Structural and functional insights into the enzymatic plasticity of the SARS-CoV-2 NiRAN domain

**Highlights**
- Cryo-EM reveals diverse nucleotide binding poses in the NiRAN domain
- GDP is the preferred substrate for mRNA capping in SARS-CoV-2
- High-resolution maps allow for insights into the chemical mechanism of catalysis
- Structures provide a platform for antiviral design targeting the NiRAN domain

**Authors**
Gabriel I. Small, Olga Fedorova, Paul Dominic B. Olinares, ..., Brian T. Chait, Seth A. Darst, Elizabeth A. Campbell

**Correspondence**
campbee@rockefeller.edu

**In brief**
The NiRAN domain in SARS-CoV-2 has been implicated as the enzyme responsible for capping of viral mRNA and NMPylation of nsp9. Small and colleagues reveal the structural basis of mRNA capping and NMPylation of nsp9 and describe how the NiRAN domain catalyzes such distinct activities in the same active site.
Structural and functional insights into the enzymatic plasticity of the SARS-CoV-2 NiRAN domain

Gabriel I. Small,1 Olga Fedorova,2,3 Paul Dominic B. Olinares,4 Joshua Chandanani,1 Anoosha Banerjee,1,5 Young Joo Choi,1,7 Henrik Molina,6 Brian T. Chait,4 Seth A. Darst,1 and Elizabeth A. Campbell1,8,*

1Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
2Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA
3Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA
4Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY, USA
5Tri-Institutional Program in Chemical Biology, The Rockefeller University, New York, New York, USA
6Proteomics Resource Center, The Rockefeller University, New York, NY 10065, USA
7Present address: University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA
8Lead contact
*Correspondence: campbee@rockefeller.edu
https://doi.org/10.1016/j.molcel.2023.10.001

SUMMARY

The enzymatic activity of the SARS-CoV-2 nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain is essential for viral propagation, with three distinct activities associated with modification of the nsp9 N terminus, NMPylation, RNAylation, and deRNAylation/capping via a GDP-polyribonucleotidyltransferase reaction. The latter two activities comprise an unconventional mechanism for initiating viral RNA 5’ cap formation, while the role of NMPylation is unclear. The structural mechanisms for these diverse enzymatic activities have not been properly delineated. Here, we determine high-resolution cryoelectron microscopy (cryo-EM) structures of catalytic intermediates for the NMPylation and deRNAylation/capping reactions, revealing diverse nucleotide binding poses and divalent metal ion coordination sites to promote its repertoire of activities. The deRNAylation/capping structure explains why GDP is a preferred substrate for the capping reaction over GTP. Altogether, these findings enhance our understanding of the promiscuous coronaviral NiRAN domain, a therapeutic target, and provide an accurate structural platform for drug development.

INTRODUCTION

Nidovirales are an order of positive-sense, single-stranded RNA viruses that includes the coronavirus (CoV) family,1 etiological agents responsible for major and deadly zoonotic events.2,3 Most recently, SARS-CoV-2 has been responsible for the ongoing COVID-19 pandemic,4,5 with catastrophic deaths and health, social, and economic impacts worldwide. CoVs possess large (~30 kb), poly-cistronic RNA genomes containing open reading frames (ORFs) for the non-structural proteins (nspS) at the 5’ end, followed by ORFs encoding structural and accessory proteins at the 3’ end.6 ORF1a and ORF1ab encode two polyproteins, PP1a and PP1ab, respectively, which are proteolytically processed into the sixteen nspS responsible for genome expression and replication.3,6 Central to replication is the conserved RNA-dependent RNA polymerase (RdRp) encoded in nsp12 that, along with essential cofactors nsp7 and two copies of nsp8, comprises the CoV holo-RdRp.7 Multiple structures of the holo-RdRp with product-template RNA scaffold have been resolved by cryoelectron microscopy (cryo-EM) and are referred to as replication-transcription complexes (RTCs).8–10

The Nidoviral RdRp-associated nucleotidyltransferase (NiRAN) domain is a unique, enigmatic domain that is conserved across all Nidovirales.11 The NiRAN domain is encoded amino-terminal (N-terminal) to the RdRp within the same nsp (nsp12 of SARS-CoV-2).12 The NiRAN domain was initially identified in equine arterivirus (EAV) as an Mn2+-dependent nucleotidyltransferase targeting an invariant lysine in the NiRAN domain. When conserved residues suggested to be involved in catalysis were mutated, viral replication was largely or completely abrogated.11 The CoV NiRAN domain was subsequently demonstrated to contain an orthologous nucleotidyltransferase activity that transferred a nucleoside monophosphate (NMP) moiety to the N-terminal amine of nsp9, a protein without an ortholog outside the CoV family, in a process termed NMPylation.13 The NMPylation of nsp9 has been robustly demonstrated in vitro across multiple CoVs.13–17 However, an in vivo role for nsp9 NMPylation has not been uncovered.
Like many eukaryotic mRNAs, CoV RNAs are 5’ capped by a 7-methylguanosine nucleotide linked by a 5’-5’-triphosphate to the 5’-RNA 2’-O-methylated nucleotide. The mRNA caps are crucial for protecting transcripts from degradation and for translation initiation. Because CoVs are not known to access the infected cell nucleus where the host capping assembly apparatus is located, CoVs encode their own enzymes for generating 5’ caps on the viral RNAs. Although nsp14 and nsp16 have long been known to harbor the N7-methyltransferase and 2’-O-methyltransferase activities, the identity of the enzyme that initially generates the 5’-G linkage was only recently identified. In a unique two-step mechanism, the NiRAN domain catalyzes the transfer of a 5’-monophosphorylated RNA (pRNA) from a nascent 5’-triphosphorylated RNA (pppRNA) to the N terminus of nsp9, forming a covalent RNA-nsp9 intermediate in a process termed RNAylation. The NiRAN then transfers the RNA from nsp9 to guanosine diphosphate (GDP), generating the core cap structure GpppN-RNA. This GDP-polyribonucleotidyltransferase (GDP-PRNTase) activity is responsible for SARS-CoV-2 mRNA capping.

To gain insight into the remarkable enzymatic plasticity of the NiRAN domain, we structurally and functionally characterized two of the three known NiRAN enzymatic activities, UMPylation and deRNAylation/capping. We used a uridine triphosphate (UTP) analog with a non-hydrolyzable α-β-phosphate bond (UMPCPP) to trap a structure of the nsp9-RTC poised for UMPylation of the nsp9 N terminus at 2.9 Å nominal resolution (~2.8 Å local resolution in the NiRAN active site). We used a slowly reacting GDP analog (GDP-βS) to visualize structures of the RNA-nsp9-RTC poised for deRNAylation and GDP transfer at 3.0 Å nominal resolution (~2.8 Å local resolution in the NiRAN active site). Our results illustrate why GDP is a strongly preferred substrate for the NiRAN GDP-transferase activity. The high local resolutions of our structures enable near atomic-resolution insights into the diverse NiRAN catalytic activities and their mechanisms.

RESULTS

Structure/function analysis of NiRAN-mediated NMPylation of nsp9

The SARS-CoV-2 NiRAN domain efficiently NMPylates the N terminus of nsp9 in vitro; UMPylation was 100% efficient, as monitored by native mass spectrometry (nMS) (Figure 1A), while NMPylation with other nucleoside triphosphates (NTPs) was less efficient. We used a UTP analog (uridine-5’-[α,β]-methylene]triphosphate [UMPCPP]) to capture a pre-catalytic intermediate of NiRAN-mediated nsp9 UMPylation (nsp9-RTC(UMPCPP)) and determined a cryo-EM structure to 2.9 Å nominal resolution (Figures 1B and S1A–S1E; Table S1). We call this complex UMP-i, for UMPylation intermediate. In UMP-i, nsp9 docks into the NiRAN domain with the conserved nsp9-GxxxG helix (Figure S1F), binding a conserved hydrophobic patch on nsp12. The eponymous glycines of the nsp9 active site. We used a slowly reacting GDP analog (GDP-βS) to visualize structures of the RNA-nsp9-RTC poised for deRNAylation and GDP transfer at 3.0 Å nominal resolution (~2.8 Å local resolution in the NiRAN active site). Our results illustrate why GDP is a strongly preferred substrate for the NiRAN GDP-transferase activity. The high local resolutions of our structures enable near atomic-resolution insights into the diverse NiRAN catalytic activities and their mechanisms.
GxxxG helix, G100 and G104, are two of six residues in the helix that form a series of hydrophobic interactions with the NiRAN domain. The five N-terminal residues of nsp9 extend through a deep groove on the NiRAN domain surface, terminating with the nsp9 N-terminal Asn (N1) deep in the NiRAN active site (Figures 1C and 1D).

Nsp9 can dimerize in solution, and nsp9 crystallizes as a dimer with the GxxxG helix at the dimer interface.23,24 Mutations in conserved residues of the nsp9 GxxxG helix disrupt nsp9 dimerization in solution, and these same mutations result in loss of viral fitness.25,26 From these results it was concluded that nsp9 dimerization was required for viral fitness. However, it is also possible that the loss in fitness is due to impaired ability to bind nsp12. More work is required to confirm whether the dimerization of nsp9 is a regulatory feature in vivo.

The N-terminal loop of nsp9 extending into the NiRAN domain active site comprises five conserved residues, most notably, the invariant N-terminal tripeptide NNE (Figure S1F). The nsp9 N-terminus is positioned for catalysis by a hydrogen bond (H-bond) between the sidechain of nsp9 N1 and the sidechain of nsp12 N39, and the interface continues with a H-bonding network between nsp9 N2, E3, and L4 and nsp12 Y38, Y728, and R733 (Figures 1D and S1G). The role of the nsp12 palm domain residues in NiRAN domain functions has previously been noted,16,27 and the residues are conserved within CoVs (Figure S1G). Due to the importance of this motif in facilitating nsp9 binding and other NiRAN domain functions, we call these conserved nsp9-palm residues the NiRAN-associated palm (NPalm) motif. Despite the absolute conservation of nsp9 E3 (Figure S1F), the sidechain does not appear to play a role in binding nsp12, suggesting a potential role in other nsp9 functions.

Within the NiRAN domain active site pocket, the UMPCPP is positioned in a “base-up” pose, distinct from previously observed “base-out” or “base-in” NiRAN domain NTP binding poses,28 with its α-phosphate positioned for catalysis (Figure 1E). Structural modeling indicates that base-in binding of guanosine triphosphate (GTP) in the NiRAN G-site29 would sterically clash with the base-up NMPylation pose, explaining the lower efficiency of GMPylation (Figures 1A and S1H). The UMPCPP is positioned by a network of polar interactions, including H-bonds between the nucleobase functional groups O2 and O4 and NPalm N713 and NiRAN N39, respectively (Figure 1E). Density features in the cryo-EM map suggested the presence of metal ions in the active site. On the basis of the presence of 2 mM MgCl₂ in the cryo-EM buffer, the Mg³⁺-ion dependence of the NMPylation reaction,16 and the observed coordination ligands and geometries, we assigned Mg²⁺-ions to two of these densities (Figure 1E). Mg³⁺ is coordinated close to the pending phosphoramidate bond, poised to catalyze the reaction that would break the α-β-phosphodiester bond in the natural substrate UTP (Figure 1E). Mg²⁺ is coordinated close to the β- and γ-phosphates, likely stabilizing the PPI leaving group.29

NiRAN domain RNA substrate preferences

In addition to nsp9 NMPylation, the CoV NiRAN domain mediates RNAylation of the nsp9 N terminus as well as the deRNAylation of nsp9 and transfer of GDP to the RNA to form the 5′-GpppA cap.16 The NiRAN-mediated RNAylation and deRNAylation reactions are inefficient in vitro16 compared with NMPylation. We hypothesized that RNAylation and/or deRNAylation may be facilitated by secondary structure in the substrate RNA. To test this hypothesis, we examined three 5′-triposphorylated model RNAs based on the 5′ UTR of the SARS-CoV-2 genome (Figure S2A), an unstructured 10-mer, a 20-mer, and a 33-mer that mimics the 5′ UTR through conserved stem-loop (SL) 1.30 The 20-mer RNA is predicted to form a stable hairpin not found in the SARS-CoV-2 genome31 but could represent a nascent RNA intermediate structure. Monitoring the time course of NiRAN-mediated nsp9 RNAylation with each pppRNA revealed a slight preference for the shorter, unstructured 10-mer ppRNA (Figures S2B and S2C). To test the effect of RNA secondary structure on the deRNAylation/capping reaction, we purified the three different RNAylated-nsp9 species and evaluated NiRAN-mediated RNA capping efficiencies using GDP as a substrate. The efficiency and rate of the deRNAylation/capping reaction was not affected by RNA length or secondary structure propensity (Figures S2D and S2E).

DeRNAylation of RNA-nsp9 has been previously demonstrated with both GDP and GTP.15,31 As observed by Park et al.16 the NiRAN domain strongly prefers GDP over GTP for 20-mer RNA-nsp9 deRNAylation (Figures 2A and 2B). The efficiency of the deRNAylation reaction with GTP was restored by the addition of a stoichiometric amount of the nsp13 helicase (Figures 2A and 2B), indicating that the nucleoside triphosphatase (NTPase) activity of nsp1332 can convert the GTP pool into mostly GDP for efficient GDP-PRNtase activity. With nMS, we monitored mRNA capping (Figures 2C and 2D) and observed the formation of the canonical 5′ mRNA cap when either GDP or GTP was used in the reaction (Figure 2C), distinct from the formation of a noncanonical mRNA cap produced in Rhabdoviruses when GTP is utilized as a substrate.33 Monitoring the reaction with nMS allows us to observe the capping reaction beginning with RNAylation of nsp9 and followed by the deRNAylation of nsp9 and the capping of the mRNA alongside the restoration of unmodified nsp9 (Figure 2D).

Structural analysis of a NiRAN domain capping intermediate

To understand the structural basis of the NiRAN domain GDP-PRNtase/capping activity, we determined 3.0 Å resolution capping intermediate cryo-EM structures of the RTC in complex with RNAylated-nsp9 (20-mer RNA; Figures 3A and S4H) with a poorly reactive GDP analog, GDP-βS (Figures S3 and S4A–S4H; Table S1). After initial steps of processing, the consensus map exhibited features indicative of structural heterogeneity. To improve the maps in the area of interest around the NiRAN active site, we constructed a soft mask encompassing the NiRAN domain and nsp9 and used masked classification with signal subtraction34 to identify six conformational states (Figure S3C). Two of the states (C00 and C02; Figure S3C) had very poor or no density for nsp9 and were not further processed. C01 contained strong cryo-EM density for RNAylated-nsp9 resolving the entire 20-mer length of the nsp9-linked RNA with the predicted SL (Figure S2A), but the GDP-βS substrate was unable to bind (Figure S5). The remaining classes revealed conformational mobility of nsp9 in its binding to nsp12. C03 and C05 were similar to C04 (except for motion in nsp9), but the maps...
in the NiRAN active site were inferior to C04. Therefore, our further analysis focused on C01 and C04 (Figure S3C).

The C04 class contained strong cryo-EM density for RNAylated-nsp9 and the GDP-βS (we designate this structure GDP-Cap-i; Figures 3A and S3C; Table S1). The nominal resolution of the structure was 3.0 Å (Figure S4H), with local features in the NiRAN active site resolved to 2.8–2.9 Å (Figure S4F). In GDP-Cap-i, RNA-nsp9 docks into the NiRAN domain with the GxxxG helix forming an extensive interface with nsp12 and the N terminus of nsp9 loaded into the NiRAN active site (Figures 3A and 3B). The first nucleotide of the nsp9-linked RNA (A1) was well resolved, binding in the same base-up pose as the UMPCPP of UMP-i (Figures 1C and 3B). Like the UMPCPP, the A1 base interacts with NPalm residue N713 and forms a direct and a water-mediated interaction with D711 (Figure 3C). Deeper in the NiRAN active site, the GDP-βS binds in the G-site in the base-in pose (Figure 3D), with the guanine base recognized by the same conserved NiRAN residues as in the GTP-bound NiRAN.28 We modeled a Mg2+-ion density located between the β-phosphate of the GDP-βS and the phosphoramidate bond of RNA-nsp9 (Figures 3B–3D), coordinated by the phosphoramidate phosphate, NiRAN residues D208, N209, and D218, and the GDP-βS β-phosphate (Figure 3C). It is likely critical for catalyzing the deRNAylation/capping activity. There was no cryo-EM density with the correct coordination geometry to allow modeling of a second Mg2+ ion, which would normally stabilize the leaving group29; leaving group stabilization may not be required for this reaction because the leaving group is the protein nsp9.

The second analyzed class, C01 (nominal resolution 3.1 Å; Figure S3C), contained strong cryo-EM density for RNAylated-nsp9, but here the entire 20-mer length of the nsp9-linked RNA, including the predicted SL (Figure S2A), was resolved (Figures S5A and S5B). We term this structure 20-mer-SL. The 5’ and 3’ ends of the SL-RNA interact with the surface of the NiRAN and NPalm (Figure S5C). To accommodate the positioning of the SL, the four 5’-nucleotides of the nsp9-linked RNA (A1-U2-U3-A4) are constrained in a sharp turn, causing a shift in the position of the nsp9 N terminus, nsp9-linked A1, the Mg2+ ion, and other elements of the NiRAN active site (Figures 3C and S5C). Thus, although the nsp9 N terminus, nsp9-linked A1, the Mg2+-ion, and other elements of the NiRAN active site are positioned similarly to the GDP-Cap-i structure (Figures 3C and S5C), GDP is unable to bind, rendering the 20-mer-SL state catalytically inactive. Thus, the catalytically active GDP-Cap-i and inactive 20-mer-SL states are in equilibrium. Surprisingly, the deRNAylation/capping reaction is not measurably slower with this 20-mer-SL RNA compared with the unstructured 10-mer (Figure S2D), suggesting that the folding/unfolding transition of the RNA-SL is not rate limiting.

Chemical mechanism of NiRAN domain catalysis

The NiRAN domain enzymatic activities have been well studied.11,13,14,16 Here, we build on these studies by resolving two
distinct pre-catalytic intermediates of NiRAN-mediated reactions at high-resolution (<3 Å in the NiRAN active site; Figures S1C and S4F). This allows for an atomic-level structural understanding of the chemical mechanisms underlying two NiRAN activities (Figure 4), providing a structure-based platform for therapeutics targeting this essential activity in CoVs.

In NiRAN-mediated NMPylation of nsp9, the N-terminal amine of nsp9 forms a phosphoramidate bond with the α-phosphate of the incoming NTP substrate, releasing PPi.13,14,16 In the NMPylation intermediate (UMP-i), the N-terminal amine of nsp9 is positioned to execute an S_{N2} nucleophilic attack on the α-phosphate of the nucleotide substrate bound in the base-up pose, catalyzed by Mg^{2+}A (Figure 4A). Completion of the reaction results in the formation of a phosphoramidate bond between the nsp9 N terminus and the NTP substrate α-phosphate, hydrolysis of the α-β phosphodiester bond, and a PPi leaving group stabilized by Mg^{2+}B (Figure 4A). This is reminiscent of the two-metal ion mechanism for nucleic acid polymerases.29

In NiRAN-mediated deRNAylation/capping, the phosphoramidate bond between an RNA chain and the nsp9 N-terminal amine is broken by a GDP, resulting in a canonical 5’-RNA GpppA cap (Figure 2C).16 In the RNA capping intermediate GDP-Cap-i, the β-phosphate is positioned to execute an S_{N2} nucleophilic attack on the phosphoramidate phosphorus, catalyzed by Mg^{2+}A (Figure 4B). This results in the formation of the GpppA cap, a broken phosphoramidate bond between nsp9 and the RNA, and the recycling of the nsp9 N-terminal amine to an unmodified state. As previously mentioned, the second Mg^{2+}ion (Mg^{2+}B) for the deRNAylation/capping reaction is not present because the capping reaction does not produce a highly negatively charged PPi leaving group.

catalytic roles (Figure 4). In the initial discovery of NiRAN NMPylation activity, the EAV NiRAN domain was shown to self-NMPylate, covalently attaching an NMP moiety through a phosphoramidate bond to K94 (corresponding to SARS-CoV-2 nsp12 K73). SARS-CoV-2 K73/EAV K94 is absolutely conserved in nidovirus NiRAN domains,11 but in the UMP-i structure, K73 is positioned to interact with the UMPCPP β-phosphate and is not close to the α-phosphate, where it would need to be to become covalently linked (Figure 4A). Although many of the NiRAN active site residues are absolutely conserved among all nidoviruses, the NiRAN domain overall is divergent in sequence; for example, the SARS-CoV-2 and EAV NiRAN domains are only about 10% identical in sequence (among 135 aligned residues).15 Moreover, arterviruses such as EAV do not encode an ortholog to CoV nsp9, so the NiRAN-mediated mechanisms for NMPylation and/or capping must be very different than CoVs. SARS-CoV-2 K50 is absolutely conserved among CoVs but is an arginine in roniviruses and arterviruses (such as EAV). Arginines in enzyme active sites typically play electrostatic stabilization roles rather than catalytic roles due to the poor reactivity of Arg.35 On this basis, we propose that SARS-CoV-2 nsp12 K50 provides a nearby positive charge to stabilize the negatively charged transition state, while K73 plays a more direct role in catalysis, such as a general acid.36

**The NiRAN domain accommodates multiple nucleotide binding poses**

The NiRAN domain has previously been characterized to bind GTP in a base-in binding mode, while binding other nucleotides in a non-specific base-out mode.15,27,28,37 In this study, we identified a distinct base-up binding pose for UMPCPP and A1 in the UMP-i (Figure 1) and GDP-Cap-i (Figure 3) structures, respectively (Figure 5A). The UMPCPP (UMP-i) and A1 (GDP-Cap-i)
bind similarly, interact with the same loop from the NPalm motif (Figure 5A), and are catalytic intermediates relevant to the NiRAN-mediated NMPylation and capping activities, respectively. The base-in pose (Figure 5A) is specific for guanosine nucleotides (GDP-\(\beta\)S, GTP, AT-9010), which are used by the NiRAN domain as capping substrates. The previously described base-out binding mode is likely not relevant to any catalytic activity.

**DISCUSSION**

NiRAN domain function is essential for CoV propagation. The NiRAN domain was initially characterized as a nucleotidyltransferase that, in EAV, self-NMPylates, or, in CoVs, NMPylates the N terminus of nsp9. Subsequently, the enzyme was shown to mediate a remarkable series of reactions—RNAylation of the nsp9 N terminus and deRNAylation/capping—which are critical first steps for the essential process of viral 5’-RNA capping.

Our detailed structural characterization of catalytic intermediates for two of these three known SARS-CoV-2 NiRAN domain enzymatic activities (Figure 5B), NMPylation (Figure 1) and deRNAylation/capping (Figure 3), provides molecular insight into how the NiRAN domain can mediate so many distinct enzymatic activities. Our results provide structural insight into NMPylation and RNA capping intermediates, revealing the determinants of substrate binding and the chemistry catalyzed by the enzyme (Figure 4), plus providing a platform for structure-based drug development.

**A potential role for NiRAN domain mediated NMPylation of nsp9**

Nsp9 has been proposed as a target for broad-spectrum antivirals targeting CoVs, and therefore understanding nsp9 functions, location, and stability is critical. In eukaryotic cells, an N-terminal asparagine is a tertiary destabilizing residue in the N-end rule ubiquitin-dependent pathway. The first enzyme in this pathway, N-terminal asparagine amidohydrolase (NTAN1), deamidates the N-terminal asparagine, converting it into aspartate, a substrate for the arginine-transferase reaction, the target for ubiquitin ligases. This pathway can reduce the half-life of proteins in eukaryotic cells with N-terminal asparagines by over 400-fold.

Nsp9, with its absolutely conserved N-terminal Asn (Figure S1F), is a potential target for the N-end rule pathway. We show that the nsp9 N-terminal Asn can serve as a substrate for hNTAN1, while NMPylation with UTP prevents hNTAN1 deamidation (Figure S6). Given the reversibility of NMPylation, we propose that a possible role for nsp9-NMPylation in linked RNA (GDP-Cap-i structure; Figures 3B and 3C). We determined the base-up pose for the non-hydrolyzable UTP analog UMPCPP, but we propose that all NMPylation proceeds through binding each NTP in a similar base-up pose seen with UMPCPP (Figures 1C and 1E). Promiscuity of the base-up position could be explained by the flexible nucleobase-binding determinants, NiRAN N39, NPalm D711, and NPalm N713, observed in UMP-i and GDP-Cap-i (Figures 1E and 2C).
GTP from 7UOB 28 onto the GDP-Cap-i structure reveals that the providing a role for NMPylation. 43 may serve as a protein primer for SARS-CoV-2 RNA synthesis, capping16 (Figures 2 A and 2B). Superimposing the G-site GTP GDP is the preferred substrate over GTP for deRNAylation/deRNAylation/capping Substrate preference for NiRAN-mediated NMPylation. While this manuscript was being prepared for publication, Schmidt et al., 2023, provided data suggesting that nsp9 may serve as a protein primer for SARS-CoV-2 RNA synthesis, providing a role for NMPylation. 43

Substrate preference for NiRAN-mediated deRNAylation/capping
GDP is the preferred substrate over GTP for deRNAylation/capping 16 (Figures 2A and 2B). Superimposing the G-site GTP from 7UOB28 onto the GDP-Cap-i structure reveals that the GTP γ-phosphate closely approaches the nsp9-linked 5’-terminal phosphate of the RNA (2.4 Å closest approach), creating strong electrostatic repulsion in the absence of a neutralizing divalent cation (Figure S4I), explaining the GDP preference. The structure of a catalytic intermediate with a non-hydrolyzable GTP analog would explain how GTP can function as a (non-preferred) capping substrate 16 (Figures 2A and 2B) and provide further insight into NiRAN domain enzymatic plasticity. The preference for GDP is difficult to understand because the cellular concentration of GTP is expected to greatly exceed that of GDP. 24 We show that the NTPase activity of nsp13 can supply GDP for the deRNAylation/capping reaction (Figures 2A and 2B).

Modeling of a GTP where the GDP-phosphate closely approaches the nsp9-linked 5′0 phosphate and the phosphoramidate bond results in a clash between the GTP γ-phosphate and the phosphoramidate bond of RNA-nsp9 (Figure S4I). Given this clash, a possible explanation for the suboptimal performance of GTP as a substrate (Figures 2A and 2B) could be that the additional phosphate forces unfavorable positioning of the GTP and the RNA-nsp9 substrates. Moreover, the mechanism for hydrolysis of the β-γ phosphodiester bond, a prerequisite for formation of the proper chemical principles; thus, it does not provide insight into the capping mechanism (Figure S4J).

NiRAN-mediated RNAylation
Questions remain regarding the NiRAN-mediated RNAylation reaction (Figure 5B). The NiRAN-mediated NMPylation and capping reactions are promiscuous to multiple divalent cations, particularly Mn2+ and Mg2+,13,14,16 but the RNAylation reaction appears to strictly require Mn2+.13,16 Moreover, the NiRAN-mediated RNAylation strongly prefers an A at the RNA 5′ end.15 The 5′ A of the RNAylation substrate may bind in the same base-up pose as A1 of the nsp9-linked RNA in the GDP-Cap-i structure. A structure of an RNAylation intermediate will be required to confirm this supposition, to understand the unique divalent cation requirement of the RNAylation reaction, and to provide further insight into the NiRAN domain enzymatic plasticity.

Conclusions
In summary, we have provided the structural basis for the NiRAN domain mediated NMPylation of nsp9 as well as proposed a function for the activity in extending the half-life of nsp9. Importantly, we provide the structural basis for the use of GDP as a substrate for deRNAylation/capping. Altogether, these studies provide insight into the binding determinants and poses of the substrates required for two of the three NiRAN-mediated enzymatic activities. Our results demonstrate the determinants of binding and the poses of the substrates to this enigmatic domain. Our results elucidate critical CoV biology while providing a high-resolution platform for structure-based therapeutic targeting and development against the essential SARS-CoV-2 NiRAN domain.

Limitations of the study
One limitation of our study is that our hypothesis that NiRAN-mediated NMPylation of nsp9 serves to protect nsp9 against
degradation is supported indirectly by an in vitro experiment. In vivo analysis of the half-life of nsp9, preferably in SARS-CoV-2-infected cells, and the effect of nsp9 NMPylation on its half-life will ultimately be required to validate this hypothesis. A second limitation is regarding the speculation as to why GDP is a better substrate than GTP for deRNAylation/capping. A high-resolution structure of a capping intermediate trapped with a GTP analog would be required to validate our hypotheses and explain the preference for GDP.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
  - Escherichia coli
- **METHOD DETAILS**
  - Expression and purification of nsp7/8
  - Expression and purification of nsp9
  - Expression and purification of nsp12
  - Expression and purification nsp13
  - Expression and purification of hNTAN1
  - Native mass spectrometry experiments
  - Purification of UMP-nsp9
  - Tandem mass spectrometry to assess hNTAN1 mediated deamidation of nsp9 and UMP-nsp9
  - Synthesis of 5′-triposphorylated RNA oligonucleotides
  - RNAylation of nsp9 time points
  - Purification of RNA-nsp9
  - Comparison of GDP and GTP as substrates for deRNAylation
  - DeRNAylation of RNA-nsp9 time points
  - Cryo-EM sample preparation
  - Cryo-EM grid preparation
  - Cryo-EM data acquisition and processing
  - Model building and refinement
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Gel quantification
  - Quantification used for cryo-EM
  - Mass spectrometry

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2023.10.001.

**ACKNOWLEDGMENTS**

We thank D. Littler for plasmids and helpful discussions, J. Perry for helpful discussions, and A.M. Pyle for materials and insightful comments. We thank M. Ebrahim, J. Sotiris, and H. Ng at the Evelyn Grass Lipper Cryo-electron Microscopy Resource Center at the Rockefeller University for assistance in grid preparation and screening. Some of the work reported here was conducted at the Simons Electron Microscopy Center and the National Resource for Automated Molecular Microscopy and National Center for CryoEM Access and Training located at the NYSBC, supported by grants from the National Institutes of Health (NIH) National Institute of General Medical Sciences (grant no. P41 GM103310), NYSTAR, the Simons Foundation (grant no. SF349247), the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy program (grant no. U24 GM129539), and NY State Assembly Majority. Some of this work was conducted in the Rockefeller University Proteomics Resource Center and thus acknowledges funding from the Leona M. and Harry B. Helmsley Charitable Trust and Sohn Conferences Foundation for mass spectrometer instrumentation. This work was supported by NIH grant nos. P41 GM109824 and P41 GM103314 to B.C. and NIH grant no. R01 AI161278 (to E.A.C. and S.A.D.).

**AUTHOR CONTRIBUTIONS**


**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: June 7, 2023
Revised: August 28, 2023
Accepted: October 3, 2023
Published: October 26, 2023

**REFERENCES**


## STAR+METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> (BL21(DE3) strain)</td>
<td>Novagen</td>
<td>Cat#: 69450</td>
</tr>
<tr>
<td><em>Escherichia coli</em> (Rosetta2 strain)</td>
<td>Novagen</td>
<td>Cat#: 71350</td>
</tr>
<tr>
<td><strong>Chemicals, peptides, and recombinant proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HiLoad 26/600 Superdex 200 pg</td>
<td>Cytiva Life Sciences</td>
<td>Cat#: 28999336</td>
</tr>
<tr>
<td>HisTrap HP 5mL</td>
<td>Cytiva Life Sciences</td>
<td>Cat#: 17524801</td>
</tr>
<tr>
<td>HiTrap Heparin HP 5 mL</td>
<td>Cytiva Life Sciences</td>
<td>Cat#: 17040701</td>
</tr>
<tr>
<td>Superose6 increase 10/300 GL</td>
<td>Cytiva Life Sciences</td>
<td>Cat#: 29091596</td>
</tr>
<tr>
<td>IPTG</td>
<td>Gold Biotechnology</td>
<td>Cat#: 1248</td>
</tr>
<tr>
<td>n-Octyl-b-D-Glucopyranoside</td>
<td>Anatrace</td>
<td>Cat#: 0311</td>
</tr>
<tr>
<td>SARS-CoV-2 nsp7/8</td>
<td>Chen et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>SARS-CoV-2 nsp9</td>
<td>Littler et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>SARS-CoV-2 nsp12</td>
<td>Malone et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>Nsp13</td>
<td>Chen et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>Anhydrotetracycline hydrochloride</td>
<td>Abcam</td>
<td>Cat# ab145350</td>
</tr>
<tr>
<td><strong>Deposited data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coordinates of UMP-i</td>
<td>This paper</td>
<td>PDB: 8SQ9</td>
</tr>
<tr>
<td>Coordinates of GDP-Cap-i</td>
<td>This paper</td>
<td>PDB: 8SQK</td>
</tr>
<tr>
<td>Coordinates of 20mer-SL</td>
<td>This paper</td>
<td>PDB: 8SQJ</td>
</tr>
<tr>
<td>CryoEM map of UMP-i</td>
<td>This paper</td>
<td>EMD: 40699</td>
</tr>
<tr>
<td>CryoEM map of GDP-Cap-i</td>
<td>This paper</td>
<td>EMD: 40708</td>
</tr>
<tr>
<td>CryoEM map of 20mer-SL</td>
<td>This paper</td>
<td>EMD: 40707</td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UMP-i primer strand:</td>
<td>IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>S′cguguagcaugcuacGUCAUUC UCcaccgcgaagC&lt;sup&gt;α&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UMP-i template strand:</td>
<td>IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>S′gcgcuaugcaugcAGUAAGACG GugcuucguucguuguuuUACCC CUAUC&lt;sup&gt;α&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP-Cap-i primer strand:</td>
<td>IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>S′cguguagcaugcuacGUCAUUC UCcaccgcgaagC&lt;sup&gt;α&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP-Cap-i template strand:</td>
<td>IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>S′gcgcuaugcaugcAGUAAGACG GugcuucguucguuguuuUACCC CUAUC&lt;sup&gt;α&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20mer-SL primer strand:</td>
<td>IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>S′cguguagcaugcuacGUCAUUC UCcaccgcgaagC&lt;sup&gt;α&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20mer-SL template strand:</td>
<td>IDT</td>
<td>N/A</td>
</tr>
<tr>
<td>S′gcgcuaugcaugcAGUAAGACG GugcuucguucguuguuuUACCC CUAUC&lt;sup&gt;α&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10mer pppRNA from SARS-CoV-2 5′ UTR: AUAUAAAGGUU</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>20mer pppRNA from SARS-CoV-2 5′ UTR: AUUAAAGGUU AUAUACUUC</td>
<td>This paper</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(Continued on next page)
RESOURCE AVAILABILITY

Lead contact
More information and requests for resources and reagents should be directed to the lead contact, Elizabeth Campbell (campbee@rockefeller.edu).

Materials availability
All unique materials generated in this study are available by contacting the lead contact.

Data and code availability
- The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank under accession codes EMD-40699 (UMP-i), EMD-40708 (GDP-Cap-i), and EMD-40707 (20mer-SL). The atomic coordinates have been deposited in the Protein Data Bank under accession codes 8SQ9 (UMP-i), 8SQK (GDP-Cap-i), 8SQJ (20mer-SL). These datasets are publicly available as of the date of publication. Raw images of SDS-PAGE gels have been deposited with Mendeley at https://doi.org/10.17632/btbjtn2ff5.1.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Escherichia coli
Two strains of *E. coli* were used in this study, BL21(DE3) or Rosetta(DE3). *E. coli* was grown at 30 °C or 37 °C in LB media or on LB-agar plates. Antibiotics and small molecules used during growth and expression were used at the following concentrations: streptomycin at 50 μg/mL, kanamycin at 50 μg/mL, chloramphenicol at 25 μg/mL, tetracycline at 10 ng/mL, nalidixic acid at 50 μg/mL, and isopropyl 

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>33mer pppRNA from SARS-CoV-2 5' UTR: AUUAAAGGUUUAAUCCCUCUCCCA GGUAAACAAACC</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Recombinant DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nsp7/8 overexpression plasmid</td>
<td>Chen et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>N-terminal SumoHis-tagged nsp9 overexpression plasmid</td>
<td>Littler et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>N-terminal Sumo-His tagged nsp12 overexpression plasmid coexpressing nsp7/8 and chaperones</td>
<td>Malone et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>N-terminal HisPPX-tagged nsp13 overexpression plasmid</td>
<td>Chen et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>N-terminal biotin and FLAG-tagged hNTAN1 overexpression plasmid</td>
<td>Cantor et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>Software and algorithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian Polishing</td>
<td>Zivanov et al.</td>
<td><a href="https://github.com/3dem/relion">https://github.com/3dem/relion</a></td>
</tr>
<tr>
<td>CryoSPARC</td>
<td>Punjani et al.</td>
<td><a href="https://cryosparc.com/">https://cryosparc.com/</a></td>
</tr>
<tr>
<td>Coot</td>
<td>Emsley and Cowtan</td>
<td><a href="https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot">https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot</a></td>
</tr>
<tr>
<td>MotionCor2</td>
<td>Zheng et al.</td>
<td>N/A</td>
</tr>
<tr>
<td>The PyMOL molecular graphics system</td>
<td>Schrodinger LLC</td>
<td><a href="https://pymol.org/2/">https://pymol.org/2/</a></td>
</tr>
<tr>
<td>RELION</td>
<td>Zivanov et al.</td>
<td><a href="https://github.com/3dem/relion">https://github.com/3dem/relion</a></td>
</tr>
<tr>
<td>Graphpad Prism</td>
<td>Graphpad</td>
<td><a href="http://www.graphpad.com/scientific-software/prism">http://www.graphpad.com/scientific-software/prism</a></td>
</tr>
<tr>
<td>UCSF ChimeraX</td>
<td>Pettersen et al.</td>
<td><a href="https://www.cgl.ucsf.edu/chimerax/">https://www.cgl.ucsf.edu/chimerax/</a></td>
</tr>
<tr>
<td>PHENIX</td>
<td>Afonine et al.</td>
<td><a href="https://phenix-online.org/">https://phenix-online.org/</a></td>
</tr>
</tbody>
</table>
β-D-1-thiogalactopyranoside (IPTG) variously at 0.1, 0.2, 0.5, and 1 mM. Small (5mL) liquid cultures were grown standing in tissue culture flasks, while larger (50 mL or greater) cultures were grown shaking in Erlenmeyer flasks at 120 rpm.

METHOD DETAILS

Expression and purification of nsp7/8
SARS-CoV-2 nsp7/8 was expressed and purified as previously reported. In brief, a pCDFDuet-1 plasmid containing His6-PPX-nsp7/8 (Addgene: 159092) was transformed into E. coli BL21(DE3) (Novagen) and plated on LB-agar containing 50 µg/mL streptomycin. Cells were grown in LB media supplemented with 10 µM ZnCl2, grown to \(OD_{600} = 0.6\) at 30°C at 200 rpm, induced with 0.1 mM IPTG, then incubated for 14 hours at 16°C. Cells were collected via centrifugation, resuspended in 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.1 mM EDTA-NaOH pH 8.0, 5 mM imidazole, 5% glycerol (v/v), 10 µM ZnCl2, 1 mM BME, 1x Protease Inhibitor Cocktail (Roche), 1 mM PMSF, and lysed through a French Press (Avestin). The cleared lysate from centrifugation was loaded on a HisTrap HP column (Cytiva), washed, and eluted. The eluate was dialyzed overnight with Precission Protease to cleave the His6-tag. The cleaved proteins were passed through another HisTrap HP column, and the resulting flow-through was injected onto a Superdex 75 Hilo 16/600 (Cytiva) for size-exclusion chromatography. Glycerol was added to purified nsp7/8 to reach 20% final concentration, aliquoted, flash-frozen with liquid N2, and stored at -80°C until use.

Expression and purification of nsp9
SARS-CoV-2 nsp9 was produced with a physiological N terminus using a synthetic gene described previously. The plasmid was transformed into BL21 (DE3) cells and plated on LB-agar containing 50 µg/mL kanamycin and grown in LB media to \(OD_{600} = 0.6\) at 37°C at 200 rpm, induced with 0.5 mM IPTG, then incubated for 4 hours at 30°C. Cells were collected via centrifugation, resuspended in 20 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole, 5% glycerol (v/v), 1 mM BME, 1x Protease Inhibitor Cocktail (Roche), 1 mM PMSF, and lysed through a French Press (Avestin). The cleared lysate from centrifugation was loaded on a HisTrap HP column (Cytiva), washed, and eluted in 20 mM HEPES pH 8.0, 300 mM NaCl, 250 mM imidazole, 5% glycerol, and 1 mM BME. The eluate was dialyzed overnight with Ulp1 to cleave the His6-affinity tag. The cleaved protein was injected onto a Superdex 75 Hilo 16/600 (Cytiva) for size-exclusion chromatography and fractions of interest were passed through another HisTrap HP column to isolate tag-free proteins. Glycerol was added to the isolated protein to reach 20% final concentration, aliquoted, flash-frozen with liquid N2, and stored at -80°C until use.

Expression and purification of nsp12
SARS-CoV-2 nsp12 was expressed53 and purified28 as previously reported. In brief, a pQE-30/pcl-ts ind+ plasmid containing a His6-PEP-like modifier (SUMO) SARS-CoV-2 nsp12 and untagged nsp7 and 8 (Addgene no. 160540) was transformed into Escherichia coli BL21(DE3) cells (Novagen). Cells were grown and protein expression was induced by the addition of 0.2 mM IPTG, 10 ng/ml tetracycline and 50 µg/mL kanamycin and grown in LB media to \(OD_{600} = 0.6\) at 37°C at 200 rpm, induced with 0.5 mM IPTG, then incubated for 4 hours at 30°C. Cells were collected by centrifugation, resuspended in 20 mM HEPES pH 8.0, 100 mM NaCl, 5% glycerol (v/v), 1 mM DTT, and lysed in a French press (Avestin). The lysate was cleared by centrifugation and purified on a HiTrap Heparin HP column (Cytiva). The fractions containing nsp12 were loaded onto a HiTrap HP column (Cytiva) for further purification. Eluted nsp12 was dialyzed, cleaved with His6-Ulp1 SUMO protease, and passed through a HiTrap HP column to remove the SUMO protease. Flow-through was collected, concentrated by centrifugal filtration (Amicon) and loaded on a Superdex 200 Hilo 16/600 (Cytiva). Glycerol was added to the purified nsp12 to a final concentration of 20%, aliquoted, flash-frozen with liquid N2 and stored at -80°C until use.

Expression and purification of nsp13
SARS-CoV-2 nsp13 was expressed and purified as previously reported. In brief, a pet28 plasmid expressing His6-nsp13 was transformed into E. coli Rosetta (DE3) cells (Novagen) and plated on LB-agar containing 50 µg/mL kanamycin and 25 µg/mL chloramphenicol. Single colonies were used to inoculate liquid LB cultures with kanamycin and chloramphenicol. Cells were grown at 37°C, induced at 0.6 OD600 by the addition of IPTG (0.2 mM final), then incubated for 17 hr at 16°C. Cells were collected by centrifugation, resuspended in 50 mM HEPES-NaOH, pH 8.0, 500 mM NaCl, 5 mM MgCl2, 5% (v/v) glycerol, 20 mM imidazole, 5 mM BME, 1 mM ATP, 1 mM PMSF and lysed in a French press (Avestin). The lysate was cleared by centrifugation then purified on a HiTrap HP column. Eluted nsp13 was dialyzed overnight into 50 mM HEPES-NaOH pH 8.0, 500 mM NaCl, 5 mM MgCl2, 5% (v/v) glycerol, 20 mM imidazole, 5 mM BME in the presence of His6-Prescission Protease to cleave the His6-tag. Cleaved nsp13 was passed through a HiTrap HP column and the flow-through was collected, concentrated by centrifugal filtration (Amicon), and loaded onto a Superdex 200 Hilo 16/600 (GE Healthcare) in 20 mM HEPES pH 8.0, 500 mM NaCl, 5 mM MgCl2, 5% glycerol, 1 mM DTT). Glycerol was added to purified nsp13 to reach 20% final concentration, aliquoted, and flash-frozen with liquid N2, and stored at -80°C until use.

Expression and purification of hNTAN1
Human NTAN1 was expressed and purified as previously reported. In brief, the plasmid, pNTAN2FLAGSII, was transformed into BL21 (DE3) cells and plated on LB-agar containing 50 µg/mL kanamycin and grown in LB media to \(OD_{600} = 0.6\) at 37°C at 200 rpm, induced with 1 mM IPTG, then incubated for 16 hours at 25°C. Cells were collected via centrifugation, resuspended in 200 mL HEPES-NaOH pH 8.0, 500 mM NaCl, 5 mM MgCl2, 5% glycerol, 1 mM DTT. Glycerol was added to purified nsp13 to reach 20% final concentration, aliquoted, and flash-frozen with liquid N2, and stored at -80°C until use.
100 mM Tris-Cl pH 7.5, 150 mM NaCl, 1 mM EDTA, and lysed in a French press (Avestin). The lysate was cleared by centrifugation and applied to a 5 mL column ( Kontes) packed with streptactin superfro high capacity resin. This was followed by purification with anti-FLAG M2 magnetic beads and elution using 100 μg/mL FLAG peptide then overnight dialysis into 50 mM Tris-Cl pH 7.5, 150 mM NaCl, 1 mM DTT. The protein was used fresh the following day.

**Native mass spectrometry experiments**

**NMPylation sample prep**

Nsp9 (10 μM) was incubated with nsp12 (0.05 μM) and one of four NTPs (200 μM) for 1.5 hours at 37°C in 50 mM HEPES pH 8, 15 mM KAc, 2 mM MgCl₂, 1 mM DTT before flash freezing the samples.

**Capping sample prep**

RNA-nsp9 constructs (4 μM) were incubated with nsp12 (0.5 μM) and either GDP or GTP (500 μM) at 37°C in 50 mM HEPES pH 8, 15 mM KAc, 2 mM MgCl₂, 1 mM DTT for 4 hours before flash freezing the samples.

**nMS analysis**

The frozen samples were thawed then buffer exchanged into either 300 mM or 500 mM ammonium acetate solution for the nucleotidylation or deRNAylation samples, respectively, using Zeba microspin desalting columns with a 40-kDa MWCO (Thermo Scientific). The nMS solutions contained 0.01% Tween-20. The sample concentrations used ranged from 2 – 4 μM. For nMS analysis, 2 – 3 μL of each sample was loaded into a gold-coated quartz capillary tip that was prepared in-house and then electrosprayed into an Exactive Plus with extended mass range (EMR) instrument (Thermo Fisher Scientific) with a static direct infusion nanospray source. The MS parameters used included: spray voltage, 1.2 kV; capillary temperature, 150 – 200°C; in-source dissociation, 0 V; S-lens RF level, 200; resolving power, 8,750 at m/z of 200; AGC target, 1 x 10⁶; maximum injection time, 200 ms; number of microscans, 5; injection flatapole, 8 V; interflatapole, 7 V; bent flatapole, 5 V; high energy collision dissociation (HCD), 45 V; ultrahigh vacuum pressure, 5 – 6 x 10⁻¹¹ mbar; total number of scans, at least 100. Mass calibration in positive EMR mode was performed using cesium iodide.

For data processing, the acquired nMS spectra were visualized using Thermo Xcalibur Qual Browser (v. 4.2.47). Data processing and spectra deconvolution were performed using UniDec version 4.2.0. The UniDec parameters used were m/z range: 1,000 – 4,000; background subtraction: subtract curved at 10; mass range: 5,000 – 30,000 Da; sample mass every 1 Da; smooth charge state distribution, on; peak shape function, Gaussian; and Beta softmax function setting, 20. The measured masses were typically off by 1 Da relative to the expected mass. The expected masses for nsp9 and the 20-mer 5'-pppRNA are 12,378 Da and 6,512 Da, respectively.

**Purification of UMP-nsp9**

Nsp9 (29 μM) was incubated with nsp12 (2 μM) and UTP (10 μM) for 30 minutes at 30°C in 50 mM HEPES pH 8, 1 mM MnCl₂, 5 mM DTT. UMP-nsp9 was then purified from nsp12 and excess UTP over a Superox 6 Increase 10/300 GL column (Cytiva) in 20 mM HEPES pH 8, 120 mM KAc, 10 mM MgCl₂, 2 mM DTT and eluted in 2 peaks that were assessed via native mass spectrometry to confirm modification.

**Tandem mass spectrometry to assess hNTAN1 mediated deamidation of nsp9 and UMP-nsp9**

Nsp9 or UMP-nsp9 (5 μM) were incubated with or without hNTAN1 for 1 hour at 37°C in 50 mM Tris-Cl pH 7.5, 150 mM NaCl, 1 mM DTT then samples were flash frozen. Protein samples were trypsinized (Promega) in 40 mM ammonium bicarbonate (FISHER SCIENTIFIC). Fractions of each digest were analyzed by LC-MS/MS. Mass spectrometer operated in high resolution/high mass accuracy mode (Orbitrap Fusion LUMOS or Q-Exactive HF, Thermo). Mass spectrometer was operated in a hybrid date dependent acquisition (DDA)/parallel reaction monitoring (PRM) mode. The following doubly charged peptides were targeted: NELSPVALR; [N(ump)NELSPVALR; (N(deam)NELSPVALR; [N(ump)[N(deam)ELSPVALR and N(deam)[N(deam)NELSPVALR, where ‘ump’ is uridine monophosphate (C9H11N2O8P, m/z of +306.025302 Da) and ‘deam’ is deamidation (H-1N-10, Δm of + 0.984016). Data were searched against a custom database containing NSP9 concatenated to a background proteome. ProteomeDiscoverer/Mascot were used to query the data. Tandem MS spectra were manually validated to assure the position of the deamidated residues. Signals of relevant peptides were extracted using Skyline(-daily) v/22.2.1.425.

**Synthesis of 5’-triphosphorylated RNA oligonucleotides**

5’-triphosphorylated RNA oligonucleotides were synthesized on the MerMade 12 RNA-DNA synthesizer (Bioautomation) as previously described (1). Base and 2’-hydroxyl group deprotection and subsequent purification was carried out essentially as described. Briefly, for base and phosphate group deprotection and removal of the oligonucleotide from the support, the polymer support was transferred to a glass vial and incubated with 4 ml of the mixture of 28%–30% aqueous ammonium hydroxide (JT Baker) and 40% aqueous methylamine (MiliporeSigma) (1:1) at 65°C for 15 minutes. Then the solution was cooled on ice for 10 minutes, transferred to a clean vial, incubated at -80°C for 1 hour and evaporated to dryness using SpeedVac. Then 500 μL of anhydrous ethanol was added and the mixture was evaporated again to eliminate all traces of water. In order to deprotect 2’-hydroxyl groups, 500 μL of 1M tetra-butylammonium fluoride in tetrahydrofuran was added and the mixture was incubated at room temperature for 36 hours. Then 500 μL
of 2M sodium acetate, pH 6.0 was added and the resulting mixture was evaporated in the SpeedVac until the volume was reduced by half. The mixture was then extracted three times with 800 µL of ethyl acetate and evaporated in the SpeedVac for 15 minutes followed by an overnight precipitation at -20°C with 1.6 ml of ethanol. Then the oligonucleotides were dissolved in 750 µl of sterile water, desalted using GlenPak 1.0 columns (Glen Research), ethanol-precipitated again, dissolved in 200 µl of the RNA storage buffer (20 mM MOPS, pH 6.5, 1 mM EDTA) and stored at -80°C.

**RNAylation of nsp9 time points**

Nsp9 (20 µM) was incubated with nsp12/7/8 (2 µM) and a pppRNA oligo (100 µM) at 37°C for the listed time points from 1 minute to 1 hour in 50 mM HEPES pH 8, 15 mM KAc, 4 mM MnCl₂, 2 mM DTT, 1 mM TCEP and 2.5 units/mL of YIPP (NEB). At each time point reaction was quenched with 4X LDS and 50 mM DTT then run on SDS-PAGE. Gels were imaged then analyzed using ImageJ to quantify the reaction efficiencies. Three independent experiments were performed for each pppRNA construct and Prism (GraphPad Software) was used to graph the results.

**Purification of RNA-nsp9**

Nsp9 (65 µM) was incubated with nsp12/7/8 (2 µM) and a pppRNA oligo (130 µM) at 37°C for 3 hours in 50 mM HEPES pH 8, 15 mM KAc, 4 mM MnCl₂, 2 mM DTT, 1 mM TCEP and 2.5 units/mL of YIPP (NEB). RNA-nsp9 product was then purified from the protein and leftover RNA components through a 1 mL Hitrap Q Fast Flow column in a buffer of 20 mM HEPES pH 7.5, 1 mM DTT with a salt gradient. Eluted product was concentrated, and buffer exchanged into buffer of 50 mM HEPES pH 8, 15 mM KAc, 1 mM DTT and flash-frozen with liquid N2, and stored at -80°C until use.

**Comparison of GDP and GTP as substrates for deRNAylation**

RNA-nsp9 constructs (4 µM) were incubated with nsp12 (0.5 µM) and either GDP or GTP (100 or 500 µM depending on the reaction) at 37°C in 50 mM HEPES pH 8, 15 mM KAc, 2 mM MgCl₂, 1 mM DTT. For the reactions including nsp13 (0.5 µM), nsp13 was added to the reaction prior to the addition of GTP. After 15 minutes, reaction was quenched with 4X LDS and 50 mM DTT then run on SDS-PAGE. Gels were imaged then analyzed using ImageJ to quantify the reaction efficiencies. Three independent experiments were performed for each pppRNA construct and Prism (GraphPad Software) was used to graph the results.

**Cryo-EM sample preparation**

Cryo-EM samples of SARS-CoV-2 RTC were prepared as previously described.²⁸,³⁷ In brief, purified nsp12 and nsp7/8 were mixed in a 1:2.5 molar ratio and incubated at room temperature for 20 minutes at 22°C. An annealed RNA scaffold (Horizon Discovery, Ltd) was added to nsp12/7/8 and incubated at 30°C for 30 minutes. The complex was then buffer exchanged into 20 mM HEPES pH 8, 80 mM KAc, 2 mM MgCl₂, 2 mM DTT, incubated at 30°C for 30 minutes again, then purified over a Superose 6 Increase 10/300 GL column (Cytiva). The eluted nsp12/7/8/RNA complex was pooled and concentrated by centrifugal filtration (Amicon).

**Cryo-EM grid preparation**

**UMPylation intermediate**

Before freezing grids beta-octyl-glucoside detergent (β-OG), nsp9, and UMPCPP (JenaBiosciences) were added to sample for a final buffer condition at time of freezing of 20 mM HEPES-NaOH, pH 8.0, 80 mM K-acetate, 2 mM MgCl₂, 2 mM DTT, 0.07% β-OG, 500 µM UpCpp. Final RTC and nsp9 concentrations were 36 and 90 µM respectively. C-flat holey carbon grids (CF-1.2/1.3-4Au, Protochips) were glow-discharged for 20 s prior to the application of 3.5 µL of sample. Using a Vitrobot Mark IV (Thermo Fisher Scientific), grids were blotted and plunge-froze into liquid ethane at 90% chamber humidity at 4°C.

**DeRNAylation/mRNA capping intermediate**

Before freezing grids beta-octyl-glucoside detergent (β-OG), RNA-nsp9, and GDP-β-S (JenaBiosciences) were added to sample for a final buffer condition at time of freezing of 20 mM HEPES-NaOH, pH 8.0, 80 mM K-acetate, 2 mM MgCl₂, 2 mM DTT, 0.07% β-OG, 500 µM GDP-β-S. Final RTC and RNA-nsp9 concentrations were 24 and 39 µM respectively. C-flat holey carbon grids (CF-1.2/1.3-4Au, Protochips) were glow-discharged for 20 s prior to the application of 3.5 µL of sample. Using a Vitrobot Mark IV (Thermo Fisher Scientific), grids were blotted and plunge-froze into liquid ethane at 90% chamber humidity at 4°C.

**Cryo-EM data acquisition and processing**

**UMP-i**

Grids were imaged using a 300 kV Titan Krios (Thermo Fisher Scientific) equipped with a GIF BioQuantum and K3 camera (Gatan). Movies were collected with Leginon⁶⁰ with a pixel size of 1.0825 Å per px (micrograph dimensions of 5,760 x 4,092 px) over a
defocus range of 0.7 to 2.5 μm with a 20 eV energy filter slit. Videos were recorded in counting mode (native K3 camera binning 2) with ~25 e−/Å²/s exposure rate in dose-fractionation mode with intermediate frames recorded every 50 ms over a 2 s exposure (40 frames per micrograph) to give an electron exposure of ~51 e−/Å². Dose-fractionated videos were gain-normalized, drift-corrected, summed and dose-weighted using MotionCor2.48 The CTF was estimated for each summed image using the Patch CTF module in cryoSPARC v3.2.0.46 Particles were picked and extracted from the dose-weighted images with box size of 256 px using cryoSPARC blob picker and particle extraction. The entire dataset consisted of 21,859 motion-corrected images with 10,257,549 particles. Particles containing the RTC were sorted from junk particles using cryoSPARC heterogeneous refinement using a template of an RTC monomer28 and junk classes that pulled out particles containing nsp9, not found in the template (Figure S1A). This was followed by further particle curation using iterative heterogeneous refinement (n=4), resulting in 1,190,409 curated particles. These particles were refined using cryoSPARC local and global CTF refinements as well as non-uniform refinements (Figure S1A). Particles were further processed through 2 rounds of RELION v.3.1 Bayesian polishing.45 Polished particles were refined using cryoSPARC non-uniform refinements then masked (Figure S1A) around the NiRAN domain and nsp9 then subtracted outside the mask for cryoSPARC local refinements around the NiRAN domain active site. Locally refined maps were combined with the consensus maps into a composite map for each class using PHENIX ‘combine focused maps’ to aid model building.50 Local resolution calculations were generated using blocros and blocfilt from the Bsoft package61 (Figure S1C). Angular distribution of particle orientations (Figure S1B) and directional resolution (Figure S1D), calculated through the 3DFSC package42 are shown for the final class. GSFSC, calculated through cryoSPARC, and map-model FSC, calculated with Phenix mtriage, are shown for the final class (Figure S1E).

**GDP-Cap-i and 20mer-SL**

Grids were imaged using a 300 kV Titan Krios (Thermo Fisher Scientific) equipped with a GIF BioQuantum and K3 camera (Gatan). Movies were collected with Leginon60 with a pixel size of 1.0825 Å per px (micrograph dimensions of 5,760 × 4,092 px) over a defocus range of 0.8 to 1.8 μm with a 20 e V energy filter slit. Videos were recorded in counting mode (native K3 camera binning 2) with ~25 e−/Å²/s exposure rate in dose-fractionation mode with intermediate frames recorded every 40 ms over a 2 s exposure (50 frames per micrograph) to give an electron exposure of ~51 e−/Å². Dose-fractionated videos were gain-normalized, drift-corrected, summed and dose-weighted using MotionCor2.48 The CTF was estimated for each summed image using the Patch CTF module in cryoSPARC v3.3.2.46 Particles were picked and extracted from the dose-weighted images with box size of 256 px using cryoSPARC blob picker and particle extraction. The entire dataset consisted of 18,565 motion-corrected images with 8,893,273 particles. Particles containing the RTC were sorted from junk particles using cryoSPARC heterogeneous refinement using a template of an RTC monomer28 and junk classes that pulled out particles containing nsp9, not present in the template (Figure S3C). This was followed by further particle curation using iterative heterogeneous refinement (n=5), resulting in 1,005,706 curated particles. These particles were refined using cryoSPARC local and global CTF refinements as well as non-uniform refinements (Figure S3C). Particles were further processed through RELION v.3.1 Bayesian polishing.45 Polished particles were then classified using cryoSPARC v4.0.2 3D classification into 6 classes. The particles in these classes were reextracted to a box size of 256 px and processed through 2 rounds of RELION v.3.1 Bayesian polishing.45 Polished particles were then masked around the NiRAN domain and nsp9 then subtracted outside the mask (Figure S3C). Subtracted particles were then classified using cryoSPARC v4.0.2 3D classification into 6 classes. The particles in these classes were reextracted to a box size of 256 px and processed through 2 rounds of RELION v.3.1 Bayesian polishing.45 Polished particles were then masked around the NiRAN domain and nsp9 then subtracted outside the mask for cryoSPARC local refinements around the NiRAN domain active site. Locally refined maps were combined with the consensus maps into a composite map for each class using PHENIX ‘combine focused maps’ to aid model building.50 Two classes were selected as the primary classes of interest for model building and analysis: 20mer-SL and GDP-Cap-i. These classes produced structures with following particle counts and nominal resolutions from cryoSPARC: 20mer-SL (112,678 particles, 3.06 Å) and GDP-Cap-i (113,000 particles, 3.01 Å). Local resolution calculations were generated using blocros and blocfilt from the Bsoft package61 (Figures S4B and S4F). Angular distribution of particle orientations (Figures S4A and S4E) and directional resolution (Figures S4C and S4G), calculated through the 3DFSC package42 are shown for the final class. GSFSC (Figures S4D and S4H), calculated through cryoSPARC, and map-model FSC, calculated with Phenix mtriage, are shown for the final class (Figures S4D and S4H).

**Model building and refinement**

An initial model of the RTC-nsp9 was derived from Protein Data Bank (PDB) 7CYQ.27 The models were manually fit into the cryo-EM density maps using ChimeraX29 and rigid-body and real-space refined using PHENIX real-space-refine.50 For real-space refinement, rigid-body refinement was followed by all-atom and B factor refinement with Ramachandran and secondary structure restraints. Models were inspected and modified in Coot v.0.9.717 and the refinement process was repeated iteratively.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

**Gel quantification**

Protein bands were quantified using ImageJ software (NIH) in Figures 2A, S2C, and S2E. Error bars are the standard error of the mean of three independent experiments.
Quantification used for cryo-EM
For calculations of Fourier shell correlations (FSC) in Figures S1E, S4D, and S4H, the FSC cut-off criterion of 0.143 was used. Quantifications and statistical analyses for model refinement and validation were calculated in PHENIX.

Mass spectrometry
Data processing and spectra deconvolution for nMS were performed using UniDec version 4.2.0. For MS/MS signals of relevant peptides were extracted using Skyline(-daily) v/22.2.1.425.