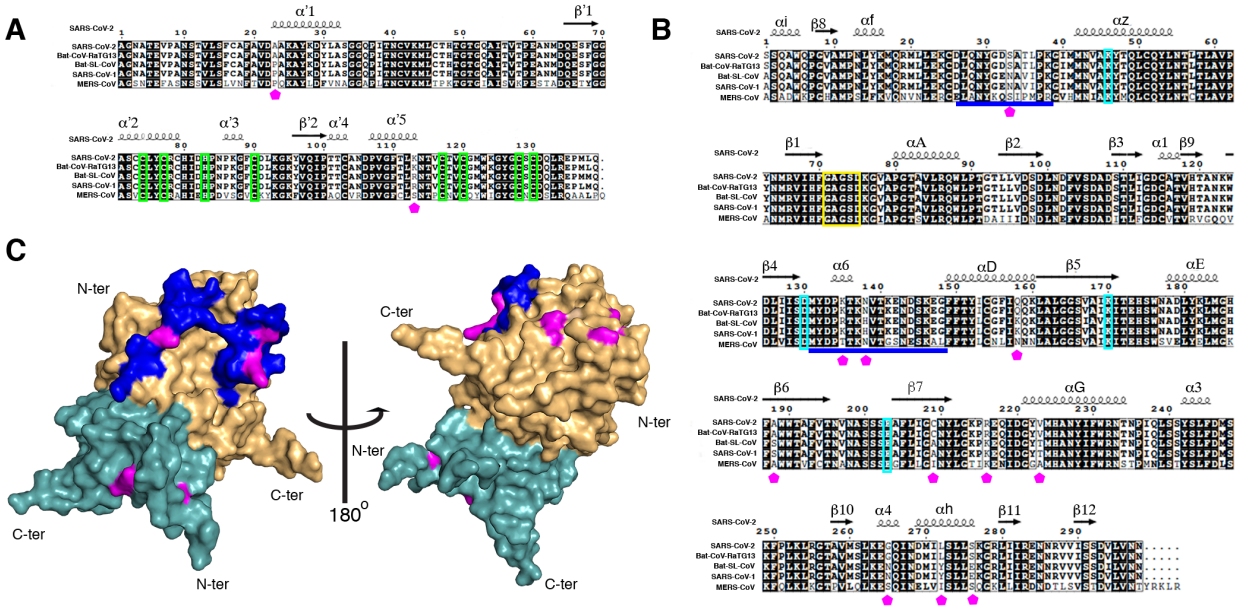

Figure 3. Sequence alignment of SARS-CoV-2 with other coronavirus and the mutations mapped in the structure. Multiple sequence alignment of nsp10 (A) and nsp16 (B). Black shading represents 100% residue identity and the punctual mutations are indicated with the magenta pentagon. In nsp10, the green boxes indicate the Zn binding site. In nsp16, the flexible loops involved in SAM binding stability and differ or are absent in SARS-CoV-1, Bat-CoV-RaTG13, Bat-LS-CoV Rs4247 and/or MERS-CoV are underlined in blue, and the cyan boxes indicate the catalytic KDKE residues. C, Cartoon representation of the structure of SARS-CoV-2, with nsp10 subunit in teal and nsp16 in tan. The single amino acid changes between SARS-CoV-1 and SARS-CoV-2, and the residues comprising nsp16 flexible loops are mapped on the surface projection, as in (A) and (B).