Phosphorylation in the Ser/Arg-rich region of the nucleocapsid of SARS-CoV-2 regulates phase separation by inhibiting self-association of a distant helix

Hannah Stuwe[1,‡], Patrick N. Reardon[2,‡] , Zhen Yu[1], Sahana Shah[1], Kaitlyn Hughes[1], and Elisar J. Barbar[1,*]
From the [1]Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon, USA; [2]OSU NMR Facility, Oregon State University, Corvallis, Oregon, USA

Reviewed by members of the JBC Editorial Board. Edited by Wolfgang Peti

The nucleocapsid protein (N) of SARS-CoV-2 is essential for virus replication, genome packaging, evading host immunity, and virus maturation. N is a multidomain protein composed of an independently folded monomeric N-terminal domain that is the primary site for RNA binding and a dimeric C-terminal domain that is essential for efficient phase separation and condensate formation with RNA. The domains are separated by a disordered Ser/Arg-rich region preceding a self-associating Leu-rich helix. Phosphorylation in the Ser/Arg region in infected cells decreases the viscosity of N:RNA condensates promoting viral replication and host immune evasion. The molecular level effect of phosphorylation, however, is missing from our current understanding. Using NMR spectroscopy and analytical ultracentrifugation, we show that phosphorylation destabilizes the self-associating Leu-rich helix 30 amino-acids distant from the phosphorylation site. NMR and gel shift assays demonstrate that RNA binding by the linker is dampened by phosphorylation, whereas RNA binding to the full-length protein is not significantly affected presumably due to retained strong interactions with the primary RNA-binding domain. Introducing a switchable self-associating domain to replace the Leu-rich helix confirms the importance of linker self-association to droplet formation and suggests that phosphorylation not only increases solubility of the positively charged elongated Ser/Arg region as observed in other RNA-binding proteins but can also inhibit self-association of the Leu-rich helix. These data highlight the effect of phosphorylation both at local sites and at a distant self-associating hydrophobic helix in regulating liquid–liquid phase separation of the entire protein.

SARS-CoV-2 is an enveloped single-stranded, positive sense RNA virus with a large 30kb genome (1). The SARS-CoV-2 virion is composed of four structural proteins: spike (S), membrane (M), envelope (E), and nucleocapsid (N). The viral membrane, with S, M, and E, surround the helical nucleocapsid containing the viral RNA genome encapsulated by N (1–3). N has several functions within the viral life cycle but is primarily involved in protecting the viral RNA genome by binding, condensing, and packaging it within the virion (2, 4). N also functions in the replicase transcriptase complexes where it mediates the synthesis of genomic RNA (gRNA) within the host cell (5–7). Additionally, N contributes to innate immune evasion *via* sequestering stress granule protein G3BP1 (8, 9). N's ability to phase separate, particularly with gRNA, has been extensively investigated (3, 4, 10–18) and undergoing this process is what allows N to participate in a vast and varied functions (12, 19). These N-containing condensates are also a promising target for drug developments (20).

N is a 419 amino-acid long protein with two independently folded domains, the N-terminal domain (NTD) and the C-terminal domain (CTD) (Fig. 1) (21). These domains are flanked by two disordered tails and separated by a central disordered linker that contains a serine/arginine rich (SR-rich) region, followed by a short Leu-rich helix (LRH) (Fig. 1). The CTD forms a domain swapped dimer, facilitating the dimerization of N (21–23) and is essential for phase separation with RNA (24) while the NTD is the primary RNA-binding domain (25–28) and its binding to RNA is strengthened in constructs that contain the intrinsically disordered N-terminal and central regions (29, 30). Upon binding to RNA, the N protein packs the NTD and the CTD closer together (31). Multiple studies using hydrogen-deuterium exchange mass spectrometry and analytical ultracentrifugation have shown that the central linker self-associates (32, 33) and using molecular dynamics modeling, have predicted the helical LRH region as the most likely site of self-association (33–35).
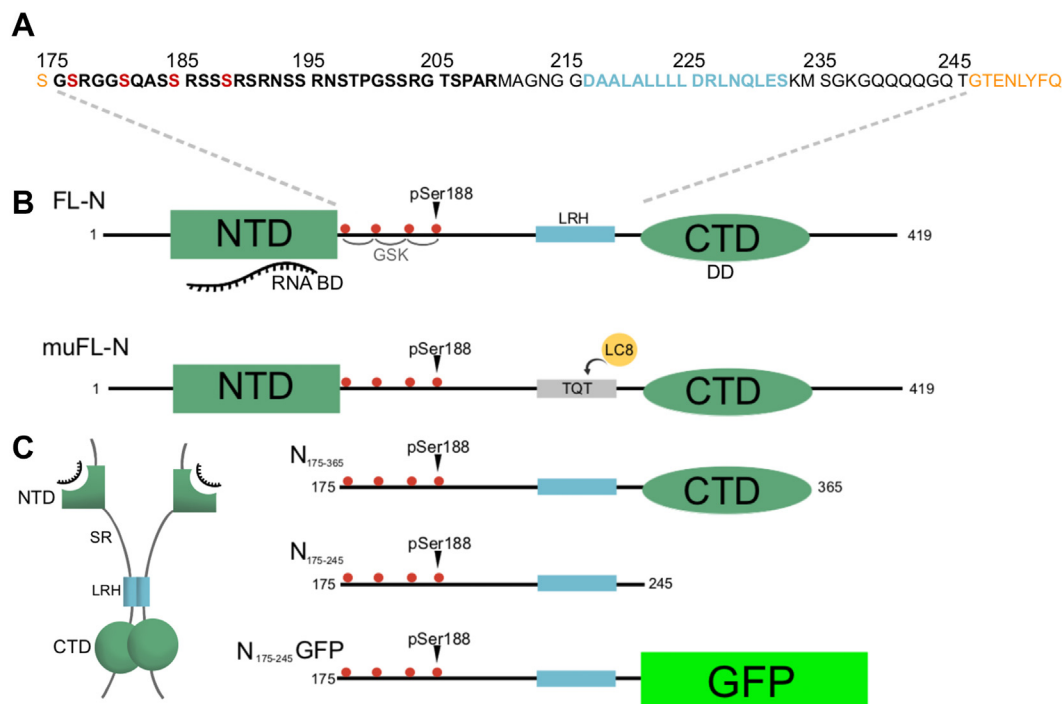
The dual roles of N in the assembly with gRNA into viral RNA-protein complexes and in localization to the replicase transcriptase complexes to enhance viral replication and transcription are regulated by phosphorylation of the SR-rich region of N (36–39). Phosphorylation of the homolog SARS-CoV-1 SR-rich region did not significantly affect RNA binding of the nucleocapsid protein but impaired its ability to form oligomers (40). In SARS-CoV-2, phosphorylation of N altered liquid-liquid phase separation (LLPS) formation, with unmodified protein forming gel-like condensates, while phosphorylated protein forming more dynamic liquid-like droplets (19). In a follow up study, the Morgan group showed that the assembly into viral particles requires the LRH and using

Figure 1. Domain maps of SARS-CoV-2N and constructs used in this work. *A*, amino acid sequence of SARS-CoV-2N construct spanning the linker region from residues 175 to 245. Residual TEV protease cleavage sites are in *orange*. The SR-rich region is in *bold black*, with phosphorylation sites investigated here in *bold red*. The Leu-rich helix (LRH) spanning residues 216 to 232 is in *bold blue*. *B*, domain maps of full-length (FL-N), mutant full-length (muFL-N), shorter constructs $N_{175-365}$, $N_{175-245}$, and GFP-tagged $N_{175-245}$ from *top* to *bottom*, respectively. Intrinsically disordered regions are represented by *black lines*, the structured RNA-binding domain NTD is represented by a *dark green rectangle*, and the dimerization domain CTD by a *dark green oval*. The LRH is represented by a *small blue rectangle*. GSK priming site pSer188 is pointed to by an *arrow*, and the subsequent phosphorylation sites are indicated by *red circles*. The muFL-N has a TQT recognition motif for LC8-binding site incorporated to replace the LRH shown as *gray rectangle*. *C*, a cartoon depiction of dimeric FL-N showing self-association of LRH- and RNA-binding site of the NTD. CTD, C-terminal domain; GSK, glycogen synthase kinase; LC8, dynein light chain 8; LRH, Leu-rich helix; NTD, N-terminal domain; SR-rich, serine/arginine-rich.

phosphomimetic SR demonstrated that phosphorylation inhibits vRNP assembly (41). The phosphorylation-mediated differences in LLPS characteristics are hypothesized to contribute to N switching between functions, with liquid-like phosphorylated N droplets potentially promoting viral replication and host immune evasion and gel-like unphosphorylated droplets promoting viral RNA packaging (10, 11). Overall, these observations demonstrate the critical role of phosphorylation in regulating the function of N during the virus life cycle. It is notable that the SR-rich region in the linker can have up to 15 phosphorylation sites depending on cell type (42) which will presumably alter the structure and interactions of the highly positive charge pattern of the unphosphorylated protein.

Here, we systematically probe the effect of a single phosphorylation event, at serine 188, a known priming site of phosphorylation (42) and of hyperphosphorylation, *via* glycogen synthase kinase 3β (GSK-3), on the structure of the linker domain, the self-association of the LRH, and the interactions of the full-length protein with gRNA, including its ability to phase separate. Our structural and functional characterization demonstrate the importance of phosphorylation in affecting not only the local environment but also in inhibiting the self-association of LRH distant from the site of phosphorylation. We expect that phosphorylation at all potential sites will cause an even more profound change. We present a model that

explains how phosphorylation could alter the structure of N in condensates and force unpackaging of viral RNA.

Results

Association of the central linker is localized to helical residues 216 to 232

A suite of BEST triple resonance NMR experiments helped assign the backbone resonances of $N_{175-245}$ at 10 °C (Fig. 2), including the 225 to 230 region missing in the assignments deposited in the BMRB (43). The ^{15}N HSQC spectrum is consistent with a primarily disordered polypeptide, with generally narrow chemical shift dispersion. Fast time scale dynamics determined by measuring nuclear spin relaxation ^{15}N R_1, ^{15}N R_2, and {^{1}H}-^{15}N Nuclear Overhauser Effect (NOE) identified residues 178 to 215 and 236 to 248 to be fully disordered with NOE's below 0.5 and negative NOE's at the termini (Fig. 2). ^{15}N R_2 and R_1 relaxation rates in these regions were relatively uniform, while ^{15}N R_2 and {^{1}H}-^{15}N NOE were elevated for residues 216 to 232, which together with the carbon chemical shifts (Fig. S1) confirm that residues 216 to 232 adopt a helical structure and that except for this short helix, $N_{175-245}$ does not adopt any regular secondary structure under our experimental conditions.

As stated above, missing peaks for the deposited assignments were attributed to self-association of the central linker
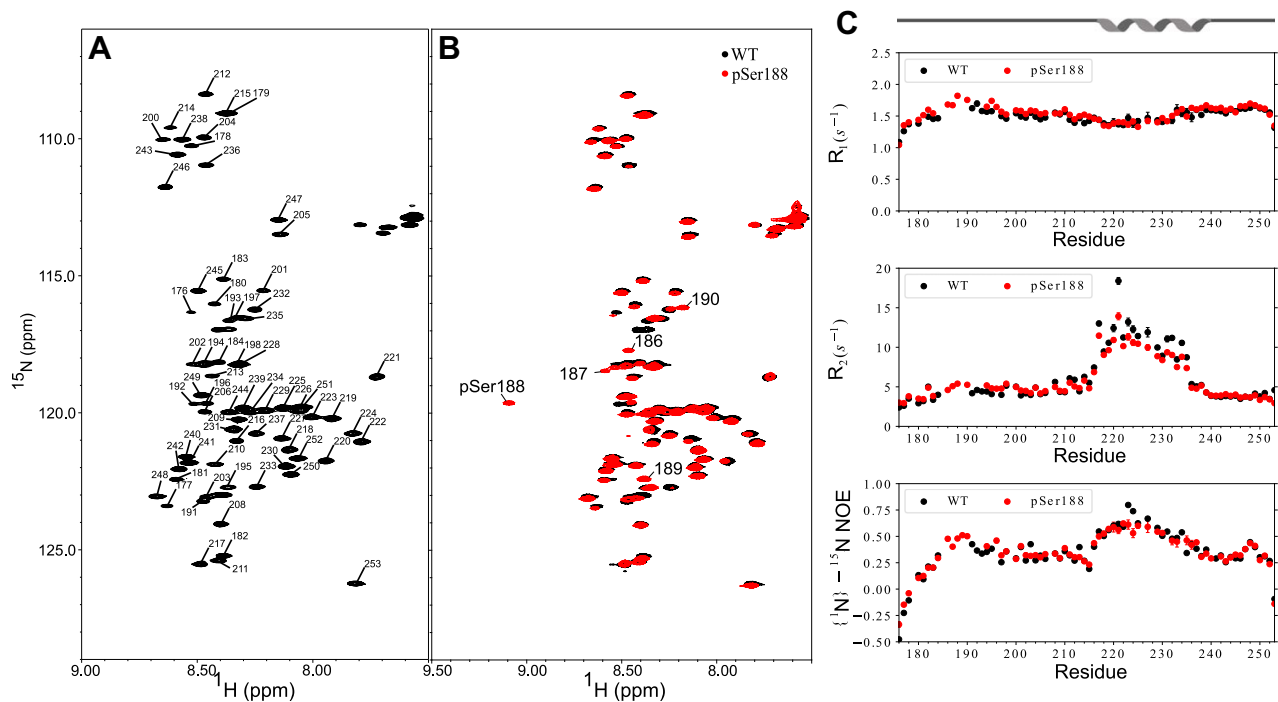
Figure 2. NMR structural characterization of WT N175 to 245 and pSer188 N175 to 245. *A*, ^{15}N-HSQC spectrum of 150 μM WT $N_{175-245}$ at 10 °C with backbone resonance assignments. *B*, overlay of ^{15}N-HSQC spectrum of WT $N_{175-245}$ (*black*) and pSer188 $N_{175-245}$ (*red*), with shifted peaks labeled. *C*, ^{15}N nuclear spin relaxation of WT $N_{175-245}$ (*black*) compared to pSer188 $N_{175-245}$ (*red*). The Leu-rich helix is indicated.

(43). Consistent with this interpretation, the peaks assigned at lower concentrations of ^{15}N-labeled $N_{175-245}$ (Fig. 3) correspond to the helical region and thus identify self-association to be localized to the helical residues. At 300 μM, these peaks disappeared from the spectrum (Fig. 3).

To confirm that peak attenuation is in fact due to self-association and to evaluate the strength of self-association, we performed sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments on GFP-tagged $N_{175-245}$ (WT-$N_{175-245}$ GFP). The GFP tag was used to provide a stronger UV absorbance. The SV-AUC profiles clearly show two peaks in the WT-$N_{175-245}$ GFP at 200 μM, corresponding to ~3.1S and 4.8S, suggesting the presence of monomeric and dimeric species (Fig. 3). The GFP control at the same concentration only gave rise to a single peak, indicating no significant dimerization under these conditions (Fig. S2). Subsequent sedimentation equilibrium analysis (SE-AUC) gave a mass average molecular weight for $N_{175-245}$ GFP of ~66 kDa near the expected molecular weight of 72 kDa for a dimer (Fig. 3) which upon analysis to a monomer-dimer equilibrium model yielded a dissociation constant of 77 μM. Together, these data indicate that the LRH can self-associate but only weakly in the linker construct.

### Self-association of LRH is strengthened in context of dimeric CTD

To probe the behavior of the LRH in the context of the full-length protein (FL-N), we collected a series of ^{15}N HSQC experiments at decreasing concentrations of a ^{15}N-labeled construct of N including the central linker and the CTD, spanning residues 175 to 365 ($N_{175-365}$) as a more tractable model of dimeric FL-N. For concentrations of $N_{175-365}$ as low as 10 μM, which is still considerably above the 1 μM dimerization constant of the CTD without the LRH (22), but significantly lower than the 77 μM dimerization constant of the LRH alone, resonances corresponding to residues 216 to 232 were still not detectable, indicating they are involved in self-association at those concentrations (Fig. 4). For comparison, these residues were observed at 200 μM in the linker construct. Additionally, SV-AUC experiments at 350 μM, the highest concentration assessed, showed two distinct peaks corresponding to ~2.7S and 4.2S, suggesting the presence of dimeric and tetrameric species (Fig. 4). At 200 μM, only a primary peak at ~2.7S and a shoulder in the ~3.3 to 4.5S range were observed. This broad shoulder suggests exchange between dimeric and tetrameric states. At 100 μM, a single broad peak centering at ~2.7S was observed, indicative of a dimeric $N_{175-365}$ as the primary species in solution, but still undergoing exchange with higher order species. SE-AUC data collected to confirm that the higher order species is indeed a tetramer were fit to a dimer-tetramer model yielding an equilibrium dissociation constant of about 300 μM. Together, these data demonstrate that the self-association of the LRH is much stronger in context of the dimeric CTD and could form tetramers bridging two dimeric N molecules.

Phosphorylation in the SR-rich region inhibits self-association of the LRH

N protein phosphorylation is a complex process involving several host kinases acting sequentially to phosphorylate the SR-rich region (36, 42, 44). We simplify this process by focusing only on Ser188 as an
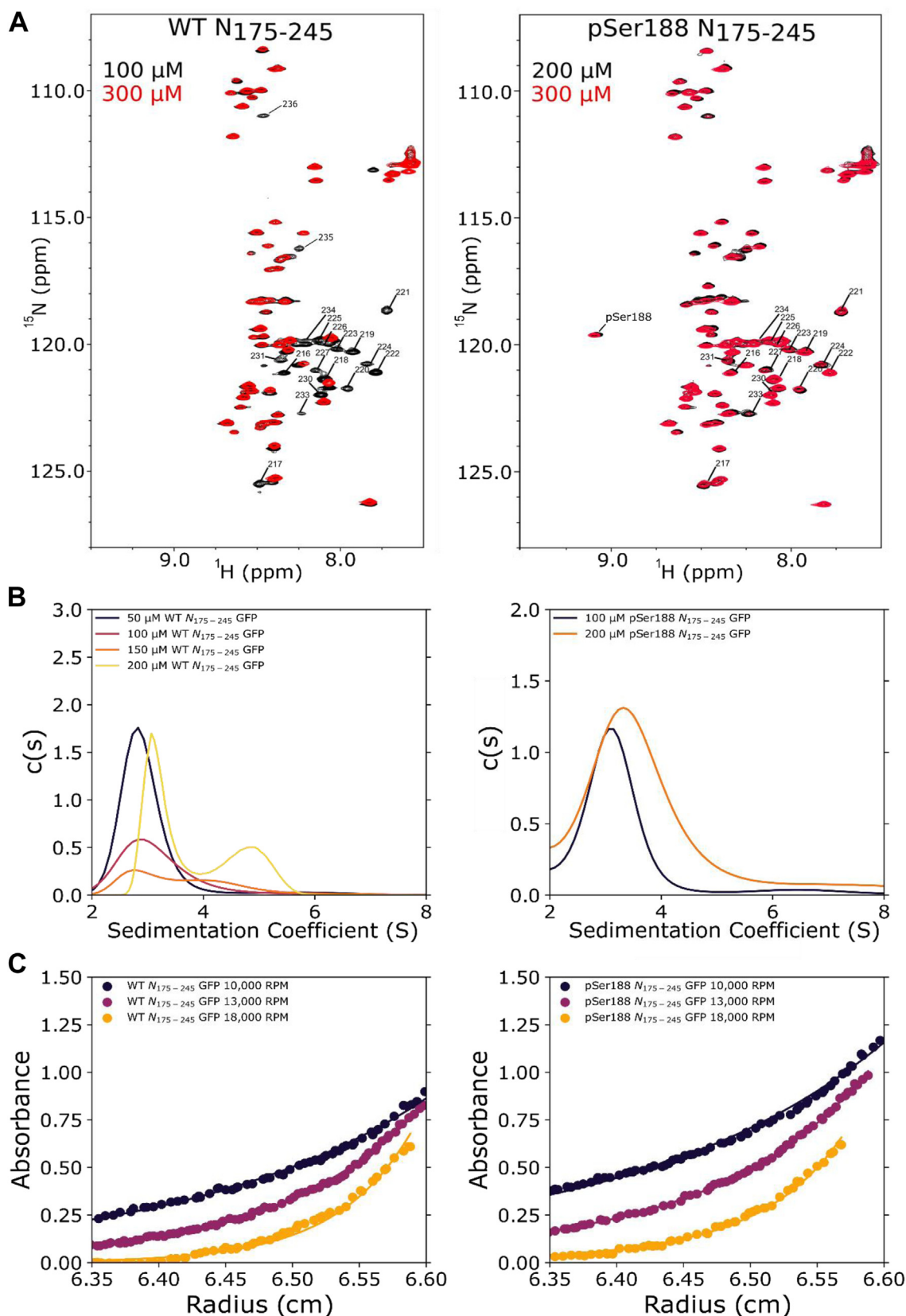
Figure 3. Concentration dependence of self-association of WT and pSer188 N175 to 245. *A*, ^{15}N-HSQC spectra of WT $N_{175-245}$ (*left*) at 300 μM (*red*) and 100 μM (*black*) concentration. Resonances that disappear at higher concentration are labeled. ^{15}N-HSQC spectra of pSer188 $N_{175-245}$ (*right*) at 300 μM (*red*) and 200 μM (*black*) concentration. The same resonances are labeled. Missing peaks for residues 233 to 235 is due to ordered structure in proximity of the self-associating helix. *B*, sedimentation velocity analytical ultracentrifugation (SV-AUC) of WT-$N_{175-245}$ GFP (*left*) at 50 to 200 μM concentration and SV-AUC of pSer188 $N_{175-245}$ GFP (*right*) at 100 and 200 μM concentration. *C*, representative sedimentation equilibrium data at three rotor speeds for WT-$N_{175-245}$ GFP (*left*) and for pSer 188 $N_{175-245}$ GFP (*right*). Monomer-dimer equilibrium fits are shown as *solid* lines. All AUC experiments with $N_{175-245}$ were performed with GFP fusion.
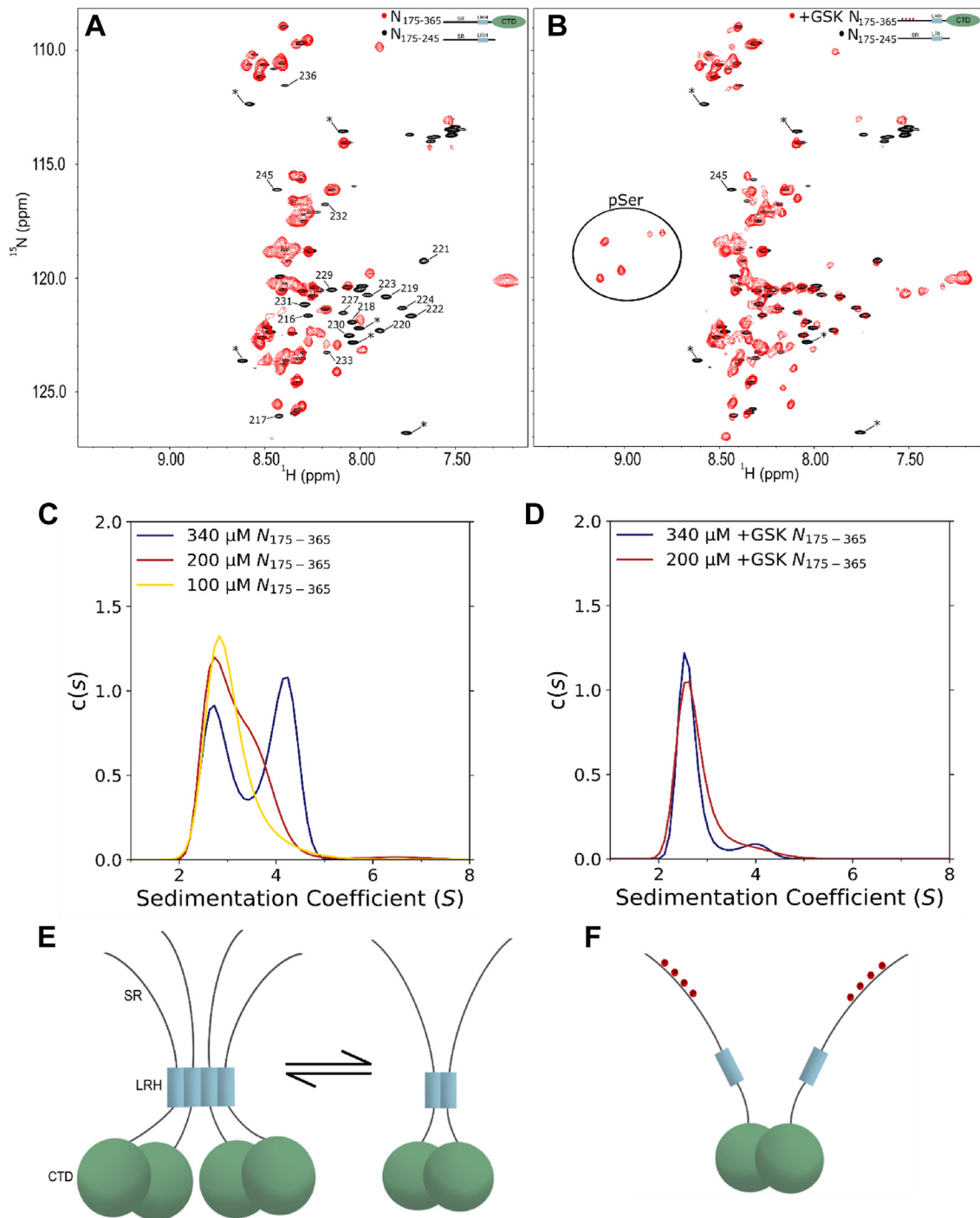
Figure 4. Concentration dependence self-association of WT and phosphorylated N175 to 365. *A*, ^{15}N-TROSY-HSQC spectra of 50 μM WT $N_{175-365}$ (*red*) compared to WT $N_{175-245}$ (*black*). Resonances of $N_{175-245}$ that are not present in $N_{175-365}$ are labeled. Resonances that correspond to residual TEV cleavage sites in $N_{175-245}$ are indicated by an *asterisk*. *B*, ^{15}N-TROSY-HSQC spectra of 50 μM GSK-hyperphosphorylated $N_{175-365}$ (+GSK $N_{175-365}$) (*red*) compared to WT $N_{175-245}$ (*black*). Down-field resonances corresponding to phosphorylated serine are circled. Resonance labeling scheme is the same as in (*A*). *C*, SV-AUC of WT-$N_{175-365}$ at 100 to 340 μM concentration. *D*, SV-AUC of +GSK $N_{175-365}$ at 200 and 340 μM concentration. *E*, representative model of proposed tetramerization of $N_{175-365}$ due to LRH self-association (*dark blue*). Tetrameric $N_{175-365}$ is in exchange with dimeric $N_{175-365}$ as indicated on the *right*. *F*, representative model of the effect of hyperphosphorylation of $N_{175-365}$ showing the CTD as an intact dimer (*green spheres*) but with dissociation of the LRH from tetramers and dimers to monomers. CTD, C-terminal domain; GSK, glycogen synthase kinase 3β; LRH, Leu-rich helix.

activating primer for GSK-3β, which phosphorylates the SR-rich region in an n-4 pattern. Based on the expected pattern of GSK-3 phosphorylation, with only Ser188 primed, we can expect a total of four phosphates

incorporated after GSK-3 activation at residues 188, 184, 180, and 176.

Phosphorylation of N in the SR-rich region changes the consistency of phase-separated droplets with RNA from dense

gel-like droplets to dynamic liquid-like droplets (19). To determine the molecular basis for this process, we sought to identify the effect of phosphorylation on the structure of linker alone and in context of the CTD. The SR-rich region preceding the self-associating LRH is a stretch of 40 residues containing 7 arginine and 14 serine residues that are targets for phosphorylation (45). We used genetic code expansion to incorporate a single phosphoserine at known priming position 188 (pSer188) (46). Spectra of ^{15}N-HSQC of pSer188 $N_{175-245}$ and WT $N_{175-245}$ are essentially identical, except for the resonances corresponding to residues proximal to the phosphorylation site (Fig. 2). Comparison of ^{15}N R_1, ^{15}N R_2 and $\{^1H\}$-^{15}N NOE revealed similar dynamic properties, with a modest increase in $\{^1H\}$-^{15}N NOE near the phosphorylation site and a modest decrease in R_2 in the alpha helical region. Since phosphorylation typically forms a hydrogen bond between the phosphate and the backbone amide causing the strong downfield shift in the amide proton resonance for pSer188 and is expected to dampen fast motion, it is not surprising to see a modest impact on the $\{^1H\}$-^{15}N NOE in the region proximal to the site of phosphorylation. The chemical shift changes for residues 186, 187, 189, and 190 suggest some minor structural changes localized to the vicinity of S188.

Even though only a modest change in chemical shifts is observed in pSer188 $N_{175-245}$ spectra, a major increase is observed in the peak intensities of the self-associating LRH that is also reflected in the decrease in R_2. NMR spectra collected at decreasing concentrations like those of the WT $N_{175-245}$ showed higher intensities for the helical resonances implicated in self-association, compared to WT $N_{175-245}$ at similar concentration (Fig. 3). SV-AUC analysis of pSer188-$N_{175-245}$ GFP also indicated weakened self-association (Fig. 3). Concentrations of 200 μM and 100 μM of pSer188-$N_{175-245}$ GFP showed a single peak in the c(s) plots, with sedimentation coefficients of ∼3.3 and ∼3.1S, respectively, indicating a weaker self-association when compared to WT $N_{175-245}$ which showed two peaks. SE-AUC of pSer188 $N_{175-245}$ GFP determined a Kd for dimer dissociation of 235 μM, significantly weaker than the 77 μM of WT $N_{175-245}$ (Fig. 3).

pSer188 $N_{175-365}$ displayed similar behavior to WT in ^{15}N HSQC spectra at decreasing concentrations (Fig. S3). As with WT, resonances corresponding to residues 216 to 232 were not detectable at concentrations as low as 10 μM (Fig. S4), indicating that a single phosphoserine incorporation is not sufficient to dissociate the LRH in the context of the CTD. We then tested if hyperphosphorylation would dissociate the LRH. Addition of GSK-3 (+GSK $N_{175-365}$) which phosphorylates serine or threonine residues N-terminal of an initial site of phosphorylation (42). While GSK-3 will canonically result in four phosphoserines, recent MS analysis of the SR-rich linker primed at S188 and subsequently reacted with GSK-3 showed partial phosphorylation up to six sites (47). These results are consistent with our observed NMR spectrum of GSK-3–treated $N_{175-365}$, which showed three strong phosphoserine resonances and additional weaker resonances corresponding to lower abundance phosphorylation of additional serines. ^{15}N

HSQC experiments at decreasing concentrations revealed that resonances corresponding to residues 216 to 232 re-appeared at concentrations as high as 100 μM, indicating that hyper-phosphorylation prevents LRH self-association in context of the CTD (Fig. 4). The dissociation of the tetramer is confirmed by SV-AUC experiments, in which there is a single broad peak centering around ∼2.7 S (which corresponds to dimeric +GSK $N_{175-365}$) at 340 μM and 200 μM.

Phosphorylation decreases RNA binding in the linker region

RNA binding to N is primarily driven by interactions with the NTD and, to a lesser extent, the CTD (24, 28). The central linker region shows direct interactions with RNA in the presence of the NTD (30). Here, we confirm using EMSA that $N_{175-245}$ corresponding to the central linker alone can directly interact with the first 1000 nts of the 5'-end of SARS-Cov-2 viral RNA (g(1–1000)) (Fig. 5). Increasing concentrations of $N_{175-245}$ caused an increasing shift of the g(1–1000) to a higher molecular weight, indicating that the protein is binding the RNA and reducing its mobility in the gel in a concentration-dependent manner.

We utilized NMR to probe the RNA interaction sites within $N_{175-245}$. Addition of increasing concentrations of g(1–1000) RNA to ^{15}N-labeled WT $N_{175-245}$ caused most of the resonances to disappear, consistent with the formation of high molecular weight complexes. Resonances that are still observed at a 1000:1 protein:RNA molar ratio are localized to the C-terminal end of the WT $N_{175-245}$ containing the residual tobacco etch virus (TEV) cleavage site, indicating these residues are not significantly interacting with g(1–1000) (Fig. 5). The significant line broadening of the remaining resonances suggests that the SR-rich region and the LRH are interacting with RNA, with only modest differences between the regions. This observation is consistent with a previous report that shows peak attenuation in this region when using a small piece of RNA (48).

We next determined if phosphorylation altered RNA binding using pSer188 $N_{175-245}$. We found that the pSer188 $N_{175-245}$ also bound RNA, based on EMSA analysis, but with weaker affinity (Fig. 5). Weaker binding was also demonstrated by NMR titrations of ^{15}N-labeled pSer188 $N_{175-245}$, with g(1–1000) RNA showing less peak attenuation at similar concentrations of RNA compared to the WT $N_{175-245}$ (Fig. S5). Residues 175 to 200 also show less peak attenuation, suggesting that the double negative charge of pSer188 is sufficient to counter the favorable electrostatic interactions between the RNA and the arginine-rich linker. In contrast to the WT, the LRH resonances are more attenuated in the pSer188 $N_{175-245}$ spectra with RNA, indicating that the RNA could still weakly interact with the helical region or that the RNA is altering the monomer-dimer exchange leading to additional broadening in the helical region.

Having observed a modest impact on RNA binding from a single phosphorylation, we next examined the impact of hyper-phosphorylation on RNA binding using GSK-3. Our EMSA results on GSK-3–hyperphosphorylated pSer188 $N_{175-245}$
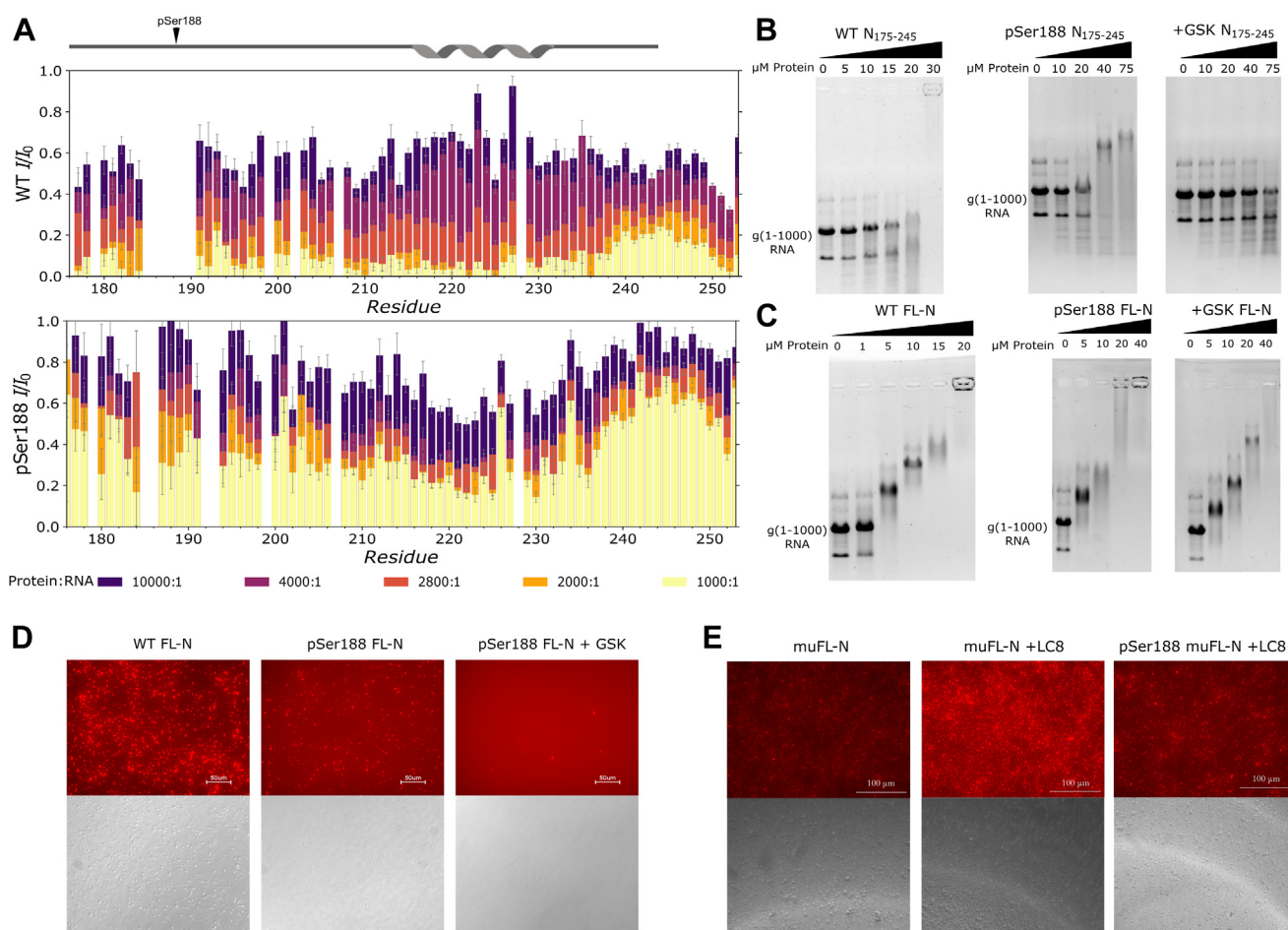
Figure 5. Phosphorylation modulation of N-RNA interactions. *A*, peak intensity ratios based on NMR titrations of g(1–1000) RNA into ^{15}N-labeled N$_{175–245}$ (*top*) and pSer188 N$_{175–245}$ (*bottom*). The protein:RNA ratio varied from 10000:1 to 1000:1. Data are plotted as intensity ratios for peaks with and without RNA. *B*, EMSA gels for WT N$_{175–245}$ (0–30 μM) and pSer188 N$_{175–245}$ (0–75 μM) with g(1–1000) RNA (0.5 μM). GSK-3 was used to hyperphosphorylate (+GSK) the pSer188 N$_{175–245}$ (−GSK). *C*, EMSA gels for WT FL-N (0–20 μM) and pSer188 FL-N (0–40 μM) with g(1–1000) RNA (0.5 μM), along with GSK-3 hyperphosphorylation. For (*B*) and (*C*), the g(1–1000) RNA migrates as multiple bands because the RNA is not denatured and can adopt secondary structures. *D*, WT FL-N (*left*), pSer188 FL-N (*middle*), and hyperphosphorylated FL-N (+GSK FL-N) (*right*) liquid-liquid phase separation (LLPS) using fluorescence imaging (*top*) and bright field (*bottom*). The g(1–1000) RNA was labeled with cy3 for fluorescence imaging. Images were taken at 40× magnification, and the scale bar represents 50 μm. All LLPS experiments were collected at 37 °C after 2 h of incubation. *E*, LLPS utilizing the same conditions as (*D*) for muFL-N, muFL-N with LC8 added, and pSer188 muFL-N with LC8 added (from *left* to *right*, respectively). The scale bar represents 100 μm. FL-N, full length N protein; GSK, glycogen synthase kinase 3β; LC8, dynein light chain 8.

show full disruption of the interaction between N$_{175–245}$ and g(1–1000) RNA even at 75 μM protein.

Phosphorylation of FL-N attenuates LLPS but not RNA binding

Having found that phosphorylation attenuated binding of N$_{175–245}$ to RNA, we performed similar experiments on the FL-N. In contrast to what we observed with N$_{175–245}$, EMSA analysis showed that RNA-binding affinity to the FL-N was not affected by phosphorylation and hyperphosphorylation (Fig. 5) clearly indicating that linker phosphorylation at these four sites does not significantly alter g(1–1000) RNA binding by FL-N.

Phosphorylation has also been consistently shown to alter the phase-separation properties of FL-N (11, 19). We thus tested the effect of phosphorylation on LLPS behavior in the context of the full-length protein by comparing the ability of pSer188 and WT FL-N proteins to form droplets. WT FL-N phase separated at 4 μM protein with 50 nM g(1–1000) RNA at 37 °C (Fig. 5 and Table S1). The droplets formed with pSer188 FL-N protein were on average similar in size to WT

FL-N but much fewer in number. The difference became even more pronounced in the GSK-hyperphosphorylated protein which showed almost no phase separation under these conditions (Fig. 5), with only a handful of droplets formed, clearly demonstrating that phosphorylation in the SR-rich region inhibits LLPS. The size of droplets formed in WT compared to pSer188 FL-N are similar, but the number of droplets formed by hyperphosphorylated FL-N was ~70× less than the singly phosphorylated pSer188 FL-N and 180× less than the number of droplets observed for WT FL-N (Table S1).

A switchable self-association domain to dissect effect of phosphorylation on LLPS

To test the contribution of the self-associating LRH to LLPS, we engineered a variant of FL-N with a switchable self-association capability by replacing the LRH with a disordered 10 amino acid segment containing a TQT recognition motif

(KAIDAATQTE) specific for the dimeric hub protein LC8. LC8 is a dimerizing hub as it binds two disordered chains of partner proteins at the TQT motif (49). By replacing the LRH with a TQT motif of similar length, we have created a construct of FL-N (muFL-N) that will only self-associate in the linker in the presence of LC8 but is quite disordered without. We tested that LC8 indeed binds muFL-N (Fig. S6) and forms a tight complex. LLPS experiments on muFL-N showed less droplet formation with g(1–1000) RNA when compared to WT FL-N at similar conditions (Fig. 5), but upon addition of LC8, muFL-N showed extensive phase separation with g(1–1000) RNA (Fig. 5 and Table S2). Addition of LC8 to phosphorylated muFL-N (pSer188 muFL-N) increased LLPS but not to the same extent as in the absence of phosphorylation (Table S2 and Fig. 5), indicating that disrupting self-association in the linker is important but is not the only determinant of reduced LLPS. These data implicate the LRH as a primary modulator of LLPS as in both phosphorylated and unphosphorylated muFL-N LLPS experiments, LC8-induced dimerization promotes phase-separation behavior.

Discussion

The structure of N is generally depicted with two folded domains, NTD and CTD, that are dimerized by the CTD and linked by a central disordered linker containing an α-helical region, LRH (Fig. 1). At the N terminus of the disordered linker is a positively charged SR-rich region that can undergo hyperphosphorylation. Given that the SR region is hyper-phosphorylated in infected cells but unphosphorylated during viral assembly in infectious virions, it is significant to understand the molecular processes that drive this switch. Here, we show that the LRH-spanning residues 216 to 232 forms self-associated dimers and higher order aggregates in the unphosphorylated form, but its self-association is significantly weakened by phosphorylation and hyperphosphorylation at sites within the SR-rich region about 30 amino-acids distant from the LRH. Phosphorylation also attenuates RNA binding to the SR region, but while it does not significantly affect RNA binding to the full-length protein, it drastically alters its LLPS.

Higher order association in FL-N is driven by the LRH-forming dimer of dimers

Since the crystal structure of the CTD was solved for a dimer (21, 23), it has been recognized that the CTD alone is the dimerization domain of N. Importantly, the CTD is a stable dimer and does not form higher order oligomers at NMR concentrations (22, 24). Work in the literature based on AUC and molecular dynamics reported a higher order trimeric coiled-coil mediated by intermolecular interactions in the LRH (34). Also reported is a mutation in N at the beginning of the LRH, G215C, exhibiting substantially stronger self-association and shifting self-association to a tetrameric oligomeric state (33). Using sedimentation equilibrium AUC on a construct containing the LRH and the CTD, we indeed identified a higher order species in addition to the dimer at concentration above 100 μM (Fig. 4). We assign this oligomerization state to

a tetramer rather than a trimer for the following reasons: the LRH self-association is coupled to CTD dimerization and therefore the dimer is always the more stable species, and at higher concentration, it will populate a dimer of dimers or a tetramer 4-helix bundle that brings two dimers of N together. In this model, the LRH cannot dissociate in the presence of the CTD dimer but forms an intra dimer helix–helix interaction at low protein concentration, while the 4-helix bundle tetramer forms at high concentration. Multivalent binding of RNA to N will increase the effective local concentration shifting the equilibrium towards a tetramer or higher.

To deconvolute the contribution of the LRH from that of the CTD to formation of the higher order structure, we used NMR and AUC on constructs of the linker alone and the linker plus CTD and compared these results to the CTD alone. LRH in the linker construct has a weak self-association of 77 μM which is significantly enhanced to about 1 μM in context of the CTD and can also form weak tetramers. A construct corresponding to the CTD alone is a stable dimer at all concentrations. Therefore, these studies show clearly that the LRH is required to form higher order assemblies but forms only multiples of dimers.

Phosphorylation decreases LLPS formation by inhibiting LRH self-association

MD simulations suggested that phosphorylation of the SR region increases inter- and intra-peptide interactions through phosphate-arginine salt bridges (50). Phosphorylation of an SR region's synthetic peptide also exhibited reduced binding to polyU RNA as well as showed more liquid-like droplets (50). More recently, hyperphosphorylation was reported to promote direct binding of the phosphorylated SR region to the RNA-binding domain of the NTD and thus providing a mechanism for inhibiting RNA binding (51). The reported work, however, did not address the effect of phosphorylation on the LRH, as we do below.

Our NMR results show that phosphorylation at serine 188 did not dramatically alter the monomeric structure of the linker. Instead, phosphorylation resulted in a lower propensity of the linker to self-associate, with a Kd, three-fold weaker than WT. To place these results in context of the full-length protein, we determined the effect of phosphorylation in the presence of the dimeric CTD which significantly enhances dimerization of the LRH. A single phosphorylation event did not affect LRH self-association, but hyperphosphorylation with GSK-3, which places a pSer at 3 to 6 sites, caused dissociation of the tetrameric form suggesting that the level of phosphorylation can be tuned in response to the level of the protein concentration and self-association needed.

How phosphorylation in the SR region, about 30 amino-acid residues distant from the beginning of the LRH, promotes its dissociation is not entirely obvious. Comparison to other RNA-binding proteins with SR-rich regions gives some clues and suggests that phosphorylation could simply be increasing the solubility of the protein, preventing oligomerization (45). Phosphorylation significantly alters the net charge of a protein, adding a −2 charge with each phosphorylated residue at

physiological pH. In TDP-43, for example, phosphorylation was proposed as a preventative cellular mechanism against aggregation (52). TDP-43 hyperphosphorylation was also shown to reduce phase separation and aggregation. One attractive explanation is that serine residues are more prone to interact with other protein residues than with solvent, while phosphoserine would interact more with water molecules than be involved in protein–protein interactions (45). Increased interactions with water would still not explain our observation of phosphorylation promoting dissociation 30 residues away, however. It is likely that solvation would change the structure of the SR region from a more extended to flexible with local structures; a change supported by chemical shift differences and a slight increase in the heteronuclear NOEs in the phosphorylated form. Local structures could interfere with stacking of the LRH in a helical bundle, thus opposing higher order formation.

Phase separation and RNA binding

While SR-rich regions of proteins are often involved in RNA interactions (53, 54), in SARS-CoV-2, the NTD is the primary RNA-binding domain, but the CTD and the disordered domains are all capable of interacting with RNA with varying affinity and specificity. We show here direct interaction between the SR region and RNA, but its phosphorylation resulted in reduced RNA binding, while hyperphosphorylation by GSK-3, primed with pSer188, dramatically abolished it. Interestingly, in contrast to the isolated linker domain, WT, pSer188, and hyperphosphorylated FL-N all interact strongly with RNA as shown by EMSA (Fig. 5). These results demonstrate that while phosphorylation of the linker inhibits RNA binding at the SR-region, it does not inhibit overall RNA binding by FL-N presumably because phosphorylation would leave the high affinity NTD-binding site unaltered. A recent preprint reported a direct interaction of the phosphorylated region with the RNA site on the NTD, potentially blocking RNA binding and abolishing binding to FL-N (51). We see no effect on RNA binding with the GSK-phosphorylated FL-N. The discrepancies between these studies and ours could be explained as due to their use of a short RNA sequence while ours uses a 1000 nt viral RNA. The 1000 nt RNA may have some allosteric effects preventing the blocking of the NTD. Further, a larger RNA with multivalent sites will compete with any weak interactions between the SR-rich region and the RNA-binding region of the NTD.

Exciting about our findings is the effect of phosphorylation on phase separation of FL-N. Phosphorylation by a combination of CDK-1 and GSK-3 was reported to alter phase-separation behavior, inducing a transition from gel-like to liquid-like droplets (19). Our GSK-hyperphosphorylated FL-N shows significant reduction in both the size and number of droplets formed. Interesting to note is that while RNA interaction is certainly required for phase separation, phase-separation behavior can be inhibited by phosphorylation in the disordered linker region even while overall RNA binding remains intact.

A model for how hyperphosphorylation in the SR region attenuates phase separation

The model of Figure 6 summarizes our data and proposes a process of how phosphorylation switches the role of N from viral packaging to viral replication. The N protein is in a dimer-tetramer equilibrium that shifts towards tetramer at high protein concentration as shown here or forms higher order oligomers when bound multivalently to RNA (55). It is well established that viral packaging would involve genome compaction through multivalent protein:RNA and protein:protein interactions (56). In the absence of phosphorylation, the LRH self-association organizes the SR regions in an elongated manner primed for higher order aggregation when bound to RNA. A single phosphorylation introduces repulsive charges that disrupt charge patterning which is normally associated with LLPS (57) preventing stacking of the SR regions and improving solubility thus reducing LLPS while keeping the self-association of LRH. The local effect of phosphorylation is confirmed by using the switchable dimer which shows that self-association is not enough to restore full LLPS. Hyperphosphorylation is required to amplify the disruption in the SR regions and significantly introduces bound water molecules to cause dissociation of the LRH about 30 amino acids downstream, disrupting LLPS.

The model of Figure 6 also explains how abolishing binding of the SR region to RNA reduces compaction while keeping the RNA bound in the full-length protein. Binding to the linker intercalates the RNA tightly in the nucleocapsid during packaging and this is reversed upon phosphorylation. Dissociation of the LRH will increase its accessibility to other binding partners during viral maturation such as the viral protein nsp3a (58) and the host protein 14-3-3 (59).

Concluding remarks

Phosphorylation in SR-rich proteins has been observed to increase solubility or disrupt alpha helical regions inhibiting their ability to self-assemble in fibrils (60). In TDP-43, phosphorylation of few key residues disrupts the helical structure and regulates condensate formation (61). Interesting to point out here, however, that in these examples, the phosphorylation effect is spread locally. The effect of phosphorylation of the N protein and its binding to RNA at multivalent sites is significantly more complicated. Our model presents an explanation of how phosphorylation in a disordered linker in addition to causing changes in the local environment can also act on the self-association of a distant helix that is easily responsive to long range site-specific changes, contributing to switching from genome replication to virus maturation.

Experimental procedures
Plasmid construction

The $N_{175-245}$ plasmid construct was prepared by inserting DNA encoding amino acid residues 175 to 245 into a pRBC SUMO vector with a N-terminal bdSUMO tag, a C-terminal TEV protease cleavage site linked to GFP, and a hexahistidine tag (62). $N_{175-365}$ plasmid construct was prepared by inserting
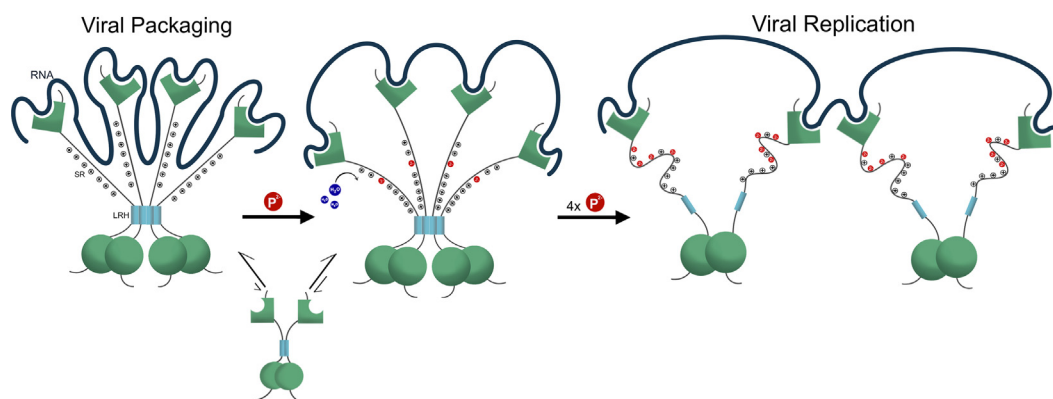
Figure 6. Proposed model of how phosphorylation acts as a molecular switch for the role of N from viral packaging to viral replication. FL-N is depicted here using similar illustration as in Figure 1C. (*Left*) Unphosphorylated FL-N is in a dimer-tetramer equilibrium which shifts to tetramer when bound to RNA. The positive charges in the SR-rich region cause elongation due to charge–charge repulsion. RNA intercalates by binding to the SR region of the linker, resulting in a compacted RNA expected to be most populated in LLPS and in viral packaging. (*Middle*) A single phosphorylation event (*red*) introduces negative charges in the SR causing some structural change. RNA remains bound to the NTD but does not bind the linker resulting in reduced compaction. (*Right*) Four phosphorylation events with GSK-3 causing significant structural changes in the SR-region of GSK hyperphosphorylated FL-N and dissociation of the 4-helix bundle tetramers. The resulting structure of FL-N is a dimer with significant flexibility in the linker after dissociation of the LRH causing significant dispersing of the RNA, consistent with the model expected in viral replication. Two N proteins are shown to illustrate multivalent binding. Phosphates are indicated by *red circles*, and arginines by circled pluses. RNA is depicted by the *dark blue line*. FL-N, full length N protein; GSK-3, glycogen synthase kinase 3β; NTD, N-terminal domain; SR-rich, serine/arginine-rich.

DNA encoding amino acid residues 175 to 365 into a pRBC vector with a C-terminal TEV protease cleavage site linked to a hexahistidine tag. pSer188 N$_{175-245}$ plasmid was generated as previously described (46). pSer188 N$_{175-365}$ was prepared from the N$_{175-365}$ plasmid by mutating S188 to an amber stop codon (TAG). pSer188- FL-N plasmid was generated by cloning the FL-N sequence into the pRBC vector, then mutating S188 to a TAG for expression with genetic code expansion. muFL-N and pSer188 muFL-N plasmids were prepared from the FL-N and pSer188-FL-N plasmids respectively, using site-directed mutagenesis to remove the DNA-encoding residues 216 to 232 and replace them with DNA encoding the sequence KAIDAATQTE, a sequence known to bind LC8 strongly (63).

Protein expression and purification

To prepare WT N$_{175-245}$ or WT N$_{175-365}$ proteins, BL21(DE3) *E. coli* transformed with the appropriate plasmid were grown in LB-rich media to an OD of 0.6 to 0.8, then induced with 1 mM IPTG at 37 °C for 6 h. Stable isotope-labeled samples of WT N$_{175-245}$ or WT N$_{175-365}$ were expressed and induced using the same procedure as natural abundance, expect that the cells were grown in MJ9 minimal media supplemented with ^{15}N ammonium chloride and ^{13}C glucose as the sole nitrogen or carbon source as appropriate. Following induction, cells were harvested by centrifugation and either used immediately or stored at −80 °C.

WT N$_{175-245}$ and WT N$_{175-365}$ were purified using the TALON His-tag purification protocol (Clontech Laboratories). Cell pellets were lysed in 50 mM tris, 500 mM NaCl, 5 mM imidazole, 1 mM NaN$_3$, pH 7.5 (high salt buffer) by sonication and centrifuged at 27,200 relative centrifugal force to remove cell debris. The supernatant was mixed with resin for 1 h and washed with 20 column volumes of high salt buffer followed by

four column volumes 50 mM tris, 150 mM NaCl, 5 mM imidazole, 1 mM NaN$_3$, pH 7.5 (low salt buffer). For WT N$_{175-245}$, the SUMO solubility tag was cleaved with 200 nM SENP1 protease (46, 64, 65) (∼1:250 protease:protein) for 1 h at 4 °C on the resin. Proteins were eluted off the resin with high salt buffer supplemented with 300 mM imidazole. Proteins were exchanged into buffer containing 50 mM Tris, 300 mM NaCl, 5 mM imidazole, pH 7.5 using a PD-10 desalting column (Cytiva). The tag was cleaved by incubating overnight with TEV protease (66) (∼1:20 protease:protein) at 4 °C and then removed by reverse affinity purification with Talon His-Tag resin. Proteins were concentrated and exchanged into 50 mM sodium phosphate, 150 mM NaCl, pH 6.5 (NMR buffer).

FL-N and muFL-N were prepared using a modified procedure based on one previously described for FL-N (22). The plasmid was transformed into *E. coli* Rosetta (DE3) cells, cultured in 2xYT media to an OD of 0.6, and induced with 1 mM IPTG at 18 °C overnight. Cells were harvested by centrifugation then resuspended in 50 mM sodium phosphate, 1M NaCl, 5 mM imidazole, 1 mM NaN$_3$, 1 mg/ml lysozyme, pH 8.0 and incubated for 1 h at 4 °C. Cells were sonicated 3× 2 min and centrifuged at 27,200 relative centrifugal force for 45 min. The clarified lysate was mixed with Talon His-Tag resin, incubated for 1 h at 4 °C, and then washed with 20 column volumes of 50 mM sodium phosphate, 3M NaCl, 10 mM imidazole, 1 mM NaN$_3$, pH 8.0 to remove nonspecifically bound RNA. The protein was eluted with 50 mM sodium phosphate, 300 mM NaCl, 350 mM imidazole, 1 mM NaN$_3$, pH 8.0 FL-N, then concentrated, and further purified on a Superdex 200 column in 50 mM sodium phosphate, 150 mM NaCl, pH 7.5 buffer. The protein was concentrated and either used immediately or stored flash frozen at −80 °C.

To prepare pSer188 N$_{175-245}$ and pSer N$_{175-365}$, BL21(DE3) ΔSerB *E. coli* cells were transformed simultaneously with the

appropriate expression plasmid and pKW2-EFsep. For natural abundance samples, cells were grown in 2xYT to an OD of 0.6 to 0.8 and induced with 1 mM IPTG for 48 h at 18 °C. Stable isotope-labeled pSer188 $N_{175-245}$ and pSer $N_{175-365}$ were expressed and purified as previously described (46). Briefly, cells were grown to an OD 0.6 to 0.8 in MJ9 minimal media supplemented with ^{15}N Celtone (0.2%) and then induced with 1 mM IPTG for 48 h at 18 °C. Cells were harvested by centrifugation and stored at –80 °C. Protein was purified using the same method as used for the WT samples above, except that buffers were supplemented with phosphatase inhibitors (10 mM NaF, 2.5 mM sodium pyrophosphate, and 1 mM orthovanadate).

To prepare pSer188 FL-N and pSer188 muFL-N proteins, the appropriate expression plasmid was transformed simultaneously with pKW2-EFsep into BL21(DE3) ΔSerB *E. coli* cells. Cells were grown in rich-Auto-inducing Media (37) at 37 °C until OD 600 reached 1.3 and then at 18 °C for 48 h. Cells were harvested and the protein was purified using the same methods as FL-N, except that the buffers were supplemented with phosphatase inhibitors (10 mM NaF, 2.5 mM sodium pyrophosphate, and 1 mM orthovanadate).

For hyperphosphorylated protein, 80 μM of target protein (pSer188 linker, pSer188 $N_{175-365}$, FL-N pSer188, or muFL-N pSer188) was incubated with 80 nM of GSK-3 in a buffer containing 20 mM Tris pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, and 1 mM ATP, at 37 °C for 20 h. Hyperphosphorylated proteins were further purified by size-exclusion chromatography using a Superdex S75 column. Phos-Tag SDS-PAGE gels were used to confirm the phosphorylation status of all phosphorylated samples.

LC8 expression and purification was performed as previously described (67). The GSK-3 protein was expressed and purified as previously described (46). Table S3 lists the final buffer conditions for each protein and experiment.

NMR spectroscopy

NMR experiments were performed using an 800 MHz Bruker Avance III HD NMR spectrometer equipped with a triple resonance (TCI) cryogenic probe. Backbone resonance assignments were made using a suite of BEST triple resonance experiments, including HNCO, HNCACB, HNCACO, and HNCOCACB (68). All NMR data were processed (apodized, zero filled, Fourier Transformed, and phased) using nmrPipe (69) and analyzed in nmrviewJ (70). Three dimensional experiments were collected using nonuniform sampling, and nonuniform sampling artifacts were suppressed using SMILE (71). Resonance assignments were deposited in the BMRB under ascension number 51904. ^{15}N nuclear spin relaxation parameters were measured using Bruker temperature compensated pulse sequences, with eight unique delay times. The 60 ms delay was collected in triplicate to aid with error estimation. Peak intensities were fit in nmrviewJ to an exponential decay model, with Monte Carlo–based error estimation. For the {^1H}-^{15}N NOE, the D1 delay was increased to 8 s to ensure complete relaxation between scans. NMR titrations

of $N_{175-245}$ with RNA were performed using 2D ^{15}N-BEST-TROSY experiments. Peak intensities were measured in nmrviewJ and normalized to the corresponding peak in the spectrum without RNA. Secondary chemical shifts were calculated using sequence, temperature, and pH-corrected chemical shifts (72). NMR concentration titrations on $N_{175-365}$ were performed using 2D ^{15}N TROSY HSQC experiments. All experiments were conducted at 10 °C.

Electrophoretic mobility shift assay

The first 1000 nucleotides from the viral genome (g(1–1000) RNA) at a final concentration of 0.5 μM was incubated with increasing concentrations of protein (range of 0–75 μM, in 20 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.5) at room temperature for 20 min. The total reaction volume was 10 μl. After incubation, 2 μl of 6× loading dye was added to the reaction before loading on a 1% agarose gel. The gel was run at 150 V for 1 h. RNA bands were stained with Midori Green Nucleic Acid staining solution (Bulldog Bio. Inc) and visualized using a Bio-Rad Gel Doc Image system.

Analytical ultracentrifugation

Analytical ultracentrifugation was performed using a Beckman Coulter Optima XL-A analytical ultracentrifuge equipped with absorbance optics. SV-AUC experiments used either $N_{175-245}$ protein tagged with GFP because $N_{175-245}$ by itself does not have a strong UV absorbance or a construct of $N_{175-365}$. All SV-AUC samples were run in standard 2-channel sectored cells using an An60-Ti rotor at 20 °C. The concentration of each protein was varied from 50 to 200 μM. Samples were spun at 42,000 rpm and with 300 scans per sample. Data were fit to the continuous c(s) model using SEDFIT (73). Buffer density and viscosity were calculated using SEDNTERP (74). For sedimentation equilibrium analysis, three concentrations of each sample were loaded into 6-well cells. Samples were centrifuged at 10,000, 13,000, and 18,000 rpm in an An60-Ti rotor for 30 to 36 h at each rotor speed. For GFP $N_{175-245}$, data were fit to either a single ideal species or monomer-dimer equilibrium using the software Hetero-analysis (https://colelab.uconn.edu/) (75). For $N_{175-365}$, data were fit to a dimer-tetramer model.

LLPS and microscopy

Fluorescence microscopy images were taken on a Keyence BZ-X700/BZ-X710 microscope with a 40× objective lens and a 384-well plate (Cellvis P384-1.5H-N); images were processed using BZ-x viewer and BZ-x analyzer software (https://www.keyence.com/landing/microscope/lp_fluorescence.jsp). For this experiment, cy3-labeled RNA was diluted into nuclease-free water to reach a final concentration of 50 nM when added to the protein sample. Stocks of unlabeled protein were prepared by diluting into 20 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.5 droplet buffer. For FL-N constructs, samples were prepared at a total concentration of 4 μM. For linker constructs, samples were prepared at a total concentration of 20 μM. Protein samples were prepared by combining 27 μl of

protein stock of the appropriate concentration with 3 µl of cy3-labeled g(1–1000) RNA for a total sample volume of 30 µl. For muFL-N samples with LC8 added, 3.6 µM LC8 was combined with muFL-N and cy3-labeled g(1–1000). The samples were incubated at 37 °C for 2 h and subsequent imaging was taken. Foci were counted using ImageJ software (https://imagej.net/ij/). Average foci diameter was calculated in ImageJ from a random sampling of 10 foci from each image.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the supporting information.

Abbreviations—The abbreviations used are: CTD, C-terminal domain; gRNA, genomic RNA; FL-N, Full length N protein; GSK-3, glycogen snthase kinase 3β; LC8, Dynein light chain 8; LLPS, liquid-liquid phase separation; LRH, Leu-rich helix; NTD, N-terminal domain; SE-AUC, sedimentation equilibrium analytical ultracentrifugation; SR-rich, serine/arginine-rich; SV-AUC, sedimentation velocity analytical ultracentrifugation; TEV, tobacco etch virus.

References

1. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Chen, H. D., *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273
2. McBride, R., van Zyl, M., and Fielding, B. C. (2014) The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991–3018
3. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., *et al.* (2020) Molecular architecture of the SARS-CoV-2 virus. *Cell* **183**, 730–738.e13
4. Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., *et al.* (2021) The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.* **12**, 3917
5. Cai, T., Yu, Z., Wang, Z., Liang, C., and Richard, S. (2021) Arginine methylation of SARS-Cov-2 nucleocapsid protein regulates RNA binding, its ability to suppress stress granule formation, and viral replication. *J. Biol. Chem.* **297**, 100821
6. Xu, W., Pei, G., Liu, H., Ju, X., Wang, J., Ding, Q., *et al.* (2021) Compartmentalization-aided interaction screening reveals extensive high-order complexes within the SARS-CoV-2 proteome. *Cell Rep.* **36**, 109482
7. Cong, Y., Ulasli, M., Schepers, H., Mauthe, M., V'kovski, P., Kriegenburg, F., *et al.* (2020) Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle. *J. Virol.* **94**, e01925-19
8. Luo, L., Li, Z., Zhao, T., Ju, X., Ma, P., Jin, B., *et al.* (2021) SARS-CoV-2 nucleocapsid protein phase separates with G3BPs to disassemble stress granules and facilitate viral production. *Sci. Bull.* **66**, 1194–1204
9. Yang, Z., Johnson, B. A., Meliopoulos, V. A., Ju, X., Zhang, P., Hughes, M. P., *et al.* (2024) Interaction between host G3BP and viral nucleocapsid protein regulates SARS-CoV-2 replication and pathogenicity. *Cell Rep.* **43**, 113965
10. Chen, H., Cui, Y., Han, X., Hu, W., Sun, M., Zhang, Y., *et al.* (2020) Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res.* **30**, 1143–1145
11. Lu, S., Ye, Q., Singh, D., Cao, Y., Diedrich, J. K., Yates, J. R., 3rd, *et al.* (2021) The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.* **12**, 502
12. Jack, A., Ferro, L. S., Trnka, M. J., Wehri, E., Nadgir, A., Nguyenla, X., *et al.* (2021) SARS-CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. *PLoS Biol.* **19**, e3001425
13. Zachrdla, M., Savastano, A., Ibáñez de Opakua, A., Cima-Omori, M. S., and Zweckstetter, M. (2022) Contributions of the N-terminal intrinsically disordered region of the severe acute respiratory syndrome coronavirus 2 nucleocapsid protein to RNA-induced phase separation. *Protein Sci.* **31**, e4409
14. Roden, C. A., Dai, Y., Giannetti, C. A., Seim, I., Lee, M., Sealfon, R., *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
15. Perdikari, T. M., Murthy, A. C., Ryan, V. H., Watters, S., Naik, M. T., and Fawzi, N. L. (2020) SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J.* **39**, e106478
16. Dong, H., Zhang, H., Jalin, J., He, Z., Wang, R., Huang, L., *et al.* (2023) Nucleocapsid proteins from human coronaviruses possess phase separation capabilities and promote FUS pathological aggregation. *Protein Sci.* **32**, e4826
17. Tayeb-Fligelman, E., Bowler, J. T., Tai, C. E., Sawaya, M. R., Jiang, Y. X., Garcia, G., Jr., *et al.* (2023) Low complexity domains of the nucleocapsid protein of SARS-CoV-2 form amyloid fibrils. *Nat. Commun.* **14**, 2379
18. 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
19. Carlson, C. R., Asfaha, J. B., Ghent, C. M., Howard, C. J., Hartooni, N., Safari, M., *et al.* (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
20. Chen, Y., Lei, X., Jiang, Z., Humphries, F., Parsi, K. M., Mustone, N. J., *et al.* (2023) Cellular nucleic acid-binding protein restricts SARS-CoV-2 by regulating interferon and disrupting RNA-protein condensates. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2308355120
21. Peng, Y., Du, N., Lei, Y., Dorje, S., Qi, J., Luo, T., *et al.* (2020) Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *EMBO J.* **39**, e105938
22. Forsythe, H. M., Rodriguez Galvan, J., Yu, Z., Pinckney, S., Reardon, P., Cooley, R. B., *et al.* (2021) Multivalent binding of the partially disordered SARS-CoV-2 nucleocapsid phosphoprotein dimer to RNA. *Biophys. J.* **120**, 2890–2901
23. Zhou, R., Zeng, R., von Brunn, A., and Lei, J. (2020) Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Mol. Biomed.* **1**, 2

ASBMB

24. Estelle, A. B., Forsythe, H. M., Yu, Z., Hughes, K., Lasher, B., Allen, P., *et al.* (2023) RNA structure and multiple weak interactions balance the interplay between RNA binding and phase separation of SARS-CoV-2 nucleocapsid. *PNAS Nexus* **2**, pgad333

25. 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

26. Caruso, I. P., Dos Santos Almeida, V., do Amaral, M. J., de Andrade, G. C., de Araujo, G. R., de Araújo, T. S., *et al.* (2022) Insights into the specificity for the interaction of the promiscuous SARS-CoV-2 nucleocapsid protein N-terminal domain with deoxyribonucleic acids. *Int. J. Biol. Macromol.* **203**, 466–480

27. Redzic, J. S., Lee, E., Born, A., Issaian, A., Henen, M. A., Nichols, P. J., *et al.* (2021) The inherent dynamics and interaction sites of the SARS-CoV-2 nucleocapsid N-terminal region. *J. Mol. Biol.* **433**, 167108

28. Dinesh, D. C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., *et al.* (2020) Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* **16**, e1009100

29. Cubuk, J., Alston, J. J., Incicco, J. J., Holehouse, A. S., Hall, K. B., Stuchell-Brereton, M. D., *et al.* (2024) The disordered N-terminal tail of SARS-CoV-2 Nucleocapsid protein forms a dynamic complex with RNA. *Nucleic Acids Res.* **52**, 2609–2624

30. 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

31. Ribeiro-Filho, H. V., Jara, G. E., Batista, F. A. H., Schleder, G. R., Costa Tonoli, C. C., Soprano, A. S., *et al.* (2022) Structural dynamics of SARS-CoV-2 nucleocapsid protein induced by RNA binding. *PLoS Comput. Biol.* **18**, e1010121

32. Ye, Q., West, A. M. V., Silletti, S., and Corbett, K. D. (2020) Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* **29**, 1890–1901

33. Zhao, H., Nguyen, A., Wu, D., Li, Y., Hassan, S. A., Chen, J., *et al.* (2022) Plasticity in structure and assembly of SARS-CoV-2 nucleocapsid protein. *PNAS Nexus* **1**, pgac049

34. Zhao, H., Wu, D., Hassan, S. A., Nguyen, A., Chen, J., Piszczek, G., *et al.* (2023) A conserved oligomerization domain in the disordered linker of coronavirus nucleocapsid proteins. *Sci. Adv.* **9**, eadg6473

35. Zhao, H., Wu, D., Nguyen, A., Li, Y., Adao, R. C., Valkov, E., *et al.* (2021) Energetic and structural features of SARS-CoV-2 N-protein co-assemblies with nucleic acids. *iScience* **24**, 102523

36. Bouhaddou, M., Memon, D., Meyer, B., White, K. M., Rezelj, V. V., Correa Marrero, M., *et al.* (2020) The global phosphorylation landscape of SARS-CoV-2 infection. *Cell* **182**, 685–712.e19

37. Cheng, N., Liu, M., Li, W., Sun, B., Liu, D., Wang, G., *et al.* (2022) Protein post-translational modification in SARS-CoV-2 and host interaction. *Front. Immunol.* **13**, 1068449

38. Carlson, C. R., Adly, A. N., Bi, M., Howard, C. J., Frost, A., Cheng, Y., *et al.* (2022) Reconstitution of the SARS-CoV-2 ribonucleosome provides insights into genomic RNA packaging and regulation by phosphorylation. *J. Biol. Chem.* **298**, 102560

39. Liu, X., Verma, A., Garcia, G., Jr., Ramage, H., Lucas, A., Myers, R. L., *et al.* (2021) Targeting the coronavirus nucleocapsid protein through GSK-3 inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2113401118

40. Peng, T. Y., Lee, K. R., and Tarn, W. Y. (2008) Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *FEBS J.* **275**, 4152–4163

41. Adly, A. N., Bi, M., Carlson, C. R., Syed, A. M., Ciling, A., Doudna, J. A., *et al.* (2023) Assembly of SARS-CoV-2 ribonucleosomes by truncated N(*) variant of the nucleocapsid protein. *J. Biol. Chem.* **299**, 105362

42. Yaron, T. M., Heaton, B. E., Levy, T. M., Johnson, J. L., Jordan, T. X., Cohen, B. M., *et al.* (2022) Host protein kinases required for SARS-CoV-2 nucleocapsid phosphorylation and viral replication. *Sci. Signal.* **15**, eabm0808

43. Guseva, S., Perez, L. M., Camacho-Zarco, A., Bessa, L. M., Salvi, N., Malki, A., *et al.* (2021) (1)H, (13)C and (15)N Backbone chemical shift assignments of the n-terminal and central intrinsically disordered domains of SARS-CoV-2 nucleoprotein. *Biomol. NMR Assign.* **15**, 255–260

44. Wu, C. H., Yeh, S. H., Tsay, Y. G., Shieh, Y. H., Kao, C. L., Chen, Y. S., *et al.* (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

45. Kundinger, S. R., Dammer, E. B., Yin, L., Hurst, C., Shapley, S., Ping, L., *et al.* (2021) Phosphorylation regulates arginine-rich RNA-binding protein solubility and oligomerization. *J. Biol. Chem.* **297**, 101306

46. Vesely, C. H., Reardon, P. N., Yu, Z., Barbar, E., Mehl, R. A., and Cooley, R. B. (2022) Accessing isotopically labeled proteins containing genetically encoded phosphoserine for NMR with optimized expression conditions. *J. Biol. Chem.* **298**, 102613

47. Franklin, R. (2023) *Revealing sequences and modifications of intact proteins using electron fragmentation*. Doctor of Philosophy (Ph.D.) Doctoral, Oregon State University, Corvallis, OR

48. Schiavina, M., Pontoriero, L., Uversky, V. N., Felli, I. C., and Pierattelli, R. (2021) The highly flexible disordered regions of the SARS-CoV-2 nucleocapsid N protein within the 1-248 residue construct: sequence-specific resonance assignments through NMR. *Biomol. NMR Assign.* **15**, 219–227

49. Barbar, E. (2008) Dynein light chain LC8 is a dimerization hub essential in diverse protein networks. *Biochemistry* **47**, 503–508

50. 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

51. [preprint] Botova, M., Camacho-Zarco, A. R., Tognetti, J., Bessa, L. M., Guseva, S., Mikkola, E., *et al.* (2024) A specific phosphorylation-dependent conformational switch of SARS-CoV-2 nucleoprotein inhibits RNA binding. *bioRxiv*. https://doi.org/10.1101/2024.02.22.579423

52. Chiang, W.-C., Fang, Y.-S., Lye, Y. S., Weng, T.-Y., Ganesan, K., Huang, S. H., *et al.* (2022) Hyperphosphorylation-mimetic TDP-43 drives amyloid formation and possesses neuronal toxicity at the oligomeric stage. *ACS Chem. Neurosci.* **13**, 2599–2612

53. Luo, H., Ye, F., Chen, K., Shen, X., and Jiang, H. (2005) SR-rich motif plays a pivotal role in recombinant SARS coronavirus nucleocapsid protein multimerization. *Biochemistry* **44**, 15351–15358

54. Nikolakaki, E., and Giannakouros, T. (2020) SR/RS motifs as critical determinants of coronavirus life cycle. *Front Mol. Biosci.* **7**, 219

55. 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

56. Cubuk, J., Alston, J. J., Incicco, J. J., Singh, S., Stuchell-Brereton, M. D., Ward, M. D., *et al.* (2021) The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **12**, 1936

57. Prasad, A., Bharathi, V., Sivalingam, V., Girdhar, A., and Patel, B. K. (2019) Molecular mechanisms of TDP-43 misfolding and pathology in amyotrophic lateral sclerosis. *Front Mol. Neurosci.* **12**, 25

58. Bessa, L. M., Guseva, S., Camacho-Zarco, A. R., Salvi, N., Maurin, D., Perez, L. M., *et al.* (2022) The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Sci. Adv.* **8**, eabm4034

59. Tugaeva, K. V., Hawkins, D., Smith, J. L. R., Bayfield, O. W., Ker, D. S., Sysoev, A. A., *et al.* (2021) The mechanism of SARS-CoV-2 nucleocapsid protein recognition by the human 14-3-3 proteins. *J. Mol. Biol.* **433**, 166875

60. Haider, R., Penumutchu, S., Boyko, S., and Surewicz, W. K. (2024) Phosphomimetic substitutions in TDP-43's transiently α-helical region suppress phase separation. *Biophys. J.* **123**, 361–373

61. Li, H.-R., Chiang, W.-C., Chou, P.-C., Wang, W.-J., and Huang, J.-R. (2018) TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. *J. Biol. Chem.* **293**, 6090–6098

62. Zhu, P., Stanisheuski, S., Franklin, R., Vogel, A., Vesely, C. H., Reardon, P., *et al.* (2023) Autonomous synthesis of functional, permanently phosphorylated proteins for defining the interactome of monomeric 14-3-3ζ. *ACS Cent. Sci.* **9**, 816–835

63. Jespersen, N., Estelle, A., Waugh, N., Davey, N. E., Blikstad, C., Ammon, Y. C., *et al.* (2019) Systematic identification of recognition motifs for the hub protein LC8. *Life Sci. Alliance* **2**, e201900366

64. Zhu, P., Gafken, P. R., Mehl, R. A., and Cooley, R. B. (2019) A highly versatile expression system for the production of multiply phosphorylated proteins. *ACS Chem. Biol.* **14**, 1564–1572

65. Frey, S., and Görlich, D. (2014) A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *J. Chromatogr. A* **1337**, 95–105

66. Young, C. L., Britton, Z. T., and Robinson, A. S. (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* **7**, 620–634

67. Jespersen, N. E., Leyrat, C., Gerard, F. C., Bourhis, J. M., Blondel, D., Jamin, M., *et al.* (2019) The LC8-RavP ensemble structure evinces A role for LC8 in regulating lyssavirus polymerase functionality. *J. Mol. Biol.* **431**, 4959–4977

68. Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. *J. Am. Chem. Soc.* **128**, 9042–9043

69. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293

70. Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352

71. Ying, J., Delaglio, F., Torchia, D. A., and Bax, A. (2017) Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR* **68**, 101–118

72. Kjaergaard, M., and Poulsen, F. M. (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **50**, 157–165

73. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619

74. Philo, J. S. (2023) SEDNTERP: a calculation and database utility to aid interpretation of analytical ultracentrifugation and light scattering data. *Eur. Biophys. J.* **52**, 233–266

75. Cole, J. L. (2004) Analysis of heterogeneous interactions. *Methods Enzymol.* **384**, 212–232