Received: 21 January 2021 | Revised: 26 July 2021 | Accepted: 2 August 2021

DOI: 10.1002/prot.26205

RESEARCH ARTICLE

PROTEINS WILEY

A distinct ssDNA/RNA binding interface in the Nsp9 protein from SARS-CoV-2

Serene El-Kamand[1] | Mar-Dean Du Plessis[1] | Natasha Breen[1] | Lexie Johnson[1] | Samuel Beard[2] | Ann H. Kwan[3] | Derek J. Richard[2] | Liza Cubeddu[1,3] | Roland Gamsjaeger[1,3]

[1]School of Science, Western Sydney University, Penrith, New South Wales, Australia

[2]School of Biomedical Research, Institute of Health and Biomedical Innovation at the Translational Research Institute, Queensland University of Technology, Woolloongabba, Queensland, Australia

[3]School of Life and Environmental Sciences and Sydney Nano Institute, University of Sydney, New South Wales, Australia

Correspondence
Roland Gamsjaeger and Liza Cubeddu, School of Life and Environmental Sciences and Sydney Nano Institute, University of Sydney, New South Wales 2006, Australia.
Email: r.gamsjaeger@westernsydney.edu.au and l.cubeddu@westernsydney.edu.au

Funding information
Western Sydney University

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel, highly infectious RNA virus that belongs to the coronavirus family. Replication of the viral genome is a fundamental step in the virus life cycle and SARS-CoV-2 non-structural protein 9 (Nsp9) is shown to be essential for virus replication through its ability to bind RNA in the closely related SARS-CoV-1 strain. Two recent studies revealing the three-dimensional structure of Nsp9 from SARS-CoV-2 have demonstrated a high degree of similarity between Nsp9 proteins within the coronavirus family. However, the binding affinity to RNA is very low which, until now, has prevented the determination of the structural details of this interaction. In this study, we have utilized nuclear magnetic resonance spectroscopy (NMR) in combination with surface biolayer interferometry (BLI) to reveal a distinct binding interface for both ssDNA and RNA that is different to the one proposed in the recently solved SARS-CoV-2 replication and transcription complex (RTC) structure. Based on these data, we have proposed a structural model of a Nsp9-RNA complex, shedding light on the molecular details of these important interactions.

KEYWORDS
coronavirus, COVID-19, Nsp9, RNA, SARS-CoV-2

1 | INTRODUCTION

The emergence of a highly contagious RNA virus, severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), in 2019 has resulted in coronavirus disease 2019 (COVID-19) and the ensuing global pandemic. The 30 kb genome of SARS-CoV-1, the agent responsible for the global SARS pandemic of 2002 to 2003, encodes two large polyproteins, which are cleaved to form 16 non-structural proteins (Nsp proteins), four structural proteins, and eight accessory proteins.[1] The genome of SARS-CoV-2 exhibits 79.5% sequence similarity to SARS-CoV-1.[2,3] It is therefore likely that the proteins of SARS-CoV-2 function with a similar mechanism to the homologous proteins of SARS-CoV-1.[4] SARS-CoV-2 Nsp9 is one such protein, exhibiting a 97% sequence similarity to SARS-CoV-1 Nsp9[4] (Figure 1).

SARS-CoV-2 Nsp9 has been found to interact with RNA-dependent RNA polymerase (RdRp/Nsp12) to form part of the replication and transcription complex (RTC), an essential component for viral replication.[4,5] Structures of Nsp9 from both SARS-CoV-1 and

Serene El-Kamand and Mar-Dean Du Plessis have contributed equally to this study.

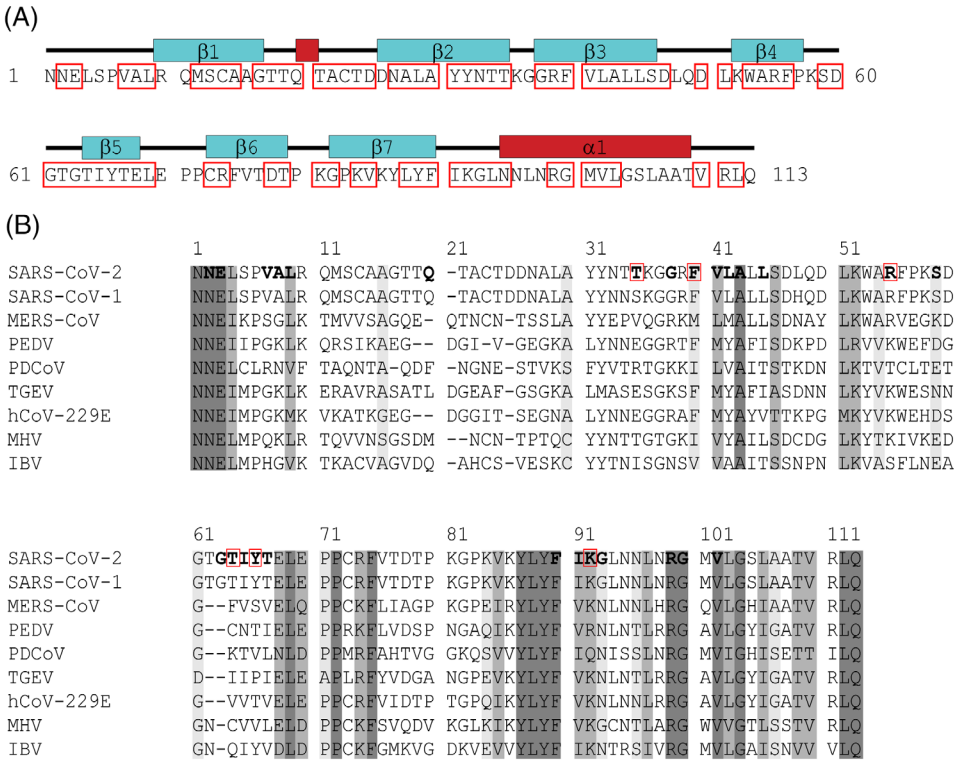


FIGURE 1 Coronavirus family Nsp9 sequence information. (A) Primary sequence of Nsp9 with structural features indicated on top (the α1 helix is part of the monomer-monomer interface). Residues in red boxes were successfully assigned in the HSQC spectra. (B) Sequence alignment of the Nsp9 proteins from SARS-CoV-2, SARS-CoV-1, MERS-CoV, porcine epidemic diarrhea virus (PEDV), porcine deltacoronavirus (PDCoV), transmissible gastroenteritis virus (TGEV), human coronavirus 229E (HCoV-229E), mouse hepatitis virus (MHV), and avian infectious bronchitis virus (IBV). Residues in bold and red boxed residues indicate residues that exhibit significant chemical shift changes ("binding residues") and residues involved in close contacts with RNA as revealed from molecular docking calculations, respectively, whereas gray areas indicate high sequence conservation

SARS-CoV-2 have indicated that the protein forms a homo-dimer in solution with a weak affinity for single-stranded DNA (ssDNA) and RNA.[4,6] Some similarities have been proposed to exist between viral Nsp9 proteins and members of the family of single-stranded DNA binding proteins (SSBs).[4] The ability to recognize and sequester ssDNA renders SSBs essential in DNA replication, repair and in the maintenance of genomic stability.[7-9] Like SSB proteins, Nsp9 proteins are thought to bind to nucleic acids in a non-specific manner to ensure the correct processing of newly synthesized viral RNA.[5,10]

Understanding the complex RNA synthesis machinery of the virus is crucial in the development of therapeutic strategies against COVID-19 and other related coronaviruses. However, for some Nsp proteins, such as Nsp9, important molecular and structural details pertinent to function remain largely unknown. In this study, we have utilized nuclear magnetic resonance spectroscopy (NMR) in combination with biolayer interferometry (BLI) to map the RNA interaction interface of SARS-CoV-2 Nsp9. We have used our biophysical data to calculate structural models of Nsp9-RNA complexes revealing a set of binding residues (including one aromatic residue) that play an important role in the recognition of RNA. Importantly, the site of RNA binding to Nsp9, according to our structural model, differs from the proposed site in the recently published structure of the SARS-CoV-2 RTC.[5]

2 | MATERIALS AND METHODS

2.1 | Protein expression and purification

Recombinant expression of 6xHis-Nsp9 from SARS-CoV-2 was induced in *E. coli* BL21(DE3) cells by the addition of 0.2 mM IPTG for 15 h at 20°C. Isotopically labeled ^{15}N-Nsp9 and ^{15}N^{13}C-Nsp9 were prepared in a bio-fermenter using the protocol described in Reference 11. Cells were lysed by sonication in lysis buffer (20 mM TRIS pH 8.0, 50 mM NaCl, 3 mM TCEP, 0.5 mM PMSF, 0.1% Triton X 100) or SEC-lysis buffer (buffer used to determine molecular weight by size exclusion chromatography; 50 mM TRIS pH 8.0, 500 mM NaCl and 0.01% v/v Igepal CA630, supplemented with protease inhibitors). Lysed cells were centrifuged, and the supernatant was subjected to Ni-NTA affinity chromatography followed by thrombin cleavage for 1 h at 25°C, and a subsequent thrombin cleavage for 15 h at 4°C. For size exclusion chromatography runs to determine molecular weight, no thrombin cleavage was carried out and the protein was eluted in SEC buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.01% v/v Igepal CA630) with 300 mM imidazole (see section below). For NMR and BLI experiments, the eluate from Ni-NTA affinity chromatography was subjected to size exclusion chromatography using a Superdex 75 column (120 ml) equilibrated with NMR buffer (25 mM sodium phosphate pH 6.0, 150 mM NaCl, 1 mM DTT). Fractions correlating to a distinct peak in UV absorbance (280 nm) were collected and analyzed by SDS-PAGE. Protein concentrations were determined using the absorbance at 280 nm and the theoretical molar extinction coefficient for Nsp9 (12 950 M^{-1} cm^{-1}).

2.2 | Size exclusion chromatography (SEC) analysis

Nsp9 was expressed and purified as described above at concentrations ~50 μM. The protein was then analyzed via size exclusion chromatography on a Superose 6 10/300 column (Cytiva) in SEC buffer at 0.3 ml/min (single peak at 18.11 mL retention volume). Standards were run on the

same column under the same conditions (β-amylase β-AML, M = 200 kDa; alcohol dehydrogenase ADH, M = 150 kDa; carbonic anhydrase CA, M = 29 kDa; bovine serum albumin BSA, M = 66 kDA and cytochrome C, M = 12 kDa) and the elution volumes were used to calculate the molecular weight of Nsp9 (23 ± 6 kDa) using the equations described in Irvine et al.[12] The calculated molecular weight corresponds to a dimer in solution (predicted molecular weight of 28 kDa).

2.3 | NMR spectroscopy

Purified SARS-CoV-2 Nsp9 samples (1.4-1.6 mM) were prepared in NMR buffer with 10% deuterium oxide (D_2O) and 20 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Spectra were acquired on a Bruker Avance III 600 or 800 MHz spectrometer at 25°C. The spectra recorded to obtain protein backbone chemical shift assignments include ^{15}N-HSQC, CBCA(CO)NH, HNCACB, HNCO, HN(CA)CO, and ^{15}N-NOESY. To map binding interfaces, HSQC experiments were carried out by adding ssDNA (oligo$[dT]_{23}$, c = 4.5 mM) and RNA (oligo$[dU]_{23}$, c = 5.1 mM) to 1.4 and 1.6 mM of ^{15}N-Nsp9, respectively, until no more chemical shift changes were observed (1:1 molecular ratio of protein: nucleic acid). HPLC-purified and lyophilised nucleic acids were purchased from Integrated DNA Technologies. Both ssDNA and RNA were dissolved in NMR buffer prior to recording NMR experiments. All spectra were processed using TOPSPIN3.x (Bruker, Biospin) with reference to the chemical shift of DSS (0 ppm) and assignments were made in Sparky (T. D. Goddard and D.G. Kneller, University of California at San Francisco). The calculation of weighted chemical shift changes was carried out as described[13] using OriginPro 2019 (OriginLab Corporation, Northampton, Massachusetts). All chemical shift assignments have been deposited at the BMRB under the accession number 50725.

2.4 | Biolayer interferometry

For each Nsp9 mutant protein and the wild type, a concentration series consisting of 4 to 5 solutions at concentrations between 90 and 2600 μM were prepared using NMR buffer. BLI steady-state analysis was carried out in which the proteins were bound to a 5′ biotinylated oligo $(dU)_{23}$ oligonucleotide in triplicate, using the BLItz biosensor system (ForteBio). Streptavidin biosensors (ForteBio) were equilibrated in NMR buffer for 30 minutes prior to use. For each individual binding curve, an initial baseline was carried out (30 s), followed by the binding of the oligonucleotide (1.3-1.6 μM) to the biosensor until saturation (120 s). A second baseline (60 s) was carried out to re-establish the baseline of the RNA-bound tip. Each sample was allowed 240 s to reach an equilibrium state, followed by a 120 s dissociation step in NMR buffer. Average corrected (protein binding without immobilized RNA) BLI equilibrium binding values (steady-state) were taken from the sensorgrams, plotted against the respective protein concentrations, and fitted using the Michaelis Menten equation ($R_{BLI} = (R_{max} \times c)/(c + K_D)$) in Origin 9.1 (Microcal).

2.5 | HADDOCK modeling

The protein structure of one monomer of Nsp9 (residue 4-113) was taken from the recently solved crystal structure of the Nsp9 dimer (PDB 6WXD)[4] and used as input for HADDOCK,[14,15] together with a model of RNA (oligo$(dU)_9$) constructed in silico using the software OPEN BABEL 2.3.90.[16] Nsp9 protein residues 7-9, 20, 35, 38, 40-43, 45, 55, 59, 63-67, 90-93, 99-100 and 102 were defined as active and semi-flexible residues based on NMR data and all nine uracils of the RNA were defined as active and flexible. 34 ambiguous interaction restraints (AIRs) for both the protein and the RNA were chosen based on our NMR data and fixed at 2 Å. For HADDOCK runs where base stacking between F40 and U8 was proposed based on the BLI data, additional restraints to maintain base planarity as well as AIRs between the aromatic residue and the RNA bases were used in the calculations. The 10 conformers with the lowest value of total energy of the lowest-energy cluster were analyzed and the lowest-energy structure visualized using PYMOL (Schrödinger, New York).

3 | RESULTS

3.1 | Nsp9 is a dimer in solution

A very recent study has already established that Nsp9 exists as a dimer in solution at concentrations as low as 1 μM with dissociation into monomers in the nM range.[17] After expression, we purified Nsp9 (for sequence information refer to Figure 1) for the purpose of NMR spectroscopy, size exclusion chromatography (SEC) and biolayer interferometry (BLI) experiments at concentrations between 50 and 2600 μM. To confirm that Nsp9 is indeed a dimer in solution at the lowest concentration used (50 μM), we carried out SEC experiments with a set of protein standards of known molecular weight (Figure 2). As expected, Nsp9 elutes at a volume that most closely corresponds to a dimer in solution with a calculated size of 23 ± 6 kDa.

3.2 | NMR backbone resonances of Nsp9 in solution

To determine the molecular details of ssDNA and RNA binding, we used NMR spectroscopy to assign backbone amide resonances of Nsp9. A combination of 2D and 3D NMR experiments enabled us to assign around 70% of the 113 residues of the SARS-CoV-2 Nsp9 monomer (indicated in red boxes in Figure 1A and shown in Table S1). In agreement with the presence of a symmetric dimer in the crystal structure,[4] only one set of peaks was observed. Relatively broad lines observed are also consistent with the dimeric state of the protein in solution and also supported by our size exclusion chromatography data (Figure 2). Apart from 10 peaks that exhibit substantial line broadening and could not be unambiguously assigned, the rest of the unassigned backbone signals were absent in the HSQC spectrum, most likely due to chemical exchange processes with the solvent.

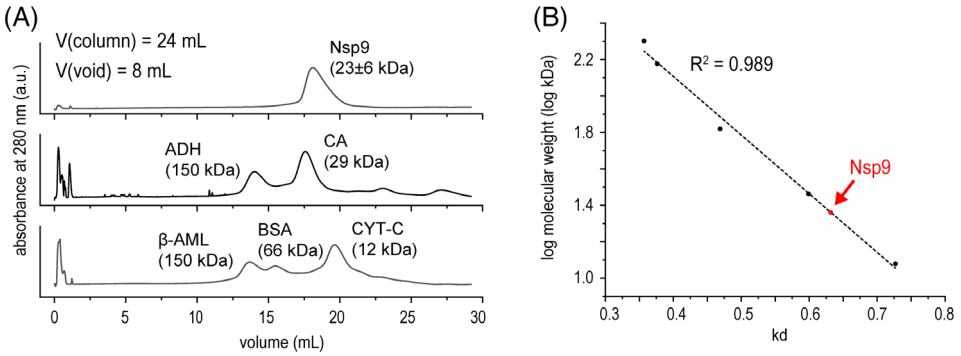


FIGURE 2 Size exclusion chromatography of Nsp9 confirming dimeric state in solution. (A) Size exclusion chromatography (SEC) of Nsp9 as well as five different proteins standards (β-amylase β-AML, alcohol dehydrogenase ADH, carbonic anhydrase CA, bovine serum albumin BSA, cytochrome C) run under the same conditions. (B) Calibration curve with the five standards and Nsp9 (denoted as red dot)

We used the backbone resonance obtained assignments to predict secondary structure elements of Nsp9 by TALOS-N[18] (Figure 3). Overall, the predictions are in very good agreement with the crystal structure[4] (shown on top of Figure 3).

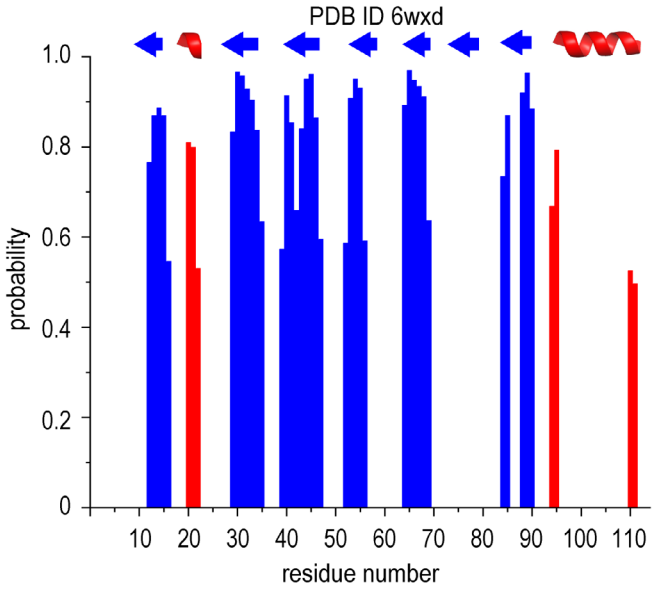


FIGURE 3 Secondary-structure prediction of Nsp9 based on backbone chemical shifts. TALOS-N prediction[18] of α-helices (red) and β-sheets (blue) in the Nsp9 structure based on NMR backbone chemical shift data. The top of the Figure depicts a schematic of the secondary structure elements as observed in the X-ray structure (PDB 6WXD)

3.3 | NMR reveals ssDNA/RNA binding residues of Nsp9

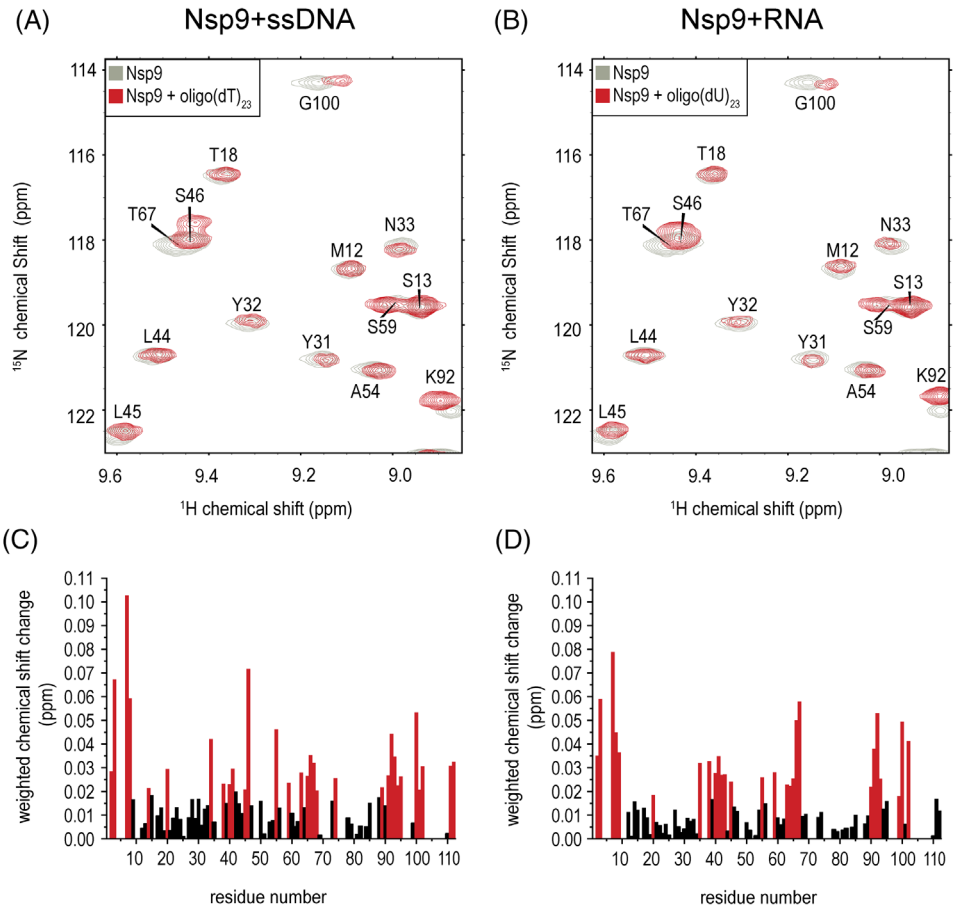
Next, to determine how ssDNA and RNA are recognized by Nsp9, we recorded HSQC spectra in the absence and presence of nucleic acids (Figure 4; for full spectra see Figure S1). RNA has been shown to bind Nsp9 extremely weakly,[4] we therefore utilized the highest concentrations of soluble protein and ssDNA/RNA that could be obtained (between 1.4-1.6 mM protein and 4.5-5.1 mM nucleic acid). To ensure that the length of the ssDNA/RNA is sufficient to recognize the entire Nsp9 dimer,[4] we utilized nucleic acid oligonucleotides comprising of 23 bases (oligo $(dT)_{23}$ and oligo $(dU)_{23}$, respectively). As is evident from Figure 4, the observed chemical shift changes are very small (Figure 4A,B show HSQC spectra at binding saturation, that is, no additional chemical changes were observed upon further addition of nucleic acid). Calculation of weighted chemical shift changes[13] for Nsp9 upon binding to ssDNA/RNA, revealed residues that undergo substantial changes in backbone structure (above the average of all observed chemical shift changes). These residues are either subject to distant conformational changes or more likely involved in nucleic acid binding (termed "binding residues", see Figure 4C,D; for a summary of all chemical shift changes see Table S1). Notably, the binding profiles of ssDNA and RNA are highly similar, indicating that the molecular details of the interaction of Nsp9 with single-stranded nucleic acids are conserved.

3.4 | RNA binding occurs via a distinct Nsp9 interaction interface

Next, we mapped the RNA binding residues as determined in Figure 4D onto the X-ray crystal structure of the Nsp9 dimer (PDB 6WXD)[4] (Figure 5, colored in red and Table S1). As seen in Figure 5, binding of RNA takes place on one side of the Nsp9 dimer (indicated as "binding interface" in Figure 5). The majority of binding residues are located on the external surface of the protein (at least 30% surface exposure), with the exception of a set of buried hydrophobic residues (V31, L42, A43, L45, F90, and V102).

3.5 | Structural model of Nsp9-RNA complex reveals important binding residues

To visualize nucleic acid binding to Nsp9, we next used our NMR data as well as the existing crystal structure of SARS-CoV-2 Nsp9[4] to calculate structural models of Nsp9-RNA complexes (Figure 6). The symmetry of the Nsp9 dimer and the fact that we did not observe any splitting of individual peaks in the HSQC upon RNA binding confirmed that each Nsp9 monomer recognizes RNA in an identical manner. Thus, only one SARS-CoV-2 Nsp9 monomer from the dimeric crystal[4] structure was utilized in our structural modeling. Preliminary docking



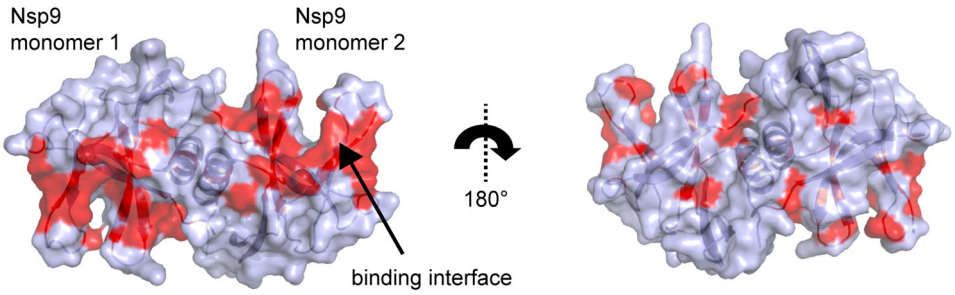


FIGURE 4 NMR analysis of Nsp9 protein in complex with ssDNA and RNA. Sections of ^{15}N-HSQC spectrum of Nsp9 in the absence (gray) and presence (1:1 mixture, red) of oligo(dT)$_{23}$ (A) and oligo (dU)$_{23}$ (B), respectively. Weighted backbone chemical shift changes of HN and N^{13} atoms for Nsp9 upon binding to ssDNA (C) and RNA (D). Residues exhibiting changes larger than the average ("binding residues") are colored in red. Note the similarity in the binding profile between ssDNA and RNA

FIGURE 5 NMR reveals distinct interaction surface. Nsp9 dimer structure (taken from PDB 6WXD) in surface and cartoon representation (colored in light blue). The RNA binding residues as determined in Figure 4C,D were mapped onto the deposited crystal structure. (red). Note that a distinct binding interface exists on one side of the dimer structure

calculations using HADDOCK[14,15] revealed that an RNA oligonucleotide nine bases in length (oligo (dU)$_9$, constructed in silico using the program BABEL[16]), is sufficient to recognize one Nsp9 protein molecule.

We first performed HADDOCK runs with all binding residues defined as active (AIRs) and oligo(dU)$_9$ RNA. Figure 6A depicts the lowest-energy structure from the lowest-energy cluster of 1000 calculated structures. Interactions between Nsp9 and the RNA occur mainly via the backbone of the nucleic acid, in agreement with the expectation that the binding is non-specific, ensuring a variety of different viral RNA sequences can be processed. Only two major electrostatic interactions were observed between positively charged R55 and K92 and the RNA backbone, which may explain the low binding affinity. Potential hydrogen bonds between two threonine residues (T35 and T64) and corresponding RNA bases were identified in most of the structural models; these residues are indicated as red boxes in Figure 1B.

3.6 | BLI experiments confirm structural model and binding mode via aromatic residue

Closer inspection of our NMR data (Figure 4) reveals that two aromatic residues present within the binding interface exhibit significant

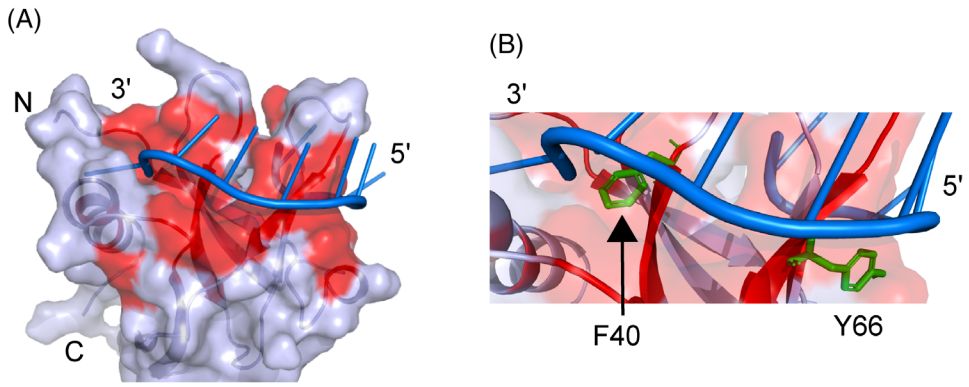


FIGURE 6 Structural models of Nsp9-RNA complex structures calculated using HADDOCK. (A) Surface and cartoon representation of an Nsp9-RNA complex model structure (containing one Nsp9 monomer) calculated using the binding residues (colored in red) as determined in Figure 4C,D in HADDOCK (Nsp9 is colored in light blue, RNA in dark blue). (B) Detailed view of the surface-exposed aromatic residues F40 and Y66 (stick representation; colored in green) that are located within the binding interface

chemical shift changes and are sufficiently exposed to the solvent: F40 and Y66 (Figure 6B; also boxed in Figure 1B). To determine if these aromatics could form Π-Π stacking interactions with RNA as well as to confirm the contribution by the other binding residues, we made a series of Nsp9 alanine mutants (F40A, R55A, Y66A, K92A, and R55A/K92A) based on our NMR data and binding model (Figures 4 and 6). Dissociation constants of the binding between Nsp9 as well as mutant Nsp9 proteins and RNA (oligo$(dU)_{23}$) were calculated using a steady-state analysis from BLI data (Figures 7 and S2). All Nsp9 mutant proteins were assessed for correct folding using 1D NMR spectroscopy (Figure 7F). The F40A mutant exhibits a substantial decrease in the binding affinity compared to the wild type protein (Figure 7A,B,E) strongly suggesting that F40 forms stacking interactions with the RNA. In contrast, replacing either R55 or Y66 with alanine did not result in a significant change in the binding strength (Figure 7C,D,E).

Notably, while binding of wild-type, F40A, R55A, and Y66 to RNA is instantaneous, interaction of both K92A and the double mutant R55A/K92A with the RNA is characterized by significantly longer association and dissociation phases with no steady-state reached within the observed time period (Figure S2). Consequently, the steady-state analysis could not be reliably performed for these two mutants, however, dissociation constants could be estimated from the kinetic binding data (K_D > 450 μM and > 1000 μM for K92A and R55/K92A, respectively) indicating that K92 plays a major role in RNA recognition, both due to the increase in the dissociation constant but also the change in the interaction kinetics. In contrast, R55 may only make a minor contribution to RNA binding as no significant effect on the RNA interaction of the single R55A mutant compared to the wild-type protein could be observed (Figure 7C).

3.7 | Refinement of Nsp9-RNA structural model

Based on our BLI data, we repeated our HADDOCK calculations in the presence of additional restraints between F40 and the RNA (AIRs and planar restraints). As seen in Figures 8A,B, the resulting structural model is very similar to the one shown in Figure 6A, however, additional base-stacking can now be observed between F40 and U8 (Figure 8C). Consistent with the significant but small chemical shift changes observed for F40, the overall structural change of the side chain of this residue upon binding is minor. Overall, apart from F40, the binding interface is mostly made up by positively charged residues including K36, K58, and K92 (Figure 8C), however, refinement of the model resulted in significantly fewer contacts of R55 compared to K92, in good agreement with our mutational data (Figure 7).

4 | DISCUSSION

4.1 | Comparison with other SSBs and viral Nsp9 proteins

It has been proposed that the structure of Nsp9 exhibits some topological similarities with OB (oligonucleotide/oligosaccharide binding) domain-containing proteins,[4] however, until now, the molecular details of RNA binding have not been resolved. In this study we present structural models of Nsp9-RNA complexes based on NMR chemical shift perturbation analysis and BLI data. Although mostly electrostatic in nature, we identify one aromatic residue (F40) within the interaction surface that most likely plays a key role in the recognition of RNA via base-stacking mechanisms.

A very recently published study used an RNA mimic (1,3-dimethyl-6H-pyrrolo[3,4-days]pyrimidine-2,4-dione; FR6) to probe nucleic acid binding by Nsp9.[19] Figure S3 depicts the determined crystal structure of that work (PDB 7KRI) revealing that FR6 engages Nsp9 via a tetrameric Π-Π stacking between F40 and the terminal bases of FR6 inducing the formation of a parallel Nsp9 trimer-of-dimers. These data confirm that F40 is indeed able to stack with RNA bases, providing further evidence of the validity of our model.

Π-Π stacking mechanisms are very common in OB domain containing proteins that recognize ssDNA or RNA where up to four

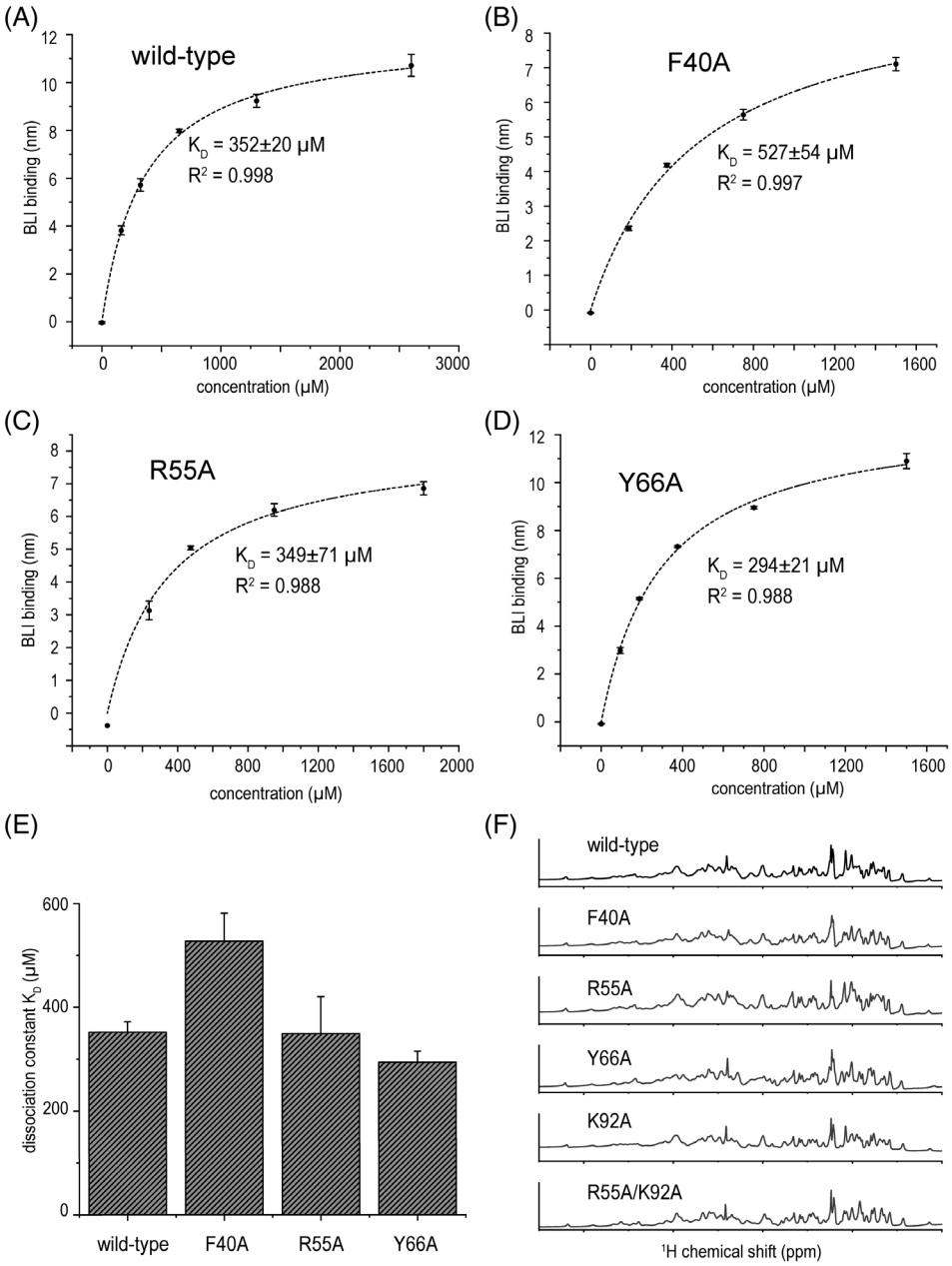


FIGURE 7 Mutational analysis revealing critical RNA binding residues of Nsp9. (A–D) Average (±SE) steady-state equilibrium BLI values and fits to the steady-state binding model from three independent binding experiments of wild-type Nsp9 and mutants (F40A, R55A, and Y66A). Adjusted R^2 as well as dissociation constants are shown. (E) Summary of dissociation constants (± SE) for Nsp9 and alanine mutants, for binding to oligo(dU)$_{23}$, as measured by BLI. (F) ^{1}H NMR spectra of Nsp9 alanine mutants show that each is correctly folded. Spectra were recorded at concentrations of between 200 and 1000 μM at 25°C

aromatic residues within one OB domain are utilized.[20] However, Nsp9 differs both in the location and composition of the RNA binding site (as determined in this study) as well as the binding affinity, suggesting that no significant structural similarities exist between the two protein families.

Although F40 is not conserved throughout the coronavirus family, other aromatics in close proximity to its position (Figure 1B) could potentially be utilized in viral Nps9 analogues to form similar base-stacking interactions to those proposed in our model. In good agreement with our BLI data, K92 is almost entirely preserved in all related Nsp9 proteins (Figure 1B).

The low interaction affinity of the Nsp9-RNA interaction as measured by our BLI data (∼350 μM) is most likely due to the few electrostatic contacts observed in our structural modeling. Similar weak RNA binding affinities have been observed recently both for SARS-CoV-2 Nsp9 (>500 μM)[4] and other related viral Nsp9 proteins such as PDCoV (∼400 μM)[21] (Figure 1B).

Several studies have shown that the binding strength depends on the oligomerisation state of Nsp9 (as mutants that disrupt dimerisation display lower affinities) as well as the length of the used oligonucleotide.[21-23] These data are consistent with our NMR experiments revealing that each protein monomer recognizes RNA in an identical manner. Furthermore, the interaction interface is located on one side of the Nsp9 dimer, allowing for one continuous strand of RNA to bind the Nsp9 dimer using the binding residues described by our structural model.

4.2 | NMR assignments and comparison with recently published work

In this study, we have used NMR to obtain backbone chemical shift assignments of Nsp9 under physiological conditions. Using standard triple resonance approaches we have been able to assign the backbone chain atoms of 77 out of the 113 residues of Nsp9 (red boxed in Figure 1A and listed in Table S1) and have identified 27 out of 29 residues that exhibit significant chemical shift changes upon binding of ssDNA/RNA (Figure 4). During the course of our work, another two groups published near-complete solution-state backbone assignments of Nsp9,[17,24] one of which was available in the BMRB database (50513), along with our data (50725). Overall, our assignments are in good agreement with the published data with only minor differences observed, which could be explained by the different pH value used in our final NMR buffer (6) compared to the other two studies (7).

Notably, one of the two studies (Buchko et al.)[17] revealed that Nsp9 exists predominantly as a dimer at protein concentrations >100 μM, in excellent agreement with our size exclusion data (Figure 2). However, although not observed at concentrations used in our NMR experiments, Buchko et al. suggested that dissociation into monomers could still play a functional role for these proteins, for example through binding to other proteins (such as Nsp12 in the context of RTC formation; see next section).

4.3 | The Nsp9-RNA model in the context of the RTC structure

A very recent study has revealed the structural basis of assembly of the SARS-CoV-2 replication and transcription complex (RTC), including the Nsp9 protein.[5] Although an interaction of Nsp9 with Nsp8 has been proposed in the past,[6] no direct binding between these two

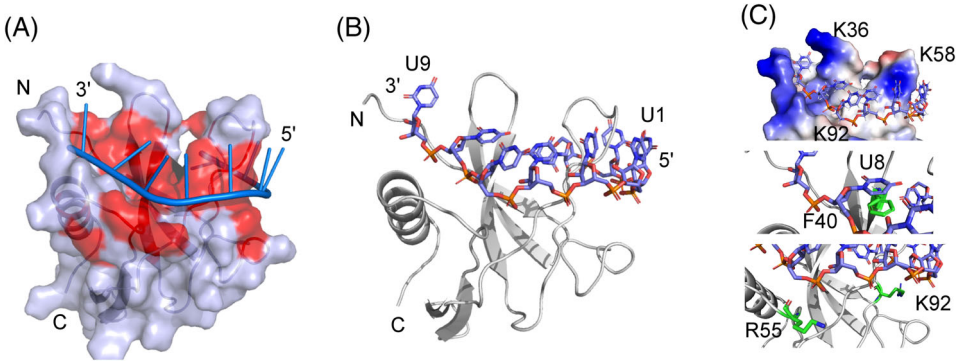


FIGURE 8 Refined HADDOCK model of Nsp9-RNA complex. (A) Surface and cartoon representation of an Nsp9-RNA complex model structure calculated as in Figure 6 with the addition of restraints to achieve base-stacking between F40 with the corresponding RNA base (Nsp9 is colored in light blue, RNA in dark blue). (B) Same as in A with the protein (light gray) and the RNA (dark blue) shown as cartoon and stick representation, respectively. (C) Electrostatic potential (blue = positive, red = negative) of RNA binding interface with positively charged binding residues indicated (top) as well as detailed view (as stick and cartoon representation) of Π- Π stacking interaction between F40 and U8 (middle) and the location of the two electrostatic residues R55 and K92 (colored in green; bottom)

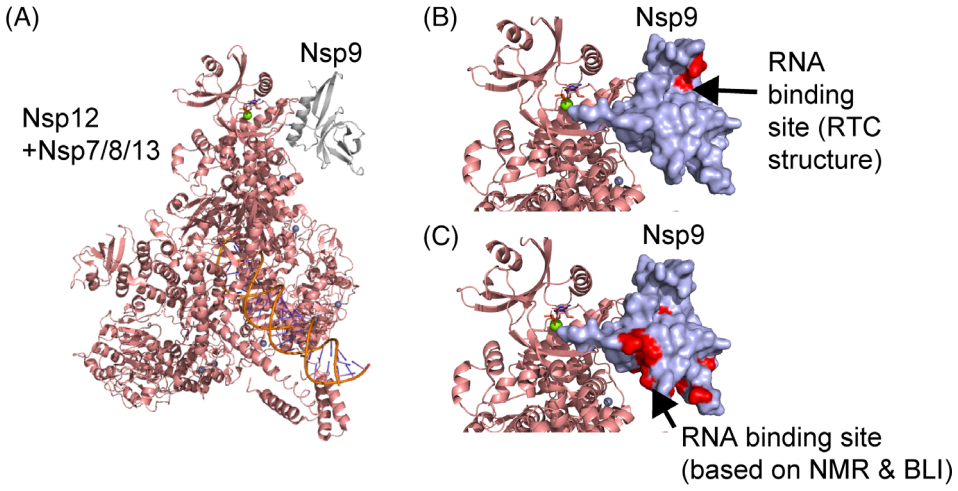


FIGURE 9 The RNA binding interface of Nsp9 within the replication and transcription complex (RTC). (A) Structure of the RTC[5] taken from PDB 7CYQ with Nsp9 colored in gray and Nsp7/8/12/13 proteins in salmon (cartoon representation). (B,C) Detailed view of RNA interaction surface as proposed in[5] (B) as opposed to in this study (C). Nsp9 is shown as surface representation (light blue) and the RNA binding site is colored in red. Note that the proposed binding site (indicated by the arrow) differs from the one determined in our structural modeling

proteins has been observed in the RTC structure. Indeed, Nsp9 appears to only interact with the RNA polymerase (RdRp, Nsp12)[5] (Figure 9A). This interaction is characterized by the insertion of the amino terminus of Nsp9 (one of the regions of very high sequence conservation; Figure 1B) to the NiRAN domain of Nsp12, affecting the catalytic activity of the polymerase. Nsp9 binds as a monomer which can be explained by the fact that apart from the amino terminus, residues that form part of the monomer-monomer interface in the dimeric crystal structure[4] play an important role in the binding to Nsp12. These data indicate that Nsp9 dimers, although being the dominant species in solution (apart from very low concentrations), are not required for RTC function.

Figure 9B,C show a comparison between the proposed[5] and our determined RNA binding interface, respectively. Importantly, two (K81 and K84) out of the three positively charged residues (K81, K84, K86) of the proposed binding interface[5] in Figure 9B (indicated by the arrow) have been assigned by NMR in our work and exhibit no significant chemical shift changes upon RNA binding (see also Table S1), further confirming that the recognition of RNA takes place on the opposite site as proposed in our model (Figure 9C). The 3′ end of the RNA in our structural model points towards the flexible amino terminus of Nsp9 (consistent with the large chemical shift changes observed in the RNA titration experiments; Figure 4) and is in close proximity to two positively charged Nsp12 residues (R733 and R735) that may be able to direct RNA binding towards the polymerase.

RNA processing within the RTC in viral replication is a complex process and the role of Nsp9 is not well understood. Although it is proposed that Nsp9 stabilizes the 5′ end of the processed RNA, it is not clear how the RNA is directed to Nsp9, given the large distance (~ 110 Å) between the catalytic centre of Nsp12 NiRAN (RNA exit site) and the Nsp9 binding site.[5] Our study resolves some of the ambiguities around the exact location of RNA binding within Nsp9, however, further studies will be required to shed light onto these important processes given the urgent need to develop new therapeutic strategies against COVID-19.

ACKNOWLEDGMENTS

This research was funded by grants from Western Sydney University and facilitated by access to Sydney Analytical, a core research facility at the University of Sydney; in particular we thank Dr Mario Torrado Del Rey for assistance and advice with work related to protein expression and purification.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/prot.26205.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Roland Gamsjaeger https://orcid.org/0000-0003-1095-2569

REFERENCES

1. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol*. 2020;92(4):418-423.
2. Wu F, Zhao S, Yu B, et al. Author correction: a new coronavirus associated with human respiratory disease in China. *Nature*. 2020;580(7803):E7-E7.
3. Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells (Basel, Switzerland)*. 2020;9(5):1267.
4. Littler DR, Gully BS, Colson RN, Rossjohn J. Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. *iScience*. 2020;23(7):101258-101258.
5. Yan L, Ge J, Zheng L, et al. Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell*. 2021;184(1):184-193.e110.
6. Sutton G, Fry E, Carter L, et al. The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure*. 2004;12(2):341-353.
7. Richard DJ, Bolderson E, Khanna KK. Multiple human single-stranded DNA binding proteins function in genome maintenance: structural, biochemical and functional analysis. *Crit Rev Biochem Mol Biol*. 2009;44:98-116.
8. Ashton NW, Bolderson E, Cubeddu L, O'Byrne KJ, Richard DJ. Human single-stranded DNA binding proteins are essential for maintaining genomic stability. *BMC Mol Biol*. 2013;14(1):9-9.
9. Touma C, Adams MN, Ashton NW, et al. A data-driven structural model of hSSB1 (NABP2/OBFC2B) self-oligomerization. *Nucleic Acids Res*. 2017;45(14):8609-8620.
10. Egloff MP, Ferron F, Campanacci V, et al. 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*. 2004;101(11):3792-3796.
11. Cai M, Huang Y, Sakaguchi K, Clore GM, Gronenborn AM, Craigie R. An efficient and cost-effective isotope labeling protocol for proteins expressed in *Escherichia coli*. *J Biomol NMR*. 1998;11(1):97-102.
12. Irvine GB. Determination of molecular size by size-exclusion chromatography (gel filtration). *Curr Protoc Cell Biol*. 2001;6:5.5.1-5.5.16.
13. Ayed A, Mulder FAA, Yi G-S, Lu Y, Kay LE, Arrowsmith CH. Latent and active p53 are identical in conformation. *Nat Struct Biol*. 2001;8:756-760.
14. de Vries SJ, van Dijk AD, Krzeminski M, et al. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*. 2007;69(4):726-733.
15. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*. 2003;125(7):1731-1737.
16. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Chem*. 2011;3:33.
17. Buchko GW, Zhou M, Craig JK, Van Voorhis WC, Myler PJ. Backbone chemical shift assignments for the SARS-CoV-2 non-structural protein Nsp9: intermediate (ms - mus) dynamics in the C-terminal helix at the dimer interface. *Biomol NMR Assign*. 2021;15:107-116.
18. Shen Y, Bax A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR*. 2013;56(3):227-241.
19. Littler DR, Mohanty B, Lowery SA, et al. Binding of a pyrimidine RNA base-mimic to SARS-CoV-2 non-structural protein 9. *J Biol Chem*. 2021;019103R2. http://dx.doi.org/10.1016/j.jbc.2021.101018

20. Dickey TH, Altschuler SE, Wuttke DS. Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure*. 2013; 21(7):1074-1084.
21. Zeng Z, Deng F, Shi K, et al. Dimerization of coronavirus nsp9 with diverse modes enhances its nucleic acid binding affinity. *J Virol*. 2018; 92(17):e00692–18.
22. Hu T, Chen C, Li H, et al. Structural basis for dimerization and RNA binding of avian infectious bronchitis virus nsp9. *Protein Sci*. 2017; 26(5):1037-1048.
23. Miknis ZJ, Donaldson EF, Umland TC, Rimmer RA, Baric RS, Schultz LW. Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J Virol*. 2009;83(7): 3007-3018.
24. Dudas EF, Puglisi R, Korn SM, et al. Backbone chemical shift spectral assignments of SARS coronavirus-2 non-structural protein nsp9. *Biomol NMR Assign*. 2021. https://doi.org/10.1007/s12104-021-10011-0

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: El-Kamand S, Du Plessis M-D, Breen N, et al. A distinct ssDNA/RNA binding interface in the Nsp9 protein from SARS-CoV-2. *Proteins*. 2021;1-10. doi: 10.1002/prot.26205