

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Assembly of SARS-CoV-2 nucleocapsid protein with nucleic acid.

### Permalink

<https://escholarship.org/uc/item/98h4w56x>

### Journal

Nucleic Acids Research, 52(11)

### Authors

Zhao, Huaying

Syed, Abdullah

Khalid, Mir

et al.

### Publication Date

2024-06-24

### DOI

10.1093/nar/gkae256

Peer reviewed

# Assembly of SARS-CoV-2 nucleocapsid protein with nucleic acid

Huaying Zhao<sup>1</sup>, Abdullah M. Syed<sup>2,3</sup>, Mir M. Khalid<sup>2</sup>, Ai Nguyen<sup>1</sup>, Alison Ciling<sup>2,3</sup>, Di Wu<sup>4</sup>, Wai-Ming Yau<sup>5</sup>, Sanjana Srinivasan<sup>1</sup>, Dominic Esposito<sup>6</sup>, Jennifer A. Doudna<sup>2,7,8,9,10,11</sup>, Grzegorz Piszczek<sup>4</sup>, Melanie Ott<sup>2,12,13</sup> and Peter Schuck<sup>1,14,\*</sup>

<sup>1</sup>Laboratory of Dynamics of Macromolecular Assembly, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD 20892, USA

<sup>2</sup>Gladstone Institutes, San Francisco, CA 94158, USA

<sup>3</sup>Innovative Genomics Institute, University of California, Berkeley, CA 94720, USA

<sup>4</sup>Biophysics Core Facility, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

<sup>5</sup>Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

<sup>6</sup>Protein Expression Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

<sup>7</sup>Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

<sup>8</sup>HHMI, University of California, Berkeley, CA 94720, USA

<sup>9</sup>Department of Chemistry, University of California, Berkeley, CA 94720, USA

<sup>10</sup>Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

<sup>11</sup>California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA

<sup>12</sup>Department of Medicine, University of California, San Francisco, CA 94143, USA

<sup>13</sup>Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

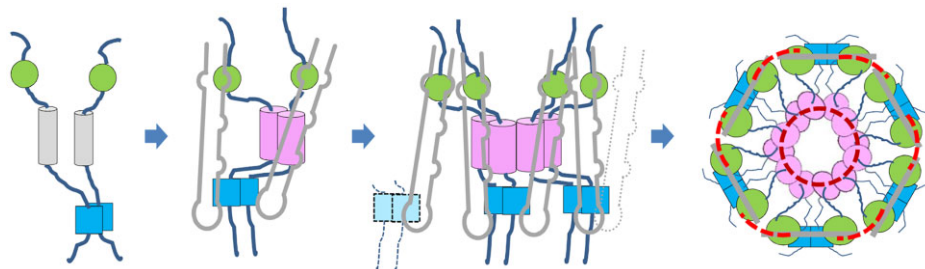
<sup>14</sup>Center for Biomedical Engineering Technology Acceleration, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD 20892, USA

\*To whom correspondence should be addressed. Tel: +1 301 435 1950; Email: [schuckp@mail.nih.gov](mailto:schuckp@mail.nih.gov)

## Abstract

The viral genome of SARS-CoV-2 is packaged by the nucleocapsid (N-)protein into ribonucleoprotein particles (RNPs), 38 ± 10 of which are contained in each virion. Their architecture has remained unclear due to the pleomorphism of RNPs, the high flexibility of N-protein intrinsically disordered regions, and highly multivalent interactions between viral RNA and N-protein binding sites in both N-terminal (NTD) and C-terminal domain (CTD). Here we explore critical interaction motifs of RNPs by applying a combination of biophysical techniques to ancestral and mutant proteins binding different nucleic acids in an *in vitro* assay for RNP formation, and by examining nucleocapsid protein variants in a viral assembly assay. We find that nucleic acid-bound N-protein dimers oligomerize via a recently described protein–protein interface presented by a transient helix in its long disordered linker region between NTD and CTD. The resulting hexameric complexes are stabilized by multivalent protein–nucleic acid interactions that establish crosslinks between dimeric subunits. Assemblies are stabilized by the dimeric CTD of N-protein offering more than one binding site for stem–loop RNA. Our study suggests a model for RNP assembly where N-protein scaffolding at high density on viral RNA is followed by cooperative multimerization through protein–protein interactions in the disordered linker.

## Graphical abstract



## Introduction

Packaging the genome into the viral particle prior to viral egress, and the reverse, uncoating after entering a new host cell, are critical steps in the viral life cycle. dsDNA viruses of-

ten employ molecular motors to pump the nucleic acid into preformed capsids generating high pressure. By contrast, ss-RNA viruses, owing to the much higher flexibility of ssRNA, usually first condense their genome through spontaneous

cooperative assembly with nucleocapsid (N-)proteins into ribonucleoprotein particles (RNPs) that are then embedded into a budding viral envelope (1,2). Coronaviruses are examples of the latter: after synthesis of their large +ssRNA genome in the confines of double membrane vesicles (DMVs), soon after the genomic RNA (gRNA) is released through membrane pores into the cytoplasm, it is thought to be condensed into discrete RNPs like beads on a string (3–5), followed by transport to viral budding sites in the endoplasmic reticulum–Golgi intermediate compartment (ERGIC), where RNPs interact with membrane-bound matrix (M-)protein and trigger formation of viral particles and their egress (6–9). Unfortunately, little molecular and atomistic detail is known about most of these processes, and while coronavirus RNA-protein interactions have been well studied in transcription and replication (10), the architectural principles and assembly pathways of the RNPs in viral packaging are largely unclear. Unlike the symmetrical nucleocapsids of many helical ssRNA viruses (2), coronavirus RNPs are pleomorphic, significantly complicating the analysis of the multivalent protein–protein and sequence non-specific protein–RNA interactions.

The present work is focused on the assembly principles of SARS-CoV-2—a persistent threat to global health that requires continued research for the development of new vaccines and antivirals counteracting ongoing viral immune evasion and mutagenesis. Viral assembly is a potential target for therapeutic development. As observed in recent cryo-electron tomography (cryo-ET) studies, virions contain  $\approx 38 \pm 10$  ribonucleoprotein particles (vRNPs) that condense the  $\approx 30$  kb gRNA into vRNPs with 14–16 nm sized pillar-shaped structures (3,4). SARS-CoV-2 N-protein is a 419 aa protein expressed in infected cells, at levels up to 1% of total protein (11,12). Besides its key role in vRNP formation it is highly multifunctional (6,13–15). It consists in two folded domains, the NTD and CTD, which are linked and flanked by large intrinsically disordered regions (IDRs) termed N-arm, linker, and C-arm (7) (Figure 1A). The NTD forms a nucleic acid (NA) binding site that is thought to harbor recognition mechanism of a packaging signal (16). The CTD provides a high-affinity dimerization interface and also binds NA (7) (Figure 1B). Ultra-weak protein–protein interactions allow N-protein to undergo liquid-liquid phase separation (LLPS), enhanced in the presence of NA, to form droplets of high macromolecular density (17–20), which are thought to facilitate the assembly of vRNPs (18,20–22). IDRs also play important roles in N-protein NA interactions (23–26). For example, the N-arm undergoes conformational changes upon NA binding to NTD, significantly increasing the affinity of the NTD for RNA (24). Similar to the N-arm, an SR-rich portion of the linker IDR flanking the NTD is also perturbed by NA binding (23), as is an L-rich sequence of the linker IDR (LRS) further downstream.

Recently, we have shown that highly conserved transient helices in the LRS can assemble cooperatively into coiled-coil trimers, tetramers, and higher oligomers, in a process that is allosterically enhanced by NA binding (26,27) (Figure 1C). This oligomerization interface is highly conserved, both within the mutational landscape of SARS-CoV-2 as well as across coronaviruses, and provides a new element for understanding the elusive architecture of vRNPs in coronavirus assembly. We previously hypothesized that this mechanism might initiate assembly by providing additional protein–protein interfaces essential for stabilizing 3D structures of the vRNPs (26,27).

To study the assembly mechanism in more detail we examine N-protein binding to short NA probes of sufficient length for multivalent binding and ask how scaffolding N-protein and different NA structure might diminish or augment simultaneous LRS oligomerization. As shown by Carlson and co-workers, mixtures of stem–loop RNA and N-protein form large RNA/protein complexes (RNPs) that may serve as a model for the viral vRNPs (20,25). In the current study, to further elucidate the assembly mechanism we utilize this system and compare the size distribution and composition of complexes formed by wildtype N-protein ( $N_{WT}$ ) binding either short linear oligonucleotides or stem–loop RNA. To probe the involvement of the LRS protein–protein interface we examine mutants that abrogate LRS oligomerization. In parallel, we study the impact of the same mutations on viral packaging and assembly in a VLP assay. Our data highlight the role of the LRS oligomerization and the interplay with nucleic acid binding of NTD and CTD in the formation of RNPs. Taken together, these elements suggest features of possible RNP assembly models on the level of protein domains and their interfaces.

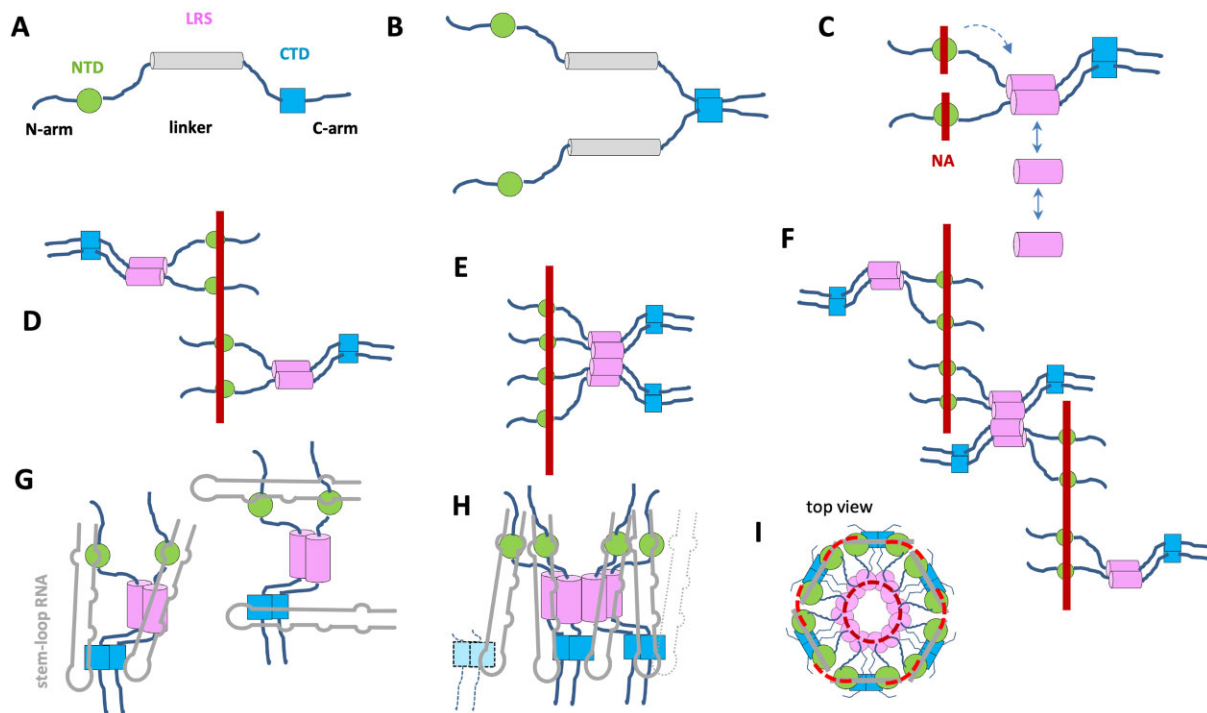
## Materials and methods

### Protein expression and purification

Wildtype and mutant full-length SARS-CoV-2 N-proteins were expressed and purified as described previously (26,27). Briefly, the full-length protein was cloned into the pET-29a(+) expression vector fused to DNA encoding a 6xHis tag followed by a Tobacco Etch Virus (TEV) cleavage site at N-terminus. The vector was transformed into One Shot BL21(DE3)pLysS *Escherichia coli* (Thermo Fisher Scientific, Carlsbad, CA). After cell lysis, the protein was bound to a Ni-NTA column and subjected to unfolding and refolding steps to remove residual protein-bound bacterial nucleic acid (20). After  $Ni^{2+}$  affinity chromatography, the 6xHis tag of the eluate was cleaved by TEV and the tag-free protein purified by affinity and size exclusion chromatography. 95% purity of the proteins was confirmed by SDS-PAGE. The ratio of absorbance at 260 nm and 280 nm of  $\sim 0.50$ – $0.55$  confirmed absence of nucleic acid.

SARS-CoV-2 N-protein NTD (48–173) was cloned into kanamycin-resistant, NdeI/XhoI-digested plasmid pET29a vector (GenScript) with 6xHis tag included at the N-terminus. After expression and lysis, NTD was purified by Ni-NTA chromatography. The C-terminal domain (247–364) was generated as previously described (28): Briefly, an N-terminal 6xHis tagged N-CTD construct preceded by a TEV protease cleavage site was expressed in *E. coli* BL21(DE3) using Dyna-mite Broth as described in (29). After expression and lysis, protein was purified by  $Ni^{2+}$  affinity chromatography followed by cleavage of the tag and purification by affinity and size exclusion chromatography. Protein purity was validated by SDS-PAGE chromatography and electrospray ionization mass spectrometry.

The oligonucleotides T<sub>40</sub>, U<sub>40</sub> and stem–loop RNA SL7 were purchased from Integrated DNA Technologies (Skokie, IL) and purified by HPLC and lyophilized by the vendor. After reconstitution, to obtain optimal RNA secondary structure, SL7 was subject to thermal denaturation at 95°C for 2 min followed by gradual cooling to room temperature over 1–2 h.



**Figure 1.** Schematic organization of N-protein and hypothetical architectures of NA complexes. **(A)** Organization of N-protein chain with folded domains NTD (green cylinder) and CTD (blue square), and the intrinsically disordered N-arm, C-arm, and linker, the latter containing the transient helix in the leucine rich sequence (LRS) capable of oligomerization (cylinder). **(B)** N-protein in solution is a dimer linked with high affinity at the CTD. **(C)** Occupation of NA binding sites in the NTD induces folding in the LRS and causes compaction (magenta) and LRS oligomerization in dimers, trimers, tetramers, and higher oligomers. **(D)** Configuration of two N-protein dimers independently scaffolded on NA  $T_{40}$  (red bar) without further LRS oligomerization. **(E)** Two N-protein dimers scaffolded on NA and stabilized through LRS oligomerization. **(F)** Similar to **(E)** with crosslink between NA strands allowing the formation of higher oligomers. **(G)** N-protein dimer with two SL7 stem-loop RNA ( $\{2N/2SL7\}$ , grey), depicted in alternate configurations occupying all major NA interfaces in the NTD and CTD creating different intra-dimer crosslinks. **(H)** Possible architecture of N-protein/stem-loop complexes allowing  $\{2N/2SL7\}$  units to oligomerize *via* LRS and simultaneous multivalent binding of SL7 in inter-dimer crosslinks. Dotted lines belong to neighboring  $\{2N/2SL7\}$  units not fully drawn. **(I)** The top view of a  $6 \times \{2N/2SL7\}$  hexamer of dimers. The dashed lines indicates two levels of inter-dimer crosslinks of neighboring  $\{2N/2SL7\}$  subunits, *via* contacts of the LRS interfaces (dark red dashed) and *via* multivalent RNA binding of CTD and/or scaffolding of NTD (light red dashed).

Prior to biophysical experiments, protein and nucleic acid samples were dialyzed into working buffer and monomeric stock concentrations were determined by UV-Vis spectrophotometry, using a theoretical extinction coefficient of 43 890 OD/(M × cm) for N-protein. A refined concentration determination was carried out after a sedimentation velocity experiment, by determining the amplitude of the sedimentation boundaries observed in interference optical data and employing a refractive-index signal increment predicted by amino acid composition (30). More detailed information on all protein constructs can be found in the [Supplementary Materials and Methods](#). Biophysical experiments were generally carried out two to four times, using protein batches from multiple independent preparations.

### Virus-like particle assay

A virus-like particle (VLP) assay was employed as a physiological model to test the efficiency of packaging and assembly as a function of the mutations on SARS-CoV-2 N protein (31). Implementation in the present work mirrors the detailed description in (32). Briefly, the VLPs were generated by co-expressing all four structural proteins of SARS-CoV-2 in HEK 293T cells. A 2 kb viral packaging signal was incorporated into the untranslated region of a luciferase reporter plasmid, resulting in the transcripts to be packaged in the VLPs, which were added

to receiver 293T cells expressing ACE2 and TMPRSS2. Luminescence in receiver cells was measured using a Promega Luciferase Assay System (Promega E1501). More details can be found in the [Supplementary Materials and Methods](#).

### Mass photometry

Mass photometry (MP) experiments were carried out using a TwoMP instrument (Refeyn, UK). Samples were loaded in mini-wells created by a silicone gasket placed on top of a microscope coverslip on the microscope stage. First, the samples were prepared by mixing the stock solutions with the working buffers in the Eppendorf tubes. Prior to MP data acquisition, the same working buffer (10–15  $\mu$ l) was loaded into the mini-well to focus the objective, after which 1 or 2  $\mu$ l of the sample was added to the buffer droplet, mixed and measured immediately.

The TwoMP instrument was calibrated with Beta-Amylase from Sweet Potato (Sigma A8781) and Thyroglobulin from Bovine Thyroid (Sigma T9145) as recommended by the manufacturer. Mass photometry data was acquired with AcquireMP software and the analysis was performed with DiscoverMP software (Refeyn, UK). For each MP data set, histograms of individual mass measurements were inspected, and the mass distribution was fitted with Gaussian curves to estimate the average molar mass of the selected distributions.

## Sedimentation velocity analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments were performed in a ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter, Indianapolis, IN) by following the standard protocol (33). The macromolecules and their mixtures were loaded in 12- or 3-mm charcoal-filled Epon double-sector centerpieces with sapphire windows. The AUC cell assemblies filled with samples were inserted into An-50 or An-60 rotors, followed by temperature equilibration at 20°C in the AUC chamber at rest. After rotor acceleration, data were acquired with Rayleigh interference optics and absorbance optics at 260 nm and/or 280 nm, depending on the solution composition. Instrument calibration factors were determined as previously described (34). The SV-AUC data were analyzed using the standard sedimentation coefficient distribution  $c(s)$  model in the software SEDFIT version 16.35 (35) (<https://sedfitsedphat.nibib.nih.gov/software>).

Multi-signal analysis was based on the analysis of integrated  $c(s)$  peaks obtained in analyses of families of sedimentation profiles recorded with different absorbance and refractometric optical signals. Signal increments were determined from the signals in the sedimentation boundaries of free NA or protein species, respectively, quantified through the integration of the respective  $c(s)$  distribution peaks and their known concentration. Partial-specific volumes of complexes were calculated as weight-averages, based on a partial-specific volume of RNA of 0.59 ml/g in potassium salts (36). Partial-specific volumes of protein species, and calculations using Stokes and Svedberg equations were carried out using the calculator functions in SEDFIT.

## Dynamic light scattering

Autocorrelation functions (ACFs) of the samples were collected in either a NanoStar instrument (Wyatt Technology, Santa Barbara, CA) or Prometheus Panta (Nanotemper, Germany) instrument at 20°C. In NanoStar, 20  $\mu$ l samples were inserted into a 1  $\mu$ l quartz cuvette (WNQC01-00, Wyatt Instruments). Laser light scattering was measured at 658 nm at a detection angle of 90°. In the Prometheus Panta instrument, the samples were loaded into a capillary (Nanotemper PR-AC002) and ACFs were acquired using the 405 nm laser at the detection angle of 140°. The ACFs from both instruments were processed and analyzed in SEDFIT.

## Results

To dissect the assembly principles of RNPs we performed experiments with full-length N-protein carrying intact or mutationally abrogated binding interfaces in the LRS of the linker. Binding partners are viral RNA stem loops and model linear ssRNA and ssDNA oligos, with their contrasting behavior revealing important aspects of RNA conformational contributions to the assembly. The investigation takes the following path: (i) We first study the interplay between protein–protein oligomerization and scaffolding of N-protein in a model system of single-stranded NA. (ii) We similarly examine complex formation on stem–loop RNA. (iii) We establish that the larger RNPs formed with stem–loop RNA require the cooperative interactions from critical protein–protein interfaces and multivalent protein–NA interactions. (iv) We identify a second site for stem–loop RNA binding on the CTD dimer as an essential

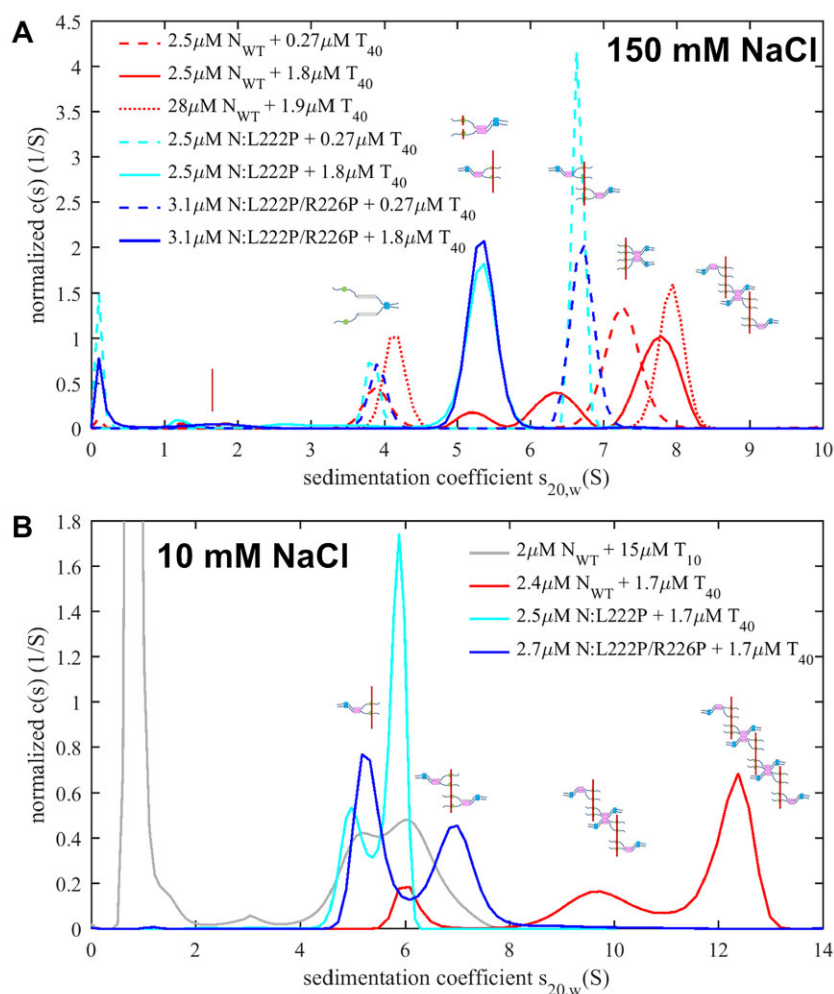
factor in the stabilization of RNPs. (v) Finally, we show that the key N-protein oligomerization interface identified *in vitro* is also essential in a viral-like particle assembly assay.

## LRS augments protein–NA complex formation in the presence of multivalent NA binding

In previous work we have studied N-protein dimer in mixtures with sufficiently short oligonucleotides ( $T_{10}$ ) to eliminate the possibility of multivalent binding. This allowed us to focus on the protein–protein interactions that arise for NA-liganded N-protein through allosteric stabilization of LRS helices and their cooperative coiled-coil oligomerization (26), as schematically indicated in Figure 1C. Occupation of the NA binding site (presumably in the NTD) causes a conformational change and improves the effective  $K_D^*$  for oligomerization by approximately three orders of magnitude from  $\approx 1$  mM to low  $\mu$ M. However, this is still significantly weaker than the N-protein affinity for NA. Therefore the question arises if the LRS interface can have a role in assembly when the N-protein dimer is more strongly and multivalently binding to NA. Possible scenarios for multivalent NA binding are sketched in Figures 1D and E, where either LRS oligomerization is not possible in N/NA complexes due to the strong NA binding interfaces constraining the complex structure (Figure 1D), or alternatively, complex architectures do permit LRS interactions between N-protein dimers while N-protein is scaffolded on NA (Figure 1E, F). The distinction is crucial for possible RNP assembly pathways.

To shed light on this question we studied N-protein binding to  $T_{40}$  oligonucleotides, which are twice the length required for cross-linking N-protein (37), and four times the size of the NTD binding site (38). Even though ssDNA oligos are not native N-protein ligands, their use at this stage of the study allows us to focus on steric aspects of multivalent binding without chemically altering the interactions previously studied with  $T_{10}$ . As we have shown previously (26) it is conveniently possible to reduce or eliminate LRS oligomerization through LRS mutations (L222P or L222P/R226P) that suppress helix formation, rendering N-protein capable only of presenting protein/NA interfaces (besides the obligatory high-affinity dimerization of the CTD). Thus, as a starting point to probe the binding capacity of  $T_{40}$  for N-protein, we carried out experiments in a concentration series with different molar ratios of mutant and wild-type (WT) N-protein binding to  $T_{40}$ . Sedimentation velocity analytical ultracentrifugation (SV-AUC) permits size-dependent hydrodynamic separation of complexes and unbound species, and can be recorded in multiple optical signals to determine the molar ratio of protein and NA in the sedimenting macromolecular complexes (39,40).

The resulting sedimentation coefficient distributions are shown in Figure 2A. In moderate ionic strength buffer  $B_{150Na}$  (20 mM HEPES, 150 mM NaCl, pH 7.4) in mixtures with ten-fold molar excess of either N:L222P or N:L222P/R226P over  $T_{40}$  we observe a  $\approx 6.7$  S peak (dashed blue and cyan curves, respectively). This is distinctly higher than the sedimentation coefficient of  $\approx 4$  S of N-protein alone, and  $\approx 1.7$  S of free  $T_{40}$ . It is also distinctly higher than complexes observed in analogous experiments with the shorter  $T_{10}$ , that when bound to the same mutant protein produces only  $\approx 5$  S species consisting of  $T_{10}$ -liganded N dimers (26). Unraveling the signal ratios of the 6.7 S N:L222P/ $T_{40}$  complex leads to a molar ratio of



**Figure 2.** Simultaneous LRS oligomerization and scaffolding on a linear oligonucleotide  $T_{40}$ . Shown are sedimentation coefficient distributions recorded at 260 nm for  $N_{\text{WT}}$  (red) and the LRS mutants N:L222P (cyan) and N:L222P/R226P (blue) at concentrations indicated. **(A)** Experiments in moderate ionic strength (PBS for  $28\mu\text{M } N_{\text{WT}} + 1.9\mu\text{M } T_{40}$ , all others buffer  $B_{150\text{Na}}$ ). **(B)** Complex formation at low ionic strength in buffer  $B_{10\text{Na}}$ . For comparison, data with  $T_{10}$  are reproduced from (26). In this and the following figures, near the distribution peaks are cartoons of examples of complexes (as in Figure 1) that would sediment approximately at similar  $s$ -value or mass, respectively (see Results).

$\approx 4$  N-proteins per  $T_{40}$ , suggesting a highly extended scaffolded configuration such as sketched in Figure 1D (with an  $s$ -value of 6.7 S implying a translational frictional ratio  $f/f_0 \approx 2$ ). At only threefold molar excess of these LRS mutants over  $T_{40}$  smaller complexes sedimenting at  $\approx 5.2$  S are observed (solid blue and cyan curves) with molar ratios protein/NA of  $\approx 2:1$ , consistent with lower saturation of NA sites with a single N-protein dimer bound per  $T_{40}$ .

By contrast, mixtures of  $N_{\text{WT}}$  with  $T_{40}$  at the same concentrations exhibit higher sedimentation coefficients throughout (Figure 2A, red curves). An increase in sedimentation coefficient with increasing total complex concentration can be discerned, as is characteristic in a ‘reaction boundary’, which signifies higher-order oligomerization that is rapidly reversible on the timescale of sedimentation ( $\approx 1000$  s). The reaction boundaries reflect the time-average sedimentation velocity of all co-sedimenting complex and free species, and therefore represent a lower limit for the  $s$ -value of the largest complex (39,41). We can attribute this complex formation to the additional LRS protein–protein interfaces in  $N_{\text{WT}}$  augmenting complex formation, as in configuration Figure 1F. In the mixtures of  $2.5\mu\text{M}$  (or  $28\mu\text{M}$ )  $N_{\text{WT}}$  with  $\approx 2\mu\text{M}$   $T_{40}$  a sedimentation boundary of  $\approx 7.8$  S is observed (Figure 2A, red solid line), as compared to the value of  $\approx 5.2$  S for the LRS mutants under the same conditions (blue and cyan solid lines). For reference, aside from complications in the interpretation of sedimentation velocities from reversible reacting systems (26), and potential conformational changes, such a 1.5-fold increase in sedimentation coefficient suggests the presence of majority species with at least twofold molecular weight. From the spectral data at the highest concentration, a complex molar ratio of 4:1  $N_{\text{WT}}/T_{40}$  can be measured, which may correspond for example to complexes sketched in Figure 1F, although the detailed configuration is undetermined by the current data and may be heterogeneous.

SARS-CoV-2 N-protein is basic and can therefore be expected to exhibit strong electrostatic contributions to the driving force of NA binding and a strong dependence on ionic strength (42). We therefore examined the same N-protein/ $T_{40}$  interaction next at low salt buffer  $B_{10\text{Na}}$  (10 mM NaCl, 10.1 mM  $\text{Na}_2\text{PO}_4$ , 1.8 mM  $\text{KH}_2\text{PO}_4$ , 2.7 mM KCl, pH 7.4) under otherwise identical conditions. Since previous studies quantifying the LRS oligomerization affinities were conducted with isolated LRS peptide (N:210–246) at 150 mM NaCl

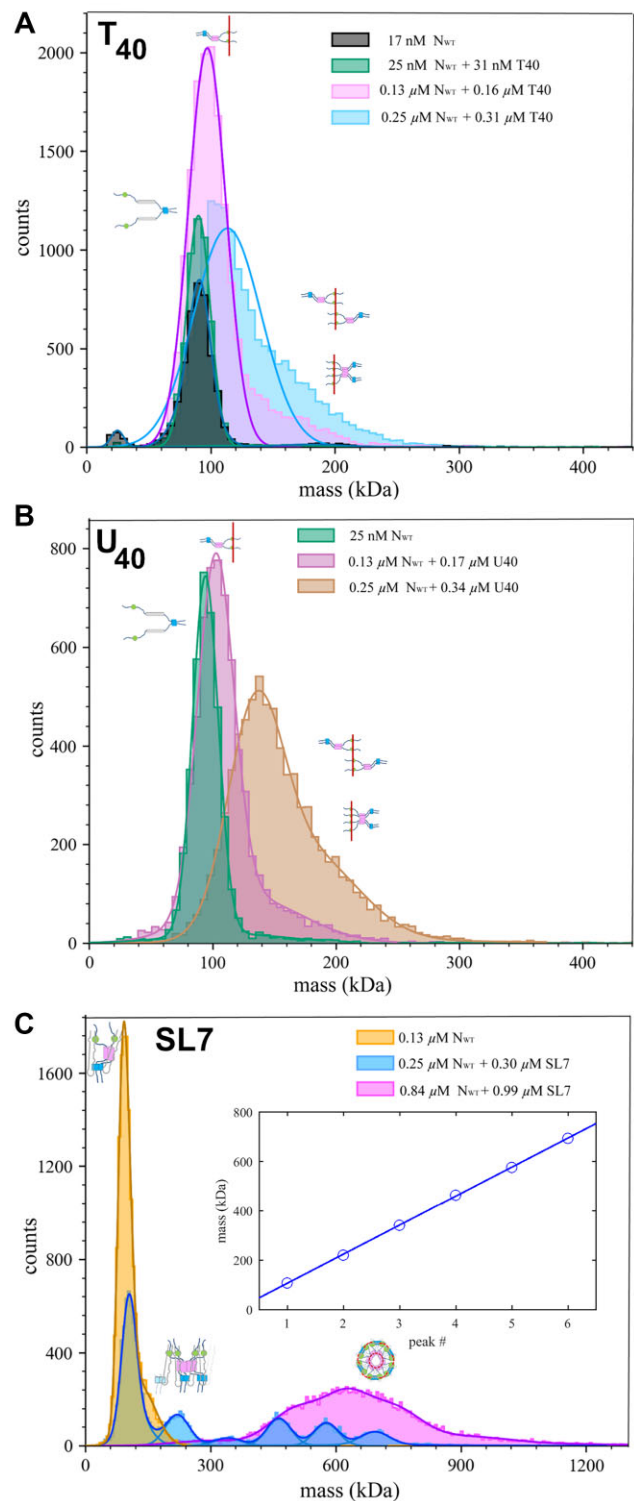
(26), we carried out preliminary SV-AUC experiments with the LRS peptide in  $B_{10Na}$ , and observed enhanced cooperative assembly of the LRS protein–protein interface in the lower ionic strength (Supplementary Figure S1). Thus, in the full-length protein both the protein–NA interfaces and the protein–protein interfaces strengthen at lower ionic strength.

We examined the impact of this on complex formation. For the LRS mutants with ablated LRS interfaces at a 3:2 molar ratio of protein/ $T_{40}$  the  $\approx 5$  S species observed previously at moderate ionic strength is at low ionic strength accompanied by majority  $\approx 6$  S peak or minority  $\approx 7$  S peaks for N:L222P and N:L222P/R226P, respectively, indicating some additional assembly (or greater  $T_{40}$  saturation) (Figure 2B, blue and cyan curves). Low salt conditions also lead to larger complex formation with  $N_{WT}$ –protein, but here exhibiting a significantly augmented effect: Under identical low salt conditions,  $N_{WT}$  forms complex species with  $T_{40}$  sedimenting at  $>12.5$  S (red curve). Based on the hydrodynamic scaling law  $s \sim M^{2/3}$ , the  $>2$ -fold increase in sedimentation coefficient introduced by the LRS in  $N_{WT}$  compared to N:L222P implies a majority species with approximately threefold higher molecular weight. A back-of-the-envelope estimate suggests moderately compact ( $f/f_0 = 1.3$ ) species with  $s$ -values of 12.5 S require complexes of  $\approx 300$  kDa, and species of higher molecular weights if complexes are in more extended conformations or in reaction boundaries. This demonstrates the potential for strong oligomerization of LRS with complexes of even higher assembly state than sketched in Figure 1E and F, involving at least three  $N_{WT}$  dimers. It should be noted that in these low salt conditions, higher protein and  $T_{40}$  concentrations (such as  $10 \mu M$  WT N-protein with  $5 \mu M$   $T_{40}$ ) promote the formation of macromolecular condensates (19,43).

SV-AUC data at intermediate ionic strength with majority potassium ions (buffer  $B_{65K}$  containing 64.8 mM KCl, 5.6 mM NaCl, 24.6 mM HEPES, pH 7.50) – a condition applied to the stem–loop RNP assembly assay described below – lead to results similar to those at the higher ionic strength  $B_{150Na}$ , reproducing the increased complex formation with  $T_{40}$  of  $N_{WT}$  as compared to both LRS mutants under otherwise identical conditions (Supplementary Figure S2A).

Finally, we can observe the molar mass distribution directly using mass photometry (MP). This technique is applicable to mixtures at nanomolar concentrations, and is therefore suitable to explore the onset of binding. As shown in Figure 3A, in moderate ionic strength (buffer  $B_{65K}$ ) at 25 nM  $N_{WT}$  with 31 nM  $T_{40}$ , the peak is indistinguishable from that of  $N_{WT}$  alone, which shows the expected dimeric state. Unbound  $T_{40}$  is below the minimum detectable mass in the MP instrument. At higher concentrations, increasingly the main peak shifts to slightly higher masses consistent with binding of  $T_{40}$ , and a secondary peak emerges with masses up to the  $\approx 200$  kDa range indicating onset of tetramerization at  $0.25 \mu M$   $N_{WT}$  with  $0.31 \mu M$   $T_{40}$ . Consistent with this, SV-AUC of this mixture shows a reaction boundary at  $\approx 6$  S, at tenfold higher concentrations growing to  $\approx 8$  S (Supplementary Figure S2A).

An analogous MP experiment was carried out with ssRNA  $U_{40}$  instead of  $T_{40}$ , in order to probe whether the different sugar component in the structurally similar oligonucleotide may lead to different oligomeric states of complexes with N-protein. However, as shown in Figure 3B, this is not the case, although the equilibrium is shifted slightly more to the assembly state, indicating a somewhat higher affinity for  $U_{40}$ . This is corroborated in SV experiments at higher concentrations,



**Figure 3.** Complex mass distributions of  $N_{WT}$  with  $T_{40}$ ,  $U_{40}$  and SL7 in mass photometry. Shown are mass histograms of  $N_{WT}$  with  $T_{40}$  (A),  $U_{40}$  (B) and SL7 (C). The inset in (C) shows peak mass values vs peak number of the  $0.25 \mu M$   $N_{WT}$  with  $0.3 \mu M$  SL7 mixture and linear fit leading to a mass increment of 118 kDa.

where qualitatively similar complex patterns for  $U_{40}$  are observed as compared to  $T_{40}$  (Supplementary Figure S2B).

### N-protein forms large RNP complexes with stem-loop RNA

Recently, the Morgan laboratory has examined binding of N-protein to viral genome segments of different length from the 5' end of the viral genome, and observed the formation of  $\approx 700$ – $800$  kDa complexes with 600 nt RNA consistent in size with vRNP particles observed in cryo-ET, and consistent in the average length of RNA that would need to be packaged by each vRNP (20,25). Similar sized protein/RNA complexes were observed for N-protein binding to stem-loops SL4a, SL7 and SL8, which were unstable at submicromolar concentrations unless chemically crosslinked (25). In the present work, we focused on non-covalent binding of N-protein to stem-loop RNA SL7 (46 nt, 14.9 kDa) to compare it with the similar-sized oligonucleotides  $T_{40}$  and  $U_{40}$ .

Under the same conditions as shown in Figure 3A/B for  $T_{40}$  and  $U_{40}$ , respectively, a mixture of  $0.25 \mu\text{M}$   $N_{\text{WT}}$  with  $0.3 \mu\text{M}$  SL7 exhibits a distinct ladder of species with peak molar masses of  $108 \pm 24$ ,  $221 \pm 29$ ,  $342 \pm 33$ ,  $464 \pm 29$ ,  $577 \pm 29$ ,  $695 \pm 90$  kDa (Figure 3C, blue histogram), analogous to the results reported by Carlson *et al.* for SL8 (25). The best-fit increment is 118 kDa, which is closest to the theoretical molar mass of 121 kDa of an  $N_{\text{WT}}$  dimer in complex with two SL7. The first peak appears at a slightly lower mass, which may indicate a mixture of unresolved liganded and free  $N_{\text{WT}}$  dimer (free SL7 is below the mass limit of detection). Thus, the MP data are most consistent with an oligomerization of up to six  $\{N_{\text{WT}}$  dimer/2SL7 $\}$  subunits. However, we attribute an estimated  $\approx 5$ – $10\%$  uncertainty to the molar mass values of MP due to differences in calibration, and differences in mass-to-contrast ratios for nucleic acids compared to proteins (44). Solvent-dependent variation may be expected due to the strong counter-ion dependence of the refractive index of nucleic acids (45). Therefore, complexes of  $N_{\text{WT}}$  dimer with one or three SL7 cannot be ruled out from the MP data.

As mentioned above, MP is limited to nanomolar macromolecular concentrations to allow discrimination of individual adsorption events. When the limit is approached at  $0.84 \mu\text{M}$   $N_{\text{WT}}$  +  $0.99 \mu\text{M}$  SL7, as shown in Figure 3C, a single broad peak is observed ranging from 400 kDa to  $>900$  kDa, suggesting the possibility of higher complexes at higher concentrations. Therefore, to study whether the complex formation with stem-loop RNA is an unlimited self-assembly or leads to a well-defined complex we turn back to hydrodynamic techniques.

Figure 4A shows a concentration series of  $N_{\text{WT}}$  and SL7 at constant molar ratio in SV-AUC. At low concentrations of  $0.25 \mu\text{M}$   $N_{\text{WT}}$  with  $0.3 \mu\text{M}$  SL7 a broad distribution of species in the range of 7–18 S can be discerned (black line). This is qualitatively consistent with the ladder of complex species observed in MP under identical conditions (Figure 3C), although consideration is required of the fact that MP reports number distributions whereas SV-AUC produces weight-based distributions, i.e. the latter causes larger signals for larger particles, and oligomers cannot be well resolved. The  $\approx 18$  S peak confirms the strongly enhanced assembly with SL7 compared to  $T_{40}$  or  $U_{40}$ . In Figure 4A, the  $0.25 \mu\text{M}$   $N_{\text{WT}}$  +  $0.3 \mu\text{M}$  SL7 sample was created both by mixing of separate stocks to achieve the final concentration, and alternatively by 90 min incubation

of a 10-fold concentrated stock mixture followed by dilution. Within error the sedimentation coefficient distributions are indistinguishable, demonstrating they reflect equilibrium between species (after further 90 min incubation prior to the start of centrifugation).

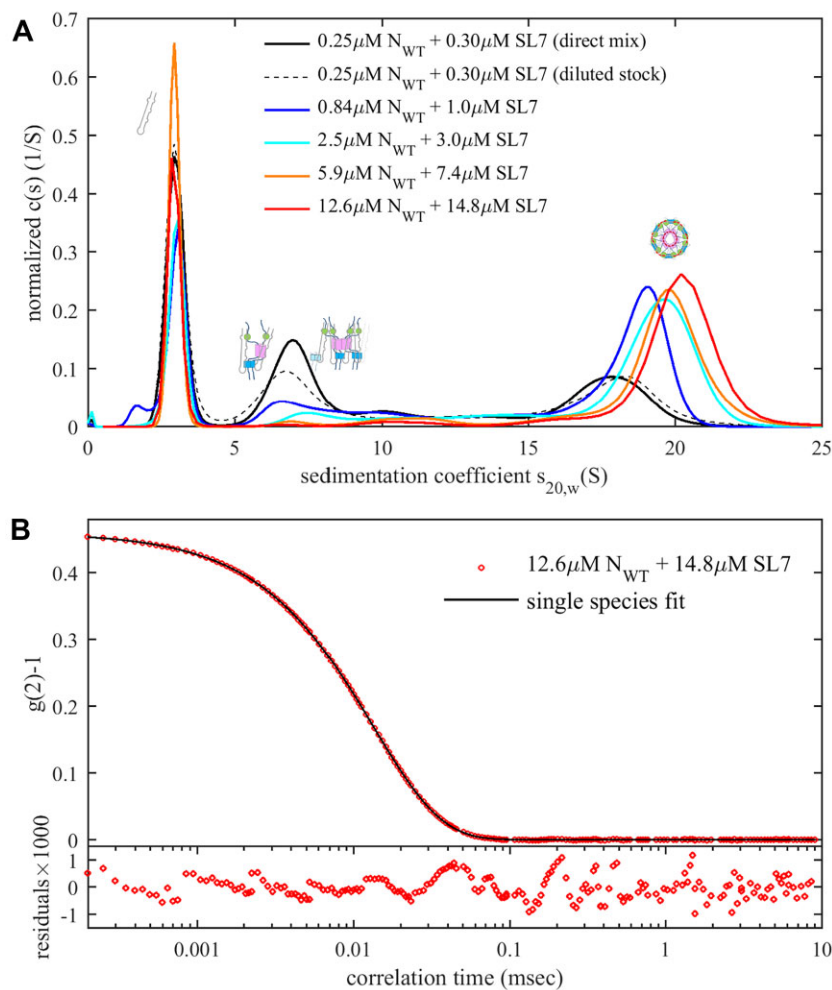
Increasing concentrations up to 50-fold ( $12.6 \mu\text{M}$   $N_{\text{WT}}$  with  $14.8 \mu\text{M}$  SL7) causes the largest species to asymptotically approach a sedimentation coefficient of  $\approx 20$  S. Multi-signal analysis of the fastest sedimenting boundary leads to a molar ratio of  $(0.95 \pm 0.1)$  SL7/ $N_{\text{WT}}$ , consistent with a model of self-assembly of  $\{N_{\text{WT}}$  dimer/2SL7 $\}$  subunits. In order to determine the mass of the largest RNP complex, we carried out DLS measurements. Even though this technique does not resolve species of similar size, the scattering intensity is weighted strongly in favor of the largest component. As shown in Figure 4B, the autocorrelation function can be fit well with a single species model with a diffusion coefficient of 2.45 F, corresponding to a Stokes radius of 8.5 nm. The diffusion coefficient can be combined with the peak sedimentation coefficient in the Svedberg equation  $M(1 - \bar{v}\rho) = RT(s/D)$  to produce a buoyant molar mass  $M(1 - \bar{v}\rho)$  of  $\approx 200$  kDa. With a 1:1 molar ratio we estimate the complex partial-specific volume to be 0.695 ml/g, leading to a complex molar mass of  $\approx 650$  kDa. This estimate is slightly below the value from MP of  $\approx 695$  kDa, as well as the theoretical value of 726 kDa for a hexamer  $\{N_{\text{WT}}$  dimer/2SL7 $\}_6$ , which may be caused by an underestimate of the largest species  $s$ -value from inspection of the reaction boundaries with incomplete saturation (39) and an overestimate of the diffusion coefficient due to the unaccounted scattering contributions of free SL7. Nevertheless, it clearly shows the finite assembly of  $N_{\text{WT}}$  with SL7 to species not exceeding a hexamer of  $\{N_{\text{WT}}$  dimer/2SL7 $\}$  subunits. Finally, from the molar mass and sedimentation coefficient we can calculate the frictional ratio and arrive at a value of 1.5; this indicates the complex is significantly more compact than the free N-protein which has a frictional ratio of 1.8 (37).

### The formation of large RNP complexes depends on LRS oligomerization and multivalent NA binding

Analogous to our studies of N-protein binding to  $T_{40}$  and  $U_{40}$  above (Figure 2 and Supplementary Figure S2), we next studied how the LRS protein-protein interfaces contribute to the formation of RNP complexes by examining the reaction products of SL7 with LRS point mutants N:L222P and N:L222P/R226P. First, Figure 5A shows the mass distribution obtained at low concentrations suitable for MP experiments ( $0.25 \mu\text{M}$  N-protein with  $0.3 \mu\text{M}$  SL7). While for  $N_{\text{WT}}$  in this mixture a ladder of complex species up to  $\approx 700$  kDa can be discerned, very similar to the experiment shown in Figure 3C (of which it is a replicate), both LRS mutants show a single peak consistent with N-protein dimer.

To study these mixtures at tenfold higher concentrations we turn to SV-AUC (Figure 5B), which for  $N_{\text{WT}}$  shows the familiar  $\approx 20$  S peak. By contrast, both LRS mutants only show drastically slower  $\approx 6.5$  S reaction boundaries. DLS data acquired for the same samples leads to hydrodynamic radii of 4.8 nm and 5.0 nm for the single and double mutant, respectively (Supplementary Figure S3). The combination of sedimentation and diffusion coefficients yields molar mass estimates of 350 kDa for SL7 complexes with N:L222P, and 370 kDa for N:L222P/R226P. Thus, at low micromolar concentrations we can observe only complexes with 3–4 N-protein





**Figure 4.** Concentration-dependent RNP assembly of  $N_{WT}$  with stem-loop SL7. **(A)** Sedimentation coefficient distributions from SV-AUC experiments recorded at 260 nm. Mixtures of  $N_{WT}$  and SL7 at concentrations indicated in  $B_{65K}$  (or  $B_{75Na}$  for the two most concentrated mixtures). **(B)** Autocorrelation data from DLS of the highest concentration mixture. The best-fit single-species model leads to a diffusion coefficient of  $2.466 \times 10^{-7} \text{ cm}^2/\text{s}$  or a Stokes radius of 8.5 nm.

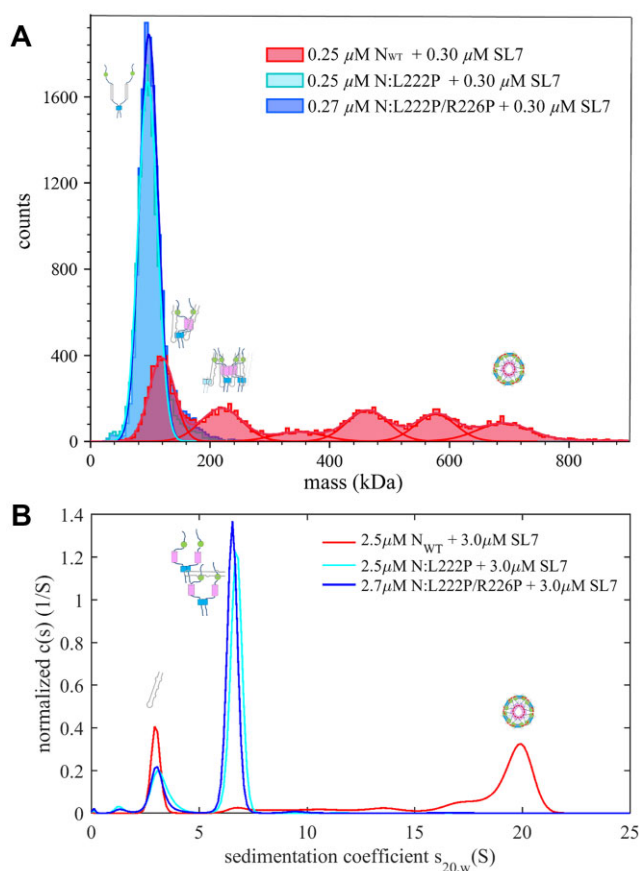
dimers bound to SL7, in contrast to the significantly larger and more compact 20 S complex with  $N_{WT}$  carrying intact LRS protein-protein interfaces.

It is useful to recapitulate the roles of the different assembly interfaces. The experiments contrasting LRS point mutants *vs*  $N_{WT}$  demonstrate the key role of LRS oligomerization in the assembly of large RNPs. However, clearly LRS oligomerization alone is not sufficient for the formation of large RNPs: When LRS oligomerization is switched on through binding of only short  $T_{10}$  NA ligands that are unable to scaffold N-protein, only  $\approx 6$  S complexes are formed at low micromolar concentrations, consistent with dimerization of dimers (Figure 2B, grey line) (26). This underlines the importance of simultaneous multivalent scaffolding for RNP formation. Even without LRS interactions, scaffolding of N-protein on NA can lead to intermediate size complexes at micromolar concentrations, as demonstrated in Figure 5B (blue and cyan lines) for binding to SL7 and in Figure 2 (blue and cyan lines) for binding to  $T_{40}$ . However, these complexes are not very stable and dissociate at nanomolar concentrations, as shown in Figure 5A (blue and cyan histograms). We can conclude that both multivalent scaffolding and LRS oligomerization is necessary for large RNPs to form.

This can be observed clearly at nanomolar concentrations in MP experiments, where only simultaneous interactions of NA-mediated scaffolding and LRS-mediated protein-protein oligomerization jointly provide sufficient stability for the oligomers of  $\{N \text{ dimer}/2SL7\}$  subunits to populate (Figure 5A, red). This can occur only if NA-mediated scaffolding acts to help cross-link  $\{N \text{ dimer}/2SL7\}$  subunits, i.e. through shared interactions of different N dimers with the same SL7, or potentially through the same N-protein domain offering binding sites for multiple SL7 (as sketched in Figure 1H/I). Such additional  $\{N \text{ dimer}/2SL7\}$  subunit interactions *via* the NA in concert with LRS interactions create the strong enhancement typical for avidity-based multivalent interactions (in a simplistic picture through added binding energies). The same avidity-based enhancement of complex stability supports the formation of large RNP complexes at low micromolar concentrations, once scaffolding and LRS interactions act jointly.

#### RNP complexes are stabilized through multivalent stem-loop RNA binding of the CTD

We next focus on the question of what supports the formation of large RNPs with SL7 as NA substrate but not the similar



**Figure 5.** RNP formation of N-protein with SL7 depends on LRS oligomerization *in vitro* and in viral assembly. **(A)** Mass distributions from MP of mixtures of 0.25  $\mu\text{M}$  N-protein with 0.3  $\mu\text{M}$  SL7 in moderate ionic strength buffer  $B_{65K}$  for  $N_{WT}$  and the mutants inhibiting LRS oligomerization N:L222P and N:L222P/R226P. **(B)** Sedimentation coefficient distributions from SV-AUC experiments at 2.5  $\mu\text{M}$  N-protein with 3.0  $\mu\text{M}$  SL7 in buffer  $B_{65K}$  recorded at 260 nm. Shown are results with  $N_{WT}$ , N:L222P and N:L222P/R226P.

sized  $T_{40}$  or  $U_{40}$ . As shown in experiments with LRS mutants, both single-stranded oligonucleotides and stem-loop RNA allow scaffolding of multiple N-protein dimers. The comparison of LRS mutants with  $N_{WT}$  shows scaffolding can act in concert with LRS oligomerization. Nevertheless, RNP complexes with SL7 are roughly twice the size than those with  $T_{40}$  and  $U_{40}$ . (A side-by-side comparison of  $T_{40}$ ,  $U_{40}$ , and SL7 complexes with  $N_{WT}$  and LRS mutants at the same concentration and under the experimental conditions of the RNP assay is provided in Supplementary Figure S4.)

Notwithstanding supporting roles of N-protein IDRs, the major nucleic acid binding sites are located in the folded NTD and CTD (23,46,47). Accordingly, we prepared constructs of isolated NTD and CTD domains to study their interactions with single-stranded oligonucleotides and stem-loop RNA. By itself, NTD is monomeric and CTD dimeric due to its high-affinity dimerization interface; and neither shows detectable further self-association in the micromolar concentration range considered here, consistent with expectations from the literature (46,48,49). By applying low ionic strength conditions ( $B_{10Na}$ ) to drive the reaction into the maximal assembly state, we measured the size and composition of the CTD/NA and

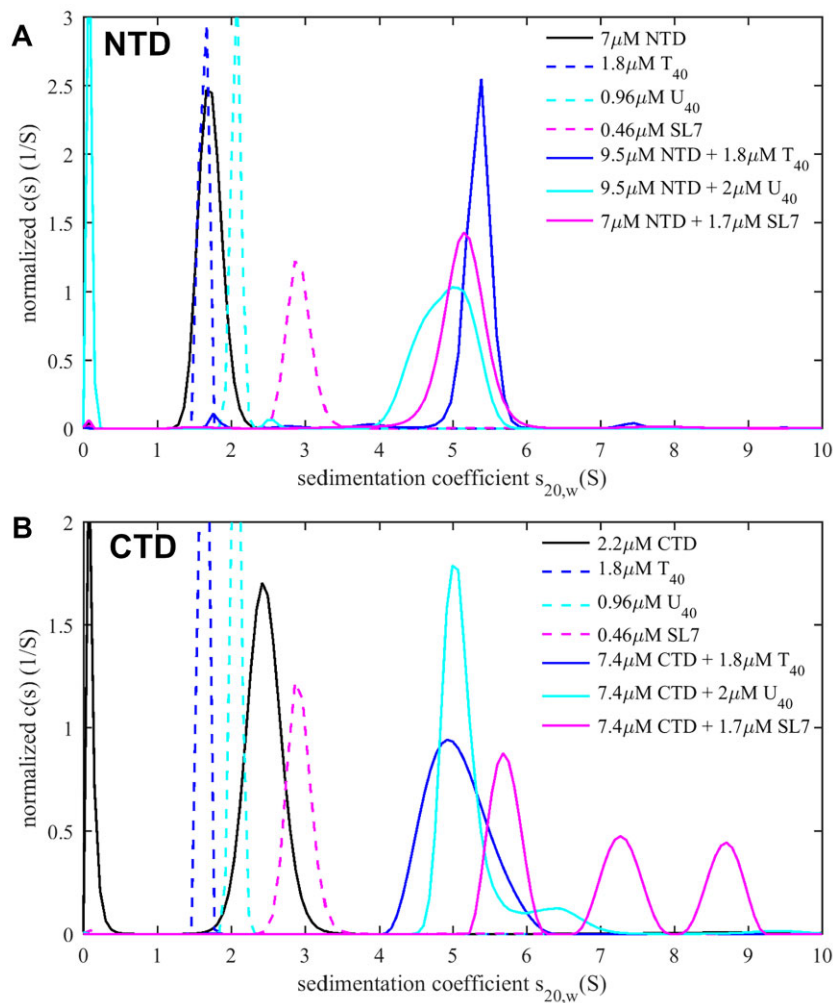
NTD/NA complexes by SV-AUC and obtain information on the protein footprint.

Figure 6A shows sedimentation coefficient distributions of  $\approx 10 \mu\text{M}$  NTD in 5-fold molar excess over  $T_{40}$ ,  $U_{40}$  or SL7, respectively. For all oligonucleotides, a  $\approx 5.0$ – $5.5$  S peak is observed with no or little free protein. The complex sediments much faster than either free NA or free NTD. Multi-signal analysis leads to an estimated protein/NA molar ratio of 5.4:1 on  $T_{40}$  and 4:1 – 5:1 on SL7. For  $T_{40}$ , allowing for end effects, we can conclude a maximum for the footprint of  $\approx 10$  bases.

Analogous experiments for the CTD dimer are shown in Figure 6B. CTD in 5-fold molar excess over  $T_{40}$  in low ionic strength conditions produces similar sized CTD/ $T_{40}$  complexes compared to NTD/ $T_{40}$ . By contrast, mixtures of CTD with SL7, under otherwise identical conditions, become cloudy upon mixing, and complexes of CTD/SL7 sediment significantly faster, up to  $\approx 8.8$  S (magenta). Based on the hydrodynamic scaling law  $s \sim M^{2/3}$  one would expect the largest complex to have approximately twice the molar mass of the CTD/ $T_{40}$  complexes. Multi-signal analysis of all CTD/SL7 complexes results in an average molar ratio of  $\approx 6$  ( $\pm 0.5$ ):1 CTD per SL7, or 3 CTD dimers per SL7. This composition was used to determine the complex partial specific volume, and to determine *via* the Stokes equation the minimum mass of a hydrated spherical complex that can sediment as fast as 8.8 S, which is 140.8 kDa, or  $\approx 181$  kDa for more realistically assuming a moderately compact particle with frictional coefficient 1.3. This clearly exceeds the smallest possible complex with 6:1 stoichiometry (94.7 kDa for 3 CTD dimers + 1 SL7), but it would be consistent with a moderately asymmetric 12:2 complex (i.e. 6 CTD dimers binding 2 SL7). Since SL7 does not oligomerize by itself, and the CTD does not oligomerize beyond the dimer, we can conclude CTD dimers must exhibit at least a weak second site for SL7, to be able to accommodate two SL7 molecules as required for the large CTD/SL7 complexes observed. Mutual multi-valency of the CTD dimer and SL7 is consistent with the observed polydispersity of complexes as well as the aggregation propensity of these mixtures.

Finally, to distinguish whether the observed different binding behavior of CTD for stem-loop RNA and the ssDNA oligomer arises from the different sugar moiety of the NA or its secondary structure in the stem-loop, we carried out analogous experiments with ssRNA  $U_{40}$ . As shown in Figure 6B (cyan), similar complexes are formed for  $U_{40}$  as for  $T_{40}$ , lacking the faster-sedimenting, higher-order complexes seen with the stem-loop RNA. This is consistent with the inability of  $U_{40}$  to form RNP-sized complexes (Figure 3B and Supplementary Figure S4). These results suggest the second NA binding site on the CTD dimer recognizes stem-loop secondary structure, and that N-protein may not differentiate between RNA and DNA substrates.

In conclusion, translated into the context of the FL N-protein, it appears the additional interactions of CTD dimer with SL7 on its second site act to weakly cross-link {N dimer/2SL7} subunits and thereby stabilize the hexameric complexes of these subunits in cooperation with scaffolding and LRS protein-protein interactions (Figure 1H/I). When binding  $T_{40}$  or  $U_{40}$ , these additional interactions are absent and only smaller RNP complexes can form in the micromolar concentration range tested.



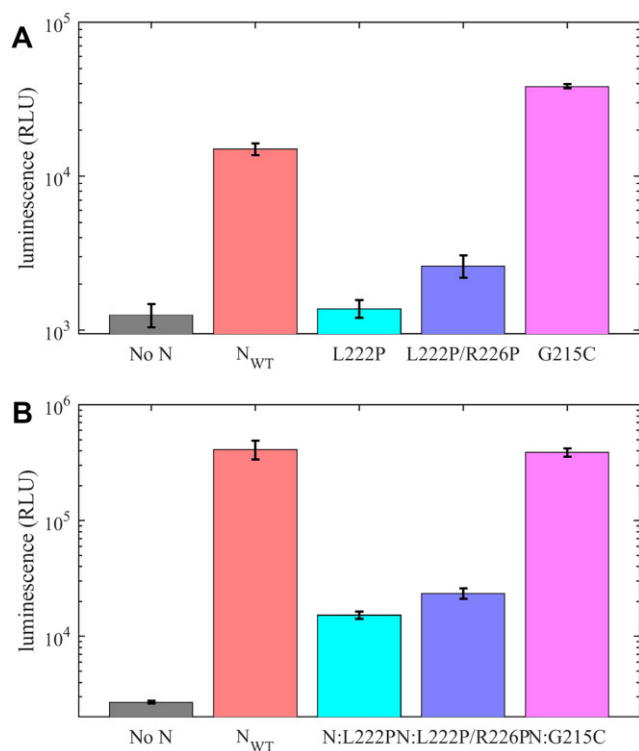
**Figure 6.** NTD and CTD binding to T<sub>40</sub>, U<sub>40</sub> and SL7. Sedimentation coefficient distributions  $c(s)$  from SV-AUC of NTD (**A**) or CTD (**B**) alone and in mixtures with T<sub>40</sub>, U<sub>40</sub> or SL7, respectively. Mixture experiments are carried out in low ionic strength buffer B<sub>10Na</sub> to promote formation of the complexes with maximum stoichiometry. Distributions of free NA are reduced by a factor 2. For NTD alone, the molecular weight determined from  $c(s)$  analysis and best-fit frictional ratio is 15.1 kDa, which compares well to the theoretically expected value of 15.2 kDa. For CTD alone, the experimental molecular weight is 27.3 kDa, which compares to the theoretical value of 26.6 kDa for a CTD dimer.

### The assembly of virus-like particles depends on the oligomerization in the LRS

To probe whether the involvement of LRS in the formation of RNPs *in vitro* extends to viral assembly *in vivo* we employed a recently described SARS-CoV-2 virus-like particle (VLP) assay (31). As shown previously, co-expression of the virus structural proteins S, E, M and N in HEK 293T cells leads to the assembly of VLPs when initiated by a suitable NA packaging signal. The latter is achieved through co-transfection with a plasmid containing a 2 kb segment of the viral RNA presenting the packaging signal T20. The packaged RNA also includes a luciferase reporter transcript. VLPs from the supernatant of these producer cells are then used to infect 293T receiver cells expressing SARS-CoV-2 entry factors ACE2 and TMPRSS2. This allows entry of the VLPs into these receiver cells, and after expression of the cargo luciferase reporter in the receiver cells their luminescence is measured to quantitate both VLP assembly in the producer cells and viral entry into the receiver cells. Thus, introduction of mutant N-protein in this assay allows to test their impact on viral packaging and assembly efficiency (31).

Figure 7A shows a strong reduction of luminescence, relative to N<sub>WT</sub>, for the point mutants N:L222P and N:L222P/L226P, demonstrating that selectively inhibiting LRS interactions strongly impairs VLP formation. To test the opposite effect, we introduced the mutation N:G215C that was previously shown to enhance LRS interactions by stabilizing the helical state in the linker IDR (27). In the VLP assay, significantly elevated luminescence was observed for this enhancer mutation of LRS interactions (Figure 7A). This is consistent with the hypothesis that the protein-protein contacts from transient helices in the LRS and their coiled-coil oligomerization are an essential component of the assembly mechanism *in vivo*.

This experiment can be repeated under slightly different conditions. Intracellular N-protein is strongly phosphorylated, in contrast to N-protein in the virions (50–53). Accordingly, it has been shown that inhibition of phosphorylation enhances the efficiency of viral packaging in the present VLP assay (32,50). This can be achieved by treatment of the VLP producer cells with an inhibitor of GSK3 (CHIR98014), for which the SR-rich region of the linker provides abundant sites



**Figure 7.** LRS oligomerization is critical for viral assembly in a VLP assay. The formation of VLPs in 293T cells containing SARS-CoV-2 structural proteins (S, E, M and N) and packaging RNA is detected through expression of a luciferase reporter in infected receiver cells. The efficiency of VLP assembly was measured for different (or lacking) N-protein species indicated and quantified in relative luminescence units. Experiments were conducted with normal intracellular phosphorylation (**A**) and in the presence of the GSK3 inhibitor CHIR98014 (1.25  $\mu$ M) inhibiting N-protein phosphorylation (**B**). The error bars are standard deviations of six experiments.

(15,51–53). Under these conditions of suppressed phosphorylation and enhanced assembly propensity, introduction of the LRS mutants again led to a significant reduction of VLP assembly, consistent with the critical role of helix oligomerization in the LRS (Figure 7B).

Comparing the VLP assembly efficiency in the presence and absence of the phosphorylation inhibitor, it is noteworthy that assembly reached a higher level in the presence of phosphorylation inhibitor even for LRS mutants as compared to the wild-type without the inhibitor. Thus, it appears the structural destabilization of the LRS helices can be partially compensated in cells by increased intracellular levels of assembly-competent unphosphorylated N-protein, reminiscent of expectations from mass action law. Furthermore, it is interesting that the double mutation L222P/R226P is slightly less efficiently inhibiting VLP assembly than the single point mutation L222P. (Similarly, close inspection of the complex formation of N-protein mutants in the presence of T<sub>40</sub> in Figure 2B also suggests slightly stronger residual oligomerization of L222P/R226P compared to L222P.) These results are in contrast to the structural predictions and previous *in vitro* biophysical characterization of the LRS oligomerization potential in the presence and absence of short oligos T<sub>10</sub> (26), where the double mutant was slightly more effective in abrogating oligomerization than the single mutant. We speculate that the different relative behavior may be due to the scaffolding on

NA and/or the higher-order oligomerization in RNP assemblies imposing slightly different LRS interface structures than those of the smaller oligomers studied previously. This points to the pleomorphic nature of the interface and limitations of the point mutants as models, as they apparently do not completely abolish LRS oligomerization, but merely substantially destabilize this protein–protein interface.

## Discussion

The assembly pathway of SARS-CoV-2 N-protein to package and condense the large gRNA has remained enigmatic, despite advances such as cryo-ET density maps of distinct vRNP particles (3,4), increasing atomistic insight in the structural and dynamic aspects of NA binding modes of the folded NTD and CTD domains (16,23,38,49,54,55), and a recently described protein–protein interface in the disordered linker (26). In part this is due to the formidable problem of compounded effects of highly multivalent protein–NA and protein–protein interactions taking place side-by-side, the limited specificity of NA sites, high flexibility of N-protein imparted by the IDRs (56), and the resulting polymorphic nature of the complexes. Recent elegant work from the Morgan laboratory has suggested a model of six N-protein dimers organizing  $\approx$ 600 bases of RNA in RNPs (25). In the present study, limited complex formation *in vitro* between model NAs and purified WT N-protein or point-mutants, along with complementary VLP assembly experiments, revealed further architectural principles of RNPs that contribute to a more complete picture of assembly.

It is likely that simultaneous binding of N-protein dimers to the 5' untranslated region of the gRNA, *via* NTD likely preferring unpaired single stranded regions and CTD preferring double stranded RNA, creates locally a high density of N-protein in close proximity (16,18,20,57,58). As the NA-bound N-protein NTD induces a conformational change in the LRS to a self-assembly-competent helical state (26), we propose the cooperative oligomerization of LRS helices ‘rolls up’ N-protein dimer complexes into hexameric assemblies with LRS oligomers on the inside, similar to the sketch Figure 1I. Potentially this may occur as gRNA exists the DMV pores. There appears to be a role for macromolecular condensates to help initiate and/or propagate vRNP formation by ensuring high local N-protein concentrations (18,20–22); and plausibly a role for the same ultra-weak attractive interactions that can lead to macromolecular condensates to also help holding semi-liquid high-density states together (25).

Our *in vitro* data on the assembly properties of LRS point mutant N-proteins clearly demonstrate LRS assembly is a key mechanism, without which only small complexes can form even at high concentrations, resulting from simple scaffolding of N-protein on the NA. The protein–protein interface in the LRS allows such NA-bound N-protein dimers to further self-assemble into three-dimensional structures. This is consistent with optical tweezer experiments by Morse et al. examining multiple copies of N-protein binding to a single molecule of ssDNA, which show a cooperative reorganization of an initially formed protein/DNA complex (59) that we would attribute to LRS oligomerization after N-protein scaffolding on NA.

In support that LRS oligomerization is essential also for assembly *in vivo*, we carried out VLP experiments that demonstrate strongly reduced or enhanced VLP formation as a result of N-protein point mutants that diminish or enhance

coiled-coil oligomerization of LRS helices, respectively. Even though the packaged NA in the VLPs is different from the longer viral genome, and is different from the stem-loop based *in vitro* model system for RNP assembly, the LRS assumes a critical role in each assembly model. Several lines of evidence suggest this conclusion also extends to live SARS-CoV-2 virus: While we have shown previously that LRS oligomerization occurs similarly in N-protein of related coronaviruses, SARS-CoV-2 Delta variant clades containing the G215C mutation that enhances LRS oligomerization dramatically outperformed those without G215C (27). Conversely, mutations of the LRS that destabilize the oligomerization of LRS helices are excluded from the exhaustively sampled mutational landscape of viable SARS-CoV-2 (26).

Interestingly, the formation of large RNPs in the *in vitro* model examined here depends, besides the LRS, also on multivalent binding of stem-loop RNA by the CTD. In experiments focused on the binding of the isolated CTD domain to either single-stranded oligonucleotides or similar-sized stem-loop RNA SL7, the major difference was the assembly of significantly larger complexes with SL7, which are of a size and composition that require the ability of a CTD dimer to bind two SL7 molecules. While a second binding site for RNA on the CTD dimer is unexpected, additional support for at least a weak second RNA site on the CTD dimer comes from our observed turbidity upon mixing of CTD with SL7 (which depends on mutually multivalent interactions for either aggregation or LLPS), and, similarly, the observation reported by Dang & Song of aggregation at high concentration in the presence of stem-loop S2m (47). It would also be consistent with multi-step binding of isolated CTD to DNA observed in optical tweezer experiments by Morse *et al.* (59), and conclusions from structural modeling by Padroni *et al.* (54). In the present RNP assembly assay with FL-N, this appears to be the key difference between the different NA substrates, where, similar to the CTD results, binding of either ssDNA oligomer T<sub>40</sub> or ssRNA oligomer U<sub>40</sub> leads to only moderately sized complexes, whereas the stem-loop RNA SL7 supports significantly larger RNP assemblies. Mechanistically, since the second SL7 binding site on the CTD dimer acts in concert with LRS oligomerization, the resulting avidity can amplify even a weak second SL7 site to provide significant complex stabilization. We speculate that in the interaction with gRNA the ability to accommodate a second RNA interaction at the CTD dimer may similarly stabilize compaction and scaffolding of stem loops on N-protein dimers which would then enhance LRS oligomerization and vRNP formation. An attractive feature of this hypothesis is that the total binding energy of the vRNP is divided between LRS and CTD interfaces, such that uncoating upon viral entry may require only relatively minor destabilization of one of the interactions.

An important aspect of the assembly model is the existence of an upper limit of size of the RNP complexes to form discrete particles. Combining SV-AUC and DLS at a range of micromolar concentrations, we found this limit to be at a hexamer of N-protein dimers, consistent with results from Carlson *et al.* by MP in dilute solution after chemical crosslinking (25). From cryo-ET data, vRNPs are heterogeneous, but a subclass appear to be hexagonally packed (4). The previously observed pleomorphic ability of LRS helices to form different oligomeric states (26) may contribute to the heterogeneity of particles. Unfortunately it is difficult to reference assembly models in detail to the electron microscopy density maps

due to the large fractions of IDRs of N-protein that are highly flexible. For example, end-to-end distances of the linker IDR alone may range from 2 to 12 nm (56). The resulting variation of the local conformation may cause lack of contrast in maps averaged across vRNP particles. Besides the configuration of the N-protein scaffold, it is also unclear in which way gRNA is packaged onto the vRNPs and contributes to its structure. Various secondary structural features involve the majority of gRNA (60), and N-protein flexibility may contribute structural plasticity to accommodate different RNA structural elements. On the other hand, ssRNA is also highly flexible with a persistence length of  $\approx 2$  nm (61), and RNA 3D structures may not be unique (62), so that it is also conceivable that RNA structures may be dynamically reshaped by N-protein interactions (63,64).

As pointed out by Carlson *et al.*, if the size of the genome is divided by the number of vRNPs, on average approximately 800 nt would be scaffolded by each vRNP (25), or somewhat less accounting for stretches connecting neighboring particles. This could be achieved if a hexamer of N-protein dimers would on average bind 60–65 nt per N-protein, which is not dissimilar to the RNP complexes obtained in the SL7 stem-loop model system in the present work, or the SL8 and other stem-loop model examined by Carlson *et al.* (25). Thus assembly of N-protein dimer/RNA stem-loop configurations as depicted in Figure 1H/I could theoretically roughly satisfy scaffolding requirements of vRNPs. Furthermore, for the RNP particles with SL7 we measured a Stokes radius of 8.5 nm, which seems reasonably consistent with the 14–16 nm sized structures observed in cryo-ET (3,4).

The pleomorphic ability of vRNPs may extend to its architectural principles. Adly *et al.* and Syed *et al.* have recently shown by MP, EM, and in the VLP assay that truncated N-protein N: $\Delta$ (1-209) can still form RNPs similar in size and shape to FL-N (32,50). Remarkably, this is despite lacking the NTD (in addition to the N-arm IDR and the SR-rich region of the linker IDR), which is thought to confer specificity of the assembly (16,57). However, the LRS region was critical for assembly of N: $\Delta$ (1-209) (32), consistent with our model of LRS oligomerization as a central mechanism for vRNP formation. N: $\Delta$ (1-209) arises as an alternate viral transcript due to the introduction of a new transcription regulatory sequence associated with the R203K/G204R mutation in Alpha, Gamma, and Omicron variants. Interestingly, a significant increase in the number of RNPs was found by cryo-ET in Omicron variants (65), suggesting altered RNA packaging. However, other factors may also contribute to this, such as new protein-protein interactions between Omicron N-arm IDRs arising from the obligatory mutation P13L of Omicron variants (43) and possibly epistatic interactions with other N-protein mutations that coexist in variants of concern. In light of the pleomorphic architectures of vRNP, it appears that the interfaces in LRS oligomerization along with the NA binding sites on the CTD would remain as the most promising antiviral targets.

Finally, from a methodological point of view, the current work highlights the strong complementarity between the relatively new method of MP and the classical biophysical methods of SV-AUC and DLS for studying interacting systems, all of which report on distributions of macromolecules and their complexes in equilibrium in solution without requiring separations, modifications, or surface immobilization. Their combination provides a powerful new approach for the study of multi-step assembly processes. MP has superior mass

resolution and applications to the analysis of interacting systems have been pioneered (66), but the technique is limited to nanomolar concentrations and species above 30 kDa. In the present case it served well to study onset of the multi-step assembly reaction. SV-AUC far extends the concentration range and therefore allowed observation of the assembly up to the largest assembly products, adding spectral resolution to distinguish composition of the protein/NA complexes. (Fluorescence detected SV-AUC also extends sensitivity into the picomolar range, but attachment of a fluorescent tag has been found detrimental in the present system (37).) Importantly, in the intermediate concentration range accessible to both MP and SV-AUC we obtained highly consistent results. While SV-AUC carries molar mass information only indirectly, folded into hydrodynamic migration, the latter can be decomposed into size and mass information with the help of independent information on translational friction, here taken from DLS for the largest reaction product. Alternatively, shape information may be derived by combination of the complex mass values from MP with sedimentation coefficients from SV. Furthermore, an additional promising aspect from this combination of methods is that the notorious dependency of sedimentation boundary patterns on the complex lifetimes in SV-AUC may be untangled with information from MP. Jointly, MP, SV-AUC and DLS should be amenable to global multi-method analysis (67) but detailed affinities of all assembly steps would be of limited interest in the current case due to the pleomorphic nature of the RNP assembly. Overall, this approach might prove valuable for the elucidation of architecture of other RNPs, for example, in assembly of other RNA viruses that lack the high degree of capsid symmetry of helical viruses, and are therefore currently inaccessible to high-resolution structural methods (2).

### Data availability

All data are available in the main text or the [supplementary materials](#). Raw data supporting this work is available in the Harvard Dataverse at <https://doi.org/10.7910/DVN/UYPEHJ>.

### Supplementary data

[Supplementary Data](#) are available at NAR Online.

### Acknowledgements

We thank Vanessa Wall, Brianna Higgins and J.-P. Denson for N-CTD protein production. This project was funded in part with federal funds from the National Cancer Institute, National Institutes of Health Contract 75N91019D00024. J.A.D. got support from by a grant from the NIH (R21AI59666) and by support from the Howard Hughes Medical Institute and the Gladstone Institutes. M.O. received support from NIH Host pathogen mapping initiative (HPMI) grant (U19 AI135990), the James B. Pendleton Charitable Trust, the Roddenberry Foundation, P. and E. Taft, and the Gladstone Institutes. M.O. is a Chan Zuckerberg Biohub – San Francisco Investigator. This work was supported by the Intramural Research Programs of NIBIB, NHLBI, NIDDK, and NCI of the National Institutes of Health, Bethesda.

### Funding

National Institutes of Health [ZIA EB000095-04 to P.S., R21AI59666 to J.A.D., U19 AI135990 to M.O., NIH Contract 75N91019D00024 to D.E.]. Funding for open access charge: National Institutes of Health.

### Conflict of interest statement

J.A.D. is a cofounder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, Intellia Therapeutics, and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Vertex, Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Synthego, Algen Biotechnologies, Felix Biosciences, The Column Group, and Inari. J.A.D. is a director at Johnson & Johnson and Tempus and has research projects sponsored by Biogen, Pfizer, AppleTree Partners, and Roche. M.O. is a cofounder of DirectBio and on the SAB for Invisishield.

### References

- Comas-Garcia, M. (2019) Packaging of genomic RNA in positive-sense single-stranded RNA viruses: a complex story. *Viruses*, **11**, 253.
- Speir, J.A. and Johnson, J.E. (2012) Nucleic acid packaging in viruses. *Curr. Opin. Struct. Biol.*, **22**, 65–71.
- Klein, S., Cortese, M., Winter, S.L., Wachsmuth-Melm, M., Neufeldt, C.J., Cerikan, B., Stanifer, M.L., Boulant, S., Bartenschlager, R. and Chlanda, P. (2020) SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat. Commun.*, **11**, 5885.
- Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., et al. (2020) Molecular architecture of the SARS-CoV-2 virus. *Cell*, **183**, 730–738.
- Wolff, G., Limpens, R.W.A.L., Zevenhoven-Dobbe, J.C., Laugks, U., Zheng, S., de Jong, A.W.M., Koning, R.I., Agard, D.A., Grünewald, K., Koster, A.J., et al. (2020) A molecular pore spans the double membrane of the coronavirus replication organelle. *Science*, **369**, 1395–1398.
- Ding, B., Qin, Y. and Chen, M. (2016) Nucleocapsid proteins: roles beyond viral RNA packaging. *Wiley Interdiscip. Rev. RNA*, **7**, 213–226.
- Masters, P.S. (2019) Coronavirus genomic RNA packaging. *Virology*, **537**, 198–207.
- Bracquemond, D. and Muriaux, D. (2021) Betacoronavirus assembly: clues and perspectives for elucidating SARS-CoV-2 particle formation and egress. *mBio*, **12**, e02388-23.
- Mendonça, L., Howe, A., Gilchrist, J.B., Sheng, Y., Sun, D., Knight, M.L., Zanetti-Domingues, L.C., Bateman, B., Krebs, A.S., Chen, L., et al. (2021) Correlative multi-scale cryo-imaging unveils SARS-CoV-2 assembly and egress. *Nat. Commun.*, **12**, 4629.
- Sola, I., Mateos-Gomez, P.A., Almazan, F., Zúñiga, S. and Enjuanes, L. (2011) RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol.*, **8**, 237–248.
- Tugaeva, K.V., Hawkins, D.E.D.P., Smith, J.L.R., Bayfield, O.W., Ker, D.-S., Sysoev, A.A., Klychnikov, O.I., Antson, A.A. and Sluchanko, N.N. (2021) The mechanism of SARS-CoV-2 nucleocapsid protein recognition by the Human 14-3-3 proteins. *J. Mol. Biol.*, **433**, 166875.
- Sender, R., Bar-On, Y.M., Gleizer, S., Bernshtein, B., Flamholz, A., Phillips, R. and Milo, R. (2021) The total number and mass of SARS-CoV-2 virions. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2024815118.
- Wu, W., Cheng, Y., Zhou, H., Sun, C. and Zhang, S. (2023) The SARS-CoV-2 nucleocapsid protein: its role in the viral life cycle, structure and functions, and use as a potential target in the development of vaccines and diagnostics. *Virol. J.*, **20**, 6.

14. Yu, H., Guan, F., Miller, H., Lei, J. and Liu, C. (2023) The role of SARS-CoV-2 nucleocapsid protein in antiviral immunity and vaccine development. *Emerg. Microbes Infect.*, **12**, e2164219.
15. Schuck, P. and Zhao, H. (2023) Diversity of short linear interaction motifs in SARS-CoV-2 nucleocapsid protein. *mBio*, **14**, e02388-23.
16. Korn, S.M., Dhamotharan, K., Jeffries, C.M. and Schlundt, A. (2023) The preference signature of the SARS-CoV-2 nucleocapsid NTD for its 5'-genomic RNA elements. *Nat. Commun.*, **14**, 3331.
17. Savastano, A., Ibáñez de Opakua, A., Rankovic, M. and Zweckstetter, M. (2020) Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat. Commun.*, **11**, 6041.
18. Roden, C.A., Dai, Y., Giannetti, C.A., Seim, J., Lee, M., Sealfon, R., McLaughlin, G.A., Boerneke, M.A., Iserman, C., Wey, S.A., et al. (2022) Double-stranded RNA drives SARS-CoV-2 nucleocapsid protein to undergo phase separation at specific temperatures. *Nucleic Acids Res.*, **50**, 8168–8192.
19. Cascarina, S.M. and Ross, E.D. (2022) Phase separation by the SARS-CoV-2 nucleocapsid protein: consensus and open questions. *J. Biol. Chem.*, **298**, 101677.
20. Carlson, C.R., Asfaha, J.B., Ghent, C.M., Howard, C.J., Hartooni, N., Safari, M., Frankel, A.D. and Morgan, D.O. (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.
21. Cubuk, J., Alston, J.J., Incicco, J.J., Singh, S., Stuchell-Breteron, M.D., Ward, M.D., Zimmerman, M.I., Vithani, N., Griffith, D., Wagoner, J.A., et al. (2021) The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.*, **12**, 1936.
22. Iserman, C., Roden, C.A., Boerneke, M.A., Sealfon, R.S.G., McLaughlin, G.A., Jungreis, I., Fritch, E.J., Hou, Y.J., Ekena, J., Weidmann, C.A., et al. (2020) Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol. Cell*, **80**, 1078–1091.
23. Pontoriero, L., Schiavina, M., Korn, S.M., Schlundt, A., Pierattelli, R. and Felli, I.C. (2022) NMR reveals specific tracts within the intrinsically disordered regions of the SARS-CoV-2 nucleocapsid protein involved in RNA encountering. *Biomolecules*, **12**, 929.
24. Cubuk, J., Alston, J.J., Incicco, J.J., Holehouse, A.S., Hall, K.B., Stuchell-Breteron, M.D. and Soranno, A. (2024) The disordered N-terminal tail of SARS-CoV-2 nucleocapsid protein forms a dynamic complex with RNA. *Nucleic Acids Res.*, **52**, 2609–2624.
25. Carlson, C.R., Adly, A.N., Bi, M., Howard, C.J., Frost, A., Cheng, Y. and Morgan, D.O. (2022) Reconstitution of the SARS-CoV-2 ribonucleosome provides insights into genomic RNA packaging and regulation by phosphorylation. *J. Biol. Chem.*, **298**, 102560.
26. Zhao, H., Wu, D., Hassan, S.A., Nguyen, A., Chen, J., Piszczek, G. and Schuck, P. (2023) A conserved oligomerization domain in the disordered linker of coronavirus nucleocapsid proteins. *Sci. Adv.*, **9**, eadg6473.
27. Zhao, H., Nguyen, A., Wu, D., Li, Y., Hassan, S.A., Chen, J., Shroff, H., Piszczek, G. and Schuck, P. (2022) Plasticity in structure and assembly of SARS-CoV-2 nucleocapsid protein. *PNAS Nexus*, **1**, pgac049.
28. Karkanitsa, M., Li, Y., Valenti, S., Spathies, J., Kelly, S., Mehalko, J., Drew, M., Denson, J., Putman, Z. and Fathi, P. (2023) Dynamics of SARS-CoV-2 seroprevalence in a large US population over a period of 12 months. medRxiv doi: <https://doi.org/10.1101/2023.10.20.23297329>, 21 October 2023, preprint: not peer reviewed.
29. Taylor, T., Denson, J.-P. and Esposito, D. (2017) Optimizing expression and solubility of proteins in *E. coli* using modified media and induction parameters. *Methods Mol. Biol.*, **1586**, 65–82.
30. Zhao, H., Brown, P.H. and Schuck, P. (2011) On the distribution of protein refractive index increments. *Biophys. J.*, **100**, 2309–2317.
31. Syed, A.M., Taha, T.Y., Tabata, T., Chen, I.P., Ciling, A., Khalid, M.M., Sreekumar, B., Chen, P.-Y., Hayashi, J.M., Soczek, K.M., et al. (2021) Rapid assessment of SARS-CoV-2-evolved variants using virus-like particles. *Science*, **374**, 1626–1632.
32. Adly, A.N., Bi, M., Carlson, C.R., Syed, A.M., Ciling, A., Doudna, J.A., Cheng, Y. and Morgan, D.O. (2023) Assembly of SARS-CoV-2 ribonucleosomes by truncated N\* variant of the nucleocapsid protein. *J. Biol. Chem.*, **299**, 105362.
33. Schuck, P., Zhao, H., Brautigam, C.A. and Ghirlardo, R. (2015) In: *Basic Principles of Analytical Ultracentrifugation*. CRC Press, Boca Raton, FL.
34. Ghirlardo, R., Balbo, A., Piszczek, G., Brown, P.H., Lewis, M.S., Brautigam, C.A., Schuck, P. and Zhao, H. (2013) Improving the thermal, radial, and temporal accuracy of the analytical ultracentrifuge through external references. *Anal. Biochem.*, **440**, 81–95.
35. Schuck, P. (2016) In: *Sedimentation Velocity Analytical Ultracentrifugation: Discrete Species and Size-Distributions of Macromolecules and Particles*. CRC Press, Boca Raton, FL.
36. Greive, S.J., Lins, A.F. and von Hippel, P.H. (2005) Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in *Escherichia coli*. *J. Biol. Chem.*, **280**, 36397–36408.
37. Zhao, H., Wu, D., Nguyen, A., Li, Y., Adão, R.C., Valkov, E., Patterson, G.H., Piszczek, G. and Schuck, P. (2021) Energetic and structural features of SARS-CoV-2 N-protein co-assemblies with nucleic acids. *iScience*, **24**, 102523.
38. Dinesh, D.C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V. and Boura, E. (2020) Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.*, **16**, e1009100.
39. Schuck, P. and Zhao, H. (2017) In: *Sedimentation Velocity Analytical Ultracentrifugation: Interacting Systems*. CRC Press, Boca Raton, FL.
40. Balbo, A., Minor, K.H., Velikovskiy, C.A., Mariuzza, R.A., Peterson, C.B. and Schuck, P. (2005) Studying multiprotein complexes by multi-signal sedimentation velocity analytical ultracentrifugation. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 81–86.
41. Schuck, P. (2010) Sedimentation patterns of rapidly reversible protein interactions. *Biophys. J.*, **98**, 2005–2013.
42. Record, M.T., Anderson, C.F. and Lohman, T.M. (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.*, **11**, 103–178.
43. Nguyen, A., Zhao, H., Myagmarsuren, D., Srinivasan, S., Wu, D., Chen, J., Piszczek, G. and Schuck, P. (2024) Modulation of biophysical properties of nucleocapsid protein in the mutant spectrum of SARS-CoV-2. *eLife*, **13**, RP94836.
44. Li, Y., Struwe, W.B. and Kukura, P. (2020) Single molecule mass photometry of nucleic acids. *Nucleic Acids Res.*, **48**, E97.
45. Casassa, E.F. and Eisenberg, H. (1961) Partial specific volumes and refractive index increments in multicomponent systems. *J. Phys. Chem.*, **65**, 427–433.
46. Chang, C.K., Hou, M.H., Chang, C.F., Hsiao, C.D. and Huang, T.H. (2014) The SARS coronavirus nucleocapsid protein - forms and functions. *Antiviral Res.*, **103**, 39–50.
47. Dang, M. and Song, J. (2022) CTD of SARS-CoV-2 N protein is a cryptic domain for binding ATP and nucleic acid that interplay in modulating phase separation. *Protein Sci.*, **31**, 345–356.
48. Wu, C., Qavi, A.J., Hachim, A., Kaviani, N., Cole, A.R., Moyle, A.B., Wagner, N.D., Sweeney-Gibbons, J., Rohrs, H.W., Gross, M.L., et al. (2021) Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain. *iScience*, **24**, 102681.
49. Zinzula, L., Basquin, J., Bohn, S., Beck, F., Klumpe, S., Pfeifer, G., Nagy, I., Bracher, A., Hartl, F.U. and Baumeister, W. (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.
50. Syed, A.M., Ciling, A., Chen, I.P., Carlson, C.R., Adly, A., Martin, H., Taha, T.Y., Khalid, M.M., Bouhaddou, M., Ummadi, M., *et al.* (2023) SARS-CoV-2 evolution balances conflicting roles of N protein phosphorylation. <https://doi.org/10.2139/ssrn.4472729>.
  51. Johnson, B.A., Zhou, Y., Lokugamage, K.G., Vu, M.N., Bopp, N., Crocquet-Valdes, P.A., Kalveram, B., Schindewolf, C., Liu, Y., Scharon, D., *et al.* (2022) Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. *PLoS Pathog.*, **18**, e1010627.
  52. Yaron, T.M., Heaton, B.E., Levy, T.M., Johnson, J.L., Jordan, T.X., Cohen, B.M., Kerelsky, A., Lin, T., Liberatore, K.M., Bulaon, D.K., *et al.* (2022) Host protein kinases required for SARS-CoV-2 nucleocapsid phosphorylation and viral replication. *Sci. Signal.*, **15**, eabm0808.
  53. Liu, X., Verma, A., Garcia, G., Ramage, H., Lucas, A., Myers, R.L., Michaelson, J.J., Coryell, W., Kumar, A., Charney, A.W., *et al.* (2021) Targeting the coronavirus nucleocapsid protein through GSK-3 inhibition. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2113401118.
  54. Padroni, G., Bikaki, M., Novakovic, M., Wolter, A.C., Rüdiger, S.H., Gossert, A.D., Leitner, A. and Allain, F.H.T. (2023) A hybrid structure determination approach to investigate the druggability of the nucleocapsid protein of SARS-CoV-2. *Nucleic Acids Res.*, **51**, 4555–4571.
  55. Redzic, J.S., Lee, E., Born, A., Issaian, A., Henen, M.A., Nichols, P.J., Blue, A., Hansen, K.C., D'Alessandro, A., Vögeli, B., *et al.* (2021) The inherent dynamics and interaction sites of the SARS-CoV-2 nucleocapsid N-terminal region. *J. Mol. Biol.*, **433**, 167108.
  56. Różycki, B. and Boura, E. (2022) Conformational ensemble of the full-length SARS-CoV-2 nucleocapsid (N) protein based on molecular simulations and SAXS data. *Biophys. Chem.*, **288**, 106843.
  57. Estelle, A.B., Forsythe, H.M., Yu, Z., Hughes, K., Lasher, B., Allen, P., Reardon, P.N., Hendrix, D.A. and Barbar, E.J. (2023) RNA structure and multiple weak interactions balance the interplay between RNA binding and phase separation of SARS-CoV-2 nucleocapsid. *PNAS Nexus*, **2**, p. gad333.
  58. Baric, R.S., Nelson, G.W., Fleming, J.O., Deans, R.J., Keck, J.G., Casteel, N. and Stohlman, S.A. (1988) Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. *J. Virol.*, **62**, 4280–4287.
  59. Morse, M., Sefcikova, J., Rouzina, I., Beuning, P.J. and Williams, M.C. (2023) Structural domains of SARS-CoV-2 nucleocapsid protein coordinate to compact long nucleic acid substrates. *Nucleic Acids Res.*, **51**, 290–303.
  60. Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., Yang, M., Xing, X., Wang, Y., Li, M., *et al.* (2021) The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.*, **12**, 3917.
  61. Chen, H., Meisburger, S.P., Pabitt, S.A., Sutton, J.L., Webb, W.W. and Pollack, L. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 799–804.
  62. Ding, J., Lee, Y.-T., Bhandari, Y., Schwieters, C.D., Fan, L., Yu, P., Tarosov, S.G., Stagno, J.R., Ma, B., Nussinov, R., *et al.* (2023) Visualizing RNA conformational and architectural heterogeneity in solution. *Nat. Commun.*, **14**, 714.
  63. Perlmutter, J.D. and Hagan, M.F. (2015) Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.*, **66**, 217–239.
  64. Larson, S.B. and McPherson, A. (2001) Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr. Opin. Struct. Biol.*, **11**, 59–65.
  65. Ma, X., Wang, Y., Gao, Y., Wang, Y., Yan, A., Chen, J., Zhang, L., Wang, P., Zhao, J. and Liu, Z. (2023) Structural characteristics of the SARS-CoV-2 Omicron lineages BA.1 and BA.2 virions. *Signal Transduct. Target. Ther.*, **8**, 131.
  66. Wu, D. and Piszczek, G. (2020) Measuring the affinity of protein–protein interactions on a single-molecule level by mass photometry. *Anal. Biochem.*, **592**, 113575.
  67. Zhao, H. and Schuck, P. (2012) Global multi-method analysis of affinities and cooperativity in complex systems of macromolecular interactions. *Anal. Chem.*, **84**, 9513–9519.