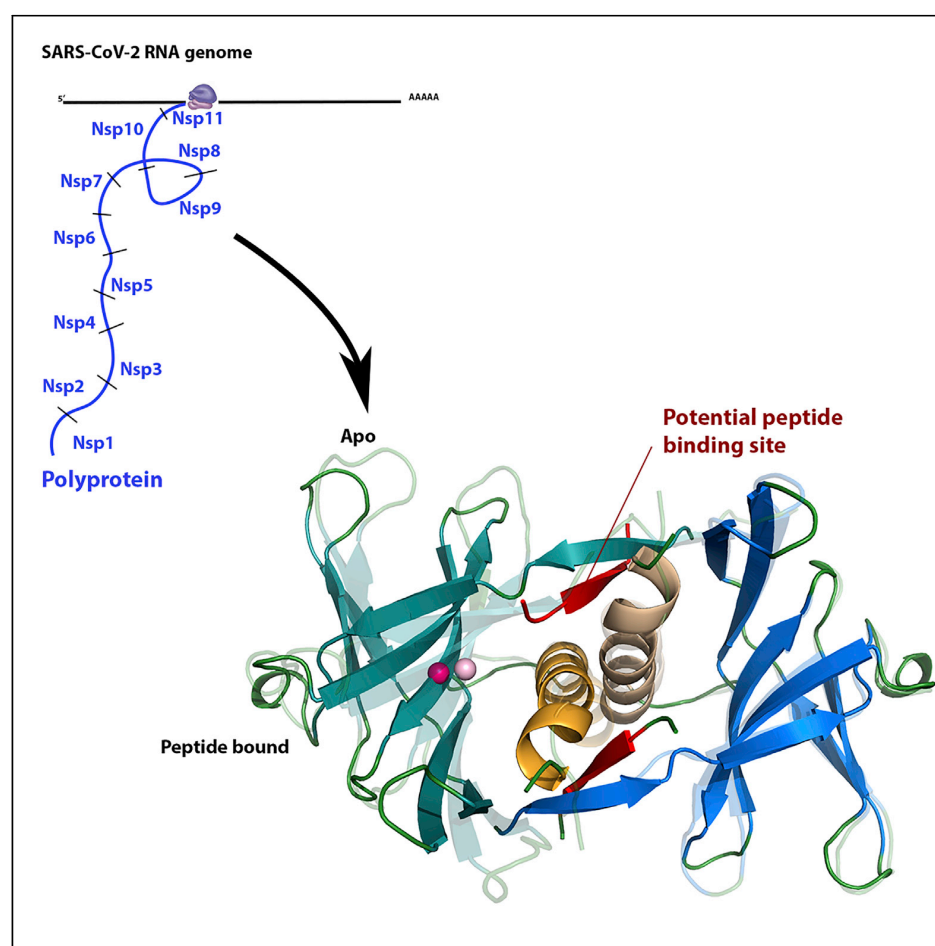# iScience

Article

Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9

Dene R. Littler,
Benjamin S. Gully,
Rhys N. Colson,
Jamie Rossjohn

dene.littler@monash.edu
(D.R.L.)
jamie.rossjohn@monash.edu
(J.R.)

HIGHLIGHTS

The SARS-CoV-2 Nsp9 protein is structurally similar to SARS-CoV

Dimerization of the coronaviral Nsp9 proteins is known to be required for its function

Oligomerization is mediated by an unusual GxxxG protein-protein interaction interface

A cavity near this Nsp9 GxxxG interaction interface may be able to bind peptides

Article

Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9

Dene R. Littler,[1,2,*] Benjamin S. Gully,[1,2] Rhys N. Colson,[1] and Jamie Rossjohn[1,2,3,4,*]

SUMMARY

Many of the SARS-CoV-2 proteins have related counterparts across the Severe Acute Respiratory Syndrome (SARS-CoV) family. One such protein is non-structural protein 9 (Nsp9), which is thought to mediate viral replication, overall virulence, and viral genomic RNA reproduction. We sought to better characterize the SARS-CoV-2 Nsp9 and subsequently solved its X-ray crystal structure, in an apo form and, unexpectedly, in a peptide-bound form with a sequence originating from a rhinoviral 3C protease sequence (LEVL). The SARS-CoV-2 Nsp9 structure revealed the high level of structural conservation within the Nsp9 family. The exogenous peptide binding site is close to the dimer interface and impacted the relative juxtapositioning of the monomers within the homodimer. We have established a protocol for the production of SARS-CoV-2 Nsp9, determined its structure, and identified a peptide-binding site that warrants further study to understanding Nsp9 function.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) comprises a large single-stranded positive polarity RNA genome that acts as messenger RNA after entering the host. The 5′ two-thirds of the genome encodes a large polyprotein that is translated into ORF1a and ORF1ab through host ribosomal frameshifting, with the remainder of the viral RNA encoding structural and accessory proteins within smaller ORFs. The viral proteins necessary for host cell infection such as the RNA polymerase along with enzymes that facilitate RNA synthesis are largely contained within the SARS-CoV-2 polyproteins and are released by the action of two internally encoded proteases. The mature proteins thus released are known as non-structural proteins (Nsps) as they are not incorporated within the virion particles. Owing to their degree of sequence conservation the enzymatic roles and essentiality of each of the Nsps within SARS-CoV-2 is likely to mimic the behavior of homologous proteins within previously studied coronaviruses such as SARS-CoV.

The development of therapeutic interventions against SARS-CoV-2 infection has focused on a number of approaches: vaccination strategies that target the structural spike glycoprotein of the envelope (Wrapp et al., 2020) and may also include a larger selection of viral proteins (Thanh Le et al., 2020), whereas small-molecule compounds have predominantly targeted two conserved viral enzymes, the main protease (Zhenming et al., 2020) (Yang et al., 2005) and the RNA-polymerase (Yan et al., 2020). Nevertheless, some of the betacoronaviral non-structural proteins appear necessary for viral replication within SARS-CoV and influence pathogenesis (Frieman et al., 2012). Despite their close homology between viruses, such non-structural proteins remain of interest as they may have conserved roles within the viral life cycle of SARS-CoV-2 that could be susceptible to inhibition.

During infection of human cells, SARS-CoV Non-structural protein 9 (Nsp9$_{SARS}$) was found to be essential for replication (Frieman et al., 2012). Homologs of the Nsp9 protein have been identified in numerous coronaviruses, including SARS-Cov-2 (Nsp9$_{COV19}$), human coronavirus 229E (Nsp9$_{HcoV}$), avian infectious bronchitis virus (Nsp9$_{IBV}$), porcine epidemic diarrhea virus (Nsp9$_{PEDV}$), and porcine delta virus (Nsp9$_{PDCoV}$). Nsp9$_{SARS}$ has been shown to have modest affinity for long oligonucleotides with binding thought to be dependent on oligomerization state (Egloff et al., 2004) (Sutton et al., 2004). Nsp9$_{SARS}$ dimerizes in solution via a conserved α-helical "GxxG" motif. Disruption of key residues within this motif reduces both RNA binding (Sutton et al., 2004) and SARS-CoV viral replication (Frieman et al., 2012). The mechanism of RNA binding within the Nsp9 protein family is not understood as these proteins have an unusual structural fold not previously seen in RNA-binding proteins (Egloff et al., 2004) (Sutton et al., 2004). The fold's Greek-key motif exhibits

[1]Infection and Immunity Program, Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia

[2]Australian Research Council Centre of Excellence for Advanced Molecular Imaging, Monash University, Clayton, VIC, Australia

[3]Institute of Infection and Immunity, Cardiff University School of Medicine, Heath Park, Cardiff, UK

[4]Lead Contact

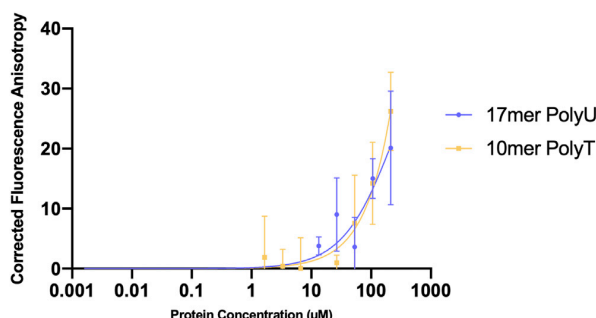*Correspondence: dene.littler@monash.edu (D.R.L.), jamie.rossjohn@monash.edu (J.R.) https://doi.org/10.1016/j.isci.2020.101258

Figure 1. Nsp9$_{COV19}$ Nucleotide-Binding Assay
Fluorescence polarization anisotropy assays were used to examine the possibility that Nsp9$_{COV19}$ could bind to labeled 17-mer and 10-mer single-stranded oligonucleotides. The plot shows corrected anisotropy for each Nsp9$_{COV19}$ protein concentration; error bars represent the SD from the mean of triplicate measurements after 60 min incubation.

topological similarities with Oligonucleotide/oligosaccharide binding proteins (OB-fold), but such vestiges have proven insufficient to provide clear insight into Nsp9 function (Egloff et al., 2004). As a consequence of the weak affinity of Nsp9$_{SARS}$ for long oligonucleotide stretches it was suggested that the natural RNA substrate may instead be conserved features at the 3′ end of the viral-genome (the stem-loop II RNA-motif) (Ponnusamy et al., 2008). Furthermore, potential direct interactions with the co-factors of the RNA polymerase have been reported (Chen et al., 2017). However, it remains to be determined how the oligonucleotide-binding activity of Nsp9 proteins promotes viral replication during infection.

The sequence of Nsp9 homologs is conserved among betacoronaviruses, suggesting a degree of functional conservation. Nsp9$_{COV19}$ exhibits 97% sequence identity with Nsp9$_{SARS}$ but only 44% sequence identity with Nsp9$_{HCoV}$. The structure of the HCoV-229E Nsp9 protein suggested a potential oligomeric switch induced upon the formation of an intersubunit disulfide bond. Here, disulfide bond formation shifts the relative orientation of the Nsp9 monomers, which was suggested to promote higher-order oligomerization (Ponnusamy et al., 2008). The resultant rod-like higher-order Nsp9$_{HCoV}$ assemblies had increased affinity for the RNA oligonucleotides. Cysteine mutants of Nsp9$_{HCoV}$ that are unable to produce the disulfide displayed reduced RNA-binding affinity (Ponnusamy et al., 2008). The observation of a redox-induced structural switch of Nsp9$_{HCoV}$ led to the hypothesis that Nsp9$_{HCoV}$ may have a functional role in sensing the redox status of the host cell (Ponnusamy et al., 2008). Although the "redox-switch" cysteine responsible for oligomer formation in Nsp9$_{HCoV}$ is conserved among different viral Nsp9 homologs, the higher-order oligomers were not observed for Nsp9$_{SARS}$ (Ponnusamy et al., 2008). Because of these potential differences between Nsp9 proteins we sought to further characterize the nature of Nsp9$_{COV19}$.

RESULTS

Expression and Purification of the SARS-CoV-2 Nsp9 Protein

The Nsp9 protein from SARS-CoV-2 (Nsp9$_{COV19}$) was cloned and recombinantly expressed in *E. coli*. The expression construct included an N-terminal Hexa-His tag attached via a rhinoviral 3C-protease site. Following Ni-affinity chromatography Nsp9$_{COV19}$ was further purified via size-exclusion chromatography to yield >95% pure and homogeneous protein. Nsp9$_{COV19}$ eluted from gel filtration columns with the apparent molecular weight of a dimer suggesting that, as with other Nsp9 proteins, Nsp9$_{COV19}$ is an obligate homodimer. The N-terminal tag was removed prior to any biochemical experiments via overnight digestion with precision protease, as reported for Nsp9$_{SARS}$ (Sutton et al., 2004).

Nucleotide Binding of Nsp9$_{COV19}$ Protein

The affinity of viral Nsp9 homologs for oligonucleotides has a range of binding affinities reported, some of which are dependent on oligomerization state and nucleotide length, ranging from 20 to 400 μM (Zeng et al., 2018). We therefore sought to assess the potential for Nsp9$_{COV19}$ to bind to fluorescently labeled oligonucleotides using fluorescence anisotropy. Preliminary experiments were performed under conditions similar to those previously identified for Nsp9$_{SARS}$ (Sutton et al., 2004). Oligonucleotide affinity was very limited under our assay conditions (Figure 1). Indeed, protein concentrations up to 200 μM of Nsp9$_{COV19}$

	3c-Nsp9$_{COV19}$	apo-Nsp9$_{COV19}$
Data Collection		
Space group	$P4_322$	$P6_122$
Cell dimensions		
$\quad$ a, b, c (Å)	59.0, 59.0, 85.7	88.8, 88.8, 134.0
$\quad$ α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0
Resolution (Å)	48.5-2.0 (2.1-2.05)	44.7-2.00 (2.06-2.00)
R_{pim}[a]	2.4 (60.4)	1.2 (40.4)
I/σ_1	13.7 (1.4)	26.4 (1.4)
Completeness (%)	100 (100)	99.4 (93.1)
Total number of observations	86,945 (12,241)	423956 (28,236)
Number of unique observations	10,055 (1,425)	21,482 (1,451)
Multiplicity	8.6 (8.6)	19.7 (19.5)
Refinement Statistics		
R_{factor}[b] (%)	23.3	21.4
R_{free}[c] (%)	24.7	25.2
Number of atoms		
$\quad$ Protein	936	1,716
$\quad$ Water (ligand)	5 (PO_4)	18 (3 SO_4)
Ramachandran plot (%)		
$\quad$ Most favored	95.0	95.9
$\quad$ Allowed region	5.0	4.1
$\quad$ Outlier	0.0	0.0
B factors (Å^2)		
$\quad$ Protein	72.1	73.8
$\quad$ RMSD bonds (Å)	0.015	0.012
$\quad$ RMSD angles (°)	1.33	1.19

Table 1. Data Collection and Refinement Statistics

Values in parentheses refer to the highest resolution bin.

[a]$R_{p.i.m} = \Sigma_{hkl} [1/(N-1)]^{1/2} \Sigma_i \mid I_{hkl, i} - <I_{hkl}> \mid /\Sigma_{hkl} <I_{hkl}>$.

[b]$R_{factor} = (\Sigma \mid |F_o| - |F_c| |)/(\Sigma |F_o|)$—for all data except as indicated in footnote c.

[c]Five percent of data was used for the R_{free} calculation.

did not result in saturated binding and thus indicated an incredibly low affinity K_D, or no affinity for these oligonucleotides at all under these assay conditions.

Crystal Structure of apo-Nsp9$_{COV19}$

We next determined the structure of apo-Nsp9$_{COV19}$ (Table 1). The apo-Nsp9$_{COV19}$ structure aligned closely to that of Nsp9$_{SARS}$ (root-mean-square deviation [RMSD] of 0.57 Å over 113 C$_\alpha$, Figures 2A–2C) (Egloff et al., 2004). Like other Nsp9 homologs it exhibits an unusual fold that is yet to be observed outside of coronaviruses (Sutton et al., 2004). The core of the fold is a small six-stranded enclosed β barrel, from which a series of extended loops project outward (Figure 2A). The elongated loops link the individual β strands of the barrel,
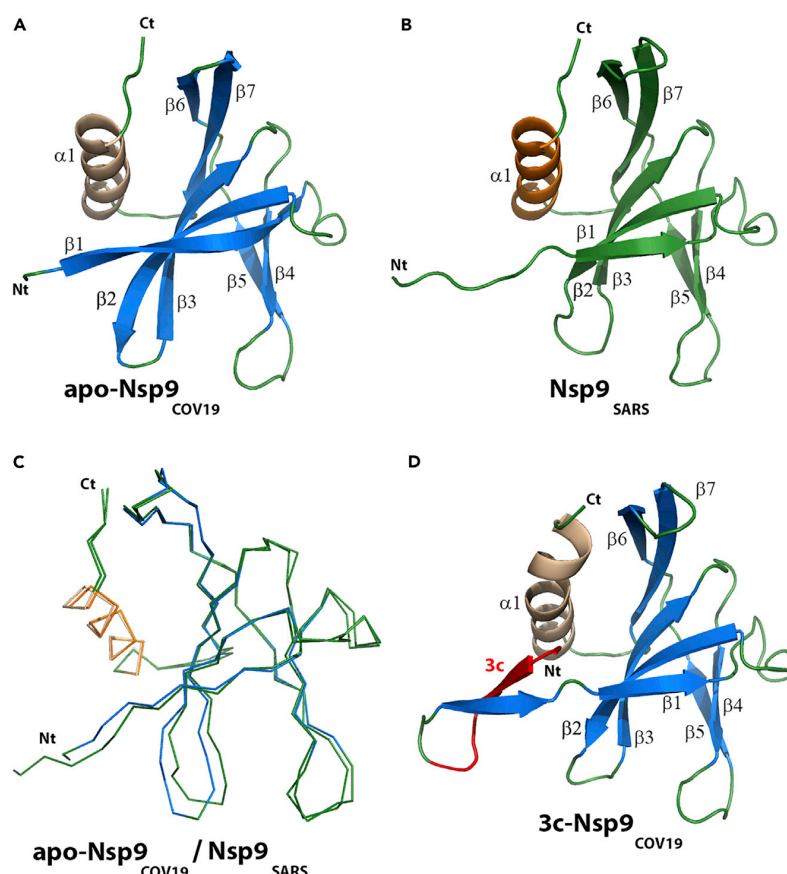
Figure 2. apo-Nsp9$_{COV19}$ Is Structurally Similar to Nsp9$_{SARS}$

(A–C) Cartoon representation of the monomeric units of (A) apo-Nsp9$_{COV19}$, (B) apo-Nsp9$_{SARS}$ (Sutton et al., 2004), and (C) a backbone alignment of the two structures. The COV19 structures are colored with β strands in *marine* and the α helix in *wheat;* the SARS structures are in *teal* and *orange*, respectively.

(D) The bound peptide is highlighted in *red.*

along with a projecting N-terminal β strand and C-terminal α1 helix; the latter two elements make up the main components of the dimer interface (Figure 3A). Two loops project from the open face of the barrel: the β2-3- and β3-4-loops are both positively charged, glycine rich, and proposed to be involved in RNA binding. The only protrusion on the enclosed barrel side is the β6-7-loop; the C-terminal half of the β7 strand is an integral part of the fold's barrel-core, but its other half extended outward to pair with the external β6 strand and create a twisted β hairpin, cupping the α1 helix and interacting with subsequent C-terminal residues.

The arrangement of monomers within Nsp9 dimers is well conserved in different viruses and is maintained within Nsp9$_{COV19}$ (RMSD of 0.66 Å over 226C$_\alpha$ compared with the dimeric unit of Nsp9$_{SARS}$). The main component of the intersubunit interaction is the self-association of the conserved GxxxG protein-protein binding motif (Figure 3C) that allowed backbone van der Waals interactions between interfacing copies of the C-terminal α1 helix (Hu et al., 2017). Here Gly-100 of the respective parallel α1 helices formed complementary backbone van der Waals interactions. These interactions were replicated after a full helical turn by Gly-104 of the respective chains, thereby forming the molecular basis of the Nsp9$_{COVID19}$ dimer interface (Figure 3C). The 2-fold axis that created the dimer ran at a ∼15° angle through the GxxxG motif allowing the 14-residue helix to cross its counterpart (Figure 4A), the N-terminal turns of the helix were relatively isolated, only making contacts with counterpart protomer residues. In contrast, the C-terminal portions were encircled by hydrophobic residues, albeit at a distance that created funnel-like hydrophobic cavities either side of the interfacing helices (Figure 3A). Strands β1 and β6 and the protein's C terminus served to provide
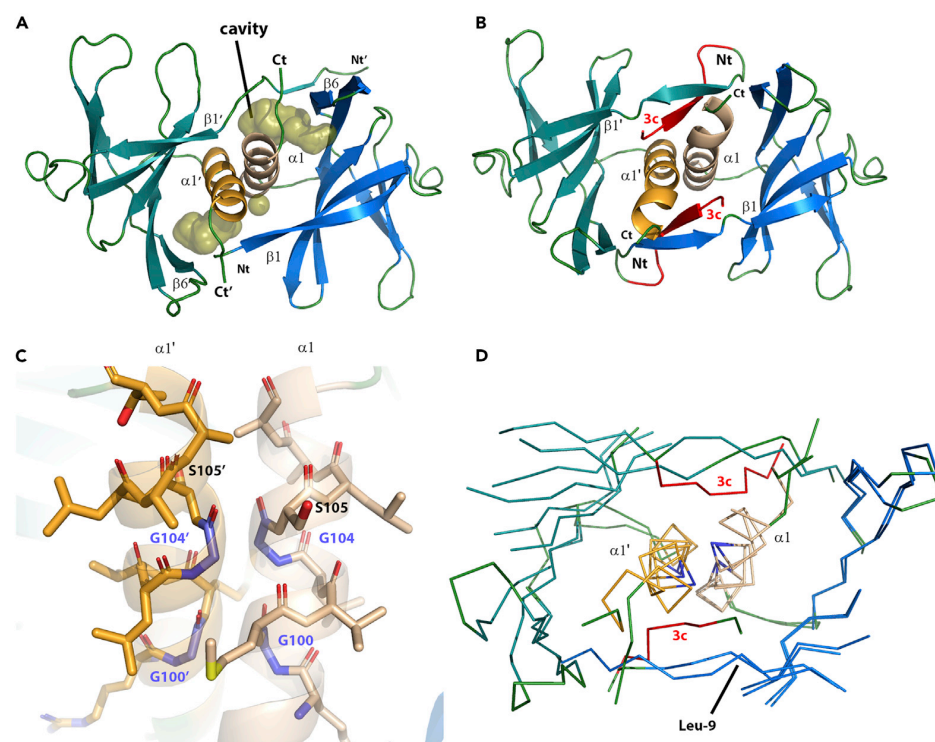
Figure 3. Peptide Binding in Nsp9_{COV19} Alters the Dimer Interface
(A and B) Top-down views of the dimer interface highlighting the interaction helices for (A) unbound Nsp9_{COV19} in which the surface of the hydrophobic interface cavity is displayed labeled; (B) an equivalent representation of peptide-occupied 3c-Nsp9_{COV19} dimer.
(C) Stick representation of the GxxG protein-protein interaction helices at the dimer interface for apo-Nsp9_{COV19}.
(D) C_α backbone overlay of the Nsp9_{COV19} interface in the apo and peptide-occupied states. The GxxG motif residues are colored *light purple*.

a ring of residues that encircled the paired helices. The first 10 residues of Nsp9_{COV19} exchanged across the dimer-interface to form a strand-like extension of β1 that ran alongside β6 from the other protomer (Figure 3A). The interaction these strands made did not appear optimal; indeed, the remaining four C-terminal residues projected sideways across the dimer interface, inserting between the two strands while contributing a hydrophobic backing to the main helix.

Extraneous Peptides Occupy the Hydrophobic Cavities of Nsp9

In a separate crystallization experiment we determined the structure of Nsp9_{COV19} that included the N-terminal tag together with a rhinoviral 3C protease sequence (termed 3C-Nsp9_{COV19}). The 3C-Nsp9_{COV19} crystal form diffracted to 2.05 Å resolution in space group P4_322 and had one molecule within the asymmetric unit, with the dimer being created across the crystallographic 2-fold axis.

Unexpectedly, the high-resolution structure of 3C-Nsp9_{COV19} diverged from that of the apo-Nsp9_{COV19} (RMSD 0.86 Å for the monomer and 2.23 Å when superimposing a dimer). The 3C sequence folded around either side of the paired intersubunit helices to fill two funnel-like hydrophobic cavities (Figures 2D, 3B, 4C, and 4D), namely, 3C residues LEVL, inserted into the opposing cavities either side of the dimer interface and ran parallel to the paired GxxG motif. Moreover, the 3C sequence formed additional β sheet interactions with the N terminus of the protein from the other protomer (Figure 3B). To accommodate the 3C residues the N-terminal strand residues moved outward by ~1.6 Å (residues 6–10). This movement allowed the N terminus to increase the number of β sheet interactions it formed with β6′. The β barrel core of the fold remained unchanged, but the increase in interactions between β1 and β6′ served to exclude the C terminus, prompting residues 106–111 to condense into a bent extension of the α helix (Figures 4A and 4B). The
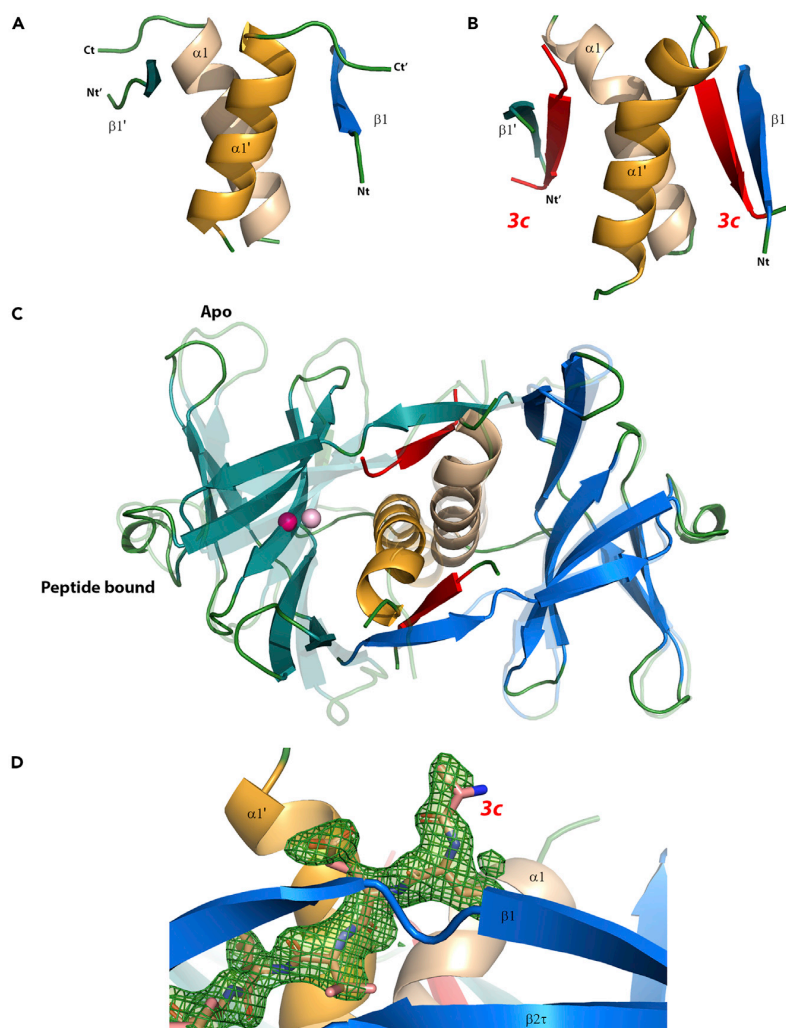
Figure 4. Movements within the GxxxG Motif

(A and B) A side view of the N- and C- terminal structural elements at the dimer interface is shown for the (A) apo and (B) peptide-occupied forms.

(C) Overlay of the Nsp9$_{COV19}$ dimer in the apo and peptide-bound forms indicating respective shifts in subunit orientation. The center of mass of the nonaligned subunit is depicted with a *light-pink* and *dark-pink* point, respectively.

(D) Unbiased omit map contoured at 3.2σ near the hydrophobic cavities into which the exogenous bound peptide was refined.

subtle structural changes near the interacting GxxxG motifs (Figure 3D) are amplified at the periphery of the dimer resulting in ∼6 Å shift in the β barrel core (Figure 4C).

Conserved Cavity Residues Accommodate a Peptide Backbone

When comparing apo-Nsp9$_{COV19}$ with 3C-Nsp9$_{COV19}$ the point where the N-terminal interface strand diverges is near Leu-9 (Figure 3D). Within the apo form it makes van der Waals interactions with the side chains of Met-101, Asn-33, and Ser-105; this latter serine is important as it immediately follows the conserved protein-binding motif (^{100}GMVLGS^{105}), while also specifically interacting with Gly104′ from the opposing protomer. Within the 3C-Nsp9$_{COV19}$ structure the extraneous LEVL residues insert at this point (Figure 5B) and the hydrophobic side chains clasp either side of Ser-105 and allowed its hydroxyl group to form backbone hydrogen bonds to the glutamate within the extraneous sequence (Figure 5B).
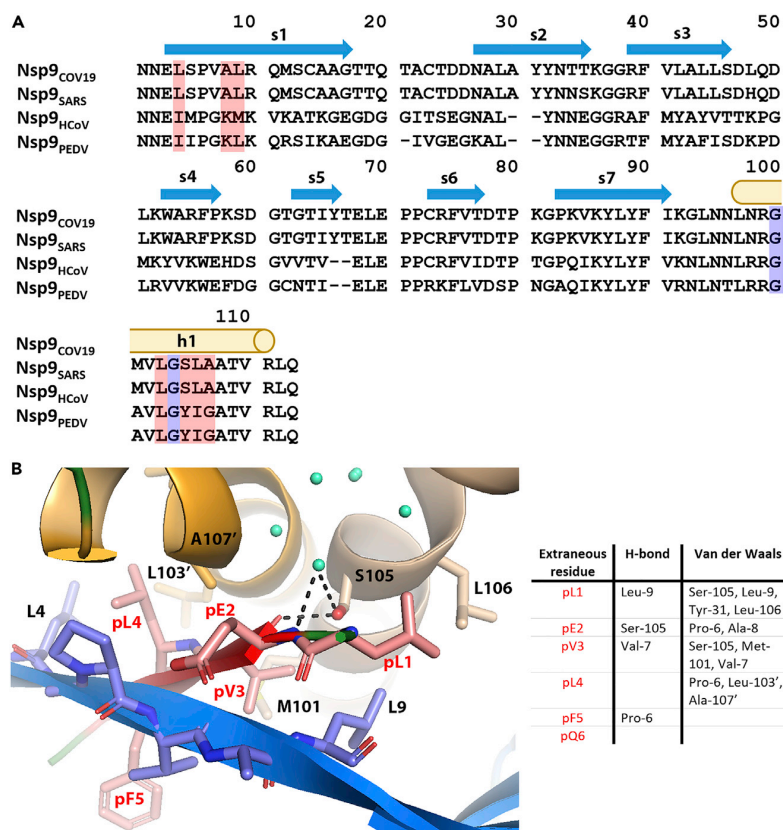
A

```
                      10              20              30        40              50
                               s1                       s2          s3
Nsp9_COV19   NNELSPVALR QMSCAAGTTQ TACTDDNALA YYNTTKGGRF VLALLSDLQD
Nsp9_SARS    NNELSPVALR QMSCAAGTTQ TACTDDNALA YYNNSKGGRF VLALLSDHQD
Nsp9_HCoV    NNEIMPGKMK VKATKGEGDG GITSEGNAL- -YNNEGGRAF MYAYVTTKPG
Nsp9_PEDV    NNEIIPGKLK QRSIKAEGDG -IVGEGKAL- -YNNEGGRTF MYAFISDKPD

                 s4      60       s5     70       s6     80        s7    90            100
Nsp9_COV19   LKWARFPKSD GTGTIYTELE PPCRFVTDTP KGPKVKYLYF IKGLNNLNRG
Nsp9_SARS    LKWARFPKSD GTGTIYTELE PPCRFVTDTP KGPKVKYLYF IKGLNNLNRG
Nsp9_HCoV    MKYVKWEHDS GVVTV--ELE PPCRFVIDTP TGPQIKYLYF VKNLNNLRRG
Nsp9_PEDV    LRVVKWEFDG GCNTI--ELE PPRKFLVDSP NGAQIKYLYF VRNLNTLRRG

                   110
                 h1
Nsp9_COV19   MVLGSLAATV RLQ
Nsp9_SARS    MVLGSLAATV RLQ
Nsp9_HCoV    AVLGYIGATV RLQ
Nsp9_PEDV    AVLGYIGATV RLQ
```

B

Extraneous residue	H-bond	Van der Waals
pL1	Leu-9	Ser-105, Leu-9, Tyr-31, Leu-106
pE2	Ser-105	Pro-6, Ala-8
pV3	Val-7	Ser-105, Met-101, Val-7
pL4		Pro-6, Leu-103', Ala-107'
pF5	Pro-6	
pQ6		

Figure 5. Sequence Conservation within Nsp9 Homologs

(A) Sequence alignment for viral Nsp9 proteins encoded by SARS-CoV-2, SARS-CoV, Human coronavirus 229E, and Porcine epidemic diarrhea virus. The extent of secondary structural elements observed in the 3C-Nsp9$_{COV19}$ structure is shown and labeled above. The GxxxG motif residues are highlighted with *purple* and those making up the extraneous peptide binding site in *pink*.

(B) Cartoon-and-stick representation of the peptide-binding site observed in 3C-Nsp9$_{COV19}$. The side chains of the extraneous 3C residues on one side of the paired helical-interface are displayed with carbon atoms colored *pink*. Nearby residues that make up the binding site are displayed and labeled and listed in the accompanying contacts table.

Meanwhile, the C-terminal leucine from the extraneous residues inserted behind the α helix of the other protomer (Figure 5B). Cumulatively these changes allow for an ~5° rotation of the protomer subunits about the 2-fold axis compared with apo-Nsp9$_{COV19}$ (Figure 4C).

Most residues involved in protein binding within the hydrophobic cavity and the structural changes needed to accommodate them appear broadly conserved among other Nsp9 viral homologs (*red* highlights in Figure 5A). The main exception to this is Ser-105, which is a tyrosine in the distantly related Nsp9$_{HCoV}$ and Nsp9$_{PEDV}$ proteins (Ponnusamy et al., 2008) (Zeng et al., 2018). However, the N-terminal interface β strand in these homologs is known to be involved in interface re-organization of the subunits (Ponnusamy et al., 2008) and thus denotes other structural differences at this site.

DISCUSSION

Here we describe the structure of the recombinantly expressed Nsp9$_{COV19}$ as part of a global effort to characterize the virus causing a current global pandemic. Nsp9 is important for virulence in SARS-CoV (Miknis et al., 2009). It remains to be understood whether Nsp9$_{COV19}$ plays a similar role in SARS-CoV-2; however, the 97% sequence identity suggests a high degree of functional conservation. The CoV Nsp9 proteins are seemingly obligate dimers comprising a unique fold that associates via an unusual α-helical GxxxG interaction motif. The integrity of this

motif is considered important for viral replication (Miknis et al., 2009), leading to a proposal that disruption of the unusual dimer interface impacts on RNA binding and function (Hu et al., 2017). Mutation of the same interaction motif in the porcine delta coronavirus Nsp9$_{PDCoV}$ also disrupted nucleotide binding capacity (Zeng et al., 2018).

We describe the ability to produce homogeneous Nsp9$_{COV19}$, which purifies as an obligate dimer, consistent with other Nsp9 proteins. Our preliminary nucleotide binding assays brought into question the RNA-binding capacity of Nsp9$_{COV19}$. The structure of the Nsp9$_{COV19}$ showed conservation of the unique Nsp9 fold when compared with homologs from SARS (Egloff et al., 2004) (Sutton et al., 2004). Indeed, the topological fold was conserved as was the Nsp9-specific α-helical GxxxG dimerization interface. This α-helical interface is encircled by hydrophobic residues but the interface includes considerable cavities as observed previously (Egloff et al., 2004). We made a serendipitous discovery in our 3C-Nsp9$_{COV19}$ structure, whereby the hydrophobic cavity captured the 3C cleavage sequence LEVL. The extraneous residues were tightly bound on all sides within the site situating themselves proximal to the conserved GxxxG motif. Coordination of the 3C sequence induced changes within interfacing residues, serving to both restructure key structural elements and cause a modest shift in subunit orientation.

At this stage it is unclear whether the bound residues within our structure have any bearing on the physiological function of Nsp9$_{COV19}$. In the first instance this would seem unlikely; however, our sequence is that of a rhinoviral 3C protease site and the SARS-CoV main protease cleaves consensus sequences following an LQ sequence (Zhu et al., 2011). The bound 3C-LE residues have hallmarks of the LQ motif and are proximal to the highly conserved GxxxG motif. There are no obvious structural features to preclude, or select for, a Glu to Gln substitution within our bound sequence. Notably the M^{pro} cleavage sequence occurs at multiple points throughout the CoV genome as the majority of viral proteins are released by its activity; thus, it remains possible that Nsp9 may associate with unprocessed viral polyproteins retaining them near the viral RNA. Within our structure Met-101 provides contacts with the bound valine side chain but the presence of Asn-33 nearby may also accommodate residues such as lysine at this position. Peptide-binding assays will need to be developed to rigorously assess if other sequences are preferred by this putative binding site. Indeed, Nsp9 may be part of the viral replication-transcriptase complex, so we may have serendipitously identified a protein-protein interaction interface for another viral or host protein.

In summary, we have established a protocol for the production and purification of SARS-CoV-2 Nsp9 protein. We determined the structure of the Nsp9$_{COV19}$ and described the conservation of the unique fold and dimerization interface identified previously for members of this protein family. We also determined the structure of Nsp9$_{COV19}$ in complex with a 3C sequence, although the significance of this is yet to be established. The structures we describe here could potentially be utilized in drug screening and targeting experiments to disrupt a dimer interface known to be important for coronavirus replication.

Limitations of the Study

The identity of the peptide bound to Nsp9$_{COV19}$ raises follow-up questions that were not addressed within this study. Namely, it remains to be determined whether a physiological peptide, either of the same sequence or a 3C peptide variant is also able to occupy this putative site. Further questions remain on whether this is a retention mechanism for pre-processed or post-processed polyproteins or another protein altogether, and if so, what its binding affinity might be.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to: jamie.rossjohn@monash.edu.

Materials Availability

Plasmids generated in this study are available upon request.

Data and Code Availability

The accession number for the atomic coordinates of the apo-Nsp9$_{COV19}$ with 3C-Nsp9$_{COV19}$ and associated diffraction data have been deposited at the protein databank (www.rcsb.org) with accession codes PDB: 6W9Q and PDB: 6WXD, respectively.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101258.

AUTHOR CONTRIBUTIONS

D.R.L, B.S.G., and J.R. designed the project and wrote the manuscript. D.R.L. cloned, purified, and crystallized $Nsp9_{Cov19}$ and refined the structures. R.N.C. performed the RNA binding assays.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

This article contains Supplemental Information online.

REFERENCES

Chen, J., Xu, X., Tao, H., Li, Y., Nan, H., Wang, Y., Tian, M., and Chen, H. (2017). Structural analysis of porcine reproductive and respiratory syndrome virus non-structural protein 7alpha (NSP7alpha) and identification of its interaction with NSP9. Front. Microbiol. 8, 853.

Egloff, M.P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J., Gorbalenya, A.E., Cambillau, C., and Canard, B. (2004). The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. Proc. Natl. Acad. Sci. U S A 101, 3792–3796.

Frieman, M., Yount, B., Agnihothram, S., Page, C., Donaldson, E., Roberts, A., Vogel, L., Woodruff, B., Scorpio, D., Subbarao, K., et al. (2012). Molecular determinants of severe acute respiratory syndrome coronavirus pathogenesis and virulence in young and aged mouse models of human disease. J. Virol. 86, 884–897.

Hu, T., Chen, C., Li, H., Dou, Y., Zhou, M., Lu, D., Zong, Q., Li, Y., Yang, C., Zhong, Z., et al. (2017). Structural basis for dimerization and RNA binding of avian infectious bronchitis virus nsp9. Protein Sci. 26, 1037–1048.

Miknis, Z.J., Donaldson, E.F., Umland, T.C., Rimmer, R.A., Baric, R.S., and Schultz, L.W. (2009). Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. J. Virol. 83, 3007–3018.

Ponnusamy, R., Moll, R., Weimar, T., Mesters, J.R., and Hilgenfeld, R. (2008). Variable oligomerization modes in coronavirus non-structural protein 9. J. Mol. Biol. 383, 1081–1096.

Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., et al. (2004). The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. Structure 12, 341–353.

Thanh Le, T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen, S., Saville, M., and Mayhew, S. (2020). The COVID-19 vaccine development landscape. Nat. Rev. Drug Discov. 19, 305–306.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263.

Yan, G., Liming, Y., Yucen, H., Fengjiang, L., Yao, Z., Lin, C., Tao, W., Qianqian, S., Zhenhua, M.,

Lianqi, Z., et al. (2020). Structure of RNA-dependent RNA polymerase from 2019-nCoV, a major antiviral drug target. bioRxiv. https://doi.org/10.1101/2020.03.16.993386.

Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., et al. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLoS Biol. 3, e324.

Zeng, Z., Deng, F., Shi, K., Ye, G., Wang, G., Fang, L., Xiao, S., Fu, Z., and Peng, G. (2018). Dimerization of coronavirus nsp9 with diverse modes enhances its nucleic acid binding affinity. J. Virol. 92, e00692-18.

Zhenming, J., Xiaoyu, D., Yechun, X., Yongqiang, D., Meiqin, L., Yao, Z., Bing, Z., Xiaofeng, L., Leike, Z., Chao, P., et al. (2020). Structure of Mpro 1 from COVID-19 virus and discovery of its inhibitors. bioRxiv. https://doi.org/10.1101/2020.02.26.964882.

Zhu, L., George, S., Schmidt, M.F., Al-Gharabli, S.I., Rademann, J., and Hilgenfeld, R. (2011). Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. Antiviral Res. 92, 204–212.

Supplemental Information

Crystal Structure of the SARS-CoV-2

Non-structural Protein 9, Nsp9

Dene R. Littler, Benjamin S. Gully, Rhys N. Colson, and Jamie Rossjohn

Synthetic cDNA for the Nsp9$_{COV19}$ protein was cloned into pET28-LIC expression vector bearing an N-terminal His-tag with a Rhinovirus 3C protease cleavage site (MAHHHHHHSAA*LEVL*FQGPG).

Synthetic cDNA for Nsp9$_{COV19}$:

```
 aat aat gaa ctg agt cct gtc
  N   N   E   L   S   P   V
gcg ctg cgt caa atg agt tgc gcc gcc ggt acg acc cag acc gca tgt act gac gac aac
 A   L   R   Q   M   S   C   A   A   G   T   T   Q   T   A   C   T   D   D   N
gct tta gcc tat tat aac acc aca aag ggg ggt cgt ttt gtt ctt gca ttg ctg tcg gat
 A   L   A   Y   Y   N   T   T   K   G   G   R   F   V   L   A   L   L   S   D
ttg cag gac ctg aaa tgg gct cgt ttt ccc aaa agc gac ggt act gga aca att tac acg
 L   Q   D   L   K   W   A   R   F   P   K   S   D   G   T   G   T   I   Y   T
gaa tta gag cca ccc tgt cgt ttc gtt aca gat acg ccc aag ggt ccc aag gtt aaa tac
 E   L   E   P   P   C   R   F   V   T   D   T   P   K   G   P   K   V   K   Y
tta tac ttc atc aag ggt ctt aac aat ctg aat cgc ggt atg gta ctg ggt tca ctg gcc
 L   Y   F   I   K   G   L   N   N   L   N   R   G   M   V   L   G   S   L
gcc aca gtt cgc ctt caa   A   T   V   R   L   Q
```

Cloning

DAY -5 - Primer design and ordering Determine the cDNA sequence of the constructs you want to make. Design primers based on this sequence aiming for a 68oC annealing temperature. Add the LIC-specific extensions "cagggacccggt" to the 5' end of the fwd primer and "cgaggagaagcccggtta" to the 5' end of the rev primer. For example:

forward	cagggacccggtaataatgaactgagtcctgtc
reverse	cgaggagaagcccggttattgaaggcgaactgtggcggc

Primers can be ordered from the Sigma Genosys website 0.025 ug quantity and desalting is sufficient. Delivery time can vary considerably allow plenty of time.

DAY0 - Growth induction

From glycerol stocks or a plate inoculate 250 -300 mL LB with 30ug/mL Kanamycin in a 1L flask with either DH5a or Nova Blue cells containing the pET-NKIb 3C/LIC empty plasmid. Grow to saturation overnight at 37oC 140rpm.

DAY1 – Vector cleavage

Perform a Qiagen-Midi or Maxi- prep of the plasmid according to instructions. 1ug of cleaved plasmid will be enough for approximately 20 LIC reactions. At this stage carefully check the size of the plasmid on a 0.8% Agarose gel, the supercoiled plasmid should run just below the MIII Marker 3.5kbp band.

Cleave the vector with the appropriate restriction enzyme (based on the LIC vector you want to use). For example the reaction for 10ug would be something like:

Example:

Reagent	Add
pET-NKIb 3C/LIC	10ug (e.g. 100uL of 100ng/uL stock)
restriction enzyme	10uL of 1U/uL stock
10x buffer	14uL
MQ H20	Make upto 140uL water

Make up in a 1.5mL eppendorf and incubate at 37oC for 2-3 hours. Remove a small sample and run on a 0.8% Agarose gel to check cleavage efficiency.

DAY1 – Insert PCR

Setup the PCR of the insert in 500uL thermowell tubes.

Reagent	Add	Final conc.
DNA template	X uL	20 ng
10x Stratagene Pfu buffer (contains Mg2+)	10 uL	1x buffer
10mM dNTPs	2.5 uL	250 uM
Fwd primer (100pmol/uL)	0.05-1.0 uL	5-100 pmol
Rev primer (100pmol/uL)	0.05-1.0 uL	5-100 pmol
Pfu Stratagene polymerase (2.5U/uL)	1.0 uL	2.5U
MQ H2O	Upto 100uL	

Run with a cycle such as:

94oC for 5 minutes	
94oC for 1 minute }	
55oC for 1 minute }	Repeat steps 30x
72oC for 2.5 minutes }	
72oC for 10 minutes	
10oC for ever	

Remove 5 uL of the PCR reaction and mix with 3 uL of 6x Loading dye. Run on a 0.8% Agarose gel and check the size of the bands. Ensure that they are correct (usually by looking for slight movements up/down compared to nearby samples).

Perform a Qiagen PCR-cleanup of the PCR reaction mixtures providing there is a single band only. If two or more bands occur a gelextraction may be necessary, but if they are clearly distinct these can be more easily separated by performing a larger number of PCRcolony screen

DAY2 – T4 Polymerase treatment

Treat the cut-vector and PCR'd insert with T4 polymerase in 1.5mL eppendorf tubes. The number of pg/pmol = (#of bp)x650. So for a 1000bp PCR-product we have 1000x650 = 650 000 pg/pmol = 650 ng/pmol which means for 0.2pmol of insert we require:

```
650ng/pmol x 0.2pmol = 130ng
```

normal yields of Qiagen miniprep are 20ng/uL – 100ng/uL so will require 8uL-1uL of PCR product.

Reagent Add PCR insert 0.2 pmol 10x buffer (NEBuffer #2) 2 uL 25 mM dATP 2 uL T4 polymerase (NEB 3U/uL) 0.8 uL MQ H2O Upto 20 uL

Reagent Add Cut, gel-extracted vector 0.2 pmol 10x buffer (NEBuffer #2) 2 uL 25 mM dTTP 2 uL T4 polymerase (NEB 3U/uL) 0.8 uL MQ H2O Upto 20 uL
1) Incubate at 22oC for 30 minutes.
2) Denature the T4 polymerase by then incubating the reaction at 75oC for 10-20 minutes. 3) Spin down the
   evaporation.

Conjoin the vector and insert

1) Add 2 uL of the T4 treated insert to a 1.5 mL eppendorf tube
2) Add 1 uL of the T4 treated vector (~50ng/uL) 3) Incubate
together at 22oC for 5 minutes.
4) Stop the reaction by adding 1 uL of sterile 25 mM EDTA.

Transform the reaction into Novagen NovaBlue's
1) Add 10-15 uL of Nova Blue competent cells to 2ul of conjoined vector/insert mix.(Store the other 2ul at -20oC). Include a T4 treated vector-only sample to assess background. 2) Incubate on ice for 30 minutes.
3) Heat shock at 42oC for 30-40 seconds.
4) Incubate on ice for 2 minutes.
5) Add 250 uL of SOC media or LB without antibiotic.
6) Allow the cells to recover by incubating at 37oC, 1400rpm.
7) Plate out on the appropriate antibiotic for the vector, leave overnight at 37oC.

DAY3 – Colony screening.

If colonies have grown (colonies take closer to 20 hours to grow) perform a PCR-screen using the sequencing primers appropriate for the vector (e.g. T7 promoter/ T7 terminator primers for the pET vector). Setup a Taq sequencing mix such as:

Reagent For 1 reaction For 25 reactions 10 x invitrogen Taq buffer (- Mg2+) 2 uL 50 uL 50 mM invitrogen MgCl2 1 uL 20 uL 10 mM dNTPs 0.4 uL 10 uL Fwd primer (10 pmol/uL) 0.3 uL 7.5 uL Rev primer (10 pmol/uL) 0.3 uL 7.5 uL Invitrogen Taq polymerase (5U/uL) 0.15 uL 3.75 uL MQ H2O Upto 20uL 401.25 uL
1) Label 2 or 3 bacterial colonies from each samples plate.
2) Aliquot 20 uL of the master PCR mix into each well of a 96-well thermowell tray.
3) Take a sterile P10 tip and scrape half of each bacterial colony and put this into the 20 uL. Leave the tip in the tray while going onto the next sample.
4) After scraping all colonies use a pipette to mix briefly and seal the reactions. 5) Run using a program such as:

96oC for 10 minutes 96oC for 1 minute | 55oC for 1 minute | Repeat steps 25x 68oC for 1 minutes/kbp | 68oC for 10 minutes 10oC for ever
6) When finished add loading dye to each PCR reaction and run half on a 0.8%-1.0% Agarose gel. Check the size of each band (remember the sequencing primers will add approx 170 extra bp compared to the original reaction). If any samples are missing colonies go back and rescreen.
7) Inoculate overnights from the remaining half-spot of the positive colonies.

DAY4 – Plasmid minipreps.

From the overnights purify plasmids by either using Qiagen mini-preps or in a 96-well block format. Transform into an expression cell-line like BL21 (DE3) or Rosetta (DE3).

Protein purification

The plasmid was transformed into *E. coli* BL21(DE3) cells which were grown in Luria Broth at 37°C until reaching an Absorbance at 600nm of ~1.0 before being induced with 0.5mM Isopropyl β- d-1-thiogalactopyranoside for 4 hours.

Cells were harvested in 20mM HEPES pH 7.0, 150mM NaCl, 20mM Imidazole, 2mM $MgCl_2$ and 0.5mM TCEP and frozen until required.

Lysis was achieved by sonicating the cells in the presence of 1mg of Lysozyme and 1mg of DNAase on ice.

The lysate was then cleared by centrifugation at 10,000xg for 20 minutes and loaded onto a nickel affinity column.

Bound protein was washed extensively with 20 column volumes of 20mM HEPES pH 7.0, 150mM NaCl, 0.5mM TCEP before being eluted in the same buffer with the addition of 400mM Imidazole.

For His-tag removal samples were incubated with precision 3c protease overnight at 4°C.

All samples were subjected to gel filtration (S75 16/60; GE Healthcare) in 20mM HEPES pH 7.0, 150mM NaCl before being concentrated to 50mg/mL for crystallization trials.

Crystallisation
$Nsp9_{COV19}$ crystallized in 2.0-2.2M NH_4SO_4 and 0.1M Phosphate-citrate buffer pH 4.0.

Crystals of the His-tag samples grew with rectangular morphology in space group $P4_122$, however if the His-tag was removed the crystals grew in space group $P6_122$ form with hexagonal morphology.

All diffraction data were collected at the Australian synchrotron's MX2 beamline at the Australian synchrotron (Aragao *et al.*, 2018) (see Table 1 for details).
Data were integrated in XDS (Kabsch, 2010), processed with SCALA, phases were obtained through molecular replacement using PDB 1QZ8 (Egloff *et al.*, 2004) as a search model.

Subsequent rounds of manual building and refinement were performed in Coot (Casanal *et al.*, 2019) and PHENIX (Liebschner *et al.*, 2019).

Nucleotide binding assay

To examine the RNA-binding affinity, an 18-point serial dilution (212-0 μM) of $Nsp9_{COV19}$ was incubated with 1 nM 5'-Fluorescein labelled 17mer poly-U single-stranded RNA (Dharmacon GE, USA) or 10mer PolyT single-stranded DNA (IDT, USA) in assay buffer (20mM HEPES pH 7.0, 150mM NaCl, 2mM $MgCl_2$) at room temperature.

The assay was performed in 96-well non-binding black plates (Greiner Bio-One), with fluorescence anisotropy measured in triplicate using the PHERAstar FS (BMG) with FP 488-520-520 nm filters.

The data was corrected using the anisotropy of RNA sample alone, then fitted by a one-site binding model using the Equation, $A = (Amax\ [L])/(KD+[L])$, where A is the corrected fluorescence anisotropy; Amax is maximum binding fluorescence anisotropy signal, [L] is the $Nsp9_{COV19}$ concentration, and K_D is the dissociation equilibrium constant.

Amax and K_D were used as fitting parameters and nonlinear regression was performed using GraphPad Prism. Measurements were taken after 60 minutes incubation between protein and RNA.

Methods References

Aragao, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N.P., Ericsson, D.J., Gee, C.L., Macedo, S., Mudie, N., *et al.* (2018). MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. J Synchrotron Radiat *25*, 885-891.

Egloff, M.P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J., Gorbalenya, A.E., Cambillau, C., and Canard, B. (2004). The severe acute respiratory syndrome coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. Proc Natl Acad Sci U S A *101*, 3792-3796.

Kabsch W. (2010) XDS. Acta Cryst. D66, p133-144.

Casanal, A., Lohkamp, B., and Emsley, P. (2019). Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. Protein Sci.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., *et al.* (2019). Macromolecular structure determination using Xrays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol *75*, 861-877.