Biomolecular NMR Assignments (2021) 15:107–116
https://doi.org/10.1007/s12104-020-09992-1

ARTICLE

Backbone chemical shift assignments for the SARS-CoV-2 non-structural protein Nsp9: intermediate (ms – µs) dynamics in the C-terminal helix at the dimer interface

Garry W. Buchko[1,2,3] · Mowei Zhou[2] · Justin K. Craig[1,4] · Wesley C. Van Voorhis[1,4] · Peter J. Myler[1,5,6]

Received: 16 October 2020 / Accepted: 24 November 2020 / Published online: 4 January 2021

Abstract

The *Betacoronavirus* SARS-CoV-2 non-structural protein Nsp9 is a 113-residue protein that is essential for viral replication, and consequently, a potential target for the development of therapeutics against COVID19 infections. To capture insights into the dynamics of the protein's backbone in solution and accelerate the identification and mapping of ligand-binding surfaces through chemical shift perturbation studies, the backbone ^{1}H, ^{13}C, and ^{15}N NMR chemical shifts for Nsp9 have been extensively assigned. These assignments were assisted by the preparation of an ~ 70% deuterated sample and residue-specific, ^{15}N-labelled samples (V, L, M, F, and K). A major feature of the assignments was the "missing" amide resonances for N96-L106 in the ^{1}H-^{15}N HSQC spectrum, a region that comprises almost the complete C-terminal α-helix that forms a major part of the homodimer interface in the crystal structure of SARS-CoV-2 Nsp9, suggesting this region either undergoes intermediate motion in the ms to µs timescale and/or is heterogenous. These "missing" amide resonances do not unambiguously appear in the ^{1}H-^{15}N HSQC spectrum of SARS-CoV-2 Nsp9 collected at a concentration of 0.0007 mM. At this concentration, at the detection limit, native mass spectrometry indicates the protein is exclusively in the monomeric state, suggesting the intermediate motion in the C-terminal of Nsp9 may be due to intramolecular dynamics. Perhaps this intermediate ms to µs timescale dynamics is the physical basis for a previously suggested "fluidity" of the C-terminal helix that may be responsible for homophilic (Nsp9-Nsp9) and postulated heterophilic (Nsp9-Unknown) protein-protein interactions.

Keywords SARS-CoV-2 · Non-structural protein · Native mass spectrometry · Viral replication · Solution NMR · Protein dynamics

✉ Garry W. Buchko
garry.buchko@pnnl.gov

1 Seattle Structural Genomics Center for Infectious Disease, Seattle, USA

2 Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington, USA

3 School of Molecular Biosciences, Washington State University, Pullman, Washington, USA

4 Department of Medicine, Division of Allergy and Infectious Disease, Center for Emerging and Re-emerging Infectious Disease, University of Washington, Seattle, Washington, USA

5 Seattle Children's Research Institute, Seattle, Washington, USA

6 Department of Pediatrics, Department of Medical Education and Biomedical Informatics, Department of Global Health, University of Washington, Seattle, Washington, USA

Biological context

A highly contagious novel coronavirus, SARS-CoV-2, is the agent responsible for the respiratory disease COVID19 and the global pandemic of 2020. This pressing global emergency demands the rapid development of a vaccine or antiviral drugs to halt the destructive impact of this disease. Almost 30-kb in length, the positive-sense single-stranded SARS-CoV-2 RNA is one of the largest known viral genomes (Wu et al. 2020), encoding 16 non-structural proteins (Nsps), four structural proteins, and potentially nine accessory proteins (open reading frames, Orfs) (Gordon et al. 2020b). The large number of proteins are suggestive of a complex life cycle and represent a large number of potential drug targets. The two main proteases (Jin et al. 2020), Nsp5 (Mpro) and Nsp3 (PLpro), and the proteins that comprise the RNA-dependent RNA-polymerase apparatus (Gordon et al. 2020a), Nsp7, Nsp8, and Nsp12, are

the primary focus for small molecule therapeutics targeting viral components. However, many of the other Nsp and Orf proteins generated by the virus interact with host proteins (Gordon et al. 2020b) and have been associated with viral replication roles (Tan et al. 2005), and consequently, may also be viable drug targets.

Nsp9 is one of the 11 proteins generated by the translation and proteolytic processing of SARS-CoV-2 Orf1a. The primary amino sequence of SARS-CoV-2 Nsp9 is 97% identical to the corresponding protein in the related SARS-CoV virus (SARS) with the latter protein shown to be essential for viral replication (Frieman et al. 2012). While the precise functional role of Nsp9 is unknown, SARS-CoV Nsp9 has affinity for single-stranded DNA and RNA oligonucleotides (Egloff et al. 2004) and has been reported to interact with the RNA-dependent RNA-polymerase apparatus (Nsp7 and Nsp8) (Chen et al. 2017; Sutton et al. 2004), and hence, is associated with viral replication. Crystal structures have been solved for SARS-CoV (Egloff et al. 2004; Sutton et al. 2004) and SARS-CoV-2 (Littler et al. 2020) Nsp9. Both Nsp9 structures are similar (backbone 113-residue Cα RMSD of 0.57 Å) (Littler et al. 2020), bearing resemblance to the oligonucleotide/oligosaccharide binding-fold (OB-fold) but with an overall topology unique to coronaviruses (Littler et al. 2020). The crystal structures suggest the biologically relevant form of the protein is a dimer that is predominantly stabilized by strong backbone van der Waals interactions between a conserved, anti-parallel, α-helical GXXXG motif at the protein-protein interface. To accelerate our understanding of the biological function of Nsp9 and the discovery of new therapeutic strategies for COVID19, we report here the near-complete backbone assignments for SARS-CoV-2 Nsp9 that will enable residue-specific tracking of interactions with other proteins and potential drug candidates through chemical shift perturbation studies (Buchko et al. 1999; Zuiderweg 2002).

Methods and experiments

Cloning, expression, and purification

A codon optimized expression construct corresponding to the processed protein Nsp9 from the *Betacoronavirus* SARS-CoV-2 (2019-nCoV; COVID-19) Wuhan-Hu-1 isolate (Genbank MN908947.3) was synthesized and inserted in the pET28a-TEV vector at the NdeI/NotI restriction enzyme sites by Genescript (Piscataway, NJ). The recombinant plasmid was then used to transform chemically competent *Escherichia coli* Rosetta BL21(DE3)pLyS cells (Novagen, Millipore Sigma, Burlington, MA) by a heat shock method. The expressed gene product contained a 22 amino acid extension, MGSSHHHHHHSSGENLYFQGHM- at the N-terminus of the native protein to enable protein purification by metal chelation chromatography (Choi et al.., 2011). The SSGCID internal ID for the SARS-CoV-2 Nsp9 construct is BewuA.18928.a.MX101 (www.ssgcid.org).

Approximately 70% deuterated, uniformly ^{15}N-, ^{13}C-labeled protein was obtained by growing the transformed cells (310 K) in 750 mL of minimal medium (Miller) containing ~ 70% $^{2}H_2O$ (v), $^{15}NH_4Cl$ (1 mg/mL) and D-[$^{13}C_6$] glucose (2.0 mg/mL), NaCl (50 µg/mL), $MgSO_4$ (120 µg/mL), $CaCl_2$ (11 µg/mL), Fe_2Cl_3 (10 ng/mL) and the antibiotics chloramphenicol (35 µg/mL) and kanamycin (34 µg/mL). When the cell culture reached an OD_{600} reading of ~ 0.8, it was transferred to a 298 K incubator and protein expression induced with isopropyl β-D-1-thiogalactopyranoside (0.026 µg/mL). Cells were harvested by mild centrifugation following overnight incubation and then frozen at 193 K. After thawing the frozen pellet, Nsp9 was purified with a conventional two-step protocol involving metal chelate affinity chromatography on a 20 mL Ni-Agarose 6 FastFlow column (GE Healthcare, Piscataway, NJ) followed by gel-filtration chromatography on a Superdex75 HiLoad 26/60 column (GE Healthcare, Piscataway, NJ). In addition to removing minor impurities, the latter step exchanged Nsp9 into the buffer used for the NMR studies: 100 mM NaCl, 20 mM Tris, 1.0 mM dithiothreitol, pH 7.0.

Nitrogen-15 labelled Nsp9 was prepared using an autoinduction protocol and the protein purified as described above. The concentrated protein was diluted 1:1 with a Tobacco Etch Virus (TEV) Buffer (50 mM TrisHCl, 150 mM NaCl, pH 7.8) to a final volume of approximately 1 mL (~ 10 mg/mL) and treated overnight at 278 K with 20 uL of TEV protease (0.5 ug/mL, prepared in house) to remove the N-terminal tag. The cleaved protein, which contained three N-terminal scar residues afterwards (GHM-), was purified by re-application on the size exclusion column. To assist amide assignments, residue-specific, ^{15}N-labeled Leu, Val, Phe, Lys, and Met samples were prepared using a modified "Redfield-medium" (Arachchige et al. 2018) and the protein purified as described above.

Nuclear magnetic resonance spectroscopy

All NMR data used for the chemical shift assignments were collected at 303 K on a triple-labeled (~ 70% ^{2}H, ^{13}C, ^{15}N) sample of Nsp9 (~ 0.5 mM) using Agilent Inova-600 spectrometers equipped with an HCN-cryoprobe and pulse field gradients. The backbone ^{1}H, ^{13}C, and ^{15}N chemical shifts were assigned from the analysis of two-dimensional ^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC spectra and three-dimensional HNCA, HNCOCA, HNCACB-(^{13}Cβ-optimized), CBCA(CO)NH, HCC-TOCSY-NNH, CC-TOCSY-NNH, HNCO, and HNCACO spectra collected with Agilent Biopack pulse programs. These experiments were also

used to assign the side chain chemical shifts for most of the β-carbons and other side chain carbons when possible. Three-dimensional ^{15}N-edited TOCSY-HSQC and NOESY-HSQC, ^{13}C-edited NOESY-HSQC (aliphathic), and HCC-TOCSY-NNH spectra were analyzed to obtain backbone ^{1}Hα and some side chain ^{1}H assignments. A two-dimensional HBCBCGCDHD experiment was used to assign tyrosine ^{13}Cδ atoms. An overall rotational correlation time, τ_c, was estimated at 293 K for Nsp9 (no tag) at 0.7, 0.07, and 0.007 mM concentrations from the ratio of collective backbone amide ^{15}N T_1 and T_{1rho} measurements (Szyperski et al. 2002). Steady-state {^{1}H}-^{15}N heteronuclear NOE values ($NOE = I_{sat}/I_{unsat}$) were measured in triplicate from the ratios of ^{1}H-^{15}N HSQC cross peak heights in spectra recorded with (I_{sat}) and without (I_{unsat}) three seconds of ^{1}H presaturation prior to the ^{15}N excitation pulse (Buchko et al. 1999). Nitrogen-15 T_1 and T_2 values were measured from ^{1}H to ^{15}N HSQC-based experiments using seven different time delays (s): 0.0, 0.1, 0.2, 0.3. 0.4, 0.6, 0.8, and 1.0 (T_1); 0.01, 0.03, 0.05, 0.07, 0.09, 0.11, and 0.13 (T_2). The T_1 and T_2 values and their associated errors were obtaining using the rate analysis function in NMRFAM-SPARKY (v1.414) measuring amide cross peak heights. Chemical shift data was obtained with 70% deuterated, uniformly ^{15}N-, ^{13}C-labeled Nsp9 containing the 22-residue, N-terminal tag while the relaxation data ({^{1}H}-^{15}N NOE, T_1, and T_2) was obtained with ^{15}N-labeled Nsp9 with the tag N-terminal tag removed. The ^{1}H, ^{13}C, and ^{15}N chemical shifts were referenced using indirect methods (DSS = 0 ppm) and deposited into the BioMagResBank database (www.bmrb.wisc.edu), without correction for any deuteration effects, under the accession number BMRB-50513.

Native mass spectrometry

Nanoelectrospray capillaries were made from borosilicate glass (OD 1 mm, ID 0.78 mm, 10 cm length, Sutter Instrument, item # BF100-78-10) using a micropipette puller (Sutter Instrument, model P-1000). The puller program was set to heat 495, pull 5, velocity 40, delay 235, pressure 200, and ramp 496. The resulting capillaries had openings of ~ 1 µm. Ten µL protein solution (~ 17 mg/mL) was buffer exchanged into 100 mM ammonium acetate (pH ~ 6.7) using a size exclusion spin column (Zeba Micro Spin Desalting Columns, 7 kDa MWCO, 75 µL) following manufacturer's protocol. The resulting solution had a concentration of ~ 13 mg/mL as estimated by Nanodrop A280 measurement. Mass spectra were collected on an ion mobility–mass spectrometer (Waters Synapt G2s-i). The solution was further diluted to a desirable concentration and loaded into the glass capillary, where a platinum wire was inserted to apply a voltage of 500 V to sustain a nano-electrospray. Source temperature was set to 303 K and the sampling cone to 50 V to minimize protein activation. Gas flows were set to 3 mL/min for trap, 120 mL/min for helium cell, and 60 mL/min for ion mobility. Bias voltages were set to 45 V for trap, 35 V for helium cell, and 3 V for IMS. Traveling waves were set to 150 m/s 3 V for trap, 500 m/s 20 V for ion mobility, and 70 m/s 3 V for transfer. Ion mobility mass spectra were average over 1 min. Data were visualized in Driftscope and annotated manually.

Extent of assignments and data deposition

Crystal structures of SARS-CoV (Egloff et al. 2004; Sutton et al. 2004) and SARS-CoV-2 (Littler et al. 2020) have been solved. On the basis of the buried surface area at dimer interfaces these structures suggest the protein is a homodimer with the protomer forming a single compact domain composed of one α-helix and seven β-strands arranged into a cone shaped β-barrel. The dimeric state for Nsp9 has been corroborated in solution at protein concentrations greater than ~ 0.1 mM by dynamic light scattering, analytical ultracentrifugation, and gel-filtration chromatography (Campanacci et al. 2003; Egloff et al. 2004; Sutton et al. 2004). NMR properties observed for SARS-CoV-2 Nsp9 at 0.5 mM further confirm this conclusion. Foremost, there was poor magnetization transfer in the three-dimensional backbone NMR experiments, an observation that would not be expected for a monomeric, ~ 14.8 kDa protein (with tag) (Buchko et al. 2015). As a consequence, it was necessary to prepare an ~ 70% perdeuterated sample to collect the three-dimensional NMR backbone assignment data. Note that the tag was left on the protein and deuteration performed at only the ~ 70% level for the sample used in the backbone NMR assignment experiments because of poor protein expression yields in minimal media. Due to the properties of the dimer and the level of deuteration of the NMR sample, unambiguous chemical shift assignments were primarily limited to backbone ($^{1}H^{N}$, ^{15}N, ^{13}C', and ^{13}Cα) and ^{13}Cβ atoms.

Figure 1a is the ^{1}H-^{15}N HSQC spectrum for SARS-CoV-2 Nsp9 with the unambiguously assigned amide resonances labeled. Because the ^{1}H-^{15}N HSQC spectrum for Nsp9 without the N-terminal tag was superimposable upon the spectrum of Nsp9 with the N-terminal tag, the former spectrum is shown and discussed for simplicity. Note that even with ~ 70% perdeuteration and data collection at 303 K, magnetization transfer was still poor through many of the amides, especially with the CC-TOCSY-NNH experiment, necessitating the use of residue-specific, ^{15}N-labelled samples to confirm some amide assignments. In the end, all the backbone amide cross peaks in the spectrum could be assigned except for the four resonances circled in cyan in Fig. 1a. As illustrated in the Nsp9 assignment summary on the top of Fig. 1b, these four amide resonances belong to the group of 20 residues with missing or unassigned amide cross peaks colored red in the primary amino acid sequence.

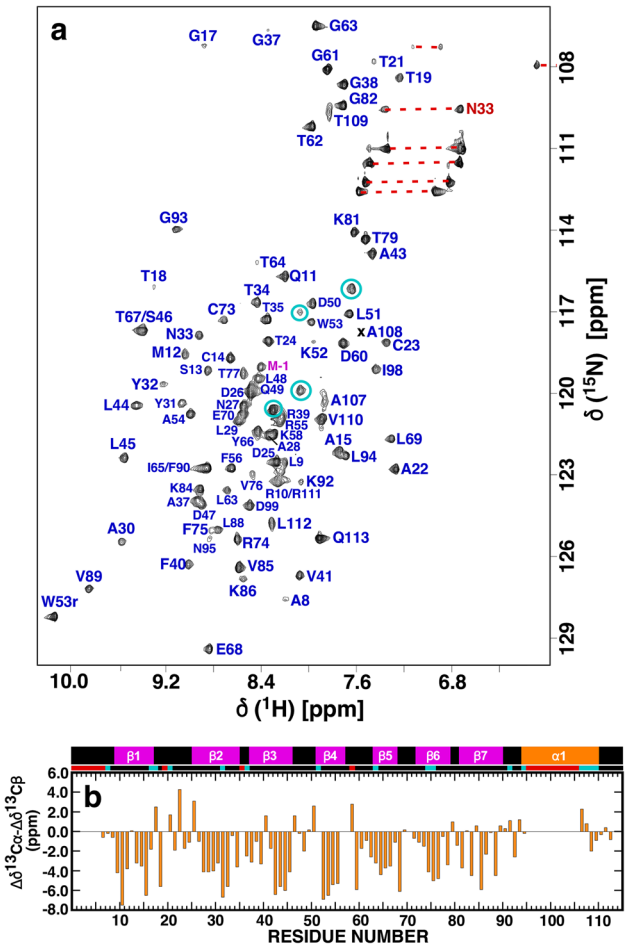

Fig. 1 **a** Assigned ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled Nsp9 (~0.5 mM) collected at a proton resonance frequency of 600 MHz, 303 K, in 100 mM NaCl, 20 mM Tris, 1 mM DTT, pH 7.0. Amide cross peaks for the three "scar" residues (G-3–M-1) and the 113-residue native protein (N1 - V113) are colored magenta and blue, respectively, with the side chain resonances identified with a red horizontal line. **b** Plot of the backbone combined Cα/Cβ chemical shift differences from random coil values, $\Delta\delta^{13}C\alpha$ - $\Delta\delta^{13}C\beta$ for SARS-CoV-2 Nsp9. On top of the plot is a schematic representation of the elements of secondary structure observed in the crystal of SARS-CoV-2 Nsp9 (PDB ID 6WXD): β-strands = magenta, α-helix = orange. Residues with missing or unassigned resonances in the ^{1}H-^{15}N HSQC spectrum of SARS-CoV-2 Nsp9 are identified with red blocks below scheme and residues with weak amide cross peak intensities relative to the majority of amide resonances are identified with cyan blocks

Hence, there are no amide cross peaks for 16 of the 109 expected amide resonances (116 – (N-terminal + 6 Pro)). These missing or unassigned residues are largely clustered in two regions of the protein, near the N- (N1-V7) and C- (N96-L106) termini. The remaining missing or unassigned residues are in loops between the elements of secondary structure observed in the crystal structure of Nsp9 (PDB ID 6WXD).

Despite the poor magnetization transfer in the three-dimensional NMR backbone assignment experiments it was possible to assign most of the backbone ^{13}Cα and side chain ^{13}Cβ atoms for residues with assigned amides through the analysis of complementary experimental pairs (primarily HNCA and HNCOCA, HNCACB-(^{13}Cβ-optimized) and CBCA(CO)NH). Deviations of the observed ^{13}Cα and ^{13}Cβ chemical shifts from random coil values reflect the secondary structure at these residue positions (Hafsa et al. 2015). In general these deviations move in opposite directions for α-helices (^{13}Cα = downfield; ^{13}Cβ = upfield) and β-strands (^{13}Cα = upfield; ^{13}Cβ = downfield), and hence, the difference between the ^{13}Cα and ^{13}Cβ chemical shift deviations ($\Delta\delta^{13}C\alpha$ - $\Delta\delta^{13}C\beta$) is a reference-independent measure of secondary structure (α-helix = positive and β-strand = negative, with four consecutive values in one direction usually the criteria for identifying a region of secondary structure). Figure 1b is a summary of these chemical shift calculations plotted below the elements of secondary structure observed in the crystal structure for SARS-CoV-2 Nsp 9 (Littler et al. 2020). In general, the negative $\Delta\delta^{13}C\alpha$ - $\Delta\delta^{13}C\beta$ values follow the pattern for the seven β-strands in the crystal structure of SARS-CoV-2 Nsp9 suggesting the protein adopts a similar structure in solution. Such a chemical shift analysis is not possible for most of a1 however due to missing or unassignable amide cross peaks for most of the residues in this region (N96-L106).

SARS-CoV-2 Nsp9 dimer interface–intermediate (ms to μs) dynamics

Missing amide cross peaks in ^{1}H-^{15}N HSQC spectra of protein provide useful information on a protein's biophysical properties because they typically identify the presence of a paramagnetic species (Shaheen et al. 2020) or regions undergoing motion or chemical exchange in the intermediate (ms to μs) timescale (Aden et al. 2011; Buchko et al. 2015). This motion or exchange can include millisecond fluctuations of side chains and/or monomer-dimer equilibria dynamics (Buchko et al. 2015). Figure 2a is a more detailed summary of the missing or unassigned amide resonances in the ^{1}H-^{15}N HSQC spectrum of SARS-CoV-2 Nsp9 and the amide cross peaks with weak intensity, colored red and cyan, respectively. The latter cross peaks may identify residues approaching motion or exchange in the intermediate (ms to μs) timescale and the majority in Nsp9 are adjacent to, or near, residues with missing or unassigned amide cross peaks. Following the same color scheme, Fig. 2b summarizes the spatial location of the residues with missing or unassigned amide resonances and the residues with weak amide resonances intensities in the ^{1}H-^{15}N HSQC spectrum of Nsp9 on a cartoon representation of the crystal structure for SARS-CoV-2 Nsp9 (PDB ID 6WXD). The missing or unassigned residues near the N-terminus and between elements of secondary are largely exposed on the surface of

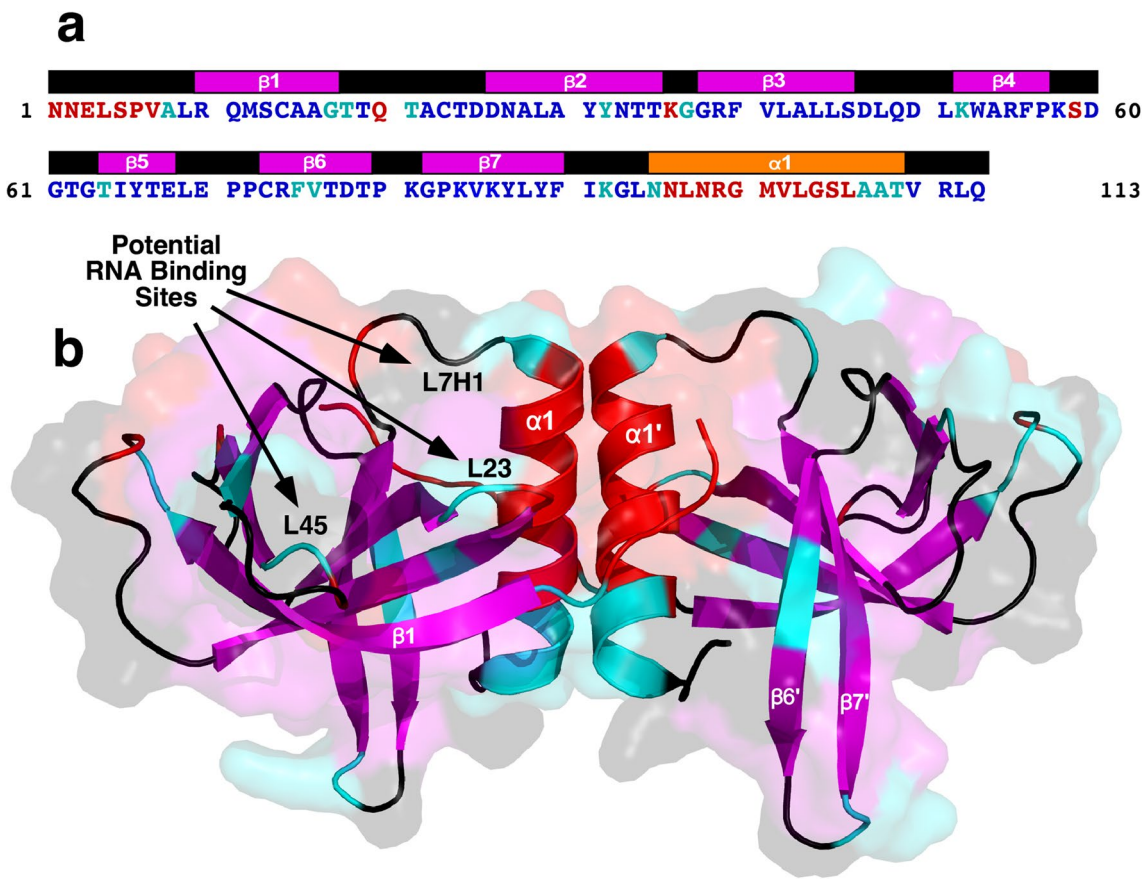

Fig. 2 **a** Primary amino acid sequence for SARS-CoV-2 Nsp9 highlighting, in red, residues for which amide cross peaks were not observed or unambiguously assignable in the ^{1}H-^{15}N HSQC spectrum at 303 K. Residues for which amide cross peak intensity were significantly reduced relative to the bulk of the cross peaks are colored cyan. Proline and residues assigned to amide cross peaks are colored blue. Above the sequence are the elements of secondary structure observed in the crystal of SARS-CoV-2 Nsp9 (PDB ID 6WXD): β-strands = magenta, α-helix = orange. **b** Mapping of the missing or unassigned resonances in the ^{1}H-^{15}N HSQC spectrum of SARS-CoV-2 Nsp9 onto a cartoon representation of the same crystal structure: residues with missing or unassigned amide resonances = red, amides with reduced cross peaks intensities = cyan, β-strands = magenta. Predicted RNA binding sites (Sutton et al. 2004) on the surface of one face of the SARS-CoV-2 Nsp9 dimer are identified with black arrows: between β2 and β3 (L23), β4 and β5 (L45), and β7 and α1 (L7H1)

the protein and their absence may be due to intermediate chemical exchange with solvent or multiple local conformations that broaden these resonances beyond detection. Note that three regions with missing or unassigned resonances and resonances with reduced intensity are in loops between elements of secondary structure (β2-β3, β4-β5, and β7-α1). These regions are aligned along one face of the dimer as highlight in Fig. 2b and are proposed to be sites for RNA binding (Sutton et al. 2004). Some plasticity in these loops, in the form of multiple local secondary structure that results in broad and intensity reduced amide cross peaks, may be required for RNA binding. Otherwise, the largest remaining region with missing or unassigned amide resonances is N96-L106 that comprises almost the complete C-terminal α-helix in the crystal structure of Nsp9. Moreover, the intensity of the resonances for four amide residues surrounding this region were significantly reduced. This α-helix is highly conserved in coronaviruses (Littler et al. 2020), and as shown in Fig. 2b, it is the main component of the homodimer interface driven by van der Waals interactions between anti-parallel GXXXG motifs (Egloff et al. 2004; Littler et al. 2020; Sutton et al. 2004). Missing or unassigned amide resonances have been observed at the interface of other self-assembled proteins including the sensor domain of the *Burkholderia pseudomallei* histidine kinase RisS (Buchko et al. 2015) and the eukaryotic biomineralization protein amelogenin (Buchko et al. 2008).

While a number of experimental techniques illustrated that SARS-CoV Nsp9 was a dimer in solution at protein concentrations above ~0.1 mM, they also showed that at lower concentrations the equilibrium favored the monomeric species (Sutton et al. 2004). Consequently, the intermediate

timescale motion responsible for the missing and intensity-reduced amide resonances in the C-terminal α-helix could be due, at least in part, to intermolecular dynamics related to chemical exchange between monomers and dimers. If intermolecular dynamics at the dimer interface were responsible for the amide resonances missing in this region of Nsp9, it was hypothesized that they would appear if the dimer-monomer equilibrium was shifted completely to either state. Figure 3 compares the ^{1}H-^{15}N HSQC spectra obtained at a concentration of 0.0007 mM, a concentration where the monomeric form of SARS-CoV Nsp9 was reported to dominate (Sutton et al. 2004), and at a concentration of 0.5 mM. The majority of the amide cross peaks overlap at the two concentrations with only subtle chemical shift perturbations. However, the amide resonances for at least four residues, circled blue, change more significantly and the resonances

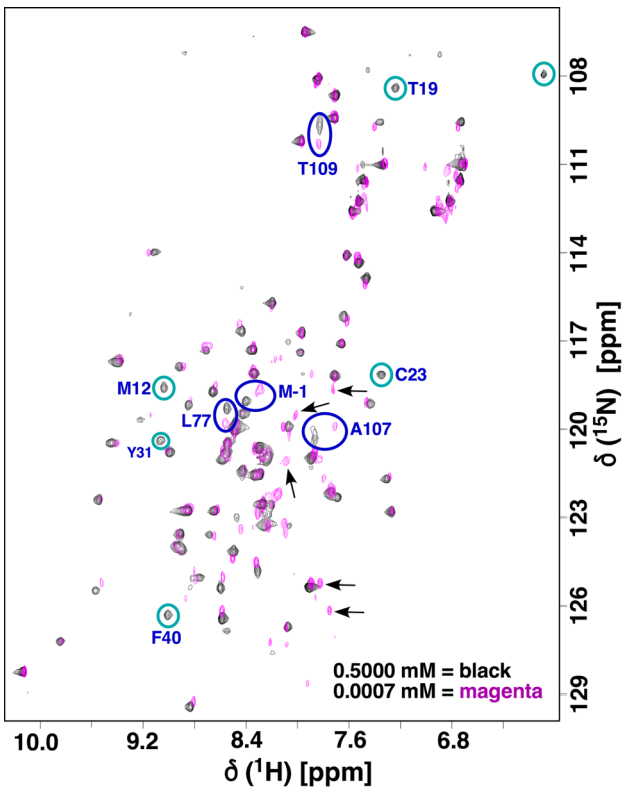

Fig. 3 Overlay of the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled SARS-CoV-2 Nsp9 collected at a concentration of 0.5 (black) and 0.0007 (magenta) mM, a proton resonance frequency of 600 MHz, 303 K, in 100 mM NaCl, 20 mM Tris, 1 mM DTT, pH 7.0. It is not possible to collect 3D NMR data at 0.0007 mM, and hence, the assignment of cross peaks with the largest chemical shift dispersion, circled blue, is tentative. Strong cross peaks at 0.5 mM that have disappeared (or alternatively, moved significantly) at 0.0007 mM are circled cyan. Some amide cross peaks that were weak at 0.5 mM (e.g.: G17, G37, and T18) cannot be detected at 0.0007 mM. The strongest, potentially "new" amide cross peaks are identified by a black arrow. Note that there are other weaker cross peaks near the noise level that may also be "new" amide cross peaks

for five previously "strong" amides, circled cyan, disappear completely at the lower protein concentration. Two of the perturbed amides, tentatively assigned to A107 and T109, are part of the C-terminal α-helix at the dimer interface in the crystal structures, the third, tentatively assigned to L77, is in a β-strand (β6) that is nestled next to the C-terminal helix, and the fourth, tentatively assigned to M-1, is also near the dimer interface in crystal structures. Such chemical shift perturbations are consistent with a progression from a dimeric to monomeric species. As for the five amide resonances that disappear at the lower protein concentration, these are distributed between M12 to F40 and are further from the dimer interface suggesting long range dynamic effects related to the dissociation of the dimers. More significant than the chemical shift perturbations and intensity changes, at 0.0007 mM only five "strong" potentially new amide cross peaks, labelled with black arrows, were observed relative to the spectrum collected at 0.5 mM, and not 11, the number of residues with missing or unassigned amides in the C-terminal α-helix at 0.5 mM (in total, cross peaks for 20 residues are unassigned or missing). Moreover, the α-helix contains two glycine residues, and clearly, there are no new glycine amide cross peaks in the expected region of the ^{1}H-^{15}N HSQC spectrum (upper left-hand corner) collected at 0.0007 mM. Overall there are more amide cross peaks with a range of intensity differences at the lower protein concentration. These observations indicate that intermediate timescale motion responsible for the missing amide resonances at the C-terminal α-helix was still present when the SARS-CoV-2 Nsp9 dimer-monomer equilibrium was shifted to the monomeric state, and therefore, suggests this motion may be intramolecular.

Because the population of a minor species may influence the relaxation properties of the major species (Hansen et al. 2008; Kovermann et al. 2016), it is important to ascertain that the monomer-dimer equilibrium had shifted exclusively to the monomer at 0.0007 mM. Ion mobility mass spectrometry (IM-MS) under native conditions is a highly sensitive method to characterize the oligomerization state of proteins, including those with dynamic structures (Marklund and Benesch 2019). Figure 4 shows the native IM-MS spectra for SARS-CoV-2 Nsp9 collected at a concentration of 0.07, 0.007, and 0.0007 mM. While the monomer is the dominant species at a 0.07 mM concentration (M7 + and top band M6+), a smaller population of dimer species is still present (D11+, D10+, and D9+) along with evidence for some trimers and tetramers. At a 10-fold lower concentration (0.007 mM) the dimer species has not fully disappeared (D11 + and D10+) with a semi-quantitative analysis of the dimer:monomer populations approximating this ratio at 1:10. At a yet 10-fold lower concentration (0.0007 mM) there was no evidence for any dimer species above the noise level in the IM-MS spectrum. Hence, at the limits of dimer

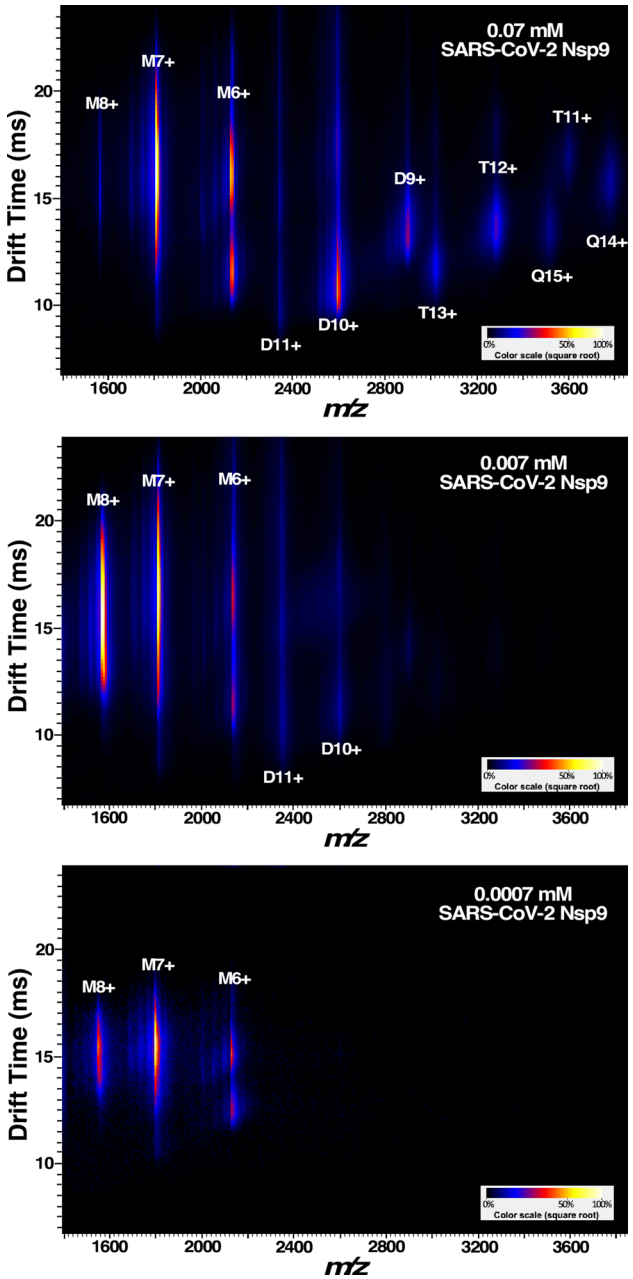

Fig. 4 Ion mobility–mass spectrum of ^{15}N-labeled SARS-CoV-2 Nsp9 at 0.07, 0.007, and 0.0007 mM concentration in 100 mM ammonium acetate, pH ~ 6.7. The horizonal axis is the *m/z*, the vertical axis is the drift time, and the color represents the relative intensity as indicated in the scale bar in the lower right corners. The peaks are annotated for monomers (M), dimers (D), trimers (T), and tetramers (Q)

detection by native mass spectrometry, SARS-CoV-2 Nsp9 was monomeric at a concentration of 0.0007 mM. However, it should be cautioned that even at 0.0007 mM there may still be an undetectable population of minor species influencing the relaxation properties of the major species.

At the protein concentration used to obtain the chemical shift assignments for SARS-CoV-2 Nsp 9 (~0.5 mM), IM-MS did not provide any insights on the monomer-dimer equilibrium because the protein appeared to aggregate in the capillary and electrospray could not be sustained (data not shown). However, at this concentration it is likely that there is exchange between dimers and monomers. Overall rotational correlation times (τ_c) were estimated from the ratio of collective backbone amide ^{15}N T_1 and T_{1rho} NMR measurements and observed to decrease from 10.3 ± 0.7 ns (0.7 mM) to 9.3 ± 0.8 ns (0.07 mM) to 8.7 ± 0.7 ns (0.007 mM). The last value is in the range expected for a 12.7 kDa monomer (Bhattacharjya et al. 2004) and the increase in τ_c with increasing concentration is evidence for dimer formation with increasing protein concentrations as previously reported (Egloff et al. 2004; Sutton et al. 2004), however, an ~25 kDa dimeric species would be expected to have a τ_c value around 15 ns (Rossi et al. 2010). Moreover, a self-association K_d of 0.16 mM estimated for SARS-CoV Nsp9 by analytical ultracentrifugation suggested the dimer association was weak (Sutton et al. 2004). Hence, while the dimer species dominates at protein concentrations greater than 0.1 mM, dissociation into monomers may still be significant if SARS-CoV and SARS-CoV-2 Nsp9 behave similarly.

Interestingly, the IM-MS drift time distributions for both the monomers and dimers of Nsp9 were broad (vertical axis in Fig. 4). Because ion mobility separates proteins based on their shape and charge, the broadness in drift time indicates multiple conformational ensembles. For M6 + and D11 + in particular, several distinct, almost discrete distributions can be noticed. Similar pattern of multiple ensembles can be seen for D10+, although it may partially overlap with M5 + and complicate the analysis. The IM-MS data thus suggests dynamics in protein structure. Note that it is unlikely that the C-terminal α-helix observed in the crystal structure is an intrinsically disordered region of the protein that folds into a helix upon dimer formation. Analysis of the primary amino acid sequence for SARS-CoV-2 Nsp9 for intrinsically disordered regions with the Protein DisOrder predictions System (PrDOS, http://prdos.hgc.jp/cgi-bin/top.cgi) predicts that only the first seven and last four residues of the protein are disordered, an observation consistent with the secondary structure analysis of the available ^{13}Cα and ^{13}Cβ chemical shifts for Nsp9 (Fig. 1a).

Different crystal forms for SARS-CoV and SARS-CoV-2 Nsp9 suggest that the dimer interface of Nsp9 may be "fluid" and biologically important (Sutton et al. 2004). Sutton et al. solved SARS-CoV Nsp9 in two crystal forms, one with two and the other with four molecules in the asymmetric unit (Sutton et al. 2004). While both forms contained a dimer with a C-terminal α-helix interface, the latter crystal form also contained a dimer stabilized by edge-to-edge β-sheet interactions with one of the two monomers disordered. In

another example the dimer interface for a SARS-CoV-2 construct containing a 20-residue N-terminal tag was shifted relative to the interface in the crystal structure of SARS-CoV-2 Nsp9 without the tag, additional evidence that this interface was adaptable to different alignments. Multiple different, heterogenous interfaces for the dimer state, even without dynamic interconversion, would generate multiple different local chemical environments for the C-terminal α-helix that could broaden these backbone amide cross peaks beyond detection.

Peptide binding site

In one of the crystal structures for SARS-CoV-2 Nsp9, the protein construct contained a 20-residue, N-terminal tag (MAHHHHHHSAALEVLFQGPG-) with a Rhinovirus 3C protease cleavage site (PDB ID 6W9Q) (Littler et al. 2020). The N-terminal tag was observed to interact near the dimer interface and, relative to the crystal structure of the protein without the N-terminal tag, affect a repositioning of the monomer orientation in the dimer and the generation of a new β-sheet between a region in the N-terminal tag and L4-A8. It was proposed that this serendipitous observation may have identified a potential peptide-binding site on the surface of the SARS-CoV-2 Nsp9 homodimer. Our construct for SARS-CoV-2 contained a 22-residue N-terminal tag with a TEV protease cleavage site (MGSSHHHHHHSSGENLYFQGHM-; 45% identical and 68% conserved relative to the 3C protease tag). As mentioned, with the TEV containing N-terminal sequence there was no significant perturbation in the ^{1}H-^{15}N HSQC spectrum of Nsp9 with and without the N-terminal tag suggesting there were no strong interactions between the TEV containing N-terminal tag and the exposed region on the Nsp9 dimer interface in solution. Hence, if the region near the Nsp9 dimer interface does have a biological role as a peptide-binding site, our observation with the TEV containing N-terminal tag suggests it is very weak in solution or sequence specific.

Fast (ns to ps) backbone dynamics

The missing amide resonances in the ^{1}H-^{15}N HSQC spectrum of SARS-CoV-2 Nsp9 are likely due to intermediate backbone dynamics on the ms to μs timescale. To probe backbone dynamics on the faster ns–ps timescale and obtain an idea on the "compactness" of the structure, steady-state {^{1}H}-^{15}N heteronuclear NOE ratios were collected in triplicate for Nsp9 at a concentration of 0.5 mM and the mean values plotted in Fig. 5a. Aside from a few residues near the termini and a couple of loops between β-strands, most of ratios are above 0.8, an average suggestive of a relatively rigid overall structure with motion on the fast timescale (ps) restricted. This is in agreement with the compact single domain for the monomer observed in the crystal structure of SARS-CoV-2 Nsp9 composed of one α-helix and seven β-strands arranged into a cone shaped β-barrel. This conclusion based on the heteronuclear NOEs is further corroborated in the backbone ^{15}N T_1 and T_2 relaxation times plotted for SARS-CoV-2 Nsp9 in Fig. 5b and c, respectively. Most of the T_1 values fall between 600 and 800 ms except for a few residues near the termini and most of the T_2 values fall between 50 and 100 ms with the exception of a few residues near termini or in loops between β-strands.

Disulfide bond oligomerization switch

In the crystal structure of human coronavirus Nsp9 (PDB ID 2J97) an intermolecular disulfide bond was observed between the lone cysteine residue in the protein, C68, that was implicated to play a role in the protein's oligomeric state in solution, and in turn, affinity for oligonucleotides (Ponnusamy et al. 2008). This cysteine is conserved in both SARS-CoV and SARS-CoV-2 Nsp9 (Littler et al. 2020), however, mutating this residue to an alanine in SARS-CoV Nsp9 did not alter its oligonucleotide binding or oligomerization properties (Ponnusamy et al. 2008). The chemical shift of the cysteine ^{13}Cβ is sensitive to its oxidation state (Sharma and Rajarathnam 2000) and at 27.2 ppm, the corresponding C73 in SARS-CoV-2 Nsp9 was in the thiol/thiolate state as expected in the presence of 1 mM DTT. A ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled SARS-CoV-2 Nsp9 prepared in the absence of any reducing agent was identical to the spectrum collected on a sample prepared with 1 mM DTT in the final buffer. If an intermolecular C73-C73' disulfide bond had formed under the oxidizing conditions present during the preparation of the latter sample, significant chemical shift perturbations would be expected for the amide cross peaks of C73 and nearby residues in the ^{1}H-^{15}N HSQC spectrum of Nsp9. Hence, at least during recombinant expression in *E. coli*, there does not appear to be any predisposition for intermolecular disulfide bond formation under oxidizing conditions.

Summary

The near-complete backbone assignments for SARS-CoV-2 Nsp9 will enable rapid, in solution, residue-specific tracking of interactions with other proteins and potential drug candidates. This may take the form of chemical shift perturbation studies that will allow mapping of the ligand-binding surface on the homodimer and allow the measurement of dissociation constants (K_d) in the mM to nM range (Buchko et al. 1999; Zuiderweg 2002). A major feature of the assignments was the 11 "missing" and intensity-reduced amide resonances in the ^{1}H-^{15}N HSQC

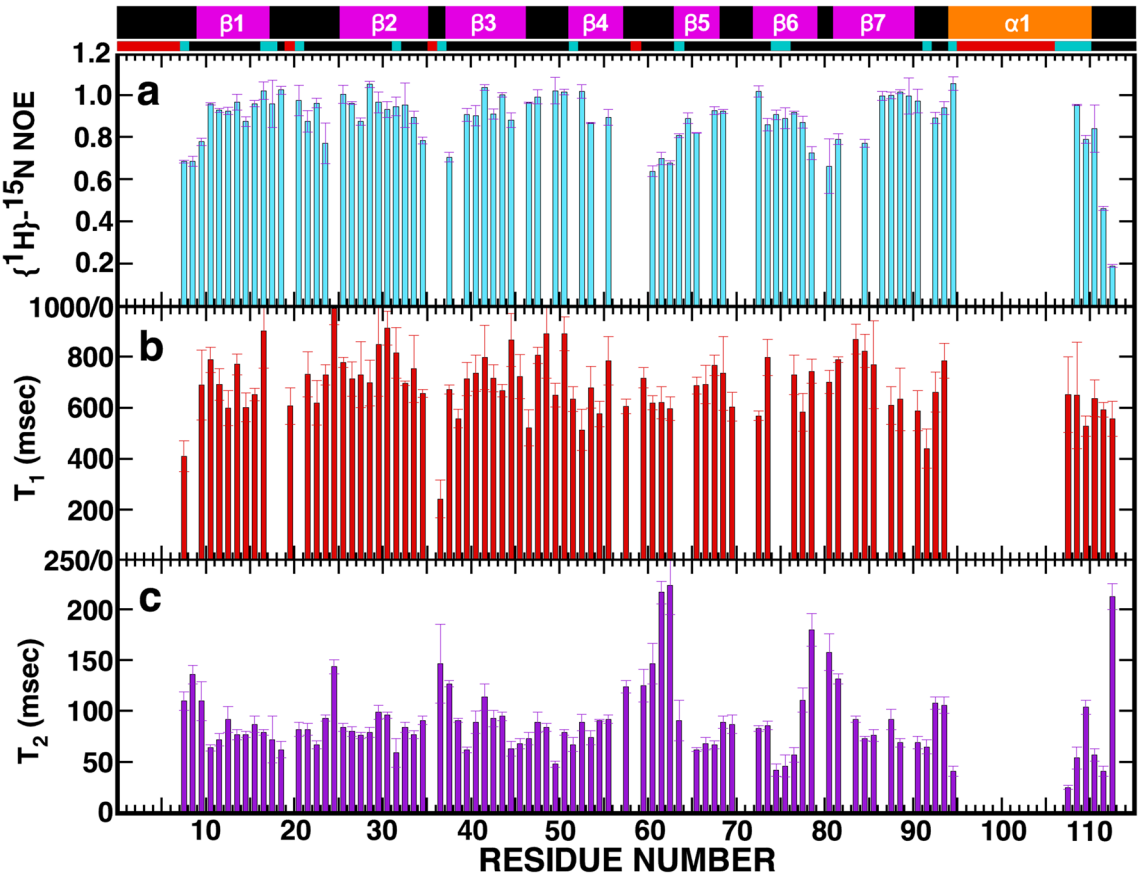

Fig. 5 Plots of the backbone **a** $\{^1H\}$-^{15}N heteronuclear NOEs (cyan) **b** T_1(red), and **c** T_2 (purple) values for SARS-CoV-2 Nsp9 (~0.5 mM) collected at a ^1H resonance frequency of 600 MHz, 303 K. On top of the plots is a schematic representation of the elements of secondary structure observed in the crystal of SARS-CoV-2 Nsp9 (PDB ID 6WXD): β-strands = magenta, α-helix = orange. Residues with missing or unassigned resonances in the ^1H-^{15}N HSQC spectrum of SARS-CoV-2 Nsp9 are identified with red blocks below scheme and residues with weak amide cross peak intensities relative to the majority of amide resonances are identified with cyan blocks

spectrum for almost the complete C-terminal α-helix that forms the dimer interface in the crystal structure of SARS-CoV-2 Nsp9. It has been suggested that the "fluidity" of this C-terminal helix in crystal structures may be indicative of potential heterophilic protein-protein interactions (Sutton et al. 2004). The absence of NMR data for this region of Nsp9, indicative of local dynamics on the ms to μs timescale different from the core of the protein, may be the physical basis for the homophilic and postulated heterophilic protein-protein interactions. In either case, if protein-protein interaction is crucial to the biological function of Nsp9, ligands that alter the dynamics of the C-terminus may also disrupt the biological functions of Nsp9 and this could be identified by the appearance of these "missing" amide resonances in the ^1H-^{15}N HSQC spectrum of Nsp9. Instead of interrupting normal backbone dynamics, ligands that bind to SARS-CoV-2 may also block binding of Nsp9 to its biological partner and this could take the appearance of perturbed resonances in the the ^1H-^{15}N HSQC spectrum of Nsp9. These and similar experiments promise to accelerate our understanding of the biological function of Nsp9 and the discovery of new therapeutic strategies to control COVID19.

Acknowledgements We thank the SSGCID cloning and protein production groups at the University of Washington and Seattle Children's Research Institute and Lynn Barrett for assistance with shipping and administration.

Funding This research was funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under Federal Contract number HHSN272201700059C and the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE National Laboratories focused on response to COVID19, with the later funding provided by the Coronovirus CARES Act. Part of the research was conducted at the W.R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by U.S. Department of Energy's Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory (PNNL). Battelle operates PNNL for the U.S. Department of Energy under contract DE-AC05-76RL01830.

Data availability The ^{1}H, ^{13}C, and ^{15}N chemicals shifts for SARS-CoV-2 Nsp9 have been deposited at the BioMagResBank (www.bmrb.wisc.edu) under accession number BMRB-50513.

References

Aden J et al (2011) Extraordinary ms-μs backbone dynamics in *Arabidopsis thaliana* peroxiredoxin. Q Biochim Biophys Acta 1814:1880–1890. https://doi.org/10.1016/j.bbapap.2011.07.011

Arachchige RJ et al (2018) Solid-state NMR identification of intermolecular interactions in amelogenin bound to hydroxyapatite. Biophys J 115:1666–1672. https://doi.org/10.1016/j.bpj.2018.08.027

Bhattacharjya S, Ping X, Gingras R, Shaykhutdinov R, Wu C, Whiteway M, Ni F (2004) Solution structure of the dimeric SAM domain of MAPKKK Ste11 and its interactions with the adaptor protein STE50 from the budding yeast: implications for Ste11 activitation and signal transmission through the Ste50-Ste11 complex. J Mol Biol 344:1071–1087

Buchko GW et al (1999) Interactions of human nucleotide excision repair protein XPA with DNA and RPA70ΔC327: chemical shift mapping and ^{15}N NMR relaxation studies. Biochemistry 38:15116–15128

Buchko GW, Tarasevich BJ, Bekhazi J, Snead ML, Shaw WJ (2008) A solution NMR investigation into the early events of amelogenin nanosphere self-assembly initiated with sodium chloride or calcium chloride. Biochemistry 47:13215–13222. https://doi.org/10.1021/bi8018288

Buchko GW, Edwards TE, Hewitt SN, Phan IQ, Van Voorhis WC, Miller SI, Myler PJ (2015) Backbone chemical shift assignments for the sensor domain of the *Burkholderia pseudomallei* histidine kinase RisS: "missing" resonances at the dimer interface. Biomol NMR Assign 9:381–385. doi:https://doi.org/10.1007/s12104-015-9614-2

Campanacci V et al (2003) Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein. Acta Crystallogr D Biol Crystallogr 59:1628–1631. https://doi.org/10.1107/s0907444903016779

Chen JP et al (2017) Structural analysis of porcine reproductive and respiratory syndrome virus non-structural protein 7α (NSP7α) and identification of its interaction with NSP9. Front Microbiol 8:853. https://doi.org/10.3389/fmicb.2017.00853

Egloff MP et al (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 USA 101:3792–3796. doi:https://doi.org/10.1073/pnas.0307877101

Frieman M 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. doi:https://doi.org/10.1128/JVI.05957-11

Gordon CJ, Tchesnokov EP, Feng JY, Porter DP, Gotte M (2020a) The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. J Biol Chem 295:4773–4779. doi:https://doi.org/10.1074/jbc.AC120.013056

Gordon DE et al (2020b) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583:459–468. https://doi.org/10.1038/s41586-020-2286-9

Hafsa NE, Arndt D, Wishart DS (2015) CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. Nucleic Acids Res 43:W370–W377. doi:https://doi.org/10.1093/nar/gkv494

Hansen DF, Vallurupalli P, Lundstrom P, Neudecker P, Kay LE (2008) Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? J Am Chem Soc 130:2667–2675. doi:https://doi.org/10.1021/ja078337p

Jin Z et al (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature 582:289–293. doi:https://doi.org/10.1038/s41586-020-2223-y

Kovermann M, Rogne P, Wolf-Watz M (2016) Protein dynamics and function from solution state NMR spectroscopy. Q Rev Biophys 49: e6.

Littler DR, Gully BS, Colson RN, Rossjohn J (2020) Crystal structure of the SARS-CoV- 2 non-structural protein 9, nsp9. iScience 23:101258. https://doi.org/10.1016/j.isci.2020.101258

Marklund EG, Benesch JL (2019) Weighing-up protein dynamics: the combination of native mass spectrometry and molecular dynamics simulations. Curr Opin Struct Biol 54:50–58. doi:https://doi.org/10.1016/j.sbi.2018.12.011

Ponnusamy R, Moll R, Weimar T, Mesters JR, Hilgenfeld R (2008) Variable oligomerization modes in coronavirus non-structural protein 9. J Mol Biol 383:1081–1096. doi:https://doi.org/10.1016/j.jmb.2008.07.071

Rossi P et al (2010) A microscale protein NMR sample screening pipeline. J Biomol NMR 46:11–22. doi:https://doi.org/10.1007/s10858-009-9386-z

Shaheen S et al (2020) Solution structure for an *Encephalitozoon cuniculi* adrenodoxin-like protein in the oxidized state. Protein Sci 29:809–817. https://doi.org/10.1002/pro.3818

Sharma D, Rajarathnam K (2000) ^{13}C NMR chemical shifts can predict disulfide bond formation. J Biomol NMR 18:165–171. https://doi.org/10.1023/A:1008398416292

Sutton G et al (2004) The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. Structure 12:341–353. doi:https://doi.org/10.1016/j.str.2004.01.016

Szyperski T, Yeh D, Sukumaran D, Moseley H, Montelione G (2002) Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment. Proc Natl Acad Sci USA 99:8009–8014

Tan YJ, Lim SG, Hong WJ (2005) Characterization of viral proteins encoded by the SARS-coronavirus genome. Antivir Res 65:69–78. doi:https://doi.org/10.1016/j.antiviral.2004.10.001

Wu F et al (2020) A new coronavirus associated with human respiratory disease in China. Nature 579:265–269. https://doi.org/10.1038/s41586-020-2008-3

Zuiderweg ERP (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. Biochemistry 41:1–7