



Published in final edited form as:

Sci Signal. ; 13(651): . doi:10.1126/scisignal.abe1202.

High-resolution structures of the SARS-CoV-2 2'-O-methyltransferase reveal strategies for structure-based inhibitor design

Monica Rosas-Lemus^{1,2,†}, George Minasov^{1,2,†}, Ludmilla Shuvalova^{1,2,†}, Nicole Inniss^{1,2}, Olga Kiryukhina^{1,2}, Joseph Brunzelle³, Karla J. F. Satchell^{1,2,*}

¹Department of Microbiology-Immunology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

²Center for Structural Genomics of Infectious Diseases, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

³Northwestern Synchrotron Research Center, Life Sciences Collaborative Access Team, Northwestern University, Argonne, IL 60439, USA.

Abstract

There are currently no antiviral therapies specific for SARS-CoV-2, the virus responsible for the global pandemic disease COVID-19. To facilitate structure-based drug design, we conducted an X-ray crystallographic study of the SARS-CoV-2 nsp16-nsp10 2'-O-methyltransferase complex, which methylates Cap-0 viral mRNAs to improve viral protein translation and to avoid host immune detection. We determined the structures for nsp16-nsp10 heterodimers bound to the methyl donor *S*-adenosylmethionine (SAM), the reaction product *S*-adenosylhomocysteine (SAH), or the SAH analog sinefungin (SFG). We also solved structures for nsp16-nsp10 in complex with the methylated Cap-0 analog m⁷GpppA and either SAM or SAH. Comparative analyses between these structures and published structures for nsp16 from other betacoronaviruses revealed flexible loops in open and closed conformations at the m⁷GpppA-binding pocket. Bound sulfates in several of the structures suggested the location of the ribonucleic acid backbone phosphates in the

This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

*Corresponding author, k-satchell@northwestern.edu.

†These authors contributed equally to this work

Author contributions: M.R.-L. and O.K. purified proteins and formed the complex, L.S. conducted crystallization and soaking experiments and G.M. and J.B. collected crystallographic data, determined and analyzed structures. M.R.-L. and N.L.I. wrote the draft of the manuscript, which was edited by all authors. K.J.F.S. supervised all aspects of the project.

Competing interests: K.J.F.S. has a significant financial interest in Situ Biosciences, LLC, a contract research organization that conducts antimicrobial testing for industrial products including antiviral testing. This work has no overlap with the interests of the company. K.J.F.S. is a consultant for a healthcare firm on public health topics related to SARS-CoV-2 and COVID-19 that are unrelated to this article. The other authors declare that they have no competing interests.

Data and materials availability: All data are publicly available in the RCSB Protein Data Bank (www.rcsb.org) with the PDB codes 6W4H, 6W75, 6WJT, 6WQK, 6WQ3, 6WVN, and 6WRZ. Plasmids have been deposited and are available from www.BEIresources.com. Raw x-ray diffraction data for small unit cell and large unit cell crystals are deposited at protein_diffraction.org. All other data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials.

ribonucleotide-binding groove. Additional nucleotide-binding sites were found on the face of the protein opposite the active site. These various sites and the conserved dimer interface could be exploited for the development of antiviral inhibitors.

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. 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). Within four months, SARS-CoV-2 rapidly spread, sparking a global pandemic. The COVID-19 pandemic has also forced governments to enact “stay-at-home” orders around the world, seriously damaging the global economy (1). According to the World Health Organization, nearly 25 million SARS-CoV-2 infections have been confirmed, of which more than 800,000 were fatal as of Aug 30, 2020 (www.who.int). These data are similar to those from the Johns Hopkins University tracking system (2).

Members of the coronaviridae family of viruses infect birds and mammals, including bats, camels, pigs, and humans. In humans, pathogenic coronaviruses cause acute and severe gastrointestinal infections, fevers, and organ failure. Three of the seven human-tropic coronaviruses – hCoV-229E, hCoV-NL63, and hCoV-OC43 – cause only asymptomatic or mild infections, including the common cold (3). Four other human coronaviruses are linked to severe infections, including hCoV-HKU1, a common cause of pneumonia; SARS-CoV, with a 10% mortality rate; Middle East Respiratory Syndrome Virus (MERS-CoV); with a 37% mortality rate; and SARS-CoV-2, currently with ~3% mortality rate among confirmed cases (3, 4). Among them, SARS-CoV-2 stands as the one with highest transmissibility, making its containment very difficult (5). As SARS-CoV-2 continues to spread, the need for effective vaccines and therapeutics increases. 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, and contains a 3′-poly-A tail. The *orf1a* and *orf1ab* coding regions are directly translated, whereas the rest of the genome serves as template to generate sub-genomic messenger RNAs (mRNAs) transcribed from the 3′-end, which are later capped and translated (6–8). The first open reading frame (*orf1a*) produces the large non-structural polyprotein 1a (pp1a), and a programmed -1 ribosomal frameshift results in translation of the larger non-structural polyprotein 1ab (pp1ab) from the reading frame *orf1ab*. These polyproteins are subsequently processed into sixteen non-structural proteins (nsp1–nsp16) that assemble to form the replication-transcription complex (RTC) or function as accessory proteins necessary for viral replication (8, 9).

The components of the RTC include enzymes that regulate mRNA and genomic RNA synthesis, proofreading, and mRNA maturation. Two of these enzymes, nsp14 and nsp16, are critical for capping viral mRNAs, a tactic employed by multiple RNA viruses to avoid

immune detection (10). 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, thus forming guanidine monophosphate (GMP). This is followed by the transfer of the α -phosphate of GMP to the diphosphate 5'-ppN transcript end, thus forming the cap core structure, methylguanine-triphosphate-ribonucleotide, referred to as GpppN. GpppN formation is followed by N⁷-methylation of the capping guanylate by a guanine-N⁷-methyltransferase (N⁷-MTase) to irreversibly generate the methylated Cap-0. Further methylation at the ribose 2'-O position of the first nucleotide of the RNA is catalyzed by a ribose 2'-O-methyltransferase (2'-O-MTase) to generate Cap-1 and sometimes at the second nucleotide to generate Cap-2. Both the N⁷-MTase and 2'-O-MTase use S-adenosyl-L-methionine (SAM) as the methyl group donor (4, 10).

For coronavirus mRNA maturation, the TPase activity is mediated by nsp13 (6, 11–13), and a still-elusive GTase guanylates the 5'-end of the nascent mRNA. The viral non-structural protein 14 (nsp14), which has N⁷-MTase activity, then generates the Cap-0 (14). Nsp14 is a bifunctional enzyme with independent N⁷-MTase and exonuclease domains (15). The association of nsp14 with viral non-structural protein 10 (nsp10) specifically stimulates nsp14 exonuclease activity but has no effect on the N⁷-MTase activity (16). The coronavirus mRNAs are further modified to have a Cap-1 by the viral 2'-O-methyltransferase (nsp16). Nsp16 is a 7-methylguanine-triphosphate-adenosine (m⁷GpppA)-specific, SAM-dependent 2'-O-MTase (17, 18) that is activated upon binding to nsp10 (16, 19). Nsp10 is a stable monomeric protein that can also form dodecamers (20, 21) in addition to binding to nsp14 and nsp16 (16, 22). Although no specific enzymatic activity has been identified for nsp10, it is known that nsp10 is a zinc-binding protein and can bind RNA (20, 23, 24). It has also been found that nsp10 interacts with human adaptor protein complex 2 when expressed in mammalian cells (25). However, the main known function of nsp10 is the stabilization of the SAM binding pockets of nsp16 and nsp14 (19). The 2'-O-methylation of coronavirus RNA that is mediated by the nsp16-nsp10 heterodimer is essential for preventing recognition by the host to evade immune responses that are triggered by viral mRNAs (17).

Structures of the nsp16-nsp10 complex have been determined for SARS-CoV and MERS-CoV (18, 23, 26, 27), and the analyses elucidated the structural basis for substrate binding and the proposed S_N2-mechanism of methyl transfer. In order to facilitate structure-based inhibitor design, we initiated a project to determine the structures of the 2'-O-MTase from SARS-CoV-2 in complex with its ligands. Herein we present a comprehensive X-ray crystallographic study of the structure of the SARS-CoV-2 nsp16-nsp10 heterodimer. The structures of the heterodimer were determined in complex with the methyl donor SAM, the product of the reaction (S-adenosylhomocysteine, SAH), and pan-methyltransferase inhibitor sinefungin (SFG). In addition, we describe the first publicly deposited SARS structures of nsp16-nsp10 in complex with the mRNA cap m⁷GpppA, which facilitates detailed analysis of the changes in the conformation of flexible loops of nsp16 upon substrate binding. Furthermore, we report crystal structures with sulfate ions in the proposed RNA binding groove as well as several additional nucleotide and sugar binding sites outside the active site. Because nsp16 is one of the most conserved proteins of SARS-CoV-2 and

related viruses, these high-resolution structures are expected to be useful as models for developing new antiviral therapeutics to treat COVID-19 and other diseases caused by coronaviruses.

Results

High-resolution structures of the SARS-CoV-2 2'-O-MTase heterodimer in two crystal forms

The SARS-CoV-2 proteins nsp10 and nsp16 are encoded by the polycistronic *orf1ab* of the (+) ssRNA (Fig. 1A) and are released from the polyproteins pp1a and pp1ab by the protease nsp5 (28). The protein nsp10 is 14.8 kDa protein released from both polyproteins pp1a and pp1ab, whereas nsp16 is a 33.3 kDa protein released only from the pp1ab polyprotein, which is created by a -1 ribosome shift (28) (Fig. 1A). We successfully expressed and purified recombinant nsp16 and nsp10 separately, combined them 1:1 in the presence of SAM to form nsp16-nsp10 complexes that were set up for crystallization, and obtained diffraction-quality crystals under several conditions (Table S1).

The first structure of the nsp16-nsp10 complex from SARS-CoV-2 was solved at 1.8 Å [RCSB Protein Data Bank (PDB) code 6W4H, Table S1, and Fig. 1B]. The crystal belonged to the space group P3₁21 with two polypeptide chains in the asymmetric unit, with chain A (nsp16) and chain B (nsp10) forming a heterodimer. We refer to this crystal form as the “small unit cell”. The heterodimer had a total solvent-exposed surface area of 19,710 Å² and a buried area of 2870 Å² estimated by the Protein, Interfaces, Structures and Assemblies tool (PISA), and it is mainly stabilized by hydrophobic interactions and hydrogen bonds at the interface of nsp16 and nsp10. In this structure, the methyl donor SAM was bound to nsp16, and Zn²⁺ was bound to nsp10 (Fig. 1B).

The second crystal form of the nsp16-nsp10 complex with SAM yielded a structure solved at 1.95 Å (PDB code 6W75, Table S2). This crystal form belonged to the P3₂21 space group and had four chains in the asymmetric unit. The four chains were arranged as a dimer of dimers with a butterfly-like shape (Fig. 1C). The two heterodimers interacted by the C-terminus of nsp16 and as well as the N-terminus of nsp10. We refer to this crystal form as the “large unit cell”.

The overall structure of nsp16-nsp10 in the large unit cell was almost identical to the small unit cell structure, including the bound ligands SAM and Zn²⁺. In order to corroborate the degree of structural identity, the chains of both crystal forms were aligned using the FATCAT server (29). Alignment of nsp16 from the large unit cell (chains A and C) with that from the small unit cell (chain A) showed substantial similarity with a raw root-mean-square deviation (r.m.s.d.) of 0.37 Å and 0.42 Å, respectively. The introduction of a flexibility factor in the alignment showed an optimized r.m.s.d. of 0.38 Å for chain A and 0.77 Å for chain C, demonstrating that the nsp16 structures in these two crystal forms are similar but have flexible regions. One of these flexible regions was a part of the loop formed by the residues Asp⁶⁹³¹-Phe⁶⁹⁴⁷, which was disordered from the residues Lys⁶⁹³³ to Lys⁶⁹³⁹ in chain C, a likely explanation for the structural differences (movie S1). The nsp10 alignment had a raw r.m.s.d. of 0.23 Å for chain B and 0.34 Å for chain D and there were no gaps in the alignment. The optimized r.m.s.d. was 0.28 Å and 0.35 Å, respectively, indicating very low

differences between these chains. Thus, both crystallographic forms were almost identical with small discrepancies caused by different conformations in the flexible loops of nsp16.

To determine which stoichiometry existed in solution, we performed analytical size-exclusion chromatography (SEC). In the elution profile, we observed a prominent elution peak at 15 ml that corresponded to a molecular weight (m.w.) of 45 kDa, which is close to the estimated m.w. of the heterodimer (49.8 kDa). We also observed a small peak at ~17.5 ml, containing mostly nsp10 (Fig. 1, D and E). No peak was detected corresponding to ~90 kDa, which would be consistent with four chains in a complex in solution. Thus, the heterodimer is the most soluble and stable form of the 2-*O*-MTase complex, and the dimer of dimers in the large unit cell formed as a result of crystal packing.

Topology of nsp16 and nsp10 and the heterodimer interface

The nsp16 protein consisted of pp1ab residues 6799–7096 plus three additional residues (Ser-Asp-Ala) at the N-terminus derived from the recombinant expression tag after tobacco etch virus (TEV) (20) protease cleavage. The 2'-*O*-MTase catalytic core was comprised of a Rossmann-like β -sheet fold with the canonical 3-2-1-4-5-7-6 arrangement, in which β 7 was the only antiparallel strand (Fig. 2A). This β -sheet was sandwiched by eleven α -helices and 20 loops (Fig. 2B).

The nsp10 protein, consisting of pp1a residues 4272–4392, has at its core three β -strands (β '1, β '2, β '3) that form a central anti-parallel β -sheet. At one side of the β -sheet is the large loop that directly interacts with nsp16 and stabilizes the heterodimer complex. At the other side of this β -sheet there are six helices and loops that form two zinc fingers (Fig. 2C). In other coronaviruses, these zinc fingers are involved in non-specific binding of RNA (23, 24). The Zn²⁺-binding site 1 is coordinated by the residues Cys⁴³²⁷, Cys⁴³³⁰, Cys⁴³³⁶, and His⁴³⁴³. The Zn²⁺-binding site 2 is coordinated by Cys⁴³⁷⁰, Cys⁴³⁷³, Cys⁴³⁸¹, and Cys⁴³⁸³ (Fig. 2D).

The residues forming the nsp16-nsp10 heterodimer interface can be divided into clusters. The clusters for nsp16 are defined as A (residues 6835–6846), B (6874–6889), C (6900–6908) and D (7042–7046). For nsp10 they are defined as clusters I (4293–4300), II (4322–4337), and III (4346–4349) (18, 23). Almost all of the interface contacts between nsp16 and nsp10 are formed by hydrophobic interactions between cluster I (Val⁴²⁹⁵, Met⁴²⁹⁷, Leu⁴²⁹⁸) of nsp10 and cluster A (Pro⁶⁸³⁵, Ile⁶⁸³⁸, Met⁶⁸³⁹, Val⁶⁸⁴², Ala⁶⁸⁴³), cluster B (Val⁶⁸⁷⁶, Pro⁶⁸⁷⁸, Ala⁶⁸⁸¹), and cluster D (Leu⁷⁰⁴², Met⁷⁰⁴⁵) of nsp16 (Fig. 2E). We observed that the remaining interactions at the interface are mediated by hydrogen bonds, and these hydrophilic interactions consisted of five direct contacts between residues Lys⁴²⁹⁶, Leu⁴²⁹⁸, Ala⁴³²⁴, Tyr⁴³⁴⁹, and Gly⁴³⁴⁷ of nsp10 with Lys⁶⁸³⁶, Gln⁶⁸⁷⁵, Ala⁶⁸⁸¹, and Asp⁶⁹⁰⁴ of nsp16, respectively, plus eight water-mediated interactions (Fig. 2F).

The binding of SAM, SAH and the inhibitor SFG to the methyl donor binding site

The nsp16 protein catalyzes the transfer of the methyl group from SAM to the Cap-0, generating the reaction products SAH and Cap-1. This reaction can be inhibited by SFG, a 5'-aminoalkyl analog of SAH used as a pan-inhibitor of methyltransferases (Fig. 3A). In order to identify potential structural differences caused by having SAM, SAH, or SFG in the

SAM-binding cleft, we also determined the structures of nsp16-nsp10 in complex with SAH (PDB code 6WKQ) and SFG (PDB code 6WJT) at 2.0 and 1.98 Å resolution, respectively. These structures showed that SAM binds to a negatively charged cleft formed by α A, α Z, α D, and the loops L5, L8, L11 in nsp16 (Fig. 3B). The adenosine moiety is stabilized by residues Phe⁶⁹⁴⁷, Asp⁶⁹¹², Leu⁶⁸⁹⁸, Cys⁶⁹¹³, and Met⁶⁹²⁸. The sugar moiety is stabilized by the residues Gly⁶⁸⁷¹ and Asp⁶⁸⁹⁷ and by two molecules of water that interact with Asn⁶⁸⁹⁹. The methionine moiety interacts with Asp⁶⁹²⁸, Tyr⁶⁸⁴⁵, Asn⁶⁸⁴¹, and Gly⁶⁸⁷¹. Notably, SAH and SFG interact with the same residues as SAM without modifications in the site or the overall structure (Fig. 3C).

Comparison of nsp16-nsp10 from SARS-CoV-2 and SARS-CoV

At the primary amino acid sequence level, nsp16 from SARS-CoV-2 is 99% identical to Bat-CoV-RaTG13, 94% identical to Bat-SARS-like coronavirus Rs4247(Bat-SL-CoV), and 93% identical to SARS-CoV, but only 66% identical to MERS-CoV (Fig. S1A). The primary amino acid sequence of nsp10 is 100% identical to Bat-CoV-RaTG13, 99% identical to SARS-CoV, and 98% identical to Bat-SL-CoV, but only 59% identical to MERS-CoV (Fig. S1B). Thus, nsp16 and nsp10 are highly conserved in the lineage B betacoronaviruses (Bat-SL-CoV, Bat-CoV-RaTG13 and SARS-CoV), and less conserved with lineage C betacoronaviruses (MERS-CoV). We compared our structures from SARS-CoV-2 with published structures from SARS-CoV (18, 23) to determine if minor differences in sequence impacted the structures of nsp16 or nsp10. The two amino acid differences in nsp10 between SARS-CoV-2 and SARS-CoV are A4276P and K4366R and do not introduce important structural changes (Fig. S1C). Furthermore, in the aligned sequences, the Zn²⁺-coordinating residues in nsp10 are 100% conserved, emphasizing the importance of the zinc fingers across the coronaviruses (Fig. S1B).

Another difference was identified when our nsp16 structure was aligned with the SARS-CoV nsp16 structure published by Chen *et al.* (23), which has three gaps and an optimized r.m.s.d of 1.19 Å. The differences were localized to loop 1 and helix η 3 of SARS-CoV-2 (Fig. S1B). The loop 1 sequence that starts at residue 6829 is DSATL in SARS-CoV-2, compared to ENAVI in SARS-CoV. This loop is flexible and is in a “closed” conformation in SARS-CoV compared with a more “open” conformation for SARS-CoV-2. In addition, η 3 in SARS-CoV-2 is modeled as a loop in SARS-CoV, which may be due to differences in the primary sequence of PKTKN compared to PRTKH. All other amino acid differences were solvent-exposed (Fig. S1C), and none of them affected the SAM binding site. Further, the nsp16-nsp10 interface residues identified for SARS-CoV-2 (Fig. 2E,F) are 100% identical to this interface in SARS-CoV (Fig. S1A,B), indicating that the interface is conserved between the two viruses.

Despite the sequence divergence, the overall structure of the SARS-CoV-2 2'-O-MTase is very similar to structures available for the MERS-CoV 2'-O-MTase (27), with no substantial changes in nsp10 (r.m.s.d. = 0.78 Å). The MERS-CoV nsp16 structure is also very conserved, but the loops L1 and L8-L9 are disordered, which causes an increase in the r.m.s.d. of 0.69 Å. One difference between the SARS-CoV-2 and MERS-CoV 2'-O-MTases exists at the nsp16-nsp10 interface. Tyr⁴³⁴⁹ in nsp10 contacts Ala⁶⁸⁸¹ in nsp16, two residues

that are conserved in SARS-CoV-2 and all other betacoronaviruses. However, in MERS-CoV the corresponding residues are Phe and Ser, respectively (Fig. S1A,B). This Phe substitution in MERS-CoV nsp10 promotes the binding of nsp10 to nsp16 and stimulates the MTase activity of nsp16 (30). The conservation of Tyr at this position in SARS-CoV-2 and SARS-CoV could indicate a positive selection against overstimulation of the MTase in the lineage B betacoronaviruses.

Binding of SAM and m⁷GpppA to the nsp16 catalytic site

In addition to the methyl donor SAM, the nsp16 catalytic reaction requires a Cap-0-mRNA with an adenosine in position 1 of the single-stranded RNA (m⁷GpppA-RNA). The crystal structure of the ternary complex of nsp16-nsp10 with m⁷GpppA and SAM has been determined for the more distantly related lineage C betacoronavirus MERS-CoV (PDB code 5YNM (27)), but not for the more closely related lineage B betacoronavirus SARS-CoV, in which the binding of the mRNA was previously only modeled (23). Herein, we describe structures of the SARS-CoV-2 nsp16-nsp10 heterodimer in complex with m⁷GpppA and either the methyl donor SAM (PDB code 6WVN) or the product SAH (PDB code 6WQ3 and 6WRZ). Overall, the three were almost identical between the SAM- and SAH-bound structures (r.m.s.d. of 0.12 Å) except for sulfates bound into the RNA-binding groove at different positions in one of the structures with SAH. One of structures with cap and SAM (PDB code 6WVN) was selected for more detailed analysis because it was the most complete, with eight additional residues at the N-terminus of nsp10 compared to the SAM-bound structure (PDB code 6W4H) and is at slightly better resolution (Table S2). The N-terminus in this structure forms an α -helix that seems to be in a more open conformation with respect to the previous structure from MERS (r.m.s.d.=0.84 Å) (27).

The cap analog m⁷GpppA bound to the cap high-affinity binding site (HBS), which is a positively charged surface on nsp16 formed by the loops L1, L8, L9, L10, and L12, α D, and the η 3 (Fig. 3D). The guanosine ring of m⁷GpppA is stacked with Tyr⁶⁸²⁸. The phosphate groups were mostly stabilized by side chain atoms of Tyr⁶⁸²⁸, Tyr⁶⁹³⁰, Lys⁶⁹³⁵, Thr⁶⁹⁷⁰, Ser⁶⁹⁹⁹, and Ser⁷⁰⁰⁰, and by the main chain atoms of His⁶⁹⁷² and Ser⁷⁰⁰⁰ of loops 10 (residues 6970–6975) and 12 (residues 6994–6997) (Fig.3E). The adenosine sugar interacted with side chain atoms of Lys⁶⁸⁴⁴, Lys⁶⁹⁶⁸ and with Asp⁶⁹²⁸ through a water molecule. The adenine moiety was stabilized by stacking interaction with the side chain of Tyr⁶⁹³⁰ (Fig. 3F), and it was in close proximity to the SAM binding cleft. These interactions are also found in the structure of MERS-CoV nsp16 in complex with Cap-0 (PDB code 5YNM (27)).

The high quality of the crystal structure of the nsp16-nsp10 complex with m⁷GpppA and SAM bound facilitated detailed analysis of the catalytic site. The protein nsp16 contains the highly conserved residues Lys⁶⁸³⁹, Asp⁶⁹²⁸, Lys⁶⁹⁶⁸, and Glu⁷⁰⁰¹, comprising the canonical catalytic motif (K-D-K-E) conserved among class I MTases (18, 23, 31). These residues are close to the SAM methyl group that is transferred to the 2'-OH on the m⁷GpppA. The amino group of Lys⁶⁹⁶⁸ is in close interaction with the 2'-OH of Cap-0 and possibly activates this oxygen for the nucleophilic attack of the methyl group in SAM (Fig. 3F). In the structure of nsp16-nsp10 with the m⁷GpppA and SAM, we detected the presence of one molecule of water (Fig. 3F) that might participate in the stabilization of intermediate catalytic states (18,

23, 32). Although the nsp16 MTase reaction was previously characterized as Mg^{2+} -dependent, we did not observe this metal bound, and it is likely that Mg^{2+} is involved only in transitory states of catalysis or stability of the protein as previously suggested for dengue virus 2'-*O*-MTase (33).

The flexibility of the m⁷GpppA site in nsp16

Previous biochemical studies of SARS-CoV nsp16 demonstrated stabilization of the SAM cleft upon nsp10 binding (18, 23) and improved m⁷GpppA-RNA interaction (23), although there is as yet no structure with cap bound to SARS-CoV nsp16 to identify the conformational changes necessary for cap binding. The HBS is surrounded by Loop 1 (residues 6824–6834) and a loop formed by L8-η-3-L9 (residues 6930–6943) in nsp16. These flexible loops, which were not visible in the previously published structure for MERS, are visible in our structures and thus we could analyze the diverse conformations of nsp16-nsp10 complexes.

First, we compared the structures of the heterodimer with only SAM bound from the large unit cell crystal form (Fig. 4A, green) with to that from the small unit cell crystal form (Fig. 4A, blue). There are only minor conformational differences for loop 1 of nsp16, however, the L8-η-3-L9 loop showed a more “open” conformation in the large unit cell and a more “closed” conformation for the small unit cell structure when only SAM was bound. This analysis corroborated that this specific region is flexible and that these loops likely transit from an “open” to a “closed” state in the absence of m⁷GpppA.

In order to evaluate the position of the loops when m⁷GpppA is bound, we analyzed the structural alignment between the high-resolution heterodimer with SAM bound (Fig. 4A, blue) and the heterodimer in complex with m⁷GpppA and SAM (Fig. 4A, orange). This alignment has an r.m.s.d. of 0.55 Å, and we observed that the presence of m⁷GpppA induces a stable open conformation of L8-η-3-L9 (residues 6930–6943), which was found also in the heterodimer structures in complex with m⁷GpppA and SAH (r.m.s.d.=0.12 Å, PDB code 6WQ3 and 6WRZ). The main conformational changes were observed near residues 6936–6939, which were displaced in the open conformation and appeared to mitigate clashes between Lys⁶⁹³⁵ of nsp16 with the ribose and the first phosphate group of the m⁷GpppA and at residue Tyr⁶⁹³⁰ that is rotated to improve stacking interaction with the adenine of the Cap-0 (Fig. S2). These changes indicated that the presence of m⁷GpppA stabilized an open state of the Cap-0 binding site, which could facilitate the release of the product upon methylation.

Furthermore, the C-α chain of nsp16 from SARS-CoV-2 with Cap-0 and SAM bound (Fig. 4B, orange) was aligned with MERS-CoV nsp16 with SAM (Fig. 4B, violet) or with Cap-0 and SAM bound (Fig. 4B, cyan) over 293 residues. In the absence of the Cap-0, the MERS-CoV nsp16 residues corresponding to residues 6934–6940 in SARS-CoV-2 were disordered, confirming that this loop is flexible. In contrast, Loop 1 and residues 6930–6943 from the MERS-CoV m⁷GpppA-nsp16 complex aligns with the open conformation observed in the SARS-CoV-2 m⁷GpppA-nsp16 complex.

Sulfates align to the RNA binding groove

The 2'-*O*-MTase nsp16-nsp10 possibly binds the viral RNA in the positively charged nucleotide binding groove, also known as the low affinity binding site (LBS). There is, as yet, no structural evidence of the arrangement of this part of the viral RNA in the protein and only predictive models have thus far been published (23, 34, 35). Sulfates are known to bind to proteins in the same positions as nucleic acid backbone phosphates and thus can be used to model nucleic acid binding regions. Because the first small unit cell structure (PDB code 6W4H) was obtained from crystallization conditions with polyethylene glycol (PEG), which did not allow for soaking with substrates, crystals were screened for suitable conditions for soaking with substrates followed by cryoprotection with 2 M lithium sulfate (see Materials and Methods). We speculated that these sulfate molecules could indicate the possible binding sites for the phosphates of the RNA molecule. All of the structures with m⁷GpppA also had molecules of sulfates bound at distinct sites and could be superimposed to analyze the position of sulfates (Fig. 5A–C). Sulfate 1 (S1) was in a position next to the SAM cleft in two alternative conformations, which could indicate the importance of charged molecules near the catalytic site (Fig. 5B). Sulfate 2 (S2) seems to mimic the phosphate group between the first and the second nucleotide in the RNA and is followed by a zig-zag line of other sulfates (S2-S5) along the positively charged LBS from nsp16 and the extension groove of nsp10. The compilation of these results suggested that the nucleotide binding groove might accommodate four to five nucleotides from the viral m⁷GpppA-RNA (Fig. 5B). This experimental and structural evidence reveals the possible position of the viral RNA in the nucleotide-binding groove.

Identification of three additional m⁷GpppA binding sites in nsp16

The present comprehensive study of the interaction of m⁷GpppA with nsp16-nsp10 resulted in the unexpected finding of nucleotides in non-catalytic sites of the structures (Fig. 5A and 5C). Although, relatively short soaking times were tested to avoid non-specific binding of m⁷GpppA, nucleotides were consistently found at three different positions additional to the active site. One of the sites showed binding of the guanine and phosphate moiety from m⁷GpppA (designated MGP, Fig. 5D) and is located on face of the protein opposite the active site (Fig. 5C). This site was found occupied by a guanine moiety in all three structures with m⁷GpppA. The guanine moiety of m⁷GpppA interacts with the hydrophobic surface formed by Trp⁶⁹⁸⁷ and with the NH₂ in the position 2 stabilized by Ser⁷⁰⁷⁴ in a small negatively charged cavity (Fig. 5E). The adenosine moiety of the m⁷GpppA ligand was disordered. In the structure without the cap (PDB code 6W4H), this same site was occupied by β-D-fructopyranose (BDF) (Fig. 1B and 5E), indicating that this site is not nucleotide-specific.

Another binding site (ADE1) was occupied by an adenine (ADE) moiety that likewise is derived from the m⁷GpppA (Fig. 5C and 5D). The ADE stacked with Trp⁶⁸⁰³ (Fig. 5F). The third binding site (ADE2) is also occupied by an adenine ring stacking with Tyr⁷⁰²⁰ (Fig. 5A and 5G). Although these non-catalytic nucleotide binding events might be crystallization artefacts, they may represent interactions that would mimic the interaction between nsp16 and the ribonucleotides of the capped mRNA.

Discussion

The SARS-CoV-2 pandemic has yielded an urgent world-wide effort to understand the molecular mechanisms involved in coronavirus transmission, virulence, and replication (1). The ultimate goal is to identify viral proteins that are amenable to drug targeting and epitopes suitable for vaccine development. Although previous studies conducted in the related betacoronaviruses SARS-CoV and MERS-CoV paved the way for drug discovery and vaccine development, no approved treatments were fully developed (36). Thus, in order to ensure an accurate approach for drug discovery, we present a comprehensive study of the first structures of the SARS-CoV-2 2'-*O*-MTase complex that were publicly available to the scientific community. In addition to the structures reported here, similar structures of the nsp16-nsp10 complex have subsequently been determined by other groups; including structures with SAM [PDB codes 6W61 (37), 7BQ7 (38), 7C2I, and 7C2J (39)] and with SFG [PDB code 6YZ1, (34)]. Independently, another structure of nsp16-nsp10 in complex with m⁷GpppA + SAM was also deposited [PDB code 6WKS (35)] but released subsequent to our structure (PDB code 6WQ3).

Previous studies determined that the 2'-*O*-MTases of SARS-CoV and MERS-CoV, which share 93–99% and 59–66% identity to SARS-CoV-2 MTase, respectively, are heterodimers formed by the binding of nsp10 to nsp16 (16, 22). In addition, their structures have provided some insight into nsp10-dependent activation of nsp16 MTase catalysis (18, 23, 27). Because variation at the primary sequence level can impact both local and overall structure, ligand binding and structure-based drug design can also be affected by small changes in amino acid sequence. Thus, high resolution structures of the 2'-*O*-MTase from SARS-CoV-2 are necessary to best inform drug discovery for COVID-19.

The present study demonstrated that the binding site for the methyl donor SAM was highly conserved, especially the canonical Gly-X-Gly motif located at the end of the β1 and αA and Phe⁶⁹⁴⁹, as found in almost all class I MTases (40, 41). SAM analogs have been proposed as antimicrobials targeting MTases of fungi and parasites (42, 43). Here we determined the structure of nsp16-nsp10 bound to the pan-MTase inhibitor SFG and showed that SFG has nearly identical interactions with amino acids side chains as the natural substrate SAM. The high resolution of the structures with bound SAM, SAH, and SFG could facilitate the computational design of small molecules that have higher affinity than SAM for the SAM-binding cleft and have specificity for SARS-CoV-2 nsp16 compared to host methyltransferases. Further, the conservation of the binding cleft residues across the betacoronaviruses (Fig. S1) suggests that an inhibitor designed for SARS-CoV-2 could also be a broader spectrum inhibitor, which could target other coronavirus 2'-*O*-MTases.

The Cap-0 binding site offers another position in nsp16 that could be a target for small molecule inhibition. Although the published structure of nsp16-nsp10 from SARS-CoV was determined with SAM bound, only a computational model of the interaction of SARS-CoV nsp16 with m⁷GpppA-RNA was previously available (23). In order to study the different arrangements in this structure upon Cap binding, crystals were soaked with m⁷GpppA in presence of both SAM and SAH. The resulting structures identified the residues that interact with the Cap-0 and showed that conformational changes in the Cap-0 binding site could

occur during catalysis. Of particular concern for the potential development of a more broad-spectrum inhibitor that could target the Cap binding site are two nearby loops that are variable in sequence across the betacoronaviruses, which could have an impact on catalysis (Fig. 5). Our analysis of overlapped structures of nsp16 from the SARS-CoV-2 nsp16-nsp10 complex with the structure of SARS-CoV nsp16 with the Cap binding site unoccupied demonstrated that these amino acid differences do not affect the overall structure of the complex, but rather are highly flexible loops that are then stabilized upon binding of the Cap. Stabilization of these loops may be critical to obtain a high-affinity small molecule inhibitor directed at this site. The high-resolution structure of SARS-CoV-2 with the cap bound should facilitate the design of such a molecule.

A computational model of SARS-CoV 2'-O-MTase in complex with RNA was proposed previously using the structure of vaccinia virus MTase (PDB code 1AV6) as the model (23, 34, 35). This model suggested that Asp⁷⁵ in SARS CoV (Asp⁶⁸⁷³ in SARS-CoV-2), confers the selectivity of the Cap binding site for m⁷GpppA over m⁷GpppG due to steric hindrance of the Asp⁶⁸⁷³ residue with the NH₂ at position 2 of the guanidyl (23). This residue is conserved in SARS-CoV-2. However, in all our structures with Cap bound, the position 2 of the adenylate from m⁷GpppA is 7 Å away from the oxygen of the Asp⁶⁸⁷³, suggesting this residue is unlikely to be involved in the selectivity of the RNA-capped substrate in SARS-CoV-2. Indeed, recent studies show that m⁷GpppG-RNA is 2'-O-methylated by nsp16-nsp10 from SARS-CoV-2, but at a lower efficiency than is m⁷GpppA-RNA (35, 44), indicating that this site can accommodate m⁷GpppG.

In addition to the ligand binding sites, we explored the positively charged nucleotide groove, or LBS, which leads from the catalytic core of nsp16 toward nsp10. Several efforts to obtain a short RNA bound into this groove of crystals have thus far been unsuccessful, but several computational models have been recently published and are already available for SARS-CoV (23) and SARS-CoV-2 (23, 34, 35). However, no structural evidence of RNA binding has been reported. Thus, in order to obtain experimental evidence of the possible accommodation of the phosphate groups for four to five ribonucleotides of the mRNA that will directly interact with the nsp16-nsp10 groove, we cryoprotected crystals with lithium sulfate. We speculate that small charged molecules could be designed to prevent binding of mRNA, which could impair the efficiency of the MTase reaction. The potential advantage of targeting a site away from the SAM or Cap binding sites could prevent potential toxicity by avoiding cross-inhibition of human MTases.

To this end, our study revealed previously unknown features in the nsp16-nsp10 structure that could be advantageous for the design of new therapeutics. First, adenine was found to bind at two different sites, stacking with Tyr⁷⁰²⁰ and Trp⁶⁸⁰³ when the crystal was soaked with m⁷GpppA. Further, another possible nucleotide binding site occupied by a guanine moiety was identified on the back of surface of nsp16. This site was also independently identified by another group as occupied by adenosine (35). In addition, we observed BDF in this site when crystals were cryoprotected with sucrose in the absence of m⁷GpppA, suggesting this site is non-specific. Thus, this binding site needs to be further studied in order to determine whether the binding of other molecules might affect the activity of the

enzyme, and, if so, whether these newly defined binding sites could be used as potential candidates for developing inhibitors.

A final focus for development of inhibitors against nsp16 is to target the interface with its activator nsp10. Peptides derived from nsp10 have been developed to target SARS-CoV nsp16 in order to impair nsp10 binding to nsp16, resulting in inhibition of nsp16 MTase activity (45). In our analysis, we found that the residues that form the interface between nsp16 and nsp10 are 100% conserved with SARS-CoV. Thus, we predict that small molecules or peptides that target the nsp16-nsp10 interface could also be highly effective inhibitors of nsp16 from SARS-CoV-2 and closely related coronaviruses. An advantage of an inhibitor that instead binds specifically to nsp10 could have even broader implications and potentially also inhibit the N⁷-MTase nsp14, which is also activated by binding to nsp10 (16, 46).

One problem for the development of antiviral compounds against SARS-CoV-2 is the potential impact of emerging mutations in its genome. Newly emerging mutations in SARS-CoV-2 have recently been mapped onto the structures of the proteins they affect (<https://coronavirus3d.org>) (7). Few mutations have emerged for nsp10, and none of those are predicted to affect the structure of the protein. Some emerging mutations have been identified for nsp16, but likewise none of these are predicted to disrupt the structure. Further, review of these data showed that no mutations of residues at the interface between nsp16 and nsp10 have yet been detected in SARS-CoV-2 viral variants isolated from around the globe (7). Mutations that affect nsp16 activity may be absent from clinical isolates because a reduction or failure in capping could result in early immune detection and accelerate the clearance of the virus. In support of this, recent studies have suggested that stimulation of an interferon response early in SARS-CoV-2 infection could result in less severe disease, particularly in younger individuals (47). In addition, mice infected with Mouse Hepatitis Virus (MHV) showed improved survival when treated with a 29 amino acid peptide, based on the loops of nsp10, that inhibits nsp16 activation. The protection was correlated with higher levels of interferon during early infection, resulting in lower viral titers (45). Altogether, this analysis suggests that immunological studies of the impact of nsp16 inhibition could benefit from the identification of an inhibitor that can be employed for study of the virus in vitro or in animals. Such an inhibitor could also be used during early infection to stimulate immunity, reducing the likelihood for development of severe disease. The structural work found in this study will help with these next stages in our understanding of MTase-mediated modification of viral mRNA and improved treatments for COVID-19.

Materials and Methods

Chemicals and synthetic DNA

Common biological chemicals were obtained from MilliporeSigma or ThermoFisher unless otherwise indicated. DNA sequences corresponding to the predicted amino acid sequence for nsp10 and nsp16 from SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512) were codon optimized for expression in *E. coli* using GenSmart™ Codon Optimization followed by manual editing. The genes were synthesized and cloned into the pMCSG53 vector (48) by Twist Biosciences (South San Francisco, CA). The vector sequences add a TEV cleavable

N-terminal 6xHis-tag to expressed proteins. The plasmids were transformed into competent *E. coli* BL21(DE3)(Magic) cells (49).

Protein expression, purification and analytical SEC

Transformed *E. coli* cells were cultured for expression in Terrific Broth medium (BD Difco) supplemented with 200 µg/ml ampicillin and 50 µg/ml kanamycin incubated at 37°C and 220 rpm. Protein expression was induced at OD₆₀₀=1.8–2 by an addition of 0.5 mM isopropyl β-d-1-thiogalactopyranoside (Research Products International) and the cultures were further incubated at 25°C, at 200 rpm for 14 h (50). The cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris, 0.5 M NaCl, 10% glycerol, 0.1% IGEPAL CA-630) and frozen at -30°C until purification.

Frozen suspensions of cells with expressed nsp10 or nsp16 were thawed and sonicated at 50% amplitude, in 5 s × 10 s cycle for 20 min at 4°C. The lysate was cleared by centrifugation at 18,000 × *g* for 40 min at 4°C, the supernatants were collected, and the protein was purified as previously described with some modifications (51). Each supernatant was loaded into a His-Trap FF (Ni-NTA) column using a GE Healthcare ÄKTA Pure system using loading buffer (10 mM Tris-HCl pH 8.3, 500 mM NaCl, 1 mM Tris(2-carboxyethyl) phosphine (TCEP), 2 mM MgCl₂ and 5% glycerol). The column was washed with loading buffer, followed by 10 mM Tris-HCl pH 8.3, 500 mM NaCl, 25 mM imidazole, and was eluted with 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, collected and incubated with TEV-protease overnight. The cleaved tag and TEV protease were separated from protein by Ni-NTA-affinity chromatography using loading buffer and Nsp10 was collected in the flow through, whereas nsp16 eluted with 50 mM imidazole. To form the nsp16-nsp10 complex, the pure proteins were mixed at a 1:1 molar ratio at approximately 2 mg/ml in loading buffer and incubated for 1 h, and then dialyzed in crystallization buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, MgCl₂, TCEP and 5% glycerol) for 2 h (26). SAM was added to a final concentration of 2 mM. The complex was concentrated to 5.5–10 mg/ml and set up for crystallization immediately.

In order to confirm that the nsp16-nsp10 complex was formed, we performed analytical SEC using a Superdex 200 10/30 column (GE Healthcare) with 10 mM Tris-HCl, pH 7.5, 150 NaCl, 2 mM MgCl₂, 1 mM TCEP and 5% glycerol. The standard calibration curve was obtained using combined low molecular weight (LMW) and high molecular weight (HMW) Gel filtration Calibration kits (GE Healthcare). The resulting peaks from the elution of the protein were fractionated in 0.5 ml. Each fraction was collected and 8 µl of sample was denatured with Laemmli buffer (Bio-Rad), then separated using 4–15% gradient SDS-PAGE (Bio-Rad).

Crystallization, soaking and cryoprotection conditions

The nsp16-nsp10 complex + SAM was set up as 2 µl crystallization drops (1µl protein:1µl reservoir solution) in 96-well (Corning) plates using commercially available Classics II, PEG's II, AmSO₄, Anions and CompAS Suites (Qiagen). Diffraction quality crystals appeared after 5–10 days in 78 conditions, 118 crystals of various complexes were frozen,

and 57 data sets were collected. The crystals were soaked, cryoprotected and flash frozen for data collection as follows.

The small unit cell crystal (6W4H) was cryoprotected with 25% of sucrose in the well solution and the large unit cell crystal (6W75) – with 4M sodium formate (Table S1). In order to obtain complexes with SAH and SFG, crystals were transferred into the 10 μ l drops with their well solutions supplemented with 5mM of SAH or SFG, soaked for 3 h, cryoprotected with 4M sodium formate or 2M LiSO₄ and flash frozen. In the attempt to observe the complexes of nsp16-nsp10 with SAM + m⁷GpppA and SAH + m⁷GpppA crystals were transferred into 10 μ l drops containing 5 mM SAM or SAH and 0.5 mM of m⁷GpppA in their respective well solutions, soaked for various amount of time from 3 min to 6 hrs and flash frozen. 2 M LiSO₄ was used as a cryoprotectant in an attempt to observe the binding of sulfates on the places of phosphates from the RNA in the RNA binding groove. Crystals grown in PEG conditions were found not suitable for these soaks due to a phase separation of m⁷GpppA in presence of PEG.

Data collection and structure determination

Almost 120 crystals were screened, and 57 data sets were collected at the Life Sciences-Collaborative Access Team (LS-CAT) beamlines D, G and F at the Advanced Photon Source (APS) at the Argonne National Laboratory. All the data sets reported here were collected at the beamline F. Images were indexed, integrated and scaled using HKL-3000 (52). Seven structures were chosen to be described in this manuscript (Table S2).

The first structure of nsp16-nsp10 from SARS-CoV-2 in complex with SAM with the small unit cell was determined by Molecular Replacement with Phaser (53) from the CCP4 Suite (54) using the crystal structure of the nsp16-nsp10 heterodimer from SARS-CoV as a search model (PDB ID 3R24, (23)). For all other crystal structures, refined structure from this crystal form was used as a search model. The initial solutions went through several rounds of refinement in REFMAC v. 5.8.0258 (55) and manual model corrections using Coot (56). The water molecules were generated using ARP/wARP (57), SAM, SAH or SFG, Zn²⁺ and ligands were added to the model manually during visual inspection in Coot. Translation–Libration–Screw (TLS) groups were created by the TLSMD server (58) and TLS corrections were applied during the final stages of refinement. MolProbity (59) was used for monitoring the quality of the model during refinement and for the final validation of the structure. A total of seven structures were deposited to the Protein Data Bank (<https://www.rcsb.org/>) with the assigned PDB codes 6W4H, 6W75, 6WJT, 6WKQ, 6WQ3, 6WVN, and 6WRZ with associated validation reports including electron density maps of all ligands of interest.

Sequence and structural alignment

The protein sequence of nsp16 and nsp10 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 Clustal-O (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and merged with the coordinates of the structure deposited as PDB code 6w4h using ESPript 3.x (60). The PDB coordinates of SARS-CoV nsp16 and nsp10 were

analyzed on the FATCAT (29) and PDBFlex servers (61) to perform structural, flexibility and sequence alignment. Structural alignments and structure figures were downloaded from the servers and modeled in PyMOL open source V 2.1 (62). The movie showing the flexibility of nsp16 was generated with the files downloaded from PDBFlex server, frames were captured in PyMOL and exported to editor iMovie editor.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

The authors thank Lukasz Jaroszewski and Adam Godzik for construct design and Grant Wiersum for protein expression.

Funding: This project has been funded in whole or in part with Federal funds from the Department of Health and Human Services, National Institutes of Health, National Institute of Allergy and Infectious Diseases under Contract No. HHSN272201700060C. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817).

References and Notes

1. Keni R, Alexander A, Nayak PG, Mudgal J, Nandakumar K, COVID-19: Emergence, Spread, Possible Treatments, and Global Burden. *Front Public Health* 8, 216 (2020). [PubMed: 32574299]
2. Dong E, Du H, Gardner L, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*, (2020).
3. Ye ZW, Yuan S, Yuen SY, Fung SY, Chan CP, Jin DY, Zoonotic origins of human coronaviruses. *Int J Biol Sci* 16, 1686–1697 (2020). [PubMed: 32226286]
4. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY, 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 (2020). [PubMed: 31987001]
5. Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, Storgaard M, Al Khalili S, Simonsen L, Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis*, (2020).
6. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H, The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181, 914–921 e910 (2020). [PubMed: 32330414]
7. Sedova M, Jaroszewski L, Alisoltani A, Godzik A, Coronavirus3D: 3D structural visualization of COVID-19 genomic divergence. *Bioinformatics*, (2020).
8. Sola I, Almazan F, Zuniga S, Enjuanes L, Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* 2, 265–288 (2015). [PubMed: 26958916]
9. Masters PS, The molecular biology of coronaviruses. *Adv Virus Res* 66, 193–292 (2006). [PubMed: 16877062]
10. Ramanathan A, Robb GB, Chan SH, mRNA capping: biological functions and applications. *Nucleic Acids Res* 44, 7511–7526 (2016). [PubMed: 27317694]
11. Furuichi Y, Shatkin AJ, Viral and cellular mRNA capping: past and prospects. *Adv Virus Res* 55, 135–184 (2000). [PubMed: 11050942]
12. Hussain S, Pan J, Chen Y, Yang Y, Xu J, Peng Y, Wu Y, Li Z, Zhu Y, Tien P, Guo D, Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. *J Virol* 79, 5288–5295 (2005). [PubMed: 15827143]
13. Ivanov KA, Thiel V, Dobbe JC, van der Meer Y, Snijder EJ, Ziebuhr J, Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol* 78, 5619–5632 (2004). [PubMed: 15140959]

14. Chen Y, Cai H, Pan J. a., Xiang N, Tien P, Ahola T, Guo D, Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proceedings of the National Academy of Sciences* 106, 3484–3489 (2009).
15. Minskaia E, Hertzog T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, Ziebuhr J, 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 (2006). [PubMed: 16549795]
16. Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, Imbert I, Canard B, Guillemot JC, Lecine P, Pfefferle S, Drosten C, Snijder EJ, Decroly E, Morelli X, Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J Biol Chem* 289, 25783–25796 (2014). [PubMed: 25074927]
17. von Grotthuss M, Wyrwicz LS, Rychlewski L, mRNA cap-1 methyltransferase in the SARS genome. *Cell* 113, 701–702 (2003). [PubMed: 12809601]
18. 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, Canard B, Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog* 7, e1002059 (2011). [PubMed: 21637813]
19. Chen Y, Guo D, Molecular mechanisms of coronavirus RNA capping and methylation. *Virology* 511, 3–11 (2016). [PubMed: 26847650]
20. Joseph JS, Saikatendu KS, Subramanian V, Neuman BW, Brooun A, Griffith M, Moy K, Yadav MK, Velasquez J, Buchmeier MJ, Stevens RC, Kuhn P, Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *J Virol* 80, 7894–7901 (2006). [PubMed: 16873246]
21. Su D, Lou Z, Sun F, Zhai Y, Yang H, Zhang R, Joachimiak A, Zhang XC, Bartlam M, Rao Z, Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. *J Virol* 80, 7902–7908 (2006). [PubMed: 16873247]
22. Pan J, Peng X, Gao Y, Li Z, Lu X, Chen Y, Ishaq M, Liu D, Dediego ML, Enjuanes L, Guo D, Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV replication. *PLoS One* 3, e3299 (2008). [PubMed: 18827877]
23. 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, Guo D, Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog* 7, e1002294 (2011). [PubMed: 22022266]
24. Matthes N, Mesters JR, Coutard B, Canard B, Snijder EJ, Moll R, Hilgenfeld R, The non-structural protein Nsp10 of mouse hepatitis virus binds zinc ions and nucleic acids. *FEBS Lett* 580, 4143–4149 (2006). [PubMed: 16828088]
25. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL, Tummino TA, Huettenhain R, Kaake RM, Richards AL, Tutuncuoglu B, Foussard H, Batra J, Haas K, Modak M, Kim M, Haas P, Polacco BJ, Braberg H, Fabius JM, Eckhardt M, Soucheray M, Bennett MJ, Cakir M, McGregor MJ, Li Q, Meyer B, Roesch F, Vallet T, Mac Kain A, Miorin L, Moreno E, Naing ZZC, Zhou Y, Peng S, Shi Y, Zhang Z, Shen W, Kirby IT, Melnyk JE, Chorba JS, Lou K, Dai SA, Barrio-Hernandez I, Memon D, Hernandez-Armenta C, Lyu J, Mathy CJ, Perica T, Pilla KB, Ganesan SJ, Saltzberg DJ, Rakesh R, Liu X, Rosenthal SB, Calviello L, Venkataramanan S, Liboy-Lugo J, Lin Y, Huang XP, Liu Y, Wankowicz SA, Bohn M, Safari M, Ugur FS, Koh C, Savar NS, Tran QD, Shengjuler D, Fletcher SJ, O'Neal MC, Cai Y, Chang J, Broadhurst DJ, Klippsten S, Sharp PP, Wenzell NA, Kuzuoglu D, Wang HY, Trenker R, Young JM, Cavero DA, Hiatt J, Roth TL, Rathore U, Subramanian A, Noack J, Hubert M, Stroud RM, Frankel AD, Rosenberg OS, Verba KA, Agard DA, Ott M, Emerman M, Jura N, von Zastrow M, Verdin E, Ashworth A, Schwartz O, d'Enfert C, Mukherjee S, Jacobson M, Malik HS, Fujimori DG, Ideker T, Craik CS, Floor SN, Fraser JS, Gross JD, Sali A, Roth BL, Ruggero D, Taunton J, Kortemme T, Beltrao P, Vignuzzi M, Garcia-Sastre A, Shokat KM, Shoichet BK, Krogan NJ, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, (2020).
26. Debarnot C, Imbert I, Ferron F, Gluais L, Varlet I, Papageorgiou N, Bouvet M, Lescar J, Decroly E, Canard B, Crystallization and diffraction analysis of the SARS coronavirus nsp10-nsp16 complex. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67, 404–408 (2011).

27. Wei SM, Yang L, Ke ZH, Liu QY, Che Y, Yang ZZ, Guo DY, Fan CP, Crystal structure of MERS-CoV nsp10/nsp16 complex bound to SAMor bound to SAM and m7GpppA. RCSB Protein Data Bank. DOI: 10.2210/pdb5YNM/pdb ; DOI: 10.2210/pdb5YNM/pdb10.2210/pdb5YN6/pdb; DOI: 10.2210/pdb5YN6/pdb . Release date 12/05/2018.
28. 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, Jiang T, Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* 27, 325–328 (2020). [PubMed: 32035028]
29. Ye Y, Godzik A, FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Res* 1, 582–585 (2004).
30. Lugari A, Betzi S, Decroly E, Bonnaud E, Hermant A, Guillemot JC, Debarnot C, Borg JP, Bouvet M, Canard B, Morelli X, Lecine P, Molecular mapping of the RNA Cap 2'-O-methyltransferase activation interface between severe acute respiratory syndrome coronavirus nsp10 and nsp16. *J Biol Chem* 285, 33230–33241 (2010). [PubMed: 20699222]
31. Bollati M, Milani M, Mastrangelo E, Ricagno S, Tedeschi G, Nonnis S, Decroly E, Selisko B, de Lamballerie X, Coutard B, Canard B, Bolognesi M, Recognition of RNA cap in the Wesselsbron virus NS5 methyltransferase domain: implications for RNA-capping mechanisms in Flavivirus. *J Mol Biol* 385, 140–152 (2009). [PubMed: 18976670]
32. Aouadi W, Blanjoie A, Vasseur JJ, Debart F, Canard B, Decroly E, Binding of the Methyl Donor S-Adenosyl-l-Methionine to Middle East Respiratory Syndrome Coronavirus 2'-O-Methyltransferase nsp16 Promotes Recruitment of the Allosteric Activator nsp10. *J Virol* 91, (2017).
33. Zhao Y, Soh TS, Lim SP, Chung KY, Swaminathan K, Vasudevan SG, Shi PY, Lescar J, Luo D, Molecular basis for specific viral RNA recognition and 2'-O-ribose methylation by the dengue virus nonstructural protein 5 (NS5). *Proc Natl Acad Sci U S A* 112, 14834–14839 (2015). [PubMed: 26578813]
34. Krafcikova P, Silhan J, Nencka R, Boura E, Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat Commun* 11, 3717 (2020). [PubMed: 32709887]
35. Viswanathan T, Arya S, Chan SH, Qi S, Dai N, Misra A, Park JG, Oladunni F, Kovalskyy D, Hromas RA, Martinez-Sobrido L, Gupta YK, Structural basis of RNA cap modification by SARS-CoV-2. *Nat Commun* 11, 3718 (2020). [PubMed: 32709886]
36. Martinez MA, Compounds with Therapeutic Potential against Novel Respiratory 2019 Coronavirus. *Antimicrob Agents Chemother* 64, (2020).
37. Kim Y, Jedrzejczak R, Maltseva N, Endres M, Godzik A, Joachimiak A, Center for Structural Genomics of Infectious Diseases (CSGID), Crystal Structure of the methyltransferase-stimulatory factor complex of NSP16 and NSP10 from SARS CoV-2. RCSB Protein Data Bank. DOI: 10.2210/pdb6W61/pdb. Release date 3/25/2020.
38. Yan LM, Huang YC, Lou ZY, Rao ZH, Crystal structure of 2019-nCoV nsp16-nsp10 complex. RCSB Protein Data Bank. DOI: 10.2210/pdb7BQ7/pdb. Release date 3/24/2020.,
39. Lin S, Chen H, Ye F, Chen Z, Yang F, Zheng Y, Cao Y, Qiao J, Yang S, Lu G, Crystal structure of SARS-CoV-2 nsp10/nsp16 2'-O-methylase and its implication on antiviral drug design. *Signal Transduct Target Ther* 5, 131 (2020). [PubMed: 32728018]
40. Schubert HL, Blumenthal RM, Cheng X, Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* 28, 329–335 (2003). [PubMed: 12826405]
41. Smietanski M, Werner M, Purta E, Kaminska KH, Stepinski J, Darzynkiewicz E, Nowotny M, Bujnicki JM, Structural analysis of human 2'-O-ribose methyltransferases involved in mRNA cap structure formation. *Nat Commun* 5, 3004 (2014). [PubMed: 24402442]
42. Nolan LL, Molecular target of the antileishmanial action of sinefungin. *Antimicrob Agents Chemother* 31, 1542–1548 (1987). [PubMed: 3124733]
43. Zheng S, Hausmann S, Liu Q, Ghosh A, Schwer B, Lima CD, Shuman S, Mutational analysis of *Encephalitozoon cuniculi* mRNA cap (guanine-N7) methyltransferase, structure of the enzyme bound to sinefungin, and evidence that cap methyltransferase is the target of sinefungin's antifungal activity. *J Biol Chem* 281, 35904–35913 (2006). [PubMed: 16971388]

44. Benoni R, Krafcikova P, Baranowski MR, Kowalska J, Boura E, Cahová H, Substrate specificity of SARS-CoV-2 nsp10-nsp16 methyltransferase. *bioRxiv*, 2020.2007.2030.228478 (2020).
45. Wang Y, Sun Y, Wu A, Xu S, Pan R, Zeng C, Jin X, Ge X, Shi Z, Ahola T, Chen Y, Guo D, Coronavirus nsp10/nsp16 Methyltransferase Can Be Targeted by nsp10-Derived Peptide In Vitro and In Vivo To Reduce Replication and Pathogenesis. *J Virol* 89, 8416–8427 (2015). [PubMed: 26041293]
46. Ferreira de Freitas R, Ivanochko D, Schapira M, Methyltransferase Inhibitors: Competing with, or Exploiting the Bound Cofactor. *Molecules* 24, (2019).
47. Zhou Z, Ren L, Zhang L, Zhong J, Xiao Y, Jia Z, Guo L, Yang J, Wang C, Jiang S, Yang D, Zhang G, Li H, Chen F, Xu Y, Chen M, Gao Z, Yang J, Dong J, Liu B, Zhang X, Wang W, He K, Jin Q, Li M, Wang J, Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. *Cell Host Microbe* 27, 883–890 e882 (2020). [PubMed: 32407669]
48. Eschenfeldt WH, Makowska-Grzyska M, Stols L, Donnelly MI, Jedrzejczak R, Joachimiak A, New LIC vectors for production of proteins from genes containing rare codons. *J Struct Funct Genomics* 14, 135–144 (2013). [PubMed: 24057978]
49. Kwon K, Peterson SN, in *Structural Genomics and Drug Discovery: Methods and Protocols*, Anderson WF, Ed. (Springer New York, New York, NY, 2014), pp. 61–74.
50. Millard CS, Stols L, Quartey P, Kim Y, Dementieva I, Donnelly MI, 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 (2003). [PubMed: 12767824]
51. Shuvalova L, in *Methods Mol Biol.* (2014), vol. 1140, pp. 137–143. [PubMed: 24590714]
52. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M, 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 (2006). [PubMed: 16855301]
53. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, Phaser crystallographic software. *J Appl Crystallogr* 40, 658–674 (2007). [PubMed: 19461840]
54. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS, Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67, 235–242 (2011). [PubMed: 21460441]
55. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67, 355–367 (2011). [PubMed: 21460454]
56. Emsley P, Cowtan K, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126–2132 (2004). [PubMed: 15572765]
57. Morris RJ, Perrakis A, Lamzin VS, ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* 374, 229–244 (2003). [PubMed: 14696376]
58. Painter J, Merritt EA, Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62, 439–450 (2006). [PubMed: 16552146]
59. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66, 12–21 (2010). [PubMed: 20057044]
60. Robert X, Gouet P, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42, W320–324 (2014). [PubMed: 24753421]
61. Hrabe T, Li Z, Sedova M, Rotkiewicz P, Jaroszewski L, Godzik A, PDBFlex: exploring flexibility in protein structures. *Nucleic Acids Res* 44, D423–428 (2016). [PubMed: 26615193]
62. DeLano WL, Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallog*, 82–92 (2002).

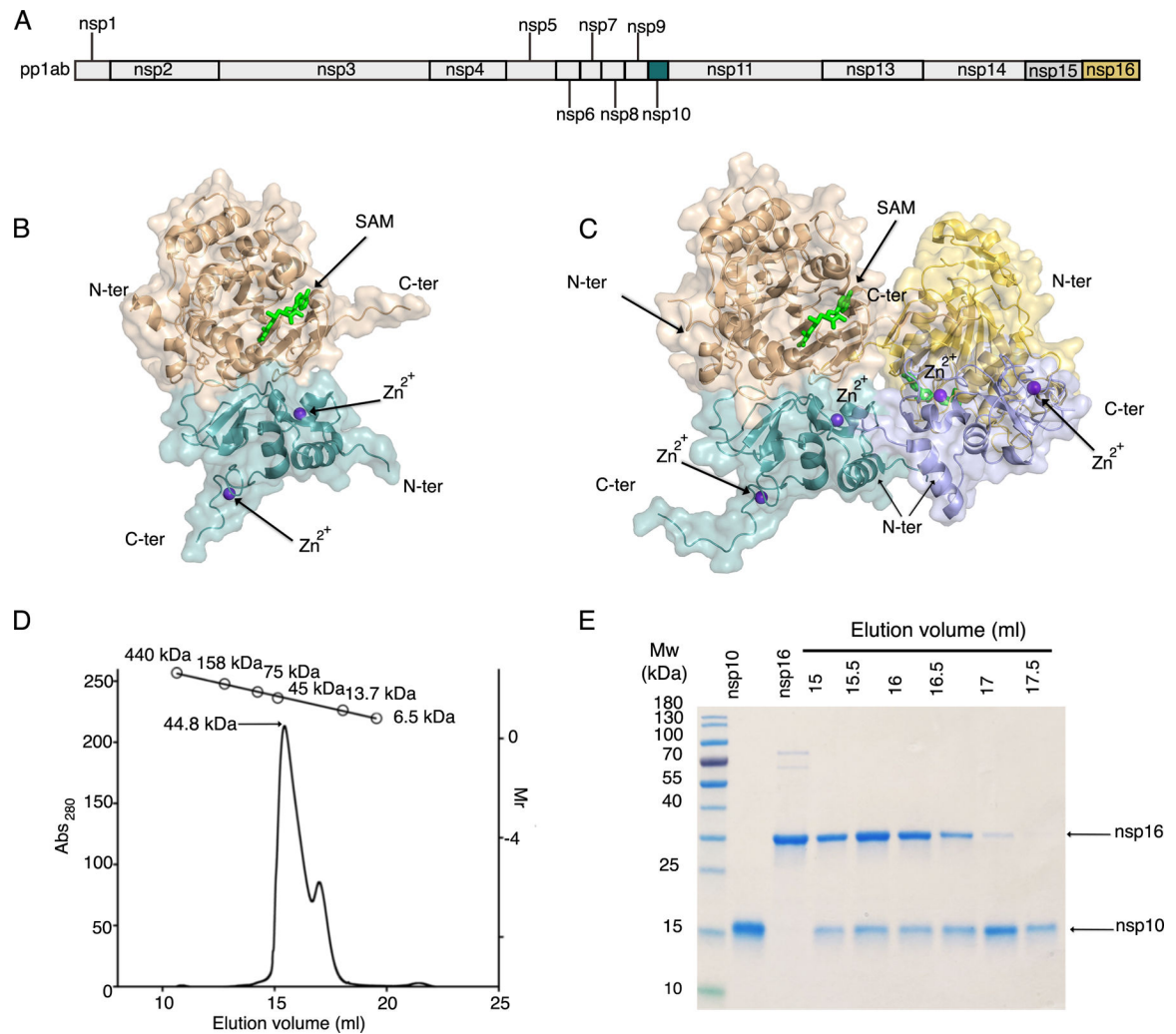


Fig. 1. Overall structure of the nsp16-nsp10 oligomers.

(A) Linear schematic of the *orf1a/orf1b* protein product pp1ab prior to proteolytic processing. (B) Cartoon representation of the nsp16-nsp10 heterodimer of the small unit cell crystal form (PDB code 6W4H). (C) Cartoon representation of the two nsp16-nsp10 heterodimers in the asymmetric unit of the large unit cell crystal form (PDB code 6W75). In (B) and (C), nsp16 is in shades of tan and yellow and nsp10 is in shades of teal and purple. Ligands are represented as sticks, with SAM in bright green and Zn²⁺ in purple. N-ter, N terminus; C-ter, C terminus. (D) Elution profile for analytical size-exclusion chromatography (SEC) with corresponding plot for molecular weight standards shown at the top. (E) Separation of elution fractions on a 4–15% gradient SDS-PAGE gel stained with Coomassie blue.

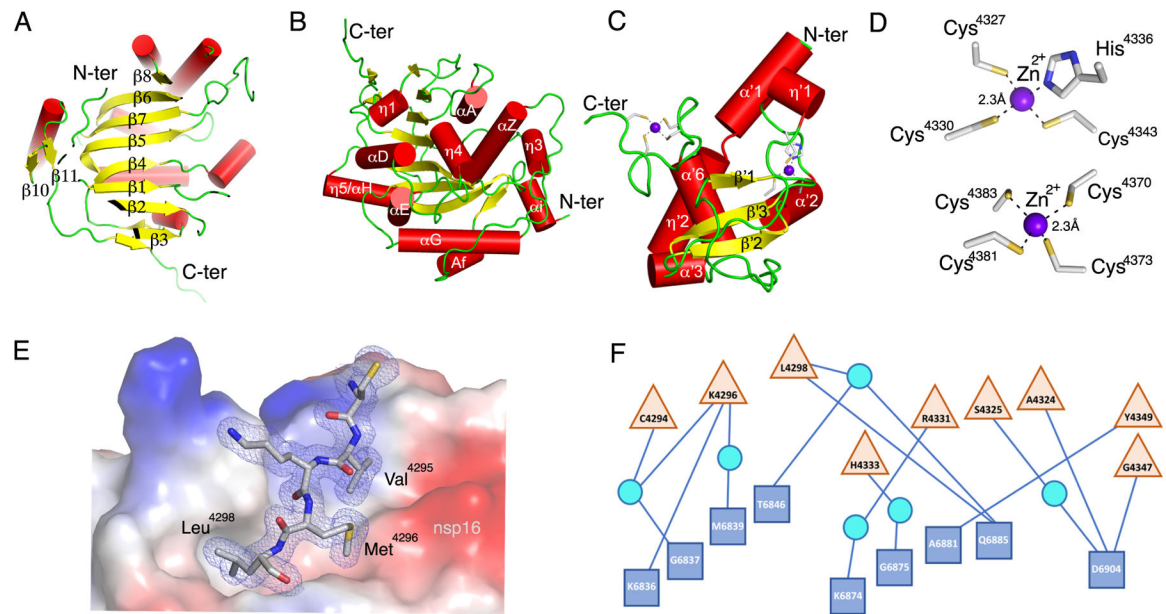


Fig. 2. Detailed representation of nsp16, nsp10, and the heterodimer interface.

(A to C) Cartoon representations of two views of nsp16 featuring the canonical β -sheet (A) and overall secondary structure of nsp16 (B) and nsp10 (C). α -helices are shown as red cylinders, β -strands as yellow arrows, loops as green strands, and zinc ions as purple spheres. (D) Close-up view of the two Zn^{2+} binding sites in nsp10. (E) Interaction of Cys4294-Leu4298 (sequence CVKML, grey sticks) from nsp10 with the hydrophobic surface of nsp16 (colored by electrostatic potential). Oxygen, red sticks; nitrogen, blue sticks; sulfur, yellow sticks. (F) Schematic representation of residues from nsp16 (blue squares) and nsp10 (tan triangles) that interact through hydrogen bonds, represented as lines. Some interactions are mediated by water molecule (cyan circles). For panels A to D, structural representations are based on the structure of the nsp16-nsp10 complex with $m^7\text{GpppA}$ and SAM (PDB code 6WVN). E and F are based on the structure of the nsp16-nsp10 complex with SAM in the small unit crystal form (PDB code 6W4H).

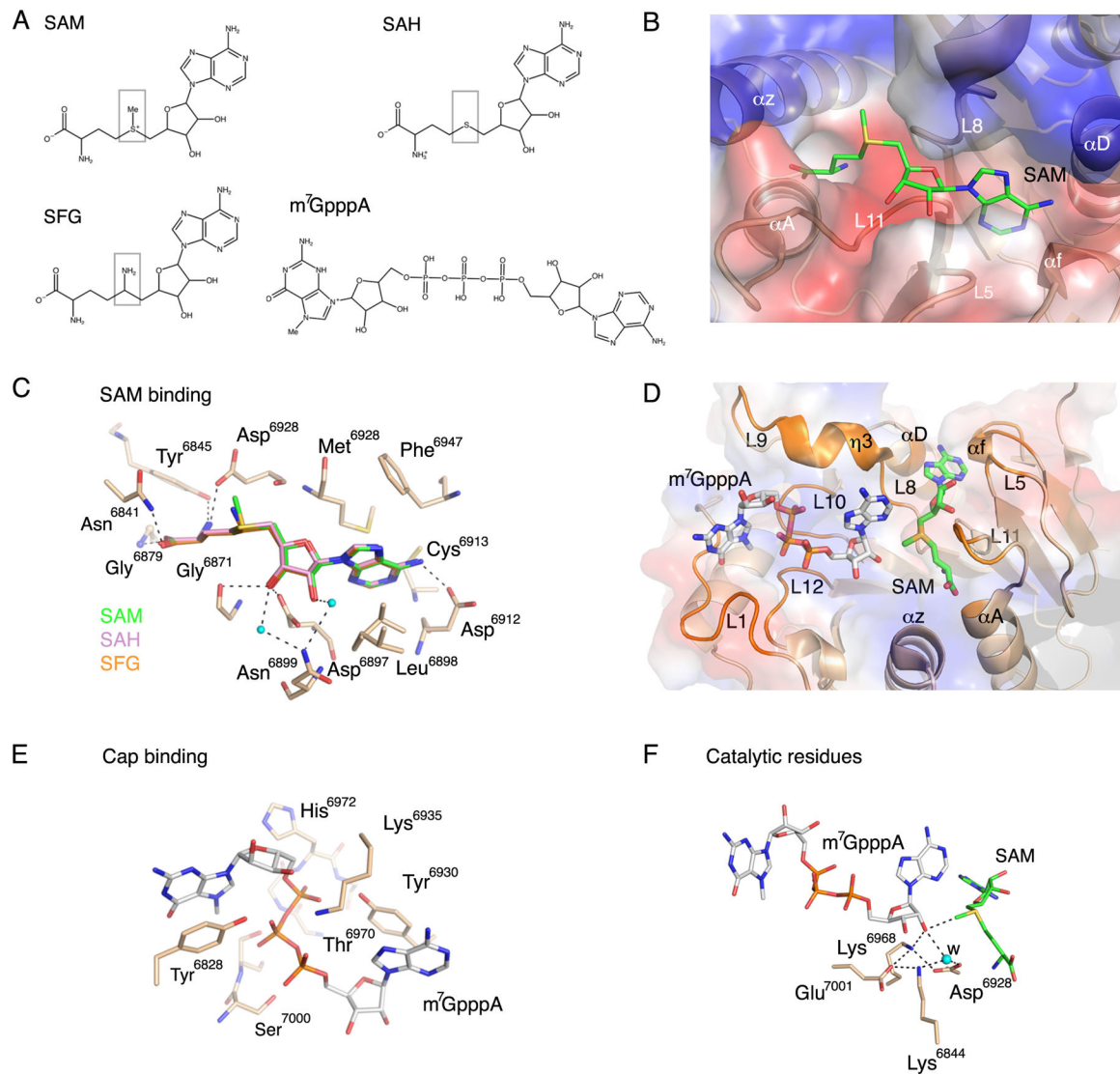


Fig. 3. Substrate interactions and catalytic site.

(A) Chemical structures of the methyl donor SAM; the product after methyl transfer, SAH; the SAH analog and inhibitor, SFG; and the Cap-0 analog, m⁷GpppA. Boxes highlight differences in the chemical structures of SAH and SFG compared to SAM. (B) Cartoon and surface charge representations of the nsp16 SAM binding cleft occupied by SAM (green sticks). (C) Close-up view of the overlay of nsp16 structures with the ligands SAM (green, PDB code 6W75), SAH (pink, PDB code 6WJT) and SFG (orange, PDB code 6WKQ). (D) Cartoon and surface charge representations of nsp16 (PDB code 6WVN) with the SAM binding cleft occupied by SAM (green sticks) and the Cap binding site occupied by m⁷GpppA (gray sticks). (E) Detailed view of the residues that coordinate m⁷GpppA in the Cap binding site as tan sticks. (F) Close-up view of the side chains of the catalytic residues showing the orientation of the methyl group in SAM in proximity to the acceptor 2'-OH group in m⁷GpppA. Dashed lines indicate the interactions between the residues in the active site, and small cyan dots indicate water molecules (w). Red sticks, oxygen; blue sticks, nitrogen; orange sticks, phosphate; yellow sticks, sulfur.

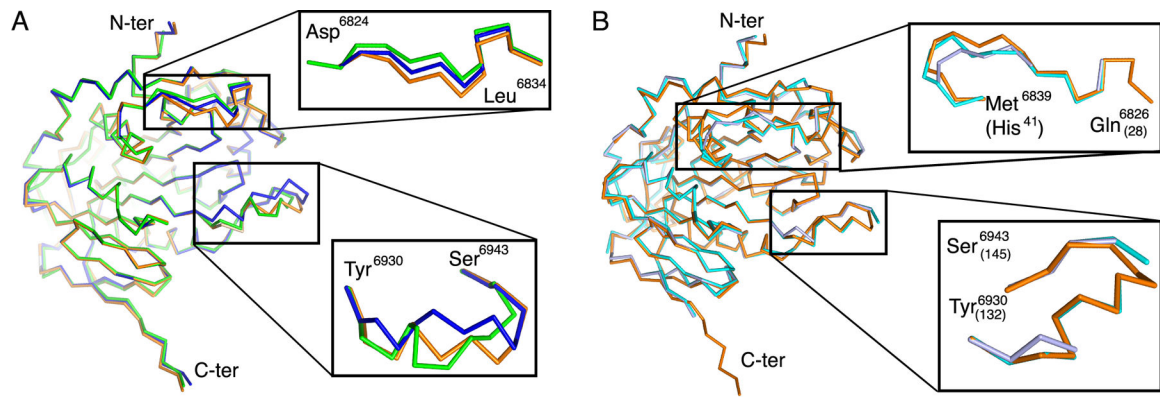


Fig. 4. Structural alignment of nsp16 in the presence and absence of m7GpppA.

(A) Alignment of the C- α chain of nsp16 from SARS-CoV-2 in complex with SAM from the small unit cell (blue, PDB ID 6W4H), in complex with SAM from the large unit cell crystal form (green, PDB code 6W75), and in complex with SAM and m7GpppA (orange, PDB code 6WVN). Two flexible loops are enlarged in insets. (B) Alignment of the C- α chain of nsp16 from SARS-CoV-2 in complex with SAM and m7GpppA (orange) with the corresponding region of MERS-CoV nsp16 in complex with SAM alone (light blue, PDB code 5YN6) or in complex with SAM and m7GpppA (cyan, PDB code 5YNM). Numbering of MERS residues is indicated in parentheses.

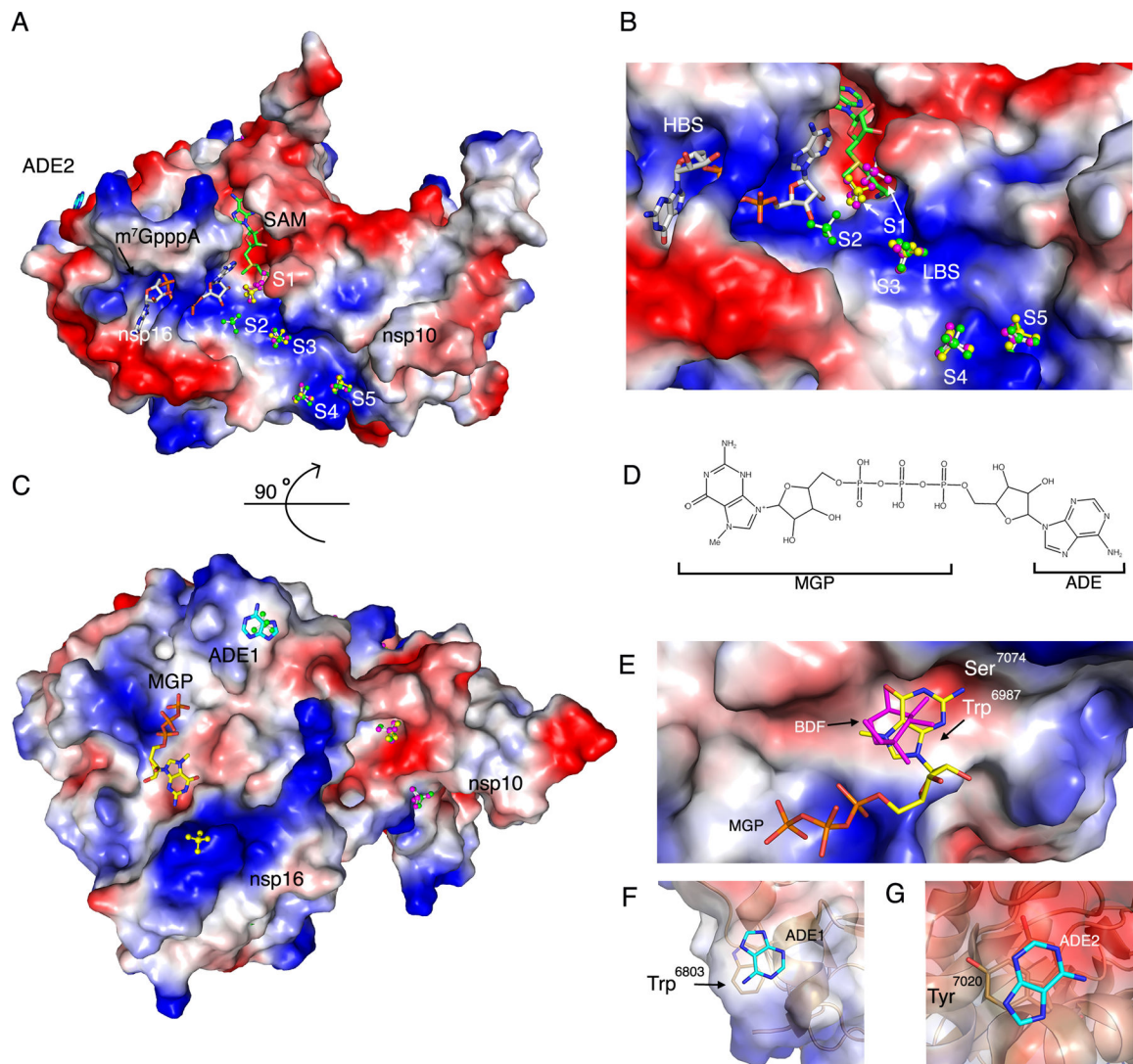


Fig. 5. Sulfate and nucleotide binding sites on nsp16.

(A) Surface charge representation of nsp16-nsp10 bound to SAM and m⁷GpppA (PDB code 6 WVN) with sulfates in balls and sticks along the nucleotide binding groove numbered from the catalytic core to the nsp10 extension (S1-S5). The sulfates in the overlaid structures are designated by color according with their corresponding PDB code: 6WRZ (green), 6WVN (yellow), and 6WQ3 (pink). m⁷GpppA is shown as gray sticks and SAM in green sticks. Positive charges are shown in blue and negative charges in red. ADE2, adenine moiety 2. (B) Close-up view of (A) showing m⁷GpppA, SAM and S1-S5 in the high-affinity binding site (HBS) and low-affinity binding site (LBS). (C) 90° rotation of the complex showing the secondary binding sites MGP and ADE1 along with additional sulfates. (D) Schematic representation of m⁷GpppA noting the MGP (m⁷GpppA guanine and phosphate) and ADE (adenine) moieties. (E) Surface charge representation of the nsp16 MGP binding site with MGP (yellow sticks) and BDF (pink sticks) from structure 6W4H. (F and G) Cartoon and surface charge representation of the adenine moieties (ADE1 and ADE2) bound

to nsp16 from structure (PDB code 6WVN). Sticks, carbon; blue, nitrogen; red, oxygen; orange, phosphate; yellow, sulfate.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript