

Multivalent binding of the partially disordered SARS-CoV-2 nucleocapsid phosphoprotein dimer to RNA

Heather M. Forsythe,¹ Joaquin Rodriguez Galvan,¹ Zhen Yu,¹ Seth Pinckney,¹ Patrick Reardon,² Richard B. Cooley,¹ Phillip Zhu,¹ Amber D. Rolland,³ James S. Prell,³ and Elisar Barbar^{1,*}

¹Department of Biochemistry and Biophysics and ²NMR Facility, Oregon State University, Corvallis, Oregon; and ³Department of Chemistry and Biochemistry, University of Oregon, Eugene, Oregon

ABSTRACT The nucleocapsid phosphoprotein N plays critical roles in multiple processes of the severe acute respiratory syndrome coronavirus 2 infection cycle: it protects and packages viral RNA in N assembly, interacts with the inner domain of spike protein, binds to structural membrane (M) protein during virion packaging and maturation, and to proteases causing replication of infective virus particle. Even with its importance, very limited biophysical studies are available on the N protein because of its high level of disorder, high propensity for aggregation, and high susceptibility for autoprolysis. Here, we successfully prepare the N protein and a 1000-nucleotide fragment of viral RNA in large quantities and purity suitable for biophysical studies. A combination of biophysical and biochemical techniques demonstrates that the N protein is partially disordered and consists of an independently folded RNA-binding domain and a dimerization domain, flanked by disordered linkers. The protein assembles as a tight dimer with a dimerization constant of sub-micromolar but can also form transient interactions with other N proteins, facilitating larger oligomers. NMR studies on the ~100-kDa dimeric protein identify a specific domain that binds 1–1000-nt RNA and show that the N-RNA complex remains highly disordered. Analytical ultracentrifugation, isothermal titration calorimetry, multiangle light scattering, and cross-linking experiments identify a heterogeneous mixture of complexes with a core corresponding to at least 70 dimers of N bound to 1–1000 RNA. In contrast, very weak binding is detected with a smaller construct corresponding to the RNA-binding domain using similar experiments. A model that explains the importance of the bivalent structure of N to its binding on multivalent sites of the viral RNA is presented.

SIGNIFICANCE There is an urgent need to accelerate research into the mechanisms of severe acute respiratory syndrome coronavirus 2 infection, transmission, and control. Here, we focus on the essential nucleocapsid phosphoprotein N, which plays critical roles in multiple processes of the infection cycle including packaging viral RNA in nucleocapsid assembly. We show the importance of dimerization of N to its binding efficiency to RNA, a process that is the foundation of currently lacking medical treatment strategies.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for one of the most significant global public health crises of the past century. The rapid global spread of SARS-CoV-2 has resulted in more than 88,000,000 cases and 1,900,000 deaths worldwide as of

January 2021, only 1 year since it was first identified (1), with those recovered experiencing serious long-term effects on their physical health (2,3). In the wake of such human, social, and economic devastation, there is an urgent need to continue growing and evolving our understanding of structures and functions of SARS-CoV-2 proteins, their role in the viral life cycle, and their potential for vaccine and drug development.

The SARS-CoV-2 virion is composed of four structural proteins, spike (S), membrane (M), envelope (E), and nucleocapsid (N). N is the only protein in the nucleocapsid and is the only viral protein that associates with the replicase

Submitted December 15, 2020, and accepted for publication March 25, 2021.

*Correspondence: barbar@oregonstate.edu

Editor: Mark C. Williams.

<https://doi.org/10.1016/j.bpj.2021.03.023>

© 2021 Biophysical Society.



transcriptase complexes (RTCs) (4). N proteins bridge a connection between E and M proteins and the ~30 kb SARS-CoV-2 genomic RNA (gRNA). Because of these unique interactions, N plays critical roles in nucleocapsid assembly, virion assembly, mature virion packaging, and replicating and transcribing the single-stranded RNA (5). Consequently, the N protein is being investigated as a potential drug target (6–8), and its low rate of mutation has made it a candidate for vaccine development (9).

The 422-amino-acid N protein is organized into two independently folded domains, the N-terminal domain (NTD) and the C-terminal domain (CTD), flanked by two disordered tails (N_{IDR} and C_{IDR}), and separated by a 65-amino-acid flexible linker (Fig. 1 A). The CTD forms a strong dimer, facilitating dimerization of N (Fig. 1 B; (10)). Experiments performed on SARS-CoV-2 N and the highly conserved SARS-CoV N show that although NTD is the primary RNA-binding domain (11), both the NTD and CTD are reported to make interactions with gRNA (5,10,12–15). The conserved disordered regions of N contain a Ser-Arg-rich region, the hnRNP1 binding site, the nonstructural protein 3 (nsp3) binding site, a K-rich region, and the putative TRIM25 E3 ligase binding site (14,16,17). These disordered regions are also likely to influence gRNA binding, oligomerization, and interactions not yet identified.

Stable dimerization of N and its organization into higher order oligomers in the nucleocapsid is thought to play an essential role in virus particle assembly (18). SARS-CoV-N forms oligomers in vitro, with a dimer being the dominant species in the absence of gRNA, as shown by native gel electrophoresis, size-exclusion chromatography (SEC), and surface plasmon resonance (19). It is estimated that there are 730–2200 copies of N per SARS-CoV-2 virion (20), which must coordinate their binding along the 30-kb genome while simultaneously forming stable interactions with other N proteins. The monomeric NTD structure was solved by both NMR and x-ray crystallography (7,11,13,14). The crystal structure for the CTD was solved for a dimer (Fig. 1 B; (10,21)). The CTD is thought to facilitate dimerization of N but can also pack as an octamer (22)

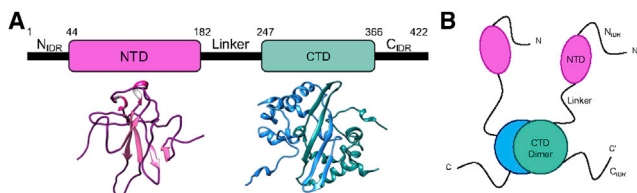


FIGURE 1 The SARS-CoV-2 nucleocapsid protein N domain organization. (A) Structures of folded domains shown under the domain diagram and predicted disorder shown as lines: NMR structures of SARS-CoV-2 NTD (*magenta*), Protein Data Bank, PDB: 6YI3, and the highly conserved SARS-CoV CTD represented as a dimer (*blue-teal*), PDB: 2JW8. (B) Model of N protein dimer based on sequence prediction and structures of NTD and CTD. To see this figure in color, go online.

or larger (23) stacking in a helical formation. What is missing is a clear picture of the structure and association state of the full-length SARS-CoV-2 N protein (FL-N), and not just the domains, along with characterization of its binding to gRNA because this information is foundational for any drug development underway to target these essential viral interactions.

Biological condensates are a nearly universal feature of cells to protect RNA during stress. There is increasing evidence for the presence of these condensates in virology, such as in measles and human immunodeficiency virus, with RNA as a common constituent of these condensates (24–27). FL-N phase separates with RNA, and the intrinsically disordered N_{IDR} and linker as well as the CTD are essential for this process (15,28–32). Different segments of the genome contribute to this phase separation, including the 1000-nucleotide (nt) segment at the 5'-end (1–1000 RNA), which under physiological conditions, condenses with N and is enhanced at human upper airway temperatures, suggesting an important role during infection (15). 1–1000 RNA contains several structural elements that are thought to constitute the binding site to N (15).

Given the importance of N in the virus assembly and replication, this work focuses on the FL-N and comparison of its structural features with the individual smaller NTD and CTD constructs. We report, to our knowledge, the first thorough characterization of FL-N and its binding to a 1000-nt segment of the viral RNA (1–1000 RNA) and propose a model that explains the importance of FL-N bivalent structure to its binding affinity on multivalent sites of a long segment of the viral RNA.

MATERIALS AND METHODS

Plasmid construction and protein expression

FL-N phosphoprotein and CTD corresponding to residues 231–361 were amplified by PCR using plasmid pLVX-EF1 α -SARS-CoV-2-N-2xStrep-IRES as the template (141391, a gift from Dr. Krogan; Addgene, Watertown, MA) (8), then cloned into pet24d(+) vector in frame with an N-terminal hexahistidine tag and TEV (“Tobacco Etch Virus”) protease cleavage site. The plasmid was expressed in *Escherichia coli* Rosetta cells and cultured in Terrific Broth media to an OD of 2.0 and induced by 0.5 mM isopropylthio- β -galactoside at 18°C overnight. For NMR, at an optical density of 0.7, cells were washed twice with MJ9 medium and cultured again at 25% volume of ^{15}N -NH $_4$ Cl MJ9 medium for 1 more h and induced with 0.5 mM isopropylthio- β -galactoside at 30°C for 4 h. The NTD corresponding to residues 47–172 was expressed in frame with an N-terminal His $_6$ -bdSUMO fusion (33). NTD was grown and expressed in C43 cells in rich autoinducing media or ^{15}N autoinducing minimal media for NMR and grown for 24 h at 37°C.

In vitro transcription (IVT) and RNA purification

Plasmids as templates to generate RNA were a generous gift from Dr. Gladfelter (15). The 5'-end RNA fragment corresponds to the first 1000 nt of viral RNA, shown to bind and cause phase separation with N (15). The in vitro transcription and RNA preparation followed protocols established

in Iserman et al. (15). Briefly, plasmids were linearized with XbaI restriction enzyme and RNA was synthesized by *in vitro* transcription (E2040S; New England Biolabs, Ipswich, MA). RNA was further purified by 2.5 M LiCl precipitation after DNase treatment, and the size was verified by agarose gel electrophoresis. The same procedure was carried out with Cy3- (PA53026; Sigma-Aldrich, St. Louis, MO) labeled RNA.

Protein purification

FL-N was purified under denaturing conditions using TALON His-tag purification protocol (Clontech Laboratories, Mountain View, CA). Eluted protein fractions in 6 M urea buffers were dialyzed in 50 mM HEPES (pH 7.5), 500 mM NaCl, and 10% glycerol to refold overnight at 4°C (34). The refolded protein was further purified by TALON His-tag purification protocol (Clontech Laboratories) with all washing steps in 1 M NaCl to ensure removal of bound RNA contaminant. FL-N was further purified and separated from soluble aggregate components by gel filtration on a Superdex 75 gel filtration column (GE Healthcare, Chicago, IL). NTD and CTD were purified under native conditions using TALON His-tag purification protocol (Clontech Laboratories). Purified NTD was eluted by on-column proteolysis with 30 nM untagged fast acting protease bdSEN1 for 1 h at 4°C. Both NTD and CTD were further purified on a Superdex 75 gel filtration column (GE Healthcare) in 50 mM sodium phosphate and 150 mM NaCl (pH 6.5). The purity of the recombinant proteins assessed by sodium-dodecyl-sulfate-polyacrylamide-gels was >95%. Purified proteins were stored at 4°C with a cocktail of pepstatin and phenylmethylsulfonyl fluoride protease inhibitors and used within 1 week.

SEC multiangle light scattering

SEC coupled to multiangle light scattering (SEC-MALS) was carried out using a Superdex 200 gel filtration column on an AKTA-FPLC (GE Healthcare) coupled with a DAWN multiple-angle light scattering and an Optilab refractive index system (Wyatt Technology, Santa Barbara, CA). Data for NTD and CTD were collected by injecting 70 and 100 μ M protein, respectively, in the FL-N NMR buffer (50 mM sodium phosphate and 150 mM NaCl (pH 6.5)) as well as the NTD NMR buffer (20 mM sodium phosphate and 20 mM NaCl (pH 6.5)). FL-N SEC-MALS data were collected with 20 μ M protein in 50 mM sodium phosphate and 150 mM NaCl (pH 7.5). Molar mass and error analysis were determined with ASTRA software package v9, employing a Zimm light scattering model (Wyatt Technology). RNA-binding experiments were carried out with addition 1–1000 RNA to protein to a final ratio of 0.003:1 RNA/protein.

Analytical ultracentrifugation

Protein samples were either dialyzed overnight or buffer exchanged into 50 mM sodium phosphate and 150 mM NaCl (pH 6.5). Concentrations were 22, 38, 59, and 15 μ M for FL-N, NTD, CTD, and FL-N + 1–1000 RNA, respectively. RNA samples in the same buffer had a final concentration of 90 ng/ μ L. All analytical ultracentrifugation (AUC) experiments were performed using a Beckman Coulter Optima XL-A ultracentrifuge equipped with absorbance optics (Brea, CA). Protein-partial-specific volumes as well as buffer densities and viscosities were estimated using the software *sednterp* (35). For sedimentation velocity AUC (SV-AUC) experiments, samples were loaded into epon-2-channel sector cells with a 12-mm optical pathlength. Protein and RNA absorbance was monitored at 280 nm and at 255 nm, respectively. SV-AUC experiments were conducted at 42,000 rpm in a four-cell Beckman Coulter AN 60-Ti rotor and 20 °C. Scans were performed continuously for a total of 300 scans per cell. Data were fit to a continuous *c*(*S*) distribution using the software *sedfit* (36). Sedimentation coefficients were expressed in Svedbergs (*S*).

Sedimentation equilibrium experiments were carried out at 5°C to reduce degradation. Three concentrations of each protein were measured, 50, 30, and 15 μ M for CTD and 23, 15, and 8 μ M for FL-N. Experiments were performed at three speeds, 12,000, 15,000, and 20,000 for CTD and 10,000 for FL-N. Scans were taken every 3 h to check that equilibrium had been achieved, with final scans taken a minimum of 30 h after reaching speed. Data were fit to a homodimerization model using the Heteroanalysis software (37).

Circular dichroism

Spectra were recorded on a JASCO 720 spectropolarimeter (Jasco, Oklahoma City, OK) using a 0.1-cm cell for NTD and FL-N and a 0.05-cm cell for CTD. Protein samples were dialyzed overnight against 20 mM sodium phosphate, 100 mM NaF, and 1 mM sodium azide (pH 7.5) before data collection. Spectra were collected at 25°C at a concentration of 14.8, 17, and 30 μ M for NTD, FL-N, and CTD, respectively. Spectra were acquired from 270 nm (data not shown) to 200 nm to rule out potential RNA signals around 260 nm. Buffer spectra were subtracted from sample spectra before the analysis.

Isothermal titration calorimetry

Samples were codialyzed overnight into 20 mM sodium phosphate, 100 mM NaF, and 1 mM sodium azide (pH 7.5) at the same conditions as for circular dichroism (CD) measurements. 25 μ L of 40,000 U/ μ L of murine RNase inhibitor (New England Biolabs) were added to the RNA dialysis membrane to a total volume of 1 mL. For the experiment, 0.3 μ M of 1–1000 RNA in the syringe was titrated into 15 μ M FL-N in the cell at 25°C. Binding thermogram was obtained with a VP-ITC microcalorimeter (Microcal, Westborough, MA). Data were fitted to a simple single-site binding model.

NMR

NMR experiments were collected on a Bruker 800 MHz Avance III HD spectrometer (Bruker Biosciences, Billerica, MA) equipped with a triple resonance cryogenic probe. NMR experiments for ¹⁵N FL-N were carried out at 25°C in 50 mM sodium phosphate and 150 mM NaCl (pH 6.5) with 10% D₂O and a protease inhibitor mixture (Roche Applied Science, Madison, WI). CLEANEX-TROSY experiments for solvent accessibility measurements were collected with a mixing time of 100 ms using a recycle delay of 1.5 s. Spectra of ¹⁵N NTD were collected in 20 mM sodium phosphate and 20 mM NaCl (pH 6.5) to transfer assignments from the Biological Magnetic Resonance Data Bank entry BMRB: 34511. Spectra of ¹⁵N CTD were collected at 25°C in 10 mM sodium phosphate, 50 mM NaCl, 1 mM NaN₃, and 1 mM EDTA (pH 6.0) to accurately compare and transfer assignments from SARS-CoV CTD (BMRB: 17217). All spectra were referenced to DSS. All two-dimensional (2D) spectra were processed using Topspin (Bruker Biosciences) and analyzed in CcpNmr. RNA-binding experiments were carried out with the same FL-N or CTD sample as TROSY and CLEANEX. Purified 1–1000 RNA was added to ¹⁵N FL-N or CTD at a molar ratio of 0.008:1 RNA/protein, followed by concentration to the sample's original volume.

Turbidity measurements

Purified protein solutions were dialyzed overnight into 30 mM NaCl and 50 mM sodium phosphate (pH 6.5 and 7.5), and samples were serially diluted to reach desired concentrations. Turbidity measurements were carried out in two different ways. First, 80 μ L of samples were transferred into a quartz cuvette, and turbidity was measured at 340 nm using a Cary 60 UV-Vis (Agilent Technologies, Santa Clara, CA) (Fig. 6A). Additional turbidity

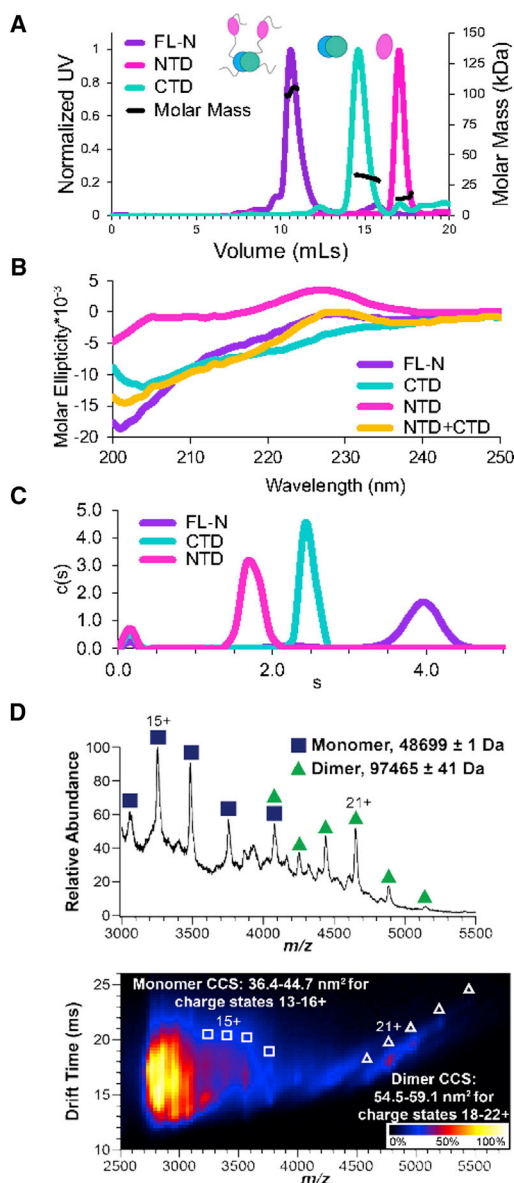


FIGURE 2 Biophysical characterization of SARS-CoV-2 N protein. (A) SEC-MALS of FL-N (purple), NTD (magenta), and CTD (teal). (B) CD spectra of FL-N (purple), NTD (magenta), CTD (teal), and the sum of NTD + CTD (yellow). (C) AUC sedimentation velocity of FL-N (purple), NTD (magenta), and CTD (teal). (D) Native IM-MS of FL-N. Peaks corresponding to N-protein monomer and dimer are marked with squares and triangles, respectively, with the most abundant charge state labeled for each species. (Top) Mass spectrum of N-protein monomers (navy squares) and dimers (green triangles) acquired with sample cone at 25 V and trap at 100 V. (Bottom) IM-MS spectrum of the same sample under gentler, more native-like acquisition conditions (sample cone at 25 V and trap at 25 V) used to measure CCS. To see this figure in color, go online.

measurements were carried out in Greiner Bio-one UV-STAR UV-VIS 96-well plates (Kremsmünster, Austria). 20 μ L of samples were plated in replicates of three. Plates were then shaken for 30 s, and turbidity was subsequently measured at 340 nm using a BioTek Synergy HT plate reader. Data points were collected and processed using Gen5 software, version 2.09 (Fig. 6, B and C). Kinetic turbidity measurements were performed by the same plate reader using the kinetic feature on Gen5 with 10 inter-

spaced measurements every 2.5 min. RNA was added to a final concentration of 200 and 20 μ g/L for 1–1000 RNA and mouse RNA, respectively.

Electrophoretic mobility shift assay

RNA was visualized by electrophoretic mobility shift assay in 1% agarose gel. RNA at 0.3 or 0.5 μ M was added to increasing concentrations of protein in the range of 0–40 μ M and incubated for 20 min at room temperature in a reaction volume of 10 μ L. 1 μ L of 30% glycerol was added to end the reaction before loading on the gel. RNA bands were stained by Midori Green Nucleic Acid staining solution (Bulldog Bio, Portsmouth, NH) and visualized by Bio-Rad Gel Doc Image system (Bio-Rad Laboratories, Hercules, CA).

Microscopy

Light microscopy imaging was done using a Zeiss Primo Vert microscope (ZEISS, Oberkochen, Germany) on samples prepared as described for the turbidity measurements. Images were taken using a camera attached to the light microscope. Fluorescence microscopy images were taken on a Keyence BZ-X700/BZ-X710 microscope and a 96-well plate; images were processed using BZ-x viewer and BZ-x analyzer software. For this experiment, Cy3-labeled RNA at 90 μ g/L was dissolved in 30 mM NaCl and 50 mM sodium phosphate (pH 6.5) to a final volume of 30 μ L. After imaging of RNA alone, protein was added to a 15 μ M final concentration, and image capture began immediately thereafter; foci became visible after 5 min.

Cross-linking experiments

Cross-linking reactions were performed by addition of a 2.3% glutaraldehyde solution in a 1:20 volume ratio. Reactions were conducted in 50 mM sodium phosphate (pH 7.5) at 37°C for 5 min and stopped by the addition of 1.5 M Tris-HCl in a 1:10 volume ratio. Concentrations used for cross-linking assays were 30 μ M for NTD and dilute FL-N, 60 μ M for NTD and concentrated FL-N, and 35 μ M for CTD (Fig. 5, C, D, and F). 1 or 2 μ M (2 $\times$) 1–1000 RNA was added to NTD or FL-N to a final ratio of 0.03:1 for dilute protein + RNA, or 0.015:1 for concentrated protein + RNA, and 0.03:1 for concentrated protein + 2 $\times$ RNA. Urea controls were incubated in either 2-, 4-, or 6-M urea solutions before cross-linking.

Native mass spectrometry

A 50- μ L aliquot of 20 μ M FL-N protein was buffer exchanged into 200 mM ammonium acetate (pH 7.50) using a Micro Bio-Spin column (Bio-Rad Laboratories). All ion-mobility-mass-spectra were collected using a Waters Synapt G2-Si time-of-flight mass spectrometer (SYNAPT, Milford, MA) with a nanoelectrospray ionization source. A small volume ($\sim$ 3–5 μ L) of buffer-swapped sample was loaded into borosilicate capillaries pulled to a $\sim$ 1 μ m tip with a Flaming-Brown P-97 micropipette puller (Sutter Instrument, Novato, CA). A platinum wire was placed in electrical contact with the solution, and a voltage of +0.7 kV was applied to initiate electrospray. Data were collected with the source at ambient temperature, sampling cone at 25 V, trap collision energy at 100 V (accurate mass determination) or 25 V (ion mobility), transfer collision energy at 5 V, and trap gas flow at 5 mL/min. A mass calibration profile was generated using cesium iodide cluster ions before data acquisition for determination of accurate mass. All ion-mobility-mass-spectra were acquired with the following conditions: trap wave velocity 300 m/s and height 2.0 V, ion-mobility mass spectrometry wave velocity 500 m/s and height 18 V, and transfer wave velocity 100 m/s and height 2.0 V. Ion-mobility arrival time data were calibrated

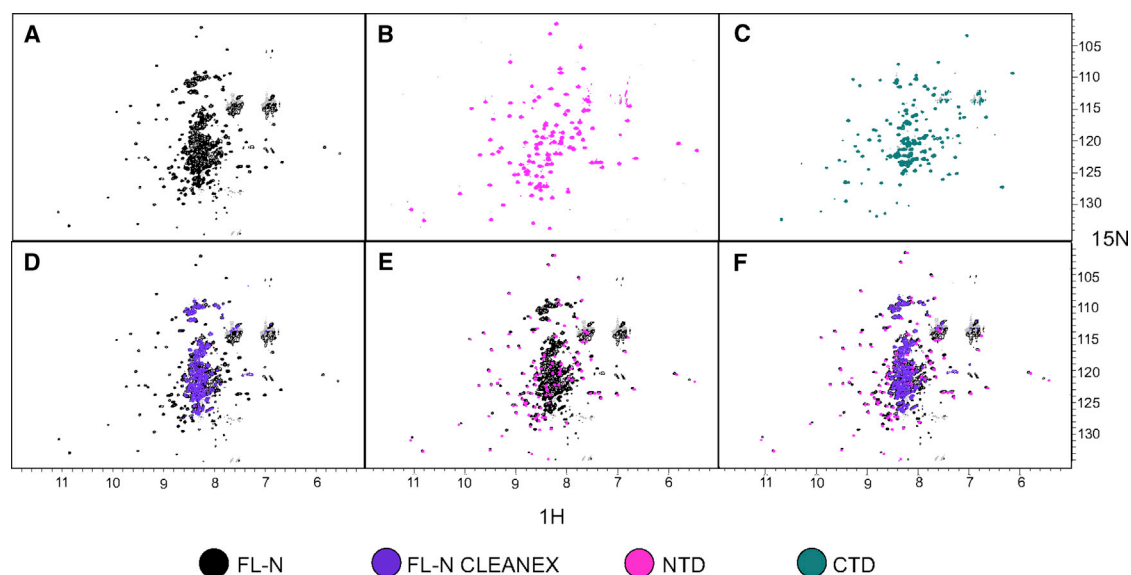


FIGURE 3 NMR spectra of SARS-CoV-2 N protein. (A) 2D- ^1H , ^{15}N -TROSY spectra of 250 μM ^{15}N FL-N at 25°C (black) (B) 2D- ^1H , ^{15}N -TROSY spectra of 200 μM ^{15}N NTD at 25°C (magenta). (C) 2D- ^1H , ^{15}N -TROSY spectra of 250 μM ^{15}N CTD at 25°C (teal). (D) Overlay of 2D- ^1H , ^{15}N -TROSY of FL-N (black) and 2D- ^1H , ^{15}N -CLEANEX spectra of FL-N (purple). (E) Overlay of 2D- ^1H , ^{15}N -TROSY of FL-N (black) and NTD (magenta) spectra. (F) Overlay of 2D- ^1H , ^{15}N -TROSY of FL-N (black) and 2D- ^1H , ^{15}N -CLEANEX of FL-N (purple) with 2D- ^1H , ^{15}N -TROSY of NTD (magenta). To see this figure in color, go online.

using β -lactoglobulin, transthyretin, avidin, bovine serum albumin, concanavalin A, and alcohol dehydrogenase as in (38–40).

RESULTS

FL-N is a dimer consisting of ordered domains flanked with disorder

Size exclusion chromatography and multi-angle light scattering (SEC-MALS) of FL-N, NTD, and CTD show that FL-N is predominantly a dimer in solution, with molar mass of $98,250 \text{ Da} \pm 3\%$ very close to the calculated molar mass of $97,584 \text{ Da}$ for a dimer. The CTD is also a stable dimer in solution with a molar mass of $32,130 \text{ Da} \pm 1.6\%$ ($35,239 \text{ Da}$, calculated from sequence), whereas NTD is a monomer with a molar mass of $14,380 \text{ Da} \pm 3.5\%$ ($14,084 \text{ Da}$, calculated from sequence) (Fig. 2 A).

Secondary structure, assessed by CD, confirms the model of FL-N (Figs. 1 and 2 B). The NTD CD spectrum shows an unusual positive ellipticity at 228 nm, consistent with a structure of an irregular β -trefoil fold (Fig. 1 A). This maximum is also present in the FL-N spectrum, indicating that the structure of the NTD is similar to that in the full-length protein. In FL-N, there is also a large negative ellipticity at $\sim 200 \text{ nm}$ indicative of long segments of disorder, consistent with prediction and with previously published CD spectra (10). The CTD spectrum shows a mixture of β -sheets and -helices. The sum of the CTD and NTD molar ellipticity spectra is equivalent to the FL-N spectrum

without the disordered regions, indicating that the structures of the individual domains are similar to those in the intact protein.

SV-AUC data show a single sharp peak for each of the NTD and CTD constructs with a sedimentation coefficient consistent with their respective monomer and dimer folded states, whereas the FL-N shows a peak of considerably larger size but somewhat broad, indicative of a more disordered structure sampling multiple conformations (Fig. 2 C).

Native electrospray-ionization-ion-mobility-mass-spectrometry (IM-MS) clearly indicates the presence of both a monomer and dimer native ion population under these solution conditions (Fig. 2 D). The measured monomer mass of $48,699 \pm 1 \text{ Da}$ compared well with the expected mass based on the sequence of the His-tagged N-protein construct ($48,792 \text{ Da}$). The measured dimer mass of $97,465 \pm 41 \text{ Da}$ compared well with the expected $97,398 \text{ Da}$ from the measured monomer mass. Both monomer and dimer populations persisted upon dilution of the sample to $\sim 1 \mu\text{M}$ protein concentration (data not shown). Ion mobility spectrometry measurements, which provide ion size and shape information in the form of ion collision cross sections (CCSs), indicated structures of the monomer and dimer consistent with folded, globular proteins (38,39,41). This conclusion is inferred from the relatively low and narrow charge state distribution observed for both the monomer and dimer ion population (40). Monomer CCSs ranged from 36.4 to 44.7 nm^2 over the observed charge states (13 – $16+$), whereas dimer CCSs ranged from 54.5 to 59.1 nm^2 for the charge states 18 – $22+$. Both sets of CCS

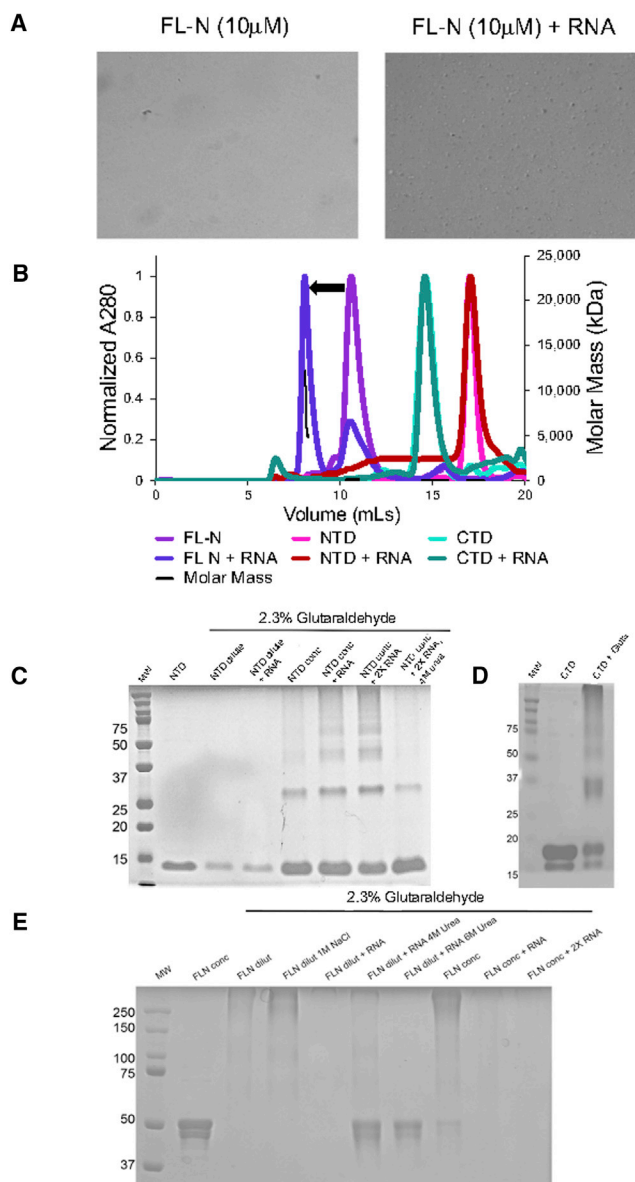


FIGURE 4 Binding of the N protein to viral RNA. (A) Light microscopy image of FL-N with and without 1–1000 RNA at 400 $\times$ augmentation. (B) SEC-MALS of FL-N (purple and blue), NTD (pink and red), and CTD (cyan and teal) with and without 1–1000 RNA. Arrow indicates peak shift in FL-N upon addition of RNA. (C) Cross-linking of 30 μ M (dilute) and 60 μ M (concentrated) NTD with and without 1–1000 RNA. (D) Cross-linking of 35 μ M CTD. (E) Cross-linking of 30 μ M (dilute) and 60 μ M (concentrated) FL-N with and without 1–1000 RNA. To see this figure in color, go online.

measurements are consistent with the expected two-thirds power law relationship between mass and CCS for folded, globular proteins, indicating that the FL-N monomer retains much of the structure observed in the individual domains.

Sedimentation equilibrium AUC shows that both the CTD construct and FL-N are stable dimers with a dimerization constant of 0.4 μ M for FL-N (Fig. S1) and 1 μ M for the CTD (Fig. S2). The dimerization constant of the CTD is

based on a global fit of three protein concentrations and four speeds, whereas the FL-N is based on three concentrations but only one speed because FL-N partially degraded with time and rendered other speeds not reliable. In general, samples for AUC were tested for absence of degradation before and after each run. FL-N degraded within 2–3 days even in the presence of protease inhibitor cocktails, but the CTD remained intact.

NMR analysis of the FL-N dimer shows regions of high disorder

Over 250 peaks are visible in the [^1H , ^{15}N]-TROSY spectra of FL-N out of the expected 446 peaks (Fig. 3 A). CLEANEX spectra collected for FL-N revealed that $\sim$ 150 residues are highly disordered (Fig. 3 D), indicating that in the FL-N, the linkers are disordered as predicted to give rise to CLEANEX peaks. TROSY spectra of NTD and CTD constructs (Fig. 3, B and C) are similar to those published and assigned earlier (11,22). Overlay of NTD TROSY and FL-N CLEANEX shows that nearly all peaks visible in FL-N can be attributed to the NTD domain and disordered regions (Fig. 3, D–F). CTD is not visible in FL-N spectra.

The fact that the NTD peaks and $\sim$ 150 peaks from the disordered linkers and the intrinsically disordered regions (IDRs) are observed in the spectra, whereas none of the CTD peaks are observed, indicates that in this $\sim$ 100-kDa protein, the linker connecting the NTD with the CTD remains significantly flexible such that the NTD in each subunit behaves independently from the rest of the protein with features similar to those of a monomer. The same is not true for the CTD because it is a folded dimer, and thus, its amide peaks are missing in the 100-kDa FL-N. This peak disappearance for the dimerization domain is similar to what we observed for the 66-kDa rabies virus phosphoprotein, which showed peaks for all the residues except for those of the dimerization domain (42).

Bivalency of N is necessary for binding to RNA

Before performing binding studies on FL-N with 1–1000 RNA, we confirmed that the samples in hand can indeed form condensates as previously shown (15). Using light microscopy, the FL-N forms droplet-like aggregates in the presence of 1–1000 RNA (Fig. 4 A), indicating that further studies with this RNA segment are relevant.

Comparison of FL-N, NTD, and CTD with and without 1–1000 RNA by SEC-MALS reveals that at low molar ratios of RNA/protein (0.003:1), FL-N easily binds to RNA, with the FL-N peak shifting to the left, resulting in a tight N/RNA complex with a corresponding molecular weight of $\sim$ 70 $\times$ greater than FL-N dimer alone (Fig. 4 B). In contrast, the same amount of RNA added to NTD or CTD does not cause detectable peak shift (Fig. 4 B).

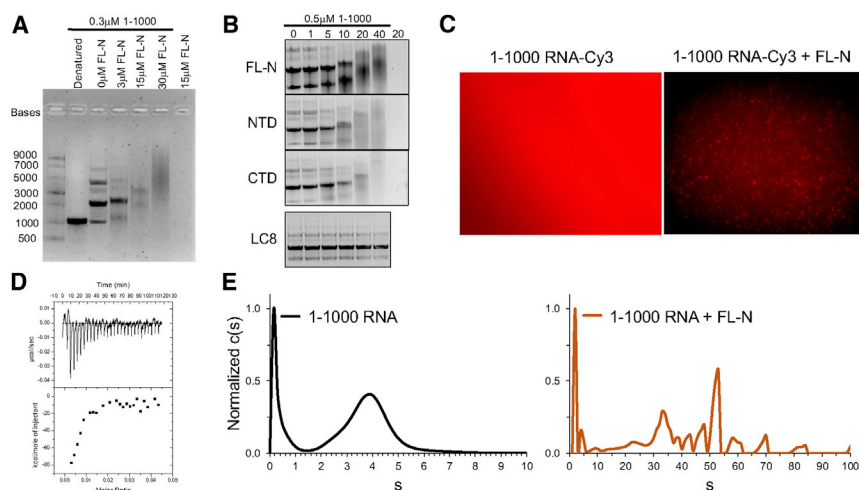


FIGURE 5 Characterization of 1–1000 RNA bound to N protein. (A) 1–1000 RNA gel shift assay with FL-N. From left to right shown as follows: 0.3 μ M denatured RNA showing a 1000 bp band, 0.3 μ M RNA in binding buffer only, 0.3 μ M RNA with 3 μ M FL-N, with 15 μ M FL-N, and with 30 μ M FL-N in binding buffer. The last lane shows 15 μ M FL-N in binding buffer as a control. (B) 1–1000 RNA gels showing binding to FL-N, NTD, and CTD at increasing protein concentration, with the same RNA concentration of 0.5 μ M. From left to right shown as follows: RNA with 1 μ M protein, 5 μ M protein, 10 μ M protein, 20 μ M protein, and 40 μ M protein. The last lane shows 20 μ M protein with no RNA. Same RNA and protein concentrations are shown for LC8 as a negative control. (C) Fluorescence microscopy images of RNA-Cy3 with and without FL-N present. (D) ITC thermogram of 1–1000 RNA titrated into FL-N. (E) Sedimentation velocity AUC at 255 nm of 1–1000 RNA (black) and with FL-N (orange). To see this figure in color, go online.

Glutaraldehyde cross-linking experiments show that NTD is primarily a monomer in solution (~ 14 kDa), with a small dimer population (~ 30 kDa) only observed at a higher concentration of 60 μ M. In the presence of RNA (0.03:1, RNA/protein), higher-order NTD populations appear (>40 kDa) (Fig. 4 C). Cross-linking of CTD without RNA shows a band corresponding to the CTD dimer, consistent with SEC-MALS (~ 35 kDa), and minor populations of larger oligomers (>50 kDa) (Fig. 4 D). FL-N cross-linking shows formation of an array of heterogeneous larger oligomers. With the addition of RNA (0.03:1, RNA/protein), FL-N forms a complex too large to migrate through the gel, but with urea, the higher aggregates are not formed for NTD and FL-N (Fig. 4 E). What is clear from this study is that when the individual NTD binds RNA, it forms higher-order aggregates, suggesting a model in which multiple NTDs, even though they do not dimerize, bind at close enough sites of the RNA to form intermolecular contacts.

Multivalent binding of FL-N to 1–1000 RNA forms heterogeneous large complexes

From SEC-MALS, it is clear that multiple FL-N dimers bind to a single RNA chain, as the FL-N/RNA complex has a molar mass that is $\sim 70\times$ greater than one FL-N dimer bound to 1–1000 RNA (measured molar mass of ~ 7000 kDa bound to RNA vs. ~ 100 kDa free dimer) (Fig. 4 B), suggesting an excess of N dimers bound to this segment of RNA. Furthermore, RNA gels demonstrate significant shifts with FL-N at a ratio of 0.015–0.02:1 RNA/protein, resulting in a smear of larger complexes (Fig. 5 A). Comparison of gel shifts of FL-N, NTD, and CTD show that for FL-N and NTD, there is a shift with 10 μ M or more protein (a ratio of 0.05–0.01:1), but for CTD, a sig-

nificant shift is not apparent until at 20 μ M protein (0.025:1), estimating a K_d of RNA binding of ~ 20 μ M (Fig. 5 B). With LC8, a non-RNA-binding protein, and shown here as a control, there is no gel shift with increasing protein concentration (Fig. 5 B). Fluorescence imaging of 1–1000 RNA with a Cy3 fluorescent tag demonstrates that RNA-Cy3 alone is uniformly dispersed throughout the solution but with the addition of FL-N becomes organized and condensed (Fig. 5 C), similar to previous reports (15).

FL-N binding to 1–1000 RNA measured by isothermal titration calorimetry gives a binding isotherm with an estimated K_d -value of 0.1 μ M ($\Delta G = -9.5$ kcal/mol) and a stoichiometry of 0.0027:1 RNA/protein. This large ratio is consistent with the ratio observed with SEC-MALS, indicating that for every 1–1000 RNA strand, 70–187 N-protein dimers are bound, depending on sample conditions and the method used (Fig. 5 D). Binding is accompanied by very large entropic and enthalpic contributions ($T\Delta S = -408$ kcal/mol and $\Delta H = -417$ kcal/mol), which is not surprising for this large number of multivalent interactions. AUC measurements recorded at 255 nm to bias signal toward RNA present a clear difference between free RNA and RNA bound to FL-N, with free RNA sedimenting as a single broad peak with a sedimentation coefficient of 4 S (Fig. 5 E). In contrast, FL-N/1–1000 RNA produced a heterogeneous sedimentation profile with s -values on a scale $10\times$ greater than free RNA, suggesting formation of multiple large complexes of variable stoichiometry (Fig. 5 F).

Turbidity of FL-N in the presence of RNA is concentration, pH, and time dependent

The turbidity of FL-N increases in the 5–20 μ M protein concentration with RNA purified from mouse skin fibroblast

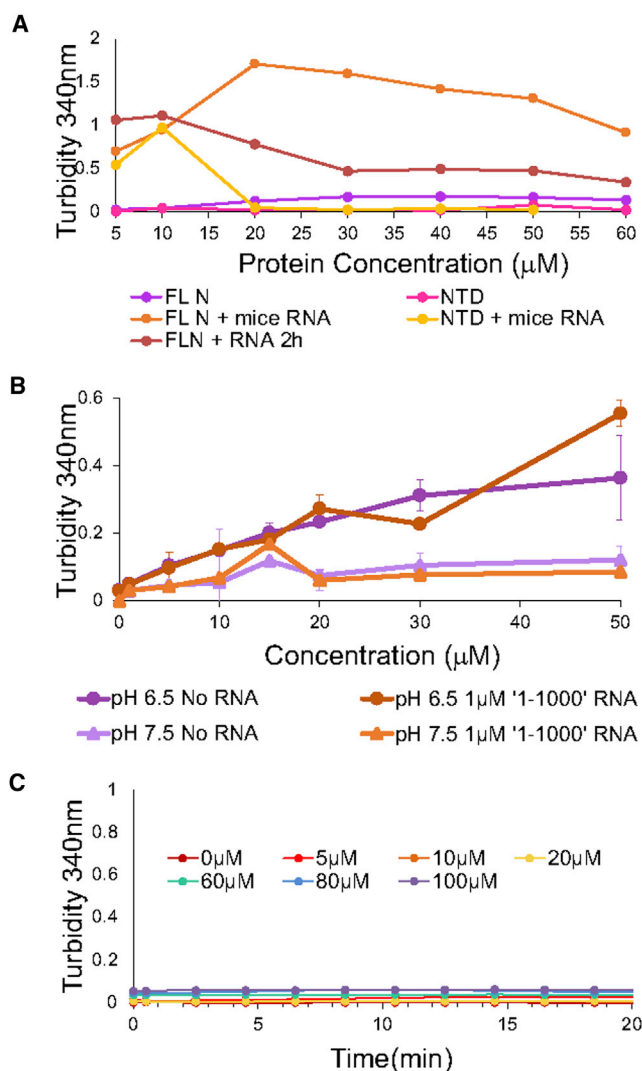


FIGURE 6 Effect of RNA binding on turbidity of N protein. (A) Turbidity measurements of FL-N and NTD with and without mouse RNA at increasing concentrations. Shown are FL-N (purple), NTD (pink), FL-N with mouse RNA (orange), NTD with mouse RNA (yellow), and FL-N incubated with mouse RNA for 2 h (red). Points represent the mean of triplicates. Standard deviation error bars are smaller than marker. (B) Turbidity measurements of FL-N with 1–1000 RNA at increasing concentrations, in pH 6.5 or 7.5. Shown are the free FL-N at pH 6.5 (dark purple), RNA + FL-N at pH 6.5 (dark orange), free FL-N at pH 7.5 (light purple), and RNA + FL-N at pH 7.5 (light orange). Points represent the mean of triplicates and error bars represent $\pm$ the SD. (C) Turbidity measurements of NTD at multiple concentrations with the addition of 1–1000 RNA at time 0. To see this figure in color, go online.

and reaches a plateau or slightly decreases with increasing concentrations (Fig. 6 A). After 2 h of incubation, turbidity is significantly decreased (FLN + RNA 2 h). When FL-N is incubated with 1.1 μ M 1–1000 viral RNA at multiple protein concentrations in pH 6.5 and 7.5, higher turbidity is observed overall at pH 6.5, with a significant difference arising only between free and RNA bound at a concentration of 50 μ M FL-N (Fig. 6 B). Higher turbidity is observed for

binding the nonspecific mouse RNA relative to 1–1000 RNA, and similar to an earlier report, higher turbidity is observed with lower pH (28). NTD alone does not show turbidity with the addition of 1–1000 RNA, regardless of protein concentration (Fig. 6 C), although there was some turbidity with the nonspecific RNA (Fig. 6 A). Therefore, what these experiments show is that turbidity is highest with FL-N and when RNA is bound at lower pH.

Disordered regions of FL-N are largely unaffected by RNA binding

Upon addition of a small molar ratio of 1–1000 RNA to 15 N-labeled FL-N (0.008:1), NMR spectra show that peaks corresponding to the NTD decrease in intensity or disappear completely (Fig. 7 A). A majority of peaks corresponding to the most disordered residues remain mostly unchanged, with only a few CLEANEX peaks decreasing in intensity (Fig. 7 B). Resonance assignments of NTD were obtained by comparing our spectra with published assignments at similar conditions (BMRB: 34511). Mapping the change in intensity in FL-N upon binding to RNA on the NTD structure shows that the majority of the peaks have decreased in intensity (orange), and only few unchanged (white) (Fig. 7 C). Residues that increase in peak intensity (dark purple) are largely due to peak overlap (Fig. 7 C). Because the majority of the peaks within the NTD disappear, we conclude that the full NTD domain is locked in a bound conformation and that the IDRs of FL-N do not significantly contribute to binding because they do not change in peak intensity or chemical shifts.

DISCUSSION

Within the SARS-CoV-2 nucleocapsid, hundreds to thousands (20) of N proteins package the gRNA while maintaining their ability to make other protein interactions, such as with the M protein within a fully formed virion (30) and with papain-like protease Nsp3, enabling transfer of gRNA to a replicase (14). In this space, the N protein must also make organized, favorable interactions with other N proteins. Although there have been a large number of reports on atomic level characterization of the N protein, the majority are focused on structural characterization of the smaller NTD and CTD (7,11,14,29,43). Characterization of FL-N is scarce and limited to conflicting reports of the association state (10,44,45). Furthermore, the few biophysical studies that incorporate RNA binding, either on individual domains or FL-N, have been limited to short segments of RNA or to synthesized polynucleotide RNA models rather than the viral RNA (11,12,14,43). These limited and conflicting reports are not surprising considering that the FL-N is a large multidomain protein with three long disordered segments and is prone to autoproteolysis (46) and aggregation as well as nonspecific binding to bacterial proteins and

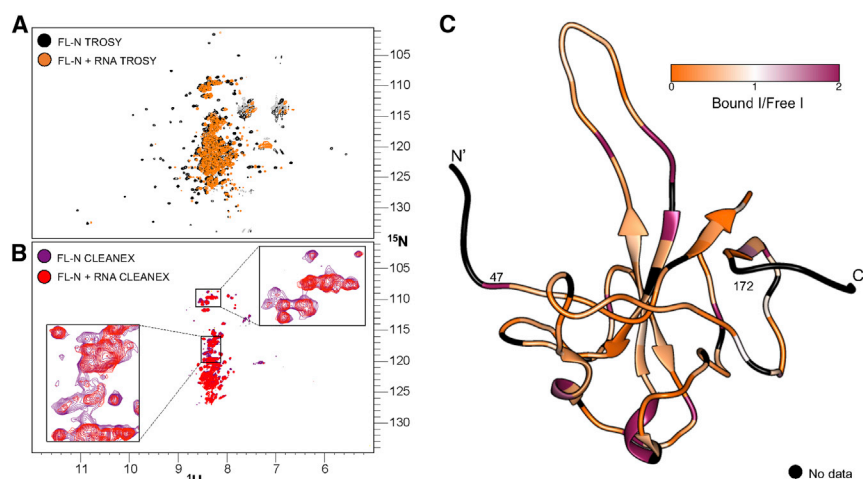


FIGURE 7 RNA binding to FL-N characterized by NMR. (A) Overlay of free FL-N 2D- $[^1\text{H}, ^{15}\text{N}]$ -TROSY (purple) with FL-N + 1–1000 RNA 2D- $[^1\text{H}, ^{15}\text{N}]$ -TROSY (orange) at 25°C. (B) Overlay of free FL-N CLEANEX (dark purple) with FL-N + 1–1000 RNA CLEANEX (red). (C) Ratio of RNA-bound/unbound peak intensities mapped onto NTD (PDB: 6YI3). Residues corresponding to peaks that decrease in intensity upon binding are colored orange. Residues corresponding to peaks that increase in intensity upon binding, colored in purple, are attributed to peak overlap. Residues not present in our construct or that do not have peaks visible in TROSY spectra are colored black. To see this figure in color, go online.

RNA during recombinant expression and purification, as we find in our studies. Our work here, to our knowledge, offers the first thorough characterization of FL-N and its binding to a 1000-nt segment of the viral RNA and presents a model that explains the importance of FL-N bivalent structure to enhancing its binding affinity on multivalent sites of a long segment of the viral RNA.

SEC-MALS and AUC show that FL-N is a stable dimer in solution (Fig. 2) with a dimerization constant that is lower than 0.5 μM estimated from the loading protein concentration by SEC and of 0.4 μM measured from AUC at a single speed (Fig. S1), whereas CD and NMR reveal a structure composed of folded NTD and CTD that retain the structures of the individual constructs. These domains are linked and flanked by disordered linkers that provide significant flexibility to the NTD within the 100-kDa dimer sufficient to show sharp peaks in the TROSY HSQC spectra (Fig. 3). Native mass spectrometry identifies a monomeric species

in addition to the dimer that has folded domains and is populated at low protein concentrations (Fig. 2), whereas cross-linking and turbidity measurements suggest the formation of larger N oligomers at high protein concentration and lower pH. These results show that a stable dimer is the primary species but also explains the conflicting reports of different association states in the literature.

The FL-N protein binds to the 1–1000 RNA to form a tight heterogeneous complex; while under similar conditions, no binding is observed to the individual NTD and CTD by SEC-MALS (Fig. 4 B). We attribute the tighter binding to the bivalent nature of FL-N, which uses the two flexible NTD domains that are anchored by the CTD dimerization domain to latch on the RNA. With cross-linking experiments, the NTD forms larger oligomeric states in the presence of higher RNA to protein ratios than what is used in SEC-MALS or NMR (0.03:1 instead of 0.003:1), suggesting that the NTD on its own can weakly binds

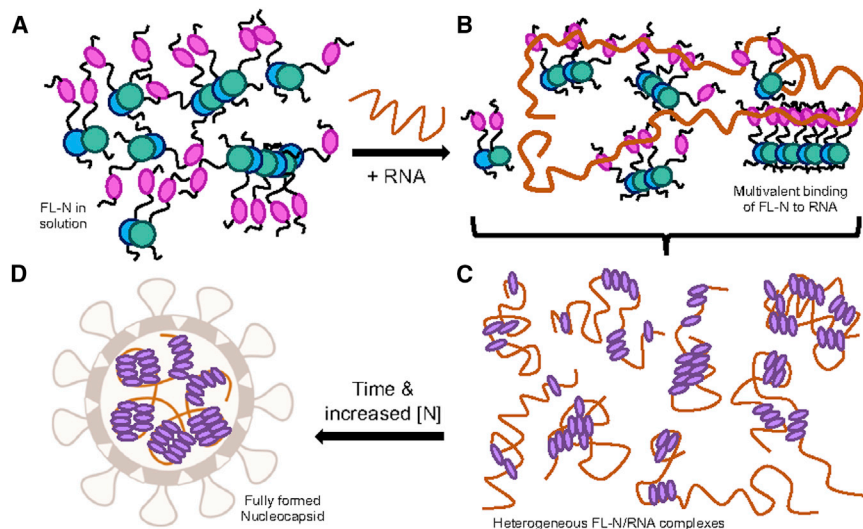


FIGURE 8 A Model of SARS-CoV-2 nucleocapsid formation. (A) In solution, N proteins exist as dimers with disordered linkers but could make larger oligomers through CTD interactions at high protein concentration. (B) In the presence of RNA, the NTD binds multivalently along the RNA, bringing N proteins into closer contact with each other, making contacts with other N proteins primarily through the CTD. Localized higher concentrations enable NTD to make intermolecular interactions with other NTDs. (C) FL-N binds to RNA at multiple sites along the strand, resulting in the formation of heterogeneous complexes. (D) As more N proteins become available, stabilizing interactions between RNA and proteins result in an organized nucleocapsid that can be later packaged into fully-assembled progeny virions. To see this figure in color, go online.

RNA. As NTD binds to RNA, localized NTD concentrations increase to favor some NTD intermolecular association, resulting in >40 kDa protein complexes (Fig. 4 C). FL-N cross-linking with the same amount of RNA results in a complex too large to enter the well of the gel (Fig. 4 E). Therefore, FL-N most readily oligomerizes without RNA and most readily binds to RNA. CTD on its own also readily oligomerizes, but oligomerization of the NTD can occur at a higher concentration and in the presence of RNA, underscoring the importance of bivalent N for efficient interactions with RNA. FL-N binding to RNA creates heterogeneous complexes observed by RNA gel shift assays, fluorescence microscopy, and AUC (Fig. 5). Therefore, taken together, these results show that FL-N binds to 1–1000 RNA at multiple locations, simultaneously forming complexes with diverse sizes rather than creating a few distinct oligomeric species.

FL-N's multivalent binding to RNA assessed by NMR reveals that in the context of FL-N and with a 1000-nt region of SARS-CoV-2 RNA, all of the NTD is impacted by binding, supporting a model that multivalent binding along RNA brings together many copies of NTD, changing the tumbling rate and chemical environment of NTD as a whole rather than only impacting key RNA-binding residues, as observed with small segments of RNA (11,14). Disordered regions remain largely unaffected by RNA binding (Fig. 7 B) and do not appear to gain structure or make significant interactions with other regions of FL-N. Because the disordered linker resides within the SR-rich region, the hnRNP1 binding site, and the nsp3 binding site, and within the C-terminal IDR is a K-rich region and the putative TRIM25 E3 ligase binding site, it seems essential for proper function of the N protein that these disordered regions remain disordered and available for binding or for phosphorylation.

The FL-N dimerization and multivalent RNA binding is critical for the viral life cycle and could inform potential drug therapy because both the NTD (6,7,47,48) and CTD (49) are potential drug targets. Therefore, developing compounds that either disrupt the RNA binding or dimerization requires understanding of not only how these compounds interfere with the individual domains but with the larger protein/RNA complexes involving many FL-N copies and longer RNA (6–8). Because we see different RNA-binding behavior with FL-N compared with individual domains, we expect that compounds developed to disrupt dimerization would also disrupt RNA binding and larger nucleocapsid assembly. Furthermore, because RNA-binding affinity is different in the FL-N and individual domains, investigations into drug compound interactions with the N protein must be done on FL-N protein along with individual domains.

Our work presented here illustrates a model of how the SARS-CoV-2 nucleocapsid phosphoprotein assembles multivalently along its ~30 kb genome (Fig. 8). Flexible and dimeric in solution, the bivalent N can maintain a propensity

to form large, heterogeneous oligomers originating in the CTD. Although both the NTD and CTD of FL-N bind to the 1–1000 RNA, the NTD is the primary RNA-binding domain and can only bind efficiently when part of the FL-N and not as an individual domain. Because NTD binds to many sites on 1–1000 RNA, the increase in effective protein concentration facilitates increased oligomerization, and organization of RNA. Organization of the ~30-kb genome by N proteins results in a ribonucleoprotein that is dynamic, with the disordered regions of N remaining disordered and primed for further interactions.

In summary, the N protein is a partially disordered duplex that binds significantly tighter to RNA than its individual domains, demonstrating that it is necessary for the NTD to be part of a bivalent duplex for efficient binding. Another major novel finding is binding to the 1000 nt RNA, the largest viral RNA studied to date by biophysical methods, results in a heterogeneous N/RNA complex with an N/RNA ratio in the range of 70–187 dimers per RNA, with 70 dimers making the most stable core and with the NTD being the primary binding domain, whereas the disordered linkers are not affected in the N/RNA complex. The remaining disorder in the N/RNA complex is essential for the multifaceted roles of N in viral replication.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2021.03.023>.

AUTHOR CONTRIBUTIONS

H.M.F. and E.B. designed the study. J.R.G., Z.Y., S.P., P.R., A.D.R., J.S.P., and H.M.F. performed the research and analyzed the data. R.B.C. and P.Z. designed expression vector and determined expression and purification protocols for NTD. H.M.F. and E.B. wrote the manuscript with input from all authors. E.B. oversaw the study.

ACKNOWLEDGMENTS

The authors wish to acknowledge the COVID team at Oregon State University, in particular Dr. David Hendrix and Brittany Lasher for many helpful discussions and Dr. Michael Hare of e-MSion for mass spectrometry sample analysis.

This work is funded by a National Science Foundation EAGER MCB 2034446 to E.B. We also acknowledge the support of the Oregon State University NMR Facility funded in part by the National Institutes of Health, HEI Grant 1S10OD018518 and by the M.J. Murdock Charitable Trust grant 2014162. Native mass spectrometry data acquisition was supported by a National Institutes of Health T32 GM007759 (to A.D.R.). A.D.R. is an ARCS Scholar supported by the ARCS Oregon Chapter.

REFERENCES

1. World Health Organization. 2020. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization, Geneva, Switzerland.

2. Carfi, A., R. Bernabei, F. Landi; Gemelli Against COVID-19 Post-Acute Care Study Group. 2020. Persistent symptoms in patients after acute COVID-19. *JAMA*. 324:603–605.
3. Greenhalgh, T., M. Knight, ..., L. Husain. 2020. Management of post-acute covid-19 in primary care. *BMJ*. 370:m3026.
4. Cong, Y., M. Ulasli, ..., F. Reggiori. 2020. Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle. *J. Virol*. 94:e01925-19.
5. McBride, R., M. van Zyl, and B. C. Fielding. 2014. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*. 6:2991–3018.
6. Peng, Y., N. Du, ..., H. Song. 2020. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *EMBO J*. 39:e105938.
7. Kang, S., M. Yang, ..., S. Chen. 2020. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B*. 10:1228–1238.
8. Gordon, D. E., G. M. Jang, ..., N. J. Krogan. 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 583:459–468.
9. Chukwudozie, O. S., R. C. Chukwuanukwu, ..., E. U. Okinedo. 2020. Attenuated subcomponent vaccine design targeting the SARS-CoV-2 nucleocapsid phosphoprotein RNA binding domain: in silico analysis. *J. Immunol. Res*. 2020:2837670.
10. Zeng, W., G. Liu, ..., T. Jin. 2020. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem. Biophys. Res. Commun*. 527:618–623.
11. Dinesh, D. C., D. Chalupska, ..., E. Boura. 2020. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog*. 16:e1009100.
12. Chen, C.-Y., C. K. Chang, ..., T. H. Huang. 2007. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol*. 368:1075–1086.
13. Saikatendu, K. S., J. S. Joseph, ..., P. Kuhn. 2007. Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J. Virol*. 81:3913–3921.
14. Khan, A., M. Tahir Khan, ..., D. Q. Wei. 2020. Structural insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein. *Comput. Struct. Biotechnol. J*. 18:2174–2184.
15. Iserman, C., C. A. Roden, ..., A. S. Gladfelter. 2020. Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol. Cell*. 80:1078–1091.e6.
16. Barik, S. 2020. Genus-specific pattern of intrinsically disordered central regions in the nucleocapsid protein of coronaviruses. *Comput. Struct. Biotechnol. J*. 18:1884–1890.
17. Wu, C.-H., S. H. Yeh, ..., P. J. Chen. 2009. Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication. *J. Biol. Chem*. 284:5229–5239.
18. Cong, Y., F. Kriegenburg, ..., F. Reggiori. 2017. Coronavirus nucleocapsid proteins assemble constitutively in high molecular oligomers. *Sci. Rep*. 7:5740.
19. Luo, H., F. Ye, ..., H. Jiang. 2004. In vitro biochemical and thermodynamic characterization of nucleocapsid protein of SARS. *Biophys. Chem*. 112:15–25.
20. Bar-On, Y. M., A. Flamholz, ..., R. Milo. 2020. SARS-CoV-2 (COVID-19) by the numbers. *eLife*. 9:e57309.
21. Zhou, R., R. Zeng, ..., J. Lei. 2020. Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Mol. Bio-med*. 1:2.
22. Takeda, M., C. K. Chang, ..., M. Kainosho. 2008. Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J. Mol. Biol*. 380:608–622.
23. Chang, C. K., M. H. Hou, ..., T. H. Huang. 2014. The SARS coronavirus nucleocapsid protein-forms and functions. *Antiviral Res*. 103:39–50.
24. Guseva, S., S. Milles, ..., M. Blackledge. 2020. Structure, dynamics and phase separation of measles virus RNA replication machinery. *Curr. Opin. Virol*. 41:59–67.
25. Sengupta, P., A. Y. Seo, ..., J. Lippincott-Schwartz. 2019. A lipid-based partitioning mechanism for selective incorporation of proteins into membranes of HIV particles. *Nat. Cell Biol*. 21:452–461.
26. Guseva, S., S. Milles, ..., M. Blackledge. 2020. Measles virus nucleocapsid and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. *Sci. Adv*. 6:eaa7095.
27. Monette, A., M. Niu, ..., A. J. Mouland. 2020. Pan-retroviral nucleocapsid-mediated phase separation regulates genomic RNA positioning and trafficking. *Cell Rep*. 31:107520.
28. Perdikari, T. M., A. C. Murthy, ..., N. L. Fawzi. 2020. SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J*. 39:e106478.
29. Savastano, A., A. Ibáñez de Opakua, ..., M. Zweckstetter. 2020. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat. Commun*. 11:6041.
30. Lu, S., Q. Ye, ..., K. D. Corbett. 2021. The SARS-CoV-2 Nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun*. 12:502.
31. Cubuk, J., J. J. Alston, ..., A. S. Holehouse. 2020. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *bioRxiv* <https://doi.org/10.1101/2020.06.17.158121>.
32. Chen, H., Y. Cui, ..., J. Lou. 2020. Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res*. 30:1143–1145.
33. Frey, S., and D. Görlich. 2014. Purification of protein complexes of defined subunit stoichiometry using a set of orthogonal, tag-cleaving proteases. *J. Chromatogr. A*. 1337:106–115.
34. Carlson, C. R., J. B. Asfaha, ..., D. O. Morgan. 2020. Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol. Cell*. 80:1092–1103.e4.
35. Laue, T. M. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Royal Society of Chemistry, pp. 90–125.
36. Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J*. 78:1606–1619.
37. Cole, J. L., J. W. Lary, ..., T. M. Laue. 2008. Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol*. 84:143–179.
38. Bush, M. F., Z. Hall, ..., B. T. Ruotolo. 2010. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem*. 82:9557–9565.
39. Salbo, R., M. F. Bush, ..., K. F. Haselmann. 2012. Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers. *Rapid Commun. Mass Spectrom*. 26:1181–1193.
40. Konijnenberg, A., A. Butterer, and F. Sobott. 2013. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta*. 1834:1239–1256.
41. Donor, M. T., S. A. Ewing, ..., J. S. Prell. 2017. Extended protein ions are formed by the chain ejection model in chemical supercharging electrospray ionization. *Anal. Chem*. 89:5107–5114.
42. Jespersen, N. E., C. Leyrat, ..., E. Barbar. 2019. The LC8-RavP ensemble structure evinces A role for LC8 in regulating lyssavirus polymerase functionality. *J. Mol. Biol*. 431:4959–4977.
43. Huang, Q., L. Yu, ..., E. T. Olejniczak. 2004. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry*. 43:6059–6063.

44. Ye, Q., A. M. V. West, ..., K. D. Corbett. 2020. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* 29:1890–1901.
45. Zinzula, L., J. Basquin, ..., W. Baumeister. 2021. High-resolution structure and biophysical characterization of the nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute respiratory syndrome coronavirus 2. *Biochem. Biophys. Res. Commun.* 538:54–62.
46. Lutomski, C. A., T. J. El-Baba, ..., C. V. Robinson. 2020. Autoproteolytic products of the SARS-CoV-2 nucleocapsid protein are primed for antibody evasion and virus proliferation. *bioRxiv* <https://doi.org/10.1101/2020.10.06.328112>.
47. Lang, Y., K. Chen, ..., H. Li. 2020. The nucleocapsid protein of zoonotic betacoronaviruses is an attractive target for antiviral drug discovery. *Life Sci* Published online November 12, 2020. <https://doi.org/10.1016/j.lfs.2020.118754>.
48. Dhankhar, P., V. Dalal, ..., P. Kumar. 2020. Computational guided identification of novel potent inhibitors of N-terminal domain of nucleocapsid protein of severe acute respiratory syndrome coronavirus 2. *J. Biomol. Struct. Dyn* Published online November 30, 2020. <https://doi.org/10.1080/07391102.2020.1852968>.
49. Ahamad, S., D. Gupta, and V. Kumar. 2020. Targeting SARS-CoV-2 nucleocapsid oligomerization: insights from molecular docking and molecular dynamics simulations. *J. Biomol. Struct. Dyn* Published online November 3, 2020. <https://doi.org/10.1080/07391102.2020.1839563>.